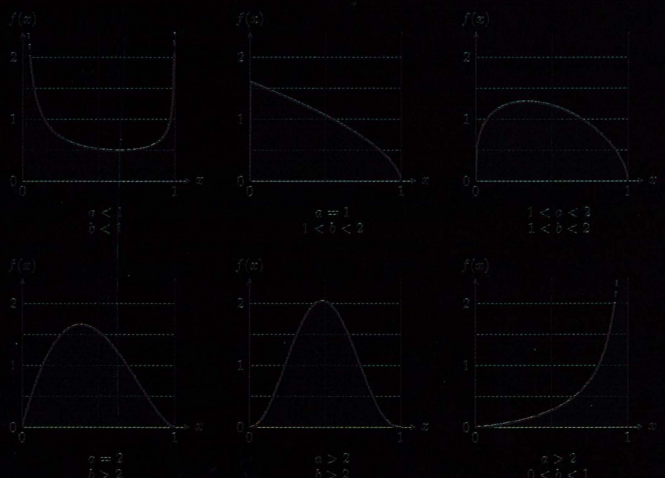
Walter Appel

Probabilités

pour les non-probabilistes

2e édition





H&K

PROBABILITÉS POUR LES NON-PROBABILISTES

Remerciements

La relecture d'un manuscrit est un moment délicat, et les nombreux collègues et amis qui ont accepté d'en relire des parties ont, par leurs critiques amicales et leurs suggestions, grandement contribué à en améliorer le contenu et la présentation. Je remercie donc ici très chaleureusement :

Pascal Alessandri, Myriam Butavant, Julie Chami, Céline Chevalier, Vincent Clapiès, Denis Choimet, Rémi Coutens, Jean-Yves Février, Nicolas François, Stéphane Gonnord, Denis Jourdan, Jean-Marie Le Gluher, Nicolas Maillard, Alain Martin, Jean-Marie Monnier, Marie Peronnier, Gabriel Peyré, Marc Pauly, Paul Pichaureau, Jacques Renault, Franz Ridde, Roland Roussillon, François Sauvageot, Laurent Schwald et Frédéric Vivien.

Je remercie tout particulièrement Jérôme Gärtner et Benjamin Monmege, qui ont relu la plus grosse partie du manuscrit — et ont demandé encore davantage de figures ! — ainsi que Bruno Calado, terrifiant (mais parfait) relecteur.

Je remercie également les étudiants qui ont testé la version initiale de ce texte, sous la forme d'un « poly », notamment Yasmine Moulehiawy.

J'ai pu, lors de la rédaction de ce manuscrit, profiter d'échanges fructueux avec Jean-Pierre Barani, Julien Brémont, Yan Doumerc, Kirone Mallick, Julien Michel et Mathilde Weill. Mon collègue Stéphane Gonnord a également beaucoup œuvré pour me faciliter la tâche dans de nombreux domaines. Enfin, Solveig, Anton et Anouck ont gentiment contribué aux illustrations. Qu'ils soient tous ici remerciés.

Je ne saurais oublier Sébastien Desreux, éditeur passionné qui m'a encouragé dès que je lui ai fait part du projet de cet ouvrage, et qui a toujours été présent pour donner un coup de main informatique. Sans lui, ce livre n'aurait jamais vu le jour.

Enfin, j'adresse un remerciement tout particulier à mon ami Emmanuel Kowalski, professeur à l'ETH de Zürich, qui a bien voulu rédiger un très intéressant chapitre sur quelques applications des probabilités au domaine — hautement non probabiliste *a priori* ! — de l'arithmétique.

Bien que l'ensemble du texte ait été consciencieusement relu, des erreurs, omissions ou imprécisions peuvent être encore présentes. L'auteur accueillera avec gratitude toute remarque, critique ou correction que le lecteur voudra bien lui faire parvenir, aux bons soins de l'éditeur et, par exemple, à l'adresse électronique `Errare.humanum.est@H-K.fr`.

Un *errata* sera régulièrement tenu à jour sur le site `www.H-K.fr`.

Imprimé en Chine par Partenaires Livres.
ISBN 978-2-35141-326-5

Probabilités pour les Non-Probabilistes

WALTER APPEL

H&K Éditions
68, boulevard de Port-Royal 75005 PARIS

Parce que le hasard, il a bon dos. Prenons un des singes, là. Objectivement, quelles sont ses chances de sortir un livre de 200 pages au hasard ? Mettons que ça fasse une combinaison de 400 000 signes. 100 caractères sur un clavier, en général. Donc ça fera 100 puissance 400 000 possibilités. Un « 1 » avec 800 000 zéros derrière.

En gros, si un milliard d'ordinateurs essayaient chacun un milliard de combinaisons par seconde depuis le big bang, ils auraient à peine eu le temps de faire suffisamment d'essais pour tomber sur le titre.

Tous ces efforts pour 200 pages. Le livre a intérêt à être bon. La haine si tu tombes sur du Dan Brown.

BOULET, *Notes*, tome 7, « Formicapunk » [12]

Sommaire

Notations 22
Introduction 24

I Introduction heuristique

1 Probabilités (heuristique) 31
2 Variable aléatoire (heuristique) 55

II Théorie et Pratique

3 Espaces probabilisé 77
4 Conditionnement 101
5 Indépendance (1/2) 111
6 Variables aléatoires discrètes 127
7 Espérance de v.a. discrètes 143
8 Couples de v.a. discrètes 173
9 Le jeu de *pile* ou *face* 203
10 Variables aléatoires à densité 223
11 Espérance d'une v.a. à densité 245
12 Couples de v.a. à densité 257

III Plus formellement...

13 Variables aléatoires (cas général) 291
14 L'espérance comme intégration... 307
15 Vecteurs aléatoires 331
16 Indépendance (2/2) 341
17 Fonctions caractéristiques 357
18 Vecteurs gaussiens (*) 365

IV Théorèmes limites

19 Convergences 383
20 Loi des grands nombres 417
21 Le théorème limite central 437
22 Techniques d'approximations 447
23 Espérance conditionnelle 469

V Quelques applications

24 Résultats élégamment établis... 495
25 Simuler informatiquement une loi 511
26 Méthodes de Monte-Carlo 533
27 Chaînes de Markov 551
28 Arithmétique et probabilités 579
29 Bosons, fermions et boltzmannions 589
30 Processus de Poisson 603
31 Mouvements browniens 611
32 Estimation bayésienne 643
33 Loi de l'Arc sinus et séries aléatoires 647
34 À propos de quelques paradoxes 655

A Annexes

A Éléments d'analyse et d'algèbre 667
B Davantage sur la mesure 697
C Théorèmes d'existence 711
D Tables 717
E Lexique français-anglais 721

Table des matières

Introduction 24
§ 1. Les lois du hasard . . . 24
§ 2. But de la théorie . . . 25
§ 3. Quelques mots sur l'ouvrage . . . 26
§ 4. Rosencrantz et Guildenstern . . . 27
Exercice . . . 28

Partie I : Introduction heuristique 29

1 Probabilités (heuristique) 31
1.1 Le langage de l'aléatoire . . . 31
§ 5. Phénomènes aléatoires . . . 31
§ 6. L'univers Ω et les réalisations possibles du hasard . . . 33
§ 7. Événements . . . 35
§ 8. Compter/mesurer . . . 35
§ 9. Langage ensembliste . . . 37
§ 10. Calcul des probabilités . . . 38
1.2 Probabilités sur un ensemble fini ou dénombrable . . . 39
§ 11. Univers fini . . . 39
§ 12. Probabilités sur un ensemble dénombrable . . . 42
1.3 Probabilités sur $\{P, F\}^{\mathbb{N}^*}$ . . . 42
§ 13. Un univers non dénombrable . . . 42
§ 14. Unions dénombrables d'événements . . . 44
§ 15. Unions non dénombrables d'événements . . . 45
§ 16. Probabilité sur $\{P, F\}^{\mathbb{N}}$ . . . 45
1.4 Probabilités conditionnelles . . . 45
§ 17. Un exemple simple . . . 45
§ 18. Second exemple . . . 46
§ 19. Formule des probabilités totales . . . 47
§ 20. Formule des probabilités composées . . . 48
§ 21. Indépendance stochastique . . . 48
§ 22. « Indépendance = produit » (*) . . . 49
Exercices . . . 50

2 Variables aléatoires (heuristique) 55
2.1 Variables réelles . . . 56
§ 23. Première définition . . . 56
§ 24. Probabilité image (ou loi) . . . 57
§ 25. Fonction de répartition . . . 58
§ 26. Extension de la notion de variables aléatoires . . . 60
§ 27. Variables discrètes . . . 61
§ 28. Variables à densité . . . 62

§ 29. Est-ce tout ? 64
§ 30. La connaissance de Ω est inutile 65
2.2 La notion d'espérance 65
§ 31. Points de vue du théoricien, point de vue de l'expérimentateur 65
§ 32. Le problème de la sommation 67
2.3 Espérance d'une variable discrète ou à densité 68
§ 33. Cas discret : première (tentative de) définition 68
§ 34. Deux expériences numériques 68
§ 35. Séries commutativement convergentes 70
§ 36. Seconde définition 71
§ 37. Espérance d'une variable à densité 72
Exercices 72

Partie II : Théorie et Pratique 75

3 Espaces probabilisés 77
3.1 Parties, algèbres, tribus 78
§ 38. Unions disjointes 78
§ 39. Algèbres 78
§ 40. Tribus 79
§ 41. Borne supérieure et borne inférieure 79
§ 42. Autres propriétés des tribus 80
§ 43. Construire une tribu. Tribu engendrée par un ensemble de parties 81
3.2 Espaces probabilisés 82
§ 44. Espaces probabilisables 82
§ 45. Espaces probabilisés 84
§ 46. Propriétés des mesures de probabilités 85
§ 47. Événements négligeables 87
§ 48. Espaces probabilisés complets (*) 87
§ 49. Systèmes complets, quasi-complets d'événements 88
§ 50. Formule du crible (ou : de Poincaré) 89
3.3 Probabilité sur $\mathbb{R}$ 90
§ 51. Fonction de répartition d'une probabilité 90
§ 52. La mesure de Lebesgue sur $[0;1]$ 92
§ 53. Probabilités finies 92
§ 54. Probabilités discrètes 93
§ 55. Probabilités à densité 93
3.4 Annexe : π-systèmes et théorème de Carathéodory 94
§ 56. π-systèmes 94
§ 57. Le théorème de Carathéodory 94
3.5 Annexe : un ensemble non mesurable 95
Exercices 96

4 Conditionnement 101
§ 58. Probabilités conditionnelles 101
§ 59. Formule des probabilités composées 103
§ 60. Formule des probabilités totales 103
§ 61. Théorème de Bayes 104
§ 62. Quelques remarques et difficultés 106
Exercices 108

5 Indépendance (1/2) 111
5.1 Indépendance d'événements et de tribus . . . 111
§ 63. Indépendance de deux événements . . . 111
§ 64. Indépendance mutuelle . . . 112
5.2 Les lemmes de Borel-Cantelli . . . 114
§ 65. Limite supérieure et limite inférieure . . . 114
§ 66. Les lemmes de Borel-Cantelli . . . 117
5.3 Application : du jeu de recouvrement au paradoxe d'Olbers . . . 119
§ 67. Le paradoxe historique d'Olbers . . . 119
§ 68. Un modèle bidimensionnel de recouvrement . . . 120
§ 69. Modèle stellaire uniforme : probabilité que le cercle... . . . 121
Exercices . . . 123

6 Variables aléatoires discrètes 127
6.1 Variables aléatoires discrètes . . . 127
§ 70. Variable simple, variable discrète . . . 127
§ 71. Système complet induit par une variable aléatoire discrète . . . 128
§ 72. Loi d'une variable aléatoire discrète . . . 129
§ 73. Fonction de répartition d'une variable aléatoire discrète . . . 130
§ 74. Histogrammes . . . 131
§ 75. Variables aléatoires au sens large . . . 132
6.2 Lois classiques . . . 133
§ 76. Loi uniforme . . . 133
§ 77. Loi de Bernoulli . . . 133
§ 78. Loi binomiale . . . 134
§ 79. Loi géométrique . . . 134
§ 80. Loi de Poisson . . . 136
§ 81. Tirages avec et sans remise. Loi hypergéométrique . . . 139
§ 82. Loi de l'image $\varphi(X)$ d'une variable aléatoire . . . 140
Exercices . . . 141

7 Espérance d'une variable aléatoire discrète 143
7.1 Espérance . . . 143
§ 83. Espérance . . . 143
§ 84. Représentations de X comme combinaisons de fonctions indicatrices . . . 145
§ 85. Espérances infinies . . . 147
§ 86. Linéarité . . . 148
§ 87. Application de la linéarité au calcul pratique d'une espérance . . . 149
§ 88. Autres propriétés de l'espérance . . . 150
§ 89. Intégration terme à terme . . . 151
7.2 Moments . . . 152
§ 90. Formule de transfert . . . 152
§ 91. Moments . . . 152
§ 92. Moments centrés . . . 152
§ 93. Variance, écart-type . . . 153
§ 94. Propriétés de la variance . . . 156
§ 95. Variable centrée réduite . . . 156
§ 96. Inégalité de Cauchy-Schwarz . . . 156
§ 97. Espérance et variance des lois classiques . . . 157
§ 98. Inégalité de Bienaymé-Tchebychev . . . 157
§ 99. Atouts et faiblesses de l'inégalite de Bienaymé-Tchebychev . . . 159

7.3 Fonctions génératrices 160
§ 100. Fonction génératrice 160
§ 101. Récupération des moments 161
§ 102. Fonctions génératrices des lois classiques 162
7.4 Un problème classique : le collectionneur 163
§ 103. Position du problème 163
§ 104. Loi du temps d'attente 164
§ 105. Une majoration de la probabilité de déviation 165
Exercices 166

8 Couples de variables aléatoires discrètes 173
§ 106. Position du problème 173
8.1 Loi d'un couple 174
§ 107. Loi conjointe, lois marginales 174
§ 108. Lois conditionnelles 175
§ 109. Somme de deux variables aléatoires à valeurs dans $\mathbb{Z}$ 176
§ 110. Théorème de transfert 176
8.2 Indépendance 177
§ 111. Indépendance de deux variables aléatoires discrètes 177
§ 112. Indépendance mutuelle 178
§ 113. Fonctions de variables indépendantes 179
§ 114. Un exemple : minimum et maximum de variables indépendantes 180
§ 115. Espérance d'un produit de variables aléatoires indépendantes 181
§ 116. Somme de variables indépendantes — formule de convolution 182
§ 117. Loi binomiale comme somme de lois de Bernoulli 182
§ 118. Théorèmes de stabilité 183
8.3 Covariance 183
§ 119. Covariance d'un couple discret 183
§ 120. Somme de variables aléatoires 185
§ 121. Coefficient de corrélation 187
§ 122. Invariance par changement d'échelle de $\rho(X, Y)$ 187
§ 123. Covariance d'un couple de données statistiques 188
8.4 Espérance conditionnelle (cas discret) 189
§ 124. Lois conditionnelles 189
§ 125. Espérance conditionnelle, première approche 190
§ 126. Espérance conditionnelle, seconde approche 191
8.5 Sommation d'un nombre aléatoire de variables 192
§ 127. Contexte historique et mathématique 192
§ 128. Identités de Wald 193
§ 129. Processus de Galton-Watson 195
§ 130. Probabilité d'extinction 196
§ 131. Évolution de la population 197
Exercices 198

9 Le jeu de pile ou face 203
9.1 Formalisation 203
§ 132. Espace probabilisé associé à une partie infinie de *pile* ou *face* 203
§ 133. Quelques notations 204
9.2 Le motif dans le tapis 205
§ 134. Longueur moyenne de la deuxième séquence homogène 205
§ 135. Vaut-il mieux parier FPP ou PPF ? 206

§ 136. Occurrence de motifs, événements régénératifs 207
§ 137. Temps d'attente d'un des motifs PPF et FPP 209
§ 138. Temps d'attente du motif PPP 210
9.3 Théorèmes limites 211
§ 139. Une certitude : la moyenne 211
§ 140. Retour à l'équilibre et marches aléatoires 213
9.4 Les nombres normaux existent ! 215
§ 141. Une suite de variables de Bernoulli indépendantes est un réel de [0; 1[(et vice-versa) 215
§ 142. Propriétés des fonctions de Rademacher 216
§ 143. Le théorème des nombres normaux de Borel 218
§ 144. Existence de nombres complètement normaux 218
Exercices 220

10 Variables aléatoires à densité **223**
§ 145. À propos de l'intégrabilité des fonctions 223
10.1 Variables aléatoires à densité 224
§ 146. Variables continues 224
§ 147. Variables absolument continues, ou à densité 224
§ 148. Propriétés de la densité de probabilité 227
§ 149. Changement d'échelle 227
§ 150. Médiane 228
§ 151. Variables symétriques 229
10.2 Lois classiques 229
§ 152. Loi uniforme 229
§ 153. Loi exponentielle 230
§ 154. Loi normale centrée réduite 231
§ 155. Loi normale 232
§ 156. Loi de Cauchy 234
§ 157. Loi Γ (Gamma) 235
§ 158. Loi γ 235
§ 159. Loi du χ^2 235
§ 160. Loi bêta $B(a, b)$ 235
10.3 Changement de variable aléatoire 237
§ 161. Loi de $\varphi(X)$ lorsque φ est injective 237
§ 162. Loi de $\varphi(X)$: cas général 238
§ 163. Simuler une loi à densité 239
Exercices 241

11 Espérance d'une variable aléatoire à densité **245**
11.1 Espérance et moments 245
§ 164. Espérance 245
§ 165. Formule de transfert 246
§ 166. Moments 247
§ 167. Moments centrés 247
§ 168. Variance 248
§ 169. Inégalité de Cauchy-Schwarz 249
§ 170. Espérance et variance des lois classiques 249
§ 171. Variables centrées réduites 252
§ 172. Autres caractéristiques numériques d'une variable aléatoire à densité . . 252

11.2 Inégalités de Markov et de Bienaymé-Tchebychev 253
§ 173. Inégalité de Markov 253
§ 174. Inégalité de Bienaymé-Tchebychev 253
Exercices 254

12 Couples de variables à densité **257**
12.1 Loi d'un couple 257
§ 175. Loi d'un couple, lois marginales 257
§ 176. Couple admettant une densité 258
§ 177. Densités marginales 261
§ 178. Somme de deux variables aléatoires à densité 261
§ 179. Théorème de transfert 263
12.2 Indépendance 263
§ 180. Indépendance de deux variables aléatoires à densité 263
§ 181. Indépendance mutuelle 264
§ 182. Somme de deux variables aléatoires indépendantes 265
§ 183. Somme de lois normales indépendantes 266
§ 184. Un exemple : quotient de deux lois normales 266
§ 185. Maximum et minimum de N variables aléatoires indépendantes 267
§ 186. Espérance du produit de variables indépendantes 268
12.3 Covariance 269
§ 187. Covariance d'un couple à densité 269
§ 188. Coefficient de corrélation 270
12.4 Lois conditionnelles 271
§ 189. Lois conditionnelles 271
§ 190. Exemple de variable conditionnée par une autre 274
§ 191. Un exemple de loi mixte (discret, continu) : temps d'attente du métro . 274
§ 192. Espérances conditionnelles 277
Exercices 278

Partie III : Plus formellement... **289**

13 Variables aléatoires : cas général **291**
13.1 Mesurabilité 291
§ 193. Définition d'une variable aléatoire 291
§ 194. Calcul des images réciproques 292
§ 195. Quelques résultats sur la mesurabilité 292
13.2 Fonction de répartition 294
§ 196. Loi d'une variable aléatoire 294
§ 197. Fonction de répartition 294
§ 198. Variables discrètes, continues, mixtes 296
§ 199. Décomposition de Lebesgue d'une fonction de répartition 297
§ 200. Densité au sens des distributions 299
§ 201. Une définition de l'espérance n'utilisant que la fonction de répartition . 300
§ 202. L'espérance et les notations de l'intégrale de Stieltjes 301
13.3 Limite simple de variables aléatoires 302
Exercices 304

14 L'espérance comme intégration sur une mesure de probabilité 307
14.1 Construction de l'espérance comme intégrale . . . 308
§ 203. Espérance d'une variable aléatoire simple . . . 308
§ 204. Espérance d'une variable aléatoire positive . . . 309
§ 205. Espérance comme limite croissante d'une suite de variables simples (*) . 309
§ 206. Espérance d'une variable aléatoire réelle . . . 310
§ 207. Espérance d'une variable aléatoire complexe . . . 311
§ 208. Propriétés de l'espérance ; espace $\mathcal{L}^1$ . . . 311
§ 209. Intégration sur une partie . . . 313
§ 210. Espaces $\mathcal{L}^1$ et L^1 . . . 313
§ 211. Moments d'ordre supérieur . . . 314
14.2 Théorèmes de convergence . . . 315
§ 212. Théorème de convergence monotone . . . 315
§ 213. Lemme de Fatou (*) . . . 315
§ 214. Théorème de convergence dominée . . . 316
§ 215. La Cavalerie Légère à l'œuvre . . . 317
14.3 Le théorème de transfert et ses applications . . . 318
§ 216. Le théorème général . . . 318
§ 217. Cas des variables à densité . . . 319
14.4 Espace L^2 ; interprétations géométriques . . . 321
§ 218. Espaces $\mathcal{L}^2$ et L^2 . . . 321
§ 219. Inégalité de Cauchy-Schwarz . . . 321
§ 220. Covariance . . . 322
§ 221. Structure de l'espace L^2 . . . 323
§ 222. Interprétations géométriques . . . 324
§ 223. Meilleur estimateur quadratique d'une variable aléatoire . . . 325
14.5 Inégalités . . . 326
§ 224. Inégalité de Markov . . . 326
§ 225. Inégalités de Bienaymé-Tchebychev . . . 327
§ 226. Inégalité de Bernstein . . . 328
§ 227. Inégalité de Jensen (convexité) . . . 328
14.6 Annexe : le théorème de convergence monotone . . . 329
Exercices . . . 329

15 Vecteurs aléatoires 331
§ 228. Notations . . . 331
15.1 Probabilités sur $\mathbb{R}^n$ . . . 332
§ 229. Boréliens de $\mathbb{R}^n$ . . . 332
§ 230. Probabilité produit . . . 332
§ 231. Densités . . . 332
15.2 Vecteurs aléatoires . . . 332
§ 232. Espérance . . . 333
§ 233. Matrice de covariance . . . 333
§ 234. Inégalité de Cauchy-Schwarz . . . 334
15.3 Fubini !!! . . . 335
§ 235. Le théorème de Fubini en pratique . . . 335
§ 236. Application : densités marginales . . . 335
§ 237. Application : expression alternative de l'espérance . . . 336
15.4 Changements de variables aléatoires . . . 337
§ 238. Changement de vecteur aléatoire . . . 337
§ 239. Somme, produit et quotient de deux variables aléatoires . . . 338
Exercices . . . 339

16 Indépendance (2/2) **341**
16.1 Indépendance de variables aléatoires 341
§ 240. Indépendance de deux variables aléatoires 341
§ 241. Indépendance de classes et de tribus d'événements 343
§ 242. Indépendance de n variables aléatoires 344
§ 243. Caractérisations de l'indépendance de variables aléatoires 344
§ 244. Coalitions 344
§ 245. Indépendance de variables aléatoires discrètes 345
§ 246. Indépendance de variables aléatoires à densité 345
16.2 Somme et produit de variables aléatoires indépendantes 346
§ 247. Espérance d'un produit 346
§ 248. Formule de convolution 348
§ 249. Formule de convolution pour les variables à densité 349
§ 250. Somme de n variables aléatoires indépendantes 350
16.3 Temps d'arrêt et identité de Wald 350
§ 251. Notion de temps d'arrêt 350
§ 252. Temps d'arrêt et identités de Wald 351
Exercices 353

17 Fonctions caractéristiques **357**
§ 253. Variables aléatoires complexes 357
§ 254. Fonction caractéristique 358
§ 255. Cas des variables aléatoires discrètes 359
§ 256. Propriétés de la fonction caractéristique 359
§ 257. Fonction caractéristique d'une somme de variables indépendantes . . . 360
§ 258. Théorème d'unicité 360
§ 259. Fonction caractéristique et indépendance 361
§ 260. Régularité de la fonction caractéristique 361
§ 261. Fonctions caractéristiques usuelles 362
§ 262. Le théorème de continuité de Paul Lévy 362
Exercices 363

18 Vecteurs gaussiens **365**
18.1 Lois normales sur $\mathbb{R}^2$ 365
§ 263. Lois normales et extension 365
§ 264. Lois normales sur $\mathbb{R}^2$: cas non dégénéré 366
§ 265. Réduction d'une loi normale non dégénérée sur $\mathbb{R}^2$ 368
18.2 Aspects numériques de lois normales sur $\mathbb{R}^2$ 370
§ 266. Lois marginales 370
§ 267. Combinaisons linéaires de X et Y 371
§ 268. Densités conditionnelles 372
§ 269. Ellipses d'égale densité 372
18.3 Vecteurs gaussiens 374
§ 270. Définition des vecteurs gaussiens 374
§ 271. Caractérisation par les combinaisons linéaires 374
§ 272. Indépendance et décorrélation 375
§ 273. Décomposition d'un vecteur gaussien 375
§ 274. Densité d'un vecteur gaussien 377
§ 275. Ellipsoïdes d'égale densité et loi du χ^2 378
§ 276. Lemme des moments (théorème de Wick) 378
Exercices 379

Partie IV : Théorèmes limites **381**

19 Convergence **383**

§ 277. Introduction aux théorèmes limites . . . 383

19.1 Convergence presque sûre, convergence en probabilité . . . 385

§ 278. Convergence presque sûre . . . 385

§ 279. Critères de convergence presque sûre (∗) . . . 386

§ 280. Convergence en probabilité . . . 388

§ 281. Une caractérisation par l'espérance . . . 388

§ 282. Convergence et images de variables aléatoires . . . 389

§ 283. Comparaison entre convergence en probabilité et convergence p.s. . . . 389

§ 284. Une distance sur l'espace des variables aléatoires (∗) . . . 391

19.2 Convergence en moyenne et dans L^p . . . 392

§ 285. Convergence en moyenne et dans L^p . . . 392

§ 286. Convergence en probabilité et convergence dans L^p (∗) . . . 392

19.3 Convergence en loi . . . 393

§ 287. Convergence en loi . . . 393

§ 288. Convergence en loi *versus* convergence étroite . . . 394

§ 289. Convergence en loi et convergence en probabilité . . . 396

§ 290. Convergence en loi de variables discrètes . . . 396

§ 291. Convergence en loi de variables à densité . . . 399

§ 292. Convergence en loi et fonctions caractéristiques . . . 401

§ 293. Convergence en loi vers une constante . . . 401

§ 294. Convergence en loi d'un couple, théorème de Slutsky . . . 402

19.4 Une application : produit scalaire de vecteurs unitaires aléatoires . 403

§ 295. Vecteur aléatoire sur la sphère S_n . . . 403

§ 296. Loi du produit scalaire . . . 403

§ 297. Loi limite . . . 404

19.5 Seconde application : suites équidistribuées . . . 405

§ 298. Suites équidistribuées . . . 405

§ 299. Critère de Weyl . . . 405

§ 300. Application aux suites ($n\alpha \mod 1$) . . . 407

19.6 Résumé synoptique . . . 409

19.7 Annexe : convergence étroite . . . 410

§ 301. Théorème de représentation de Skorokhod . . . 410

§ 302. Équivalence de la convergence en loi et de la convergence étroite . . . 410

Exercices . . . 412

20 Loi des grands nombres **417**

20.1 La loi du tout ou rien . . . 417

§ 303. Événements de queue, tribu asymptotique . . . 417

§ 304. Loi du 0–1 de Kolmogorov . . . 419

§ 305. Loi du tout ou rien pour les événements . . . 420

§ 306. Loi du tout ou rien de Hewitt-Savage (∗) . . . 420

§ 307. Une application aux séries aléatoires . . . 421

§ 308. Une application aux marches aléatoires . . . 421

20.2 Lois des grands nombres . . . 423

§ 309. La loi faible des grands nombres . . . 424

§ 310. La loi forte des grands nombres . . . 425

§ 311. Cas de la divergence (∗) . . . 427

20.3 Théorème de Glivenko-Cantelli . . . 428

§ 312. Fonction de répartition empirique 428
§ 313. Application : le test de Kolmogorov-Smirnov 430
20.4 Annexe : démonstration de la loi des grands nombres 432
Exercices 434

21 Le théorème limite central **437**
§ 314. Au-delà de la loi des grands nombres 437
§ 315. Le théorème limite central 438
§ 316. Autres formes du théorème limite central (*) 440
§ 317. Forme locale du théorème limite central 441
§ 318. Évaluation de la vitesse de convergence 442
§ 319. Extension aux variables vectorielles (*) 443
§ 320. Utilité du théorème limite central 443
§ 321. La physique et la loi du $1/\sqrt{n}$ 443
§ 322. Au-delà du théorème limite central : la loi du logarithme itéré 444
Exercices 446

22 Techniques d'approximation **447**
22.1 Approximations poissoniennes 447
§ 323. Convergence vers une loi de Poisson 447
§ 324. Approximation poissonienne d'une loi binomiale 448
22.2 Approximation normale 449
§ 325. Approximation normale d'une loi binomiale ; correction de continuité . . 449
§ 326. Approximation normale d'une loi de Poisson 451
§ 327. Approximation normale et inégalité de Bienaymé-Tchebychev 452
§ 328. Estimer un nombre de lancers nécessaires 453
§ 329. Résumé des conditions d'approximation 456
22.3 Régression linéaire 456
§ 330. Droite de régression 456
§ 331. Droite de régression par rapport à y 457
§ 332. Corrélation et causalité 459
22.4 Estimation 460
§ 333. Un exemple : moyenne empirique 460
§ 334. Un exemple : variance empirique 460
§ 335. Notion d'estimateur 462
§ 336. Risque quadratique 462
§ 337. Intervalles de confiance 463
Exercices 464

23 Espérance conditionnelle **469**
23.1 Le cas fini ou dénombrable 469
§ 338. Espérance conditionnelle sachant une partition 470
§ 339. Un autre point de vue 471
§ 340. La propriété fondamentale de l'espérance conditionnelle 472
§ 341. Propriétés élémentaires 472
§ 342. Espérance conditionnelle sachant une variable aléatoire discrète 474
§ 343. Fonctions mesurables selon une partition dénombrable 475
23.2 Le cas général 476
§ 344. Espérance d'une variable aléatoire sachant un événement 476
§ 345. Espérance conditionnelle sachant une tribu 476
§ 346. Espérance conditionnelle sachant une variable aléatoire 478

§ 347. Lemme de factorisation 478
§ 348. Propriétés de l'espérance conditionnelle 478
§ 349. Théorèmes de convergence 479
§ 350. Densités conditionnelles (*) 480
§ 351. Le point de vue hilbertien : l'espérance conditionnelle comme projection 480
23.3 Martingales 482
§ 352. Ma première martingale 482
§ 353. Filtrations, martingales 482
§ 354. Sur- et sous-martingales 484
§ 355. Théorème de convergence des martingales 484
§ 356. Applications 487
23.4 Annexe : probabilités conditionnelles (*) 487
§ 357. Probabilité conditionnelle sachant une partition dénombrable 487
§ 358. Probabilité conditionnelle sachant une variable aléatoire 488
§ 359. Probabilité conditionnelle sachant une tribu 488
Exercices 489

Partie V : Quelques applications 493

24 Quelques résultats établis par la théorie des probabilités 495
§ 360. La formule de Bernstein 495
§ 361. L'existence de nombres complètement normaux 498
§ 362. Points singuliers du cercle de convergence d'une série entière 499
§ 363. Le théorème de Weierstrass et les polynômes de Bernstein 501
§ 364. En guise de conclusion 506
Exercices 506

25 Simuler informatiquement une loi de probabilité 511
25.1 Simuler une loi discrète 511
§ 365. Simuler une loi de Bernoulli 512
§ 366. Simuler une loi binomiale 512
§ 367. Simuler une loi discrète finie 512
§ 368. Simuler une loi géométrique 513
§ 369. Simuler une loi de Poisson 514
25.2 Simuler une loi continue 515
§ 370. Méthode de l'inverse 515
§ 371. Méthode du rejet 516
§ 372. Mélange de fonctions de répartition 518
25.3 Simuler une loi normale 520
§ 373. Méthode découlant du théorème limite central 520
§ 374. Méthode de la transformée inverse 520
§ 375. Méthode de Box-Muller 520
§ 376. Simulation d'une loi normale par la méthode du rejet 522
§ 377. Simulation d'un vecteur gaussien sur $\mathbb{R}^2$ 523
25.4 Remarques sur les générateurs d'une loi uniforme 525
§ 378. Remarques préliminaires 525
§ 379. Générateur à récurrence linéaire 527
§ 380. Récurrences d'ordre supérieur, ou non linéaires 528
§ 381. Brève remarque historique 529
Exercices 530

26 Méthodes de Monte-Carlo **533**
26.1 Estimation d'une moyenne 534
§ 382. L'aiguille de Buffon 534
26.2 Intégration numérique 536
§ 383. Position du problème 536
§ 384. Convergence de la méthode 539
§ 385. Effet de la dimension 539
26.3 Laissons faire le hasard ! 540
§ 386. Un exemple arithmétique : tests de primalité 540
§ 387. Une machine à deviner des conjectures : le modèle de Cramér 544
§ 388. Le hasard et les algorithmes de recherche de solution optimale 545
§ 389. Falsification de données comptables et la loi de Benford 546
Exercices 549

27 Chaînes de Markov **551**
27.1 Matrices de transition 551
§ 390. Matrices de transition, chaînes de Markov 552
§ 391. Mise en garde : probabilités conditionnelles et mémoire du passé 554
§ 392. Graphe associé à une chaîne de Markov 554
§ 393. Propriétés des matrices stochastiques 555
§ 394. Mesures sur l'espace des états 555
§ 395. Évolution à n pas et équation de Chapman-Kolmogorov 556
§ 396. Évolution du système aux grandes échelles de temps 557
27.2 Classification immédiate des états 559
§ 397. États communicants 559
§ 398. Notion de classes. Graphes et matrices irréductibles. 561
§ 399. Parties fermées. États absorbants 562
§ 400. Périodicité 563
§ 401. Ergodicité (irréductibilité forte) 564
§ 402. Théorème de Perron-Frobenius et théorème ergodique 565
27.3 Classification asymptotique des états 567
§ 403. États transitoires, états récurrents 567
§ 404. Matrice réduite 568
27.4 Méthode « un pas en avant » 569
§ 405. Temps moyen de retour 569
§ 406. Temps moyen d'atteinte d'une partie 570
§ 407. Temps de retour et loi stationnaire 570
§ 408. Probabilité d'absorption 571
27.5 Trois exemples 572
§ 409. Attente d'un motif dans le jeu de *pile* ou *face* 572
§ 410. Une chaîne réductible 573
§ 411. Le modèle des urnes d'Ehrenfest 574
Exercices 575

28 Arithmétique et probabilités : une randonnée **579**
28.1 Exemple élémentaire 580
28.2 Le théorème d'Erdős-Kac et la loi normale 581
28.3 Écarts entre nombres premiers et lois de Poisson 583
28.4 Valeurs propres de matrices aléatoires 585

29 Bosons, fermions et boltzmannions 589
29.1 Le facteur de Boltzmann . . . 589
§ 412. Théorie cinétique des gaz et statistiques non quantiques . . . 589
§ 413. « Nombre de complexions » d'un état . . . 591
§ 414. Un exemple (très) élémentaire . . . 592
§ 415. Le facteur de Boltzmann . . . 593
29.2 Statistiques quantiques . . . 596
§ 416. L'entropie et le paradoxe de Gibbs . . . 596
§ 417. Fermions et bosons . . . 598
§ 418. Nombre d'états occupés . . . 600

30 Processus de Poisson 603
§ 419. Introduction : comment placer des points uniformément sur $\mathbb{R}^+$? . . . 603
§ 420. Processus de comptage . . . 604
§ 421. Processus de Poisson . . . 604
§ 422. Caractérisation des processus de Poisson . . . 605
§ 423. Processus marqués . . . 606
§ 424. Distribution des instants d'arrivée . . . 607
§ 425. Effet physique : le bruit de grenaille . . . 607
§ 426. Âge et temps de vie résiduel . . . 607
§ 427. Processus de Poisson sur $\mathbb{R}$ . . . 608
§ 428. Processus de Poisson sur $\mathbb{R}^d$ . . . 609
Exercices . . . 609

31 Mouvements browniens 611
31.1 Une approche heuristique . . . 611
§ 429. L'observation de Robert Brown . . . 611
§ 430. Marche aléatoire discrète et passage à la limite continue . . . 612
§ 431. Limite continue — première approche . . . 613
§ 432. Limite continue — seconde approche . . . 614
§ 433. Une propriété du noyau de la chaleur . . . 616
§ 434. Limite continue d'une marche aléatoire asymétrique . . . 617
§ 435. Marche aléatoire en dimension supérieure . . . 617
§ 436. Temps moyen de visite . . . 618
§ 437. Dernières remarques . . . 619
31.2 Modélisation probabiliste . . . 620
§ 438. Définition . . . 621
§ 439. La loi du mouvement brownien est la loi limite de marches aléatoires . . . 622
§ 440. Loi du mouvement brownien ; mesure de Wiener . . . 623
§ 441. Construction canonique du mouvement brownien . . . 624
§ 442. Covariance . . . 625
31.3 Propriétés des chemins browniens . . . 626
§ 443. Invariances . . . 626
§ 444. Propriétés de régularité . . . 626
§ 445. Récurrence : le théorème de Pólya . . . 627
§ 446. Autres propriétés (*) . . . 627
§ 447. Construction de Lévy-Ciesielski du mouvement brownien . . . 628
31.4 Mouvement brownien, physique et potentiels . . . 630
§ 448. Le modèle d'Einstein et la détermination du nombre d'Avogadro . . . 630
§ 449. Problème de Dirichlet et méthode de Monte-Carlo . . . 630
§ 450. Mouvement brownien et potentiel capacitif . . . 633

§ 451. Théorie du potentiel et propriétés de récurrence ($d = 1$) 633
§ 452. Théorie du potentiel et propriétés de récurrence ($d = 2$) 634
§ 453. Théorie du potentiel et propriétés de récurrence ($d \geqslant 3$) 635
31.5 Annexes 637
§ 454. Détermination du noyau de la chaleur 637
§ 455. Démonstration du théorème de Paley-Wiener-Zygmund 637
Exercices 639

32 Estimation bayésienne **643**
§ 456. Distribution *a priori* uniforme, et distribution *a posteriori* 643
§ 457. Choix d'une autre distribution *a priori* 646
Exercices 646

33 Retour à pile ou face : loi de l'Arc sinus et séries aléatoires **647**
33.1 L'iniquité flagrante d'un jeu équitable 647
§ 458. La loi de l'Arc sinus 647
33.2 Séries harmoniques de signe aléatoire 649
§ 459. Cas d'une pièce déséquilibrée 649
§ 460. Cas équilibré 650
33.3 Rayon d'une série entière aléatoire 651
§ 461. Le rayon est presque sûrement constant 651
§ 462. Trois rayons sont possibles 652

34 À propos de quelques paradoxes **655**
34.1 Biais d'équiprobabilité 655
§ 463. Biais d'équiprobabilité : le paradoxe des deux cassettes 655
§ 464. Le problème de Monty Hall 656
§ 465. Retour au problème des deux cassettes 657
34.2 Problèmes dus à l'interprétation de l'espérance 658
§ 466. L'espérance d'un jeu n'est pas un critère suffisant 658
§ 467. Paradoxe de Saint-Pétersbourg 658
34.3 Pseudo-paradoxes 661
§ 468. Le paradoxe de Walter Penney 661
Exercices 662

Annexes

A Éléments d'analyse et d'algèbre **667**
A.1 Suites 667
A1. Fini, dénombrable, indénombrable 667
A2. Limite supérieure, limite inférieure 668
A3. Suites de Cauchy 668
A4. Comparaison de suites 669
A5. Formule de Stirling 669
A.2 Séries numériques 669
A6. Définitions 669
A7. Séries positives 669
A8. Théorème de comparaison de séries positives 670
A9. Produits infinis 670
A10. Convergence absolue 670
A11. Exponentielle complexe 670
A12. Changement d'ordre de sommation 671
A13. Familles sommables positives 672

A14. Familles sommables de signes quelconques . . . 672
A15. Théorème de Fubini pour les séries doubles . . . 672
A16. Sommation par paquets . . . 674
A17. Théorème de convergence dominée discrète . . . 674
A.3 Fonctions . . . 674
A18. Continuité, continuité à droite . . . 674
A19. Fonctions à variation bornée . . . 675
A20. Fonctions absolument continues . . . 675
A21. Fonction Γ d'Euler . . . 675
A22. Fonction d'erreur, fonction de survie de la loi normale . . . 675
A.4 Fonctions croissantes . . . 676
A23. Continuité, continuité à gauche, à droite . . . 676
A24. Discontinuités de 1re et 2e espèce. . . . 676
A25. Sauts de discontinuité d'une fonction croissante . . . 677
A26. Régularité d'une fonction croissante . . . 677
A27. Décomposition de Lebesgue . . . 677
A28. Une fonction singulière : la fonction de Cantor . . . 678
A29. Pseudo-inverse d'une fonction de répartition . . . 679
A.5 Fonctions convexes . . . 680
A30. Fonctions convexes, strictement convexe, concaves . . . 680
A31. Caractérisation par la pente . . . 681
A32. Caractérisation par la dérivée . . . 681
A33. Cordes, tangentes et demi-tangentes . . . 681
A34. Inégalités de Jensen . . . 682
A.6 Modes de convergence de suites et séries de fonctions . . . 682
A35. Convergence simple et convergence uniforme . . . 682
A36. Théorèmes de Weierstrass . . . 683
A37. Séries de fonctions . . . 683
A.7 Séries entières . . . 683
A38. Rayon . . . 683
A39. Fonctions définies par une série entière . . . 684
A40. Lemme radial d'Abel . . . 684
A41. Produit de Cauchy de deux séries entières . . . 684
A42. Quelques sommes . . . 685
A.8 Calcul intégral . . . 685
A43. Théorème de Fubini . . . 685
A44. Changement de variable dans $\mathbb{R}^n$ . . . 686
A45. Produit de convolution . . . 686
A46. Quelques intégrales utiles . . . 687
A47. Méthode des multiplicateurs de Lagrange . . . 687
A.9 Espaces de Hilbert . . . 687
A48. Espaces préhilbertiens . . . 687
A49. Espaces de Hilbert, bases hilbertiennes . . . 688
A50. Théorème de projection orthogonale . . . 689
A.10 Transformée de Fourier . . . 689
A51. Transformée de Fourier d'une fonction intégrable . . . 689
A52. Distributions. Distributions tempérées . . . 689
A53. Transformée de Fourier d'une distribution . . . 690
A.11 Dénombrement . . . 690
A54. Coefficients binomiaux . . . 690
A55. Formulaire pour les coefficients binomiaux . . . 691
A56. Nombres de Stirling . . . 691
A.12 Topologie élémentaire . . . 693
A57. Ouverts, fermés . . . 693
A58. Espaces complets . . . 693
A.13 Réduction, théorème spectral . . . 693
A59. Adjoint . . . 693
A60. Théorème spectral . . . 694
A61. Matrices symétriques positives, définies positives . . . 694

A.14 Matrices positives 694
A62. Rayon spectral 694
A63. Matrices à coefficients positifs 694
A64. Matrices irréductibles, matrices ergodiques 695
Exercices 696

B Davantage sur la mesure **697**
B.1 Boréliens 697
A65. Boréliens de $\mathbb{R}$ 697
A66. Boréliens de $\mathbb{R}^n$ 699
A67. Boréliens de $\mathbb{R}^\infty$ 699
A68. Tribus produits, espaces produits 699
B.2 Fonctions mesurables 700
A69. Définition et premières propriétés 700
A70. Tribu engendrée par une variable aléatoire 701
B.3 Un complément au jeu de *pile* ou *face* 702
A71. Additivité dénombrable de $\mathbf{P}_0$ sur la classe $\mathfrak{C}$ des cylindres (*) 702
B.4 Le lemme de classe monotone 703
A72. π-systèmes et classes monotones 703
A73. Première application : le lemme des coalitions 705
A74. Deuxième application : rôle central de la fonction de répartition 706
A75. Le théorème d'extension de Carathéodory 707
B.5 Le théorème de Radon-Nikodým 707
A76. Mesure absolument continue par rapport à une autre 707
A77. Le théorème de Radon-Nikodým 707
Exercices 708

C Théorèmes d'existence **711**
C.1 Quelques exemples simples de théorèmes d'existence 711
A78. Existence d'une variable aléatoire de loi donnée 711
A79. Un exemple de suite $(X_n)_{n\geqslant 0}$ de variables aléatoires indépendantes 712
A80. Une suite de variables uniformes indépendantes 713
A81. Une suite de variables indépendantes de lois quelconques données 714
C.2 Le théorème fondamental de Kolmogorov 714
A82. Conditions de compatibilité et théorème d'extension 714

D Tables **717**
D.1 Langage ensembliste et langage probabiliste 717
D.2 Probabilités 718
A83. Axiomes 718
A84. Théorèmes d'intégration 718
A85. Formules utiles 718
A86. Tables des principales lois 719
A87. Loi normale centrée réduite 720

E Lexique français-anglais **721**

Références **723**

Index **727**

Notations

Abréviations	
CNS	condition nécessaire et suffisante
CP	critère pratique
p.s.	presque sûre, presque sûrement
p.p.	presque partout

Ensembles	
$\mathbb{N}$	ensemble des entiers naturels $\mathbb{N} = \{0, 1, 2, \dots\}$
$\mathbb{P}$	ensemble des entiers premiers $\mathbb{P} = \{2, 3, 5, 7, 11, \dots\}$
$\mathfrak{P}(\Omega)$	ensemble des parties de Ω
$A \cap B$	intersection de A et B
$A \cup B$	réunion de A et B
$A + B$	réunion disjointe de A et B ($A \cap B = \varnothing$, $A + B := A \cup B$)
$\sum A_n$	réunion disjointe
$A \setminus B$	$= A \cap B^c$
$A - B$	$= A \setminus B$, étant entendu que $B \subset A$
A^c	complémentaire de A (dans Ω)
$A_n \uparrow A$, $A = \lim \uparrow A_n$	limite croissante
$A_n \downarrow A$, $A = \lim \downarrow A_n$	limite décroissante
$A^* = \limsup A_n$	
$A_* = \liminf A_n$	
$\mathfrak{M}_n(\mathbb{K})$	ensemble des matrices carrées d'ordre n
$\mathfrak{M}_{n,1}(\mathbb{K})$	ensemble des vecteurs colonnes d'ordre n
$\mathscr{C}^k(I, \mathbb{K})$	ensemble des fonctions $f : I \to \mathbb{K}$ de classe $\mathscr{C}^k$

Fonctions	
$f(x^+)$, $f(x + 0)$	limite à droite $\lim_{h \to 0+} f(x + h)$
$f(x^-)$, $f(x - 0)$	limite à gauche $\lim_{h \to 0-} f(x + h)$
$f(+\infty)$	limite en $+\infty$: $\lim_{x \to +\infty} f(x)$
$f(-\infty)$	limite en $-\infty$: $\lim_{x \to -\infty} f(x)$
$f \otimes g$	produit tensoriel $(x, y) \mapsto f(x)\, g(y)$
$f * g$	produit de convolution

Lettres gothiques															
$\mathfrak{A}$	$\mathfrak{B}$	$\mathfrak{C}$	$\mathfrak{D}$	$\mathfrak{E}$	$\mathfrak{F}$	$\mathfrak{G}$	$\mathfrak{H}$	$\mathfrak{I}$	$\mathfrak{J}$	$\mathfrak{M}$	$\mathfrak{m}$	$\mathfrak{O}$	$\mathfrak{P}$	$\mathfrak{T}$	$\mathfrak{X}$
A	B	C	D	E	F	G	H	I	J	M	m	O	P	T	X

Probabilités

Ω	univers, espace des épreuves	
$\mathfrak{T}$, $\mathfrak{F}$, $\mathfrak{A}$	tribus (« T » et « F » gothiques) ou algèbres (« A »)	
ω	réalisation du hasard (élément de Ω)	
$\mathbf{P}$, $\mathbf{Q}$	probabilités	
$X, Y, Z, T, \ldots$	variables aléatoires	
$\mathbf{Cov}(X, Y)$	covariance du couple (X, Y)	
$\mathbf{E}(X), \mathbf{E}X$	espérance de la variable X	
$\mathbf{P}(A), \mathbf{P}\{X \in B\}$	probabilité	
$\mathbf{V}(X)$	variance de la variable X	
σ_X	écart-type de X	
$\mathfrak{N}(x)$	fonction de répartition de la loi $\mathcal{N}(0\,;\,1)$	§ 155
$\mathfrak{n}(x)$	densité de probabilité de la loi $\mathcal{N}(0\,;\,1)$	§ 154
$\{X \in A\}$	événement lié à une variable aléatoire	
$\mathbf{P} \otimes \mathbf{Q}$	probabilité produit	§ 230
$f \otimes g$	produit direct,ou tensoriel $(f \otimes g)(x, y) = f(x)\, g(y)$	§ 22, § 180
$X \wedge Y$, $X \wedge m$	$\min(X, Y)$, $\min(X, m)$	
X'	variable centrée $X' = X - \mathbf{E}(X)$	
X^*	variable centrée réduite $X^* = (X - \mathbf{E}X)/\sigma_X$	
X^+	$\max(X, 0)$	§ 206
X^-	$\max(-X, 0)$	§ 206
$X_n \xrightarrow{\text{ps}} X$	convergence presque sûre	§ 278
$X_n \xrightarrow{\text{P}} X$	convergence en proba	§ 280
$X_n \xRightarrow{\text{L}} X$	convergence en loi (convergence étroite)	§ 287
$X_n \xrightarrow{L^p} X$	convergence dans L^p	§ 285
$X \rightsquigarrow \mathscr{L}$	X suit la loi $\mathscr{L}$	
$\mathscr{B}(p)$	loi de Bernoulli de paramètre p	$p \in [0\,;\,1]$
$\mathscr{B}(n\,;\,p)$	loi binomiale de paramètres n et p	$n \in \mathbb{N}$,
$\mathscr{E}(\alpha)$	loi exponentielle de paramètre α	
$\mathscr{G}_{\mathbb{N}}(p)$	loi géométrique de paramètre p	
$\mathscr{N}(m\,;\,\sigma^2)$	loi normale de moyenne m et de variance σ^2	
$\mathscr{P}(\lambda)$	loi de Poisson de paramètre λ	
$\mathfrak{B}$, $\mathfrak{B}(\mathbb{R})$, $\mathfrak{B}(\mathbb{R}^n)$	tribu des boréliens de $\mathbb{R}$, de $\mathbb{R}^n$	
$\mathfrak{T}$, $\mathfrak{E}$, ...	tribus	
$\mathfrak{D}_X$	système complet induit par une variable discrète	
$\mathfrak{A} \preccurlyeq \mathfrak{B}$	$\mathfrak{A}$ est plus fine que $\mathfrak{B}$	
$\sigma(X)$	tribu engendrée par X	
$\sigma(X_1, X_2, \ldots, X_n)$	tribu engendrée par $X_1, X_2, \ldots, X_n$	
δ_a	mesure de Dirac en a	
$\Delta(x, y)$	$\{(s, t) \in \mathbb{R}^2\,;\; s \leqslant x,\; t \leqslant y\}$	§ 175
$\mathcal{L}^0(\Omega)$	espace des variables aléatoires	§ 193
$\mathcal{L}^0_+(\Omega)$	espace des variables aléatoires positives	§ 193
$\mathcal{L}^0_b(\Omega)$	espace des variables aléatoires bornées	§ 193
$\mathcal{L}^1$, L^1	espaces des variables intégrables : $\mathbf{E}\lvert X\rvert < +\infty$	§ 210
$\mathcal{L}^2, L^2$	espaces des variables de carre intégrables : $\mathbf{E}\lvert X\rvert^2 < +\infty$	§ 218
$\mathbf{P}$, $\mathbf{Q}$	matrices de transition	§ 390
$P(i, j)$, $P^{(i, j)}$	coefficients de $\mathbf{P}$, de $\mathbf{P}^n$	
$\boldsymbol{\pi}_n$	loi instantanée	§ 395
$\mathbf{P}_i$	probabilité conditionnée par $\{X_0 = i\}$	§ 394
$i \to j$, $i \leftrightarrow j$	i mène à j, communique avec j	§ 397
$E_g(\lambda, \mathbf{P})$	sous-espace propre à gauche	§ 396

Introduction

Je considère que la théorie des probabilités n'est pas seulement une branche des mathématiques, et ne se limite pas à une collection harmonieuse d'axiomes et de théorèmes. À mon idée, le calcul des probabilités est un mode de pensée.

Mark Kac, *1961.*

§1 Les lois du hasard

Les termes qui reviennent le plus souvent dans un ouvrage de probabilité sont, paradoxalement, « loi » et « presque sûrement » — des locutions associées à la certitude plutôt qu'à l'aléatoire.

Les lois du hasard,

quel bel oxymore ! Et pourtant, c'est bien à étudier les lois (prévisibles) du hasard (par nature imprévisible) que Pascal, plusieurs Bernoulli, Tchebychev, Kolmogorov et autres grands mathématiciens se sont attachés. Que des certitudes puissent apparaître est sans doute un phénomène connu depuis longtemps : au cours d'un grand nombre de tirages, une pièce équilibrée doit mener au résultat *pile* avec une fréquence proche de 1/2, et même d'autant plus proche que le nombre de tirages est grand. C'est presque une tautologie — si l'on observe une fréquence limite qui n'est pas 1/2, on dira tout simplement que la pièce n'est pas équilibrée. Mais ce n'en est pas une : l'intérêt de la mathématisation du problème est de montrer que la fréquence observée admet *effectivement* une limite. Ce théorème, appelé *loi des grands nombres*, fut une révolution en soi : de qualitatif, l'étude du comportement du hasard est devenu quantitatif, et permet donc de prédire des nouveaux phénomènes.

Voici un exemple graphique, plus élaboré mais frappant : ci-dessous, à gauche, une représentation d'une matrice carrée A, dont les coefficients ont été tirés « au hasard » dans l'intervalle $[-1\,;\,1]$. Chaque valeur de gris correspond à une valeur réelle (-1 pour noir, 1 pour blanc). Maintenant, inversons cette matrice[1] et ob-

(1). Elle est sans doute inversible, comme doivent l'être la plupart des matrices dont les

servons, à droite, le résultat (le noir correspond maintenant à la valeur du plus petit coefficient, le blanc à celle du plus grand).

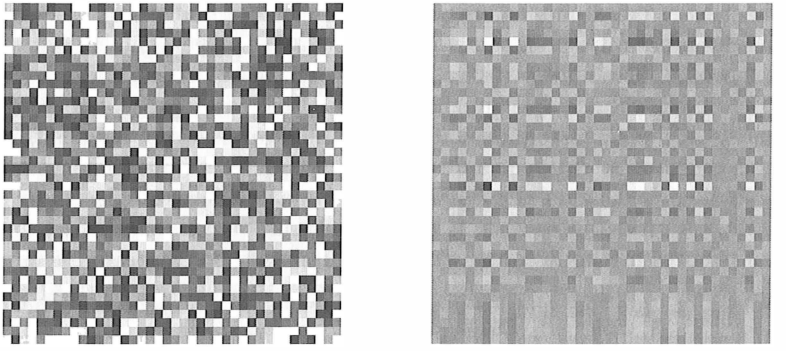


FIGURE 0.1 — À gauche : *Une matrice carrée aléatoire à coefficients dans* $[-1\,;\,1]$. À droite : *L'inverse de cette matrice.*

Que s'est-il passé ? L'aspect « nettement aléatoire » de la première figure a disparu, remplacé par une sorte de motif écossais délavé. Un ordre a surgi du hasard !

But de la théorie §2

On dit parfois en manière de plaisanterie que *le but de la théorie des probabilités est de calculer des probabilités.*

Cette galéjade cache un réel malaise : la notion de probabilité telle qu'étudiée par un mathématicien est sans aucun doute très éloignée de la notion qu'en auront un physicien, un philosophe, ou le fameux homme de la rue (quand il n'est pas philosophe ou mathématicien). La notion elle-même est suffisamment fuyante pour qu'il soit confortable de se cacher derrière les formules et les techniques de calcul et éluder la question de la signification, en la laissant à d'autres.

- La première interprétation, et la plus fréquente, est statistique, et résumée par la formule « nombre de cas favorables sur nombre total de cas ». Mais très rapidement, on en vient à manipuler des nombres infinis et un appareillage mathématique plus sophistiqué est requis. On doit alors faire un saut conceptuel et construire, en confiance, un *modèle*, dont on espère qu'il décrira le monde avec la précision et les propriétés voulues. Pour tester ce modèle, on peut tenter de répéter une expérience un grand nombre de fois, et calculer une fréquence limite. Plusieurs problèmes se posent alors : que veut dire un grand nombre de fois ? que veut dire répéter une expérience ? La réponse mathématique est de considérer des *suites de variables aléatoires indépendantes*, et d'établir des théorèmes affirmant que, sous certaines conditions, la fréquence limite observée est bien la valeur attendue.
- Il existe d'autres interprétations. Pour certains, les probabilités objectives n'existent pas, il n'existe que des probabilités subjectives, qui mesurent notre degré de confiance en certains événements.
- Enfin, on peut se débarrasser des difficultés conceptuelles en les ignorant tout simplement : on décide de s'abstraire entièrement du contexte historique des probabilités, et de n'envisager la théorie que comme un pro-

coefficients sont des réels choisis « au hasard ». Le lecteur sera invité à réfléchir à ce problème ultérieurement.

longement de la théorie de la mesure et de l'intégration. On peut alors axiomatiser entièrement la théorie, la débarrasser de tous les problèmes d'interprétation, de toute référence à des situations concrètes, et réduire la théorie des probabilités à une branche mathématique aussi pure, aussi nette que les autres. Ce point de vue a un côté stérile. Bien que le véritable essor des probabilités n'ait pu avoir lieu qu'une fois l'axiomatisation correctement faite et une théorie de l'intégration profondément ancrée dans le sol mathématique, il n'en reste pas moins que la motivation à la base des travaux les plus cruciaux se rattache à des situations concrètes : phénomènes physiques ou biologiques, calculs de risques, détermination de stratégies. Quoi que l'on fasse, les probabilités restent une branche unique des mathématiques, avec ses propres méthodes et, surtout, ses propres types de raisonnements[(2)].

Comme à chaque fois que les mathématiciens s'approprient un concept — le hasard n'a appartenu, des siècles durant, qu'au champ de la philosophie, de la métaphysique ou des controverses de comptoir —, les outils mathématiques qu'ils créent deviennent en eux-mêmes un objet d'étude. Une part importante de tout ouvrage de probabilités est consacrée à l'étude des concepts et des méthodes spécifiques au champ d'étude du hasard, qui ont pour noms *variables aléatoires*, *indépendance*, *fonction de répartition*, *espérance*, *convergence presque sûre*, *convergence en loi*, etc.

§ 3 Quelques mots sur l'ouvrage

Aucune connaissance spécifique aux probabilités n'est requise pour aborder l'ouvrage, qui commence par une introduction heuristique aux idées et au langage de l'aléatoire. Cette partie est indépendante du reste de l'ouvrage[(3)].

Les mathématiques nécessaires à la compréhension du texte sont rappelées en annexe[(4)].

L'ouvrage est découpé en plusieurs parties :

- La partie I présente, de manière assez informelle, les idées essentielles de la théorie (probabilités conditionnelles, indépendance, variables aléatoires, espérance), ainsi que les problèmes théoriques auxquels on est confronté lorsque l'on manipule des univers infinis.
- La partie II débute par un chapitre assez formel sur les espaces probabilisés. Ce chapitre peut être lu en deux temps : dans un premier temps, seuls les passages balisés par le symbole « ▶ » peuvent être étudiés. Suivent les notions de probabilité conditionelle et d'indépendance. Le reste de la partie traite de deux grandes classes de variables aléatoires (discrètes et à densité), et est tourné vers l'étude pratique. Il correspond globalement au niveau des probabilités exigées en début de cycle universitaire, en classes préparatoires ou au CAPES de mathématiques. Un interlude sur le jeu de *pile* ou *face* présente, par anticipation, quelques résultats intéressants.

---

(2). De nombreux résultats, d'une grande profondeur, ont été établis *avant* (ou *sans*) le cadre axiomatique désormais universellement utilisé. Le lecteur est très vivement incité à lire le merveilleux ouvrage de William FELLER [33], totalement dénué de théorie de la mesure.

(3). Le lecteur pressé d'en découdre avec des définitions plus formelles peut la sauter ; on trouvera des renvois aux exemples importants qui s'y trouvent dans le texte principal.

(4). Pour la partie II, seules la théorie des séries et les rudiments de calcul différentiel et intégral sont nécessaires.

Les parties III et IV font parfois référence à l'intégrale de Lebesgue, sans qu'une connaissance approfondie du sujet soit nécessaire ; notamment, la construction explicite de l'espérance est effectuée sans appel à des connaissances extérieures. Dans cet esprit, nous avons réduit l'importance de l'utilisation des fonctions caractéristiques, sans tout à fait les supprimer.

- La partie III généralise, de manière plus théorique, les résultats de la partie précédente. On y présente la construction générale de l'espérance d'une variable aléatoire, permettant de prouver différentes propriétés admises pour les variables aléatoires à densité, et permettant également de « passer à la limite ». (Deux chapitres peuvent être omis en première lecture : celui sur les fonctions caractéristiques et celui sur les vecteurs gaussiens.) Cette partie est plus du niveau de fin de cycle de licence ou de début de maîtrise, et pourra intéresser également les candidats au concours de l'Agrégation externe.
- La partie IV présente les « théorèmes limites » : après avoir abordé la notion générale de convergence, on étudie les lois de type « zéro-un », les lois des grands nombres et le théorème limite central.
- La partie V présente des applications diverses des probabilités, et les chapitres sont totalement indépendants les uns des autres. Le choix des sujets reflète sans doute davantage les goûts de l'auteur qu'autre chose. Nous conseillons entre autres au lecteur de lire le chapitre sur la simulation informatique d'une loi de probabilités, et de jouer avec un ordinateur...

Nous n'avons pas cherché à *unifier* cette présentation succinte de la vaste théorie des probabilités ; au contraire, nous avons cherché à en *diversifier* les points de vue. Ainsi, certaines notions peuvent être définies jusqu'à trois ou quatre fois, selon le degré de généralité ou d'abstraction cherché, selon le but que l'on recherche[(5)], voire selon le public visé[(6)]. Un désagrément possible pour le lecteur est de devoir lire, parfois, des explications trop détaillées de passages qu'il jugera triviaux, à côté de passages nettement plus difficiles — signalés comme tels par une taille de caractères plus petite, ou par un astérisque (∗).

Enfin, dernière chose, mais non des moindres, nous avons essayé d'*illustrer* le texte au maximum ; nous accueillerons avec beaucoup de gratitude toute remarque, suggestion ou proposition d'ajout concernant les figures.

Un *index*, que nous espérons le plus complet possible, permet de s'y retrouver. Les entrées peuvent être multiples (par exemple, la notion de *convergence en loi* peut être recherchée aussi bien à l'entrée *Convergence* qu'à celle de *Loi*).

Le symbole de la main qui écrit « ✍ », dans la marge extérieure du texte, informe le lecteur de l'existence d'un exercice proche de ce qui vient d'être dit, et qui peut être traité sur-le-champ. Les exercices sont regroupés en fin de chapitres, et suivis, la plupart du temps, d'indications et/ou d'une correction détaillée. ✍0.1

- Les paragraphes du texte principal sont numérotés en continu, de § 1 à § 468.
- Les paragraphes des annexes sont également numérotés en continu, de A1 à A87.

Rosencrantz et Guildenstern §4

Un dernier mot : lors d'expériences (de pensée) mettant en scène deux protagonistes, ceux-ci seront conventionnellement appelés Rosencrantz et Guildenstern, en référence à une pièce de Tom Stoppard, *Rosencrantz and Guildenstern are*

(5). Par exemple, une définition abstraite de l'espérance d'une variable aléatoire telle que celle donnée à la page 311 possède un haut degré de généralité. Elle ravira ceux qui y verront un parallèle avec la théorie de l'intégration selon Lebesgue. Mais elle n'est pas effective, au sens où elle ne permet pas de calculer cette espérance de manière simple. Une autre définition, utilisant l'intégrale de Stieltjes, permet d'enrichir ce point de vue.

(6). Bien que les probabilistes usent de préférence de la fonction de répartition — et ils ont bien raison — les physiciens préféreront sans doute utiliser la densité de probabilité, fût-elle définie au sens des distributions — et ils ont raison également.

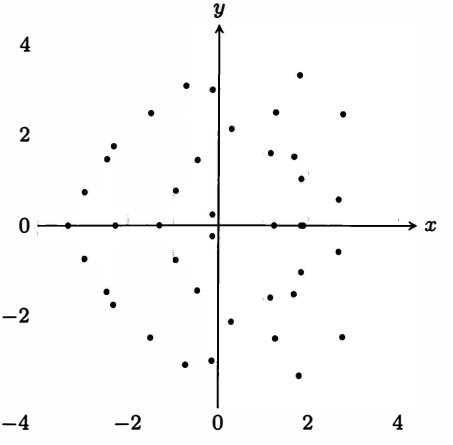
dead [95, 96]. La scène initiale de la pièce représente nos deux héros — personnages mineurs de *Hamlet* — en train de disputer une interminable partie de *pile* ou *face*, dont tous les tirages se soldent par un *face* et le passage d'une pièce de la bourse de Guildenstern à celle de Rosencrantz.

EXERCICE

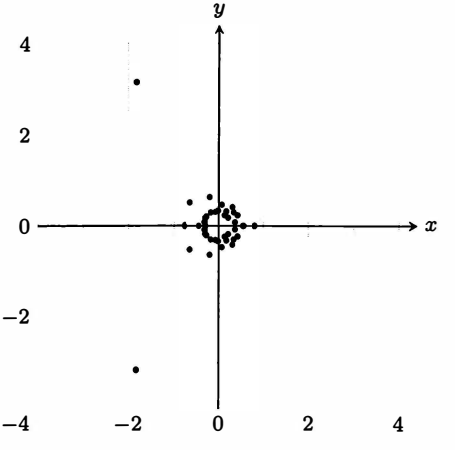
♦ ***Exercice 0.1*** Expliquer *qualitativement* l'allure de la matrice de droite dans la figure 0.1 : pourquoi ce motif écossais ? Pourquoi la matrice est-elle d'aspect plus terne que la précédente ?

Indications

◊ *Indication 0.1 :* La matrice M initiale a de grandes chances d'être diagonalisable (en tant que matrice complexe). Voici son spectre dans le plan complexe.



Que devient le spectre lorsque l'on calcule l'inverse de la matrice ? Et notamment, que deviennent les « petites » valeurs propres ? La figure ci-dessous est représentative...



Enfin, représenter (informatiquement) une matrice « aléatoire » de petit rang (1, 2, 3,...)

Première partie

Introduction heuristique

Chapitre 1

Probabilités

(approche heuristique)

Hasard ?
Mets que font les fripons pour les sots qui le mangent.
Point de hasard !

Victor HUGO
Ruy Blas, *acte IV, scène* VII.

Ce chapitre, ainsi que le suivant, présente un panorama rapide du langage et des concepts les plus importants dans la théorie des probabilités :

- *l'univers des possibles, qui est la collection de toutes les « possibilités de réalisation du hasard » ;*
- *notion d'événement — une partie de l'univers des possibles, dont on va quantifier l'importance, soit par comptage, soit par une mesure plus sophistiquée ;*
- *notion de probabilité conditionnée par un événement, et notion d'indépendance d'événements.*

Des exemples simples sont présentés, ainsi qu'un premier exemple non trivial, permettant de modéliser une succession infinie de tirages à pile *ou* face *— expérience que l'on peut qualifier de fondamentale dans l'approche théorique des probabilités.*

1.1 LE LANGAGE DE L'ALÉATOIRE

Phénomènes aléatoires §5

On parle de phénomène aléatoire lorsque, en recommençant une expérience dans des conditions aussi semblables que possible, le résultat de cette expérience est variable et échappe à toute prévision absolue. À ce titre, le jeu de dé est l'archétype même de l'expérience que l'on associe au hasard[1].

(1). Le nom latin du dé, *alea*, a donné le terme « aléatoire ». Le mot arabe *az-zahr*, signifiant « le dé » et passant par l'espagnol *azar*, a lui-même conduit à « hasard ». Quant à la manière

Généralement, recommencer l'expérience permet de constater une certaine régularité : la notion intuitive de probabilité vient du fait que la fréquence d'apparition d'un résultat semble se stabiliser quand le nombre d'expériences réalisées est grand. Le but de la théorie des probabilités est de *reproduire* ce comportement, c'est-à-dire de fabriquer un modèle qui puisse *prédire* des comportements statistiques. En ce sens, la théorie des probabilités ressemble aux théories physiques : modélisation, vérification de l'adéquation du modèle, prévision de nouveaux phénomènes.

Exemple 1.1 L'étude répétée de lancers d'une pièce équilibrée (tirage à *pile* ou *face*) permet de se rendre compte que :

- chaque face tombe un nombre de fois approximativement égal au nombre total de lancers, divisé par 2 ;
- la fréquence relative d'obtention d'une face s'approche d'autant mieux de 1/2 que l'on effectue un grand nombre de lancers ;
- le résultat d'un lancer est indépendant du résultat des lancers précédents : par exemple, avoir eu deux *pile* de suite ne diminue en rien la probabilité d'obtenir un nouveau *pile* au lancer suivant.

On modélise une succession finie de n tirages à *pile* ou *face* par un modèle mathématique relativement simple : chaque issue possible d'une telle succession de tirages se voit attribuer la même probabilité $1/2^n$. Ce modèle permet d'obtenir des résultats nouveaux, et parfois inattendus : par exemple, s'il s'agit d'une partie dans laquelle deux joueurs s'affrontent, chacun donnant un euro à l'autre lorsqu'il perd, alors chaque joueur a environ une chance sur 5 d'être globalement gagnant durant 90 % du temps de la partie[2].

Remarque 1.2 (Probabilités et confiance) Il arrive que des mathématiciens posent des questions saugrenues comme la suivante : *Quelle est la probabilité qu'un entier* N *donné soit premier ?* Si cette question peut paraître raisonnable pour la valeur N = 16 254 426 798 623 524 463, elle sonne bizarrement pour N = 21.

Ici, la « probabilité » que N soit premier désigne essentiellement le degré de confiance que l'on peut accorder à l'hypothèse « N est premier ». Si l'on a besoin d'un nombre premier pour chiffrer un message extrêmement sensible, un risque de 1 % sera évidemment beaucoup trop important ; un risque de 10^{-7} paraît plus raisonnable. La quantification de ce degré de confiance ne requiert que des outils de base en probabilité ; le lecteur est invité à se reporter au chapitre 26, paragraphe § 386, et plus spécifiquement à la remarque 26.2 page 543.

Depuis ses balbutiements au XVII^e siècle, la théorie mathématique des probabilités s'est peu à peu étoffée et a pris un essor fulgurant dans les années 30, après que Kolmogorov eut proposé une base axiomatique d'une solidité à toute épreuve[3], se substituant au corpus existant. Les relations avec l'expérience et la modélisation ont pu être mises de côté pour développer des outils extrêmement performants[4], parfois très abstraits.

Reste que le but initial de la théorie est malgré tout de fournir un *modèle* mathématique cohérent de la notion de résultat aléatoire. Sans entrer dans le détail de cette notion difficile, précisons simplement quelques points.

---

dont les dés tombent, et *via* le latin *cadere* (qui a donné « choir »), le français du XII^e siècle l'appelait *chaance*, ce qui a donné « chance » en français moderne. Ce terme signifiait donc « hasard », avant de devenir « heureux hasard ». On retrouve ce sens, par exemple, dans *Le Comte de Monte-Cristo* : « ... j'ai renoncé à mon projet comme trop chanceux », ou « Vous risquez la diffamation, ce me semble, et [...] un procès fort chanceux. »

(2). Ce résultat peut sembler étonnant, et montre que l'état de « gagnant » est relativement stable. Il n'est valable que si la partie est suffisamment longue, et est exposé dans le chapitre 33.

(3). Andreï KOLMOGOROV, *Foundations of the theory of Probability* [62], chapitre I, paragraphe §1 pour le cas fini, chapitre II, paragraphe §2 pour le cas général.

(4). L'époque était prête : la théorie de l'intégration avait elle aussi subi de profonds bouleversements avec l'intégrale de Lebesgue et les travaux d'éminents mathématiciens comme Borel, Carathéodory, Radon, etc. Les questions que se posaient les probabilistes étaient finalement très proches : en substance, développer des outils pour mesurer des parties d'un ensemble, et intégrer des fonctions selon cette mesure.

- Une expérience retourne un résultat dit *aléatoire* lorsque l'on ne peut pas prédire ce résultat. Cela peut être dû à notre incapacité à décrire un système trop complexe, comme la trajectoire précise d'un dé lancé sur un tapis, ou encore être une conséquence de la physique sous-jacente, comme dans le cas de la désintégration d'une particule instable. Que la Nature procède ou non du hasard dans son essence même n'est pas notre problème ; c'est pour l'heure une question encore ouverte, aucune expérience ne permettant de trancher la question[(5)].
- Précisons que le terme d'expérience, qui est consacré, est à prendre dans un sens large. Il peut s'agir aussi bien d'un lancer de dé ou d'un tirage à pile ou face, que de l'observation de la désintégration d'un atome radioactif. Il peut également s'agir d'une expérience totalement fictive, décrite de façon purement mathématique.
- Il sera donc supposé qu'une telle expérience peut effectivement être menée un grand nombre de fois dans des conditions semblables, et que le résultat pourra varier d'une expérience à l'autre. On peut alors — au moins d'un point de vue théorique — comparer des fréquences empiriques aux probabilités calculées.
  Notamment, des questions comme « Quelle est la probabilité que la vie apparaisse dans l'Univers ? » ne sont pas concernées par cette approche, et n'ont peut-être aucun sens[(6)].
- On prendra garde à ne pas confondre la notion de prédictibilité d'un système avec celle de déterminisme physique. Un système physique est dit **déterministe** lorsque son évolution dépend uniquement de son état à un instant initial choisi. Un système déterministe mais complexe peut ne pas être prédictible à cause d'une extrême sensibilité aux conditions initiales : plusieurs tirages de dé, même réalisés soigneusement par un automate dans des conditions aussi raisonnablement semblables que possible, donneront des résultats différents. La meilleure modélisation d'un tirage de dé n'est alors pas la modélisation déterministe, vouée à l'échec, mais la modélisation probabiliste.
  Toutefois, le vocable « déterministe » a, pour les mathématiciens probabilistes, un sens particulier. Une variable est dite **déterministe** si elle est constante, par opposition avec les variables **aléatoires** qui dépendent de la variable ω représentant le hasard.

L'univers Ω et les réalisations possibles du hasard §6

Pour modéliser la notion de résultat aléatoire d'une expérience, le plus simple, pour un mathématicien (qui n'a pas à économiser sur le matériel), est de construire des mondes différents, chacun donnant un des résultats possibles de l'expérience

Prenons un exemple simple : on lance un dé à six faces. La main jette le dé, qui décrit une élégante parabole avant de frapper la surface d'une table. À ce moment-là, l'histoire bifurque, et de nombreux futurs sont possibles. Si l'on se contente d'une description partielle du lancer de dé, on a six « futurs possibles »,

(5). Même les expériences d'Alain Aspect concernant les fameuses inégalités de Bell ne prouvent pas la nature intrinsèquement probabiliste de la physique ; des modèles entièrement déterministes de mécanique quantique sont en parfait accord avec les inégalités de Bell et les résultats de l'expérience (« *I saw the impossible done* », dit John Bell lui-même [5] à propos du fameux article de David Bohm de 1952).

(6). Pour la petite histoire, des estimations de cette « probabilité » ont été proposées. Les résultats varient de presque zéro à presque un. Dans un cas comme dans l'autre, cela permet moultes interprétations, évidemment contradictoires.

chacun d'eux étant caractérisé par la face visible une fois le dé immobile. Une description plus complète des rebonds successifs du dé entraînera au contraire une multiplication des futurs possibles. Les futurs possibles sont désignés par la variable générique ω ; ils portent, pour des raisons historiques, de nombreux noms : épreuve, aléa, résultat de l'expérience, **réalisation du hasard**, état du monde, événement élémentaire, événement atomique... Les amateurs de science-fiction peuvent songer à tous les ω comme à des *mondes parallèles.*

Pour des raisons historiques, la collection de toutes ces réalisations du hasard s'appelle **univers des possibles**(7) ou encore **espace des épreuves** et est noté Ω.

Une expérience est modélisée par le choix d'un univers Ω ; dans le cas d'un jet de dé, l'univers le plus simple est $\Omega := \{1, 2, 3, 4, 5, 6\}$, mais on peut souhaiter un univers plus « détaillé ». Cet univers contient donc toutes les « versions » du monde possible, chacune de ces versions illustrant une issue différente à l'expérience réalisée. Le *résultat* de l'expérience est l'un des éléments ω de l'univers Ω des possibles(8).

On prendra bien garde au fait que, entre une expérience réelle et sa transcription mathématique (choix de Ω puis description d'une probabilité sur Ω), il y a une étape de *modélisation.* Cette étape est la seule partie non mathématique de la théorie des probabilités ; elle est évidemment sujette à caution et procède de choix (nous le verrons notamment dans le chapitre consacré à l'indépendance de variables aléatoires). Une fois cette modélisation faite, on peut presque oublier l'aspect intuitif de la théorie des probabilités(9) ; il suffit simplement de s'adapter au vocabulaire local.

Exemple 1.3 Donnons sans plus tarder quelques exemples fondamentaux. Dans le premier, l'univers est fini ; dans le deuxième, il est infini mais dénombrable ; dans le troisième et dans le quatrième, il est non dénombrable(10).

- **Tirage d'un dé** Considérons un personnage qui, à un instant t_0, lance un dé à six faces. On peut, par exemple, décider que les éléments de Ω sont les descriptions complètes du monde réel (avec tous ses atomes, leur évolution dans le temps). De (très) nombreux éléments de Ω correspondent à un même tirage de dé. Bien entendu, une telle description est bien trop complexe et, pour tout dire, inutile.

  De manière beaucoup plus économique, on peut « oublier » la plus grosse partie du monde, n'en conserver que les propriétés essentielles et choisir une modélisation plus simple, dans laquelle on décrète que $\Omega := \{1, 2, 3, 4, 5, 6\}$. C'est évidemment le point de vue qui sera adopté dans la suite.
- **Arrêt d'un programme** On considère un programme acceptant en entrée une série de données de taille arbitraire, et on note T le temps (compté en nombre de calculs élémentaires effectués par l'ordinateur) avant l'arrêt du programme. On pose alors $\Omega := \{1, 2, 3, \ldots\} \cup \{\infty\} = \mathbb{N}^* \cup \{\infty\}$.
- **Succession de tirages à pile ou face** Si l'on accepte l'idée d'une succession infinie de tirages au sort, on prend comme univers l'ensemble $\Omega := \{\mathrm{P}, \mathrm{F}\}^{\mathbb{N}^*}$ des applications de $\mathbb{N}^*$ dans $\{\mathrm{P}, \mathrm{F}\}$, c'est-à-dire des suites $(r_n)_{n \geqslant 1}$ à valeurs dans $\{\mathrm{P}, \mathrm{F}\}$. Dans une suite $\omega = (r_n)_{n \geqslant 1}$ donnée, la valeur de r_k est le résultat du k-ième tirage.
- **Arrivée d'un train** L'instant d'arrivée d'un train est une quantité qui peut varier de manière continue ; on la modélise par un univers du type $\Omega := [0 ; +\infty[$, ou $\Omega := [0 ; \mathrm{T}]$ si l'on est certain que l'arrivée a lieu avant une date ultime T.

---

(7). Ou, en abrégé : **univers**.

(8). Les amateurs de mythologie peuvent imaginer que le Destin — ou Dame Fortune, ou les Nornes, ou Lachesis... — attribue, à chaque fois qu'une expérience est réalisée, une issue ω de son choix.

(9). Presque. Nous ne le ferons pas, ne serait-ce que pour garder le sujet un peu vivant. Les ouvrages traitant la théorie des probabilités comme un cas particulier de la théorie de la mesure et de l'intégration sont nombreux et très bien faits.

(10). L'annexe A1 rappelle la notion de dénombrabilité ainsi que les propriétés élémentaires des ensembles dénombrables (ou non).

Événements

§7

Il arrive fréquemment que l'on ne s'intéresse pas à une épreuve élémentaire ω donnée, mais à une partie de Ω, que l'on appelle **événement** (par exemple, l'événement « le dé sort un nombre pair » est la partie $\{2, 4, 6\}$ de l'univers $\Omega := \{1, 2, 3, 4, 5, 6\}$).

Plus précisément, l'ensemble A des ω vérifiant une propriété $\mathcal{P}$ donnée (comme « ω est un entier pair ») caractérise la réalisation de cette propriété lors d'une expérience, ce qui permet l'identification de cet ensemble et de la propriété. Si ω est élément de A, on dit que **l'événement A se réalise dans l'épreuve ω.**

Très couramment, on ne fait pas mention de ω, et l'on dit simplement que « l'événement A est réalisé ».

Exemple 1.4 Revenons au cas du tirage d'un dé ; l'univers choisi est $\Omega = \{1,2,3,4,5,6\}$. Voici quelques événements :

- **« le résultat est pair »** qui est l'événement $\{2, 4, 6\}$. On dira alors que, lors d'une expérience de lancer de dé (c'est-à-dire lors du choix d'une réalisation particulière ω du hasard), « le résultat est pair » pour signifier « $\omega \in \{2, 4, 6\}$ ».
- **« le résultat est impair »** qui est bien sûr $\{1,3,5\}$.
- **« le résultat est supérieur à 5 »** qui est l'événement $\{5, 6\}$.
- **l'événement certain** qui est l'événement Ω tout entier. Il apparaît de façon naturelle (par exemple « le résultat est inférieur à 42 »).
- **l'événement impossible** qui est l'événement $\varnothing$: aucun résultat d'expérience n'en fait partie. Lui aussi apparaît de façon naturelle (par exemple : « le résultat est négatif »).

Toutes les parties de Ω (il y en a $2^6 = 64$) sont des événements.

Compter/mesurer

§8

Une fois défini un univers, ensemble Ω dont les éléments sont des réalisations possibles du monde, on cherche à évaluer, *quantitativement*, celles qui sont plus ou moins susceptibles de se produire. On cherche ainsi à associer à chaque ω une grandeur numérique, modèle mathématique de la notion intuitive de probabilité. Deux situations se présentent, que nous appellerons génériquement le *cas discret* et le *cas continu.*

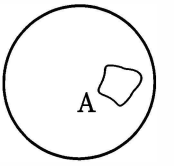
- *Cas discret* : dans le cas où Ω est un ensemble fini ou dénombrable, on associe, à chaque réalisation ω du monde, un nombre $p(\omega)$. Si A est un événement, on lui associe alors une probabilité $\mathbf{P}(\mathrm{A})$ en *comptant* : la probabilité de l'événement est la somme des probabilités des événements élémentaires le composant (voir section 1.2) :

$$\mathbf{P}(\mathrm{A}) = \sum_{\omega \in \mathrm{A}} p(\omega). \qquad (\star)$$

  Le cas le plus simple et le plus intuitif est celui d'une expérience ne menant qu'à un nombre *fini* de résultats possibles, tous équiprobables (cas d'un dé équilibré). Si l'on note n le cardinal de Ω, on a donc $p(\omega) = 1/n$ pour tout ω. Un événement A a ainsi pour probabilité $\mathbf{P}(\mathrm{A}) = \mathrm{Card}(\mathrm{A})/n$ (nombre de cas favorables sur nombre total de cas).
- *Cas continu* : dans le cas où Ω est non dénombrable, des difficultés apparaissent.
  Considérons le résultat d'un tir de fléchette sur une cible ronde, assimilée à un disque D de rayon R du plan. L'issue de l'expérience est un point quelconque ω de D, et l'univers est simplement $\Omega = \mathrm{D}$. On veut modéliser la situation où la fléchette atteint la cible avec une probabilité *uniforme.* Ce que l'on entend par là, c'est qu'il y a une chance sur deux que la fléchette atteigne la moitié supérieure comme la moitié inférieure, ou bien celle de

droite comme celle de gauche ; et que, plus généralement, si l'on délimite une partie A du disque, la probabilité que la fléchette atteigne un point de A est proportionnelle à l'aire de A, et plus précisément égale à

$$\mathbf{P}(\mathrm{A}) = \frac{\mathrm{Aire}(\mathrm{A})}{\mathrm{Aire}(\mathrm{D})} = \frac{1}{\pi\,\mathrm{R}^2} \cdot \mathrm{Aire}(\mathrm{A}).$$

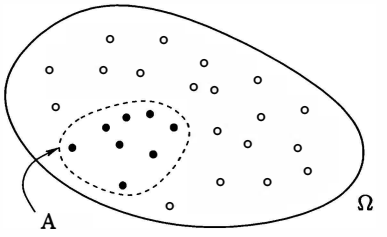


Considérons un unique point ω du disque ; la probabilité que la fléchette tombe exactement sur ce point est nulle : $p(\omega) = 0$. C'est en effet l' « aire » d'un point[(11)]. Si maintenant l'on cherche à écrire une formule telle que $(\star)$, ou si l'on écrit que la somme des probabilités est égale à 1, on tombe sur un os :

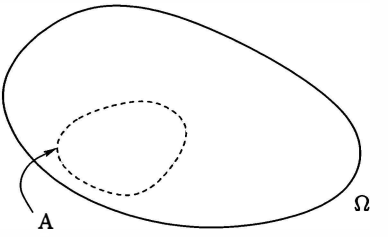
$$1 = \mathbf{P}(\mathrm{D}) = \text{« } \sum_{\omega \in \mathrm{D}} p(\omega) \text{ » } = \text{« } \sum_{\omega \in \mathrm{D}} 0 \text{ ».}$$

Contrairement au cas discret, il n'est en général pas possible de décrire la fonction de probabilité en prenant pour éléments de base les événements élémentaires ω (qui, dans l'exemple donné, sont tous « de probabilité nulle »). Une description cohérente exige que l'on se concentre sur les parties de Ω, les *événements*. On ne peut plus alors simplement compter, comme avec la formule $(\star)$, on doit *mesurer* des événements $\mathrm{A} \subset \Omega$.
Pour cela, l'outil « naturel » est évidemment l'intégrale[(12)], généralisation continue d'une somme discrète ; de fait, on pourra n'utiliser que des connaissances d'analyse élémentaire pour étudier des probabilités sur $\mathbb{R}$ ou sur $\mathbb{R}^n$ ou, plus généralement, des variables aléatoires « continues » (en fait, des variables « à densité », qui seront l'objet du chapitre 10).

Dans un cas plus général, on peut utiliser la théorie de la mesure, dont l'introduction par Kolmogorov et Fréchet dans la théorie des probabilités a véritablement révolutionné la discipline ; les difficultés techniques de cette méthode risquant de faire perdre de vue les idées importantes, on ne l'introduira que tardivement, afin que le lecteur ait le temps de se familiariser avec les concepts, le langage, les idées et les méthodes des probabilités.



Compter



Mesurer

Remarque 1.5 Sur un ensemble non dénombrable, la situation sera d'autant plus compliquée que, pour des raisons techniques incontournables, il n'est pas possible de définir **P** sur l'ensemble

(11). On peut également inclure ce point dans une suite de disques de rayons tendant vers 0, et majorer la probabilité cherchée par une suite de limite nulle.

(12). Elle n'est pas cachée bien loin dans l'exemple de la cible et de la fléchette : l'aire d'une partie du plan peut être définie comme une intégrale double. Les problèmes inhérents à la notion d'intégrabilité d'une partie quelconque sont momentanément laissés de côté.

$\mathfrak{P}(\Omega)$ des parties de Ω, mais sur une partie $\mathfrak{T}$ (« T » gothique, comme *Tribu*) plus restreinte. Nous n'en disons pas davantage pour l'instant ; voir la discussion de la remarque 3.19 page 83 et l'exemple de Vitali présenté dans l'annexe 3.5 page 95.

Langage ensembliste §9

La définition d'un événement comme partie de Ω nous permet de transcrire, en langage ensembliste, certaines locutions intuitives. Ainsi, la phrase : *l'événement* A *est réalisé* se traduit simplement par « ω est élément de A ».

Exemple 1.6 Notons $\Omega := \{1, 2, \ldots, 2n\}$, et $A := \{2, 4, \ldots, 2n\}$. Pour une réalisation ω du hasard donnée, dire « le résultat est pair » équivaut à affirmer « $\omega \in A$ ».

De même, la phrase « A implique B » se traduit par

$$\forall \omega \in \Omega \qquad \omega \in A \implies \omega \in B$$

c'est-à-dire $A \subset B$. La notion d'implication[13] se traduit donc en terme d'inclusion.

Exemple 1.7 Notons $\Omega := \{1, 2, \ldots, 16\}$ et construisons les événements $D := \{2, 4, \ldots, 16\}$ (« le résultat est pair ») et $Q := \{4, 8, 12, 16\}$ (« le résultat est multiple de 4 »). Être multiple de 4 implique le fait d'être pair, ce qui se traduit par $Q \subset D$.

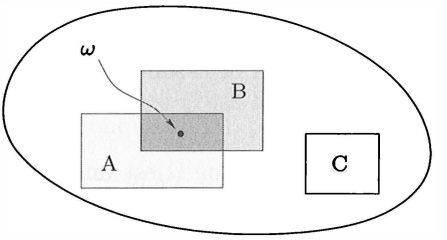


FIGURE 1.1 — *Lorsque le résultat de l'expérience est l'événement élémentaire* ω, *les événements* A *et* B *sont réalisés, mais pas l'événement* C. *Aucun résultat de l'expérience n'amène la réalisation des événements* A *et* C *: ils sont incompatibles.*

Le tableau 1.1 résume ces équivalences[14], qui seront très largement exploitées, dans le cas discret comme dans le cas continu.

Langage probabiliste	Notation	Langage ensembliste
épreuve	ω ($\omega \in \Omega$)	élément
événement	A ($A \subset \Omega$)	partie de Ω
A est réalisé	$\omega \in A$	ω est élément de A
A implique B	$A \subset B$	A inclus dans B
A ou B	$A \cup B$	réunion
A et B (simultanéité)	$A \cap B$	intersection
non-A (événement contraire)	A^c ou $\Omega \setminus A$	complémentaire[15]
A mais pas B	$A \setminus B := A \cap B^c$	différence
événement impossible	$\varnothing$	partie vide
événement certain	Ω	
événements incompatibles	$A \cap B = \varnothing$	parties disjointes

TABLE 1.1 — *Langage des probabilités et langage ensembliste correspondant.*

(13). Qu'on ne confondra pas avec de la causalité, et qui ne signifiera rien de plus que : chaque fois que l'événement A est réalisé, l'événement B est aussi réalisé. Il ne faut pas nécessairement y voir un lien « de cause à effet » entre les événements A et B.

(14). Un tableau plus complet est donné page 717.

On rappelle quelques propriétés des opérations sur les événements :

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \qquad A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup A^c = \Omega \qquad A \cap A^c = \varnothing$$
$$(A^c)^c = A \qquad \Omega^c = \varnothing$$
$$(A \cup B)^c = A^c \cap B^c \qquad (A \cap B)^c = A^c \cup B^c,$$

ces deux dernières égalités portant le nom de *règles de de Morgan*[16].

§ 10 Calcul des probabilités

Une fois défini notre univers, nous devons maintenant répondre à la question : quelles sont les « versions du monde » qui sont les plus probables ? Plus précisément, quelle probabilité associer, si possible, à chaque élément ω de Ω ou, du moins, à chaque événement A ?

Pour ce faire, on définit une fonction **P** sur l'ensemble des événements (que nous noterons $\mathfrak{T}$ comme « tribu ») et qui vérifie, au minimum, les propriétés élémentaires suivantes, qui ne font que reprendre l'idée intuitive que l'on se fait d'une probabilité (« nombre de cas favorables sur nombre de cas total ») :

- ⋆ **Axiome *a* :** **P** est à valeurs dans $[0\,;1]$;
- ⋆ **Axiome *b* :** l'univers Ω est de probabilité unité : $\mathbf{P}(\Omega) = 1$;
- ⋆ **Axiome *c* :** **P** est additive : si A et B sont des événements disjoints, alors la probabilité associée à leur union est la somme de leurs probabilités : $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.

Cette liste d'axiomes est suffisante lorsque Ω est un ensemble fini ; nous devrons la compléter par un axiome supplémentaire lorsque Ω est infini.

De ces axiomes, on déduit immédiatement quelques propriétés importantes, dont la preuve est laissée en exercice :

■ **Proposition 1.8 (Propriétés élémentaires)** *Soient* A *et* B *deux parties de* Ω.

i) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

ii) $\mathbf{P}(\varnothing) = 0$.

iii) $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.

iv) *Si* $A \subset B$, *alors* $\mathbf{P}(A) \leqslant \mathbf{P}(B)$.

Ces relations permettent de *calculer* des probabilités — inconnues — en fonction d'autres probabilités — connues.

Avant de terminer, quelques mots de vocabulaire :

Définition 1.9 Un événement A est dit **négligeable** s'il est de probabilité nulle : $\mathbf{P}(A) = 0$. Il est dit **presque sûr** s'il est de probabilité 1. Une propriété est dite vraie **presque sûrement** si l'ensemble des éléments de Ω vérifiant cette propriété est un événement de probabilité 1.

(15). La notation $\overline{A}$ est également employée pour désigner l'événement contraire.

(16). Augustus DE MORGAN (1806–1871), mathématicien et logicien britannique.

1.2 PROBABILITÉS SUR UN ENSEMBLE FINI OU DÉNOMBRABLE

Nous l'avons dit : le Calcul des Probabilités est l'art de calculer des probabilités inconnues, en fonction de probabilités connues. Pour modéliser mathématiquement une expérience, on choisit un univers Ω, puis on le munit d'une probabilité $\mathbf{P}$, qui est une application de l'ensemble des événements dans $[0\,;\,1]$. La construction générale d'une telle fonction est expliquée ici dans le cas où Ω est fini ou dénombrable.

Univers fini § 11

Munir un univers fini $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ d'une probabilité revient à associer, à chaque élément ω_k, un réel $p_k \geqslant 0$, avec pour seule contrainte que $p_1 + p_2 + \cdots + p_n = 1$. Insistons sur le fait que l'on *choisit* effectivement ces nombres, et qu'on ne les calcule pas. Le choix peut être plus ou moins pertinent, c'est-à-dire que la modélisation est plus ou moins bien réussie.

Exemple 1.10 Dans le cas d'un lancer de dé, on peut choisir $\Omega = \{1, 2, 3, 4, 5, 6\}$ et poser $p_k = 1/6$ pour $k = 1, \ldots, 6$; cette situation modélise un dé équilibré. Si l'on se rend compte que, pour un dé donné, les prédictions du modèle ne se réalisent pas, on peut modifier les valeurs des coefficients p_k. Pour déterminer expérimentalement chaque coefficient, une méthode est de lancer un très grand nombre de fois le dé et de calculer des fréquences expérimentales. Des résultats théoriques permettent de donner une estimation du nombre de lancers à effectuer pour obtenir des valeurs suffisamment précises des p_k (voir chapitre 22).

Très peu de mystère (mathématique) là-dessous, et nul besoin d'une théorie élaborée : l'addition et la multiplication suffisent. En effet, si A est un événement (c'est-à-dire une partie quelconque de Ω), poser

$$\mathbf{P}(\mathrm{A}) := \sum_{k\,;\,\omega_k \in \mathrm{A}} p_k$$

suffit à définir une probabilité répondant aux exigences des axiomes *a* à *c*.

Il ne faudrait pas croire cependant que tout soit simple : même dans des univers finis, des difficultés peuvent surgir pour résoudre un problème de probabilités, et tout d'abord au niveau de la modélisation.

Par exemple, dans la modélisation d'une succession (finie) de tirages à *pile* ou *face* avec une pièce équilibrée, on *postule* l'indépendance des tirages — parce que c'est ce qui est observé en pratique, et que toutes les expériences le confirment. On modélise donc cette succession de tirages par l'univers $\Omega := \{\mathrm{P}, \mathrm{F}\}^n$, que l'on munit de la probabilité uniforme : chaque élément particulier se voit attribuer la probabilité $1/2^n$. Cependant, le monde aurait pu être différent[17] ; si un jour des expériences démontrent que le tirage numéro un milliard n'est pas indépendant du tirage numéro un, il faudra bien évidemment revoir ce modèle. Pour l'instant, personne n'a eu le courage de lancer un milliard de fois la pièce.

Remarque 1.11 (L'erreur de d'Alembert) D'Alembert, qui eut l'esprit très clair dans de nombreux domaines des mathématiques et de la physique, s'était aventuré sans précautions dans le domaine des probabilités à une époque où pratiquement aucun formalisme n'existait, et où les raisonnements considérés comme basiques de nos jours n'étaient pas toujours clairs dans l'esprit de la communauté scientifique.

Ainsi, à la question :

(17). Dans la pièce *Rosencrantz and Guildenstern are dead*, de Tom Stoppard [96], le monde dans lequel évoluent les deux personnages principaux ne suit manifestement pas les règles usuelles des probabilités ; dès la première scène, où une interminable partie de *pile* ou *face* a lieu, une centaine de tirages donne *face* et Guildenstern vide sa bourse au profit de Rosencrantz. S'ensuivent évidemment de nombreuses réflexions des protagonistes quant à la nature des probabilités — et, plus généralement, quant à la nature de l'univers.

Quelle est la probabilité, lorsque l'on effectue deux tirages à pile *ou* face, *d'obtenir, au moins une fois, le résultat* pile *?*

d'Alembert argumenta qu'il n'y a que trois possibilités de tirage :

— deux fois *face* ;
— un *face* puis un *pile* ;
— un *pile* (auquel cas on ne continue pas le tirage !).

Il avança que ces trois possibilités étaient équiprobables, pour conclure finalement que la probabilité d'avoir au moins un *pile* était de 2/3.

Ce résultat est contredit par l'expérience : dans un grand nombre de paires de lancers, la proportion de tirages contenant au moins un *pile* converge (assez rapidement) vers 3/4. C'est le résultat auquel conduit le modèle présenté précédemment, à savoir l'univers $\Omega = \{\mathrm{P}, \mathrm{F}\}^2$ muni de la probabilité uniforme :

$$\mathbf{P}(\mathrm{PF}) = \mathbf{P}(\mathrm{PP}) = \mathbf{P}(\mathrm{FP}) = \mathbf{P}(\mathrm{FF}) = \frac{1}{4}.$$

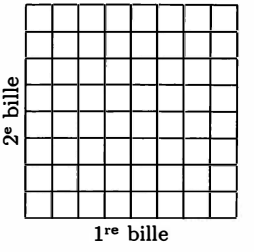
L'événement recherché est le complémentaire de « FF », de probabilité $1 - \mathbf{P}(\mathrm{FF}) = 3/4$.

Bien sûr, on pourrait rétorquer que trouver le bon espace et la bonne probabilité pour modéliser une succession de n tirages à *pile* ou *face* n'est pas une tâche bien compliquée — et on peut même légitimement s'étonner de l'erreur de d'Alembert, qui nous paraît bien naïve aujourd'hui. Voici un second problème. On se donne N boîtes (numérotées de 1 à N), et on permet à deux billes de s'y ranger, de manière aléatoire et sans que les billes interagissent physiquement entre elles. On demande la probabilité que les billes se retrouvent toutes deux dans la même boîte.

Le raisonnement suivant paraît convaincant :

> La probabilité qu'elles soient toutes deux dans la boîte 1 est égale à $1/\mathrm{N}^2$; la probabilité qu'elles soient toutes deux dans la boîte 2 est encore $1/\mathrm{N}^2$; et ainsi de suite jusqu'à la boîte numérotée N. Ces événements étant deux à deux incompatibles, la probabilité que les particules soient dans la même boîte est la somme des probabilités calculées, et donc égale à $1/\mathrm{N}$.

Si l'on effectue soigneusement l'expérience un grand nombre de fois, alors — surprise ? — on trouve une proportion de résultats positifs (« les deux billes sont dans la même boîte ») qui tend vers $1/\mathrm{N}$, comme prévu par le raisonnement que nous avons présenté, et qui revient à adopter la formalisation suivante : on choisit comme univers l'ensemble $\Omega = \{1, 2, \ldots, n\}^2$, muni de la probabilité uniforme. Un couple (i, j) représente l'état du système décrit par « la première bille est dans la boîte i, la deuxième est dans la boîte j ». Les résultats positifs sont colorés en gris sur le schéma ci-dessous ; la proportion est bien $1/\mathrm{N}$.



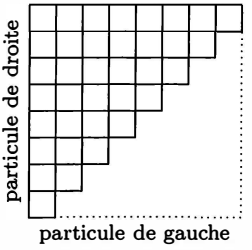
Un point crucial dans ce raisonnement est que l'on a pu différencier les deux billes, c'est-à-dire décider que l'une s'appelle *la première* et l'autre *la seconde* : alors, dans le processus de comptage, les états (i, j) et (j, i) sont distincts lorsque $i \neq j$, si bien que le nombre total d'états possibles est N^2. Souvent, pour bien montrer que les billes sont différenciables, on dira que l'une est, par exemple, rouge, tandis que la seconde est verte. Mais, même si l'on utilisait des billes aussi identiques

que possible, le raisonnement précédent resterait valable et en adéquation avec l'expérience.

Cependant, si l'on effectue une expérience similaire au niveau microscopique, un changement spectaculaire intervient. Remplaçons les boîtes par des « états quantiques », et les billes par des *bosons*, c'est-à-dire des particules obéissant à la statistique de Bose-Einstein[18]. Les deux particules sont maintenant *indiscernables*, et ce que dit la physique, c'est que dans la manière de compter les configurations distinctes, on ne doit prendre en compte qu'une seule fois la configuration (i,j) et la configuration (j,i). Autrement dit : la première particule dans la boîte i et la seconde dans la boîte j, ou bien la première dans la boîte j et la seconde dans la boite i, c'est le *même état.*

Remarque 1.12 Insistons sur le fait que la nature de la physique est très différente au niveau microscopique de ce que l'on connaît au niveau macroscopique. Fabriquer des billes macroscopiques aussi semblables que possible ne les rendra indiscernables que pour un œil humain, avec toutes ses imperfections, mais ne modifiera en rien la méthode de comptage que l'on *doit* employer pour rendre compte des propriétés statistiques du système. À l'inverse, au niveau des particules élémentaires, l'indiscernabilité est intrinsèque : non seulement il est impossible de « marquer » deux photons, ou de les peindre d'une couleur chacun, mais la Nature elle-même ne fait pas la différence entre ces deux particules et « compte » comme un même état les configurations (i,j) et (j,i).

Faisons alors un grand nombre de fois l'expérience de mettre les particules au hasard dans les boîtes : la proportion de résultats positifs tend vers $2/(\mathrm{N}+1)$. Il est important de voir qu'aucune interaction entre les particules n'a été rajoutée ; simplement, la Nature compte les états différemment — c'est sans doute étonnant, mais c'est ainsi — et l'univers permettant la modélisation de cette situation n'est plus l'ensemble des couples (i,j), où i et j parcourent $\{1,2,\ldots,n\}$, mais l'ensemble, plus restreint, des couples (i,j) avec la contrainte supplémentaire $i \leqslant j$. L'indice i est celui de « la particule la plus à gauche », et l'indice j celui de la particule « la plus à droite » (avec possibilité d'avoir $i=j$).



La figure ci-dessus illustre alors le fait que le nombre de configurations où les deux particules sont dans la même boîte est N, sur un total de $\mathrm{N}(\mathrm{N}+1)/2$ configurations ; d'où une probabilité de $2/(\mathrm{N}+1)$.

Le chapitre 29 page 589 expose d'autres conséquences de la statistique bosonique.

Exemple 1.13 (Polynômes de $\mathbb{F}_q[\mathrm{X}]$ agissant bijectivement (*)) Voici un exemple un peu plus élaboré du fait que la modélisation n'est pas toujours une étape évidente ; il est tiré de l'algèbre, mais la partie probabiliste est en fait très simple. Soit $\mathbb{F}_q$ un corps fini ; on considère l'ensemble

(18). Exemple de telles particules : des mésons, des photons. Il n'est pas facile de mettre des photons dans des boîtes, et les expériences réelles diffèrent sensiblement de la grossière modélisation que l'on donne ici, mais l'idée générale est la bonne.

V des polynômes de degré au plus égal à $q-1$, à coefficients dans $\mathbb{F}_q$. On le munit de la probabilité uniforme, c'est-à-dire qu'à chacun des q^q polynômes de V est associé la probabilité $1/q^q$. On demande la probabilité qu'un polynôme de V agisse bijectivement sur $\mathbb{F}_q$. Si l'on note W l'ensemble des polynômes bijectifs, on cherche donc $\mathbf{P}\{\mathrm{P} \in \mathrm{W}\}$.

On définit pour cela une application de V dans $\mathbb{F}_q{}^q$ par

$$\varphi : \mathrm{P} = \sum_{k=0}^{q-1} a_k \mathrm{X}^k \longmapsto \big(\mathrm{P}(x_0), \mathrm{P}(x_1), \ldots, \mathrm{P}(x_{q-1})\big),$$

où $x_0, x_1, \ldots, x_{q-1}$ sont les q éléments du corps $\mathbb{F}_q$. Un polynôme de V qui admet q racines distinctes étant forcément nul, l'application φ est injective entre deux ensembles de même cardinal, elle est donc bijective. Enfin, un polynôme agit bijectivement si et seulement si son image par φ est dans l'ensemble D des q-uplets d'éléments distincts, de cardinal $q!$. En conclusion, $\mathbf{P}\{\mathrm{P} \in \mathrm{W}\} = \mathbf{P}\{\varphi(\mathrm{P}) \in \mathrm{D}\} = q!/q^q$.

§ 12 Probabilités sur un ensemble dénombrable

On suppose ici que Ω est un ensemble *infini*, mais *dénombrable* : on peut donc numéroter chacun des éléments de Ω, c'est-à-dire écrire que

$$\Omega = \{\omega_n \ ; \ n \in \mathbb{N}\}.$$

On considérera fréquemment les cas particuliers $\Omega = \mathbb{N}$ ou $\Omega = \mathbb{N}^*$.

Associer une probabilité à chacun des éléments ω de Ω revient donc à se donner une suite $(p_n)_{n \geqslant 0}$ de réels positifs, tels que la série $\sum p_n$ converge et

$$\sum_{n=0}^{\infty} p_n = 1.$$

On définit ensuite une application $\mathbf{P} : \mathfrak{P}(\Omega) \to \mathbb{R}^+$ sur l'ensemble des parties de Ω de la manière suivante :

— pour tout événement atomique ω_k, on pose $p(\omega_k) := p_k$;

— si A est une partie de Ω, on définit la probabilité de A comme la somme des probabilités associées à chacun des éléments de A :

$$\mathbf{P}(\mathrm{A}) := \sum_{\omega \in \mathrm{A}} p(\omega) = \sum_{\substack{k \text{ tel que} \\ \omega_k \in \mathrm{A}}} p_k.$$

Dans cette dernière expression, la convergence absolue de la série $\sum p_n$ nous assure de la convergence de la série restreinte à un sous-ensemble A de Ω.

Notamment, $\mathbf{P}\big(\{\omega\}\big) = p(\omega)$ pour tout événement atomique ω.

Remarque 1.14 Par commodité, on écrit souvent $p_k = \mathbf{P}(\omega_k)$ au lieu de $p_k = \mathbf{P}\big(\{\omega_k\}\big)$. C'est conforme à l'idée que l'on peut se faire intuitivement de la fonction de probabilité, que l'on définit d'abord pour chacun des éléments de Ω.

1.3 PROBABILITÉS SUR $\{\mathrm{P}, \mathrm{F}\}^{\mathbb{N}^*}$

§ 13 Un univers non dénombrable

Considérons le cas d'une succession infinie de tirages à *pile* ou *face*, la pièce étant réputée honnête — c'est-à-dire que, pour chaque tirage, on supposera que la probabilité d'obtenir *pile* vaut $1/2$, tout comme la probabilité d'obtenir *face*.

Une expérience est ici, par définition, une telle succession de tirages. Le résultat d'une expérience est donc un élément de $\{\mathrm{P}, \mathrm{F}\}^{\mathbb{N}^*}$, c'est-à-dire une suite indexée par $\mathbb{N}^* = \{1, 2, 3, \ldots\}$ et à valeurs dans $\{\mathrm{P}, \mathrm{F}\}$. Si ω est une telle suite, on notera $u_n(\omega)$ son n-ième terme ; ainsi, $\omega = \big(u_n(\omega)\big)_{n \geqslant 1}$.

■ **Proposition 1.15** *L'ensemble* $\{\mathrm{P},\mathrm{F}\}^{\mathbb{N}^*}$ *n'est pas dénombrable.*

La démonstration de ce résultat suit un argument classique, extrêmement ingénieux, appelé « diagonale de Cantor » ; il est exposé page 667. Le fait que Ω ne soit pas dénombrable a des conséquences remarquables.

Tout d'abord, notons que chaque suite ω est « aussi probable qu'une autre ». On cherche donc à munir Ω d'une probabilité uniforme[19]. C'est ici qu'apparaît le premier problème : chaque événement atomique $\{\omega\}$ est de probabilité nulle[20]. Puisque nous ne pouvons pas caractériser la probabilité uniforme par les événements atomiques, il est nécessaire de considérer des événements beaucoup plus gros, contenant chacun une infinité (non dénombrable !) d'événements élémentaires.

Notons symboliquement avec une étoile « $*$ » (**joker**) un tirage quelconque. La notation

$$\big[\mathrm{P}\,\mathrm{F}\,\mathrm{F}\,\mathrm{P}\,\mathrm{F}\,\mathrm{F} * * * \cdots * \cdots\big]$$

représente l'ensemble des tirages commençant par la succession PFFPFF, et dont la suite est quelconque. Un tel événement est de probabilité $1/2^6$. De même, si un événement s'écrit, symboliquement, comme une succession de symboles de l'ensemble $\{\mathrm{P},\mathrm{F},*\}$, avec n symboles P ou F, le reste étant constitué de *jokers*, on lui attribue naturellement une probabilité $1/2^n$. Pouvons-nous, à partir de ces quelques propriétés, déterminer les probabilités de tous les événements qui peuvent nous intéresser[21] ?

Par exemple, considérons un événement atomique $\{\omega\}$, où $\omega = (x_n)_{n\geqslant 1}$ est une suite fixée. Pour tout entier n, l'inclusion

$$\{\omega\} \subset \big[x_1 \cdots x_n * * \cdots * \cdots\big]$$

montre que $\mathbf{P}\big(\{\omega\}\big) \leqslant 1/2^n$, et donc que

$$\mathbf{P}\big(\{\omega\}\big) = 0$$

comme annoncé. Pour effectuer ce raisonnement, nous n'avons pas eu besoin de plus que les axiomes a à c et leurs conséquences immédiates.

---

(19). Uniforme au sens où toutes les suites de tirages jouent un rôle équivalent, la suite constante « FFFFFF$\cdots$ » comme toute autre suite « générique » que l'on aura écrite à l'avance. Si l'on ne peut se défaire de l'impression que la suite constante en *face* est extrêmement improbable — Rosencrantz et Guildenstern spéculent longuement sur la signification d'un tel résultat — c'est parce que cette suite est particulière, non au niveau des probabilités, mais de sa définition : il faut très peu de mots pour la décrire entièrement. Alors qu'une suite « générique », justement, nous ne pouvons pas la décrire : toute régularité (essentiellement, le fait de la décrire par un algorithme fini) nous paraîtrait suspecte. L'étude approfondie des suites à valeurs dans $\{\mathrm{P},\mathrm{F}\}$ montre que l'ensemble des suites « génériques », qui ne peuvent être décrites par un algorithme fini, est de probabilité 1, et que donc l'ensemble de toutes les suites possédant au contraire une certaine régularité est de probabilité nulle.

(20). On peut donner un argument rapide : une somme de réels positifs, étendue à une famille non dénombrable, ne peut être finie que si tous les réels sont nuls, à l'exception peut-être d'une quantité dénombrable ; puisque tous les ω se valent, ils sont tous de probabilité nulle. Cet argument ne se généralise pas au cas d'une pièce truquée, aussi nous établirons la propriété « $\mathbf{P}\big(\{\omega\}\big) = 0$ » de manière plus formelle.

(21). Notons bien que tout le travail fait jusque là est un travail de *modélisation*. Nous n'avons pas démontré de formules du genre : « la probabilité associée à l'événement $\big[\mathrm{PF} * \mathrm{F} * * \mathrm{P} * \cdots * \cdots\big]$ est $1/2^4$ » ; nous les avons *postulées*. Nous ne pouvons pas les démontrer, sauf à faire de nouvelles hypothèses sur la nature des tirages et leur absence de liens, hypothèse que l'on nomme « indépendance ». À ce stade, il est tout aussi simple de définir, conformément à l'intuition, des probabilités *a priori* sur les événements qui nous intéressent, en vérifiant que ces probabilités sont bien compatibles avec les axiomes énoncés précédemment.

§ 14 Unions dénombrables d'événements

Cherchons un événement un peu plus compliqué : *quelle est la probabilité de l'événement* A *décrit par la phrase « la suite des tirages n'est pas stationnaire*[22] *» ?*

Pour cerner A, notons, pour tout k entier, S_k l'événement : « la suite est stationnaire à partir du rang k, mais pas à partir d'un rang inférieur ». Si $k \geqslant 2$, l'événement S_k est donc l'ensemble des tirages tels que :

— au rang $k-1$, on a eu un « F », et ensuite uniquement des « P » ;

— ou, au contraire, au rang $k-1$, on a eu un « P », et ensuite uniquement des « F ».

On peut écrire S_k sous la forme symbolique de l'union disjointe

$$S_k = [\underbrace{* \ * \ \cdots \ *}_{k-1 \text{ fois}} \ P \ F \ F \ \cdots \ F \ \cdots] \ + \ [\underbrace{* \ * \ \cdots \ *}_{k-1 \text{ fois}} \ F \ P \ P \ \cdots \ P \ \cdots].$$

Chacun des deux événements du membre de droite est de probabilité nulle (suivant le même raisonnement que précédemment), donc S_k est également de probabilité nulle. De même, S_1 est évidemment de probabilité nulle. Or, les suites non stationnaires sont celles qui ne sont ni stationnaires à partir de 1, ni stationnaires à partir de 2, etc., c'est-à-dire que

$$A = \Omega \setminus \left(\bigcup_{k=1}^{\infty} S_k \right)$$

On a donc besoin de définir la probabilité de $\bigcup_{k=1}^{\infty} S_k$; celle-ci, malheureusement, ne se déduit pas de l'application des trois axiomes du paragraphe § 10, l'union disjointe d'une infinité d'ensembles ne se déduisant pas d'un processus d'unions finies. Afin d'attribuer une probabilité nulle à l'événement $\bigcup_{k=0}^{\infty} S_k$, comme le veut l'intuition[23], on introduit un axiome supplémentaire :

★ **Axiome c' : P est dénombrablement additive** : si $(A_n)_{n \geqslant 0}$ est une suite d'événements deux à deux disjoints, alors la probabilité de leur union est la somme (au sens des séries) de leurs probabilités :

$$\mathbf{P}\left(\bigcup_{n=0}^{\infty} A_n \right) = \sum_{n=0}^{\infty} \mathbf{P}(A_n).$$

On notera que la série du terme de droite est nécessairement convergente[24]. La propriété de l'axiome c' s'appelle également la « **σ-additivité** », la lettre σ désignant ici la propriété de dénombrabilité de la famille d'événements disjoints. Elle implique évidemment l'axiome c (il suffit de noter $A_0 := A$, $A_1 := B$ et $A_n := \varnothing$ pour tout $n \geqslant 2$).

---

(22). Rappelons qu'une suite est **stationnaire** s'il existe un rang à partir duquel elle est constante.

(23). Quelle somme accepteriez-vous de parier sur le fait qu'une suite infinie de tirages à *pile* ou *face* sera stationnaire ?

(24). Pour tout $n \in \mathbb{N}$, on a la majoration

$$\sum_{k=0}^{n} \mathbf{P}(A_k) = \mathbf{P}\left(\bigcup_{k=0}^{n} A_k \right) \leqslant 1.$$

La suite des sommes partielles est croissante (tous les termes sont positifs) et majorée, elle converge donc. On remarquera que le membre de gauche est invariant par permutation des A_n, et donc que le membre de droite l'est également ; c'est une propriété bien connue des séries de réels positifs.

Unions non dénombrables d'événements § 15

Pouvons-nous pousser notre chance, et étendre la notion d'additivité à des familles plus grandes que les familles dénombrables ? La réponse est « non ». En effet, l'univers Ω lui-même peut s'écrire sous la forme d'une union

$$\Omega = \bigcup_{\omega \in \Omega} \{\omega\},$$

cette union étant disjointe. On sait de plus que $\mathbf{P}\big(\{\omega\}\big) = 0$ pour tout ω de Ω. Si la propriété d'additivité s'étendait aux familles non dénombrables, on aurait alors

$$\mathbf{P}(\Omega) = \text{« } \sum_{\omega \in \Omega} \mathbf{P}\big(\{\omega\}\big) \text{ » } = 0,$$

en contradiction avec l'axiome *b*. Ainsi, *la propriété d'additivité est vraie pour une union finie ou dénombrable — mais pas davantage.*

Probabilité sur $\{\mathbf{P}, \mathbf{F}\}^{\mathbb{N}}$ § 16

Pour l'instant, nous n'avons construit une fonction de probabilité que sur une partie $\mathfrak{F}$ très restreinte de l'ensemble $\mathfrak{P}(\Omega)$, à savoir l'ensemble des tirages dont on spécifie un nombre *fini* de termes.

Un théorème (connu sous le nom de *théorème de Carathéodory*, voir page 94) nous assure que cette probabilité s'étend, de manière unique, sur la plus petite tribu d'événements contenant les parties précédentes[25]. C'est là la première occurrence d'un tel théorème, qui se révélera crucial dans la suite.

1.4 PROBABILITÉS CONDITIONNELLES

Un exemple simple § 17

Massetard et Fortdubras sont des fournisseurs d'arbalètes pour le Guet d'Ankh-Morpork. 30 % des achats viennent de chez Massetard, et 70 % de chez Fortdubras. Seules 10 % des arbalètes de Massetard sont défectueuses, alors que 30 % de celles provenant de chez Fortdubras le sont.

Ainsi, le tableau de répartition des arbalètes selon leur provenance (Massetard ou Fortdubras) et leur état (défectueux ou non) est le suivant :

	Massetard	Fortdubras
défectueux	0,03	0,21
non défectueux	0,27	0,49

On se pose alors la question : *quelle est la probabilité qu'une arbalète vienne de chez Fortdubras, sachant qu'elle est défectueuse ?*

Dans cet exemple, les valeurs numériques associées aux « probabilités » sont simplement celles des « proportions » entre une classe d'arbalètes et le nombre total d'arbalètes ; il s'agit donc d'un simple comptage. Remplaçons donc le tableau précédent par un nouveau tableau, relatif à 100 arbalètes :

(25). La signification précise de cette phrase sera précisée dans le chapitre 3 ; la tribu en question, appelée *tribu engendrée* par $\mathfrak{F}$, comprend *notamment* tous les événements que l'on peut décrire en termes d'intersections ou d'unions, finies ou dénombrables, d'éléments de $\mathfrak{F}$.

Cette tribu est également la tribu engendrée par les événements de la forme $\mathrm{F}_n = \{\omega \,;\, \omega_n = \mathrm{F}\}$; il n'est pas évident *a priori* que cette tribu soit différente de $\mathfrak{P}(\Omega)$, mais c'est bien le cas.

	Massetard	Fortdubras
défectueux	3	21
non défectueux	27	49

Reprenons la question : quelle est la probabilité qu'une arbalète vienne de chez Fortdubras, *sachant qu'elle est défectueuse* ? La deuxième partie de la phrase indique que l'on ne considère que les arbalètes défectueuses, et que l'on élimine les autres de notre raisonnement. Le tableau des arbalètes à considérer est donc maintenant :

	Massetard	Fortdubras
défectueux	3	21

Les arbalètes défectueuses venant de Fortdubras sont au nombre de 21, non pas sur un total de 100 comme précédemment, mais sur un total de $3 + 21 = 24$. La probabilité qu'une arbalète vienne de Fortdubras sachant qu'elle est défectueuse est donc de $21/24$, soit $p = 0,875$.

Formalisons un peu ces résultats. Notons « M » l'événement : « l'arbalète vient de chez Massetard » et « F » l'événement : « l'arbalète vient de chez Fortdubras ». Notons « D » l'événement : « l'arbalète est défectueuse » et « $\overline{\mathrm{D}}$ » l'événement : « l'arbalète n'est pas défectueuse ».

La première probabilité calculée se note $\mathbf{P}(\mathrm{F} \mid \mathrm{D})$ et se lit : « probabilité de (l'événement) F sachant (l'événement) D ». Nous l'avons calculée comme étant la proportion de $\mathrm{F} \cap \mathrm{D}$ sur la proportion de D :

$$p = \mathbf{P}(\mathrm{F} \mid \mathrm{D}) = \frac{\mathbf{P}(\mathrm{F} \cap \mathrm{D})}{\mathbf{P}(\mathrm{D})}.$$

Exemple 1.16 Suite à un stage organisé par la société Massetard, le commissaire divisionnaire Vimaire repère instantanément les arbalètes défectueuses de chez Fortdubras et les fait mettre à la casse. Quelle est la probabilité qu'une arbalète, prise au hasard dans les stocks, soit défectueuse ?

On suppose que toutes les arbalètes défectueuses de chez Fortdubras ont été éliminées. Sur 100 arbalètes initiales, il reste donc :

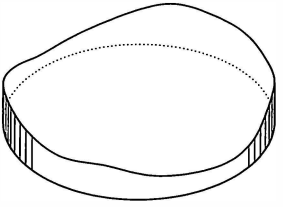
	M	F
défectueux	3	
non défectueux	27	49

La proportion d'arbalètes non défectueuses est donc de $27 + 49 = 76$ sur $3 + 27 + 49 = 79$, c'est-à-dire $76/79 \approx 96,2\,\%$. La probabilité qu'une arbalète soit défectueuse sachant que les arbalètes défectueuses de chez Fortdubras ont été éliminées est donc de $p' = 3/79 \approx 0,038$.

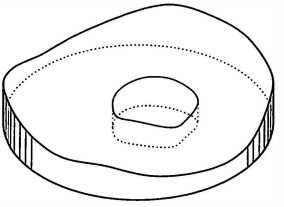
Formalisez ce résultat comme précédemment.

§ 18 Second exemple

Représentons symboliquement l'univers Ω comme une partie du plan et affectons un « poids » de probabilité variable à chaque point de Ω. Le poids total de la galette ainsi représentée est 1. Si A est un événement de Ω, la probabilité de A est égale au « poids » de galette au-dessus de A.



La galette de poids 1

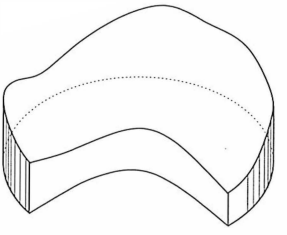


Probabilité de A

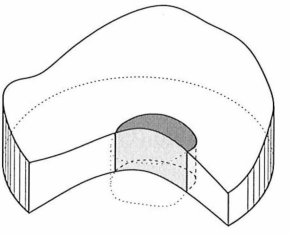
Choisissons maintenant un événement B, et intéressons-nous au nouvel univers B — il suffit donc de couper un morceau de galette et de considérer que seuls les événements de B sont acceptables. Sur cet univers, nous pouvons encore calculer des probabilités, en respectant les proportions de la galette. Problème : le poids total n'est plus égal à 1. La belle affaire ! Il suffit d'épaissir la galette, en la multipliant par le facteur $1/\mathbf{P}(B)$: ainsi, le poids total est bien égal à 1, et les proportions respectives sont préservées. On définit de cette manière une nouvelle fonction de probabilité $\mathbf{P}_B$ sur l'univers B. Si A est un événement de Ω, l'événement de B qui suit la même description[26] est $A \cap B$. Sa probabilité « sachant qu'on est dans B » est alors

$$\mathbf{P}_B(A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

Il est fréquent de noter $\mathbf{P}(A \mid B) = \mathbf{P}_B(A)$.



La nouvelle galette de poids 1, *plus épaisse*



Probabilité de A *sachant* B

Définition 1.17 Soient A et B deux événements ; on suppose $\mathbf{P}(B) > 0$. La **probabilité conditionnelle de A sachant B** est

$$\mathbf{P}_B(A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

Formule des probabilités totales § 19

Lorsqu'une expérience ne peut donner que des résultats $A_1, A_2, \ldots$ en nombre fini ou dénombrable, c'est-à-dire si l'on a une partition

$$\Omega = A_1 \cup A_2 \cup \cdots = \bigcup_k A_k$$

(l'union étant disjointe), alors tout événement B peut se décomposer sous la forme d'une union disjointe

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \cdots = \bigcup_k (B \cap A_k).$$

L'additivité (finie ou dénombrable, selon le cardinal de la famille des A_k) de l'application $\mathbf{P}$ permet d'écrire la **formule des probabilités totales** :

$$\mathbf{P}(B) = \sum_k \mathbf{P}(B \cap A_k) = \sum_k \mathbf{P}(A_k)\,\mathbf{P}(B \mid A_k).$$

(26). C'est-à-dire caractérisée par les mêmes propriétés.

Exemple 1.18 On considère l'énoncé suivant.

> *La probabilité qu'il pleuve un matin donné est* $p = 1/5$. *Lorsqu'il pleut, la probabilité que je prenne mon parapluie est* $9/10$. *Lorsqu'il ne pleut pas, la probabilité que je le prenne est* $1/10$. *Quelle est la probabilité que je prenne mon parapluie un matin donné ?*

La première étape dans la résolution d'un tel problème est de formaliser les données. On note A l'événement « il pleut », et A^c l'événement « il ne pleut pas » ; l'univers Ω est bien sûr égal à l'union disjointe $A \cup A^c$. Notons B l'événement « je prends mon parapluie ». La phrase « Lorsqu'il pleut, la probabilité que je prenne mon parapluie est 9/10 » doit s'interpréter comme une *probabilité conditionnelle*, c'est-à-dire comme $\mathbf{P}(B \mid A)$ (« sachant qu'il pleut »), et non comme la probabilité d'une conjonction $\mathbf{P}(B \cap A)$ (« il pleut *et* je prends mon parapluie »). On a ainsi

$$\mathbf{P}(B \mid A) = \frac{9}{10} \qquad \text{et} \qquad \mathbf{P}(B \mid A^c) = \frac{1}{10}.$$

La formule des probabilités totales donne alors

$$\mathbf{P}(B) = \mathbf{P}(A) \cdot \mathbf{P}(B \mid A) + \mathbf{P}(A^c) \cdot \mathbf{P}(B \mid A^c) = \frac{1}{5} \cdot \frac{9}{10} + \frac{4}{5} \cdot \frac{1}{10} = \frac{13}{50}.$$

§ 20 Formule des probabilités composées

Supposons donnés des événements $A_1, A_2, \ldots, A_n$, dont l'intersection est supposée de probabilité non nulle. On a alors

$$\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1)\,\mathbf{P}(A_2 \mid A_1)$$

par simple définition d'une probabilité conditionnelle. De même, on peut écrire

$$\begin{aligned} \mathbf{P}(A_1 \cap A_2 \cap A_3) &= \mathbf{P}(A_1 \cap A_2)\,\mathbf{P}(A_3 \mid A_1 \cap A_2) \\ &= \mathbf{P}(A_1)\,\mathbf{P}(A_2 \mid A_1)\;\mathbf{P}(A_3 \mid A_1 \cap A_2) \end{aligned}$$

et, par une récurrence simple[(27)], on aboutit à

$$\mathbf{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2 \mid A_1) \cdot \mathbf{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbf{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}).$$

C'est ce que l'on appelle la **formule des probabilités composées.**

Exemple 1.19 Lorsque Paul invite Virginie à dîner, il y a une chance sur deux qu'elle accepte de l'accompagner au restaurant (événement « R »). Si elle accepte, il y a une chance sur trois qu'elle consente, après le restaurant, à monter boire un dernier verre chez Paul (événement « V »). Et dans ce cas, il y a une chance sur cinq qu'elle ne rentre pas chez elle (événement « L »).

On notera que les deux dernières probabilités sont en fait des *probabilités conditionnelles* : 1/3 est le probabilité que Virginie invite Paul à un dernier verre, *sachant qu'elle a d'abord accepté l'invitation au restaurant.* De même pour 1/5, probabilité que Virginie reste chez Paul, *sachant qu'elle a accepté le restaurant et le dernier verre.*

La probabilité que Paul finisse la nuit avec Virginie est donc

$$\mathbf{P}(R \cap V \cap L) = \mathbf{P}(R) \cdot \mathbf{P}(V \mid R) \cdot \mathbf{P}(L \mid V \cap R) = \tfrac{1}{2} \cdot \tfrac{1}{3} \cdot \tfrac{1}{5} = \tfrac{1}{30}.$$

(Finalement, Paul invitera Virginie à faire du bateau, avec les conséquences que l'on sait.)

§ 21 Indépendance stochastique

Soient A et B deux événements, et supposons que $\mathbf{P}(B) > 0$. On dira que A et B sont indépendants lorsque la connaissance de la réalisation (ou non) de B ne modifie pas notre connaissance de la probabilité de A, c'est-à-dire que la probabilité de A *sachant* B est égale à celle de A :

$$\mathbf{P}(A \mid B) = \mathbf{P}(A),$$

(27). Voir § 59 page 103 pour une démonstration explicite.

ce qui, en multipliant de part et d'autre par $\mathbf{P}(\mathrm{B})$, fournit la condition équivalente

$$\mathbf{P}(\mathrm{A} \cap \mathrm{B}) = \mathbf{P}(\mathrm{A}) \cdot \mathbf{P}(\mathrm{B}).$$

Cette relation est plus avantageuse que la précédente car, outre son caractère symétrique, elle a encore un sens dans le cas où $\mathbf{P}(\mathrm{B}) = 0$. On dira donc que deux événements A et B sont **indépendants** quand $\mathbf{P}(\mathrm{A} \cap \mathrm{B}) = \mathbf{P}(\mathrm{A}) \cdot \mathbf{P}(\mathrm{B})$.

« Indépendance = produit » (∗) §22

Lorsque l'on souhaite modéliser une situation où deux événements sont indépendants, le plus simple est de construire un univers adapté sous la forme d'un produit cartésien. Considérons l'expérience consistant à tirer deux dés (équilibrés ou non) à la suite. On veut modéliser le fait que les deux tirages sont indépendants. Cela signifie que, après le premier tirage, tous les univers parallèles ω sont *a priori* équivalents pour le second tirage, quelle que soit l'issue du premier. (Dans le cas équilibré, un sixième de ces univers amènera le « 1 » au second tirage, un sixième le « 2 », etc.) Si l'on représente horizontalement l'univers $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$ modélisant le premier tirage, et verticalement $\Omega_2 = \{1, 2, 3, 4, 5, 6\}$ celui modélisant le second, alors l'univers $\Omega_1 \times \Omega_2$ permet de modéliser la succession des deux tirages ; il est muni de la probabilité produit, notée $\mathbf{P} = \mathbf{P}_1 \otimes \mathbf{P}_2$, caractérisée par

$$\mathbf{P}(\mathrm{A}_1 \times \mathrm{A}_2) = \mathbf{P}_1(\mathrm{A}_1) \cdot \mathbf{P}_2(\mathrm{A}_2) \qquad (*)$$

pour tous événements $\mathrm{A}_1 \subset \Omega_1$ et $\mathrm{A}_2 \subset \Omega_2$ (voir figure 1.2). On se convaincra du fait que $\mathbf{P}$ est bien entièrement définie par cette relation, *bien que tous les événements de* $\Omega_1 \times \Omega_2$ *ne soient pas de la forme* $\mathrm{A}_1 \times \mathrm{A}_2$! En effet, dans le cas présent, ils sont tous de la forme d'une union disjointe de tels événements produits (la décomposition n'étant pas unique). Par exemple, on peut écrire

$$\begin{aligned}\big\{(1,2),(1,3),(2,3),(6,4)\big\} &= \{1\} \times \{2\} \cup \{1,2\} \times \{3\} \cup \{6\} \times \{4\} \\ &= \{1\} \times \{2,3\} \cup \{2\} \times \{3\} \cup \{6\} \times \{4\}.\end{aligned}$$

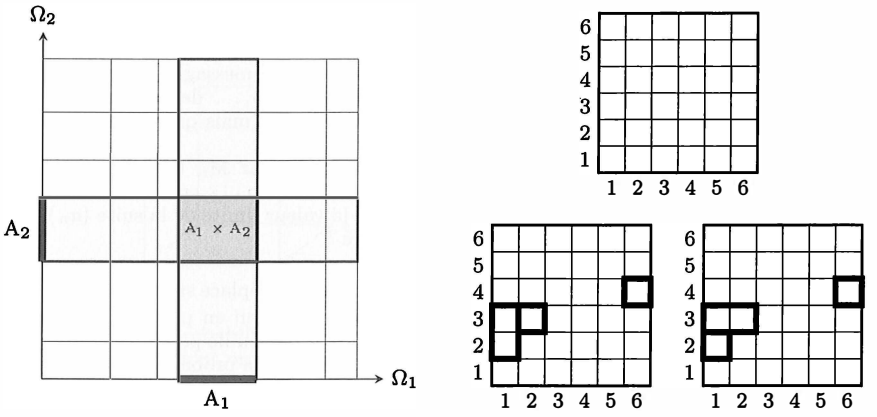


FIGURE 1.2 — *Probabilité produit.* À gauche : *cas générique de deux univers* Ω_1 *et* Ω_2. À droite : *la décomposition d'un événement en union de produits cartésiens n'est pas unique.*

Dans le cas général, la situation est sensiblement plus compliquée, mais on montrera que la relation (∗) suffit malgré tout à définir entièrement la probabilité produit ; ce sera établi au paragraphe § 230.

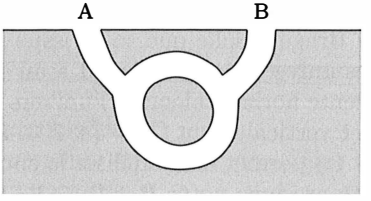
EXERCICES

Les exercices qui suivent ne nécessitent qu'un minimum de formalisation. Le seul théorème probabiliste à utiliser est la formule des probabilités totales (§ 19). Des indications suivent le groupe des énoncés. Les solutions suivent le groupe des indications.

♦ ***Exercice 1.1 (Le paradoxe des anniversaires de von Mises)*** Calculer le nombre minimal de personnes dans une classe donnée tel que la probabilité qu'au moins deux personnes aient leur anniversaire le même jour soit au moins 1/2.

(Ce résultat n'est pas à proprement parler un paradoxe — le lecteur pourra se rapporter au chapitre 34 pour d'authentiques paradoxes — mais plutôt un résultat qui contredit l'intuition.)

♦ ***Exercice 1.2*** Une petite taupe entre dans son terrier par l'entrée A. À chaque intersection, elle a (indépendamment des choix précédents) une chance sur deux de prendre à gauche et une chance sur deux de prendre à droite. Quelle est la probabilité qu'elle sorte en B ?



On pourra noter, par exemple, « DGGD » le chemin suivi par la taupe lorsqu'elle tourne à droite au premier croisement, puis à gauche au deuxième et au troisième, et à droite au quatrième ; un tel chemin la fait sortir en B.

♦ ***Exercice 1.3 (PP avant PF)*** On dispose d'une pièce, amenant *pile* avec la probabilité $p \in]0\,;\,1[$. On effectue une séquence infinie de lancers. Quelle est la probabilité d'obtenir deux *pile* consécutifs sans avoir eu auparavant la séquence *pile-face* ?

♦ ***Exercice 1.4 (Trois face d'affilée)*** On dispose d'une pièce truquée. La probabilité qu'elle tombe sur *pile* est p et la probabilité qu'elle tombe sur *face* est $q = 1 - p$. On considère une succession de lancers de cette pièce.

Pour tout entier $n \geqslant 1$, on nomme B_n l'événement :

B_n = « aucune séquence de *face* de longueur 3 n'apparaît dans la suite des n premiers lancers »

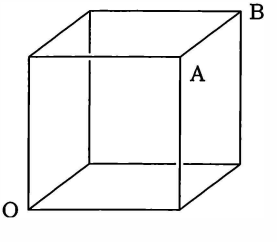
et on note b_n la probabilité de B_n. Calculer b_1, b_2 et b_3. Montrer que :

$$\forall n \geqslant 4 \qquad b_n = p\,b_{n-1} + pq\,b_{n-2} + p\,q^2\,b_{n-3}. \qquad (*)$$

♦ ***Exercice 1.5 (La chaîne des menteurs)*** On suppose qu'un message binaire (0 ou 1) est transmis depuis un émetteur M_0 à travers une chaîne M_1, M_2, M_3,... de messagers menteurs, qui transmettent correctement le message avec une probabilité p, mais qui changent sa valeur avec la probabilité $1 - p$.

Si l'on note a_n la probabilité que l'information transmise par M_n soit identique à celle envoyée par M_0 (avec comme convention que $a_0 = 1$), déterminer a_{n+1} en fonction de a_n, puis une expression explicite de a_n en fonction de n, ainsi que la valeur limite de la suite $(a_n)_{n\geqslant 1}$. Le résultat est-il conforme à ce à quoi l'on pouvait s'attendre ?

♦ ***Exercice 1.6 (La chenille et le cube [29])*** Une chenille se déplace sur les arêtes d'un cube. À chaque sommet, elle emprunte au hasard une des trois arêtes qui en partent, avec la même probabilité 1/3. Les points A et B sont enduits de colle forte. La chenille part du point O. Quelle est la probabilité qu'elle se retrouve collée en A ? En B ? Quelle est la probabilité qu'elle continue à marcher indéfiniment sans se faire coller ?



♦ ***Exercice 1.7 (Le problème des huit tours)*** On place aléatoirement huit tours sur un échiquier 8×8. Quelle est la probabilité qu'aucune tour ne puisse en attaquer une autre (i.e. qu'aucune tour ne soit sur la même ligne ou la même colonne qu'une autre) ?

♦ ***Exercice 1.8 (Sondage délicat)*** Un institut désire effectuer un sondage pour déterminer qui, du vizir Iznogoud ou du calife Haroun el Poussah, est le plus populaire. Toutefois, il est possible que les personnes interrogées ne se sentent pas libres de répondre avec sincérité face à l'enquêteur. On propose donc aux sondés le protocole suivant : six cartes sont présentées, face cachée, la personne interrogée en tire une au hasard et garde secret le message écrit dessus. Une des cartes porte le message « Dites le contraire de la vérité » ; les cinq dernières « Dites la vérité ». Ainsi, quelle que soit la réponse donnée, l'enquêteur ne peut pas savoir avec certitude si la personne votera pour le vizir ou pour le calife.

À l'issue du sondage, une proportion p' de l'échantillon interrogé a répondu « Iznogoud ». Quelle est la proportion p comptant voter effectivement pour l'infâme vizir ? *(On supposera l'échantillon suffisamment important pour identifier p et p' à des probabilités d'intention de vote.)*

Indications

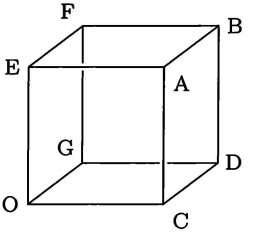
◊ *Indication 1.1 :* Dans une classe de n personnes, calculer plutôt la probabilité de l'événement inverse : chaque personne a une date d'anniversaire différente. À l'aide d'un ordinateur, calculer les valeurs numériques de la probabilité cherchée en augmentant n jusqu'à ce que le résultat dépasse $1/2$.

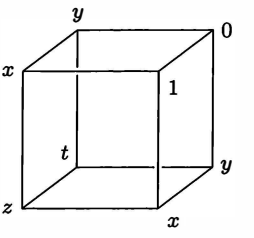
◊ *Indication 1.2 :* On listera tous les chemins commençant par « G » et qui font sortir la taupe en B, ainsi que tous les chemins commençant par « D » qui la font également sortir en B. On calculera ensuite la probabilité associée à chacun de ces chemins.

◊ *Indication 1.3 :* Lister tous les tirages possibles donnant les deux *pile* avant la séquence *pile-face*.

◊ *Indication 1.4 :* Il eixste plusieurs types de raisonnements pour établir une relation de récurrence. On peut par exemple réfléchir conditionnellement à la valeur des trois derniers tirages, mais la relation de récurrence obtenue est alors d'ordre plus élevé que voulue. Pour obtenir la relation d'ordre 3 demandée, il convient, pour $n \geqslant 4$, de décomposer B_n selon la position du dernier *face*, qui ne peut être qu'aux instants n, $n-1$, ou $n-2$.

◊ *Indication 1.6 :* On appliquera des principes de symétrie seulement (pas de théorie de la mesure ici !) et on considérera, pour chaque sommet, la probabilité que la chenille finisse en A si elle part de ce sommet. On vérifiera que le nombre de coefficients inconnu n'est que de 4 :





On reliera enfin ces probabilités entre elles par la formule des probabilités totales afin de former un système linéaire, dont on vérifiera qu'il est inversible.

◊ *Indication 1.7 :* Le plus simple est de compter le nombre de configurations possibles.

◊ *Indication 1.8 :* En utilisant l'indépendance entre le choix de vote et la carte tirée, calculer p' en fonction de p (plutôt que p en fonction de p').

SOLUTIONS

♦ **Solution 1.1** Les années bissextiles sont momentanément ignorées. Notons n le nombre de personne dans la classe. Calculons la probabilité q_n de l'événement contraire : toutes les personnes ont leur anniversaire des jours distincts.

Pour cela, calculons d'abord le nombre de possibilités de répartir n personnes dans $\mathrm{N} = 365$ cases, à raison d'une personne au maximum par case. Une façon de procéder est de commencer par numéroter ces personnes, puis d'effectuer le calcul en les considérant comme discernables ; il y a N possibilités de placer la première dans les N cases, puis $\mathrm{N}-1$ de placer la seconde,... jusqu'à $\mathrm{N}-n+1$ possibilités de placer la dernière. Ce qui nous donne $\mathrm{N}!/(\mathrm{N}-n)!$ possibilités de répartir ces n personnes dans les N cases.

Enfin, le nombre total de possibilités de répartir *sans contrainte* ces n personnes dans N cases est de N^n. La probabilité que les n personnes aient leur anniversaire un jour distinct est

$$q_n = \frac{\mathrm{N}!}{(\mathrm{N}-n)!\,\mathrm{N}^n},$$

et la probabilité cherchée est donc

$$p_n = 1 - q_n = 1 - \frac{\mathrm{N}!}{(\mathrm{N}-n)!\,\mathrm{N}^n}.$$

On peut dès lors chercher numériquement à partir de quelle valeur de n le nombre p_n devient supérieur à 0, 5. Voici le tableau des 30 premières valeurs, approchées à 10^{-6} près :

n	p_n	n	p_n	n	p_n
1	0, 000000	11	0, 141141	21	0, 443688
2	0, 002740	12	0, 167025	22	0, 475695
3	0, 008204	13	0, 194410	23	0, 507297
4	0, 016356	14	0, 223103	24	0, 538344
5	0, 027136	15	0, 252901	25	0, 568700
6	0, 040462	16	0, 283604	26	0, 598241
7	0, 056236	17	0, 315008	27	0, 626859
8	0, 074335	18	0, 346911	28	0, 654461
9	0, 094624	19	0, 379119	29	0, 680969
10	0, 116948	20	0, 411438	30	0, 706316

Dans une classe de 23 personnes ou plus, la probabilité que deux personnes aient leur anniversaire le même jour est supérieure ou égale à $1/2$.

Le lecteur pourra vérifier que, si l'on considère les années bissextiles, le résultat (23) est inchangé.

♦ **Solution 1.2** La taupe sort en B si et seulement si elle suit l'un des chemins suivants :

GG GDDG GDDDDG ou DD DGGD DGGGGD

qui sont des événements incompatibles. Les chemins GG et DD sont tous deux de probabilité $1/4$, les chemins GDDG et DGGD sont de probabilité $1/16$, etc.

Le probabilité pour que la taupe prenne l'un de ces chemins est donc la somme des probabilités de chacun des chemins, à savoir

$$p = 2\left[\frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots\right] = \frac{2}{3}.$$

♦ **Solution 1.3** Les seuls tirages possibles sont, en notant $q := 1 - p$,

événement	probabilité
PP	p^2
FPP	qp^2
FFPP	q^2p^2
⋮	⋮
$\underbrace{\mathrm{FFF}\cdots\mathrm{FF}}_{n \text{ fois}}\mathrm{PP}$	$q^n p^2$
⋮	⋮

et la probabilité d'avoir réalisé l'un des événements disjoints précédents est

$$\sum_{n=0}^{\infty} q^n\, p^2 = \frac{1}{1-q}\, p^2 = p.$$

♦ **Solution 1.4** Les événements B_1 et B_2 sont certains, par définition ! Quant à l'événement B_3, il est plus simple de calculer la probabilité de son complémentaire ; en effet, $\overline{B_3}$ correspond à l'événement « au moins une séquence de trois *face* a lieu dans la suite des 3 premiers lancers », autrement dit :

$$\overline{B_3} = \text{« la suite des tirages commence par FFF »}$$

événement qui a pour probabilité q^3. On en déduit que $\mathbf{P}(B_1) = \mathbf{P}(B_2) = 1$ et $\mathbf{P}(B_3) = 1 - q^3$.

Passons maintenant à la relation de récurrence. Soit n un entier tel que $n \geqslant 4$ et considérons un tirage correspondant à l'événement B_n, c'est-à-dire tel qu'aucune séquence de *face* de longueur 3 ne soit apparue dans la suite des n premiers lancers. Notamment, il est certain que les trois derniers lancers n'ont pas donné trois *face*, c'est-à-dire que l'on a obtenu *pile* au moins une fois.

Nous allons décomposer l'événement B_n en une réunion de trois événements incompatibles, selon la position du dernier *pile* — qui est forcément à l'une des positions n, $n-1$ ou $n-2$. On notera F_k l'événement « on a obtenu *face* au tirage numéro k » et P_k l'événement « on a obtenu *pile* au tirage numéro k ».

❶ Le dernier *pile* est au tirage n. Alors, aucune séquence de trois faces n'a eu lieu durant les $n-1$ premiers tirages, et l'événement est donc $B_{n-1} \cap P_n$. La probabilité associée est $p\, b_{n-1}$.

❷ Le dernier *pile* est au tirage $n-1$. Alors, le dernier tirage est *face*, et aucune séquence de trois faces n'a eu lieu durant les $n-2$ premiers tirages ; l'événement est donc $B_{n-2} \cap P_{n-1} \cap F_n$. La probabilité associée est $pq\, b_{n-2}$.

❸ Le dernier *pile* est au tirage $n-2$. Alors, les deux derniers tirages sont *face*, et aucune séquence de trois faces n'a eu lieu durant les $n-3$ premiers tirages ; l'événement est donc $B_{n-2} \cap P_{n-2} \cap F_{n-1} \cap F_n$. La probabilité associée est $pq^2\, b_{n-3}$.

	$\cdots$	$n-2$	$n-1$	n	**P**
❶	$\cdots$	*	*	P	pb_{n-1}
❷		*	P	F	pqb_{n-2}
❸	$\cdots$	P	F	F	pq^2b_{n-3}

Ces trois événements étant incompatibles, on a donc

$$\forall n \geqslant 4 \qquad b_n = p\, b_{n-1} + pq\, b_{n-2} + pq^2\, b_{n-3}.$$

Remarque 1.20 On pouvait également utiliser la relation $B_{n+1} \subset B_n$, et donc

$$\mathbf{P}(B_{n+1}) = \mathbf{P}(B_n) - \mathbf{P}(B_n \smallsetminus B_{n+1}).$$

Or,

$$\mathbf{P}(B_n \smallsetminus B_{n+1}) = \mathbf{P}(B_{n-3} \cap P_{n-2} \cap F_{n-1} \cap F_n \cap F_{n+1}) = pq^3\, b_{n-3},$$

ce qui montre que $b_{n+1} = b_n - pq^3\, b_{n-3}$. La relation voulue se démontre alors par récurrence en utilisant cette propriété.

♦ **Solution 1.5** La formule des probabilités totales permet d'écrire

$$a_{k+1} = p\, a_k + (1-p)(1-a_k) = (2p-1)\, a_k + (1-p).$$

On reconnaît une suite arithmético-géométrique de point fixe $\ell = 1/2$: pour tout entier n, on a donc $(a_{n+1} - \ell) = (2p-1)\,(a_n - \ell)$. Cela nous donne la valeur explicite $a_n = \frac{1}{2}(2p-1)^{n-1} + \frac{1}{2}$ et, bien entendu, $\lim a_n = \frac{1}{2}$. Cela signifie que, après un grand nombre de transmissions hasardeuses, on a perdu toute information !

♦ **Solution 1.6** Pour des raisons de symétrie, x et y apparaissent deux fois chacun. Supposons la chenille au point C. Elle peut se diriger vers les points O, A ou D avec une probabilité égale, ce qui nous fournit un système complet d'événements. D'après la formule des probabilités totales, on peut donc écrire

$$x = \frac{1}{3} \cdot z + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot y.$$

En supposant la chenille aux points D, O, ou G, on obtient trois autres relations, ce qui nous donne au final le système

$$\begin{cases} 3x = y + z + 1 \\ 3y = x + t \\ 3z = 2x + t \\ 3t = 2y + z, \end{cases}$$

dont la résolution mène à

$$x = \frac{9}{14} \qquad y = \frac{5}{14} \qquad z = \frac{4}{7} \qquad t = \frac{3}{7}.$$

Ainsi, *la probabilité que la chenille finisse collée en* A *lorsqu'elle part de* O *est* $z = 4/7$.

Maintenant, pour des raisons de symétrie, *la probabilité qu'elle finisse collée en* B *en partant de* O *est la même que celle qu'elle finisse collée en* A *en partant de* G, *c'est-à-dire* $t = 3/7$.

Puisqu'enfin $z + t = 1$, on en déduit que, conformément à l'intuition *la probabilité qu'elle marche indéfiniment est nulle.*

♦ **Solution 1.7** On commence par numéroter chacune des tours. Le nombre total de configurations possibles, c'est-à-dire de manières de placer successivement les 8 tours sur l'échiquier, est $64!/56!$.

Comptons maintenant le nombre de configurations possibles. On doit d'abord attribuer à chaque tour une colonne distincte, ce qui donne 8! possibilités. Ensuite, chacune des tours restant sur sa propre colonne, on doit leur attribuer à chacune sa propre ligne ; ce qui donne encore 8! possibilités. Au total, le nombre de possibilités est de $(8!)^2$, et la probabilité cherchée est

$$p = \frac{8! \cdot 8! \cdot 56!}{64!} \approx 9,11 \cdot 10^{-6}.$$

On pouvait également ne pas numéroter les tours mais les considérer comme indiscernables. Le nombre de possibilités de choisir 8 cases sur l'échiquier est $\binom{64}{8}$; le nombre de configurations possibles est 8! (8 possibilités pour la case de la première colonne, 7 pour la case de la deuxième colonne etc). On trouve — heureusement — le même résultat :

$$p = \frac{8!}{\binom{64}{8}} \approx 9,11 \cdot 10^{-6}.$$

♦ **Solution 1.8** Notons I l'événement correspondant à une intention de vote pour Iznogoud, et V celui correspondant au tirage de la carte enjoignant de dire la vérité.

L'événement R : « la personne répond qu'elle va voter pour Iznogoud » est égale à l'union *disjointe*

$$\mathrm{R} = (\mathrm{I} \cap \mathrm{V}) \cup (\mathrm{I}^c \cap \mathrm{V}^c),$$

et, par indépendance de I et V, on a

$$p' = \mathbf{P}(\mathrm{R}) = \mathbf{P}(\mathrm{I} \cap \mathrm{V}) + \mathbf{P}(\mathrm{I}^c \cap \mathrm{V}^c) = \mathbf{P}(\mathrm{I})\,\mathbf{P}(\mathrm{V}) + \mathbf{P}(\mathrm{I}^c)\,\mathbf{P}(\mathrm{V}^c) = p\,\frac{5}{6} + (1-p)\,\frac{1}{6} = \frac{2p}{3} + \frac{1}{6}.$$

(On a utilisé le fait que, si I et V sont indépendants, alors I^c et V^c le sont également.) Ainsi, $p = \dfrac{3p'}{2} - \dfrac{1}{4}$.

Chapitre 2

Variables aléatoires

(approche heuristique)

Ce calcul délicat s'étend aux questions les plus importantes de la vie, qui ne sont en effet, pour la plupart, que des problèmes de probabilité.

Pierre-Simon, marquis de Laplace,
dédicace (à Napoléon) de sa Théorie analytique des probabilités,
1812.

Pour décrire une quantité X *pouvant prendre des valeurs « aléatoires », il suffit de définir une fonction* X *sur l'ensemble* Ω. *Ainsi, à chaque « possibilité du hasard »* ω, *la quantité* $X(\omega)$ *est déterminée.*

Contrairement à ce qui se passe souvent en analyse, ce n'est pas X *en tant que fonction qui présente un intérêt. Les notions usuelles de continuité ou de dérivabilité, centrales en analyse, n'ont ici aucun sens, puisque l'on n'a pas défini* a priori *de notion de « proximité » entre les* ω, *c'est-à-dire de topologie sur* Ω. *D'ailleurs, la plupart du temps,* Ω *est inconnu et on ne se soucie pas de le définir correctement : tel le système d'exploitation de qualité, il travaille en tâche de fond et se fait oublier de l'utilisateur. Pour le probabiliste, ce qui compte est : avec quelle fréquence* X *prend-elle telle ou telle valeur ? On cherche ainsi à répondre à des questions du type :*

- *quelle est la probabilité que* $X(\omega)$ *prenne ses valeurs dans l'intervalle* $[a\,;b]$ *?*
- *quelle est la probabilité que* $X(\omega)$ *soit égale précisément à* 0 *?*
- *quelle est la probabilité que* $|X(\omega)| > 42$ *?*
- *ou plus généralement, si* A *est une partie de* $\mathbb{R}$, *quelle est la probabilité que* $X(\omega)$ *soit dans* A *?*

On résume l'ensemble de ces question sous la formule : comment les valeurs prises par X sont-elles distribuées ? *La donnée quantitative de cette propriété s'appelle la* **loi** *(de distribution) de* X. *Nous verrons — chose remarquable — qu'elle est entièrement déterminée par la réponse à la question : pour chaque réel* x, *quelle est la probabilité que* $X(\omega)$ *soit inférieur ou égal à* x *?*

2.1 VARIABLES RÉELLES

§ 23 **Première définition**

Dans de nombreux cas, l'univers Ω est un ensemble tellement compliqué que sa simple description est trop complexe pour qu'on s'y attarde sérieusement. Pour donner un exemple concret, considérons un duel entre deux cow-boys A et B. À midi, ils se retrouvent sur l'avenue principale ; leur cache-poussière est ouvert, ils ont la main à quelques centimètres de leurs pistolets respectifs. Chacun des deux, jusqu'à l'instant où il va tirer, reste parfaitement immobile. La question est : qui tirera le premier, et à quel instant ?

Si l'on désire rendre compte d'une telle expérience avec la plus grande précision possible, l'univers Ω doit prendre en compte la description complète des cow-boys et de leur psychologie (ce qui implique peut-être une description de la configuration atomique de nos cow-boys et l'évolution de cette configuration à partir de midi) ainsi qu'une description de leur environnement plus ou moins proche (un coup de tonnerre dans une vallée voisine peut rendre l'un des cow-boys nerveux et le pousser à tirer plus tôt qu'il ne le souhaitait). La simple description mathématique de l'univers et d'une fonction de probabilité sur l'ensemble des événements paraît totalement inaccessible. Dans l'étape de *modélisation*, il est possible de simplifier le choix de l'univers Ω.

Par exemple, on peut choisir l'univers restreint $\Omega' := [0\,;\,+\infty]^2$; une réalisation ω du hasard est le couple $\omega = (t_1, t_2)$ des instants de tir de chaque cow-boy (où par convention $t_i = +\infty$ si le cow-boy numéro i meurt avant d'avoir pu tirer). Le choix d'un tel univers, s'il est adapté à la question (qui va tirer en premier, et quand), restreint la possibilité de se poser de nouvelles questions sur l'issue du duel (par exemple en étudiant l'éventualité qu'un des pistolets s'enraye).

Si l'on ne s'intéresse qu'à une donnée numérique — l'instant du premier tir —, il est possible de simplifier les choses sans toucher nécessairement à Ω. Pour cela, définissons une application

$$X : \Omega \longrightarrow \mathbb{R}^+$$

qui, à chaque événement atomique ω, associe l'instant du tir du cow-boy A. Une nouvelle question peut alors être : quelle est la probabilité que le cow-boy A ait tiré avant un instant t donné ? Cette quantité est la probabilité associée à l'ensemble $\mathscr{A}_t$ des configurations correspondant à la description : « A a tiré avant l'instant t », à savoir

$$\mathscr{A}_t = \{\omega \in \Omega\ ;\ X(\omega) \leqslant t\}.$$

L'événement $\mathscr{A}_t$ n'est rien d'autre que l'image réciproque, par X, de l'intervalle $]-\infty\,;\,t]$:

$$\mathscr{A}_t := X^{-1}\big(]-\infty\,;\,t]\big)$$

et il est d'usage de le noter

$$\{X \in]-\infty\,;\,t]\} \qquad \text{ou} \qquad \{X \leqslant t\}.$$

Dans l'étape de modélisation de la situation, on cesse de se poser des questions sur la structure de Ω, et on ne s'intéresse plus qu'à la distribution des valeurs de X, distribution qui sera d'abord définie *a priori*, confrontée à l'expérience lorsque c'est possible[1], etc.

(1). Une fois le cow-boy mort, on ne peut plus refaire l'expérience ; la modélisation probabiliste d'une expérience *unique par essence* est donc tout à fait discutable. C'est le cas par exemple lorsque l'on traite de questions cosmologiques.

Probabilité image (ou loi) §24

L'application X envoyant Ω dans $\mathbb{R}$, utilisons-la pour définir, à partir de la probabilité **P**, une nouvelle probabilité, notée $\mathbf{P}_X$, sur l'univers $\Omega' := \mathbb{R}$. On introduit d'abord une notation bien utile : si A est une partie de $\mathbb{R}$, on note

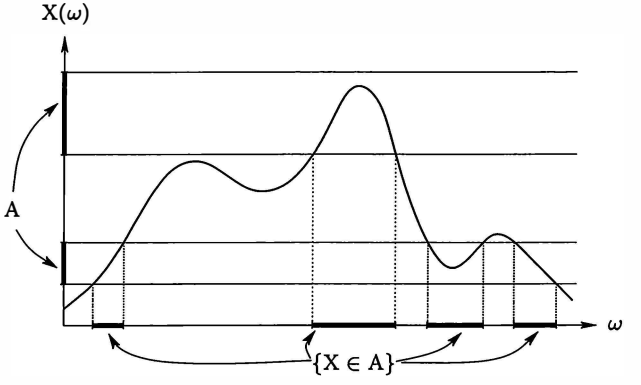
$$\{X \in A\} := \{\omega \in \Omega \,;\, X(\omega) \in A\}.$$

et on définit

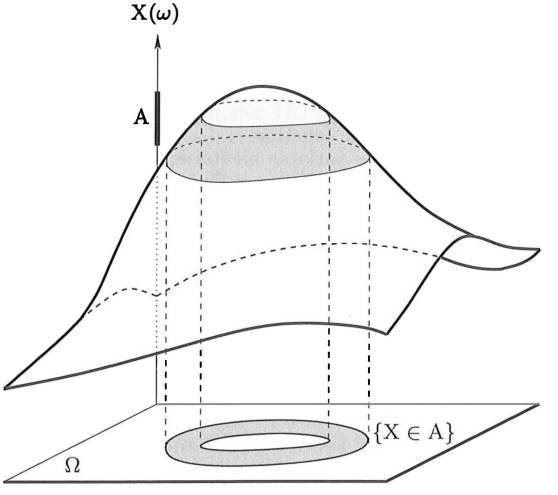
$$\mathbf{P}_X(A) := \mathbf{P}\{X \in A\}.$$

Exemple 2.1 On modélise un lancer de dé ; l'univers Ω est l'ensemble des trajectoires du dé au cours du temps depuis l'instant de lancer. Chaque élément ω est donc une application continue de $\mathbb{R}^+$ dans $\mathbb{R}^6$ (puisqu'il faut trois coordonnées et trois angles pour décrire l'état du dé à chaque instant). On définit enfin une application $X : \Omega \to \{1, 2, 3, 4, 5, 6\}$ qui, à une trajectoire donnée, associe la face visible du dé une fois celui-ci stabilisé. L'événement $\{X = 6\}$ est donc l'ensemble de toutes les trajectoires possibles telles que le dé finisse par montrer la face « 6 ».

Une représentation graphique de $\{X \in A\}$ peut-être la suivante, où Ω est symboliquement représenté sur l'axe des abscisses, et où la courbe est le « graphe » de X :



ou, mieux encore, celle-là :



L'application $\mathbf{P}_X$ est appelée **probabilité image**, ou encore **loi (de distribution) de X**. Nous verrons, dans les chapitres de formalisation, que c'est effectivement une « probabilité » au sens mathématique.

Remarque 2.2 On peut formaliser cette démarche en utilisant le langage habituel de l'analyse. Lorsque l'on se donne une application $X : \Omega \to \mathbb{R}$, on peut construire une application $X^{(-1)}$ de l'ensemble des parties de $\mathbb{R}$ sur l'ensemble des parties de Ω, par

$$X^{(-1)} : A \longmapsto X^{-1}(A).$$

L'ensemble $X^{-1}(A) = \{X \in A\}$ est l'**image réciproque** de A par X. (Rappelons que « X^{-1} » ne désigne pas l'application réciproque de X — qui n'est *a priori* pas bijective.) La loi de X est alors $\mathbf{P}_X = \mathbf{P} \circ X^{(-1)}$.

La loi d'une variable aléatoire est un objet central dans la théorie des probabilités. Il importe de bien comprendre que, pour un probabiliste, ce n'est pas la valeur que prend X en tel ou tel ω qui importe, mais uniquement la « fréquence » à laquelle les différentes valeurs de X peuvent arriver.

Donnons une analogie informatique. Considérons le cas d'un programme qui retourne à l'utilisateur un nombre aléatoire, selon une certaine loi donnée. Pour fixer les idées, imaginons que ce programme retourne un nombre entier entre 1 et 6, avec des fréquences identiques (il simule donc le jet d'un dé équilibré). Pour l'utilisateur, peu importe la manière dont le programme est écrit, et deux programmes différents se valent bien à ses yeux, du moment qu'ils se comportent, non pas identiquement, mais selon la même loi. Le premier programme pourra retourner « 2 » tandis que le second retournera « 5 » : ils ne sont pas identiques, on peut les considérer comme deux variables aléatoires distinctes. Si l'on répète l'expérience (on choisit un nouvel ω), les programmes continueront à fournir des données distinctes mais, s'ils sont bien écrits, on pourra établir qu'ils renvoient des réponses qui suivent la même distribution statistique (c'est-à-dire que les variables aléatoires qu'ils modélisent suivent la même loi).

Remarque 2.3 (Un point de terminologie) Une variable aléatoire n'est donc pas une « variable » (comme dans l'expression analytique $y = f(x)$ par exemple), mais une fonction. Cette curiosité terminologique a cependant un sens : un observateur effectuant une expérience est dans un des mondes parallèles ω de Ω, et pour lui, seul existe $X(\omega)$, qui est un nombre, et non une fonction. S'il réfléchit à son expérience de manière théorique, le résultat peut prendre plusieurs valeurs possibles, et il est naturel de le considérer comme une grandeur « variable ».

Remarque 2.4 (Une difficulté) Une contrainte supplémentaire va compliquer — légèrement — la donne. Il n'est pas possible d'assigner, à toute partie de $\mathbb{R}$, une mesure de probabilité[(2)]. On devra donc se restreindre à un sous-ensemble de l'ensemble des parties de $\mathbb{R}$. Ce sous-ensemble devra vérifier quelques propriétés simples qui en feront une *tribu*. Un ensemble de parties mesurables souvent utilisé est la **tribu $\mathfrak{B}$ des boréliens de $\mathbb{R}$**. Pour définir correctement les probabilités $\mathbf{P}\{X \in A\}$, on va donc demander que $X^{-1}(A)$ soit un événement pour tout borélien A. Cette propriété s'appelle la **mesurabilité** d'une fonction.

Ainsi, on affine la définition d'une variable aléatoire, comme une fonction *mesurable* de Ω (muni de sa tribu $\mathfrak{T}$ des événements) dans $\mathbb{R}$ (muni de la tribu des boréliens).

§25 Fonction de répartition

D'une manière générale, la loi d'une variable aléatoire réelle X est entièrement contenue dans la **fonction de répartition** de X, définie par

$$F_X(x) := \mathbf{P}_X\big(]-\infty\,;\,x]\big) = \mathbf{P}\{X \leqslant x\}.$$

On dit également que *la fonction de répartition caractérise entièrement la loi* $\mathbf{P}_X$: la connaissance de F_X permet de reconstruire entièrement la fonction $\mathbf{P}_X$.

(2). Voir l'annexe 3.5 page 95 du chapitre 3 pour plus de précisions.

Exemple 2.5 En notant $F_X(x^-) := \lim_{h\to 0^+} F(x-h)$ la limite à gauche en x, on établira des formules du type

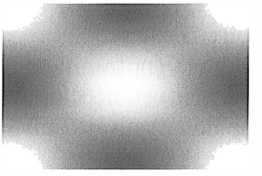
$$\mathbf{P}\{X < x\} = \lim_{h\to 0^+} \mathbf{P}\{X \leqslant x-h\} = F_X(x^-), \qquad \mathbf{P}\{X \geqslant x\} = 1 - F_X(x^-).$$

(Ces formules seront établies proprement au chapitre 3.) Si A est une partie de $\mathbb{R}$ s'écrivant sous la forme d'union d'intervalles, la quantité $\mathbf{P}\{X \in A\}$ s'écrit donc sous forme de sommes et différences de la valeur ou de la limite de F aux points extrémités de ces intervalles, par exemple

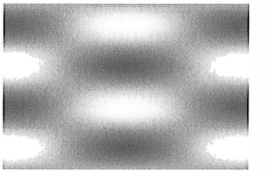
$$\mathbf{P}\{a \leqslant X \leqslant b\} = F(b) - F(a^-).$$

On prendra bien garde au fait que *deux variables aléatoires totalement différentes peuvent avoir la même fonction de répartition, c'est-à-dire suivre la même loi.*

Exemple 2.6 On représente symboliquement l'univers Ω comme un carré d'aire unité, et on donne graphiquement les valeurs de deux variables aléatoires X et Y sur Ω (muni d'une probabilité uniforme, c'est-à-dire que la probabilité d'un événement est proportionnel à son aire) ; ces deux variables prennent leurs valeurs entre -1 et 1, chaque niveau de gris est une valeur différente, le noir correspondant à -1 et le blanc à 1. Il apparaît immédiatement que ces variables sont des applications différentes (leurs graphes sont distincts) :

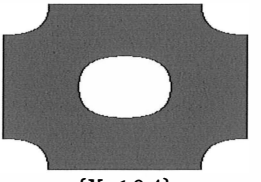


X

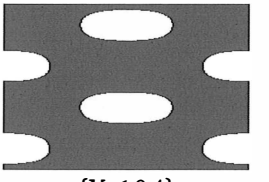


Y

Elles ont pourtant la même loi : la distribution (la fréquence) des différents niveaux de gris est la même. On représente ci-dessous les ensembles $\{X \leqslant 0{,}4\}$ et $\{Y \leqslant 0{,}4\}$; ils ont même aire, et donc même probabilité.



$\{X \leqslant 0{,}4\}$



$\{Y \leqslant 0{,}4\}$

Il en serait de même pour les ensembles $\{X \leqslant x\}$ et $\{Y \leqslant x\}$ pour toute valeur de x. Les deux variables X et Y ont donc même fonction de répartition, et donc même loi.

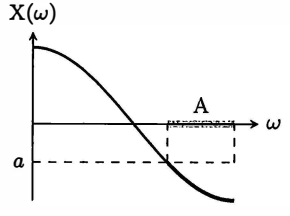
Exemple 2.7 Choisissons $\Omega = [0\,;1]$, muni de la probabilité suivante : pour toute partie A de $[0\,;1]$, $\mathbf{P}(A)$ est défini comme la « longueur » de A. En termes plus précis, la probabilité $\mathbf{P}$ n'est rien d'autre que la mesure de Lebesgue, sur la tribu des boréliens. On définit alors les deux variables aléatoires suivantes :

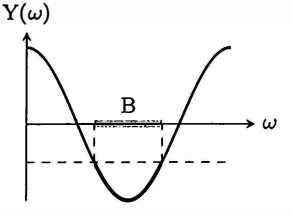
$$X(\omega) = \cos(\pi\omega) \qquad \text{et} \qquad Y(\omega) = X(2\omega) = \cos(2\pi\omega).$$

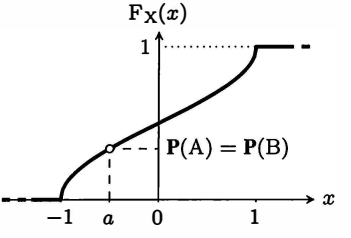
Sur les figures ci-dessous, on choisit une valeur quelconque (par exemple $a = -1/2$), et on note $A := \{X \leqslant a\}$ et $B := \{Y \leqslant a\}$: alors

$$F_X(a) = F_Y(a) = 1 - \tfrac{1}{\pi}\operatorname{Arc}\cos a = \tfrac{1}{3}.$$

Un raisonnement semblable peut être effectué pour toute valeur de a.







§ 26 Extension de la notion de variables aléatoires

Considérons, encore une fois, l'exemple de la succession d'une infinité de tirages d'un dé équilibré. L'univers que l'on choisit pour modéliser l'expérience est l'ensemble non dénombrable $\Omega := \{1,2,3,4,5,6\}^{\mathbb{N}^*}$ des suites $\omega = (u_n)_{n\geqslant 1}$ à valeurs dans $\{1,2,3,4,5,6\}$. Définissons X comme le numéro du premier tirage dont le résultat est un « 6 ». Cette fonction X n'est *pas* une variable aléatoire ! En effet, elle n'est pas définie sur Ω tout entier : certaines suites de tirages ne donnent jamais le résultat 6. On peut contourner la difficulté de deux manières :

— Ou bien l'on admet qu'une variable aléatoire puisse n'être définie que sur un univers « réduit » Ω', strictement inclus dans Ω, mais de probabilité égale à 1. Autrement dit, la probabilité que X ne soit pas définie doit être nulle (événement négligeable) ; ainsi, ces occurrences ne devraient pas interférer dans les calculs.

— Ou bien l'on définit X sur Ω tout entier, en posant $\mathrm{X}(\omega) = +\infty$ pour tout tirage ω pour lequel aucun « 6 » ne sort jamais. On dira alors que X est « presque sûrement » à valeurs réelles si $\mathbf{P}\{\mathrm{X} = +\infty\} = 0$, et dans ce cas X méritera le nom de « variable aléatoire réelle ».

Pour que ces deux points de vue soient équivalents (notamment, lorsque l'on mènera des calculs d'espérance, de variance, etc.), il est nécessaire d'introduire de nouvelles règles de calcul ; on conviendra donc des résultats suivants :

opération	résultat	condition
$x \times (+\infty)$	$+\infty$	$x > 0$
$x \times (+\infty)$	$-\infty$	$x < 0$
$x \times (-\infty)$	$-\infty$	$x > 0$
$x \times (-\infty)$	$+\infty$	$x < 0$
$0 \times (+\infty)$	0	
$0 \times (-\infty)$	0	
$x + \infty$	$+\infty$	$x \neq -\infty$
$x - \infty$	$-\infty$	$x \neq +\infty$
$+\infty - \infty$	indéfini	

TABLE 2.1 — *Règles de calcul avec* $+\infty$ *et* $-\infty$.

La règle la plus importante à retenir (la seule qui ne soit pas entièrement intuitive) est :

$$0 \times (+\infty) = 0 \times (-\infty) = 0.$$

Nous verrons que c'est bien la règle naturelle à poser pour assurer la cohérence des deux points de vue précédents.

On peut maintenant prouver que X est « presque sûrement à valeurs réelles », c'est-à-dire que $\mathbf{P}\{\mathrm{X} < +\infty\} = 1$; le plus simple est sans doute ici de décomposer cette quantité sous la forme

$$\mathbf{P}\{\mathrm{X} < +\infty\} = \sum_{n\in\mathbb{N}^*} \mathbf{P}\{\mathrm{X} = n\}.$$

Or, si $n \geqslant 1$, la probabilité que la première occurrence d'un « 6 » soit au n-ième tirage est

$$\mathbf{P}\{\mathrm{X} = n\} = \frac{1}{6} \cdot \left(\frac{5}{6}\right)^{n-1}$$

(puisque l'on doit obtenir un chiffre entre 1 et 5 sur $n-1$ tirages, et 6 sur le

dernier), ce qui mène à

$$\mathbf{P}\{X<+\infty\} = \frac{1}{6}\sum_{n=1}^{\infty}\left(\frac{5}{6}\right)^{n-1} = \frac{1}{6}\cdot\frac{1}{1-5/6} = 1.$$

Variables discrètes §27

Une variable aléatoire réelle est dite discrète si son image $X(\Omega)$ est un ensemble « discret », c'est-à-dire fini ou dénombrable. On peut donc l'écrire sous la forme

$$X(\Omega) = \{x_k\ ;\ k \in K\},$$

où K est un ensemble valant $\{1,\dots,n\}$ dans le cas fini, et $\mathbb{N}$ dans le cas dénombrable. Comme nous l'avons vu à la section 1.2 page 39, la simple connaissance des probabilités $\mathbf{P}\{X = x_k\}$ suffit à caractériser la loi de X. En effet, pour toute partie $A \subset \mathbb{R}$, on accède à la probabilité de l'événement $\{X \in A\}$ par une simple sommation de série :

$$\mathbf{P}\{X \in A\} = \sum_k \mathbf{P}\{X = x_k\}$$

où la sommation a lieu sur l'ensemble des indices k tels que $x_k \in A$. Cette égalité est justifiée par la σ-additivité de la fonction **P**, et le fait que la famille des $\{X = x_k\}$ vérifie

$$\Omega = \bigcup_{k\in K}\{X = x_k\} \qquad \textit{union disjointe !}$$

Ainsi,

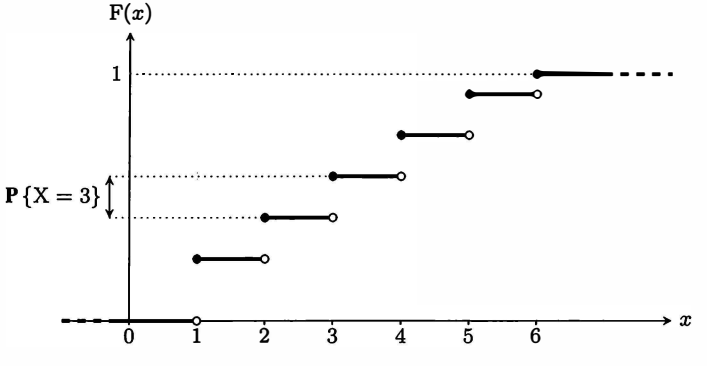
$$\mathbf{P}\{X\in A\} = \mathbf{P}\Big(\{X\in A\}\cap\bigcup_{k\in K}\{X=x_k\}\Big) = \mathbf{P}\Big(\bigcup_{k\in K}\{X\in A\}\cap\{X=x_k\}\Big)$$
$$= \sum_{k\in K}\mathbf{P}\big(\{X\in A\}\cap\{X=x_k\}\big) = \sum_{\substack{k\in K\ ;\\ x_k\in A}}\mathbf{P}\{X=x_k\}.$$

La fonction de répartition d'une variable aléatoire discrète est une fonction continue à droite, présentant un éventuel saut de discontinuité en chaque valeur x_k, constante sur tout intervalle ne contenant aucune des valeurs prises par X.

Exemple 2.8 Considérons par exemple le tirage d'un dé truqué, dont les probabilités sont, pour les nombres 1 à 6, respectivement

$$\frac{1}{4},\ \frac{1}{6},\ \frac{1}{6},\ \frac{1}{6},\ \frac{1}{6} \quad \text{et} \quad \frac{1}{12}.$$

Notons X la variable égale au résultat d'un tirage de ce dé. Sa fonction de répartition est représentée sur le graphe ci-dessous, sur lequel on a fait explicitement apparaître la continuité à droite :



§ 28 Variables à densité

Certaines variables aléatoires peuvent prendre un ensemble continu de valeurs, comme un intervalle de $\mathbb{R}$: c'est ainsi par exemple que l'on peut modéliser la vitesse d'une particule d'un gaz. Dans de nombreux cas, la probabilité de trouver la valeur de la variable X dans un intervalle $]x\,;\,x+\delta x]$ est (au premier ordre), proportionnelle à δx :

$$\mathbf{P}\{\mathrm{X}\in\,]x\,;\,x+\delta x]\,\}\approx f(x)\,\delta x,$$

le facteur de proportionnalité, ici noté $f(x)$, pouvant évidemment varier avec x.

Exemple 2.9 (Distribution des vitesses dans un gaz) Le spectre d'émission d'un gaz chaud comporte un certain nombre de raies, bien identifiées depuis l'apparition des techniques de spectroscopie. Ces raies correspondent, en principe, à une fréquence bien précise du spectre. Or, lorsqu'on les observe, elles présentent un certain étalement. Une partie de cet étalement s'explique par l'agitation, d'origine thermique, des atomes. La vitesse d'un atome selon la direction (dite *radiale*) d'observation provoque, par effet Doppler, un léger décalage de la fréquence observée. Ainsi, le profil des raies d'émission traduit-il, pour une partie du moins, la distribution des valeurs de la vitesse des atomes. Celle-ci suit la loi de Maxwell-Boltzmann : la probabilité qu'un atome ait une vitesse radiale comprise entre v et $v+\mathrm{d}v$ est

$$\mathrm{d}p=\mathrm{e}^{-mv^2/2k_{\mathrm{B}}\mathrm{T}}\,\mathrm{d}v,$$

où k_{B} est la constante de Boltzmann, T la température du gaz et m la masse de l'atome. Cette loi de distribution, appelée maxwellienne par les physiciens, est nommée *gaussienne* ou *normale* par les mathématiciens. La mesure de l'étalement des raies renseigne donc sur la température, la largeur typique d'étalement étant proportionnelle à $\sqrt{\mathrm{T}}$.

Ce qui est remarquable, c'est que l'agitation thermique, qui est un phénomène hautement imprévisible par essence, se laisse modéliser avec une précision stupéfiante par un modèle probabiliste. Cela est dû au fait que le nombre de particules dans un fragment macroscopique de matière est gigantesque (des milliards de milliards au bas mot), ce qui permet à certains théorèmes limite de jouer à plein.

En réalité, d'autres phénomènes physiques, comme l'effet Zeeman, font que l'étalement d'une raie spectrale est plus large — mais c'est une autre histoire, et cela ne grève en rien l'approche probabiliste de l'étude de cet étalement.

La quantité

$$f(x)=\lim_{\delta x\to 0}\frac{\mathbf{P}\{\mathrm{X}\in\,]x\,;\,x+\delta x]\,\}}{\delta x}=\lim_{\delta x\to 0}\frac{\mathrm{F}(x+\delta x)-\mathrm{F}(x)}{\delta x}$$

n'est rien d'autre que la dérivée de la fonction de répartition : $f=\mathrm{F}'$.

Lorsqu'une variable aléatoire possède cette propriété, on dit qu'elle **admet une densité**, et que f est une **densité** de X. La probabilité que X prenne une valeur dans un ensemble A se calcule par

$$\mathbf{P}\{\mathrm{X}\in\mathrm{A}\}=\int_{\mathrm{A}}f(t)\,\mathrm{d}t.$$

Cette propriété est évidente si A est un intervalle. Par σ-additivité, elle est vraie pour toute réunion finie ou dénombrable d'intervalles et, finalement, pour tout ensemble A « raisonnable[3] ».

Exemple 2.10 (Loi exponentielle : une modélisation) On cherche à décrire mathématiquement la constatation physique suivante : un atome radioactif a, à chaque instant, la possibilité de se désintégrer spontanément. On suppose que, pour des durées très brèves δt, la probabilité qu'il se

(3). En termes techniques : pour tout élément de la tribu des événements, qui est ici la tribu de boréliens, engendrée par les intervalles. En pratique : toutes les parties qui intéressent l'homme de la rue.

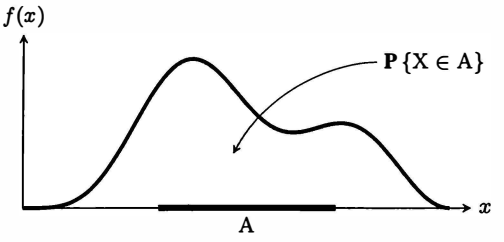


FIGURE 2.1 — *La densité de probabilité f d'une variable aléatoire permet, par intégration, de calculer* $\mathbf{P}\{X \in A\}$ *pour tout événement* A.

désintègre entre t et $t+\delta t$ — sachant qu'à l'instant t il n'est évidemment pas désintégré — ne dépende pas de t, mais soit simplement, au premier ordre, proportionnelle à δt : on l'écrit donc sous la forme $\approx \alpha\,\delta t$. Pour formaliser cela, supposons que, à l'instant $t=0$, l'atome soit présent ; notons alors $G(t)$ la probabilité qu'il ne soit pas encore désintégré à l'instant $t>0$. La probabilité qu'il se soit désintégré entre les instants t et $t+\delta t$ est donc $G(t)-G(t+\delta t)$; la probabilité du même événement, conditionnée par le fait que l'atome n'est pas désintégré à l'instant t, est

$$\frac{\mathbf{P}\{\text{désintégration avant } t+\delta t, \text{ mais pas avant } t\}}{\mathbf{P}\{\text{pas de désintégration avant } t\}} = \frac{G(t)-G(t+\delta t)}{G(t)} \approx \alpha\,\delta t.$$

Faisons tendre δt vers 0, on obtient

$$\frac{G'(t)}{G(t)} = -\alpha,$$

ce que l'on intègre, sur $\mathbb{R}^+$, en $G(t) = Ce^{-\alpha t}$. Puisque l'on a supposé que l'atome ne s'était pas désintégré à l'instant 0, on a donc $G(0)=1$, ce qui fixe la constante à 1.

Si l'on note X la variable aléatoire égale à l'instant de désintégration de l'atome, la fonction de répartition de X est :

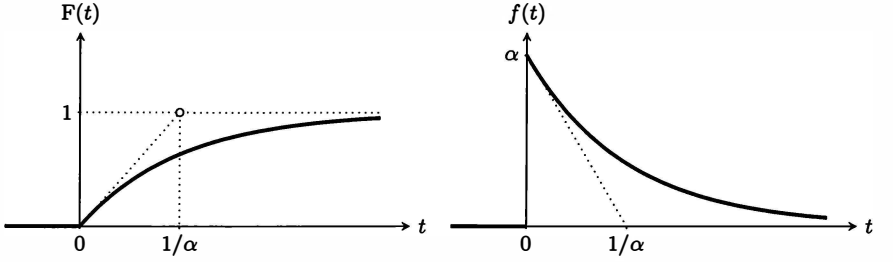
$$\forall t>0 \qquad F(t) = \mathbf{P}\{X \leqslant t\} = 1-G(t) = 1-e^{-\alpha t}.$$

La densité de probabilité de désintégration est, pour $t>0$:

$$f(t) = \frac{dF}{dt} = F'(t) = \alpha e^{-\alpha t},$$

et elle est nulle pour $t<0$. On dit que X suit la loi exponentielle de paramètre α, et on écrit $X \rightsquigarrow \mathscr{E}(\alpha)$.

Les graphes ci-dessous représentent respectivement la fonction de répartition et la densité d'une variable aléatoire de loi exponentielle $\mathscr{E}(\alpha)$.



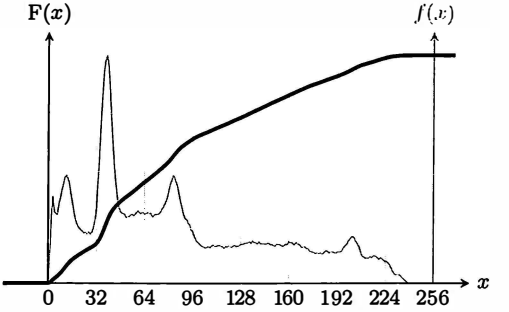
Cette modélisation nous permet de retrouver, par le calcul, un résultat physiquement évident (« l'atome finit, avec une probabilité égale à 1, par se désintégrer au cours du temps »). Si cela n'avait pas été le cas, X n'aurait pas reçu la dénomination de « variable aléatoire réelle », puisqu'elle aurait pris la valeur « $+\infty$ » avec une probabilité non nulle.

Remarque 2.11 (Que reste-t-il de X *dans* F_X *?)* Comme son nom l'indique, la fonction de répartition d'une variable aléatoire X contient l'information sur la « fréquence » des valeurs qu'elle peut prendre ; toute autre information n'apparaît pas dans la fonction F_X.

La plupart du temps, l'univers Ω est dépourvu de structure, ou du moins de structure ayant une signification importante, les objets ω peuvent être entièrement abstraits (comme on l'a dit, il arrive même que l'on modélise une expérience en postulant l'existence d'une variable aléatoire

de loi donnée, sans que l'univers Ω lui-même soit précisé). Notamment, la notions de proximité entre deux points ω_1 et ω_2, même lorsqu'elle peut être définie (par exemple si $\Omega = \mathbb{R}$) n'a pas de signification intéressante. La figure ci-dessous montre, à gauche, une variable aléatoire X définie sur un univers Ω représenté par un rectangle du plan ; les valeurs possibles de X sont des entiers entre 0 et 255 (que l'on assimilera gaillardement à des réels variant continûment entre 0 et 255), et sont représentées par des niveaux de gris (noir pour 0, blanc pour 255). À droite, la fonction de répartition de cette variable nous renseigne sur la statistique des valeurs prises par X. Sur le même graphe, on représente la « densité » correspondante (l'échelle est ajustée pour faire tenir le graphe). Il paraît clair qu'il est impossible de reconstituer X, *en tant qu'application*, à partir de sa fonction de répartition F_X ou de sa densité f.





§29 Est-ce tout ?

Non, loin de là. Certaines variables ne sont ni à densité, ni discrètes, mais un mélange des deux : un exemple classique est celui du temps d'attente à un feu de circulation, qui peut prendre des valeurs continues sur un intervalle $[0\,;\,T]$, où T est la durée du feu rouge dans le cycle des couleurs, mais tel que $\mathbf{P}\{X = 0\}$ est strictement positive (probabilité d'arriver pendant que le feu est au vert[4]). D'autres encore sont d'un type plus délicat à décrire[5]. Et puis, bien évidemment, on peut considérer des applications à valeurs vectorielles, ce qui fera l'objet de plusieurs chapitres de cet ouvrage.

Exemple 2.12 (Probabilité sur l'espace des phases) Considérons un système de N particules classiques[6], dont nous repérons les positions par un vecteur de $\mathbb{R}^{3N}$: $R := (r_1, r_2, \ldots, r_N)$. Leurs impulsions[7] sont décrites par un second vecteur $P := (p_1, p_2, \ldots, p_n) \in \mathbb{R}^{3N}$. Ce système est supposé ne pas avoir une énergie constante, mais pouvoir procéder librement à des échanges énergétiques avec une source de chaleur de température T, avec laquelle il est en équilibre thermique. L'énergie totale du système est donnée par une fonction $\mathscr{H}(R, P)$ de ces variables. On postule, en mécanique statistique, que la densité de probabilité de trouver le système décrit par la variable $X := (R, P)$, à valeurs dans $\mathbb{R}^{6N}$, est proportionnelle au **facteur de Boltzmann** $\exp\{-\beta\,\mathscr{H}(R, P)\}$, où le facteur β est défini par $\beta := (k_B T)^{-1}$, k_B étant la constante de Boltzmann et T la température absolue. Plus précisément, si l'on pose

$$Z := \iint_{\mathbb{R}^{6N}} e^{-\beta\,\mathscr{H}(R,P)}\, dR\, dP$$

(**fonction de partition**), la probabilité de trouver le système dans un volume V de l'espace des phases $\mathbb{R}^{6N}$ est

$$\mathbf{P}\{X \in V\} = \frac{1}{Z} \iint_V e^{-\beta\,\mathscr{H}(R,P)}\, dR\, dP.$$

(4). Ce modèle est détaillé dans l'exemple 13.13 page 296.

(5). Voir la remarque 13.17 page 298 pour un exemple explicite.

(6). Non quantiques, non relativistes, pour simplifier.

(7). Pour simplifier, l'impulsion d'une particule est sa quantité de mouvement, c'est-à-dire le produit $m\dot{x}$ de sa masse par sa vitesse. Cette définition est inexacte en présence d'un champ électromagnétique, ou dans un cadre relativiste ; cependant, si l'on se restreint à un système de particules entièrement décrit par la mécanique newtonienne et en l'absence d'interactions électromagnétiques, alors l'impulsion est exactement la quantité de mouvement.

Si la quantité Z est bien définie (c'est-à-dire si le facteur de Boltzmann est intégrable sur l'espace des phases), la formule précédente définit bien une loi de probabilité pour X. Le chapitre 29 propose une explication de l'origine du facteur de Boltzmann.

La connaissance de Ω est inutile § 30

Elle est même le plus souvent illusoire. Si nous cherchons à modéliser un phénomène complexe — par exemple, l'instant d'arrivée du dernier élève à un cours de mathématiques commençant à 8 h du matin — l'univers Ω est totalement hors de notre portée[8]. Il est bien plus commode — et fécond — de laisser Ω dans l'ombre, et de tenter une modélisation d'une variable aléatoire X par sa seule loi. Ainsi, on peut choisir de décréter que, pour un élève k donné (où k est un indice variant entre 1 et l'effectif n de la classe ou de l'amphithéâtre), l'instant d'arrivée X_k, mesuré en minutes à partir de 8 h, suit une certaine loi, disons par exemple une loi exponentielle[9] de paramètre $\alpha = 5$, c'est-à-dire que

$$\forall t \geqslant 0 \qquad F_k(t) := \mathbf{P}\{X_k \leqslant t\} = 1 - e^{-\alpha t}.$$

On admet maintenant qu'il existe un univers Ω convenable, muni d'une probabilité $\mathbf{P}$, ainsi que des variables aléatoires réelles $X_k : \Omega \to \mathbb{R}^+$ dont les fonctions de répartition F_k sont précisément celles que nous venons de décrire.

La modélisation peut alors continuer : on admet que cet univers peut être construit tel que les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes les unes des autres[10]. On peut enfin définir la variable aléatoire $T := \max(X_1, \ldots, X_n)$ et déterminer sa loi[11].

Remarque 2.13 Bien sûr, le mathématicien exigera un théorème autorisant une telle modélisation, c'est-à-dire affirmant l'existence de variables aléatoires $X_1, \ldots, X_n$, indépendantes, et suivant une loi donnée à l'avance. Un tel théorème existe bel et bien, comme le lecteur intéressé le verra dans l'annexe C, théorème C.5 page 715.

2.2 LA NOTION D'ESPÉRANCE

Points de vue du théoricien, point de vue de l'expérimentateur § 31

Considérons une variable aléatoire $X : \Omega \to \mathbb{R}$; cette variable aléatoire est censée représenter le résultat numérique d'une certaine expérience (tirage de dé, mesure d'une vitesse, position d'une particule...). On désire définir, si cela est possible, la notion de « valeur moyenne » de X. Deux approches sont possibles.

- On joue le *deus ex machina*, c'est-à-dire que l'on prend assez de hauteur pour observer toutes les réalisations possibles du hasard (tous les mondes parallèles possibles). On calcule alors une moyenne sur toutes les

(8). La répétition est à la base de la pédagogie : rappelons que l'univers doit permettre de décrire une foule de phénomènes, dont les pannes de réveil, l'humeur de chaque élève, les aléas de la circulation, les tentatives de l'éventuel(le) petit(e) ami(e) de retenir l'élève encore un peu sous les couvertures, etc. Toutes choses dont on conviendra qu'elles sont difficilement mathématisables.

(9). Les lois exponentielles et leurs propriétés seront exposées au chapitre 10, § 153, page 230.

(10). Là encore, la définition mathématique d'*indépendance* de variables aléatoires sera donnée ultérieurement, chapitre 16 page 341, la notion intuitive sous-jacente étant que l'instant d'arrivée d'un élève ou d'un groupe d'élèves ne dépend pas de celui des autres élèves. Hélas, en ce qui concerne l'indépendance, l'intuition est bien souvent trompeuse, et il est fréquent de trouver des raisonnements ayant une apparence de vérité, mais qui sont faux. On pourra se reporter au chapitre 34 pour plus de détails.

(11). C'est une situation classique, qui sera examinée au paragraphe § 185 et dans l'exercice 12.8.

valeurs $X(\omega)$, pour ω parcourant Ω, ces valeurs étant évidemment pondérées par la probabilité associée à chaque événement atomique ω. En d'autres termes, on tente une formule du genre

$$\text{« } m = \sum_{\omega \in \Omega} X(\omega)\, \mathbf{P}\big(\{\omega\}\big) \text{ ».} \qquad (*)$$

Ce point de vue est celui d'un mathématicien théoricien, pour qui cela ne pose aucun problème de considérer toutes les réalisations possibles du hasard ; c'est le point de vue qui sera le nôtre dans tous les chapitres de construction de l'espérance mathématique. Ce ne saurait être le point de vue d'un autre mathématicien, confronté à une ou plusieurs expériences, et qui tente de construire un modèle de ce qui est observé : par définition, il n'a accès qu'aux informations résultant d'une seule réalisation ω du hasard.

- Dans le second point de vue, on reste donc confiné à un seul choix de ω. L'idée est alors de multiplier les expériences, toujours dans la même occurrence ω de l'ensemble Ω (par exemple reprendre le même dé et le lancer un grand nombre de fois). On suppose que, lorsqu'une expérience est répétée, cela se traduit par la *duplication* de la variable aléatoire X : répéter l'expérience un nombre indéfini de fois revient à créer une suite $X_1, X_2, \ldots, X_n, \ldots$ de variables aléatoires, toutes de même loi que X. Insistons bien : si on lance plusieurs fois le dé, cela ne se traduit pas par un choix de différents ω, mais bien par la définition de plusieurs variables aléatoires, dont le résultat est pris pour la *même* réalisation ω du hasard. Si les expériences sont menées avec le soin nécessaire, on supposera que ces variables X_n sont indépendantes entre elles (et avec X)[12]. On peut alors chercher la valeur de

$$\lim_{n\to\infty} \tfrac{1}{n} \left(X_1 + X_2 + \cdots + X_n\right),$$

c'est-à-dire, pour le monde parallèle ω dans lequel on vit, chercher — si elle existe — la valeur de

$$m(\omega) := \lim_{n\to\infty} \tfrac{1}{n} \big(X_1(\omega) + X_2(\omega) + \cdots + X_n(\omega)\big).$$

Exemple 2.14 Prenons le cas d'un lancer de dé. On ne décrit pas l'univers Ω, mais celui-ci est suffisamment riche pour décrire une infinité de lancers de dés, c'est-à-dire qu'il est au moins « aussi gros » que $\{1,2,3,4,5,6\}^{\mathbb{N}^*}$. Chaque X_k est une variable aléatoire, pouvant prendre une valeur entière entre 1 et 6, et les X_k sont supposés indépendants entre eux. Que se passera-t-il, intuitivement ?

Si l'on croise les doigts, peut-être $m(\omega)$ est-elle définie ; si l'on croise les doigts encore plus fort, peut-être m est-elle une fonction constante sur Ω ? Disons-le tout de suite : la fonction m n'a aucune raison d'être constante[13] mais, si X suit une loi raisonnable, elle le sera *presque*. Plus précisément, et

---

(12). Cette hypothèse d'indépendance est absolument primordiale : poser $X_n = X$ pour tout $n \in \mathbb{N}$ conduirait à des variables aléatoires de même loi que X, mais non indépendantes de X, et inutilisable pour essayer de définir une valeur moyenne. Dans le cas d'un tirage de dé, toutes les expériences faites depuis des siècles semblent prouver que les résultats d'un tirage sont effectivement indépendants des tirages précédents.

(13). Considérons un simple tirage à *pile* ou *face* avec une pièce honnête. Dans un des mondes parallèles ω, chacun des tirages mène à *face*. Certes, à cet univers est associé une probabilité nulle, ce qui semble écarter l'argument. Remarquons tout de même que *chaque* succession infinie de tirages à *pile* ou *face* a également une probabilité nulle ; on ne peut donc pas écarter les cas gênants sans argument plus fin. Ce qu'il faut retenir est que chaque expérimentateur de chaque monde parallèle possible pourra trouver une valeur différente à m, voire ne pourra pas la définir.

ce sera le contenu du théorème appelé « Loi forte des grands nombres[14] », dans « presque toutes » les réalisations ω du hasard, la valeur de m sera égale à une certaine constante ; ainsi, presque tous les expérimentateurs de nos mondes parallèles seront d'accord. Seul une infime minorité trouvera une valeur « aberrante » de m — voire ne trouvera pas de limite à la suite des valeurs moyennes. Il reste évidemment à préciser les conditions sur X ainsi que la notion précise de « presque tous » (qui, on s'en doute, signifiera « une majorité vraiment écrasante »).

Ce point de vue est celui de l'expérimentateur, dont la tâche sera, notamment, d'évaluer des quantités numériques (probabilités, espérances, etc.).

Le problème de la sommation §32

Laissons pour l'instant de côté le point de vue de l'expérimentateur et concentrons-nous sur celui du théoricien, dont la tâche est de donner un sens à la formule (∗).

— Si Ω est un ensemble fini, aucun problème particulier ne se pose.

— Si Ω est dénombrable, on peut *a priori* sommer en passant par la théorie des séries numériques ou, plus précisément, des familles sommables[15].

Il est d'ailleurs possible d'effectuer les calculs, non pas en sommant sur les éléments ω de l'espace de départ, mais en considérant les valeurs prises par $X(\omega)$; c'est le point de vue qu'a adopté Henri Lebesgue pour définir sa théorie de l'intégrale. En regroupant les ω menant à la même valeur, on écrira

$$m = \sum_x x\, \mathbf{P}\{X = x\}\,.$$

Ce changement de point de vue permet notamment de calculer la valeur moyenne de X même si Ω n'est pas dénombrable, pour peu que l'ensemble des valeurs prises par X soit dénombrable[16].

— si Ω est infini et non dénombrable, il nous faudra sans doute remplacer la sommation discrète par une sommation *continue*, c'est-à-dire par une intégrale du genre

$$m = \int_\Omega X(\omega)\, d\mathbf{P}(\omega)$$

où tout est encore à définir dans ce qui n'est pour l'instant qu'une notation symbolique !

Cette opération est quelque peu délicate, puisqu'il faut définir sur Ω un procédé d'intégration ; c'est l'objet de la *théorie de la mesure*. Dans de nombreux cas pratiques, on peut définir la valeur moyenne de X sans passer par cette intégrale abstraite, en n'utilisant que la fonction de répartition de X, ou encore, si elle existe, sa fonction de densité. L'idée est encore celle de Lebesgue : plutôt que de sommer « horizontalement » en regardant les valeurs de X pour chaque élément ω, on évalue plutôt la taille de l'ensemble de ω tels que $X(\omega)$ soit dans un certain intervalle $]x\,;\,x+\delta x]$, qui est $\mathbf{P}\{X \in]x\,;\,x+\delta x]\} = F(x+\delta x) - F(x) \approx f(x)\,\delta x$. Il ne reste plus qu'à écrire que

$$m = \sum_x x\,[F(x+\delta x) - F(x)] \approx \sum x\, f(x)\,\delta x,$$

(14). Théorème 20.18 page 425.

(15). On verra qu'un problème surgit dès l'origine, car l'addition d'une infinité de nombres n'est pas une opération commutative : le résultat obtenu en sommant une famille dénombrable de réels ou de complexes peut dépendre de l'ordre de la sommation.

(16). La variable X est alors dite **discrète** ; les chapitres 6 à 8 sont consacrés à l'étude de telles variables.

où la sommation a lieu sur tous les x séparés de la même quantité δx. À la limite $\delta x \to 0$, on obtient alors une intégrale

$$m = \int_{-\infty}^{+\infty} x\, f(x)\, \mathrm{d}x.$$

Remarque 2.15 Lorsqu'il n'existe pas de fonction de densité, le passage à la limite $\delta x \to 0$ conduit à une autre intégrale, notée

$$m = \int_{-\infty}^{+\infty} x\, \mathrm{dF}(x)$$

appelée *intégrale de Riemann-Stieltjes* ; nous la retrouverons plus loin, lorsque l'espérance sera définie comme une intégrale abstraite sur un espace muni d'une mesure de probabilité.

2.3 ESPÉRANCE D'UNE VARIABLE DISCRÈTE OU À DENSITÉ

§ 33 **Cas discret : première (tentative de) définition**

Soit $\mathrm{X} : \Omega \to \mathbb{R}$ une variable aléatoire réelle discrète, c'est-à-dire prenant ses valeurs dans un ensemble dénombrable ; on peut donc indicer les éléments de $\mathrm{X}(\Omega)$: $\mathrm{X}(\Omega) = \{x_k \ ; \ k \in \mathbb{N}\}$. On note

$$p_k := \mathbf{P}\{\mathrm{X} = x_k\}$$

pour tout entier k. Si la série $\sum x_k\, p_k$ converge, définissons l'**espérance** de X par

$$\mathbf{E}(\mathrm{X}) := \sum_{k=0}^{\infty} x_k\, p_k.$$

C'est une « valeur moyenne de X », où chaque valeur possible x_k de X est pondérée par sa probabilité p_k d'apparition.

Si cette définition est conforme à l'intuition, la répétition d'un grand nombre d'expériences devrait mener, en moyenne, à la valeur $\mathbf{E}(\mathrm{X})$. Nous allons voir immédiatement que cela peut, dans certains cas, échouer lamentablement.

§ 34 **Deux expériences numériques**

Tout d'abord, voyons le cas où la répétition d'une expérience mène au résultat attendu. Considérons une variable aléatoire X à valeurs dans $\mathbb{N}^*$, telle que

$$\mathbf{P}\{\mathrm{X} = n\} = \frac{n}{(n+1)!}$$

et posons $\mathrm{G} = (-1)^{\mathrm{X}+1}\, \mathrm{X}$. Cette variable peut être obtenue dans une situation « concrète » de la manière suivante.

> *Rosencrantz et Guildenstern disposent d'une urne contenant une boule blanche et une boule noire. Guildenstern effectue des tirages dans cette urne en respectant le protocole suivant. Il commence par tirer aléatoirement une boule ;*
>
> - *si elle est noire, il arrête de tirer, la partie est terminée ;*
> - *si elle est blanche, il la remet dans l'urne et il y rajoute une boule noire ; puis il poursuit les tirages avec la même règle.*
>
> *Si le nombre* X *de tirages effectués par Guildenstern est impair et égal à n, il donne n euros à Rosencrantz — et on considère que le gain en euros de Rosencrantz vaut* $\mathrm{G} = +n$. *Si* X *est pair et égal à n, Rosencrantz donne n euros à Guildenstern — et on considère que le gain de Rosencrantz vaut* $\mathrm{G} = -n$.

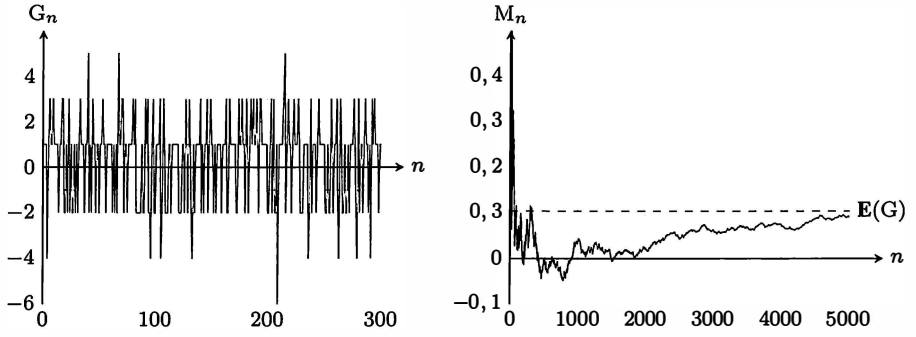


FIGURE 2.2 — À gauche : *gains successifs* G_n *de Rosencrantz au cours de 300 parties.* À droite : *évolution de la moyenne* $M_n = [G_1 + \cdots + G_n]/n$ *des gains au cours de 5000 parties. La convergence vers la valeur* $\mathbf{E}(G)$*, quoique lente, est bien visible.*

Le lecteur pourra calculer, en utilisant la formule des probabilités composées, qu'alors G admet une espérance et que

$$\mathbf{E}(G) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}\, n^2}{(n+1)!} = \frac{3}{e} - 1 \approx 0{,}104.$$

Rosencrantz et Guildenstern disputent, selon ces règles, un grand nombre de parties successives. Sur le graphe de gauche de la figure 2.2, on représente le gain G_n de Rosencrantz à la n-ième partie, pour $n = 1, \ldots, 300$; on voit que le résultat oscille globalement entre -5 et 5 : les gains ou les pertes plus importants sont très rares. Sur le graphe de droite, on affiche les moyennes successives des gains au cours de 5000 parties :

$$M_n = \frac{G_1 + G_2 + \cdots + G_n}{n}.$$

La suite des moyennes semble bien converger vers $\mathbf{E}(G)$. (On pourrait simuler une nouvelle série de 5000 parties : le graphe aurait la même allure).

Passons maintenant au cas défavorable ; pour cela, changeons les règles en modifiant le protocole.

Guildenstern tire une boule :
- *si elle est noire, il arrête de tirer ;*
- *si elle est blanche, il la remet dans l'urne et rajoute dans l'urne une boule blanche ; puis il poursuit les tirages avec la même règle.*

Notons Y le nombre de tirages effectués par Guildenstern ; on peut calculer que

$$\mathbf{P}\{Y = n\} = \frac{1}{n(n+1)}.$$

La série relative à l'espérance du gain $G' := (-1)^{Y+1} Y$ est la série harmonique alternée $\sum (-1)^n/n$. C'est une série convergente, mais pas absolument convergente. Sa somme vaut $1 - \ln(2) \approx 0{,}307$.

Dans le premier graphe de la figure 2.3, on affiche le gain de Rosencrantz au cours de 300 parties. Le résultat est très irrégulier et présente notamment des pics de très forte amplitude : on voit, à la 132e partie, une perte de 306 € ! Dans les trois graphes suivants, on affiche l'évolution de la moyenne des gains au cours de

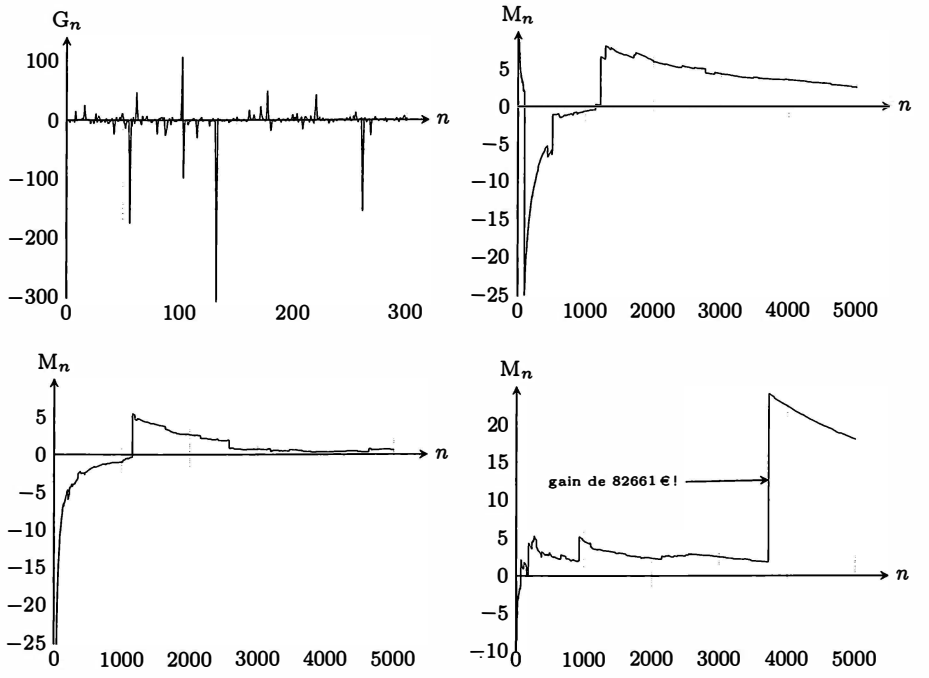


FIGURE 2.3 — (En haut à gauche) : *gains successifs* G_n *de Rosencrantz au cours de 300 parties. Comparativement à la figure 2.2, les gains exceptionnels sont nettement plus fréquents.* — (Sur les trois autres figures) : *évolution de la moyenne* M_n *des gains au cours de 5000 parties. Les gains exceptionnels empêchent la convergence de la suite* $(M_n)_{n\geqslant 1}$.

trois séries différentes de 5000 parties[17]. La suite des valeurs moyennes ne semble pas du tout converger : les pics de forte amplitude brouillent le signal et, même après avoir moyenné sur 5000 parties, le gain moyen de Rosencrantz est encore loin de $1-\ln(2)$. Notamment, sur le dernier graphe, le dernier grand saut correspond à un gain exceptionnel d'à peu près 83 000 € (moyenné sur 3700 parties, le saut est évidemment plus petit). Enfin, les trois graphes ont des allures très différentes.

§ 35 Séries commutativement convergentes

Pour définir la quantité $\mathbf{E}(X)$, la première chose que nous avons faite a été de choisir un dénombrement de $X(\Omega)$. Une fois ce choix fait, on a pu s'interroger sur la nature de la série $\sum x_k\, p_k$, et sur le calcul de sa somme[18]. Comme on le voit, la définition de la convergence et de la somme dépend donc, *a priori*, de l'ordre de sommation. La question naturelle qui surgit alors est : *l'ordre de sommation est-il, en réalité, indifférent ?*

---

(17). On admettra que la simulation informatique produit les mêmes résultats qu'une suite de tirages réels ! L'auteur avoue piteusement ne pas avoir eu le courage d'effectuer cinq mille parties de ce jeu sans aucun intérêt...

(18). Rappelons rapidement la technique : pour tout entier n, on note S_n la **somme partielle**

$$S_n := \sum_{k=0}^{n} x_k\, p_k$$

et on étudie la nature de la suite $(S_n)_{n\geqslant 0}$. Si cette suite converge et a pour limite S, on dit que la série $\sum x_k\, p_k$ **converge** et le nombre S est appelé **somme** de la série.

On pourrait penser que la réponse est positive[19] : hélas il n'en est rien, et l'on connaît de nombreux exemples de séries dont on peut modifier la somme (ou la nature) en en réordonnant simplement les éléments[20]. Les séries dont la somme ne dépend pas du choix de l'ordre de sommation sont les séries absolument convergentes.

Dans l'exemple précédent, la valeur moyenne de la variable aléatoire Y était « définie » par une série *semi-convergente* ; quelle que soit la signification que l'on donne à la somme de cette série, elle ne pourra pas être égale à une « valeur moyenne obtenue par répétition d'un grand nombre de tirages ».

Seconde définition § 36

Afin d'éviter ce qui semble être à la fois un écueil numérique (§ 34) et une inconsistance de définition (§ 35), on pose la définition suivante :

Définition 2.16 Soit X une variable aléatoire discrète, prenant un ensemble de valeurs $X(\Omega) = \{x_k \ ; \ k \in \mathbb{N}\}$ et de loi $p_k = \mathbf{P}\{X = x_k\}$. Si la série $\sum x_k p_k$ converge *absolument*, on dit que X **admet une espérance**, et on définit l'**espérance** de X comme la quantité

$$\mathbf{E}(X) = \sum_{k=0}^{\infty} x_k p_k.$$

On note $\boldsymbol{\mathcal{L}^1(\Omega, \mathfrak{T}, \mathbf{P})}$ — ou, plus familièrement, $\boldsymbol{\mathcal{L}^1}$ — l'ensemble des variables aléatoires admettant une espérance.

Remarque 2.17 Cette définition permettra l'interprétation classique de l'espérance, asymptotiquement égale à la moyenne arithmétique d'un grand nombre de résultats obtenus par répétition d'une même expérience (loi des grands nombres). Elle a également l'avantage de coïncider avec la notion de fonction intégrable au sens de Lebesgue.

Avec cette définition, le lecteur peut se persuader, crayon en main, de la validité des propriétés suivantes (qui seront toutes démontrées au chapitre 6).

■ **Proposition 2.18 (Propriétés des variables de $\mathcal{L}^1$)** *Soient* X *et* Y *des variables aléatoires réelles discrètes, et* α, β *des réels.*

i) Si X *est bornée, alors* X *admet une espérance.*

ii) Si X *et* Y *admettent une espérance, alors* $\alpha X + \beta Y$ *admet une espérance. Notamment, l'ensemble* $\mathcal{L}^1$ *est un espace vectoriel.*

iii) L'application $\mathbf{E} : \mathcal{L}^1 \to \mathbb{R}$ *est linéaire :* $\mathbf{E}(\alpha X + \beta Y) = \alpha\,\mathbf{E}(X) + \beta\,\mathbf{E}(Y)$.

iv) S'il existe un réel a tel que $\mathbf{P}\{X = a\} = 1$ *(on dit que* X *est* **presque certaine***), alors* $\mathbf{E}(X) = a$. *Notamment, si* X *est une variable aléatoire admettant une espérance et si l'on note de la même manière le réel* $\mathbf{E}(X)$ *et la variable aléatoire certaine prenant cette valeur, alors* $\mathbf{E}\big(\mathbf{E}(X)\big) = \mathbf{E}(X)$.

---

(19). Dans une somme finie, l'ordre de sommation est évidemment indifférent...

(20). Un exemple vérifiable à la main est donné dans l'annexe A, paragraphe A12. D'une manière générale, si $\sum u_n$ est une série convergente, mais pas absolument convergente, et si ℓ est un réel quelconque, il existe un ordre de sommation particulier tel que la série $\sum u_{\varphi(n)}$ converge vers ℓ.

§ 37 Espérance d'une variable à densité

Par analogie avec la définition 2.16, si X est une variable aléatoire admettant une densité f, la valeur moyenne de X est l'intégrale de ses valeurs possibles, pondérées par la densité f :

$$\mathbf{E}(\mathrm{X}) := \int_{-\infty}^{+\infty} x\, f(x)\, \mathrm{d}x.$$

Pour les mêmes raisons que précédemment, on exigera que cette intégrale soit convergente en valeur absolue ; par positivité de f, il revient au même d'exiger que les deux intégrales

$$\int_0^{+\infty} x\, f(x)\, \mathrm{d}x \qquad \text{et} \qquad \int_{-\infty}^{0} x\, f(x)\, \mathrm{d}x$$

soient finies.

La notion d'espérance sera définie plusieurs fois dans cet ouvrage. Au chapitre 11, la définition présente de l'espérance d'une variable aléatoire à densité sera reprise et approfondie, et au chapitre 14 une définition plus générale, mais également plus abstraite, sera donnée, qui s'appliquera à n'importe quelle variable aléatoire.

EXERCICES

♦ ***Exercice 2.1 (Le paradoxe de l'inspection (1/2))*** Depuis des temps immémoriaux, les dix mille temples des adorateurs de Cthulhu sont éclairés chacun par une ampoule basse consommation, économiques, écologiques et fournissant un éclairage glauque plébiscité par lesdits adorateurs. Dans chaque temple, un employé change l'ampoule aussitôt qu'elle rend l'âme, ayant à sa disposition un stock considérable d'ampoules. Celles-ci sont de deux types, absolument indiscernables à l'œil nu : celles dont la durée de vie est de 1 jour *(type A)*, celles dont la durée de vie est de 30 jours *(type B)*. Il y a deux fois plus d'ampoules de type A que de type B. Ainsi, lorsqu'il remplace l'ampoule défunte, l'employé a une probabilité 2/3 de choisir une de celles de type A, et donc une probabilité 1/3 de choisir une de celles de type B.

i) Quelle est la durée moyenne de vie d'une ampoule ?

ii) Le syndicat des employés se plaint de ce que la durée de vie des ampoules est trop courte et que ses membres croulent sous leur charge. Le grand prêtre de Cthulhu décide d'envoyer, simultanément, un inspecteur dans chaque temple. Chaque inspecteur reste alors sur place jusqu'à ce que l'ampoule claque. Ensuite, ils reviennent faire leur rapport.

(a) Si toutes les ampoules avaient une même durée de vie de $\Delta = 30$ jours, quelle serait la « durée de vie restante » moyenne observée par les inspecteurs ?

(b) Dans la situation effective où les deux tiers des ampoules ont une durée de vie de 1 jour, quelle est la « durée de vie restante » moyenne observée par les inspecteurs ? Quelle réponse recevra le syndicat des employés ?
(On prendra garde à calculer d'abord la probabilité qu'un inspecteur donné observe une ampoule de type A : il y a quand même beaucoup plus de chances d'arriver dans un temple éclairé par une ampoule résistante !)

(Le lecteur plus avancé pourra se reporter à l'exerice 30.3 page 609.)

♦ ***Exercice 2.2 (Politique nataliste aberrante)*** Les seigneurs d'un petit pays du Moyen-âge veulent un fils (pour transmettre leur héritage) et un seul (pour ne pas diviser l'héritage) et font des enfants en s'arrêtant au premier fils.

On modélise le nombre d'enfants d'une famille par une variable aléatoire N, celui du nombre de filles par F et celui de garçons par G. On suppose que, à chaque naissance, il y a une chance sur deux que le bébé soit un garçon.

i) Déterminer la loi de N, celle de F et celle de G. En déduire $\mathbf{E}(\mathrm{N})$, $\mathbf{E}(\mathrm{G})$ et $\mathbf{E}(\mathrm{N})$. Quel est le pourcentage de garçons dans la population ?

ii) On note Q la proportion de garçons dans une famille. Calculer la moyenne de la proportion de garçons, c'est-à-dire $\mathbf{E}(\mathrm{Q})$.

♦ ***Exercice 2.3 (Modélisation informatique)*** On désire effectuer une modélisation informatique de l'exercice 2.2.

i) Écrire une fonction réalisant une série de naissances et retournant le premier rang où un garçon est né, c'est-à-dire le nombre N d'enfants de la famille.

ii) Sur un million de tirages, évaluer la moyenne du nombre de filles, la proportion moyenne de garçons dans la population et la moyenne de la proportion de garçons dans chaque famille.

On comparera ces nombres avec, respectivement, l'espérance du nombre de filles $\mathbf{E}(\mathrm{F}) = 1$, le rapport $\mathbf{E}(\mathrm{G})/\mathbf{E}(\mathrm{N}) = 0{,}5$ et l'espérance de la proportion de garçons $\mathbf{E}(\mathrm{G}/\mathrm{N}) = \ln 2 \approx 0{,}693$.

SOLUTIONS

♦ **Solution 2.1**

i) La durée de vie moyenne d'une ampoule est obtenue en moyennant sur l'ensemble des ampoules... On trouve donc une durée de $(2 \times 1 + 1 \times 30)/3 = 32/3 \approx 10,66$ jours. (En gros, les employés doivent changer l'ampoule trois fois par mois.)

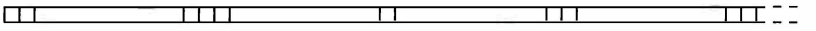
ii) (a) Si l'on arrive aléatoirement (et uniformément) dans un intervalle de temps de largeur Δ donné, le temps moyen avant la fin de l'intervalle doit être de $\Delta/2$; on peut le calculer explicitement à l'aide de la notion d'espérance d'une loi uniforme de densité $1/\Delta$ sur l'intervalle $[0\,;\,\Delta]$:

$$\mathbf{E}(\mathrm{T}) = \int_0^\Delta \frac{t}{\Delta}\,\mathrm{d}t = \frac{\Delta}{2}.$$

(On peut également invoquer un argument de symétrie : si l'on renverse le sens du temps, le temps moyen de vie écoulé doit être égal au temps moyen restant, dont la somme vaut Δ.)

Ainsi, les inspecteurs trouveraient une durée de vie moyenne résiduelle de 15 jours, et concluraient avec raison que la durée de vie moyenne est de 30 jours.

(b) Dans chaque temple, on observe une suite d'intervalles de temps dont deux tiers durent 1 jour, le dernier tiers durant 30 jours. Si l'on arrive à un moment quelconque de cette suite, il y a donc une probabilité $p = 2/32$ que ce soit pendant un des petits intervalles, et une probabilité $q = 1 - p = 30/32$ que ce soit pendant un grand intervalle ! Ci-dessous, une figure représentant une succession typique de durées de vie d'ampoules (l'échelle n'est pas respectée).



Ainsi, le temps moyen observé est

$$\langle t \rangle = \frac{p}{2}\,1 + \frac{1-p}{2}\,30 = \frac{902}{64} \approx 14,09$$

Les rapports remis au Grand Prêtre montrent que la durée de vie restante moyenne est de 14 jours environ quand on arrive au hasard ; le Grand Prêtre en déduira (faussement) qu'une ampoule dure en moyenne 28 jours et ne répondra pas favorablement à la requête du syndicat.

♦ **Solution 2.2**

i) N suit une loi géométrique $\mathscr{G}(1/2)$, donc $\mathbf{E}(\mathrm{N}) = 2$. Presque sûrement, on a $\mathrm{G} = 1$[21] donc $\mathbf{E}(\mathrm{G}) = 1$. Enfin, $\mathrm{N} = \mathrm{G} + \mathrm{F}$ donc F suit la loi géométrique $\mathscr{G}_{\mathbb{N}}(1/2)$, et $\mathbf{E}(\mathrm{F}) = 1$.

Le pourcentage de garçons dans la population totale est

$$p = \frac{\mathbf{E}(\mathrm{G})}{\mathbf{E}(\mathrm{N})} = \frac{1}{2}.$$

ii) Lorsque le garçon arrive à la k-ième naissance (ce qui arrive avec la probabilité $1/2^k$, la proportion de garçons dans la famille est $1/k$. Ainsi,

$$\mathbf{E}(\mathrm{Q}) = \mathbf{E}\left(\frac{\mathrm{G}}{\mathrm{N}}\right) = \sum_{k=1}^{\infty} \frac{1}{k \cdot 2^k} = \ln 2.$$

---

[21]. G peut prendre la valeur 0 lorsque seules des filles naissent... jusqu'à la fin des temps... ce qui doit arriver avec une probabilité nulle.

La proportion moyenne de garçons (dans la population totale) n'est donc pas égale à la moyenne des proportions des garçons (dans chaque famille). Et pour cause : s'il y a bien en moyenne autant de garçons que de filles (la Nature ne se laisse pas berner par des visées politiques), une famille sur deux affiche déjà une proportion de garçons égale à 1 !

♦ **Solution 2.3** On propose ici un programme élémentaire en Python 3.x.

i) Le rang du premier garçon (et donc le nombre total d'enfants) peut être simulé par le programme suivant :

```
def RangG():
    rang = 1
    while randint(0,1) == 0:    # une chance sur deux
        rang += 1
    return rang
```

ii)

```
def StatNaissances(n):
    filles = 0                      # pas encore de filles
    garcons = n                     # on sait déjà ce que sera le nombre de garçons
    prop_somme = 0.                 # on sommera les valeurs des proportions
                                    # (nombre flottant) dans chaque famille
    for _ in range(n):
        rang = RangG()              # nombre d'enfants dans la famille
        filles += rang - 1          # on met à jour le nombre de filles
        prop = 1 / rang             # proportion locale de garçons
        prop_somme += prop          # on met à jour la somme des proportions
    enfants = filles + n            # nombre total d'enfants
    prop_moyenne = prop_somme / n
    return filles/n, garcons/enfants, prop_moyenne

>>> StatNaissances(10**6)
(0.99913, 0.5002175946536743, 0.6934705452757103)
```

La concordance est ici assez bonne, avec une précision de 10^{-3}.

Deuxième partie

Théorie et Pratique

Chapitre 3

Espaces probabilisés

Si les mathématiques — système spécialisé de quelques signes, fondé sur l'intelligence et gouverné par elle — impliquent des situations insaisissables, et sont l'objet de permanentes discussions, que dire des obscurités du langage, amas confus de milliers de symboles maniés presque au hasard ?

Jorge Luis Borges, La jouissance littéraire

Le formalisme désormais employé dans la théorie des probabilités est celui que Kolmogorov introduisit dans ses Méthode analytiques de la théorie des probabilités *[61] de 1931, et surtout dans son opuscule* Fondements de la théorie des probabilités *[62] paru en 1933*[1]*; bien que cette axiomatisation soit aujourd'hui d'un usage bien établi*[2]*, elle mit longtemps à s'imposer, et ne fut guère connue avant les années 50.*

L'un des grands mérites de Kolmogorov fut d'avoir fourni un cadre entièrement abstrait à la théorie — qui n'est mise en parallèle avec l'expérience commune que dans le cas des probabilités finies —, cadre qui permet la généralisation de quelques notions intuitives et simples à des situations où, a priori, *aucun calcul ne semblait possible. Pour cela, il fait en permanence appel à des éléments de théorie des ensembles, et notamment à la notion de* tribu *(ou* σ-algèbre*), déjà évoquée au chapitre 1.*

Dans ce chapitre, nous passons en revue le vocabulaire nécessaire à la bonne compréhension du reste de l'ouvrage ; le lecteur est invité à parcourir, sans forcément s'y attarder en première lecture, les grandes lignes de ce formalisme — mais à s'y reporter chaque fois que le besoin s'en fera ressentir. Les points les plus importants sont signalés, en marge, par un ou deux symboles « ▶ » ou « ◀ ». Au contraire, certains passages signalés par un astérisque $(*)$ *peuvent être omis en première lecture.*

---

(1). Ce petit livre est encore très lisible malgré son âge et en dépit de notations surprenantes pour le lecteur d'aujourd'hui, comme $\bigcup_n A_n$ pour l'union, ou $\bigcap_k A_k = A_1 A_2 \cdots A_n$ pour l'intersection. La structure de la présentation a manifestement influencé un nombre incalculable d'ouvrages sur le sujet ! L'influence considérable de Kolmogorov en mathématiques est présentée de manière très agréable dans l'ouvrage de Charpentier, Lesne et Nikolski [16].

(2). Ce qui n'empêche pas d'autres types de présentation ; l'excellent ouvrage de Rényi [86] propose une approche alternative des espaces probabilisés, en prenant comme modèle de base les espaces de probabilités conditionnelles, dont les espaces usuels de Kolmogorov ne sont que des cas particuliers.

3.1 PARTIES, ALGÈBRES, TRIBUS

§38 Unions disjointes

▶▶ Nous rappelons quelques notations usuelles en théorie des ensembles.

- Si A et B sont des parties de Ω, on note $A \setminus B$ l'ensemble des éléments de A qui ne sont pas dans B :

$$A \setminus B = A \cap B^c = \{x \in A \,;\, x \notin B\}.$$

- Si B est de plus inclus dans A, on notera $A - B$ l'ensemble $A \setminus B$
- Si A et B sont des parties disjointes, on note $A + B$ leur union disjointe[3] :

$$A + B := A \cup B \qquad \text{quand } A \cap B = \varnothing.$$

De même, si $(A_i)_{i\in I}$ est une famille de parties deux à deux disjointes, on note $\sum_i A_i$ la réunion disjointe $\bigcup_i A_i$.

⇙3.1

- Enfin, il est commode d'étendre les notations précédentes aux familles vides et de définir, si I est l'ensemble vide,

$$\bigcup_{i\in I} A_i = \varnothing \qquad \sum_{i\in I} A_i = \varnothing \qquad \bigcap_{i\in I} A_i = \Omega.$$

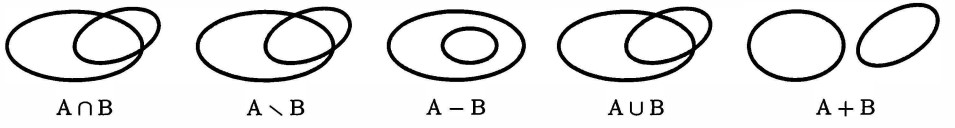


FIGURE 3.1 — *Opérations ensemblistes.*

§39 Algèbres

Un ensemble $\mathfrak{A}$ de parties de Ω est appelée **algèbre** sur Ω si elle est stable par union et passage au complémentaire, et si elle contient Ω :

i) $\Omega \in \mathfrak{A}$;

ii) pour tous $A, B \in \mathfrak{A}$, $A \cup B$ et A^c sont dans $\mathfrak{A}$.

De ces propriétés, on déduit également que

iii) $\mathfrak{A}$ est stable par intersection, car $A \cap B = (A^c \cup B^c)^c$;

iv) $\mathfrak{A}$ est stable par différence, car $A \setminus B = A \cap B^c$;

v) $\mathfrak{A}$ contient la partie vide $\varnothing = \Omega^c$.

Par itération,

vi) $\mathfrak{A}$ est également stable par unions et intersections finies.

Remarque 3.1 Le nom d'« algèbre » vient de ce que la loi « $\cup$ » peut être vue comme une addition (d'élément neutre $\varnothing$), et loi « $\cap$ » comme une multiplication (d'élément neutre Ω). Le quadruplet $(\mathfrak{A}, \cup, \cap, \cdot^c)$ porte le nom d'*algèbre de Boole.*

Exemple 3.2 Posons $\Omega = \mathbb{R}$, notons $\mathfrak{A}_f$ l'ensemble des parties de cardinal fini de Ω, et $\mathfrak{A}_c$ l'ensemble des parties dont le complémentaire est de cardinal fini. Alors $\mathfrak{A} := \mathfrak{A}_f \cup \mathfrak{A}_c$ est une algèbre.

---

(3). Cette union est parfois notée $A \sqcup B$ ou $A \coprod B$.

Remarque 3.3 Si $\mathfrak{E}$ est une partie de $\mathfrak{P}(\Omega)$, l'intersection de toutes les algèbres contenant $\mathfrak{E}$ est une algèbre, appelée **algèbre engendrée par** $\mathfrak{E}$ et notée $\alpha(\mathfrak{E})$. C'est la plus petite algèbre contenant $\mathfrak{E}$, et elle peut se décrire comme l'ensemble de toutes les parties s'écrivant avec un nombre fini d'unions, d'intersections et de passage au complémentaire appliqués à des éléments de $\mathfrak{E}$.

Tribus

§ 40

De manière générale, une algèbre n'est pas stable par union ou intersection dénombrable ; cette propriété étant intéressante pour une théorie des probabilités sur un ensemble infini, on introduit un axiome supplémentaire pour définir une structure plus riche que l'algèbre.

Une **tribu** sur Ω est une partie $\mathfrak{T}$ de $\mathfrak{P}(\Omega)$ et vérifiant les propriétés suivantes : ◀

i) $\Omega \in \mathfrak{T}$;

ii) pour tout A de $\mathfrak{T}$, son complémentaire A^c est dans $\mathfrak{T}$;

iii) $\mathfrak{T}$ est stable par unions (finies ou) dénombrables : si $(A_n)_{n\geqslant 0}$ est une suite d'éléments de $\mathfrak{T}$, alors

$$\bigcup_{n=0}^{\infty} A_n \in \mathfrak{T}.$$

Une tribu est également appelée **σ-algèbre**, la lettre σ signifiant que la propriété *vi)* des algèbres est étendue au cas dénombrable.

Le passage au complémentaire permet de prouver que $\mathfrak{T}$ est encore stable par intersections dénombrables : si $(A_n)_{n\geqslant 0}$ est une suite d'éléments de $\mathfrak{T}$, alors

$$\bigcap_{n=0}^{\infty} A_n \in \mathfrak{T}.$$

Si l'on ne suppose que la stabilité par unions ou intersections dénombrables, l'utilisation répétée de l'ensemble vide ou de Ω permettent de prouver la stabilité par unions ou intersections finies.

Remarque 3.4 (Tribus et atomes) Les axiomes définissant une tribu n'impliquent pas que $\mathfrak{T}$ comporte des singletons. Par exemple, si $\Omega = \mathbb{R}$, on peut construire une tribu de la façon suivante : pour tout $n \in \mathbb{Z}$, on note $I_n = [n\,;\, n+1\,[$, et on définit les éléments de $\mathfrak{T}$ comme les ensembles de la forme $A_K := \bigcup_{n\in K} I_n$, où K est une partie quelconque de $\mathbb{Z}$. Alors $\mathfrak{T}$ est une tribu, mais aucun élément non vide de $\mathfrak{T}$ n'est « plus petit » que les intervalles $[n\,;\, n+1\,[$; notamment, $\mathfrak{T}$ ne comporte pas de singletons $\{x\}$.

Plus généralement, on appelle **atomes** d'une tribu les éléments que la tribu ne peut pas découper, c'est-à-dire des éléments minimaux pour la relation d'inclusion : si A est un atome de $\mathfrak{T}$, alors il n'existe pas d'élément $B \in \mathfrak{T}$ tel que $B \subsetneq A$, à part bien sûr l'ensemble vide. La tribu précédente possède des atomes : les intervalles $[n\,;\, n+1\,[$. Rien n'oblige cependant une tribu à posséder des atomes. ⇙3.8 ⇙3.9

Exemple 3.5 La plus pauvre (petite) tribu de Ω est la tribu $(\varnothing, \Omega)$, qu'on appelle également **tribu grossière** ; la plus riche (grande) tribu est $\mathfrak{P}(\Omega)$, appelée **tribu triviale**, ou **tribu discrète**.

Définition 3.6 Soit $\mathfrak{F}$ une tribu sur Ω. Une partie $\mathfrak{G}$ de $\mathfrak{P}(\Omega)$ est une **sous-tribu** de $\mathfrak{F}$ si c'est une tribu et si $\mathfrak{G} \subset \mathfrak{F}$.

Borne supérieure et borne inférieure

§ 41

Si $(A_n)_{n\geqslant 0}$ est une suite de parties de Ω, alors on définit sa **borne supérieure** et sa **borne inférieure** comme

$$\sup_n A_n := \bigcup_{n=0}^{\infty} A_n \qquad \text{et} \qquad \inf_n A_n := \bigcap_{n=0}^{\infty} A_n.$$

Si $(A_n)_{n\geqslant 0}$ est une suite *croissante* de parties de Ω (c'est-à-dire que $A_n \subset A_{n+1}$ pour tout entier n), sa borne supérieure est également appelée **limite** de $(A_n)_{n\geqslant 0}$ et notée $\lim \uparrow A_n$:

$$\lim_n \uparrow A_n := \bigcup_{n=0}^{\infty} A_n.$$

En notant $A := \lim \uparrow A_n$, on écrit parfois

$$A_n \uparrow A.$$

De même, si $(A_n)_{n\geqslant 0}$ est *décroissante*, on écrit

$$A_n \downarrow A \qquad \text{avec} \qquad A := \lim_n \downarrow A_n := \bigcap_{n=0}^{\infty} A_n.$$

Remarque 3.7 Une suite $(A_n)_{n\geqslant 0}$ de parties quelconques n'admet pas forcément de limite. Pour l'instant, nous avons montré que les suites croissantes ou décroissantes en admettent une ; nous définirons plus tard les notions de limite supérieure et de limite inférieure, ce qui permettra d'étendre la notion de limite à des suites plus générales de parties de Ω (voir § 65 et notamment page 116).

Par définition d'une tribu, on a donc la propriété suivante :

■ **Proposition 3.8** *Une tribu est stable par passage à la borne supérieure et à la borne inférieure : si $(A_n)_{n\geqslant 0}$ est une suite d'éléments d'une tribu $\mathcal{T}$, alors* $\sup A_n$ *et* $\inf A_n$ *sont éléments de $\mathcal{T}$.*

Les règles de de Morgan se généralisent immédiatement aux unions et intersections infinies :

$$(\inf A_n)^c = \sup A_n^c \qquad \text{et} \qquad (\sup A_n)^c = \inf A_n^c.$$

Remarque 3.9 (Extension aux familles non dénombrables) Plus généralement, si $(A_i)_{i\in I}$ est une famille de parties de Ω, indexée par un ensemble I, dénombrable ou non, on définit la **borne inférieure** de cette famille comme l'intersection de tous les A_i :

$$\inf_{i\in I} A_i := \bigcap_{i\in I} A_i$$

et sa **borne supérieure** comme leur union :

$$\sup_{i\in I} A_i := \bigcup_{i\in I} A_i.$$

Ces opérations sont toujours possibles, en théorie des ensembles, notamment parce qu'elles ne font pas appel à un ordre particulier dans I. Cependant, la borne inférieure ou supérieure d'une famille d'éléments d'une tribu *n'est pas nécessairement dans cette tribu* lorsque I est non dénombrable.

§ 42 Autres propriétés des tribus

Le résultat suivant montre que la structure de tribu est stable par intersection (finie, dénombrable ou non dénombrable).

■ **Proposition 3.10 (Intersection de tribus)** *L'intersection d'une famille de tribus de Ω est encore une tribu.*

DÉMONSTRATION : Soit $(\mathcal{T}_i)_{i\in I}$ une famille quelconque de tribus, et notons $\mathcal{T} = \bigcap_{i\in I} \mathcal{T}_i$.

- L'ensemble Ω étant élément de toutes les tribus $\mathcal{T}_i$, il est également élément de $\mathcal{T}$.

- Soit $A \in \mathfrak{T}$. Alors, pour tout $i \in I$, A est élément de $\mathfrak{T}_i$, donc A^c l'est également. Ainsi, A^c est élément de $\mathfrak{T}$.
- Montrons maintenant la stabilité par union dénombrable. Soit $(A_n)_{n\geqslant 0}$ une suite d'éléments de $\mathfrak{T}$, et notons $A = \bigcup_{n=0}^{\infty} A_n$. Pour tout $i \in I$, tous les A_n sont éléments de $\mathfrak{T}_i$, donc leur union A l'est aussi, puisque $\mathfrak{T}_i$ est stable par union dénombrable. Ainsi, A est élément de $\mathfrak{T}$.

Cela suffit pour conclure que $\mathfrak{T}$ est une tribu. ∎

Remarque 3.11 Une tribu est, en général, un ensemble « très gros ». Si $\mathfrak{T}$ comporte une partie A d'éléments disjoints de cardinal n, alors le cardinal de $\mathfrak{T}$ est au moins de 2^n. En particulier, si $\mathfrak{T}$ comporte une famille dénombrable $(A_n)_{n\geqslant 0}$ de parties non vides et deux à deux disjointes, alors elle est de cardinal au moins $2^{\aleph_0} = \mathfrak{c}$, la puissance du continu. On montre par dichotomie qu'une tribu infinie comporte nécessairement une telle suite de parties disjointes non vides[4], ce qui prouve le résultat suivant :

■ **Proposition 3.12** *Une tribu est soit finie, soit non dénombrable.*

Cette propriété est bien connue pour être vérifiée par la tribu triviale $\mathfrak{P}(\Omega)$ de toutes les parties de Ω.

Construire une tribu. Tribu engendrée par un ensemble de parties

§ 43

Soit $\mathfrak{F} = (F_i)_{i\in I}$ un ensemble[5] non vide de parties de Ω. Comme vu dans la proposition 3.10, l'intersection de toutes les tribus contenant $\mathfrak{F}$ est encore une tribu (et elle est non vide puisque $\mathfrak{P}(\Omega)$ est bien une tribu contenant $\mathfrak{F}$). On l'appelle **tribu engendrée par $\mathfrak{F}$**, et on la note $\sigma(\mathfrak{F})$. C'est, par construction, la plus petite tribu contenant $\mathfrak{F}$.

Remarque 3.13 (Pourquoi ce concept ?) Lorsque l'ensemble Ω à probabiliser est dénombrable, la tribu $\mathfrak{P}(\Omega)$ de toutes ses parties est évidemment *la* tribu à utiliser. Dès que Ω est plus gros, par exemple lorsque $\Omega = \mathbb{R}$, il faut se limiter (voir remarque 3.19). Une méthode pratique est de décider tout d'abord que certaines des parties de Ω sont intéressantes et de les baptiser « événements » — par exemple, sur $\mathbb{R}$, on aimerait assigner une probabilité à tous les intervalles, donc on décrète que tous les intervalles sont des événements. Ensuite, afin de pouvoir appliquer les règles « naturelles » de calcul des probabilités, on travaille avec la plus petite tribu contenant tous ces événements.

Une telle tribu est en général difficile à décrire.

- Un cas relativement simple est celui où $\mathfrak{F} = (F_n)_{n\geqslant 0}$ est une partition dénombrable de Ω ; alors $\sigma(\mathfrak{F})$ est l'ensemble des parties de la forme $\bigcup_k F_{i_k}$, où $(i_k)_{k\in I}$ est une partie (finie ou non) de $\mathbb{N}$. ⇙3.3
- Dans la plupart des cas, en revanche, la tribu engendrée par $\mathfrak{F}$ devient énorme et ses éléments ne se laissent pas décrire, en terme d'intersection, unions ou passages au complémentaire dénombrable, en fonction de ceux de $\mathfrak{F}$[6].

Une des tribus les plus importantes que nous aurons à utiliser est la tribu $\mathfrak{B} = \mathfrak{B}(\mathbb{R})$ des **boréliens de $\mathbb{R}$**, qui est la tribu engendrée par les intervalles de $\mathbb{R}$. Une description plus précise de cette tribu est présentée dans l'annexe B ; retenons toutefois que :

---

(4). On choisit un élément A_0 de $\mathfrak{T}$; puisque $\mathfrak{T}$ est infinie, on peut trouver un élément non vide $B \in \mathfrak{T}$ tel que $B \neq A_0$ et $B \neq A_0^c$; alors $B \cap A_0$ et $B \cap A_0^c$ ne sont pas tous deux vides et sont éléments de $\mathfrak{T}$, ce qui montre que A_0 ou $A_0' := A_0^c$ contient un élément de $\mathfrak{T}$ et se partitionne en deux parties A_1 et A_1'. On recommence la même opération : l'un des deux ensembles A_1 ou A_1' admet une partition en deux éléments A_2 et A_2', etc.

(5). On dira parfois : classe de parties. Les mots *classe* et *ensemble* seront, dans tout l'ouvrage, synonymes, étant entendu qu'une classe a pour éléments des parties de Ω.

(6). Il faut alors quitter le domaine rassurant du dénombrable, et utiliser des unions transfinies, ce qui nous amènerait bien loin.

- $\mathfrak{B}$ est la tribu engendrée par les intervalles de la forme $]-\infty\,;\,a]$, pour a parcourant $\mathbb{R}$;
- $\mathfrak{B}$ est la tribu engendrée par tous les intervalles de $\mathbb{R}$;
- $\mathfrak{B}$ est la tribu engendrée par les ouverts de $\mathbb{R}$[7].

Ces quelques caractérisations suffisent à extraire les propriétés voulues de la tribu $\mathfrak{B}$. On retiendra que font partie des boréliens : tous les intervalles, les réunions
⇗3.7 dénombrables d'intervalles, les singletons $\{a\}$, les ouverts, les fermés.

Remarque 3.14 Un des intérêts bien connu des boréliens est qu'ils permettent de définir une intégrale : si A est un borélien, alors $\int \mathbf{1}_A$ a un sens. Si $f : \mathbb{R} \to \mathbb{R}^+$ est une fonction borélienne, c'est-à-dire telle que $f^{-1}(B)$ est un borélien pour tout borélien B, alors $\int f$ a un sens. Une telle construction de l'intégrale sera exposée dans le chapitre 14.

Nous donnons ci-dessous deux propriétés simples des tribus engendrées, conséquences immédiates du fait que $\sigma(\mathfrak{A})$ est l'intersection de toutes les tribus contenant $\mathfrak{A}$.

■ **Proposition 3.15** *Soit $\mathfrak{A}$ une partie de $\mathfrak{P}(\Omega)$.*

i) Si $\mathfrak{A}$ est une tribu, alors $\sigma(\mathfrak{A}) = \mathfrak{A}$.

ii) Si $\mathfrak{A} \subset \mathfrak{G}$ et si $\mathfrak{G}$ est une tribu, alors $\sigma(\mathfrak{A}) \subset \mathfrak{G}$.

DÉMONSTRATION : On suppose que $\mathfrak{A}$ est une tribu. On sait que $\sigma(\mathfrak{A})$ est l'intersection de toutes les tribus contenant $\mathfrak{A}$, elle est donc incluse dans toute tribu contenant $\mathfrak{A}$, et notamment dans $\mathfrak{A}$; ainsi, $\mathfrak{A} \subset \sigma(\mathfrak{A}) \subset \mathfrak{A}$ ce qui permet de conclure que $\sigma(\mathfrak{A}) = \mathfrak{A}$.

On suppose que $\mathfrak{G}$ est une tribu et que $\mathfrak{A}$ est une partie de $\mathfrak{P}(\Omega)$ incluse dans $\mathfrak{G}$. Alors $\sigma(\mathfrak{A})$ est l'intersection d'un certain ensemble de tribus, au rang desquelles on trouve $\mathfrak{G}$, donc $\sigma(\mathfrak{A}) \subset \mathfrak{G}$. ∎

On en déduit la propriété de croissance ci-dessous, élémentaire mais très utile.

■ **Proposition 3.16 (Croissance)** *Si $\mathfrak{F}$ et $\mathfrak{G}$ sont des ensembles de parties de Ω et si $\mathfrak{F} \subset \mathfrak{G}$, alors $\sigma(\mathfrak{F}) \subset \sigma(\mathfrak{G})$.*

DÉMONSTRATION : On sait déjà que $\mathfrak{F} \subset \mathfrak{G} \subset \sigma(\mathfrak{G})$; puisque $\sigma(\mathfrak{G})$ est une tribu, la proposition 3.15 permet alors de conclure que $\sigma(\mathfrak{F}) \subset \sigma(\mathfrak{G})$. ∎
⇗3.5

Dans de nombreux cas, et notamment pour la tribu des boréliens, la tribu engendrée est compliquée, voire impossible à décrire globalement de manière explicite. C'est pourquoi d'autres types de structures sont utiles, comme les « π-systèmes » (voir, en annexe, le paragraphe A72, page 703), qui sont des familles de parties de Ω, stables par intersection finies. Connaître une probabilité sur un π-système $\mathfrak{C}$ suffit à la caractériser sur toute la tribu engendrée par $\mathfrak{C}$.

3.2 ESPACES PROBABILISÉS

§44 **Espaces probabilisables**

▶ On appelle **espace probabilisable** un couple $(\Omega, \mathfrak{T})$ où Ω est un ensemble quelconque, et $\mathfrak{T}$ une tribu sur Ω. (On dit également que $(\Omega, \mathfrak{T})$ est un **espace mesurable**.) L'ensemble Ω est appelé **univers** ou **espace des épreuves**[8], et ses éléments, génériquement notés ω, sont les **épreuves**[9], ou **aléas**, ou **réalisations du**

(7). Et donc également la tribu engendrée par les fermés.

(8). *Sample space* en anglais.

(9). *Sample points.*

hasard, ou encore **résultat d'une expérience**. Les éléments de la tribu $\mathfrak{T}$ sont appelés des **événements observables**[10] ou, plus succinctement, **événements**. On dit que deux événements A et B sont **incompatibles** lorsqu'ils sont disjoints, c'est-à-dire $A \cap B = \varnothing$.

Un événement est souvent décrit par une propriété L, pouvant être ou non vérifiée par les éléments ω ; on assimilera dans toute la suite la propriété L et l'*ensemble de vérité* de L, qui est l'événement

$$\{\omega \in \Omega \,;\, L(\omega)\}.$$

Exemple 3.17 Considérons le cas simple de deux tirages successifs à *pile* ou *face*. L'univers est $\Omega = \{P, F\}^2 = \{PP, PF, FP, FF\}$, ensemble à 4 éléments. La tribu naturelle sur Ω est $\mathfrak{T} = \mathfrak{P}(\Omega)$, à savoir

$$\mathfrak{T} = \{\varnothing, \{PP\}, \{PF\}, \{FP\}, \{FF\}, \{PP, PF\}, \{PP, FP\}, \{PP, FF\}, \{PF, FP\}, \{PF, FF\},$$
$$\{FP, FF\}, \{PP, PF, FP\}, \{PP, PF, FF\}, \{PP, FP, FF\}\{PF, FP, FF\}, \{PF, FP, PF, FF\}\}$$

et possède $2^4 = 16$ éléments. L'événement « les deux tirages ont donné le même résultat » est $A = \{PP, FF\}$.

S'il est pratique de choisir un univers Ω « minimal » pour modéliser une expérience donnée, il est parfois bien plus enrichissant de choisir un univers beaucoup plus gros.

Exemple 3.18 Considérons maintenant la modélisation d'une expérience de pensée : celle où l'on tire *une infinité* de fois à *pile* ou *face*. On pose $\Omega = \{P, F\}^{\mathbb{N}^*}$; chaque point de Ω est une suite

$$\omega = (\omega_1, \omega_2, \ldots, \omega_n, \ldots)$$

où chaque ω_n est à valeurs dans $\{P, F\}$. L'application $\varphi_2 : \omega \mapsto (\omega_1, \omega_2)$ permet de décrire l'expérience de l'exemple précédent (deux tirages successifs seulement). De même, pour tout entier $n \geqslant 1$, l'application $\varphi_n : \omega \mapsto (\omega_1, \omega_2, \ldots, \omega_n)$ permet de décrire une succession de n tirages.

Remarque 3.19 (Pourquoi une tribu — et pas $\mathfrak{P}(\Omega)$ tout entier ? (*))
Si l'on ne se limitait qu'aux espaces finis ou dénombrables, il n'y aurait aucune difficulté à choisir comme tribu sur Ω la tribu $\mathfrak{T} = \mathfrak{P}(\Omega)$ de l'ensemble de toutes les parties de Ω — et aucune difficulté théorique ne surgissant, nous serions débarrassés de tout l'attirail des tribus, fonctions mesurables, etc., qui peuvent rebuter de prime abord. Seulement, dès que l'on passe aux univers non dénombrables, les difficultés surgissent.

Considérons l'espace $\mathbb{R}$, muni de la tribu triviale $\mathfrak{A} = \mathfrak{P}(\mathbb{R})$ de toutes les parties de $\mathbb{R}$. Giuseppe Vitali (1875–1932) a montré qu'il n'existait pas de mesure sur $(\mathbb{R}, \mathfrak{A})$ qui soit invariante par translation. Si l'on quotiente par $\mathbb{Z}$, cela montre qu'il n'existe pas de mesure sur l'espace $\Omega = [0\,;\,1[$ muni de $\mathfrak{T} = \mathfrak{P}(\Omega)$, qui soit de plus invariante par translation. Ainsi, *en choisissant une tribu trop riche, nous nous trouvons dans l'incapacité de définir ne serait-ce qu'une probabilité uniforme sur* $[0\,;\,1[$. Le mathématicien polonais Stanisław Ulam a montré, dans les années 30, que si un univers Ω a la puissance du continu, les seules probabilités sur l'espace $(\Omega, \mathfrak{P}(\Omega))$ sont les probabilités discrètes, c'est-à-dire qu'il existe une partie dénombrable D telle que $\mathbf{P}(\Omega - D) = 0$.

Disons-le tout de suite : *dans la plupart des applications*, toute partie à laquelle on peut raisonnablement penser sera, en pratique, un élément de la tribu $\mathfrak{T}$, et donc un événement ; les parties qui ne sont pas dans $\mathfrak{T}$ requièrent l'utilisation de l'axiome du choix — ce qui est évidemment très éloigné des préoccupations d'un utilisateur occasionnel de la théorie des probabilités[11].

---

(10). La terminologie « événement observable » n'est pas très répandue ; le lecteur aura cependant bien du bénéfice à considérer les éléments de la tribu comme des parties effectivement *observables* par l'instrument de mesure qu'est la (mesure de) probabilité **P**. Ce point de vue prendra toute son importance lorsque seront introduites les notions de probabilité conditionnelle, de temps d'arrêt, ou de martingales (voir la section 16.3 et le chapitre 23).

(11). Robert Solovay (1938–) a montré que l'hypothèse de mesurabilité de toutes les parties de $\mathbb{R}$ était consistante avec l'axiomatique de Zermelo-Fraenkel *sans* l'axiome du choix ; ce dernier est donc nécessaire pour exhiber un ensemble non mesurable (un exemple est donné dans l'annexe

Toutefois, cette complication de prime abord indésirable, est en réalité une vraie aubaine. On peut *choisir* à dessein une tribu restreinte : il est alors aisé de construire des parties de Ω qui ne sont pas dans cette tribu, et ce sans recourir à l'axiome du choix, comme par exemple la *tribu engendrée par une variable aléatoire*, qui est la tribu $\sigma(X)$ engendrée par les événements de la forme $\{X \in B\}$ pour $B \in \mathcal{B}$. Nous verrons que cette tribu qui contient, d'une certaine manière, toute l'information véhiculée par la variable X, au sens où une fonction Y est (théoriquement) calculable à partir de la valeur prises par des variables $X_1, X_2, \dots, X_n$, si et seulement si elle est mesurable par rapport à la tribu $\sigma(X_1, X_2, \dots, X_n)$ engendrée par ces variables[12]. Si Y est le résultat d'une stratégie ne dépendant que de la connaissance de $X_1, X_2, \dots, X_n$, c'est donc une fonction mesurable par rapport à la tribu $\sigma(X_1, X_2, \dots, X_n)$. La notion de tribu est le bon outil pour modéliser l'information (croissante avec le temps) ; elle est et au cœur de nombreux problèmes (voir par exemple les chapitres 23 et 31).

§45 Espaces probabilisés

► Soit $(\Omega, \mathcal{T})$ un espace probabilisable. Une **probabilité** (ou encore : **mesure de probabilité**) sur cet espace est une application $\mathbf{P} : \mathcal{T} \to \mathbb{R}$ vérifiant les propriétés suivantes :

- ⋆ **Axiome 1 :** $\mathbf{P}$ est à valeurs positives ;
- ⋆ **Axiome 2 :** $\mathbf{P}(\Omega) = 1$;
- ⋆ **Axiome 3 :** $\mathbf{P}$ est **σ-additive** : pour toute suite $(A_n)_{n \geqslant 0}$ d'éléments deux à deux disjoints de $\mathcal{T}$,

$$\mathbf{P}\left(\sum_{n=0}^{\infty} A_n \right) = \sum_{n=0}^{\infty} \mathbf{P}(A_n).$$

Le triplet $(\Omega, \mathcal{T}, \mathbf{P})$ est appelé **espace probabilisé**.

Exemple 3.20 À titre d'exercice, le lecteur vérifiera que les structures suivantes sont des espaces probabilisés.

- $\Omega = \{a_1, a_2, \dots, a_n\}$ est un ensemble fini à n éléments, $\mathcal{T}$ est l'ensemble $\mathfrak{P}(\Omega)$ de toutes les parties de Ω, et la probabilité choisie est la probabilité uniforme, caractérisée par $\mathbf{P}(\{a_k\}) = 1/n$ pour $k = 1, \dots, n$, ce qui mène à la formule explicite

$$\mathbf{P}(A) = \frac{\mathrm{Card}(A)}{n}$$

pour toute partie A de Ω.
- Ω est quelconque, $\mathcal{T} = \mathfrak{P}(\Omega)$; on choisit un élément $a \in \Omega$ particulier et on pose

$$\mathbf{P}(A) = \begin{cases} 1 & \text{si } a \in A, \\ 0 & \text{sinon.} \end{cases}$$

La probabilité $\mathbf{P}$ ainsi définie s'appelle la **masse de Dirac**, ou **mesure de Dirac**, centrée en a, et on la note habituellement δ_a.
- Ω est quelconque, $\mathcal{T} = \mathfrak{P}(\Omega)$; on choisit n éléments distincts $a_1, a_2, \dots, a_n$ dans Ω, ainsi que n réels positifs $p_1, p_2, \dots, p_n$ dont la somme vaut $p_1 + p_2 + \cdots + p_n = 1$. On définit enfin

$$\mathbf{P} = p_1 \, \delta_{a_1} + p_2 \, \delta_{a_2} + \cdots + p_n \, \delta_{a_n},$$

prototype d'une **probabilité finie**.
- $\Omega = \mathbb{R}$ est muni de la tribu $\mathcal{T} = \mathcal{B}$ des boréliens. Pour tout événement $A \in \mathcal{B}$, on pose

$$\mathbf{P}(A) = \frac{1}{\sqrt{2\pi}} \int_A \mathrm{e}^{-t^2/2} \, \mathrm{d}t = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathbf{1}_A(t) \, \mathrm{e}^{-t^2/2} \, \mathrm{d}t,$$

où l'on a noté $\mathbf{1}_A$ l'indicatrice de A, définie par $\mathbf{1}_A(t) = 1$ si $t \in A$ et $\mathbf{1}_A(t) = 0$ sinon.

---

qui clôt ce chapitre). Encore mieux : Solovay a prouvé que (sous réserve de consistance de l'axiomatique ZF et de l'axiome du choix), alors ZF muni d'une version légèrement affaiblie de l'axiome du choix, appelée *Axiome du choix dépendant* est consistant avec l'hypothèse de mesurabilité de toute partie de $\mathbb{R}$. L'Axiome du choix dépendant permet de fonder quasiment toute l'analyse habituelle.

(12). C'est le résultat du lemme de Doob-Dynkin, énoncé et démontré page 701.

Propriétés des mesures de probabilités § 46

Les trois axiomes définissant une mesure de probabilité permettent de déduire bien d'autres propriétés ; les suivantes sont élémentaires, et nous n'en démontrons que quelques unes.

■ **Proposition 3.21 (Propriétés élémentaires)** *Si* $\mathbf{P}$ *est une probabilité sur* $(\Omega, \mathfrak{T})$, ◀ *si* A *et* B *sont des éléments de* $\mathfrak{T}$, *alors des axiomes précédents, on déduit les propriétés suivantes :*

i) $\mathbf{P}(\varnothing) = 0$;

ii) $\mathbf{P}$ *est à valeurs dans* $[0\,;\,1]$;

iii) (**croissance**) *: si* $A \subset B$, *alors* $\mathbf{P}(A) \leqslant \mathbf{P}(B)$;

iv) Si A *et* B *sont disjoints, alors* $\mathbf{P}(A + B) = \mathbf{P}(A) + \mathbf{P}(B)$;

v) Si $A \subset B$, *alors* $\mathbf{P}(B - A) = \mathbf{P}(B) - \mathbf{P}(A)$;

vi) $\mathbf{P}(A \setminus B) = \mathbf{P}(A) - \mathbf{P}(A \cap B)$;

vii) $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$;

viii) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

ix) (**additivité**) *: si* $A_1, A_2, \ldots, A_n$ *sont des événements deux à deux disjoints, alors*

$$\mathbf{P}(A_1 + A_2 + \cdots + A_n) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \cdots + \mathbf{P}(A_n).$$

x) (**sous-additivité**) *: si* $A_1, A_2, \ldots, A_n$ *sont des événements quelconques,*

$$\mathbf{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \leqslant \mathbf{P}(A_1) + \mathbf{P}(A_2) + \cdots + \mathbf{P}(A_n).$$

DÉMONSTRATION : *i)* : on pose $A_0 := \Omega$ et $A_n := \varnothing$ pour tout $n \geqslant 1$. La suite $(A_n)_{n \geqslant 0}$ est composée d'éléments deux à deux disjoints, l'axiome 3 permet donc d'écrire

$$1 = \mathbf{P}(\Omega) = \mathbf{P}\Big(\sum_{n=0}^{\infty} A_n\Big) = \mathbf{P}(\Omega) + \sum_{n=1}^{\infty} \mathbf{P}(\varnothing),$$

et donc $\mathbf{P}(\varnothing) = 0$.

iv) et *ix)* : soit n un entier non nul ; soient $A_1, A_2, \ldots, A_n$ des éléments de $\mathfrak{T}$. On pose $A_k := \varnothing$ pour tout entier $k \geqslant n + 1$; la propriété de σ-additivité et la propriété *i)* permettent de conclure.

iii) et *v)* : on écrit $B = A + (B - A)$ donc $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B - A)$ grâce à la propriété d'additivité.

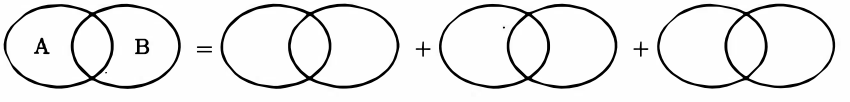
ii) et *vii)* : pour tout élément A de $\mathfrak{T}$, $\mathbf{P}(A) + \mathbf{P}(A^c) = 1$.

vi) : découle de la formule précédente et de la relation $A \setminus B = A - (A \cap B)$.

viii) : on écrit

$$A \cup B = [A - (A \cap B)] + [B - (A \cap B)] + A \cap B$$

et on applique les propriétés *iv)* et *v)*.



x) : se démontre par récurrence sur n en utilisant la propriété *viii)*. ▮

Les propriétés énoncées ne faisaient intervenir que des opérations *finies*. Or, l'intérêt de la notion de tribu et de σ-additivité est justement de permettre des opérations dénombrables ; on en déduit les trois propriétés utiles suivantes, qui sont en fait équivalentes à la propriété de σ-additivité :

▶ ■ **Théorème 3.22** *Soit* $\mathbf{P} : \mathfrak{T} \to \mathbb{R}^+$ *une application vérifiant les axiomes* 1 *et* 2, *et qui est de plus* additive. *Alors, l'axiome* 3 *(*σ*-additivité) est équivalent à chacune des propriétés suivantes :*

★ **Axiome 3′ :** *(*continuité monotone croissante*) pour toute suite croissante* $(\mathrm{B}_n)_{n\geqslant 0}$ *d'éléments de* $\mathfrak{T}$,

$$\mathbf{P}\Big(\bigcup_{n=0}^{\infty} \mathrm{B}_n\Big) = \lim_{n\to\infty} \mathbf{P}(\mathrm{B}_n)$$

ce que l'on écrit aussi

$$\mathbf{P}\big(\lim_{n\to\infty} \uparrow \mathrm{B}_n\big) = \lim_{n\to\infty} \mathbf{P}(\mathrm{B}_n).$$

★ **Axiome 3″ :** *(*continuité monotone décroissante*) pour toute suite* $(\mathrm{C}_n)_{n\geqslant 0}$ *décroissante d'éléments de* $\mathfrak{T}$,

$$\mathbf{P}\Big(\bigcap_{n=0}^{\infty} \mathrm{C}_n\Big) = \lim_{n\to\infty} \mathbf{P}(\mathrm{C}_n).$$

ce que l'on écrit aussi

$$\mathbf{P}\big(\lim_{n\to\infty} \downarrow \mathrm{C}_n\big) = \lim_{n\to\infty} \mathbf{P}(\mathrm{C}_n).$$

★ **Axiome 3‴ :** *(*continuité monotone en $\varnothing$*) pour toute suite décroissante* $(\mathrm{D}_n)_{n\geqslant 0}$ *d'événements telle que* $\bigcap_n \mathrm{D}_n = \varnothing$, *on a* $\lim_{n\to\infty} \mathbf{P}(\mathrm{D}_n) = 0$.

DÉMONSTRATION : Supposons la σ-additivité (en plus des autres propriétés) et montrons l'axiome 3′. Soit $(\mathrm{B}_n)_{n\geqslant 0}$ une suite croissante d'éléments de $\mathfrak{T}$; on pose $\mathrm{B}_{-1} = \varnothing$ et $\mathrm{A}_n = \mathrm{B}_n \smallsetminus \mathrm{B}_{n-1}$ pour tout n entier. Ainsi, les A_n sont deux à deux disjoints et

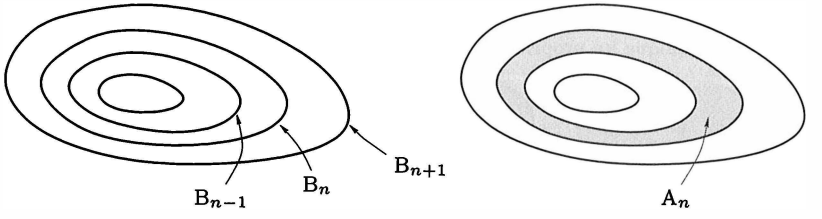
$$\mathrm{B}_n = \mathrm{A}_0 + \mathrm{A}_1 + \mathrm{A}_2 + \cdots + \mathrm{A}_n$$

pour tout n, donc notamment $\mathbf{P}(\mathrm{B}_n) = \mathbf{P}(\mathrm{A}_0)+\mathbf{P}(\mathrm{A}_1)+\cdots+\mathbf{P}(\mathrm{A}_n)$. Par double inclusion, on a

$$\bigcup_{n=0}^{\infty} \mathrm{B}_n = \sum_{n=0}^{\infty} \mathrm{A}_n$$

donc

$$\mathbf{P}\Big(\bigcup_{n=0}^{\infty} \mathrm{B}_n\Big) = \mathbf{P}\Big(\sum_{n=0}^{\infty} \mathrm{A}_n\Big) = \sum_{n=0}^{\infty} \mathbf{P}(\mathrm{A}_n) = \lim_{\mathrm{N}\to\infty} \sum_{n=0}^{\mathrm{N}} \mathbf{P}(\mathrm{A}_n) = \lim_{\mathrm{N}\to\infty} \mathbf{P}(\mathrm{B}_\mathrm{N}).$$



Réciproquement, de l'axiome 3′, on déduit la σ-additivité par un raisonnement du même style, en posant $\mathrm{B}_n = \mathrm{A}_0 \cup \mathrm{A}_1 \cup \cdots \cup \mathrm{A}_n$ pour tout $n \in \mathbb{N}$.

L'axiome 3″ se traite de manière analogue en posant $\mathrm{A}_n := \mathrm{C}_n - \mathrm{C}_{n+1}$, ce qui revient à $\mathrm{C}_n = \bigcup_{k=n}^{\infty} \mathrm{A}_k$. On peut également passer au complémentaire dans l'axiome 3′.

Enfin, l'équivalence entre la σ-additivité et l'axiome 3‴, fait l'objet de l'exercice 3.12. ▮

On pourra retenir ces résultats en utilisant la notation « $C_n \downarrow C$ » pour indiquer que la suite $(C_n)_{n\geqslant 0}$ est décroissante et de limite C (de même $B_n \uparrow B$ pour une suite croissante de limite B) :

Axiome 3	si les A_n sont disjoints	alors	$\mathbf{P}(\sum_n A_n) = \sum_n \mathbf{P}(A_n)$
Axiome 3′	si $B_n \uparrow B$,	alors	$\mathbf{P}(B_n) \uparrow \mathbf{P}(B)$
Axiome 3″	si $C_n \downarrow C$,	alors	$\mathbf{P}(C_n) \downarrow \mathbf{P}(C)$
Axiome 3‴	si $D_n \downarrow \varnothing$,	alors	$\mathbf{P}(D_n) \downarrow 0$

Enfin, la propriété de sous-additivité se généralise à une union dénombrable (propriété de **σ-sous-additivité**, démontrée en exercice) : ⇗3.10

■ **Théorème 3.23 (Inégalité de Boole / propriété de σ-sous-additivité)**
Si $(A_n)_{n\geqslant 1}$ *est une suite d'événements quelconques,* ◀

$$\mathbf{P}\left(\bigcup_{n=0}^{\infty} A_n\right) \leqslant \sum_{n=0}^{\infty} \mathbf{P}(A_n),$$

le deuxième membre de l'inégalité pouvant être fini ou égal à $+\infty$.

Événements négligeables §47

On appelle **événement négligeable** tout élément N de $\mathfrak{T}$ de probabilité nulle, c'est-à-dire tel que $\mathbf{P}(N) = 0$. À l'opposé, un événement **presque sûr** est un élément Ω' de $\mathfrak{T}$ tel que $\mathbf{P}(\Omega') = 1$. ◀

Un événement est donc négligeable si et seulement si il est complémentaire d'un événement presque sûr.

■ **Proposition 3.24** *Si* N *est un événement négligeable, et si* A *est un événement tel que* $A \subset N$, *alors* A *est négligeable. Si des événements* $N_1, N_2, \ldots$, *en quantité dénombrable, sont négligeables, alors* $N' := \bigcup_{k=1}^{\infty} N_k$ *est négligeable.*

DÉMONSTRATION : Par croissance, $0 \leqslant \mathbf{P}(A) \leqslant \mathbf{P}(N) = 0$, donc A est négligeable. Par propriété des tribus, N′ est un événement et, par σ-sous-additivité, $\mathbf{P}(N') \leqslant \sum_k \mathbf{P}(N_k) = 0$. ∎

Exemple 3.25 Nous avons évoqué l'espace $\Omega = \{0,1\}^{\mathbb{N}}$, servant à modéliser des suites infinies de tirages à *pile* ou *face* (chapitre 1, section 1.3). Supposons la pièce équilibrée. Le résultat d'une expérience est une suite $\omega = (x_n)_{n\geqslant 0}$ à valeurs dans $\{0,1\}$ (où, arbitrairement, 1 représente *pile* et 0 représente *face*). La mesure de probabilité choisie se construit ainsi : si l'on se fixe des résultats $a_{k_1}, \ldots, a_{k_n}$, où $0 \leqslant k_1 < k_2 < \cdots < k_n$, l'événement

$$C = \{\omega = (x_n)_{n\geqslant 0} \in \Omega \,;\, x_{k_1} = a_{k_1}, \ldots, x_{k_n} = a_{k_n}\}$$

(appelé *cylindre*) a pour probabilité $1/2^n$; on admettra provisoirement que cette règle de calcul permet de définir une unique mesure de probabilité sur la tribu $\mathfrak{T}$ engendrée par les cylindres. Si ω est un tirage particulier, alors $\{\omega\}$ est négligeable. Sont également négligeables tous les ensembles de successions de tirages dont on fixe une infinité de résultats, comme par exemple l'ensemble des successions de tirages pour lesquels tous les résultats pairs sont égaux à 1.

Espaces probabilisés complets (*) §48

On peut généraliser la notion de négligeabilité aux parties de Ω qui ne sont pas forcément des événements. On dira ainsi qu'une partie A de Ω est **négligeable** si elle est contenue dans un événement négligeable, c'est-à-dire s'il existe $B \in \mathfrak{T}$ tel que $A \subset B$ et $\mathbf{P}(B) = 0$.

Avec cette définition, le lecteur pourra vérifier qu'une union dénombrable de parties négligeables est toujours négligeable.

L'ennui de cette définition, c'est qu'une partie négligeable n'est pas forcément de probabilité nulle, pour la simple raison que sa « probabilité » n'est pas forcément définie. On étend alors la probabilité **P** (définie sur la tribu $\mathfrak{T}$) afin qu'elle englobe les ensembles négligeables (et leur affecte évidemment une probabilité nulle). Pour cela, notons $\mathfrak{N}$ la classe de tous les ensembles négligeables et définissons

$$\widehat{\mathfrak{T}} = \{A \cup N\,;\ A \in \mathfrak{T},\ N \in \mathfrak{N}\}.$$

On peut alors montrer que c'est une tribu[13]. Si $\widehat{\mathfrak{T}} = \mathfrak{T}$, on dit que $\mathfrak{T}$ est **complète** (relativement à **P**). Sinon, c'est-à-dire si $\widehat{\mathfrak{T}}$ est effectivement plus grande que $\mathfrak{T}$, on l'appelle la **complétée de** $\mathfrak{T}$.

On peut alors étendre, de manière unique, l'application **P** à la tribu complétée, en posant

$$\mathbf{P}(A \cup N) := \mathbf{P}(A)$$

pour tout $A \in \mathfrak{T}$ et tout $N \in \mathfrak{N}$. On doit bien entendu vérifier que cette définition ne dépend pas de la décomposition d'un élément de $\widehat{\mathfrak{T}}$ en l'union $A \cup N$. L'espace $(\Omega, \widehat{\mathfrak{T}}, \mathbf{P})$ est appelé **espace probabilisé complet.**

Remarque 3.26 (Tribu de Lebesgue) L'espace $[0\,;1]$, muni de la tribu des boréliens et de la probabilité égale à la mesure de Lebesgue, n'est pas complet ; on complète cet espace en fabriquant la **tribu de Lebesgue**, bien plus « grosse » que la tribu des boréliens.

Pourquoi les probabilistes n'utilisent-ils pas systématiquement la tribu de Lebesgue (comme les analystes), mais plutôt celle des boréliens ? D'une part, la tribu $\mathfrak{B}([0\,;1])$, tout comme la tribu $\mathfrak{B}$ des boréliens de $\mathbb{R}$, possède une propriété, celle d'être **séparable**, c'est-à-dire d'être engendrée par une famille dénombrable (par exemple celle des intervalles $[a\,;b]$, où a et b parcourent $\mathbb{Q}$) ; la tribu de Lebesgue ne possède pas cette propriété.

D'autre part, la notion d'espace probabilisé complet n'est pas une « bonne » notion probabiliste. Une même expérience peut être modélisée dans un espace probabilité $(\Omega, \mathfrak{T}, \mathbf{P})$, puis modélisée de nouveau sur un espace « plus grand » $(\Omega', \mathfrak{T}', \mathbf{P}')$. Le premier espace peut être complet sans que le second le soit.

§ 49 Systèmes complets, quasi-complets d'événements

Lorsque l'on étudie une variable aléatoire à valeurs entières $X : \Omega \to \mathbb{Z}$, on obtient une « partition » naturelle de Ω, sous la forme

$$\Omega = \bigcup_{n \in \mathbb{Z}} X^{-1}(\{n\}).$$

Le lecteur se convaincra que cette décomposition est bien une partition[14] : l'union est disjointe et est égale à Ω. La notion de découpage fini ou dénombrable de Ω est tellement importante en probabilités qu'on lui donne un nom spécifique, celui de *système complet d'événements.*

► *Définition 3.27* Une famille $(A_i)_{i \in I}$, finie ou dénombrable, de parties de Ω est appelée **système complet d'événements** si Ω est la réunion disjointe des A_i :

$$\Omega = \sum_{i \in I} A_i.$$

Interprétation : si $(A_i)_{i \in I}$ est un système complet d'événements, alors pour toute réalisation ω du hasard, *un et un seul* des événements A_i est réalisé.

► Une famille $(A_i)_{i \in I}$, finie ou dénombrable, est appelé **système quasi-complet d'événements** si ses éléments sont incompatibles, ($A_i \cap A_j = \varnothing$ si $i \neq j$) et que leur réunion est presque certaine :

$$\mathbf{P}\left(\sum_{i \in I} A_i\right) = 1.$$

(13). C'est même précisément la tribu engendrée par $\mathfrak{T} \cup \mathfrak{N}$.

(14). Pas tout à fait : traditionnellement, une partition n'est formée que de parties *non vides.* Ici, les parties vides sont acceptées, puisque X n'est pas supposée surjective.

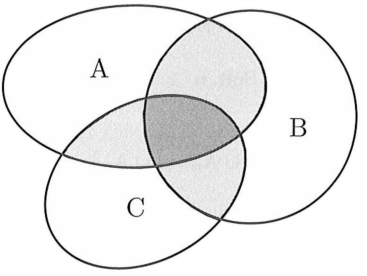


FIGURE 3.2 — *La formule de Poincaré pour* $n = 3$. *En gris foncé,* $A \cap B \cap C$. *En gris (clair et foncé),* $A \cap B$, $A \cap C$ *et* $B \cap C$.

Si l'on note $N = \Omega \setminus \sum_{i \in I} A_i$, alors N est un événement négligeable ; de plus, la famille formée des A_i et de N est un système complet d'événements.

Interprétation : si $(A_i)_{i \in I}$ est un système quasi-complet d'événements, alors avec une probabilité égale à 1, un et un seul des événements A_i est réalisé.

■ **Proposition 3.28** *Si* $(A_i)_{i \in I}$ *est un système quasi-complet d'événements, et si* B *est un événement, alors*

$$\mathbf{P}(B) = \sum_{i \in I} \mathbf{P}(B \cap A_i).$$

DÉMONSTRATION : En effet, B est l'union disjointe des $B \cap A_i$ et de $B \cap N$, ce dernier événement étant de probabilité nulle. ▮

Formule du crible (ou : de Poincaré) §50

La relation

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$$

permet de calculer la probabilité de l'union de deux événements non incompatibles. Se généralise-t-elle à davantage d'événements ? Essayons avec trois événements A, B, C ; si l'on calcule

$$\mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) - \mathbf{P}(A \cap B) - \mathbf{P}(B \cap C) - \mathbf{P}(C \cap A),$$

on voit sur la figure 3.2 que ce n'est pas le résultat juste, puisque l'on a compté trois fois la partie $A \cap B \cap C$, puis on l'a retirée trois fois ; la formule exacte est

$$\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) - \mathbf{P}(A \cap B) - \mathbf{P}(B \cap C) - \mathbf{P}(C \cap A) + \mathbf{P}(A \cap B \cap C).$$

Sa généralisation est connue sous le nom de **formule de Poincaré**, ou **du crible**.

■ **Théorème 3.29 (Formule de Poincaré)** *Pour tout entier n non nul et pour tous événements* $A_1, \ldots, A_n$ *(pas nécessairement disjoints),* ◀

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \mathbf{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}).$$

En passant au complémentaire, on obtient également

$$\mathbf{P}\left(\bigcap_{i=1}^{n} A_i^c\right) = \sum_{k=0}^{n} (-1)^{k} \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \mathbf{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}),$$

où le terme $k = 0$ *donne, par convention, la valeur* 1.

DÉMONSTRATION : Par récurrence sur n. Notons $\mathscr{H}_n$ la propriété au rang n (valable pour tous événements $A_1, A_2, \dots, A_n$).

Elle est vraie pour $n = 2$. Soit $n \geqslant 3$, et supposons la propriété vraie jusqu'à un rang $n-1$. Alors

$$\begin{aligned}
\mathbf{P}\Big(\bigcup_{i=1}^{n} A_i\Big) &= \mathbf{P}\big((A_1 \cup A_2 \cup \cdots \cup A_{n-1}) \cup A_n\big) \\
&= \mathbf{P}(A_1 \cup A_2 \cup \cdots \cup A_{n-1}) + \mathbf{P}(A_n) - \mathbf{P}\big((A_1 \cup A_2 \cup \cdots \cup A_{n-1}) \cap A_n\big) \\
&= \mathbf{P}(A_1 \cup A_2 \cup \cdots \cup A_{n-1}) + \mathbf{P}(A_n) - \mathbf{P}\big((A_1 \cap A_n) \cup \cdots \cup (A_{n-1} \cap A_n)\big) \\
&= \sum_{k=1}^{n-1} (-1)^{k+1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n-1} \mathbf{P}(A_{i_1} \cap \cdots \cap A_{i_k}) \\
&\quad - \sum_{k=1}^{n-1} (-1)^{k+1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n-1} \mathbf{P}(A_{i_1} \cap \cdots A_{i_k} \cap A_n) + \mathbf{P}(A_n)
\end{aligned}$$

Il ne reste plus qu'à constater que toute suite strictement croissante $(i_1, i_2, \dots, i_k)$ de $[1\,;\,n]$ s'écrit soit comme une suite croissante de $[1\,;\,n-1]$, soit comme une suite croissante de $[1\,;\,n-1]$ suivie de $i_k = n$, soit enfin comme le terme seul n (lorsque $k = 1$), ce qui montre la formule de $\mathscr{H}_n$ et achève la récurrence[15]. ∎

3.3 PROBABILITÉ SUR $\mathbb{R}$

Nous donnons ici quelques exemples de probabilités sur l'espace probabilisable $\big(\mathbb{R}, \mathfrak{B}(\mathbb{R})\big)$. Ce paragraphe est particulièrement important : une variable aléatoire X transporte une probabilité $\mathbf{P}$ (définie sur un espace probabilisable $(\Omega, \mathfrak{T})$) sur l'espace $\mathbb{R}$, et définit ainsi une nouvelle probabilité $\mathbf{P}_X$ sur l'espace probabilisable $\big(\mathbb{R}, \mathfrak{B}(\mathbb{R})\big)$.

§51 Fonction de répartition d'une probabilité

Pour étudier une probabilité $\mathbf{P}$ sur $\mathbb{R}$, l'outil fondamental est sa **fonction de répartition**, application $F : \mathbb{R} \to [0\,;\,1]$ définie par

$$F(x) := \mathbf{P}\big(]-\infty\,;\,x]\big).$$

---

(15). On peut donner une autre présentation de cette récurrence sous la forme suivante. On considère les variables aléatoires $\mathbf{1}_{A_i}$ indicatrices des événements A_i, dont l'espérance vaut $\mathbf{E}(\mathbf{1}_{A_i}) = \mathbf{P}(A_i)$. Ces variables aléatoires vérifient également

$$\mathbf{1}_{A_i \cap A_j} = \mathbf{1}_{A_i} \cdot \mathbf{1}_{A_j} \qquad \text{et} \qquad \mathbf{1}_{A_i^c} = 1 - \mathbf{1}_{A_i}.$$

On a donc, en utilisant conjointement ces deux propriétés :

$$\mathbf{P}(A_1 \cup \cdots \cup A_n) = \mathbf{E}\left[1 - \prod_{i=1}^{n} (1 - \mathbf{1}_{A_i})\right].$$

Il ne reste plus qu'à développer le produit intervenant dans le membre de droite ; de la même façon que l'on montre (par récurrence par exemple) que

$$\prod_{i=1}^{n} (1 - x_i) = 1 + \sum_{k=1}^{n} (-1)^k \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} x_{i_1} x_{i_2} \dots x_{i_k},$$

on trouve

$$\prod_{i=1}^{n} (\mathbf{1}_\Omega - \mathbf{1}_{A_i}) = \mathbf{1}_\Omega + \sum_{k=1}^{n} (-1)^k \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \mathbf{1}_{A_{i_1}} \cdot \mathbf{1}_{A_{i_2}} \dots \mathbf{1}_{A_{i_k}},$$

On évalue ensuite l'espérance pour trouver la formule souhaitée.

■ **Proposition 3.30** *Si* F *est la fonction de répartition d'une probabilité* **P**, *alors :* ◄

i) F *est croissante ;*

ii) F *est continue à droite ;*

iii) $\mathrm{F}(-\infty) := \lim_{x\to-\infty} \mathrm{F}(x) = 0$;

iv) $\mathrm{F}(+\infty) := \lim_{x\to+\infty} \mathrm{F}(x) = 1$.

Réciproquement, si une fonction F *vérifie ces quatre propriétés, alors c'est la fonction de répartition d'une unique probabilité* **P** *sur* $\mathbb{R}$ *muni de la tribu des boréliens.*

↫3.14

Démonstration :

i) Cette propriété vient de la croissance de la mesure de probabilité (si $A \subset B$, alors $\mathbf{P}(A) \leqslant \mathbf{P}(B)$) et du fait que, si $x \leqslant y$, alors $]-\infty ; x] \subset]-\infty ; y]$.

ii) Pour montrer que F est continue à droite en un point x, il suffit de prouver que, pour toute suite $(x_n)_{n\geqslant 0}$ décroissante et de limite x, on a $\mathrm{F}(x_n) \to \mathrm{F}(x)$. Or, la suite de parties de $\mathbb{R}$, de terme général $\mathrm{I}_n =]-\infty ; x_n]$, est décroissante pour l'inclusion ; d'après l'axiome 3″ (page 86), on a donc $\lim_{n\to\infty} \mathbf{P}(\mathrm{I}_n) = \mathbf{P}(\lim_{n\to\infty} \mathrm{I}_n)$, avec

$$\lim_{n\to\infty} \mathrm{I}_n = \bigcap_{n=0}^{\infty}]-\infty ; x_n] =]-\infty ; x],$$

ce qui permet de conclure.

iii) Toujours d'après l'axiome 3″, pour toute suite décroissante $(x_n)_{n\geqslant 0}$ de limite $-\infty$:

$$\lim_{n\to\infty} \mathrm{F}(x_n) = \lim_{n\to\infty} \mathbf{P}(]-\infty ; x_n]) = \mathbf{P}\left(\bigcap_{n=0}^{\infty}]-\infty ; x_n]\right) = \mathbf{P}(\varnothing) = 0.$$

iv) D'après l'axiome 3′, pour toute pour toute suite croissante $(x_n)_{n\geqslant 0}$ de limite $+\infty$:

$$\lim_{n\to\infty} \mathrm{F}(x_n) = \lim_{n\to\infty} \mathbf{P}(]-\infty ; x_n]) = \mathbf{P}\left(\bigcup_{n=0}^{\infty}]-\infty ; x_n]\right) = \mathbf{P}(\mathbb{R}) = 1.$$

La dernière propriété est plus délicate à montrer. L'idée est de commencer par assigner une probabilité à tout intervalle semi-ouvert $]a ; b]$ par $\mathbf{P}_0(]a ; b]) := \mathrm{F}(b) - \mathrm{F}(a)$. On étend cette application à l'algèbre $\mathfrak{A}$ des unions disjointes finies d'intervalles semi-ouverts, en posant ↫3.2

$$\mathbf{P}_0\left(\sum_{k=1}^{n}]a_k ; b_k]\right) := \sum_{k=1}^{n} \big(\mathrm{F}(b_k) - \mathrm{F}(a_k)\big).$$

On peut montrer que cette application $\mathbf{P}_0$, définie sur l'algèbre $\mathfrak{A}$ est σ-additive — c'est assez technique, et le lecteur intéressé pourra trouver la démonstration dans de nombreux ouvrages, par exemple Shiryaev [91], Jacod et Protter [53] ou Billingsley [9]. Pour conclure, on invoque un théorème issu de la théorie de la mesure, le *théorème de Carathéodory*, qui affirme que sous ces conditions, il existe une est une seule manière d'étendre l'application $\mathbf{P}_0$ sur la tribu $\sigma(\mathfrak{A})$ engendrée par $\mathfrak{A}$, et qui n'est autre que $\mathfrak{B}(\mathbb{R})$; ce théorème, rapidement présenté à la fin du chapitre, est détaillé dans l'annexe B. On notera que **P** n'est, *a priori*, pas définie sur la tribu $\mathfrak{P}(\mathbb{R})$ tout entière (sauf si **P** est une probabilité finie ou discrète). ▮

On retiendra que : *la fonction de répartition d'une probabilité* **P** *concentre toute l'information contenue dans* $\mathbf{P}$[16] :

■ **Proposition 3.31** *Une probabilité sur* $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ *est entièrement caractérisée par sa fonction de répartition : si* **P** *et* **Q** *sont des probabilités sur* $\mathbb{R}$, *et si leurs fonctions de répartition sont égales, alors* $\mathbf{P} = \mathbf{Q}$.

---

(16). Ou encore : il y a une bijection entre l'ensemble des mesures de probabilités sur $(\mathbb{R}, \mathfrak{B})$ et l'ensemble des fonctions croissantes, continues à droite, de limites 0 en $-\infty$ et 1 et $+\infty$. Cette correspondance a surtout un intérêt théorique ; il n'y a pas d'« algorithme » permettant de décrire la probabilité d'un événement quelconque en fonction de F, mis à part le cas très simple où cet événement est réunion dénombrable d'intervalles.

Comme toute fonction croissante, la fonction de répartition F d'une probabilité admet en tout point a une limite à gauche, notée $F(a^-)$; de plus, le **saut de discontinuité** en un point a, défini par

$$\Delta F(a) = F(a) - F(a^-)$$

est égal à la probabilité du singleton $\{a\}$. En effet, on peut écrire que

$$\{a\} = \bigcap_{n=1}^{\infty}]\, a - \tfrac{1}{n}\, ;\, a\,] = \lim_n \downarrow]\, a - \tfrac{1}{n}\, ;\, a\,]$$

et, d'après l'axiome 3″,

$$\mathbf{P}(\{a\}) = \lim_{n\to\infty} \mathbf{P}\,(]\, a - \tfrac{1}{n}\, ;\, a\,]) = \lim_{n\to\infty} \left(F(a) - F\left(a - \tfrac{1}{n}\right)\right) = F(a) - F(a^-).$$

On montre de même les formules suivantes (liste non exhaustive !) :

$$\mathbf{P}([\, a\, ;\, b\,]) = F(b) - F(a^-) \qquad \mathbf{P}([\, a\, ;\, b\,[) = F(b^-) - F(a^-)$$
$$\mathbf{P}(]\, a\, ;\, b\,]) = F(b) - F(a) \qquad \mathbf{P}(]\, a\, ;\, b\,[) = F(b^-) - F(a)$$
$$\mathbf{P}(]-\infty\, ;\, b\,[) = F(b^-) \qquad \mathbf{P}([\, a\, ;\, +\infty\,[) = 1 - F(a^-)$$

§ 52 **La mesure de Lebesgue sur [0 ; 1]**

On munit l'espace $\Omega = [\,0\,;\,1\,]$ de la tribu des boréliens, c'est-à-dire la tribu engendrée par les intervalles (ou les ouverts, etc) de Ω. On note traditionnellement μ la mesure définie sur cet espace par la fonction de répartition $F(x) = x$, et on l'appelle **mesure de Lebesgue sur [0 ; 1]**. Elle vérifie donc

$$\mu([\, a\, ;\, b\,]) = \mu([\, a\, ;\, b\,[) = \mu(]\, a\, ;\, b\,]) = \mu(]\, a\, ;\, b\,[) = b - a.$$

D'après le théorème 3.30, c'est une probabilité sur $[\,0\,;\,1\,]$ muni de la tribu des boréliens ; sauf précision supplémentaire, c'est toujours cette probabilité que l'on utilisera sur cet intervalle. Attirons l'attention sur le fait que c'est la probabilité la plus naturelle qui se puisse imaginer sur $[\,0\,;\,1\,]$, celle qui « ne privilégie aucun point spécifique » ; on l'appelle également la **probabilité uniforme sur [0 ; 1]**.

Dans les paragraphes suivants, nous introduisons trois cas fondamentaux de probabilités.

§ 53 **Probabilités finies**

Soient n un entier naturel non nul, $a_1, a_2, \ldots, a_n$ des réels distincts et $p_1, p_2, \ldots, p_n$ des réels positifs tels que $p_1 + p_2 + \cdots + p_n = 1$. On définit, sur l'espace probabilisable $(\mathbb{R}, \mathfrak{P}(\mathbb{R}))$, la probabilité $\mathbf{P}$ en posant, pour toute partie A de $\mathbb{R}$:

$$\mathbf{P}(A) = \sum_{k\,;\, a_k \in A} p_k,$$

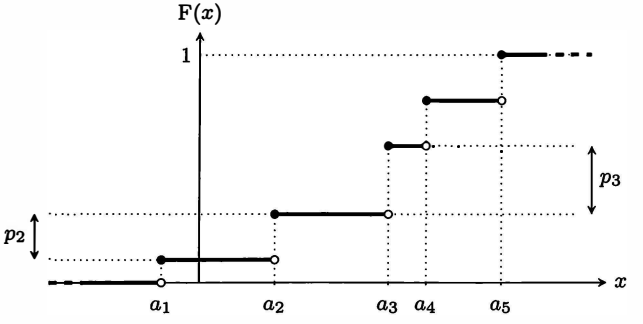
où la somme est définie sur l'ensemble des indices k tels que $a_k \in A$. Cette probabilité est notée

$$\mathbf{P} = \sum_{k=1}^{n} p_k\, \delta_{a_k},$$

où δ_{a_k} est la mesure de Dirac centrée en a_k :

$$\delta_{a_k}(A) = \begin{cases} 1 & \text{si } a_k \in A, \\ 0 & \text{sinon.} \end{cases}$$

La fonction de répartition de **P** a l'allure générique suivante : une fonction en escaliers, présentant un nombre fini de discontinuités aux points $a_1, a_2, \ldots, a_n$.



Probabilités discrètes § 54

Soit $\{a_n \ ; \ n \in \mathbb{N}\}$ une famille de réels distincts, et $\{p_n \ ; \ n \in \mathbb{N}\}$ une famille de réels positifs tels que la série $\sum p_n$ converge et a pour somme 1. On définit, sur l'espace probabilisable $\big(\mathbb{R}, \mathfrak{P}(\mathbb{R})\big)$, la probabilité **P** en posant, pour toute partie A de $\mathbb{R}$:

$$\mathbf{P}(\mathrm{A}) = \sum_{k \ ; \ a_k \in \mathrm{A}} p_k,$$

où la somme est définie sur l'ensemble des indices k tels que $a_k \in \mathrm{A}$ (cette somme est bien définie). Cela revient à poser

$$\mathbf{P} = \sum_{k=0}^{\infty} p_k \, \delta_{a_k}.$$

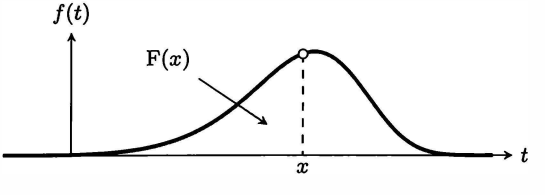
Là encore, la fonction de répartition ne prend qu'un ensemble dénombrable de valeurs, et effectue en chaque a_k un « saut » de discontinuité

$$\mathrm{F}(a_k) - \mathrm{F}(a_k^-) = \mathbf{P}\big(\{a_k\}\big) = p_k.$$

Probabilités à densité § 55

À l'opposé d'une probabilité finie ou discrète, on trouve la famille des probabilités dites **absolument continues**, ou encore **probabilités à densité** : ce sont les probabilités **P** telles qu'il existe une fonction $f : \mathbb{R} \to \mathbb{R}^+$ telles que

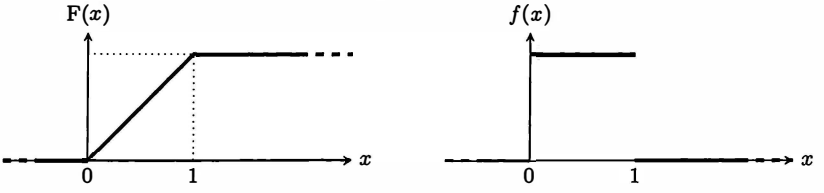
$$\forall x \in \mathbb{R} \qquad \mathrm{F}(x) = \mathbf{P}\big(\,]-\infty \, ; \, x\,]\big) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t.$$



La fonction f, appelée **densité** de la probabilité **P**, est à valeurs positives et est égale, en presque tous les réels, à la dérivée de F. Elle vérifie de plus

$$\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = \mathbf{P}(\mathbb{R}) = 1.$$

Exemple 3.32 Si l'on pose $f(x) = 1$ pour tout $x \in [0;1]$ et $f(x) = 0$ sinon, la fonction F correspondante vérifie $F(x) = x$ si $0 \leqslant x \leqslant 1$, $F(x) = 0$ si $x \leqslant 0$ et $F(x) = 1$ si $x \geqslant 1$. La probabilité définie par cette fonction de répartition est la *probabilité uniforme sur* $[0;1]$, ou *mesure de Lebesgue sur* $[0;1]$ déjà vue au paragraphe §52.



3.4 ANNEXE : π-SYSTÈMES ET THÉORÈME DE CARATHÉODORY (*)

§56 **π-systèmes**

Une classe $\mathfrak{C}$ de parties de Ω est appelée **π-système** si elle est stable par intersections finies :

$$\forall A, B \in \mathfrak{C} \qquad A \cap B \in \mathfrak{C}.$$

Parmi les classes de parties de Ω engendrant une tribu $\mathfrak{T}$ donnée, les π-systèmes sont, d'une certaine manière, les classes minimales caractérisant une mesure de probabilité :

■ **Théorème 3.33** *Une probabilité sur* $(\Omega, \mathfrak{T})$ *est entièrement caractérisée par ses valeurs sur un π-système engendrant* $\mathfrak{T}$.

En d'autres termes, si **P** *et* **Q** *sont deux probabilités sur* $(\Omega, \mathfrak{T})$, *qui coïncident sur un π-système* $\mathfrak{C}$ *tel que* $\mathfrak{T} = \sigma(\mathfrak{C})$, *alors elles sont nécessairement égales sur* $\mathfrak{T}$ *tout entière.*

La démonstration de ce théorème est présentée en annexe (théorème B.18 page 706).

§57 **Le théorème de Carathéodory**

Dans le théorème précédent, il est supposé que **P** et **Q** sont déjà définies sur $\mathfrak{T}$. Le théorème de Carathéodory, bien connu en théorie de la mesure, permet d'étendre à une tribu — de manière unique — une probabilité définie sur une algèbre. C'est donc avant tout un théorème d'*existence*, dont la partie « unicité » découle du théorème 3.33 (toute algèbre est un π-système).

■ **Théorème 3.34 (Carathéodory)** *Soit* Ω *un ensemble, muni d'une algèbre* $\mathfrak{A}$ *de parties. On note* $\mathfrak{T} = \sigma(\mathfrak{A})$ *la tribu engendrée par* $\mathfrak{A}$. *Soit* $\mathbf{P}_0$ *une mesure σ-additive*[17] *sur* $(\Omega, \mathfrak{A})$. *Alors, il existe une unique mesure* **P** *sur* $(\Omega, \mathfrak{T})$ *dont la restriction à* $\mathfrak{A}$ *soit égale à* $\mathbf{P}_0$, *c'est-à-dire telle que*

$$\forall A \in \mathfrak{A} \qquad \mathbf{P}(A) = \mathbf{P}_0(A).$$

Remarque 3.35 Nous pouvons, avec ce théorème, finir de prouver la dernière assertion du théorème 3.30, à savoir : *Soit* F *une fonction vérifiant les quatre hypothèses : croissance, continuité à droite, limites* 1 *en* $+\infty$ *et* 0 *en* $-\infty$ *; alors il existe une unique probabilité* **P** *sur* $\mathfrak{B}$ *telle que*

$$\forall x \in \mathbb{R} \qquad F(x) = \mathbf{P}(]-\infty; x])$$

(17). Ou encore, ce qui revient au même, vérifiant la propriété de continuité monotone en $\varnothing$: si $A_n \downarrow \varnothing$, alors $\mathbf{P}(A_n) \to 0$.

DÉMONSTRATION : On note $\mathfrak{A}$ l'algèbre des parties de $\mathbb{R}$ qui sont union finies d'intervalles semi-ouverts $]a\,;\,b]$, où $-\infty \leqslant a < b < +\infty$. On définit sur $\mathfrak{A}$ l'application $\mathbf{P}_0$, à valeurs dans $[0\,;\,1]$, par

$$\mathbf{P}_0\Big(\sum_{k=1}^{n}]a_k\,;\,b_k]\Big) := \sum_{k=1}^{n} \big(\mathrm{F}(b_k) - \mathrm{F}(a_k)\big).$$

Cette application est évidemment additive sur $\mathfrak{A}$ et, comme déjà dit, on admet qu'elle vérifie la propriété de continuité monotone en $\varnothing$, ce qui revient à dire qu'elle est σ-additive.

De plus, $\mathfrak{A}$ est évidemment incluse dans $\mathfrak{B}$; puisqu'elle contient tous les intervalles de la forme $]-\infty\,;\,x]$, $\sigma(\mathfrak{A})$ contient la tribu engendrée par ces intervalles, qui est $\mathfrak{B}$. Ainsi, $\sigma(\mathfrak{A}) = \mathfrak{B}$, ce qui permet de construire $\mathbf{P}$ avec le théorème de Carathéodory. ∎

3.5 ANNEXE : UN ENSEMBLE NON MESURABLE (∗)

Vu la correspondance naturelle entre $[0\,;\,1[$ et le cercle S^1, nous allons prouver qu'il n'existe pas de mesure invariante par translation (c'est-à-dire, géométriquement, par rotation) définie sur l'ensemble $\mathfrak{P}$ de toutes les parties de S^1.

Tout d'abord, identifions le cercle S^1 avec le cercle unité de $\mathbb{C}$: $\mathrm{S}^1 \simeq \mathbb{U} = \{\mathrm{e}^{\mathrm{i}\theta}\ ;\ \theta \in \mathbb{R}\}$. On définit alors sur S^1 la relation d'équivalence $\sim$ par : $z \sim w$ si et seulement si il existe des réels α, β tels que $z = \mathrm{e}^{\mathrm{i}\alpha}$, $w = \mathrm{e}^{\mathrm{i}\beta}$ et $\alpha - \beta \in \mathbb{Q}$. L'axiome du choix permet de choisir, dans chacune des classes d'équivalences de S^1, un (et un seul) représentant ; notons A l'ensemble de ces représentants. La classe d'équivalence d'un élément z de A est

$$\overline{z} = \left\{\mathrm{e}^{\mathrm{i}q}\, z\ ;\ q \in \mathbb{Q}\right\}.$$

Si l'on définit

$$\mathrm{A}_q = \mathrm{e}^{\mathrm{i}q}\,\mathrm{A} = \left\{\mathrm{e}^{\mathrm{i}q}\, z\ ;\ z \in \mathrm{A}\right\},$$

alors

- les A_q sont deux à deux disjoints ;
- $\mathrm{S}^1 = \sum_{q\in\mathbb{Q}} \mathrm{A}_q$.

Pour démontrer le premier point, supposons que q et q' soient des rationnels distincts. Supposons qu'il existe un nombre w dans l'intersection de A_q et de $\mathrm{A}_{q'}$, alors on peut l'écrire $w = \mathrm{e}^{\mathrm{i}q}\, z = \mathrm{e}^{\mathrm{i}q'}\, z'$, où z et z' sont deux éléments de A. Notamment, puisque π est irrationnel, $\mathrm{e}^{\mathrm{i}(q-q')} \neq 1$ et donc $z \neq z'$. Pourtant, $z \sim z'$, ce qui contredit le fait que A ne possède qu'*un* seul représentant par classe d'équivalence.

Le deuxième point est immédiat, tout élément de S faisant partie d'une classe d'équivalence.

On peut maintenant conclure. Si $\mathbf{P}$ est une mesure de probabilité sur $(\mathrm{S}^1, \mathfrak{P})$, invariante par « translation », alors par σ-additivité, on a

$$1 = \mathbf{P}(\mathrm{S}^1) = \sum_{q\in\mathbb{Q}} \mathbf{P}(\mathrm{A}_q) = \sum_{q\in\mathbb{Q}} \mathbf{P}(\mathrm{A}).$$

Si $\mathbf{P}(\mathrm{A}) = 0$, on obtient $1 = 0$. Si $\mathbf{P}(\mathrm{A}) > 0$, on obtient $1 = +\infty$.

De cette construction, on déduit qu'il n'existe pas de probabilité invariante par « translation modulo 1 » sur l'ensemble des parties de $[0\,;\,1[$. La translation modulo 1 se construit ainsi : si x et y sont éléments de $[0\,;\,1[$, on pose $x \oplus y = x + y - \lfloor x + y \rfloor$, partie fractionnaire de $x + y$. Le translaté d'une partie A par une quantité t est $\mathrm{A} \oplus t = \{a \oplus t\ ;\ a \in \mathrm{A}\}$. Il n'existe donc pas de probabilité $\mathbf{P}$ telle que $\mathbf{P}(\mathrm{A}) = \mathbf{P}(\mathrm{A} \oplus t)$ pour toute partie A de $[0\,;\,1[$ et pour tout $t \in [0\,;\,1[$.

EXERCICES

♦ ***Exercice 3.1*** Vérifier que, si A et B sont des parties de Ω, $A \cup B = (A \setminus B) + B$.

♦ ***Exercice 3.2 (Un exemple d'algèbre)*** On note $\mathscr{A}$ la collection de toutes les parties de $\mathbb{R}$ qui sont des réunions finies d'intervalles semi-ouverts $]a\,;\,b]$ $(a, b \in \mathbb{R} \cup \{-\infty, +\infty\},\ a \leqslant b$, avec pour convention que $]a\,;\,a] = \varnothing$). Montrer que $\mathscr{A}$ est une algèbre, mais pas une tribu.

♦ ***Exercice 3.3*** On se donne une partition dénombrable de Ω : $\Omega = \bigcup_{n=0}^{\infty} A_n$ et $A_i \cap A_j = \varnothing$ dès que $i \neq j$. Décrire la tribu $\mathfrak{T} = \sigma(A_n\ ;\ n \in \mathbb{N})$ engendrée par $(A_n)_{n \geqslant 0}$.

♦ ***Exercice 3.4*** Soit Ω un ensemble. Montrer que l'ensemble des parties $A \subset \Omega$ telles que A est (fini ou) dénombrable *ou* A^c est (fini ou) dénombrable, est une tribu.

♦ ***Exercice 3.5*** Quelle est la tribu engendrée par l'ensemble S des singletons de $\mathbb{R}$?

♦ ***Exercice 3.6*** On pose $\Omega = \mathbb{R}$ et on note $\mathfrak{T}$ l'ensemble des parties de $\mathbb{R}$ qui sont dénombrables, ou de complémentaire dénombrable. Si A est dénombrable, on pose $\mathbf{P}(A) = 0$; si A^c est dénombrable, on pose $\mathbf{P}(A) = 1$. L'espace $(\Omega, \mathfrak{T}, \mathbf{P})$ est-il un espace probabilisé ?

♦ ***Exercice 3.7*** Montrer que la tribu $\mathfrak{B} = \mathfrak{B}(\mathbb{R})$ des boréliens de $\mathbb{R}$ est invariante par translation.

♦ ***Exercice 3.8 (Une algèbre sans atomes)*** Prenons pour univers $\Omega = \mathbb{Z}$, ensemble des entiers relatifs. Convenons d'appeler *partie périodique* de Ω une partie A telle qu'il existe un entier $p > 0$ et une partie $K \subset \{0, 1, \ldots, p-1\}$ vérifiant

$$A = \bigcup_{n \in \mathbb{Z}} \tau_{np}(K),$$

où $\tau_{np}(K)$ est le translaté de K de la quantité np :

$$\tau_{np}(K) := K + np := \{k + np\ ;\ k \in K\}.$$

On note enfin $\mathfrak{A}$ l'ensemble des parties périodiques de Ω.

i) Montrer que $\mathfrak{A}$ est une algèbre sur Ω.

ii) On veut montrer que $\mathfrak{A}$ ne possède pas d'atomes. Soit $A \neq \varnothing$ un élément de $\mathfrak{A}$; il existe un entier p et une partie $K \subset \{0, 1, \ldots, p-1\}$ tels que $A = \bigcup_n \tau_{np}(K)$. On pose

$$B = \bigcup_{n \in \mathbb{Z}} \tau_{2np}(K).$$

Vérifier que $\varnothing \subsetneq B \subsetneq A$ et conclure.

♦ ***Exercice 3.9 (Une tribu sans atomes)*** Notons $\Omega = \mathscr{F}(\mathbb{R}, \{0,1\})$, ensemble des applications définies sur $\mathbb{R}$ et à valeurs dans $\{0,1\}$. Pour toute partie D finie ou dénombrable de $\mathbb{R}$ et toute partie B non vide de l'espace $\mathscr{F}(D, \{0,1\})$ des fonctions définies sur D et à valeurs dans $\{0,1\}$, notons

$$A(D, B) := \left\{ f \in \Omega\ ;\ f_{|_D} \in B \right\}.$$

Dans le cas particulier où D est un singleton, introduisons la notation

$$U_x := A\big(\{x\}, \{x \mapsto 1\}\big) = \{f \in \Omega\ ;\ f(x) = 1\}$$

et

$$Z_x := A\big(\{x\}, \{x \mapsto 0\}\big) = \{f \in \Omega\ ;\ f(x) = 0\} = U_x^c$$

Notons enfin $\mathfrak{T}$ l'ensemble contenant tous les $A(D, B)$ (pour D dénombrable et B non vide tel que précédemment) ainsi que la partie vide.

i) Que vaut $A(D, B)^c$?

ii) Que vaut $A(D, B) \cap A(D', B')$?

iii) Montrer que $\mathfrak{T}$ est une tribu, et que c'est la tribu engendrée par les U_x.

iv) Si $A(D, B)$ est un élément de $\mathfrak{T}$, et si x est un réel n'appartenant pas à D, posons $D' := D \cup \{x\}$ et B' l'ensemble des fonctions f telles que $f_{|_D} \in B$ et $f(x) = 1$. Vérifier que $A(D', B') = A(D, B) \cap U_x$ est strictement inclus dans $A(D, B)$. En déduire que $\mathfrak{T}$ ne possède pas d'atomes.

♦ ***Exercice 3.10 (Inégalité de sous-additivité Boole)***
Soit $(A_n)_{n\geqslant 0}$ une suite d'événements d'un espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$. Montrer que

$$\mathbf{P}\Big(\bigcup_{n=0}^{\infty} A_n\Big) \leqslant \sum_{n=0}^{\infty} \mathbf{P}(A_n),$$

le second membre pouvant éventuellement diverger.

En déduire qu'une union dénombrable d'ensemble négligeables est négligeable ; puis, qu'une intersection dénombrable d'événements presque sûrs est presque sûre.

♦ ***Exercice 3.11 (Utilisation de la formule du crible)*** Une loterie communale réunit n personnes ; r lots sont à distribuer et on choisit pour cela r fois une des personnes présentes, de façon équiprobable. Une personne peut être désignée pour plusieurs lots (c'est un *tirage avec remise*).

i) Quelle est la probabilité qu'une personne P_1 ne reçoive aucun lot ?

ii) Soit k un entier, $1 \leqslant k \leqslant n-1$. On choisit k personnes distinctes $P_1, P_2, \ldots, P_k$. Quelle est la probabilité qu'aucune de ces personnes ne reçoive de lot ?

iii) On suppose que le nombre de lots est supérieur ou égal au nombre de participants. Quelle est la probabilité que chacune des n personnes reçoive au moins un lot ?

♦ ***Exercice 3.12 (Continuité monotone en $\varnothing$)*** Montrer que, si $\mathbf{P} : \mathfrak{T} \to \mathbb{R}^+$ est une application additive et telle que $\mathbf{P}(\Omega) = 1$, l'axiome de σ-additivité est équivalent à l'axiome de continuité monotone en $\varnothing$: pour toute suite $(A_n)_{n\geqslant 0}$ d'événements, décroissant vers $\varnothing$, c'est-à-dire telle que $\bigcap_n A_n = \varnothing$, on a $\lim \mathbf{P}(A_n) = 0$.

♦ ***Exercice 3.13*** Soit $(\mathfrak{T}_n)_{n\geqslant 1}$ une suite croissante de tribus. On pose $\mathfrak{A} = \bigcup_{n=1}^{\infty} \mathfrak{T}_k$. Montrer que $\mathfrak{A}$ est une algèbre. *Remarque : ce n'est pas forcément une tribu...*

♦ ***Exercice 3.14*** Soient F et G des fonctions de répartition. Soit α un réel strictement positif, et soit λ un réel de $[0\,;\,1]$. Montrer que les fonctions suivantes sont encore des fonctions de répartition :

i) $x \mapsto F(x)^\alpha$;

ii) $x \mapsto \dfrac{F(x)^2}{2} + \dfrac{\ln(1+F(x))}{\ln 4}$;

iii) $\lambda F + (1-\lambda)G$.

Si $F_1, F_2, \ldots, F_n$ sont des fonctions de répartition, montrer que $G : x \mapsto F_1(x)\,F_2(x)\cdots F_n(x)$ est encore une fonction de répartition.

Indications

◊ *Indication 3.3 :* Montrer que c'est l'ensemble des parties de la forme $\bigcup_{i\in J} A_i$, où J est une partie quelconque de $\mathbb{N}$.

◊ *Indication 3.5 :* C'est l'ensemble des parties de $\mathbb{R}$ qui sont dénombrables, ou complémentaires d'un ensemble dénombrable ; le fait que cet ensemble soit une tribu est l'objet de l'exercice 3.4.

◊ *Indication 3.6 :* Oui.

◊ *Indication 3.7 :* Fixer a. Noter $\mathfrak{T}_a$ l'ensemble des éléments de la forme $a+B$, B parcourant $\mathfrak{B}$. Montrer que $\mathfrak{T}_a$ est une tribu.

◊ *Indication 3.8 :* Si A est périodique de période p et B périodique de période q, montrer que $A \cup B$ est périodique de période pq.

◊ *Indication 3.10 :* Introduire la suite croissante de terme général $B_n = A_0 \cup \cdots \cup A_n$.

◊ *Indication 3.12 :* Supposer la continuité monotone en $\varnothing$ et montrer l'axiome 3′ ; pour cela, poser $B = \lim_n \uparrow B_n$ et observer que $B = (B - B_n) + B_n$. Pour la réciproque, utiliser l'axiome 3″.

SOLUTIONS

♦ **Solution 3.3** La tribu $\mathfrak{T}$ est la plus petite tribu contenant tous les A_n, donc contient, par additivité dénombrable, tous les éléments de la forme $\bigcup_{i\in I} A_i$, pour I parcourant l'ensemble des parties de $\mathbb{N}$. Notons $\mathfrak{T}'$ l'ensemble des parties de cette forme ; on a donc $\mathfrak{T}' \subset \mathfrak{T}$. Il ne reste qu'à remarquer que $\mathfrak{T}'$ est une tribu, pour conclure que $\mathfrak{T} = \mathfrak{T}'$.

♦ **Solution 3.4** Dans tout l'exercice, par dénombrable, on entendra « dénombrable au sens large », c'est-à-dire « fini ou en bijection avec $\mathbb{N}$ ».

Notons $\mathfrak{D}$ l'ensemble des parties dénombrables, et $\mathfrak{C}$ l'ensemble des parties dont le complémentaire est dénombrable. On doit montrer que $\mathfrak{T} := \mathfrak{D} \cup \mathfrak{C}$ est une tribu.

— Ω est élément de $\mathfrak{C}$, donc de $\mathfrak{T}$.

— $\mathfrak{T}$ est évidemment stable par passage au complémentaire : si $A \in \mathfrak{D}$, alors $A^c \in \mathfrak{C}$ et réciproquement.

— Soit $(A_n)_{n\geqslant 1}$ une suite dénombrable d'éléments de $\mathfrak{T}$. Si les A_n sont tous des éléments de $\mathfrak{D}$, alors $A := \bigcup_{n=1}^{\infty} A_n$, union dénombrable d'ensembles dénombrables, est dénombrable, c'est-à-dire $A \in \mathfrak{D}$. Si au contraire il existe au moins un indice i tel que A_i est élément de $\mathfrak{C}$, alors le complémentaire de A est dénombrable : en effet,

$$A^c = \left(\bigcup_{n=1}^{\infty} A_n\right)^c = \left(\bigcap_{n\neq i} A_n^c\right) \cap A_i^c$$

est inclus dans A_i^c, qui est dénombrable, donc est dénombrable ; le point précédent montre qu'alors A est élément de $\mathfrak{C}$.

Dans les deux cas, A est élément de $\mathfrak{T}$.

♦ **Solution 3.5** Notons $\mathfrak{D}$ l'ensemble des parties dénombrables de $\mathbb{R}$, et $\mathfrak{C}$ l'ensemble des parties qui sont complémentaires d'un ensemble dénombrable. Notons enfin $\mathfrak{T} := \mathfrak{C} \cup \mathfrak{D}$. Il est clair que $\sigma(S)$ contient $\mathfrak{D}$ (dont les éléments sont union dénombrables d'éléments de S) et $\mathfrak{C}$, donc $\mathfrak{T} \subset \sigma(S)$. Il ne reste plus, pour conclure, qu'à noter que $\mathfrak{T}$ est une tribu d'après l'exercice 3.4.

♦ **Solution 3.7** Soit a un réel. Notons $\mathfrak{T}_a$ l'ensemble des éléments de la forme $a + B$, pour a parcourant $\mathbb{R}$ et B parcourant $\mathfrak{B}$. C'est bien une tribu puisque $(a+B)^c = a+B^c$ et $\bigcup_n (a+B_n) = a + \bigcup_n B_n$. De plus, elle contient tous les intervalles $]-\infty\,;\,x]$, puisque l'on peut écrire

$$]-\infty\,;\,x] = a +]-\infty\,;\,x-a].$$

Elle contient donc la tribu engendrée par ces intervalles, qui n'est autre que la tribu $\mathfrak{B}$ des boréliens. En d'autres termes, tout borélien B étant dans $\mathfrak{T}_a$ s'écrit sous la forme $B = a + B'$, avec $B' \in \mathfrak{B}$. Réciproquement, soit C un élément de $\mathfrak{T}_a$, on l'écrit sous la forme $C = a + B$, où B est un borélien ; le raisonnement fait précédemment ne dépendant pas de la valeur de a, on peut appliquer la conclusion pour la valeur $-a$: B peut s'écrire sous la forme $B = -a + B'$ avec B borélien, et donc $C = a + B = B'$ est un borélien. Au total, on a bien montré que $\mathfrak{T}_a = \mathfrak{B}$.

♦ **Solution 3.9**

i) $A(D, B)^c = A(D, B^c)$.

ii) Notons B'' l'ensemble des fonctions de $D \cup D'$ dans $\{0, 1\}$ dont la restriction à D est dans B et dont la restriction à D' est dans B'. Si D et D' ne sont pas disjoints, il est possible que cette condition soit impossible à réaliser, auquel cas B'' est vide et $A(D, B) \cap A(D', B') = \varnothing$. Dans le cas contraire, $A(D, B) \cap A(D', B') = A(D \cup D', B'')$.

iii) Tout d'abord, $\Omega = A(\{0\}, B)$ où $B = \mathscr{F}(\{0\}, \{0, 1\})$, donc Ω est élément de $\mathfrak{T}$. La question *i)* montre que $\mathfrak{T}$ est stable par passage au complémentaire, et *ii)* que $\mathfrak{T}$ est stable par intersection finie. Le même raisonnement montre que, lorsque cette intersection n'est pas vide,

$$\bigcap_{n=1}^{\infty} A(D_n, B_n) = A\left(\bigcup_{n=1}^{\infty} D_n, B\right),$$

où $D := \bigcup D_n$ est encore une partie dénombrable de $\mathbb{R}$, et B l'ensemble des fonctions f de D dans $\{0, 1\}$ telles que la restriction de f à D_n est dans B_n pour tout entier n.

Ainsi, $\mathfrak{T}$ est bien une tribu. Elle contient tous les U_x, donc contient la tribu $\mathfrak{U}$ engendrée par ces parties. Réciproquement, tout $A(D, B)$ est union dénombrable de U_x et de $Z_x = U_x^c$, ce qui prouve que $\mathfrak{T}$ est incluse dans $\mathfrak{U}$ et achève la preuve.

iv) Si D est dénombrable, alors D est strictement inclus dans $\mathbb{R}$ (non dénombrable), donc on peut trouver un réel $x \notin D$ et effectuer la construction proposée. Il est immédiat de vérifier que $A(D, B) \cap Z_x$ est non vide, et donc $A(D, B) \cap U_x$ est strictement inclus dans $A(D, B)$; l'ensemble B et l'ensemble $\{x \mapsto 1\}$ étant compatibles (au sens précédemment vu), $A(D, B) \cap U_x$ est non vide.

Ainsi, tout élément non vide de $\mathfrak{T}$ admet un élément non vide qui lui est strictement inclus : *il n'existe pas d'atomes dans la tribu* $\mathfrak{T}$.

♦ **Solution 3.10** On sait que $\mathbf{P}(A \cup B) \leqslant \mathbf{P}(A) + \mathbf{P}(B)$ pour tous événements A et B. On en déduit, par récurrence, que $\mathbf{P}(A_0 \cup \cdots \cup A_n) \leqslant \mathbf{P}(A_0) + \cdots + \mathbf{P}(A_n)$ pour tout $n \in \mathbb{N}$.

Si l'on note maintenant $B_n = A_0 \cup \cdots \cup A_n$, la suite $(B_n)_{n \geqslant 0}$ est croissante, et

$$\mathbf{P}(B_n) = \mathbf{P}(A_0 \cup \cdots \cup A_n) \leqslant \mathbf{P}(A_0) + \cdots + \mathbf{P}(A_n)$$

pour tout $n \in \mathbb{N}$. De plus,

$$\mathbf{P}\Big(\bigcup_{n=0}^{\infty} A_n\Big) = \mathbf{P}\Big(\bigcup_{n=0}^{\infty} B_n\Big) = \mathbf{P}\big(\lim_{n\to\infty} \uparrow B_n\big) = \lim_{n\to\infty} \mathbf{P}(B_n)$$

d'après l'axiome 3′. Enfin, l'inégalité précédente donne

$$\lim_{n\to\infty} \mathbf{P}(B_n) \leqslant \sum_{n=0}^{\infty} \mathbf{P}(A_n).$$

De manière alternative, on peut également transformer l'union des A_n en une union disjointe : pour cela, on note

$$A_0' = A_0 \qquad A_n' = A_n - \bigcup_{k=0}^{n-1} A_k$$

Par construction, les A_n' sont disjoints deux à deux, $A_n' \subset A_n$ et $A_0 \cup \cdots \cup A_n = A_0' \cup \cdots \cup A_n'$. On peut alors utiliser la propriété de σ-additivité de la mesure de probabilité pour obtenir

$$\mathbf{P}\Big(\bigcup_{n=0}^{\infty} A_n\Big) = \mathbf{P}\Big(\sum_{n=0}^{\infty} A_n'\Big) = \sum_{n=0}^{\infty} \mathbf{P}(A_n') \leqslant \sum_{n=0}^{\infty} \mathbf{P}(A_n).$$

Si $(A_n)_{n \geqslant 0}$ est une suite d'événement négligeables, alors

$$\mathbf{P}\Big(\bigcup_{n=0}^{\infty} A_n\Big) \leqslant \sum_{n=0}^{\infty} \mathbf{P}(A_n) = \sum_{n=0}^{\infty} 0 = 0.$$

En passant au complémentaire, on démontre la dernière propriété.

♦ **Solution 3.11**

i) On considère la personne P_1. Pour qu'elle ne reçoive aucun lot, il faut qu'elle soit perdante à chaque tirage, ce qui, si les tirages sont indépendants (ce qui est implicitement assumé), se produit avec la probabilité $\left(\dfrac{n-1}{n}\right)^r = \left(1 - \dfrac{1}{n}\right)^r$.

ii) $P_1, \ldots, P_k$ ne reçoivent aucun lot au premier tirage avec la probabilité $\frac{n-k}{n}$ (le lot numéro 1 a été distribué à l'une quelconques des $n - k$ autres personnes). La probabilité qu'elles n'aient jamais de lot est donc $\left(\frac{n-k}{n}\right)^r$.

iii) On suppose donc que $r \geqslant n$. Notons A l'événement

$$A : \text{« chacune des } n \text{ personnes reçoit un lot »}.$$

Pour $i = 1, \ldots, n$, notons B_i l'événement

$$B_i : \text{« la personne } P_i \text{ n'a pas eu de lot »}.$$

Le complémentaire de A est $A^c = B_1 \cup B_2 \cup \cdots \cup B_n$. Malheureusement, ces événements ne sont pas incompatibles : on ne peut donc pas sommer les probabilités correspondantes. Il ne nous reste, hélas, plus qu'une solution de dernière extrémité : utiliser la formule du crible.

$$\mathbf{P}(A^c) = \mathbf{P}(B_1 \cup B_2 \cup \cdots \cup B_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \mathbf{P}(B_{i_1} \cap \cdots \cap B_{i_k}).$$

Dans cette dernière expression, on peut simplifier bien des choses. Tout d'abord, pour une valeur de k fixée et pour des indices $i_1, \ldots, i_k$ également fixés, on remarque que

$$\mathbf{P}(B_{i_1} \cap \cdots \cap B_{i_k}) = \mathbf{P}(B_1 \cap \cdots \cap B_k) = \left(\frac{n-k}{n}\right)^r,$$

puisque toutes les personnes ont des rôles symétriques. De plus, toujours pour k fixé, il y a $\binom{n}{k}$ manières de choisir les indices ordonnés (en effet, il y a autant de manières de

choisir k indices ordonnés dans le sens croissant que de manières de choisir un ensemble de k indices distincts). On en déduit que

$$\mathbf{P}(\mathrm{A}^c) = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \left(\frac{n-k}{n}\right)^r.$$

Cela mène, par passage au complémentaire, à

$$\mathbf{P}(\mathrm{A}) = 1 - \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \left(\frac{n-k}{n}\right)^r = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left(\frac{n-k}{n}\right)^r.$$

♦ **Solution 3.12** Supposons la continuité monotone en $\varnothing$. Soit $(\mathrm{B}_n)_{n\geqslant 0}$ une suite croissante d'éléments de $\mathfrak{T}$, et posons

$$\mathrm{B} = \lim_{n\to\infty} \uparrow \mathrm{B}_n = \bigcup_{n=0}^{\infty} \mathrm{B}_n.$$

On peut alors écrire pour tout $n \in \mathbb{N}$:

$$\mathrm{B} = (\mathrm{B} - \mathrm{B}_n) + \mathrm{B}_n.$$

La suite de terme général $(\mathrm{B} - \mathrm{B}_n)$ est décroissante et de limite $\varnothing$ et la propriété d'additivité permet d'écrire que $\mathbf{P}(\mathrm{B}) = \mathbf{P}(\mathrm{B} - \mathrm{B}_n) + \mathbf{P}(\mathrm{B}_n)$, donc

$$\mathbf{P}(\mathrm{B}_n) = \mathbf{P}(\mathrm{B}) - \mathbf{P}(\mathrm{B} - \mathrm{B}_n) \xrightarrow[n\to\infty]{} \mathbf{P}(\mathrm{B}).$$

Réciproquement, la σ-additivité implique l'axiome 3″ qui implique évidemment l'axiome 3‴.

Résumé du chapitre

On munit un ensemble Ω (appelé *univers*) de sa *tribu des événements*, ensemble de parties de Ω stable par intersection et unions finies ou dénombrables et par passage au complémentaire, et tel que $\Omega \in \mathfrak{T}$. On définit alors une probabilité sur $(\Omega, \mathfrak{T})$ comme une application de $\mathfrak{T}$ dans $[0\,;1]$ telle que $\mathbf{P}(\Omega) = 1$ et telle que $\mathbf{P}(\sum \mathrm{A}_n) = \sum \mathrm{A}_n$: la probabilité d'une union disjointe dénombrable est la somme des probabilités. Le triplet $(\Omega, \mathfrak{T}, \mathbf{P})$ est un *espace probabilisé*.

Un cas particulier d'univers est $\Omega := \mathbb{R}$: on montre alors que toutes les probabilités sur $\mathbb{R}$ sont entièrement caractérisées par leur *fonction de répartition* $\mathrm{F} : x \mapsto \mathbf{P}(]-\infty\,;x])$.

En annexe, on montre qu'on ne peut pas espérer munir tout espace de sa tribu la plus riche, à savoir l'ensemble $\mathfrak{P}(\Omega)$ de toutes les parties de Ω.

Chapitre 4

Conditionnement

On a une chance sur dix mille, quand on prend un avion, qu'il y ait une bombe à l'intérieur. Mais on n'a qu'une chance sur cent millions qu'il y en ait deux. Alors, pour plus de sécurité, emportez votre bombe.

Notre intuition des probabilités est souvent basée sur des probabilités conditionnées par un événement (« quelle est la probabilité qu'il pleuve aujourd'hui sachant qu'on est en novembre » n'obtient pas la même réponse que la même question « sachant qu'on est en juin »). La formalisation d'une probabilité conditionnelle, pour aisée qu'elle soit, permet d'établir les premiers outils de calcul de la trousse de survie du probabiliste (occasionnel ou enragé) :

- *formule des probabilités composées ;*
- *formule des probabilités totales ;*
- *théorème de Bayes.*

Probabilités conditionnelles

§ 58

Soit $(\Omega, \mathfrak{T}, \mathbf{P})$ un espace probabilisé. Si A est un événement, on souhaite donner un sens à « la probabilité qu'un événement B soit réalisé, *sachant* que A est réalisé », comme il a été présenté dans le chapitre 1. On ne s'intéresse qu'aux expériences dont le résultat réalise A, les autres étant purement et simplement écartées du décompte. Sous ces conditions, B est réalisé si et seulement si $A \cap B$ l'est. Tout se passe comme si l'on choisissait le nouvel univers $\Omega' := A$ et qu'on le munissait d'une nouvelle probabilité $\mathbf{P}_A$, telle que $\mathbf{P}_A(B)$ soit proportionnelle à $\mathbf{P}(A \cap B)$. Pour cela, il est donc nécessaire de supposer que $\mathbf{P}(A) > 0$[1], et la condition de normalisation $\mathbf{P}_A(A) = 1$ mène naturellement à la définition qui suit[2].

Définition 4.1 Soit A un événement de probabilité non nulle : $\mathbf{P}(A) > 0$. Pour tout événement B, on définit la **probabilité conditionnelle de B sachant A** comme le réel, noté $\mathbf{P}_A(B)$ ou $\mathbf{P}(B \mid A)$, valant

$$\mathbf{P}_A(B) = \mathbf{P}(B \mid A) := \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)}.$$

(1). Le cas où $\mathbf{P}(A) = 0$ est beaucoup plus délicat à gérer ; des techniques spécifiques sont expliquées au chapitre 23.

(2). On notera que l'univers est toujours Ω, et non A ; tout ce qui est en dehors de A est cependant considéré comme *négligeable*.

Le point essentiel de cette définition est que :

■ **Théorème 4.2** *L'application*

$$\begin{aligned} \mathbf{P}_A : \ \mathfrak{T} &\longrightarrow [0;1], \\ B &\longmapsto \mathbf{P}(B \mid A) \end{aligned}$$

est une probabilité sur $(\Omega, \mathfrak{T})$ *; on l'appelle* **probabilité conditionnelle sachant A**, *ou encore* **probabilité conditionnée par A**.

DÉMONSTRATION : Vérifions explicitement que les trois axiomes définissant une probabilité sont bien respectés. Tout d'abord *(axiome 1)*, $\mathbf{P}_A$ est bien à valeurs positives. De plus, *(axiome 2)*

$$\mathbf{P}_A(\Omega) = \frac{\mathbf{P}(\Omega \cap A)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A)}{\mathbf{P}(A)} = 1.$$

Enfin *(axiome 3)*, si $(B_n)_{n \geqslant 0}$ est une suite d'éléments deux à deux disjoints de $\mathfrak{T}$, alors les $B_n \cap A$ sont également deux à deux disjoints, donc

$$\mathbf{P}_A\left(\bigcup_{n=0}^{\infty} B_n\right) = \frac{\mathbf{P}\left[\left(\bigcup_{n=0}^{\infty} B_n\right) \cap A\right]}{\mathbf{P}(A)} = \frac{\mathbf{P}\left(\bigcup_{n=0}^{\infty} B_n \cap A\right)}{\mathbf{P}(A)}$$

$$= \sum_{n=0}^{\infty} \frac{\mathbf{P}(B_n \cap A)}{\mathbf{P}(A)} = \sum_{n=0}^{\infty} \mathbf{P}_A(B_n).$$

∎

De la définition de la probabilité conditionnelle, on déduit immédiatement
⇙4.2 quelques propriétés simples lorsque $\mathbf{P}(A) > 0$:

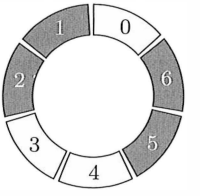
$$\mathbf{P}(A \mid A) = 1 \qquad \mathbf{P}(B \mid A) = 1 \quad \text{si } A \subset B$$
$$\mathbf{P}(\varnothing \mid A) = 0 \qquad \mathbf{P}(B \mid A) + \mathbf{P}(B^c \mid A) = 1.$$

Remarque 4.3 Attention, de manière générale, on a

$$\mathbf{P}(B \mid A) + \mathbf{P}(B \mid A^c) \neq 1,$$

cette somme pouvant prendre une valeur quelconque entre 0 et 2. (Exemple : je joue au loto ; je note B l'événement « je perds ma mise » et A l'événement « il fait beau » ; la somme précédente est dramatiquement proche de 2.)

Exemple 4.4 (Une mini-roulette) On considère une roulette très simplifiée, ne comportant que sept chiffres (de 0 à 6), et deux couleurs (Noir et Blanc), disposés selon le schéma ci-dessous.



On lance une bille qui peut tomber, avec une probabilité uniforme, dans chacune des sept cases de la roulette. La probabilité que la bille tombe sur un chiffre impair est $\mathbf{P}(\text{impair}) = 3/7$. La probabilité qu'elle tombe sur un chiffre impair *sachant qu'elle est tombée sur une case blanche* est

$$\mathbf{P}(\text{impair} \mid \text{blanc}) = \frac{\mathbf{P}(\text{impair, blanc})}{\mathbf{P}(\text{blanc})} = \frac{1/7}{3/7} = \frac{1}{3}.$$

♦ ***Exercice 4.1*** Calculer, dans le modèle de la mini-roulette précédente, la probabilité de tirer un nombre impair sachant que la bille est tombée sur une case noire. Calculer de même les probabilités conditionnelles $\mathbf{P}\{\text{couleur} \mid \text{parité}\}$, pour les deux couleurs et les deux parités possibles.

Remarque 4.5 (Sur les notations $\mathbf{P}_A(B)$ *et* $\mathbf{P}(B \mid A)$*)* Dans les ouvrages anglo-saxons, les probabilités conditionnelles sont presque toujours notées « $\mathbf{P}(B \mid A)$ » ; l'usage s'est répandu en France de les noter « $\mathbf{P}_A(B)$ ». À ceci, on trouve deux avantages. Le premier est de pouvoir parler de la mesure de probabilité elle-même, notée $\mathbf{P}_A$. Le second est d'ordre pédagogique : il permet de rappeler qu'il s'agit de travailler avec une autre probabilité que celle qui est initialement définie, et de ne pas oublier qu'il n'existe pas d'événement « $(B \mid A)$ ». Nous emploierons les deux notations, selon les exigences contextuelles ou typographiques.

Formule des probabilités composées § 59

Il résulte immédiatement de la définition d'une probabilité conditionnelle que, si A est un événement de probabilité non nulle, alors

$$\mathbf{P}(\mathrm{A} \cap \mathrm{B}) = \mathbf{P}(\mathrm{A}) \cdot \mathbf{P}(\mathrm{B} \mid \mathrm{A}). \tag{4.1}$$

Cette formule s'interprète souvent en terme de connaissance croissante des événements : pour calculer la probabilité que les événements A et B aient lieu tous deux, on calcule d'abord la probabilité de réalisation de A, puis la probabilité de réalisation de B sachant que A est réalisé.

Exemple 4.6 Les jours de semaine, lorsqu'il fait beau, je vais manger un sandwich au Parc de la Tête d'Or avec une probabilité $p = 1/3$ (c'est-à-dire, en moyenne, un jour sur trois). Au printemps, il fait beau avec une probabilité $q = 3/4$ (en moyenne trois jours sur quatre). Je peux donc calculer la probabilité de manger un sandwich au soleil en notant A l'événement « il fait beau » et B l'événement « je vais manger un sandwich au Parc ». On a $\mathbf{P}(\mathrm{A}) = q = 3/4$ et $\mathbf{P}(\mathrm{B} \mid \mathrm{A}) = p = 1/3$, donc $\mathbf{P}(\mathrm{A} \cap \mathrm{B}) = 1/4$.

On remarquera sur cet exemple que la modélisation des différents événements (qualité du temps, nature du repas) se fait naturellement avec des probabilités conditionnelles. Il arrive parfois d'hésiter dans les interprétations de certaines probabilités : est-ce $\mathbf{P}(\mathrm{A} \cap \mathrm{B})$ ou $\mathbf{P}(\mathrm{B} \mid \mathrm{A})$ que l'on connaît ? Le plus souvent, c'est la probabilité conditionnelle $\mathbf{P}(\mathrm{B} \mid \mathrm{A})$ qui est la donnée « naturelle ».

La formule (4.1) se généralise à une famille finie d'événements :

■ **Théorème 4.7 (Formule des probabilités composées)** *Soient $n \geqslant 2$ un entier et soient $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_n$ des événements ; on suppose que $\mathrm{A} = \mathrm{A}_1 \cap \mathrm{A}_2 \cap \cdots \cap \mathrm{A}_{n-1}$ est de probabilité non nulle. Alors*

$$\mathbf{P}(\mathrm{A}_1 \cap \mathrm{A}_2 \cap \cdots \cap \mathrm{A}_n) = \mathbf{P}(\mathrm{A}_1) \cdot \mathbf{P}(\mathrm{A}_2 \mid \mathrm{A}_1) \cdot \mathbf{P}(\mathrm{A}_3 \mid \mathrm{A}_1 \cap \mathrm{A}_2) \cdots \\ \cdots \mathbf{P}(\mathrm{A}_n \mid \mathrm{A}_1 \cap \cdots \cap \mathrm{A}_{n-1}). \tag{4.2}$$

DÉMONSTRATION : Une simple récurrence ; le cas $n = 2$ correspond à l'équation (4.1). Supposons la propriété démontrée à un rang $n \geqslant 2$ et soient $\mathrm{A}_1, \ldots, \mathrm{A}_{n+1}$ des événements tels que $\mathbf{P}(\mathrm{A}_1 \cap \cdots \cap \mathrm{A}_n) > 0$. Notons $\mathrm{A}' = \mathrm{A}_1 \cap \cdots \cap \mathrm{A}_n$; puisque $\mathrm{A}_1 \cap \cdots \cap \mathrm{A}_{n-1}$ est de probabilité non nulle (car au moins égale à $\mathbf{P}(\mathrm{A}') > 0$), la propriété de récurrence permet d'écrire

$$\mathbf{P}(\mathrm{A}') = \mathbf{P}(\mathrm{A}_1) \cdot \mathbf{P}(\mathrm{A}_2 \mid \mathrm{A}_1) \cdot \mathbf{P}(\mathrm{A}_3 \mid \mathrm{A}_1 \cap \mathrm{A}_2) \cdots \mathbf{P}(\mathrm{A}_n \mid \mathrm{A}_1 \cap \cdots \cap \mathrm{A}_{n-1}).$$

Enfin, la relation à l'ordre 2 s'écrit

$$\mathbf{P}(\mathrm{A}' \cap \mathrm{A}_{n+1}) = \mathbf{P}(\mathrm{A}') \cdot \mathbf{P}(\mathrm{A}_{n+1} \mid \mathrm{A}'),$$

ce qui est la formule voulue à l'ordre $n + 1$. ∎

Formule des probabilités totales § 60

Le résultat suivant, bien qu'élémentaire, est d'une grande importance dans le calcul des probabilités.

■ **Théorème 4.8 (Formule des probabilités totales)** *Soient $(\mathrm{A}_i)_{i \in \mathrm{I}}$ une famille, finie ou dénombrable, formant un système complet d'événements. Soit B un événement. On suppose que $\mathbf{P}(\mathrm{A}_i) > 0$ pour tout $i \in \mathrm{I}$. Alors*

$$\mathbf{P}(\mathrm{B}) = \sum_{i \in \mathrm{I}} \mathbf{P}(\mathrm{A}_i)\, \mathbf{P}(\mathrm{B} \mid \mathrm{A}_i).$$

DÉMONSTRATION : Puisque $(A_i)_{i\in I}$ est une partition, on peut écrire

$$\mathbf{P}(B) = \mathbf{P}(B \cap \Omega) = \mathbf{P}\Big(B \cap \bigcup_{i\in I} A_i\Big) = \mathbf{P}\Big(\bigcup_{i\in I} B \cap A_i\Big).$$

Or les événements $B \cap A_i$ sont disjoints deux à deux ; par σ-additivité, on a donc

$$\mathbf{P}(B) = \sum_{i\in I} \mathbf{P}(B \cap A_i) = \sum_{i\in I} \mathbf{P}(A_i)\,\mathbf{P}(B \mid A_i). \qquad \text{4.3}$$

∎

§61 Théorème de Bayes

La relation $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B \mid A)$ introduit une dissymétrie dans la manière d'envisager les événements A et B. Dans le membre de droite, A est considéré comme une « cause »[3] de l'événement B. On peut également l'interpréter par l'introduction d'une notion de chronologie entre les événements : la probabilité d'avoir les événements A et B s'obtient en évaluant tout d'abord la probabilité d'avoir l'événement A puis, une fois A établi, d'avoir l'événement B sachant A.

Rien n'est plus simple pourtant que de rétablir la symétrie, en écrivant la double égalité

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B \mid A) = \mathbf{P}(B) \cdot \mathbf{P}(A \mid B),$$

valable dès que A et B sont des événements de probabilité non nulle. Cela permet de tirer le résultat, toujours élémentaire et fondamental :

■ **Théorème 4.9 (Théorème de Bayes, v.1)** *Soient* A *et* B *deux événements de probabilité non nulle. On a alors*

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A) \cdot \mathbf{P}(B \mid A)}{\mathbf{P}(B)}.$$

Remarque 4.10 (Causes, antécédence, etc.) Le théorème de Bayes[4] est parfois appelé *formule de probabilité des causes (connaissant les conséquences)*, puisqu'il permet de renverser le sens de la causalité entre les événements. On peut préférer le terme d'*antécédence* plutôt que celui de causalité, ce dernier véhiculant un lot trop important d'idées en arrière plan. La formule de Bayes permet donc de *renverser la chronologie.*

Exemple 4.11 (« Tu finiras sur l'échafaud ! ») C'est ce que s'exclame une grand-tante excédée à l'adresse de son petit-neveu qui s'est gavé de bonbons et s'obstine à le nier. « On commence par mentir, et on devient criminel ! », ajoute-t-elle, car elle a de la culture et a lu dans son hebdomadaire favori que 90 % des criminels ont commencé par mentir effrontément durant leur jeunesse (alors que seuls 20 % des gens honnêtes ont menti étant jeunes).

Sachant qu'il n'y a que 1 % de criminels dans la population, peut-on calculer la probabilité qu'a le petit enfant de devenir criminel et rassurer ainsi la brave grand-tante ?

Notons C l'événement « être criminel » et M l'événement « mentir durant son enfance ». Les données chiffrées connues sont :

$$\mathbf{P}(C) = 0,01 \qquad \mathbf{P}(M \mid C) = 0,9 \qquad \text{et} \qquad \mathbf{P}(M \mid C^c) = 0,2.$$

Le théorème de Bayes nous permet d'écrire

$$\mathbf{P}(C \mid M) = \frac{\mathbf{P}(C) \cdot \mathbf{P}(M \mid C)}{\mathbf{P}(M)},$$

mais il nous manque la valeur de $\mathbf{P}(M)$. On peut l'obtenir par la formule des probabilités totales, en utilisant le système complet d'événements (C, C^c) :

$$\mathbf{P}(M) = \mathbf{P}(M \mid C)\,\mathbf{P}(C) + \mathbf{P}(M \mid C^c)\,\mathbf{P}(C^c),$$

---

(3). Dans un sens purement probabiliste, et non au sens habituel de la causalité

(4). Du nom du prêtre et théologien protestant anglais Thomas BAYES (1702–1761), qui étudia les mathématiques durant son temps libre, notamment la théorie des probabilités. Il fut élu membre de la Royal Society en 1742. Son ouvrage le plus célèbre, *Essay towards solving a problem in the doctrine of chances* [4] fut publié, de manière posthume, en 1764.

ce qui donne

$$\mathbf{P}(\mathrm{C}\,|\,\mathrm{M}) = \frac{\mathbf{P}(\mathrm{C})\cdot\mathbf{P}(\mathrm{M}\,|\,\mathrm{C})}{\mathbf{P}(\mathrm{M}\,|\,\mathrm{C})\,\mathbf{P}(\mathrm{C}) + \mathbf{P}(\mathrm{M}\,|\,\mathrm{C}^c)\,\mathbf{P}(\mathrm{C}^c)}.$$

L'application numérique donne

$$\mathbf{P}(\mathrm{C}\,|\,\mathrm{M}) = \frac{0,01\cdot 0,9}{0,01\cdot 0,9 + 0,2\cdot 0,99} \approx 0,043 \approx 4,3\,\%.$$

La grand-tante peut s'endormir tranquille.

Remarque 4.12 (Sur certains types d'énoncés) On trouve çà et là des énoncés d'exercices censés illustrer le théorème de Bayes, donc voici un exemple.

> *On a deux tiroirs pleins de chaussettes. Le premier contient 90 chaussettes noires et 10 blanches, le second 10 noires et 90 blanches. On choisit au hasard un des tiroirs. On tire (sans regarder), on obtient une chaussette noire. On demande la probabilité p que ce soit dans le premier tiroir*
> — *avant d'avoir tiré la chaussette ;*
> — *après avoir tiré la chaussette noire.*

Ce genre d'énoncés est propre à entretenir ou à créer des confusions (sans parler de la situation grotesque qu'ils décrivent). Les probabilités *changent-elles* lorsque l'on obtient un renseignement supplémentaire ? Il nous semble voir là une confusion entre *probabilité mathématique* et *crédit* que l'on peut apporter à une certaine hypothèse, crédit qui va évidemment augmenter si l'on tire une chaussette noire.

Effectuer une modélisation probabiliste est une chose, dire qu'elle va traduire la réalité en est une autre. Comment trancher ? Réponse : en effectuant un grand nombre d'expériences. Soit. On réunit une bande d'amis à qui on bande les yeux et à qui on fournit un meuble garni de chaussettes ainsi que décrit précédemment. Chacun tire au hasard un tiroir et une chaussette dans le tiroir (selon une probabilité uniforme, sans doute, mais n'aurait-il pas mieux valu le préciser dans l'énoncé ?).

Dans ce cas, quel est le sens de la deuxième question ? Que faire des amis qui ont tiré une chaussette blanche ? Ils recommencent à tirer une chaussette jusqu'à en obtenir une noire ? (On trouvera alors $p = 1/2$). On leur signifie que leur intervention n'est plus utile, on leur paye une bière et on leur demande de rentrer chez eux en oubliant tout ce qui s'est passé ? Dans ce cas, l'expérience a changé entre la première question (tous les amis étaient là) et la seconde (certains sont repartis après avoir bu leur bière). Chacune des deux questions est légitime, mais la succession des deux masque le fait que toute la modélisation est à refaire.

Comme on le voit, si l'on s'en tient à une interprétation statistique des probabilités en tant que modélisation de la réalité, il est indispensable de décider à quelle expérience précisément l'énoncé fait référence. Ce n'est que lorsque l'on aura répondu à cette question que l'on pourra écrire des formules (sans doute celle de Bayes) et donner la réponse attendue (sans doute 0, 9).

Le théorème de Bayes se généralise sous la forme suivante. On se donne un système complet d'événements $(\mathrm{A}_i)_{i\in\mathrm{I}}$, c'est-à-dire une partition de Ω :

$$\Omega = \sum_{i\in\mathrm{I}} \mathrm{A}_i.$$

Choisissons un indice p particulier ; le théorème précédent permet de mettre en relation les quantités $\mathbf{P}(\mathrm{A}_p\,|\,\mathrm{B})$ et $\mathbf{P}(\mathrm{B}\,|\,\mathrm{A}_p)$:

$$\mathbf{P}(\mathrm{A}_p\,|\,\mathrm{B}) = \frac{\mathbf{P}(\mathrm{A}_p)\cdot\mathbf{P}(\mathrm{B}\,|\,\mathrm{A}_p)}{\mathbf{P}(\mathrm{B})}.$$

En écrivant B sous la forme $\mathrm{B} = \bigcup_i \mathrm{B}\cap\mathrm{A}_i$, on obtient :

■ **Théorème 4.13 (Théorème de Bayes, v.2)** *Soient $(\mathrm{A}_i)_{i\in\mathrm{I}}$ un système complet d'événements et* B *un événement. On suppose que $\mathbf{P}(\mathrm{A}_i) > 0$ pour tout $i \in \mathrm{I}$ et $\mathbf{P}(\mathrm{B}) > 0$; alors*

$$\forall p \in \mathrm{I} \qquad \mathbf{P}(\mathrm{A}_p\,|\,\mathrm{B}) = \frac{\mathbf{P}(\mathrm{A}_p)\cdot\mathbf{P}(\mathrm{B}\,|\,\mathrm{A}_p)}{\sum_{i\in\mathrm{I}} \mathbf{P}(\mathrm{A}_i)\cdot\mathbf{P}(\mathrm{B}\,|\,\mathrm{A}_i)}.$$

§ 62 **Quelques remarques et difficultés**

Un exercice bien connu portant sur les probabilités conditionnelles est le suivant (le contexte et les noms peuvent évidemment changer).

> *Madame Durand et Madame Dupont discutent, dans le train, de leur progéniture respective. Madame Durand dit : « J'ai deux enfants, attendez, je dois avoir des photos... ». Elle cherche, mais ne trouve malheureusement qu'une seule photo : « C'est ma petite Solveig, en train de jouer du hautbois. »*
>
> *On demande la probabilité p que Solveig ait un frère.*

Le problème est classique et a fait couler beaucoup d'encre — il n'y a d'ailleurs pas encore de consensus. En effet, un véritable problème d'interprétation se pose. Voici deux raisonnements menant à des conclusions incompatibles ($p = 2/3$ et $p = 1/2$ respectivement).

- Si l'on considère que l'événement A : « la photo trouvée est celle d'une fille » est équivalent à l'événement A′ : « Madame Durand a (au moins) une fille », alors la probabilité que Solveig ait un frère est de 2/3. En effet, sur toutes les familles de deux enfants, environ un quart ont deux garçons, un quart ont deux filles et, par conséquent, la moitié ont un garçon et une fille. Or, on sait qu'il n'y a pas deux garçons. Les différentes possibilités de composition d'une famille (en notant successivement le sexe de l'aîné et celui du cadet) sont les suivantes — et elles sont équiprobables :

FF	FG
GF	—

  La probabilité d'avoir deux filles est donc de 1/4 divisé par $1/2 + 1/4$, soit $1/3$[(5)] et la probabilité que Solveig ait un frère est $p = 2/3$.
- Cependant, il n'est pas évident que les événements A et A′ soient équivalents ! Décrivons l'univers lié à l'expérience comme l'ensemble des triplets (S_1, S_2, S_3), où S_1 est le sexe du premier enfant, S_2 celui du second, et S_3 celui de l'enfant dont une photo a été piochée. Ainsi,

$$\Omega = \{\mathrm{FFF}, \mathrm{FGF}, \mathrm{FGG}, \mathrm{GFF}, \mathrm{GFG}, \mathrm{GGG}\}.$$

  On calcule alors les probabilités élémentaires suivantes (en arguant que la photo tirée est prise au hasard entre les deux enfants) :

$$\mathbf{P}(\mathrm{FFF}) = \mathbf{P}(\mathrm{GGG}) = \tfrac{1}{4} \qquad \text{et} \qquad \mathbf{P}(\mathrm{FGF}) = \mathbf{P}(\mathrm{FGG}) = \mathbf{P}(\mathrm{GFF}) = \mathbf{P}(\mathrm{GFG}) = \tfrac{1}{8}.$$

  L'événement « la photo représente une fille » est donc $\mathrm{A} = \{\mathrm{FFF}, \mathrm{FGF}, \mathrm{GFF}\}$, de probabilité 1/2, et la probabilité conditionnelle recherchée est alors

$$p = \frac{\mathbf{P}(\{\mathrm{FGF}, \mathrm{GFF}\})}{\mathbf{P}(\mathrm{A})} = \frac{1/8 + 1/8}{1/2} = 1/2.$$

  On peut représenter sous forme graphique les familles de deux enfants (4 cases principales de probabilité 1/4 chacune), avec une subdivision selon que la photo montrée est celle de l'aîné ou du cadet (chaque petite case a alors une probabilité 1/8). L'événement A (« la photo montrée est celle d'une fille ») est représenté en blanc, son complémentaire en gris.

FF (F)	FF (F)	FG (F)	FG (G)
GF (G)	GF (F)	GG (G)	GG (G)

  Au sein des cases blanches (ce qui traduit « sachant A »), la moitié des filles photographiées a effectivement un frère.

(5). Le problème eût été différent si Madame Durand avait différencié ses enfants (en disant par exemple : « Voici mon aînée au piano ») : la probabilité cherchée aurait été de 1/2.

Voilà donc deux raisonnements intéressants — et contradictoires. Une expérience permettrait-elle de trancher ? D'affirmer que $p = 2/3$ ou que $p = 1/2$?

La réponse, évidemment, est plus subtile : tout dépend des modalités de l'expérience. Selon la manière dont on va interpréter le « sachant que Mme Durand montre une photo de fille », les modalités de l'expérience vont différer.

— Supposons que l'on réunisse un grand nombre de familles de deux enfants, et que l'on ne garde finalement que celles qui ont au moins une fille (les autres familles rentrent chez elles). On demande à toutes ces familles de montrer une photo de fille. Les deux tiers des familles affirmeront que leur autre enfant est un garçon : $p = 2/3$.

— Supposons que l'on réunisse un grand nombre de familles de deux enfants, à qui l'on demande de choisir une photo d'un de leurs enfants (en leur faisant tirer à *pile* ou *face* pour savoir s'ils prennent celle de l'aîné ou celle du cadet), et que l'on ne garde finalement que celles qui ont une photo de fille (les autres familles rentrent chez elles). On tente alors l'expérience. La moitié des familles affirmera que leur autre enfant est un garçon : $p = 1/2$.

— Mais on peut également imaginer que, statistiquement, si elles n'ont la photo que d'un seul enfant, et qu'elles ont un garçon et une fille, les mamans ont, avec une probabilité q, une photo de leur fille plutôt que de leur garçon. Sous ces modalités, lorsqu'une photo de garçon est présentée, la famille est éliminée. On peut alors calculer la proportion p de familles ayant un garçon en plus de la fille montrée par la formule de Bayes.
Pour cela, notons

F_1 :	« l'aînée est une fille »	G_1 :	« l'aîné est un garçon »
F_2 :	« la cadette est une fille »	G_2 :	« le cadet est un garçon »
F^* :	« la photo montre une fille »	G^* :	« la photo montre un garçon ».

La probabilité cherchée est $p = \mathbf{P}(G_1 \cup G_2 \mid F^*)$. Puisque les événements G_1 et G_2 ne sont pas disjoints, on effectue une partition de leur union :

$$G_1 \cup G_2 = G_1F_2 + G_1G_2 + F_1G_2,$$

ce qui permet d'écrire que

$$p = \mathbf{P}(G_1 \cup G_2 \mid F^*) = \mathbf{P}(G_1F_2 + G_1G_2 + F_1G_2 \mid F^*)$$
$$= \frac{\mathbf{P}(G_1F_2F^*) + \mathbf{P}(G_1G_2F^*) + \mathbf{P}(F_1G_2F^*)}{\mathbf{P}(F^*)}.$$

Au numérateur, on a

$$\mathbf{P}(G_1F_2F^*) = \mathbf{P}(F^* \mid G_1F_2)\,\mathbf{P}(G_1F_2) = q \cdot \frac{1}{4} = \frac{q}{4},$$

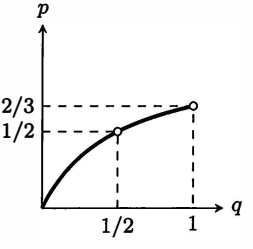
où, rappelons-le, q est la probabilité de montrer la photo d'une fille lorsque les deux enfants sont un garçon et une fille ; de même $\mathbf{P}(F_1G_2F^*) = q/4$. Le terme $\mathbf{P}(G_1G_2F^*)$ est évidemment nul.
Quant au dénominateur, on le développe en

$$\mathbf{P}(F^*) = \mathbf{P}(F^* \mid F_1F_2)\,\mathbf{P}(F_1F_2) + \mathbf{P}(F^* \mid F_1G_2)\,\mathbf{P}(F_1G_2) + \mathbf{P}(F^* \mid G_1F_2)\,\mathbf{P}(G_1F_2)$$
$$= \frac{1}{4} + \frac{q}{4} + \frac{q}{4}.$$

On trouve alors

$$p = \frac{q/2}{q/2 + 1/4} = \frac{2q}{2q+1}.$$

Le premier raisonnement supposait implicitement $q = 1$ et mène donc à $p = 2/3$. Dans le second raisonnement, on supposait implicitement $q = 1/2$, ce qui mène à $p = 1/2$. Si l'on suppose que $q = 0$ (les mamans sont plus fières d'avoir un garçon et montrent systématiquement un garçon lorsqu'elles en ont un), alors $p = 0$: si la photo montrée est celle d'une fille, c'est qu'aucun des enfants n'est un garçon !



EXERCICES

♦ ***Exercice 4.2*** Soient A et B deux événements non négligeables. Montrer que

$$\mathbf{P}(\mathrm{A} \mid \mathrm{B}) \geqslant \mathbf{P}(\mathrm{A}) \iff \mathbf{P}(\mathrm{B} \mid \mathrm{A}) \geqslant \mathbf{P}(\mathrm{B}).$$

♦ ***Exercice 4.3*** On considère un dé blanc équilibré, et un dé noir pipé, de façon qu'il sorte le « 6 » avec une probabilité 1/3, les autres numéros sortant avec la même probabilité 2/15. Deux joueurs s'affrontent. Le premier joueur choisit le dé qu'il veut et le lance. Le second prend le dé restant et le lance. Le gagnant est celui qui obtient le plus haut score ; en cas d'égalité, le lanceur du dé blanc est réputé vainqueur.

Quelle stratégie adopter pour le premier joueur : choisir le dé blanc, ou le dé noir ?

♦ ***Exercice 4.4*** Un étudiant doit répondre à une question à choix multiple où cinq réponses sont proposées, une seule étant correcte. Quand l'événement A : « l'étudiant a bien travaillé » est réalisé, la réponse fournie est la bonne réponse. Dans le cas contraire l'étudiant répond au hasard. Si l'événement B : « la réponse fournie est correcte » est réalisé, calculer la probabilité $\mathbf{P}(\mathrm{A} \mid \mathrm{B})$ en fonction de $p = \mathbf{P}(\mathrm{A})$.

♦ ***Exercice 4.5 (Une aventure de Renart)*** Le roi Noble, le lion, décide d'inviter tous les mois ses sujets Ysengrin, Tibert et Renart, à un grand banquet ; ils peuvent manger tant qu'ils veulent, mais l'étiquette exige qu'ils n'emportent rien chez eux. Tibert et Ysengrin ont chacun une probabilité p de chaparder un peu de nourriture à un banquet, indépendamment des autres fois et des autres convives. Renart, qui est plus fourbe que les autres, chapardera quant à lui avec une probabilité $p' > p$.

Noble fait discrètement surveiller le banquet par ses gardes (qui sont efficaces et repèrent toujours un voleur quand il agit).

i) Les gardes choisissent de se concentrer sur un seul des trois convives au hasard. Quelle est la probabilité q qu'ils observent un vol ?

ii) Les gardes annoncent discrètement à Noble qu'ils ont surpris un voleur ; quelle est la probabilité que ce soit Renart ?

iii) Noble décide de ne rien dire pour cette fois, et d'attendre le banquet suivant ; il recommande aux gardes de surveiller le *même* convive. Quelle est la probabilité q' qu'ils observent un vol ? Comparer q' à q.

Dans les applications numériques, on prendra $p = 0{,}1$ et $p' = 0{,}8$.

Indications

◊ *Indication 4.3 :* Calculer la probabilité que Noir gagne sachant que Blanc sort 1, puis 2, etc. En déduire la probabilité que Noir gagne.

◊ *Indication 4.5 :* Dans la troisième question, la probabilité demandée est une probabilité conditionnée par le fait qu'un premier vol a été observé ! Or, si un vol a été observé, il y a pas mal de chances que le convive soit Renart... et qu'il recommence au nouveau banquet.

SOLUTIONS

♦ **Solution 4.1** On trouve respectivement

$$\mathbf{P}\{\text{blanc} \mid \text{pair}\} = \tfrac{1}{2} \qquad \mathbf{P}\{\text{blanc} \mid \text{impair}\} = \tfrac{1}{3}$$
$$\mathbf{P}\{\text{noir} \mid \text{pair}\} = \tfrac{1}{2} \qquad \mathbf{P}\{\text{noir} \mid \text{impair}\} = \tfrac{2}{3}.$$

♦ **Solution 4.2** Chacune de ces inégalités est équivalente à $\mathbf{P}(A \cap B) \geqslant \mathbf{P}(A)\,\mathbf{P}(B)$.

♦ **Solution 4.3** Conditionnellement au fait que Blanc sorte 1, Noir gagne s'il sort un nombre entre 2 et 6, donc

$$\mathbf{P}\{\text{Noir gagne} \mid \text{Blanc} = 1\} = \frac{13}{15}.$$

De même, si Blanc sort 2, Noir gagne s'il sort un nombre entre 3 et 6,

$$\mathbf{P}\{\text{Noir gagne} \mid \text{Blanc} = 2\} = \frac{11}{15}$$

et ainsi de suite jusqu'à

$$\mathbf{P}\{\text{Noir gagne} \mid \text{Blanc} = 5\} = \frac{5}{15}.$$

Bien entendu, si Blanc sort 6, il gagne. Enfin, la probabilité que Noir gagne est

$$\begin{aligned}\mathbf{P}\{\text{Noir gagne}\} &= \sum_{k=1}^{5} \mathbf{P}\{\text{Noir gagne} \mid \text{Blanc} = k\}\ \mathbf{P}\{\text{Blanc} = k\} \\ &= \frac{1}{6}\left(\frac{13}{15} + \frac{11}{15} + \frac{9}{15} + \frac{7}{15} + \frac{5}{15}\right) = \frac{1}{2}.\end{aligned}$$

Finalement, choisir le dé noir ou le dé blanc est indifférent.

♦ **Solution 4.4** On applique la formule de Bayes, en écrivant que

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(B \mid A)\,\mathbf{P}(A)}{\mathbf{P}(B \mid A)\,\mathbf{P}(A) + \mathbf{P}(B \mid A^c)\,\mathbf{P}(A^c)} = \frac{p}{p + \frac{1}{5}(1-p)} = \frac{5p}{4p+1}.$$

♦ **Solution 4.5** On note Y, T et R les événements « le convive surveillé est Ysengrin » (respectivement Tibert, Renart). On note V l'événement « les gardes observent un premier vol », et W l'événement « les gardes observent un deuxième vol ».

i) Par la formule des probabilité totales,

$$q = \mathbf{P}(V) = \mathbf{P}(V \mid Y)\,\mathbf{P}(Y) + \mathbf{P}(V \mid T)\,\mathbf{P}(T) + \mathbf{P}(V \mid R)\,\mathbf{P}(R) = \frac{p}{3} + \frac{p}{3} + \frac{p'}{3} = \frac{2p+p'}{3}.$$

Application numérique : $q \approx 0{,}333$.

ii) Par la formule de Bayes,

$$\mathbf{P}(R \mid V) = \frac{\mathbf{P}(R \cap V)}{\mathbf{P}(V)} = \frac{\mathbf{P}(V \mid R)\,\mathbf{P}(R)}{\mathbf{P}(V)} = \frac{p'/3}{q} = \frac{p'}{2p+p'} = \frac{8}{10} = 0{,}8.$$

iii) Le nombre q' cherché n'est pas $\mathbf{P}(W)$, mais $\mathbf{P}(W \mid V)$, probabilité d'observer un second vol *sachant qu'on en a observé un premier.* Du coup, ce nombre n'est pas égal à q, car il y a plus de chances que ce soit Renart qui soit observé, plutôt que Tibert ou Ysengrin. De fait, par la formule des probabilités totales,

$$\mathbf{P}(W \mid V) = \frac{\mathbf{P}(W \cap V)}{\mathbf{P}(V)} = \frac{\mathbf{P}(W \cap V \mid Y)\,\mathbf{P}(Y) + \mathbf{P}(W \cap V \mid T)\,\mathbf{P}(T) + \mathbf{P}(W \cap V \mid R)\,\mathbf{P}(R)}{\mathbf{P}(V)}.$$

On nous dit que chacune des convives va chaparder avec une certaine probabilité, indépendamment de ce qu'il aura fait les autres fois. Ainsi, $\mathbf{P}(W \cap V \mid T) = \mathbf{P}(W \cap V \mid Y) = p^2$ tandis que $\mathbf{P}(W \cap V \mid R) = p'^2$, et

$$q' = \mathbf{P}(W \mid V) = \frac{2p^2 + p'^2}{2p+p'}.$$

Application numérique : $q' = \frac{0{,}02+0{,}64}{2\cdot 0{,}1+0{,}8} = 0{,}66$. De manière générale, si $p' > p$, on trouve que $q' > q$.

Chapitre 5

Indépendance (1/2)

Les chiffres sont dénués de toute imagination, tu devrais donc te garder de leur accorder trop d'importance.

Jón Kalman Stefánsson, *Entre ciel et terre*

D'une certaine manière, la notion d'indépendance est la *notion qui fait que la théorie des probabilités ne se réduit pas à un cas particulier de la théorie générale des espaces mesurés, et a développé, au cours de plus de trois siècles d'existence, ses méthodes propres.*

5.1 INDÉPENDANCE D'ÉVÉNEMENTS ET DE TRIBUS

Indépendance de deux événements

§ 63

D'un point de vue intuitif, si un événement A semble n'avoir pas d'effet, du point de vue probabiliste, sur un événement B, alors on s'attend à ce que le calcul de la probabilité de B, conditionné par la connaissance de la réalisation (ou non) de l'événement A, conduise au même résultat que le calcul absolu de cette même probabilité, c'est-à-dire

$$\mathbf{P}(\mathrm{B} \mid \mathrm{A}) = \mathbf{P}(\mathrm{B} \mid \mathrm{A}^{\mathrm{c}}) = \mathbf{P}(\mathrm{B}).$$

Multipliant cette égalité par la probabilité de A, on obtient $\mathbf{P}(\mathrm{B} \cap \mathrm{A}) = \mathbf{P}(\mathrm{B})\,\mathbf{P}(\mathrm{A})$; cette dernière forme est sans doute meilleure que la première, car elle est symétrique et met sur le même plan les événements A et B. De plus, elle ne fait pas appel à une probabilité conditionnelle pour laquelle le cas $\mathbf{P}(\mathrm{A}) = 0$ est problématique.

Exemple 5.1 (Cas fini : nombre de cas favorables sur nombre de cas total) Revenons au cas, extrêmement simple, mais fondateur du raisonnement sur les probabilités, où Ω est un ensemble fini de cardinal N, muni d'une probabilité uniforme. À un événement A de cardinal k, on associe la probabilité $\mathbf{P}(\mathrm{A}) = k/\mathrm{N}$ (nombre de cas favorables sur nombre de cas total).

Dire que les événements A et B sont indépendants revient à dire que la connaissance de la réalisation (ou non) de A n'influe pas sur la probabilité de B, conditionnée par cette réalisation : en d'autres termes, que les rapports

$$\frac{\text{nombre de cas où B est réalisé en même temps que A}}{\text{nombre de cas où A est réalisé}}$$

et

$$\frac{\text{nombre de cas où B est réalisé, mais pas A}}{\text{nombre de cas où A n'est pas réalisé}}$$

sont égaux, c'est-à-dire

$$\frac{\mathrm{Card}(A \cap B)}{\mathrm{Card}(A)} = \frac{\mathrm{Card}(A^c \cap B)}{\mathrm{Card}(A^c)},$$

ou encore, en écrivant que $A^c = \Omega - A$ et en divisant numérateur et dénominateur par N :

$$\frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = \frac{\mathbf{P}(B) - \mathbf{P}(B \cap A)}{1 - \mathbf{P}(A)}.$$

Multiplions par les dénominateurs, on obtient

$$\mathbf{P}(A \cap B) - \mathbf{P}(A \cap B)\,\mathbf{P}(A) = \mathbf{P}(A)\,\mathbf{P}(B) - \mathbf{P}(A \cap B)\,\mathbf{P}(A),$$

et donc $\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B)$.

Définition 5.2 Deux événements A et B d'un espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ sont **indépendants** (ou : **stochastiquement indépendants**) s'ils vérifient

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B).$$

Remarque 5.3 L'indépendance de deux événements est une notion qui dépend de la probabilité **P** dont on a muni l'espace $(\Omega, \mathfrak{T})$, au contraire de l'incompatibilité, qui est une propriété ne dépendant que de Ω.

Remarque 5.4 L'indépendance de deux événements est une propriété *numérique* ; il ne s'agit que de comparer deux nombres et vérifier s'ils sont égaux. Notamment, cette notion n'est pas à confondre avec celle de causalité (A influe sur B) ni avec celle de corrélation. Des contre-exemples seront donnés plus loin (voir § 119 pour la corrélation et § 332 pour la causalité).

■ **Proposition 5.5** *Soient* A *et* B *deux événements de* $(\Omega, \mathfrak{T}, \mathbf{P})$.

i) Si A *est un événement de probabilité non nulle, alors* A *et* B *sont indépendants si et seulement si* $\mathbf{P}(B \mid A) = \mathbf{P}(B)$.

ii) Les événements Ω *et* $\varnothing$ *sont indépendants de tout autre événement.*

iii) A *est indépendant de lui-même si et seulement si il est de probabilité* 0 *ou* 1.

iv) Si A *et* B *sont indépendants, alors* A *et* B^c *le sont, ainsi bien sûr que* A^c *et* B, *et que* A^c *et* B^c.

v) Si A *est négligeable ou presque certain, alors* A *est indépendant de* B.

DÉMONSTRATION : Les deux premières propriétés sont immédiates. Pour la troisième, A est indépendant de lui-même si et seulement si $\mathbf{P}(A \cap A) = \mathbf{P}(A)^2$, c'est-à-dire $\mathbf{P}(A) = \mathbf{P}(A)^2$, ce qui équivaut à $\mathbf{P}(A) \in \{0, 1\}$. Pour la quatrième propriété, supposons A et B indépendants, alors

$$\mathbf{P}(A \cap B^c) = \mathbf{P}\big(A \cap (\Omega - B)\big) = \mathbf{P}\big(A - (A \cap B)\big) = \mathbf{P}(A) - \mathbf{P}(A \cap B)$$
$$= \mathbf{P}(A) - \mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(A) \cdot \big[1 - \mathbf{P}(B)\big] = \mathbf{P}(A) \cdot \mathbf{P}(B^c).$$

Enfin, pour la dernière propriété, il suffit de remarquer que, si $\mathbf{P}(A) = 0$, alors $0 \leqslant \mathbf{P}(A \cap B) \leqslant \mathbf{P}(A)$ donc $\mathbf{P}(A \cap B) = 0 = \mathbf{P}(A)\,\mathbf{P}(B)$. ∎

§ 64 Indépendance mutuelle

Si A, B, C sont des événements deux à deux indépendants, peut-on affirmer que $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\,\mathbf{P}(B)\,\mathbf{P}(C)$? Un simple exemple montre la négative :

Exemple 5.6 Considérons un lancer de deux dés et les événements

• **A** : la somme des deux dés est paire ; • **B** : le premier dé est pair ; • **C** : le deuxième dé est pair.

Ces événements sont bien deux à deux indépendants, puisque

$$\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = \tfrac{1}{2} \qquad \text{et} \qquad \mathbf{P}(A \cap B) = \mathbf{P}(B \cap C) = \mathbf{P}(C \cap A) = \tfrac{1}{4}.$$

En revanche, $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(B \cap C) = \frac{1}{4}$ alors que $\mathbf{P}(A)\,\mathbf{P}(B)\,\mathbf{P}(C) = \frac{1}{8}$.

On doit introduire une définition plus restrictive pour caractériser l'indépendance de plusieurs événements.

Définition 5.7 Des événements $A_1, A_2, \dots, A_n$ sont dit **(mutuellement) indépendants**[1] si, pour tout entier $p \leqslant n$ et pour tout p-uplet d'indices $(i_1, \dots, i_p)$ deux à deux distincts, on a

$$\mathbf{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_p}) = \mathbf{P}(A_{i_1})\,\mathbf{P}(A_{i_1}) \cdots \mathbf{P}(A_{i_p}). \tag{5.1}$$

On peut également adopter le point de vue suivant : on définit des événements $B_1, B_2, \dots, B_n$ tels que pour chaque indice k, on ait $B_k = A_k$ ou $B_k = \Omega$. Les événements $A_1, A_2, \dots, A_n$ sont alors indépendants si et seulement si, pour tout choix des B_k,

$$\mathbf{P}(B_1 \cap B_2 \cap \cdots \cap B_n) = \mathbf{P}(B_1)\,\mathbf{P}(B_2) \cdots \mathbf{P}(B_n).$$

Plus généralement, soit I un ensemble, fini ou infini, d'indices, et soit $(A_i)_{i\in I}$ une famille d'événements. On dit que ces événements sont **mutuellement indépendants** si, pour toute partie *finie* $J \subset I$ non vide, on a

$$\mathbf{P}\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} \mathbf{P}(A_j).$$

Remarque 5.8 Lorsque $n = 2$, on retrouve évidemment la définition 5.2. Lorsque $n = 3$, les événements A, B et C sont mutuellement indépendants si et seulement si l'on a les quatre relations

$$\begin{aligned} &\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B \\ &\mathbf{P}(A \cap C) = \mathbf{P}(A)\,\mathbf{P}(C) \qquad\qquad \mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\,\mathbf{P}(B)\,\mathbf{P}(C). \\ &\mathbf{P}(B \cap C) = \mathbf{P}(B)\,\mathbf{P}(C) \end{aligned}$$

Dans le cas d'une famille finie de cardinal n, il y a $2^n - n - 1$ relations à vérifier : il y a en effet 2^n choix possibles de parties de $\{1, 2, \dots, n\}$, mais les parties à un élément (n choix) et la partie vide sont exclues de l'étude.

Le lecteur vérifiera que l'indépendance mutuelle implique l'indépendance deux à deux ; nous avons vu que la réciproque n'est pas nécessairement vraie. Enfin, la propriété : si A et B sont indépendants alors A et B^c le sont, se généralise à une famille d'événements : ⇗5.2

■ **Proposition 5.9** *Soit $(A_i)_{i\in I}$ une famille quelconque d'événements mutuellement indépendants. On considère une application $\varepsilon : I \to \{-1, +1\}$, et on pose*

$$B_i = \begin{cases} A_i & si\ \varepsilon_i = +1 \\ A_i^c & si\ \varepsilon_i = -1. \end{cases}$$

Alors $(B_i)_{i\in I}$ est une famille d'événements mutuellement indépendants.

(1). On rencontre parfois la dénomination **indépendants dans leur ensemble**, que nous n'utiliserons pas car elle prête à confusion.

5.2 LES LEMMES DE BOREL-CANTELLI (*)

Cette section est consacrée à un outil, extrêmement utile, permettant de vérifier si une suite $A_1, A_2, \dots, A_n, \dots$ d'événements arrive « infiniment souvent ». Nous aurons donc besoin de définir l'événement

$$\{A_n \ ; \ \text{i.s.}\}$$

signifiant qu'il existe une famille infinie d'indices k tels que A_k est réalisé ; cet événement est également appelé « limite supérieure » des A_n, et noté A^* ou $\limsup A_n$.

§ 65 **Limite supérieure et limite inférieure**

On considère une suite $(A_n)_{n\geqslant 0}$ d'événements ; on en définit la **limite supérieure** comme l'ensemble A^* des ω qui appartiennent à un nombre *infini* de A_n.

Détaillons cette notion. Dire qu'un élément ω n'appartient qu'à un nombre *fini* de A_n, c'est dire qu'il existe un rang N tel que ω n'appartient à aucun des ensembles $A_N, A_{N+1}, \dots$. À l'inverse, dire qu'il appartient à un nombre infini de A_n, c'est dire que, quel que soit l'entier N choisi, il fera partie d'un (au moins) des ensembles $A_N, A_{N+1}, \dots$, ou encore qu'il est élément de leur union $U_N = \bigcup_{n=N}^{\infty} A_n$; puisque ceci doit être vrai pour tout $N \geqslant 0$, cela revient à dire que ω fait partie de l'intersection $\bigcap_{N=0}^{\infty} U_N$. Ainsi, on pose

$$A^* = \bigcap_{N=0}^{\infty} \bigcup_{n=N}^{\infty} A_n,$$

ou encore, puisque la suite de terme général $U_N = \bigcup_{n=N}^{\infty} A_n = \sup_{n\geqslant N} A_N$ est décroissante,

$$A^* = \lim_N \downarrow \sup_{n\geqslant N} A_n.$$

Cela explique la notation usuelle

$$A^* = \limsup_{n\to\infty} A_n.$$

La limite supérieure d'une suite d'événements intervient dans de très nombreux problèmes[(2)]. Il est d'usage, dans les ouvrages anglo-saxons, de la noter $\{A_n ; \text{i.o.}\}$ pour *infinitely often*, ce qui est très pratique, mais n'est pas (encore) très fréquent dans les ouvrages français. Nous noterons $A^* = \{A_n \ ; \ \text{i.s.}\}$ pour « infiniment souvent » :

$$\{A_n \ ; \ \text{i.s.}\} = A^* = \limsup_{n\to\infty} A_n = \bigcap_{N=0}^{\infty} \bigcup_{n=N}^{\infty} A_n = \lim_N \downarrow \sup_{n\geqslant N} A_n.$$

Exemple 5.10 Pour tout $n \in \mathbb{N}^*$, on note I_n l'intervalle

$$I_n := \left[0 \, ; \, \tfrac{5}{4} + (-1)^{n+1} \left(\tfrac{1}{4} + \tfrac{2}{n}\right)\right] = \begin{cases} \left[0 \, ; \, 1 - \frac{2}{n}\right] & \text{si } n \text{ est pair,} \\ \left[0 \, ; \, \frac{3}{2} + \frac{2}{n}\right] & \text{si } n \text{ est impair.} \end{cases}$$

Alors pour tout entier N,

$$U_N = \bigcup_{n=N}^{\infty} I_n = \begin{cases} \left[0 \, ; \, \frac{3}{2} + \frac{2}{N}\right] & \text{si N est impair,} \\ \left[0 \, ; \, \frac{3}{2} + \frac{2}{N+1}\right] & \text{si N est pair.} \end{cases}$$

La suite $(U_N)_{N\geqslant 1}$ est bien décroissante, comme on le voit sur la figure 5.1 *(à gauche)*.

(2). Une question typique dans l'étude d'une marche aléatoire peut être : le marcheur va-t-il revenir à son point de départ et, si oui, reviendra-t-il plutôt un nombre fini de fois, ou bien une infinité de fois ?

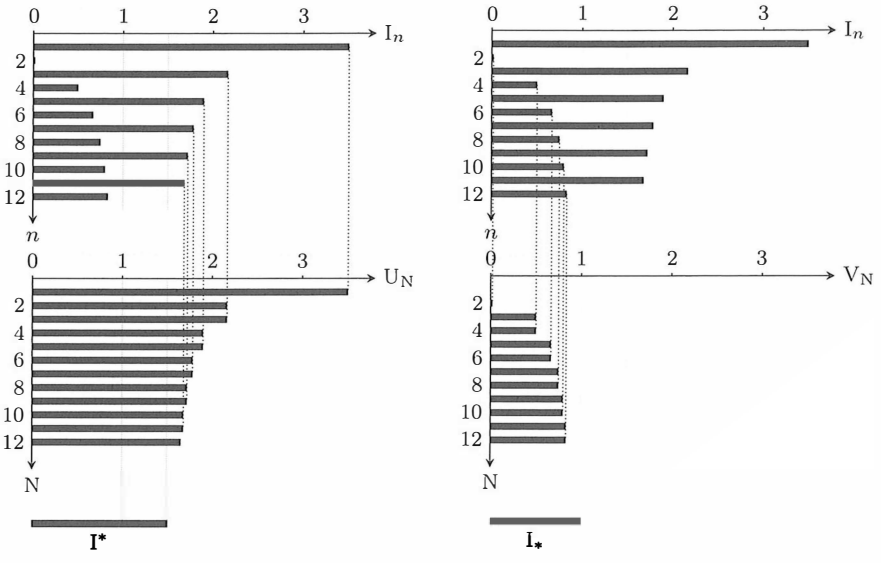


FIGURE 5.1 — (À gauche) : *les intervalles* I_n *et* U_N, *pour* $n, N = 1, \ldots, 12$.
(À droite) : *les intervalles* I_n *et* V_N.

Ainsi,

$$I^* = \{I_n \ ; \ \text{i.s.}\} = \lim_N \downarrow U_N = \left[0\,; \frac{3}{2}\right].$$

De même, la **limite inférieure** de $(A_n)_{n\geqslant 0}$ est l'ensemble A_* des éléments ω qui appartiennent à tous les A_n à partir d'un certain rang $N^{(3)}$; c'est donc également l'ensemble des ω appartiennent à tous les A_n, à l'exception éventuelle d'un nombre fini d'entre eux. Du point de vue ensembliste, on peut l'écrire sous la forme

$$A_* = \liminf_{n\to\infty} A_n = \bigcup_{N=0}^{\infty} \bigcap_{n=N}^{\infty} A_n.$$

Noter que la suite de terme général $\inf_{n\geqslant N} A_n = \bigcap_{n=N}^{\infty} A_n$ est croissante, ce qui fait que l'on peut écrire

$$\liminf_{n\to\infty} A_n = \lim_N \uparrow \inf_{n\geqslant N} A_n$$

et introduire la notation $\{A_n \ ; \ \text{à.p.c.r.}\}$ (« à partir d'un certain rang ») pour désigner $\liminf A_n$.

Exemple 5.11 Pour tout $n \in \mathbb{N}^*$, on note encore I_n l'intervalle $\left[0\,; \frac{3}{2} + (-1)^{n+1}\left(\frac{1}{2} + \frac{1}{n}\right)\right]$. Alors pour tout entier N, notons

$$V_N = \bigcap_{n=N}^{\infty} I_n = \begin{cases} \left[0\,; 1 - \frac{2}{N+1}\right] & \text{si N est impair,} \\ \left[0\,; 1 - \frac{2}{N}\right] & \text{si N est pair.} \end{cases}$$

La suite $(V_N)_{N\geqslant 1}$ est bien croissante, comme on le voit sur la figure 5.1 *(à droite)*. Ainsi,

$$I_* = \lim_N \uparrow V_N = [0\,; 1].$$

On remarquera que, sur cet exemple, I_* est strictement inclus dans I^*.

(3). Ce rang pouvant dépendre de ω.

Les propriétés suivantes sont laissées en exercice au lecteur (exercice 5.4 page 124).

■ **Proposition 5.12** *Soit $(A_n)_{n\geqslant 0}$ une suite d'événements. On note $\mathbf{1}_{A_n}$ la fonction indicatrice de A_n, c'est-à-dire la fonction définie par $\mathbf{1}_{A_n}(\omega) = 1$ si $\omega \in A_n$ et $\mathbf{1}_{A_n}(\omega) = 0$ si $\omega \notin A_n$.*

i) $\liminf A_n \subset \limsup A_n$.

ii) $\liminf A_n = (\limsup A_n^c)^c$.

iii) Si $(A_n)_{n\geqslant 0}$ est croissante, alors $\liminf_n A_n = \limsup_n A_n = \lim_n \uparrow A_n$.

iv) Si $(A_n)_{n\geqslant 0}$ est décroissante, alors $\liminf_n A_n = \limsup_n A_n = \lim_n \downarrow A_n$.

v) $\mathbf{1}_{A_*} = \liminf \mathbf{1}_{A_n}$ *et* $\mathbf{1}_{A^*} = \limsup \mathbf{1}_{A_n}$.

Cette dernière propriété justifie les dénominations de « limite supérieure » et « limite inférieure ».

Enfin, pour les applications probabilistes, il est indispensable de remarquer que A^* et A_* sont des événements :

■ **Théorème 5.13** *Soit $\mathfrak{T}$ une tribu. Alors $\mathfrak{T}$ est stable par passage à la limite supérieure et à la limite inférieure : si $(A_n)_{n\geqslant 0}$ est une suite d'éléments de $\mathfrak{T}$, alors A_* et A^* sont également des éléments de $\mathfrak{T}$*

DÉMONSTRATION : Avec les notations précédentes, les éléments $U_N = \sup_{n\geqslant N} A_n$ sont dans la tribu (stabilité par union dénombrable), donc A^* l'est également (stabilité par intersection dénombrable). La démonstration est identique pour A_*. ∎

Si $\limsup A_n = \liminf A_n$, on appelle **limite** de la suite $(A_n)_{n\geqslant 0}$ cette valeur commune :

$$\lim_{n\to\infty} A_n = A_* = A^*.$$

On dit alors que $(A_n)_{n\geqslant 0}$ **converge** vers A.

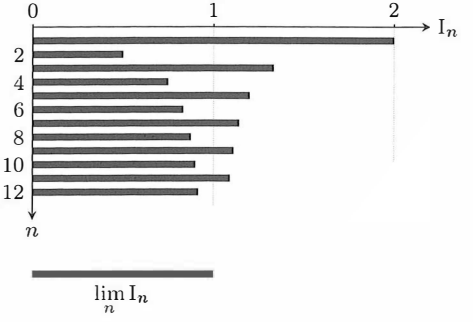
Exemple 5.14 Pour tout entier $n \geqslant 1$, on note $I_n = \left[0\,;\, 1 + \frac{(-1)^{n+1}}{n}\right]$. Alors pour tout entier N,

$$U_N = \bigcup_{n=N}^{\infty} I_n = \left[0\,;\, 1 + \frac{1}{N'}\right] \quad \text{où} \quad N' = \begin{cases} N+1 & \text{si N est pair,} \\ N & \text{si N est impair.} \end{cases}$$

On en déduit que $I^* = \{I_n \text{ ; i.s.}\} = [0\,;\,1]$. De même,

$$V_N = \bigcap_{n=N}^{\infty} I_n = \left[0\,;\, 1 - \frac{1}{N''}\right] \quad \text{où} \quad N'' = \begin{cases} N & \text{si N est pair,} \\ N+1 & \text{si N est impair.} \end{cases}$$

Ainsi, $I_* = [0\,;\,1] = I^*$: la suite $(I_n)_{n\geqslant 1}$ converge et $\lim_{n\to\infty} I_n = [0\,;\,1]$.



Les lemmes de Borel-Cantelli

§ 66

Les lemmes de Borel-Cantelli donnent un critère univoque pour la réalisation d'une infinité de A_n, selon la nature de la série $\sum \mathbf{P}(A_n)$. De manière un peu étonnante de prime abord, un sens est toujours vrai, alors que l'autre requiert l'indépendance mutuelle des événements. L'examen de la preuve permet d'éclaircir l'origine de cette différence.

■ **Lemme 5.15 (Lemmes de Borel-Cantelli)** *Soit* $(A_n)_{n\geqslant 0}$ *une suite (quelconque) d'événements.*

- *Si la série* $\sum \mathbf{P}(A_n)$ *converge, alors* $\mathbf{P}\{A_n \text{ ; } i.s.\} = 0$, *ou encore : presque sûrement, seul un nombre* fini *d'événements* A_n *se réalisent.*
- *Supposons les événements* A_n *sont deux à deux indépendants. Si la série* $\sum \mathbf{P}(A_n)$ *diverge, alors* $\mathbf{P}\{A_n \text{ ; } i.s.\} = 1$, *ou encore : presque sûrement, une infinité d'événements* A_n *se réalisent.*

Le premier résultat sera désigné comme le sens « facile », ou « trivial », le second étant le sens « difficile ».

Avant de donner la démonstration, commentons ce résultat, dans le cas où les événements A_k sont effectivement indépendants : alors l'événement $\{A_n \text{ ; i.s.}\}$ ne peut être que négligeable, ou presque sûr : il n'existe pas de cas intermédiaire. C'est un cas particulier de « loi du tout-ou-rien » (nous en verrons d'autres exemples dans le chapitre 20). Notamment, si l'on répète, de manière indépendante, une certaine expérience une infinité de fois — par exemple si l'on effectue une succession infinie de tirages à *pile* ou *face* — alors un événement « probable », c'est-à-dire de probabilité aussi petite que l'on veut, mais non nulle, sera réalisé (presque sûrement) une infinité de fois.

Exemple 5.16 (Peu réaliste... mais classique !) Donnez à un singe immortel une machine à écrire et du papier sans restriction ; alors non seulement il écrira les œuvres complètes de Shakespeare / Molière / Cervantès / Valmiki / Borges, mais il les écrira une infinité de fois.

Pour démontrer ce résultat, il convient d'abord de faire des hypothèses sur la procédure. On supposera donc que chaque pression d'une touche de la machine se fait de manière indépendante des autres, et que chaque touche a une probabilité non nulle d'être pressée. On note N la longueur totale du texte T que l'on a choisi. On considère alors A_k l'événement : « le texte compris entre les caractères $kN+1$ et $(k+1)N$ est exactement le texte T ». Les A_k sont alors indépendants, car leurs supports sont deux à deux disjoints[4], de probabilité non nulle (quoiqu'extrêmement faible) et constante, donc $\sum \mathbf{P}(A_k)$ diverge et $\mathbf{P}\{A_n \text{ ; i.s.}\} = 1$.

Cet exemple est taxé de « peu réaliste ». Le lecteur est invité, s'il ne l'a jamais fait, à estimer le temps nécessaire à un singe, frappant le clavier d'une machine à écrire à 36 touches (26 lettres et 10 signes de ponctuation usuels) de manière uniforme, pour obtenir *Les Fourberies de Scapin* (92 048 caractères), et à le comparer à l'âge de l'Univers (estimé aujourd'hui à 10 ou 15 milliards d'années).

Démonstration des lemmes de Borel-Cantelli :

- *Sens facile :* notons $B_n = \bigcup_{k=n}^{\infty} A_k$, de telle sorte que

$$\{A_i \text{ ; i.s.}\} = \bigcap_{n=1}^{\infty} B_n$$

et, notamment, $\{A_i \text{ ; i.s.}\} \subset B_n$ pour tout n. Par sous-additivité, (exercice 3.10 page 97),

$$\mathbf{P}\{A_i \text{ ; i.s.}\} \leqslant \mathbf{P}(B_n) = \mathbf{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \leqslant \sum_{k=n}^{\infty} \mathbf{P}(A_k).$$

Le membre de droite étant le reste d'une série convergente, il tend vers 0 quand $n \to \infty$, et donc $\mathbf{P}\{A_i \text{ ; i.s.}\} = 0$.

(4). C'est tout à fait intuitif ; formellement, on peut le démontrer en appliquant le lemme des coalitions 16.10 page 343.

• *Sens difficile :* les A_k étant indépendants, alors leurs complémentaires le sont également. Ainsi

$$\mathbf{P}\left(\bigcap_{k=n}^{N} A_k^c\right) = \prod_{k=n}^{N} \mathbf{P}(A_k^c) = \prod_{k=n}^{N} (1 - \mathbf{P}(A_k)).$$

Or la série $\sum \mathbf{P}(A_n)$ diverge, donc (A9) le produit du membre de droite tend vers 0 quand N tend vers $+\infty$, et puisque

$$\mathbf{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) \leqslant \mathbf{P}\left(\bigcap_{k=n}^{N} A_k^c\right)$$

pour tout $N \in \mathbb{N}$, on en déduit que

$$\mathbf{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right) = 0.$$

Une union dénombrable d'ensemble négligeables est négligeable (toujours par inégalité de sous-additivité), donc

$$\mathbf{P}\left(\bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k^c\right) = 0.$$

Enfin, par passage au complémentaire,

$$\mathbf{P}\{A_n \text{ ; i.s.}\} = \mathbf{P}\left(\bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = 1. \qquad \blacksquare$$

Remarque 5.17 (Nécessité de l'hypothèse d'indépendance dans le sens non trivial) Si l'on omet l'hypothèse d'indépendance, on peut donner un contre-exemple immédiat au sens difficile. Choisissons un événement A de probabilité $0 < \mathbf{P}(A) < 1$ et posons $A_n = A$ pour tout $n \in \mathbb{N}$. Alors $\sum \mathbf{P}(A_n)$ diverge (grossièrement), mais $\{A_n \text{ ; i.s.}\} = A$ et donc $\mathbf{P}\{A_n \text{ ; i.s.}\} < 1$.

Remarque 5.18 (Démonstration alternative du sens « facile ») Avec la notion d'espérance d'une variable aléatoire, on peut donner une démonstration alternative du sens « facile ». On note $X_k = \mathbf{1}_{A_k}$ la variable aléatoire valant 1 sur A_k et 0 en dehors ; on suppose que $\sum \mathbf{P}(A_n)$ converge, et l'on définit $S = \sum_{k=1}^{\infty} X_k$, variable aléatoire à valeurs dans $\mathbb{N} \cup \{+\infty\}$. Le théorème de convergence monotone montre que $\mathbf{E}(S) = \sum_{n=1}^{\infty} \mathbf{E}(X_n) = \sum_{n=1}^{\infty} \mathbf{P}(A_n) < +\infty$; notamment, la variable aléatoire positive S est presque sûrement finie. En d'autres termes, $\mathbf{P}\{S = +\infty\} = 0$, ou encore $\mathbf{P}\{A_n \text{ ; i.s.}\} = 0$.

Exemple 5.19 (Retour à l'équilibre dans un jeu de pile *ou* face*)* Deux joueurs — immortels eux aussi, appelons-les Thor et Loki — disputent une partie infinie au jeu de *pile* ou *face*. Leur mise est constamment de 1 pièce d'or, et Thor a commis la bêtise de laisser Loki fournir la pièce — qui, bien entendu, n'est pas équilibrée et assure la victoire à Loki avec une probabilité $p = 0,5 + 10^{-100}$, c'est-à-dire *très* légèrement supérieure à 1/2.

Notons X_n le gain de Loki à l'issue de la n-ième partie et $S_n = X_1 + X_2 + \cdots + X_n$ le gain cumulé au cours des n premières parties. Un retour à l'équilibre du gain total est un événement du type $\{S_n = 0\}$. La question est : peut-il y avoir une infinité de retours à l'équilibre au cours du jeu ?

La réponse, conforme à l'intuition, est « non », presque sûrement. D'après le premier lemme de Borel-Cantelli, il suffit pour cela de prouver que la série $\sum \mathbf{P}\{S_n = 0\}$ converge. Or, si n est impair, S_n ne peut jamais être nul ; si $n = 2k$ est pair, alors

$$\mathbf{P}\{S_{2k} = 0\} = \binom{2k}{k} p^k (1-p)^k.$$

En utilisant la formule de Stirling[5], on obtient l'équivalent

$$\mathbf{P}\{S_{2k} = 0\} \underset{n\to\infty}{\sim} \sqrt{\frac{2}{k\pi}}\, \alpha^k,$$

où $\alpha = 4p(1-p) < 1$; ainsi, $\sum \mathbf{P}\{S_{2k} = 0\}$ converge et on en déduit que : *presque sûrement, il n'y aura qu'un nombre fini de retours à l'équilibre dans la partie entre Thor et Loki.*

Ainsi, à partir d'un certain instant, le gain cumulé de Loki restera définitivement strictement positif ; la loi des grands nombres montre même que, presque sûrement, il tendra vers $+\infty$ (chapitre 20).

---

(5). On peut alternativement utiliser le critère de d'Alembert pour prouver la convergence de la série.

Exemple 5.20 (Application à la convergence de suites de variables aléatoires (*))
Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires réelles, et soit X une variable aléatoire. On suppose que la suite $(X_n)_{n\geqslant 0}$ *converge en probabilité* vers X c'est-à-dire que, pour tout $\varepsilon > 0$,

$$\lim_{k\to\infty} \mathbf{P}\{\,|X_k - X| > \varepsilon\} = 0.$$

Nous allons montrer qu'alors il existe une sous-suite $(X_{\varphi(n)})_{n\in\mathbb{N}}$ qui converge presque sûrement vers X, c'est-à-dire telle que

$$\lim_{n\to\infty} X_{\varphi(n)}(\omega) = X(\omega)$$

pour presque tout $\omega \in \Omega$.

On fabrique une suite strictement croissante des $\varphi(n)$ par récurrence, en choisissant d'abord $\varphi(0)$ arbitrairement. Supposons $\varphi(0), \ldots, \varphi(n-1)$ construits ; on peut alors choisir un entier $k > \varphi(n-1)$ tel que

$$\mathbf{P}\left\{|X_k - X| > \frac{1}{n}\right\} \leqslant \frac{1}{n^2},$$

et on note $\varphi(n)$ cet indice. Une fois cette suite φ construite, on note A_n l'événement $A_n = \left\{\,|X_{\varphi(n)} - X| > 1/n\right\}$; puisque la série $\sum 1/n^2$ converge, on en déduit que $\mathbf{P}\{A_n \text{ ; i.s.}\} = 0$. Ainsi, pour presque tout ω, il n'existe qu'un nombre fini d'événements A_n qui se réalise, et donc, pour tout n supérieur à un certain indice N (dépendant de ω), on a $\left|X_{\varphi(n)}(\omega) - X(\omega)\right| \leqslant 1/n$, et donc $X_{\varphi(n)}(\omega) \to X(\omega)$.

5.3 APPLICATION : DU JEU DE RECOUVREMENT AU PARADOXE D'OLBERS (*)

Comme application directe du lemme de Borel-Cantelli, nous allons étudier un problème de recouvrement d'un segment ou d'un cercle par des petits segments ou de petits arcs. Plus tard dans l'ouvrage, nous reprendrons cette situation en la compliquant un peu, afin de donner un cadre probabiliste satisfaisant à un problème qui a préoccupé les astronomes depuis le XVIII^e^ siècle, et connu sous le nom de paradoxe d'Olbers.

Le paradoxe historique d'Olbers

§ 67

Dans l'hypothèse d'une répartition uniforme des étoiles dans un univers statique — une telle hypothèse est évidemment outrageusement simplificatrice, puisque les étoiles sont, on le sait, regroupées en structures nombreuses, et ce à différentes échelles : amas, galaxies, amas de galaxies, etc., et que de plus l'univers est en expansion — l'astronome allemand Heinrich Olbers montra que, quelle que soit la direction dans laquelle on regarde, notre regard est arrêté par une étoile, éventuellement très éloignée[6]. Or, *la luminosité d'une étoile ne dépend pas de sa distance*[7]. Conclusion : le fond du ciel nocturne devrait avoir une luminosité de l'ordre de celle du Soleil ou d'une étoile moyenne.

Bien sûr, le fait que le ciel nocturne continue d'être noir après une démonstration aussi ingénieuse est quelque peu agaçant, et les scientifiques durent chercher des solutions à ce paradoxe. Ce n'est pas ce qui nous occupe ici[8] ; nous voulons simplement donner une preuve probabiliste de l'argument d'Olbers (« dans toutes les directions, notre regard doit être arrêté par la surface d'une étoile »).

---

(6). Un argument rapide est le suivant : considérons des coquilles sphériques, centrées sur l'observateur, d'épaisseur constante e — disons quelques années-lumière pour fixer les idées — , et de rayons successifs e, $2e$, $3e$, etc. Chacune de ces coquilles contiendra, en moyenne, un même nombre n d'étoiles ; chacune de ces étoiles a une surface angulaire proportionnelle à $1/n^2$; ainsi, chaque coquille contribue pour une surface angulaire à peu près constante. Puisqu'on suppose une infinité de telles coquilles, on en conclut que le ciel est entièrement recouvert.

(7). Précisons : sa luminosité *par unité d'angle solide*. Le fait qu'une étoile lointaine paraisse moins lumineuse est simplement dû au fait qu'elle présente une taille angulaire inversement proportionnelle au carré de sa distance.

(8). Disons simplement que divers facteurs permettent de mettre en défaut le modèle d'Olbers. Outre la répartition non uniforme, qui ne joue guère en vérité, le premier argument d'importance

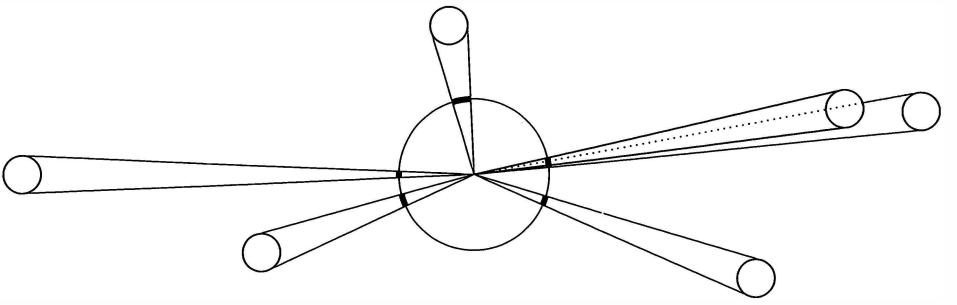
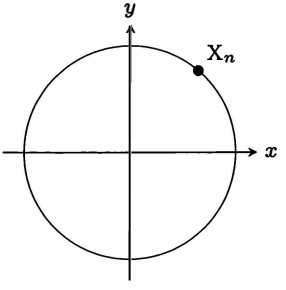
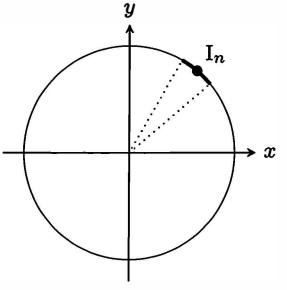


FIGURE 5.2 — *Représentation du paradoxe d'Olbers en deux dimensions : six étoiles et leur projection sur la « voûte céleste ».*

§ 68 Un modèle bidimensionnel de recouvrement

Pour l'instant, contentons-nous d'un problème beaucoup plus simple, d'un *toy-model.* On se donne une suite de points aléatoires $(X_n)_{n\geqslant 0}$ sur le cercle $\mathcal{C}$ de rayon $1/2\pi$ et de périmètre unité. Autour de chacun de ces points, on construit un arc de cercle I_n de largeur a_n, la suite $(a_n)_{n\geqslant 0}$ étant fixée à valeurs dans $[0\,;1[$. On se demande alors, en fonction de la suite $(a_n)_{n\geqslant 0}$, ce que l'on peut dire du recouvrement du cercle par ces petits arcs.





Nous avons un premier résultat de type *ponctuel.* Fixons, une fois pour toute, une direction, c'est-à-dire un point c du cercle ; la probabilité qu'il ne soit pas recouvert par I_1 est $\mathbf{P}\{c \notin I_1\} = 1-a_1$. Par indépendance des événements $\{c \in I_k\}$, la probabilité que c ne soit recouvert par aucun des arcs $I_1, \ldots, I_n$ est

$$\mathbf{P}\{c \notin I_1, \ldots, c \notin I_n\} = \prod_{k=1}^{n}(1 - a_k).$$

Quant à la limite du terme de droite, lorsque l'on fait tendre n vers l'infini, elle est dictée par la nature de la série $\sum a_n$: si $\sum a_n$ converge, alors la suite des produits converge vers une constante $p > 0$; si au contraire $\sum a_n$ diverge, elle converge vers 0 (annexe A9). Par application de l'axiome 3″, on en déduit que, dans le cas de la divergence, $\mathbf{P}\{a \notin \cup_n I_n\} = 0$. On a donc prouvé le résultat suivant :

---

fut celui-ci : si l'Univers n'a qu'un âge fini, alors nous ne pouvons en voir qu'une portion finie, puisque la vitesse de la lumière est finie. De plus, la lumière ne se propage pas impunément dans l'espace : si l'absorption par la matière interstellaire ne compte pas, car elle est nécessairement suivie d'une ré-émission par échauffement de la matière, en revanche le décalage spectral des astres lointains, prédit par la relativité générale, fait que la luminosité du fond du ciel est celle d'un astre à 3 K, et non 3000 K comme une étoile moyenne. Au total, l'intense rayonnement lumineux se situe dans les domaine des micro-ondes ; il a été observé par Penzias et Wilson en 1964 ; dans le domaine visible, il est totalement négligeable.

■ **Proposition 5.21** *Supposons que $\sum a_n$ diverge. Pour tout point c du cercle, la probabilité que c soit recouvert par au moins un intervalle est égale à 1.*

Ce résultat, élémentaire, ne concerne qu'un point à la fois.

Voici un second résultat, de type global cette fois-ci — c'est-à-dire concernant, non pas *un* point (quelconque) du cercle, mais *tous les points* à la fois. On choisit le cercle $\mathcal{C}$ comme espace probabilisé, on le munit de la tribu des boréliens ainsi que de la mesure de Lebesgue habituelle[9]. Dans cet espace, les intervalles I_n[10] ne sont pas indépendants (il n'y a aucune raison pour que $\mathbf{P}(\mathrm{I}_n \cap \mathrm{I}_m)$ soit égale à $\mathbf{P}(\mathrm{I}_n)\,\mathbf{P}(\mathrm{I}_m)$...) mais cela n'empêche pas d'utiliser le premier lemme de Borel-Cantelli :

■ **Proposition 5.22** *Si $\sum a_n$ converge, alors l'ensemble des points qui sont recouverts une infinité de fois par des petits arcs est de mesure nulle sur le cercle.*

DÉMONSTRATION : Les événements I_n sont de probabilité a_n dans l'espace $(\mathcal{C}, \mathfrak{B}(\mathcal{C}), \mu)$ où μ est la mesure habituelle de Lebesgue ; puisque $\sum a_n$ converge, l'événement $\{\mathrm{I}_n \text{ ; i.s.}\}$ est de mesure nulle ; en d'autres termes, l'ensemble des points c de $\mathcal{C}$ tels que c appartient à une infinité de I_n est de mesure nulle. ∎

Modèle stellaire uniforme : probabilité que le cercle soit entièrement recouvert § 69

Qu'en est-il maintenant du cas où la série $\sum a_n$ diverge ? Nous n'allons pas donner de résultat général, mais nous contenter du cas où $a_n = 2/\sqrt{n}$, ce qui est le « cas générique » d'un modèle stellaire uniforme en deux dimensions[11].

On reprend un espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ abstrait, sur lequel on se donne une suite $(\mathrm{X}_n)_{n\geqslant 0}$ de variables aléatoires indépendantes, de même loi uniforme sur $\mathcal{C}$. On note symboliquement $\mathrm{I}_n = \left[\mathrm{X}_n - \frac{1}{\sqrt{n}}\,;\, \mathrm{X}_n + \frac{1}{\sqrt{n}}\right]$ l'arc centré en X_n et de largeur $2/\sqrt{n}$.

■ **Théorème 5.23** *Avec une probabilité 1, le cercle $\mathcal{C}$ est entièrement recouvert par les intervalles I_n et ce, une infinité de fois.*

Remarque 5.24 Insistons sur la différence entre cet énoncé et celui de la proposition 5.21. Dans la proposition 5.21, dire que la « probabilité qu'un point a soit recouvert est égale à 1 » signifie qu'il existe une partie Ω'_a, de probabilité 1, telle que a est recouvert pour toutes les réalisations $\omega \in \Omega'_a$; ceci étant vrai pour tout $a \in \mathcal{C}$. Le point crucial est que, *a priori*, la partie Ω'_a dépend de a. Au contraire, dans le théorème 5.23, il existe une partie Ω' de probabilité 1 telle que, pour tout $a \in \mathcal{C}$ et tout $\omega \in \Omega'$, le point a est recouvert.

Démontrons maintenant ce résultat. Pour cela, fixons un entier n et découpons le cercle en n parties égales, notées $\mathrm{J}_n^k = \left[\frac{k}{n}\,;\,\frac{k+1}{n}\right]$ pour $k = 0, \ldots, n-1$. On note A_n l'événement

A_n : « il existe un J_n^k non complètement recouvert par l'union des arcs I_p pour p variant de $m_n := 2^{n-1}+1$ à 2^n ».

---

(9). On peut imaginer que les points X_n appartiennent au segment $[0\,;\,1[$, les boréliens et la mesure de Lebesgue sur $\mathcal{C}$ étant l'image de ceux sur $[0\,;\,1[$ par l'application $\theta \mapsto (\cos 2\pi\theta, \sin 2\pi\theta)$.

(10). Qui sont maintenant des ensembles fixés, des parties de l'espace *mesuré* $\mathcal{C}$; bien que de nature géométrique, on peut les considérer comme les événements de l'espace *probabilisé* $\mathcal{C}$. Dans la question précédente, les intervalles I_n étaient au contraire des parties aléatoires, dépendant d'un paramètre ω ; les événements étaient de la forme $\{c \in \mathrm{I}_n\} = \{\omega \in \Omega\,;\, c \in \mathrm{I}_n(\omega)\}$.

(11). Un disque de rayon k contient un nombre d'étoiles (bidimensionnelles) proportionnel à k^2 ; ainsi, la n-ième étoile est à une distance globalement proportionnelle à $1/\sqrt{n}$ et présente donc un diamètre angulaire apparent proportionnel à $1/\sqrt{n}$. Nous raffinerons ce point de vue ultérieurement.

Dans une première étape, majorons $\mathbf{P}(A_n)$. Pour un k donné, remarquons d'abord que, si J_n^k n'est pas entièrement recouvert par l'union des arcs $I_{m_n}, \ldots, I_{2^n}$, alors nécessairement aucun des arcs individuels ne le recouvre entièrement. Or, il est simple de minorer la probabilité que J_n^k soit recouvert par un certain I_p (pour $m_n \leqslant p \leqslant 2^n$), puisque cela arrive dès que la distance de X_p au centre de J_n^k est inférieure à $1/\sqrt{p}$:

$$\mathbf{P}(J_n^k \text{ recouvert par } I_p) \geqslant \mathbf{P}\left\{\left|X_p - \frac{k+1/2}{n}\right| \leqslant \frac{1}{\sqrt{p}}\right\} = \frac{2}{\sqrt{p}} \geqslant \frac{2}{2^{n/2}}.$$

Par passage au complémentaire,

$$\mathbf{P}(J_n^k \text{ non entièrement recouvert par } I_p) \leqslant 1 - \frac{2}{2^{n/2}}.$$

Par indépendance des X_p,

$$\mathbf{P}\big(J_n^k \text{ non entièrement recouvert par l'un des } I_{m_n}, \ldots, I_{2^n}\big) \leqslant \left(1 - \frac{2}{2^{n/2}}\right)^{2^{n-1}} \quad (*)$$

(il y a exactement 2^{n-1} tirages d'indice entre m_n et 2^n). Notre remarque préliminaire ayant montré l'inclusion

$$\big(J_n^k \text{ non recouvert par l'union des } I_{m_n}, \ldots, I_{2^n}\big) \subset \big(J_n^k \text{ non entièrement recouvert par l'un des } I_{m_n}, \ldots, I_{2^n}\big),$$

on en déduit

$$\mathbf{P}\big(J_n^k \text{ non recouvert par l'union des } I_{m_n}, \ldots, I_{2^n}\big) \leqslant \left(1 - \frac{2}{2^{n/2}}\right)^{2^{n-1}}$$

Enfin, par sous-additivité de $\mathbf{P}$

$$\mathbf{P}(A_n) \leqslant n\left(1 - \frac{2}{2^{n/2}}\right)^{2^{n-1}} \leqslant n \exp\left(-2^{1-n/2}\right),$$

la dernière inégalité provenant de la convexité de l'exponentielle ($1 - x \leqslant e^{-x}$). La série $\sum n \exp\left(-2^{1-n/2}\right)$ étant convergente, on peut appliquer le premier lemme de Borel-Cantelli[12] : avec une probabilité *zéro*, il existe une infinité d'entiers n tels que A_n est réalisée, c'est-à-dire $\mathbf{P}\{A_n \text{ ; i.s.}\} = 0$.

Passons maintenant au complémentaire. L'ensemble $\Omega' := \{A_n \text{ ; i.s.}\}^c$ est de probabilité 1, et pour tout $\omega \in \Omega'$, il existe un entier $N(\omega)$ tel que $\omega \in \bigcap_{n \geqslant N(\omega)} \overline{A_n}$. Ainsi, pour tout $\omega \in \Omega'$, le cercle entier est entièrement recouvert par les arcs $I_{m_n}, \ldots, I_{2^n}$, et ce pour tout choix de $n \geqslant N(\omega)$.

Pour terminer, remarquons que les supports $\{m_n, \ldots, 2^n\}$ sont deux à deux disjoints ; ce qui prouve que, pour tout $\omega \in \Omega'$, $\mathcal{C}$ est entièrement recouvert par une infinité de séries de tirages de la forme $\{m_n, m_n + 1, \ldots, 2^n\}$, donc entièrement
A5.7 recouvert une infinité de fois, ce qui achève la preuve du théorème.

(12). Le lecteur ayant cherché seul à montrer le théorème 5.23 aura peut-être eu l'intuition — trompeuse — que c'était au second lemme de Borel-Cantelli de jouer un rôle. Ici, l'indépendance des tirages n'intervient que dans l'établissement de l'équation (*).

Remarque 5.25 (Du toy-model *à un modèle plus élaboré* (∗)*)* On peut raffiner le résultat précédent pour proposer un modèle beaucoup plus réaliste.

— Le théorème 5.23 reste vrai si l'on prend la condition plus souple $a_n \sim \beta/\sqrt{n}$, la démonstration étant essentiellement la même ; il suffit d'ailleurs de remarquer que $a_n \geqslant \beta/2\sqrt{n}$ à partir d'un certain rang.

— Il reste encore vrai si le diamètre angulaire a_n est le résultat d'un processus de Poisson dans le plan, c'est-à-dire donné par

$$a_n = \frac{\beta}{\mathrm{R}_n},$$

où R_n est la distance de la n-ième étoile,

$$\mathrm{R}_n = \sqrt{\mathrm{P}_1 + \mathrm{P}_2 + \cdots + \mathrm{P}_n}$$

et les P_k sont des variables aléatoires exponentielles indépendantes. Dans un tel modèle, une étoile est placée à la distance R_n et selon un angle θ_n de loi uniforme sur $[0\,;\,2\pi[$. La densité moyenne d'étoiles est une constante et les positions des étoiles sont indépendantes les unes des autres : c'est donc un modèle satisfaisant d'univers infini, statique et homogène (*i.e.* de densité moyenne constante).
La loi des grands nombre montre que, presque sûrement, $a_n \sim \gamma/\sqrt{n}$ où γ est une constante (dépendant des paramètres du modèle de Poisson).

— Enfin, considérons le cas où le diamètre réel des étoiles n'est pas constant ; on doit alors écrire

$$a_n = \frac{\mathrm{B}_n}{\mathrm{R}_n},$$

où $\mathrm{R}_n \sim \sqrt{n}$ presque sûrement comme avant, et les B_n sont des variables aléatoires indépendantes et de même loi (quelconque).
On choisit deux réels a et b tels que

$$p = \mathbf{P}\{a \leqslant \mathrm{B} \leqslant b\} > 0.$$

On élimine maintenant de la sélection les étoiles telles que B_n n'est pas dans l'intervalle $[a\,;\,b]$; le résultat — processus de Poisson de paramètre λ, conditionné par une loi de Bernoulli $\mathscr{B}(p)$ — est encore un processus de Poisson, de paramètre λp (théorème 30.7 page 607).
Or, dans le cas d'un processus de Poisson et d'un rayon constant $\mathrm{B}_n = a$ pour toutes les étoiles, on a montré que l'on aurait, presque sûrement, recouvrement de tout le cercle ; avec un rayon $\mathrm{R}_n \in [a\,;\,b]$, on a (encore mieux) ce recouvrement. Ainsi, les seules étoiles sélectionnées suffisent pour assurer le recouvrement de tout le cercle.

— Enfin, tout ce qui a été fait ici en deux dimensions fonctionne en trois dimensions. L'angle solide de la n-ième étoile est $a_n \sim \beta/n^{2/3}$ et la série $\sum a_n$ diverge encore dans le modèle le plus simple. Dans le modèle plus élaboré donné par un processus de Poisson, on pose

$$\mathrm{R}_n = \sqrt[3]{\mathrm{P}_1 + \mathrm{P}_2 + \cdots + \mathrm{P}_n} \qquad \text{et} \qquad a_n = \frac{\mathrm{B}_n}{\mathrm{R}_n^2}.$$

On peut alors conclure de la même manière.

EXERCICES

♦ ***Exercice 5.1 (Limite d'une suite de parties de*** $\mathbb{R}$***)*** On pose $\mathrm{A}_n = [n\,;\,2n]$ pour tout $n \geqslant 1$, de sorte que la largeur de A_n soit égale à n. La suite $(\mathrm{A}_n)_{n\geqslant 0}$ admet-elle une limite ? Si oui, quelle est longueur de cette limite ?

♦ ***Exercice 5.2*** Un véhicule possède trois moteurs ; il ne peut avancer que si deux au moins des moteurs fonctionnent. La probabilité que le premier moteur fonctionne est p_1 ; celle que le deuxième moteur fonctionne est p_2, et celle que la troisième fonctionne est p_3. Quelle est la probabilité que le véhicule puisse avancer ? (On supposera que chaque moteur est indépendant des autres.)

♦ ***Exercice 5.3 (Un petit piège)*** On lance deux dés honnêtes. On note A et B les événements

A : « un des dés marque 1 » B : « la somme des points des deux dés vaut 7 ».

Montrer que A et B ne sont pas indépendants.

On suppose que l'un des dés est rouge, l'autre noir. On note

A′ : « le dé rouge marque 6 » B : « la somme des points des deux dés vaut 7 ».

Montrer que A′ et B sont indépendants.

♦ ***Exercice 5.4*** Démontrer la proposition 5.12 page 116.

♦ ***Exercice 5.5 (Suite de variables gaussiennes de variance tendant vers $+\infty$)***
N.B. : Cet exercice, application directe du lemme de Borel-Cantelli, requiert la notion de variable aléatoire gaussienne (chapitre 10).

Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires gaussiennes centrées telles que $\mathbf{V}(X_n) \to +\infty$.

i) Montrer que $\mathbf{P}\{|X_n| < M\} \to 0$ pour tout $M > 0$.

ii) Montrer que, pour tout $k \in \mathbb{N}^*$, il existe $n(k)$ tel que $\mathbf{P}\{|X_{n(k)}| < k\} \leqslant 2^{-k}$.

iii) En déduire que la suite $(X_n)_{n\geqslant 0}$ est presque sûrement non bornée.

♦ ***Exercice 5.6 (Inégalité de Fatou)*** Soit $(A_n)_{n\geqslant 0}$ une suite d'événements de $(\Omega, \mathcal{T}, \mathbf{P})$. Montrer que

$$\mathbf{P}(\liminf A_n) \leqslant \liminf \mathbf{P}(A_n) \qquad \text{et} \qquad \limsup \mathbf{P}(A_n) \leqslant \mathbf{P}(\limsup A_n).$$

En déduire que, si la suite $(A_n)_{n\geqslant 0}$ converge (au sens où $\liminf A_n = \limsup A_n$) alors la suite de terme général $\mathbf{P}(A_n)$ converge et $\lim\limits_{n\to\infty} \mathbf{P}(A_n) = \mathbf{P}(\lim\limits_{n\to\infty} A_n)$.

♦ ***Exercice 5.7 (Olbers en trois dimensions)*** Reprendre le *toy-model* du paragraphe § 69 page 121, et l'adapter pour un modèle tridimensionnel d'étoile.

Indications

◊ *Indication 5.2 :* Écrire l'événement « le véhicule peut avancer » sous la forme d'une union *disjointe* d'événements.

◊ *Indication 5.3 :* Ne pas forcément se fier à l'intuition, mais calculer paisiblement les probabilités de A, B, $A \cap B$, A' et $A' \cap B$.

◊ *Indication 5.5 :* Si $Y \rightsquigarrow \mathcal{N}(0\,;\,1)$, vérifier que $\mathbf{P}\{|Y| < a\} \leqslant a$ pour tout $a > 0$. Pour conclure, utiliser le lemme de Borel-Cantelli dans le seul sens autorisé (les variables ne sont pas supposées indépendantes !)

◊ *Indication 5.7 :* Une sphère de rayon k contient, dans un modèle uniforme, un nombre d'étoile proportionnel à k^3 ; ainsi, le diamètre angulaire de la n-ième étoile sera pris égal à $\alpha/n^{1/3}$, et son angle solide sera $\approx \alpha^2/n^{2/3}$.

SOLUTIONS

♦ **Solution 5.1** En appliquant les définitions, on trouve que $\liminf A_n = \limsup A_n = \varnothing$, et donc que $A_n \to \varnothing$... ensemble limite de longueur évidemment nulle.

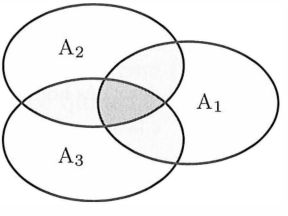
♦ **Solution 5.2** Notons A_k l'événement « le moteur k est en bon état ». Le véhicule peut avancer si et seulement si l'événement

$$B := (A_1 \cap A_2) \cup (A_2 \cap A_3) \cup (A_1 \cap A_3)$$

est réalisé. Comme l'union n'est pas disjointe, récrivons-la sous la forme

$$B = (A_1 \cap A_2 \cap A_3^c) + (A_2 \cap A_3 \cap A_1^c) + (A_3 \cap A_1 \cap A_2^c) + (A_1 \cap A_2 \cap A_3)$$

correspondant au découpage suivant :



L'indépendance des A_k implique l'indépendance de A_1, A_2 et A_3^c, donc $\mathbf{P}(A_1 \cap A_2 \cap A_3^c) = p_1 p_2 (1 - p_3)$; de même pour les deux termes suivants. On trouve finalement

$$\mathbf{P}(B) = p_1 p_2 (1 - p_3) + p_2 p_3 (1 - p_1) + p_3 p_1 (1 - p_2) + p_1 p_2 p_3 = p_1 p_2 + p_2 p_3 + p_1 p_3 - 2 p_1 p_2 p_3.$$

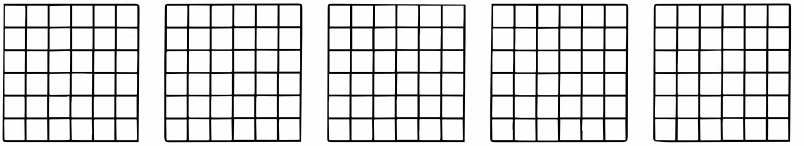
♦ **Solution 5.3** Il y a 11 combinaisons de dés menant à l'événement A, sur un total de 36. Ainsi, $\mathbf{P}(A) = 11/36$. De même, il y a 6 combinaisons menant à l'événement B, donc $\mathbf{P}(B) = 1/6$. Enfin, il n'y a que deux combinaisons menant à $A \cap B$ (1 rouge et 6 noir, ou 1 noir et 6 rouge), donc

$$\mathbf{P}(A \cap B) = \tfrac{1}{18} \neq \mathbf{P}(A) \cdot \mathbf{P}(B).$$

L'événement A' est évidemment de probabilité 1/6, et $A' \cap B$ n'a lieu que si le dé noir marque 1, c'est-à-dire

$$\mathbf{P}(A' \cap B) = \tfrac{1}{36} = \mathbf{P}(A') \cdot \mathbf{P}(B).$$

Les événements A' et B sont indépendants, tandis que A et B ne le sont pas.



A $\quad$ $A \cap B$ $\quad$ B $\quad$ $A' \cap B$ $\quad$ A'

♦ **Solution 5.4**

i) Si $\omega \in \liminf A_n$, cela signifie que, à partir d'un certain rang, il est dans tous les A_n ; il est donc dans une infinité d'entre eux — et donc dans $\limsup A_n$.

ii) Si $\omega \in \Omega$, il y a équivalence entre les énoncés

(a) $\omega \notin \liminf A_n$

(b) la propriété « ω est dans tous les A_n à partir d'un certain rang » est fausse ;

(c) quel que soit l'entier N, il existe un entier $n \geqslant N$ tel que $\omega \notin A_n$:

(d) quel que soit l'entier N, il existe un entier $n \geqslant N$ tel que $\omega \in A_n^c$:

(e) ω est dans un nombre infini de A_n^c ;

(f) $\omega \in \limsup A_n^c$.

iii) Si $(A_n)_{n \geqslant 0}$ est croissante, alors

$$\bigcap_{n \geqslant N} A_n = A_N$$

et donc $\liminf A_n = \lim_n \uparrow A_n$. De plus, pour tout entier n on a l'égalité

$$\bigcup_{k=n}^{\infty} A_k = \bigcup_{k=0}^{\infty} A_k,$$

donc $\limsup A_n = \bigcup_{k=0}^{\infty} A_k = \lim_n \uparrow A_n$.

iv) De même, si $(A_n)_{n \geqslant 0}$ est décroissante, $\bigcup_{n \geqslant N} A_n = A_N$ donc $\limsup A_n = \lim_n \downarrow A_n$.

v) Soit $\omega \in A_*$, alors $\mathbf{1}_{A_*}(\omega) = 1$. De plus, à partir d'un certain rang, ω appartient à tous les A_n, donc $\mathbf{1}_{A_n}(\omega) = 1$ et $\liminf \mathbf{1}_{A_n}(\omega) = \lim \mathbf{1}_{A_n}(\omega) = 1$.
De même, si $\omega \notin A_*$, cela signifie que ω est dans une infinité de A_n^c, et donc que $\mathbf{1}_{A_n}(\omega)$ s'annule une infinité de fois, donc 0 est valeur d'adhérence de la suite de terme général $\mathbf{1}_{A_n}(\omega)$. Comme cette dernière est à valeurs dans $\{0, 1\}$, sa limite inférieure est au moins 0, donc finalement $\liminf \mathbf{1}_{A_n}(\omega) = 0 = \mathbf{1}_{A_*}(\omega)$.
On peut effectuer un travail semblable pour la deuxième égalité, ou bien passer au complémentaire et utiliser la deuxième propriété de la proposition.

♦ **Solution 5.5**

i) Si $Y \rightsquigarrow \mathcal{N}(0\,;\,1)$, alors pour tout $a > 0$, on a $\sqrt{2\pi} > 2$ et donc

$$\mathbf{P}\{\,|Y| < a\,\} = \frac{1}{\sqrt{2\pi}} \int_{-a}^{+a} \mathrm{e}^{-t^2/2}\,\mathrm{d}t \leqslant \frac{2a}{2} = a.$$

Soit $M > 0$. Si l'on pose $Y_n = X_n / \sigma_{X_n}$ pour tout $n \in \mathbb{N}$, alors Y_n suit la loi normale centrée réduite et l'inégalité précédente nous donne

$$\mathbf{P}\{\,|X_n| < M\,\} = \mathbf{P}\left\{\,|Y_n| \leqslant \frac{M}{\sigma_{X_n}}\,\right\} \leqslant \frac{M}{\sigma_{X_n}},$$

ce qui prouve que $\mathbf{P}\{\,|X_n| < M\,\} \to 0$.

ii) Conséquence immédiate de la question précédente.

iii) La série $\sum \mathbf{P}\{|X_{n(k)}| < k\}$ converge, et donc $\mathbf{P}\{|X_{n(k)}| < k \text{ ; i.s.}\} = 0$. Ou encore : avec une probabilité 1, un nombre fini d'événements $\{|X_{n(k)}| < k\}$ se réalise. Notamment, avec une probabilité égale à 1, la suite $(X_{n(k)})_k$ est non bornée, et donc la suite $(X_n)_{n\geqslant 0}$ également.

♦ **Solution 5.6** Notons $A_* = \liminf A_n$. On pose $B_n = \bigcap_{k=n}^{\infty} A_n$ pour tout $n \in \mathbb{N}$. Alors la suite $(B_n)_{n\geqslant 0}$ est croissante et $A_* = \lim_n B_n$; l'axiome 3′ permet d'écrire que $\mathbf{P}(A_*) = \lim_n \mathbf{P}(B_n)$. Par ailleurs, $B_n \subset A_n$, donc $\mathbf{P}(B_n) \leqslant \mathbf{P}(A_n)$. En passant à la limite inférieure, on obtient $\lim_n \mathbf{P}(B_n) \leqslant \liminf \mathbf{P}(A_n)$, ce qui démontre la première inégalité.

La seconde inégalité se montre de même en posant $C_n = \bigcup_{k=n}^{\infty} A_n$: la suite $(C_n)_{n\geqslant 0}$ est décroissante, $A^* = \lim_n C_n$ et $\mathbf{P}(A^*) = \lim_n \mathbf{P}(C_n)$. De plus, $A_n \subset C_n$ donc $\mathbf{P}(A_n) \leqslant \mathbf{P}(C_n)$, et il ne reste qu'à passer à la limite supérieure pour conclure.

Enfin, si $A_* = A^*$, alors $\liminf_n \mathbf{P}(A_n) = \limsup_n \mathbf{P}(A_n)$, ce qui prouve (A2) que la suite de terme général $\mathbf{P}(A_n)$ converge et a pour limite $\mathbf{P}(A^*) = \mathbf{P}(\lim_n A_n)$.

Résumé du chapitre

Quelques formules résument les techniques de calcul les plus utilisées :

Probabilité d'une union

- Si les événements sont deux à deux incompatibles,

$$\mathbf{P}\Big(\sum_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} \mathbf{P}(A_k).$$

- Sinon, on peut utiliser la formule de Poincaré :

$$\mathbf{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n} \mathbf{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}).$$

- Pour une union dénombrable, on a

$$\mathbf{P}\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \lim_{n\to\infty} \mathbf{P}\Big(\bigcup_{k=1}^{n} A_k\Big).$$

Probabilité d'une intersection

- Si les événements sont mutuellement indépendants,

$$\mathbf{P}\Big(\bigcap_{k=1}^{n} A_k\Big) = \prod_{k=1}^{n} \mathbf{P}(A_k).$$

- Sinon, on utilise la formule des probabilités composées :

$$\mathbf{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2 \mid A_1) \cdot \mathbf{P}(A_3 \mid A_1 \cap A_2) \cdots \cdots \mathbf{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}).$$

De plus, les méthodes sur les unions peuvent servir aux intersection par passage au complémentaire, puisque

$$\Big(\bigcap_{k=1}^{n} A_k\Big)^{c} = \bigcup_{k=1}^{n} A_k^{c}.$$

Chapitre 6

Variables aléatoires discrètes

Les variables ne prenant qu'un nombre fini ou dénombrable de valeurs jouent un rôle particulier dans la théorie des probabilités : outre qu'elle sont naturelles dans les problèmes de dénombrement, elles permettent également de comprendre et de démontrer de nombreux résultats généraux de la théorie des probabilités sans avoir recours aux outils sophistiqués de la théorie de la mesure.

Les trois chapitres qui suivent traitent de ces variables, du point de vue théorique comme du point de vue pratique.

6.1 VARIABLES ALÉATOIRES DISCRÈTES

Dans tout ce chapitre, $(\Omega, \mathcal{T}, \mathbf{P})$ est un espace probabilisé ; de plus, les variables aléatoires considérées seront toujours, sans que l'on prenne la peine de le préciser, à valeurs réelles.

Variable simple, variable discrète § 70

Soit $(\Omega, \mathcal{T})$ un espace probabilisable (qui pourra la plupart du temps être considéré comme un ensemble fini ou dénombrable muni de sa tribu la plus riche, à savoir $\mathcal{T} = \mathfrak{P}(\Omega)$). Une **variable aléatoire (réelle) discrète** sur $(\Omega, \mathcal{T})$ est une variable aléatoire X telle que $X(\Omega)$ est un ensemble réel fini ou dénombrable. On dit que X est une **variable aléatoire simple** si $X(\Omega)$ est un ensemble fini.

L'ensemble $X(\Omega)$ est appelé **univers-image** ou **ensemble des valeurs prises par X** ou **ensemble des réalisations de X** ; on le notera génériquement

$$X(\Omega) = \{x_k \ ; \ k \in K\}$$

où $K = \{1, 2, \ldots, n\}$ si $X(\Omega)$ est un ensemble fini, et $K = \mathbb{N}$ ou $K = \mathbb{N}^*$ si $X(\Omega)$ est strictement dénombrable.

Pour tout $x \in \mathbb{R}$ et pour toute partie $A \subset \mathbb{R}$, on note

$$\{X = x\} = X^{-1}(\{x\}) = \{\omega \in \Omega \ ; \ X(\omega) = x\}$$

et

$$\{X \in A\} = X^{-1}(A) = \{\omega \in \Omega \ ; \ X(\omega) \in A\}.$$

La probabilité associée à ces événements sera notée $\mathbf{P}\{X = x\}$ et $\mathbf{P}\{X \in A\}$ plutôt que $\mathbf{P}(\{X = x\})$ et $\mathbf{P}(\{X \in A\})$.

Remarque 6.1 (Sur la mesurabilité des variables aléatoires) La tribu $\mathcal{T}$ dont on munit Ω n'est pas toujours $\mathcal{P}(\Omega)$, notamment lorsque Ω est un ensemble non dénombrable. Dans ce cas, une variable aléatoire sur $(\Omega, \mathcal{T}, \mathbf{P})$ ne peut pas être n'importe quelle application, mais doit vérifier la propriété suivante : pour tout $x \in \mathbb{R}$, l'ensemble $\{X = x\}$ est un élément de $\mathcal{T}$, de telle sorte que l'expression $\mathbf{P}\{X = x\}$ ait un sens. Une application vérifiant cette propriété est dite **mesurable** ; une définition généralisant le cas présent (où $X(\Omega)$ est dénombrable) sera donnée dans les chapitres suivants, notamment, définition 13.1 page 291.

En attendant, et tant que l'on n'étudiera qu'une seule variable aléatoire (ou un nombre fini de variables aléatoires), il est toujours possible de prendre comme univers modèle un ensemble Ω dénombrable, et de le munir de la tribu $\mathcal{T} = \mathcal{P}(\Omega)$. Aussi, nous reportons les questions de mesurabilité des variables aléatoires aux chapitres 13 et suivants.

Exemple 6.2 (Tirage à pile *ou* face*)* On peut modéliser un tirage à *pile* ou *face* par une variable aléatoire $X : \Omega \to \{0, 1\}$, vérifiant $\mathbf{P}\{X = 1\} = p$ et $\mathbf{P}\{X = 0\} = 1 - p$. La valeur de p caractérise la pièce ; une pièce « honnête » sera modélisée par $p = 1/2$. Une pièce déséquilibrée (ou un lanceur habile) sera mieux modélisée par une autre valeur de p (le lecteur curieux pourra se reporter à l'article [24]).

Remarque 6.3 (Sur l'espace de départ) Dans l'exemple précédent, nous avons omis de préciser quel était l'espace probabilisé $(\Omega, \mathcal{T}, \mathbf{P})$ sur lequel la variable aléatoire X était définie, et comment elle était définie. Il est pourtant aisé de construire un tel couple espace/variable :

— on choisit $\Omega = \{0, 1\}$;
— on pose $\mathcal{T} = \mathcal{P}(\Omega) = \{\varnothing, \{0\}, \{1\}, \Omega\}$;
— on munit $(\Omega, \mathcal{T})$ de la probabilité définie par $\mathbf{P}(\varnothing) = 0$, $\mathbf{P}(\{0\}) = 1 - p$, $\mathbf{P}(\{1\}) = p$ et $\mathbf{P}(\Omega) = 1$;
— on définit enfin $X : \Omega \to \mathbb{R}$ comme l'application identique : $X(0) = 0$ et $X(1) = 1$.

C'est tellement trivial, au sens propre, que c'en est à pleurer. Mais, très bientôt, nous aurons besoin de considérer, sur un même espace probabilisé, d'autres variables aléatoires, en nombre fini ou non, dénombrable ou non, et de postuler diverses hypothèses sur ces variables (hypothèses d'indépendance notamment) : toutes choses qui rendront difficile, voire pratiquement impossible une description utilisable de l'espace $(\Omega, \mathcal{T}, \mathbf{P})$, lequel devra être « suffisamment gros ». Le mieux à faire, et le plus court, est de ne pas chercher à connaître Ω, mais à admettre qu'un espace probabilisé répondant à nos attentes existe toujours[1]. On prendra l'habitude de se concentrer, non sur « X » en tant qu'application définie sur un espace de départ Ω, mais uniquement sur la répartition des valeurs prises par X dans l'espace d'arrivée $\mathbb{R}$.

Remarque 6.4 (Sur la valeur des paramètres de certaines lois) Dans l'exemple 6.2, nous avons *postulé* une certaine loi pour la variable aléatoire X. Dans certaines situations, les valeurs numériques proviennent de modèles (plus ou moins discutables). Parfois, elles sont directement issues d'observations, comme par exemple dans le cas d'une variable suivant une loi de Bernoulli $\mathscr{B}(p)$[2]. La difficulté est alors d'évaluer le nombre p. Pour cela, il existe plusieurs approches possibles. L'une est la *loi des grands nombres*, qui nous apprend que la répétition de l'expérience dont le résultat numérique est X nous permet d'approcher, asymptotiquement, le nombre p (c'est la limite des moyennes des valeurs observées) ; le chapitre 20 est consacré à ces résultats de convergence. Une autre possibilité est l'*estimation bayésienne*, que nous commenterons au chapitre 32.

§71 Système complet induit par une variable aléatoire discrète

L'ensemble $\big\{\{X = x\}\ ;\ x \in X(\Omega)\big\}$ est un système complet d'événements, appelé **système complet induit par X**. Dans toute la suite, on le notera

$$\mathfrak{D}_X := \big\{\{X = x_k\}\ ;\ k \in K\big\}.$$

Du fait que $X(\Omega) = \{x_k\ ;\ k \in K\}$ est dénombrable, le système $\mathfrak{D}_X$ est pratique d'utilisation, car il permet d'exploiter la propriété de σ-additivité de la mesure de probabilité.

---

(1). Le théorème d'extension de Kolmogorov est l'outil nous permettant de calmer nos inquiétudes légitimes ; il est présenté dans l'annexe C.

(2). Rappelons que c'est bien l'observation de la répétition statistique de certains phénomènes qui a permis la mathématisation des probabilités ; ainsi du prince de Toscane qui avait remarqué que l'on obtient plus souvent 10 que 9 en lançant trois dés à six faces. La mathématisation des probabilités a d'abord été basée sur une interprétation statistique.

Exemple 6.5 (Rayons cosmiques) On note N le nombre d'impacts de rayons cosmiques détectés par une chambre à bulle durant un intervalle d'une heure. On peut alors modéliser N par une variable aléatoire à valeurs dans $\mathbb{N}$ (il est, non seulement de probabilité nulle, mais physiquement exclu qu'une infinité de particules frappe le détecteur). La loi de N reste à déterminer, le meilleur modèle étant une loi de Poisson (voir § 80). Si A est un événement quelconque, il pourra être pratique de le décomposer comme l'union disjointe

$$A = \sum_{n=0}^{\infty} A \cap \{N = n\}.$$

Exemple 6.6 (Premier succès dans un tirage à pile *ou* face*)* On considère une succession infinie de tirages à *pile* ou *face*, que l'on modélise par une suite $(X_n)_{n \geqslant 1}$ de variables aléatoires définies par : $X_k = 1$ si le $k$ième tirage donne *pile* et $X_k = 0$ sinon. On note Y le rang du premier tirage amenant *pile*. Les valeurs prises par Y sont les entiers positifs ; il faut cependant considérer le cas où $X_k = 0$ pour tout $k \in \mathbb{N}$, pour lequel on donnera une valeur spéciale à Y, par exemple $Y = 0$:

$$\mathfrak{D}_Y = \{\{Y = 0\}, \{Y = 1\}, \{Y = 2\}, \{Y = 3\}, \ldots, \{Y = n\}, \ldots\}.$$

L'usage est répandu de considérer que Y prend ses valeurs dans $\{1, 2, \ldots\} \cup \{+\infty\}$ et de poser $Y = +\infty$ si tous les tirages mènent à *face* (voir § 75), ce qui fait que

$$\mathfrak{D}_Y = \{\{Y = 1\}, \{Y = 2\}, \{Y = 3\}, \ldots, \{Y = n\}, \ldots, \{Y = +\infty\}\}.$$

Loi d'une variable aléatoire discrète. § 72

Selon la définition générale de la loi d'une variable aléatoire (voir § 24), la **loi** d'une variable aléatoire discrète X est la probabilité image $\mathbf{P}_X = \mathbf{P} \circ X^{\langle -1 \rangle}$, c'est-à-dire l'application

$$\mathbf{P}_X : \mathfrak{P}(\mathbb{R}) \longrightarrow \mathbb{R}^+$$
$$A \longmapsto \mathbf{P}\{X \in A\}.$$

Décrire la loi de X, c'est — en principe — décrire les $\mathbf{P}\{X \in A\}$ pour toutes les parties A de $\mathbb{R}$. Dans le cas d'une variable aléatoire discrète, on peut toutefois simplifier cette description ; en effet, pour toute partie A de $\mathbb{R}$, $\{X \in A\}$ s'écrit comme une union *disjointe*, finie ou dénombrable, d'éléments de $\mathfrak{D}_X$:

$$X^{-1}(A) = \sum_k \{X = x_k\}$$

où l'union disjointe porte sur l'ensemble des indices k tels que $x_k \in A$; le caractère σ-additif de la mesure de probabilité implique que

$$\mathbf{P}_X(A) = \mathbf{P}\{X \in A\} = \mathbf{P}\left(\sum_{k \,;\, x_k \in A} \{X = x_k\} \right) = \sum_{k \,;\, x_k \in A} \mathbf{P}\{X = x_k\}.$$

En résumé :

■ **Proposition 6.7** *La loi d'une variable aléatoire discrète* X *d'image* $\{x_k \,;\, k \in K\}$ *est entièrement caractérisée par les probabilités atomiques* $p_k = \mathbf{P}\{X = x_k\}$. *Pour toute partie* A *de* $\mathbb{R}$, *on a*

$$\mathbf{P}_X(A) = \mathbf{P}\{X \in A\} = \sum_{k \,;\, x_k \in A} \mathbf{P}\{X = x_k\}.$$

En pratique, déterminer la loi d'une variable aléatoire discrète, ce sera donc :

- déterminer son image $X(\Omega) = \{x_k \,;\, k \in K\}$;
- déterminer les probabilités atomiques $p_k = \{X = x_k\}$ associées.

Remarque 6.8 L'image exacte d'une variable aléatoire est parfois délicate à déterminer, et une simple inclusion peut faire l'affaire (voir, paragraphe § 81, le cas de la loi hypergéométrique). De plus, par définition, lorsque l'on détermine la loi d'une variable, on ne se préoccupe pas des événements de probabilité nulle. Un tirage à *pile* ou *face* est aussi bien modélisé par une variable aléatoire X prenant les valeurs 0 et 1 avec une probabilité 1/2 chacune, que par une variable aléatoire Y prenant les valeurs 0 et 1 avec la probabilité 1/2 et la valeur -1 avec la probabilité 0 (« la pièce tombe sur la tranche ») ; les lois de X et de Y sont égales.

§ 73 Fonction de répartition d'une variable aléatoire discrète

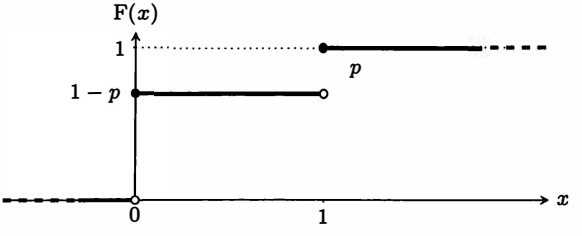
Si X est une variable aléatoire réelle, sa **fonction de répartition** est l'application $F : \mathbb{R} \to [0\,;1]$ définie par

$$\forall x \in \mathbb{R} \qquad F(x) = \mathbf{P}\{X \leqslant x\}.$$

Exemple 6.9 (Variable de Bernoulli) Considérons une variable aléatoire X ne prenant que les valeurs 0 et 1, avec les probabilités respectives $1-p$ et p. Sa fonction de répartition est

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1-p & \text{si } x \in [0\,;1[\,, \\ 1 & \text{si } x \geqslant 1. \end{cases}$$

Elle a l'allure suivante :



On remarquera que $F(0) = 1-p$, que $F(1) = 1$ et que la fonction F est continue à droite en tous points.

■ **Proposition 6.10** *Soit* X *une variable aléatoire réelle discrète ; si* F *est sa fonction de répartition, alors :*

i) F *est croissante sur* $\mathbb{R}$;

ii) $\lim_{x \to -\infty} F(x) = 0$;

iii) $\lim_{x \to +\infty} F(x) = 1$;

iv) F *est continue à droite :* $F(x) = F(x^+)$ *pour tout* $x \in \mathbb{R}$;

v) si a *est un point de* $X(\Omega)$ *tel que* $\mathbf{P}\{X = a\} > 0$, *alors* F *admet en* a *un saut de discontinuité et*

$$F(a) - F(a^-) = \mathbf{P}\{X = a\}.$$

DÉMONSTRATION : Ces premières propriétés sont générales (vraies pour des variables aléatoires autres que discrètes).

i) Si $x \leqslant y$, alors $]-\infty\,;x] \subset]y\,;+\infty]$ et donc $\{X \leqslant x\} \subset \{X \leqslant y\}$; par croissance de **P**, on a $F(x) \leqslant F(y)$.

ii) Soit $(x_n)_{n \geqslant 0}$ une suite décroissante de limite $-\infty$. Alors la suite de terme général $[-\infty\,;x_n[$ décroît vers $\varnothing$, et de même $\{X \leqslant x_n\}$ décroît vers $\varnothing$ (l'intersection des $\{X \leqslant x_n\}$ est vide). D'après l'axiome 3″ page 86, on a donc $F(x_n) \to 0$.

iii) Soit $(x_n)_{n \geqslant 0}$ une suite croissante de limite $+\infty$, alors la suite de terme général $]-\infty\,;x_n]$ croît vers $\mathbb{R}$ (son union est $\mathbb{R}$ tout entier) ; notamment, la suite de terme général $\{X \leqslant x_n\}$ croît vers Ω et l'axiome 3′ permet de conclure que $F(x_n) \to 1$.

iv) Soit $x \in \mathbb{R}$ et soit $(x_n)_{n\geqslant 0}$ une suite décroissante de réels de limite x ; alors $\lim_n \downarrow \{X \leqslant x_n\} = \{X \leqslant x\}$ et l'axiome 3'' permet de conclure que $F(x_n) \to F(x)$.

v) Enfin, on considère la suite d'intervalles $I_n =]\, a - 2^{-n}\, ;\, a\,]$, suite décroissante dont la limite est $\lim_{n\to\infty} \downarrow I_n = \{a\}$. Alors, d'une part, l'axiome 3'' nous assure que

$$\mathbf{P}(I_n) = F(a) - F(a - 2^{-n}) \xrightarrow[n\to\infty]{} \mathbf{P}\{X = a\}\,.$$

D'autre part, F étant monotone, elle admet une limite à gauche en a et

$$F(a - 2^{-n}) \to F(a^-),$$

ce qui conclut la preuve.

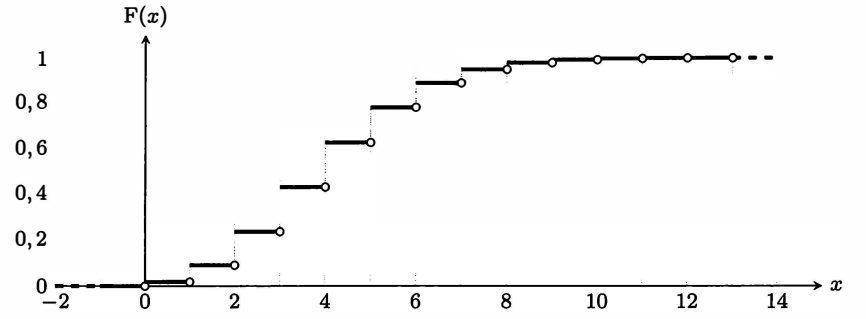
∎

Remarque 6.11 Pour démontrer les quatre premières propriétés, on aurait pu remarquer plus simplement que F est la fonction de répartition de la probabilité $\mathbf{P}_X$, image de $\mathbf{P}$ par X :

$$F(x) = \mathbf{P}\{X \leqslant x\} = \mathbf{P}\big(X^{-1}(]-\infty\,;\,x\,])\big) = \mathbf{P}_X\big(]-\infty\,;\,x\,]\big) = F_X(x),$$

et se reporter à la proposition 3.30 page 91.

Exemple 6.12 (Fonction de répartition d'une loi de Poisson) Ci-dessous, la fonction de répartition d'une variable aléatoire X, ne prenant que des valeurs k entières avec les probabilités respectives $\mathbf{P}\{X = k\} = e^{-4}\,4^k/k!$ (c'est une *loi de Poisson* de paramètre $\lambda = 4$).



Remarque 6.13 (Fonctions de répartition difficiles à représenter) Bien des variables aléatoires que nous aurons à considérer prendront leurs valeurs dans $\mathbb{N}$ ou dans $\mathbb{Z}$, parfois dans un ensemble à peine plus compliqué, comme l'ensemble des inverses des entiers non nuls. Dans ces cas, la fonction de répartition se représente sans trop de difficultés. On peut cependant construire des cas (artificiels) pour lesquels une telle représentation graphique est impossible. À titre d'exemple, puisque $\mathbb{Q}$ est dénombrable, on peut en choisir une énumération

$$\mathbb{Q} = \{r_1, r_2, r_3, \ldots\} = \{r_k\ ;\ k \in \mathbb{N}^*\}\,.$$

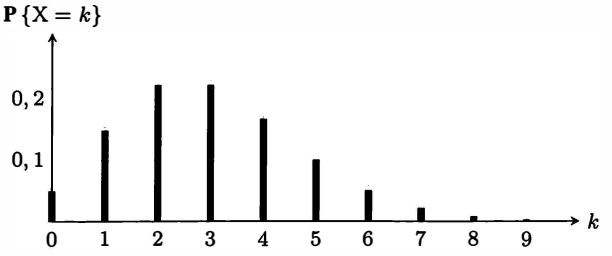
Munissons maintenant l'espace $\Omega = \mathbb{N}^*$ de la tribu $\mathcal{T} = \mathcal{P}(\mathbb{N}^*)$ et de la probabilité discrète $\mathbf{P}(\{k\}) = 1/2^k$. On définit enfin la variable aléatoire X sur Ω par $X(k) = r_k$; la loi de X est donc $\mathbf{P}\{X = r_k\} = 1/2^k$ et la fonction de répartition F de X admet, en chaque rationnel r_k, un saut de discontinuité

$$F(r_k) - F(r_k^-) = \frac{1}{2^k}.$$

Bien que F soit continue en presque tous points (en tous les irrationnels), ses points de discontinuité forment un ensemble dense, empêchant toute représentation graphique précise.

Histogrammes §74

Lorsque l'ensemble $X(\Omega)$ des valeurs prises par X est constitué de points isolés, il est pratique de représenter sur un graphe l'ensemble des ces valeurs par des barres d'abscisse x_k et hauteur $\mathbf{P}\{X = x_k\}$. Cette représentation « en histogramme » est en général bien plus parlante que celle de la fonction de répartition, c'est celle que nous privilégierons désormais.



§75 Variables aléatoires au sens large

Certaines situations nous obligent à ne définir une variable aléatoire X que sur une partie Ω' strictement incluse dans Ω, mais de probabilité 1 (c'est-à-dire que $N := \Omega \setminus \Omega'$ est négligeable). Cette indétermination sur un ensemble N de probabilité nulle n'est pas gênante, que ce soit pour des estimations de probabilité d'événements — comme $\mathbf{P}\{X \leqslant a\}$ — ou dans le calcul d'espérances mathématiques[3].

Exemple 6.14 (L'athlète et le saut de haies) Un athlète doit sauter une succession de haies, de plus en plus difficiles, et s'arrête au premier échec. On suppose que la probabilité qu'il réussisse à sauter la n-ième haie (s'il n'a pas été éliminé avant) est égale à $p_n = 1/n$, et on modélise l'épreuve par une succession infinie de haies. On demande la loi de X, numéro de la dernière haie franchie avec succès.

Les valeurs possibles prises par X sont les entiers naturels non nuls, à moins que *tous* les sauts soient réussis, auquel cas la valeur de X n'est pas définie. Nous allons montrer que la probabilité que X ne soit pas définie est nulle, en calculant tout simplement $\mathbf{P}\{X \in \mathbb{N}^*\}$.

Pour tout n entier, X prend la valeur n si et seulement si la n-ième haie est franchie (et toutes les haies précédentes aussi), et si la $(n+1)$-ième haie ne l'est pas. La probabilité de franchir les n premières haies est

$$1 \cdot \frac{1}{2} \cdot \frac{1}{3} \cdots \frac{1}{n} = \frac{1}{n!},$$

tandis que la probabilité d'échouer à la n-ième est

$$1 - \frac{1}{n+1} = \frac{n}{n+1},$$

ce qui mène à

$$\mathbf{P}\{X = n\} = \frac{n}{(n+1)!}.$$

Notamment, par σ-additivité,

$$\mathbf{P}\{X \in \mathbb{N}^*\} = \sum_{n=1}^{\infty} \frac{n}{(n+1)!} = \sum_{n=1}^{\infty} \left[\frac{n+1}{(n+1)!} - \frac{1}{(n+1)!} \right] = 1,$$

et X est donc définie presque sûrement[4].

---

(3). On peut faire le parallèle suivant. Si l'on ne s'intéresse qu'à des quantités de la forme $\int_A f$, changer la valeur de la fonction f en *un* point n'entraîne aucun changement, puisque la valeur de l'intégrale n'en sera pas modifiée ; c'est encore le cas si l'on ne modifie f qu'en un nombre fini ou dénombrable de points. De même (pour le lecteur familier avec l'intégrale de Lebesgue) si l'on ne modifie la valeur de f que sur un ensemble de mesure nulle. Du point de vue du calcul intégral, seule compte la classe d'équivalence de f à une égalité presque partout près. En langage probabiliste, lors du calcul d'une espérance, seule compte la classe d'équivalence d'une variable aléatoire à une égalité presque sûre près.

(4). On peut, alternativement, décider de la définir sur Ω tout entier en posant $X(\omega) = +\infty$ sur le complémentaire de $\{X \in \mathbb{N}^*\}$ (voir § 26).

6.2 LOIS CLASSIQUES

Loi uniforme §76

Si a et b sont des entiers, une variable aléatoire suit la **loi uniforme (discrète) sur** $[\![a\,;\,b]\!]$ si la probabilité qu'elle prenne comme valeur n'importe quel entier $a \leqslant k \leqslant b$ est la même :

$$\forall k \in \{a, a+1, \ldots, b\} \qquad \mathbf{P}\{\mathrm{X} = k\} = \frac{1}{b-a+1}.$$

On écrit alors « $\mathbf{X} \rightsquigarrow \mathscr{U}([\![a\,;\,b]\!])$ ». (Voir figure 6.1.)

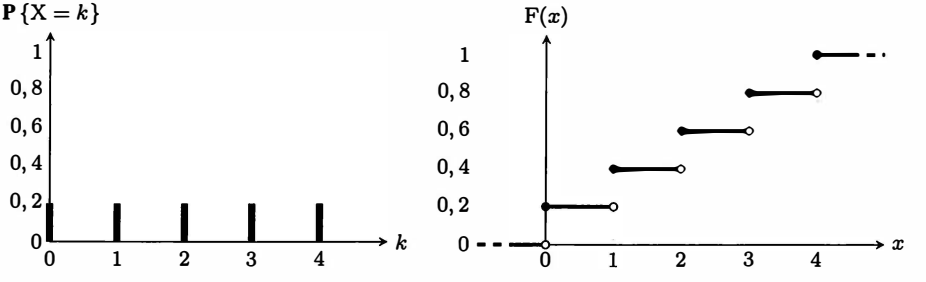


FIGURE 6.1 — *Histogramme et fonction de répartition de la loi uniforme sur* $[0\,;\,4]$.

Loi de Bernoulli §77

La loi la plus fondamentale des probabilités est la loi dite de Bernoulli, et modélise une expérience n'ayant que deux issues possibles, baptisées conventionnellement **succès** et **échec**. Une telle expérience est appelée **épreuve de Bernoulli**. Soit p un réel, $0 < p < 1$, et associons au succès la probabilité p, à l'échec la probabilité $1-p$; on peut alors créer une variable aléatoire X qui prend la valeur 1 en cas de succès, la valeur 0 en cas d'échec. Le succès est alors l'événement $\{\mathrm{X} = 1\}$, tandis que l'échec est l'événement $\{\mathrm{X} = 0\}$.

Plus généralement, une variable aléatoire X suit la **loi de Bernoulli de paramètre** p, et on écrit « $\mathbf{X} \rightsquigarrow \mathscr{B}(p)$ » lorsque $\mathrm{X}(\Omega) = \{0, 1\}$ et

$$\mathbf{P}\{\mathrm{X} = 1\} = p \qquad \text{et} \qquad \mathbf{P}\{\mathrm{X} = 0\} = 1 - p.$$

Remarque 6.15 Par extension, on parle parfois de variable de Bernoulli pour indiquer une variable aléatoire ne possédant que deux valeurs possibles (comme -1 et 1).

Si A est un événement de Ω, la **fonction indicatrice de A**, notée $\mathbf{1}_{\mathrm{A}}$, est la variable aléatoire définie par

$$\forall \omega \in \Omega \qquad \mathbf{1}_{\mathrm{A}}(\omega) = \begin{cases} 1 & \text{si } \omega \in \mathrm{A} \\ 0 & \text{sinon.} \end{cases}$$

Une telle variable aléatoire suit la loi de Bernoulli de paramètre $p = \mathbf{P}(\mathrm{A})$.

Réciproquement, toute variable de Bernoulli est évidemment une fonction indicatrice : celle de l'événement $\mathrm{A} = \{\mathrm{X} = 1\}$ — et que cette tautologie nous soit pardonnée.

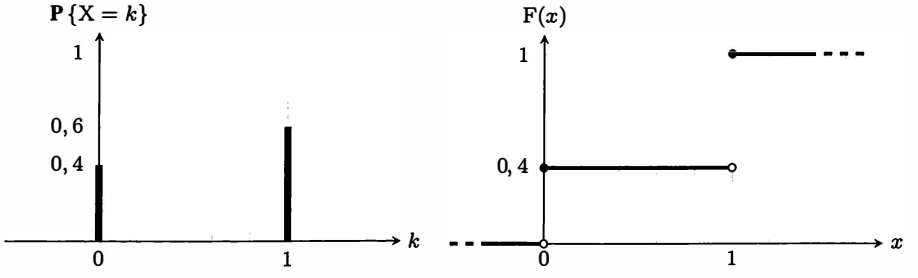


FIGURE 6.2 — *Histogramme et fonction de répartition de la loi de Bernoulli de paramètre* $p = 0,6$.

Exemple 6.16 Soit Y une variable aléatoire à valeurs dans $\mathbb{R}$. On définit la variable aléatoire $X = \mathbf{1}_{\{Y \geqslant 0\}}$, qui vaut 1 lorsque Y est positive et 0 sinon :

$$\forall \omega \in \Omega \qquad X(\omega) = \begin{cases} 1 & \text{si } Y(\omega) \geqslant 0, \\ 0 & \text{si } Y(\omega) < 0. \end{cases}$$

La variable X suit la loi de Bernoulli de paramètre $p = \mathbf{P}\{Y \geqslant 0\}$.

■ **Proposition 6.17** *Soient* A *et* B *deux événements de* Ω. *Alors*

$$\mathbf{1}_{A \cap B} = \mathbf{1}_A \cdot \mathbf{1}_B \qquad \textit{et} \qquad 1 - \mathbf{1}_{A \cup B} = (1 - \mathbf{1}_A) \cdot (1 - \mathbf{1}_B)\,.$$

La démonstration est immédiate et montre que : le produit de deux variables de Bernoulli est encore une variable de Bernoulli.

§ 78 **Loi binomiale**

On modélise une expérience consistant en la réalisation de n épreuves de Bernoulli indépendantes de même loi $\mathscr{B}(p)$; on note X le nombre total de succès. Par exemple, on peut lancer n fois de suite une pièce déséquilibrée, donnant *pile* avec la probabilité p, et compter le nombre de *pile* obtenus au cours des n lancers.

La variable X prend toutes les valeurs comprises entre 0 et n ; la probabilité qu'elle soit égale à un entier k est égale à la probabilité $p^k\,(1-p)^{n-k}$ d'obtenir k *pile* et $n-k$ *face*, multipliée par le nombre de façons d'obtenir cette configuration, c'est-à-dire le coefficient binomial « k parmi n »

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

Ainsi, $X(\Omega) = \{0, 1, 2, \ldots, n\}$ et

$$\forall k \in [\![0\,;\,n]\!] \qquad \mathbf{P}\{X = k\} = \binom{n}{k} p^k\,(1-p)^{n-k}.$$

On dit que variable aléatoire X suit la **loi binomiale de paramètres n et p**, et on écrit « $\mathbf{X} \rightsquigarrow \mathscr{B}\,(n\,;\,p)$ ». Des histogrammes de la loi binomiale sont présentés sur la figure 6.3 page ci-contre.

§ 79 **Loi géométrique**

Considérons de nouveau une succession de tirages à *pile* ou *face*, supposés indépendants et de même loi de Bernoulli $\mathscr{B}(p)$; cette fois-ci, on se donne une infinité dénombrable de tels tirages. On note X le numéro du premier tirage dont l'issue est *pile*. Cette quantité n'étant pas définie lorsque tous les tirages ont pour issue

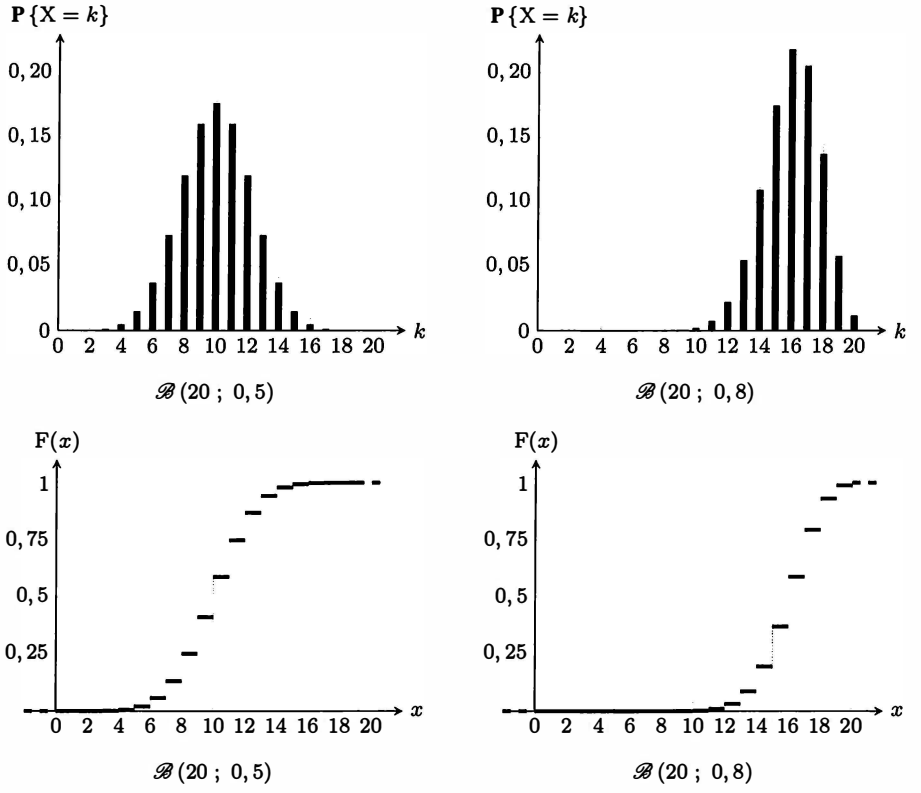


FIGURE 6.3 — En haut : *histogrammes de la loi binomiale pour une même valeur* $n = 20$ *et pour des paramètres* $p = 0,5$ *et* $p = 0,8$. *On notera la symétrie du premier et la forte dissymétrie du second.* En bas : *fonctions de répartition des lois binomiales précédentes.*

face[5], on pose, conformément à l'intuition, $X(\omega) = +\infty$ pour un tel ω. Si $n \geqslant 1$ est un entier, le rang du premier *pile* est n si et seulement si on a obtenu *face* à chacun des $n-1$ premiers tirages, ainsi que *pile* au n-ième tirage. Compte tenu de l'indépendance des tirages, la probabilité de cette configuration est

$$\mathbf{P}\{X = n\} = (1-p)^{n-1}\, p.$$

La probabilité que X prenne une valeur finie est

$$\mathbf{P}\{X < +\infty\} = \sum_{n=1}^{\infty} (1-p)^{n-1}\, p = \frac{p}{1-(1-p)} = 1,$$

ce qui montre que l'événement $\{X = +\infty\}$ est négligeable, et que X est bien une variable aléatoire (au sens large).

On dit que X suit la **loi géométrique de paramètre p**, ou encore **loi du premier succès**, et on écrit « $\mathbf{X} \rightsquigarrow \mathcal{G}(p)$ ».

(5). On rappelle qu'un tel événement fait partie de l'univers Ω, même s'il est évidemment de probabilité nulle.

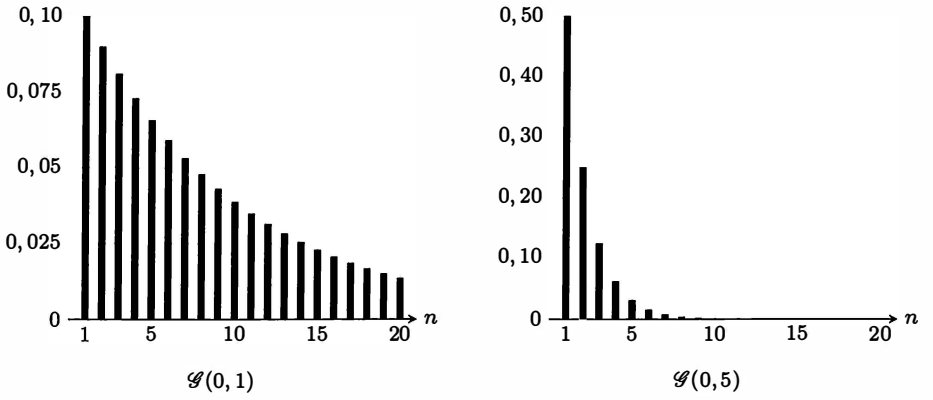


FIGURE 6.4 — *Histogrammes de la loi géométrique $\mathcal{G}(p)$ pour un paramètre $p = 0,1$ (à gauche) et $p = 0,5$ (à droite). L'échelle verticale n'est pas la même entre les deux graphes.*

Remarque 6.18 (Loi du nombre d'échecs) On appelle également loi géométrique de paramètre p la **loi du nombre d'échecs (avant le premier succès)**, qui est le nombre de *face* obtenus avant le premier pile. La variable aléatoire Y comptant le nombre de *face* est donc égale à $X - 1$, et elle a pour loi

$$\forall n \in \mathbb{N} \qquad \mathbf{P}\{Y = n\} = (1-p)^n p.$$

Nous la désignerons toujours comme la « loi du nombre d'échecs », afin d'éviter les confusions, et nous la noterons $\mathcal{G}_{\mathbb{N}}(p)$. La loi géométrique (du premier succès) est parfois notée $\mathcal{G}_{\mathbb{N}^*}(p)$.

Les lois géométriques jouissent d'une propriété commune, appelée propriété de **non vieillissement** (on dit aussi qu'elles sont **sans mémoire**). La probabilité que X soit strictement supérieur à un entier k est

$$\mathbf{P}\{X > k\} = \sum_{n=k+1}^{\infty} (1-p)^{n-1} p = (1-p)^k.$$

Notamment, si k, ℓ sont des entiers,

$$\mathbf{P}\{X > k+\ell \,|\, X > \ell\} = \frac{\mathbf{P}\{X > k+\ell, X > \ell\}}{\mathbf{P}\{X > \ell\}} = \frac{\mathbf{P}\{X > k+\ell\}}{\mathbf{P}\{X > \ell\}}$$
$$= \frac{(1-p)^{k+\ell}}{(1-p)^{\ell}} = (1-p)^k = \mathbf{P}\{X > k\}.$$

Exemple 6.19 Dans le cas d'attente d'un tirage gagnant au cours d'une partie (infinie) de *pile* ou *face*, la probabilité de devoir attendre (strictement) au moins 5 lancers avant le premier succès est la même que la probabilité de devoir attendre au moins 15 lancers lorsque l'on n'a eu aucun succès pendant les 10 premiers.

Cette propriété d'absence de mémoire est en fait caractéristique des lois géométriques au sein des lois discrètes à valeurs dans $\mathbb{N}$ (voir exercice 6.5).

§80 Loi de Poisson

La dernière grande loi discrète classique provient du type de modélisation suivant.

On considère un événement se produisant *rarement*, mais sur un grand nombre d'épreuves. Typiquement, on peut penser à un grand nombre d'atomes instables, pouvant se désintégrer (indépendamment les uns des autres) avec émission d'un rayonnement, que l'on détecte au moyen d'un capteur. Le nombre total d'atomes étant de l'ordre de 10^{23}, on peut le considérer comme

constant même après quelques désintégrations. On fait l'hypothèse que la probabilité de détection d'un rayonnement durant un intervalle de temps Δt petit est proportionnelle à Δt ; on la prend de la forme $\alpha\,\Delta t$. On se pose la question de déterminer la loi du nombre X de désintégrations que l'on va mesurer durant un temps T fixé.

La modélisation consiste à découper l'intervalle T en un nombre n de petits intervalles de durée $\Delta t = \mathrm{T}/n$. On va supposer Δt tellement petit que la probabilité de mesurer *deux* désintégrations durant cet intervalle est complètement négligeable. Le problème est alors équivalent à celui d'une répétition de n épreuves de Bernoulli, de probabilité de succès égale à $p(n) = \alpha\mathrm{T}/n$. Ainsi,

$$\mathbf{P}\{\mathrm{X} = k\} \approx \binom{n}{k} p(n)^k \left(1 - p(n)\right)^{n-k}.$$

On obtiendra une estimation d'autant plus précise de $\mathbf{P}\{\mathrm{X} = k\}$ que l'on prendra n grand.

Or, si l'on maintient k fixé, on a les équivalents suivants quand n tend vers l'infini :

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} \underset{n\to\infty}{\sim} \frac{n^k}{k!}$$

et

$$\begin{aligned}\left(1 - p(n)\right)^{n-k} &= \exp\Big\{(n-k)\,\ln\left(1-p(n)\right)\Big\} = \exp\Big\{-(n-k)\left(p(n) + \mathrm{o}\left(p(n)\right)\right)\Big\} \\ &= \exp\left\{-\alpha\mathrm{T} + \mathrm{o}(1)\right\} \xrightarrow[n\to\infty]{} \mathrm{e}^{-\alpha\mathrm{T}}\end{aligned}$$

ce qui nous donne l'expression suivante, dans la limite $n \to \infty$:

$$\mathbf{P}\{\mathrm{X} = k\} \approx \mathrm{e}^{-\alpha\mathrm{T}}\,\frac{(\alpha\mathrm{T})^k}{k!}.$$

Il est immédiat de vérifier que la somme de ces valeurs, pour $n = 0, 1, 2, \ldots$, est bien égale à 1, ce qui justifie la définition suivante.

Définition 6.20 Soit λ un réel, $\lambda > 0$. On dit qu'une variable aléatoire suit la **loi de Poisson de paramètre $\boldsymbol{\lambda}$**, et on écrit « $\mathbf{X} \rightsquigarrow \mathscr{P}(\boldsymbol{\lambda})$ », si $\mathrm{X}(\Omega) = \mathbb{N}$ et si

$$\forall k \in \mathbb{N} \qquad \mathbf{P}\{\mathrm{X} = k\} = \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}.$$

Les lois de Poisson permettent de modéliser de nombreuses situations où l'on découpe un intervalle de temps en intervalles indéfiniment petits, comme précédemment, pour déterminer le nombre d'événements d'un certain type se produisant dans cet intervalle[(6)]. Citons, outre les problèmes de désintégration, quelques cas classiques : ≏6.1

— le nombre de connexions à un serveur web durant un intervalle de temps T ;
— le nombre de clients se présentant à une caisse de supermarché durant T ;
— le nombre de passagers se présentant à une station de métro durant T ;
— le nombre de coquilles typographiques par page d'un livre de probabilités. ≏6.2

Comme nous l'avons vu, la loi de Poisson a été obtenue comme une « loi limite » ; en conséquence, elle est souvent utilisée comme approximation d'une loi exacte (binomiale), ce qui permet d'effectuer numériquement certains calculs ; voir le chapitre 22 et spécialement l'exercice 22.3 page 464.

Le calcul effectué précédemment se généralise immédiatement au cas où p est une fonction de n telle que $n\,p(n)$ admette une limite λ en l'infini :

(6). Chaque événement, rapporté à une section de temps Δt, est donc « rare » ; la loi de Poisson est parfois appelée « loi des événements rares ». Pour l'anecdote, la première utilisation répertoriée de la loi de Poisson se trouve dans un opuscule d'une soixantaine de pages, *Das Gesetz der kleinen Zahlen* (*La loi des petits nombres*) de Władisław Bortkiewicz, parue en 1898, où il effectue une étude statistique du nombre de décès annuels par ruade de cheval dans la cavalerie prussienne. Il remarque également, dans cet ouvrage, que des événements rares de probabilités variables sont encore décrits par des lois de Poisson, ce qui est une conséquence du théorème 8.13 page 183.

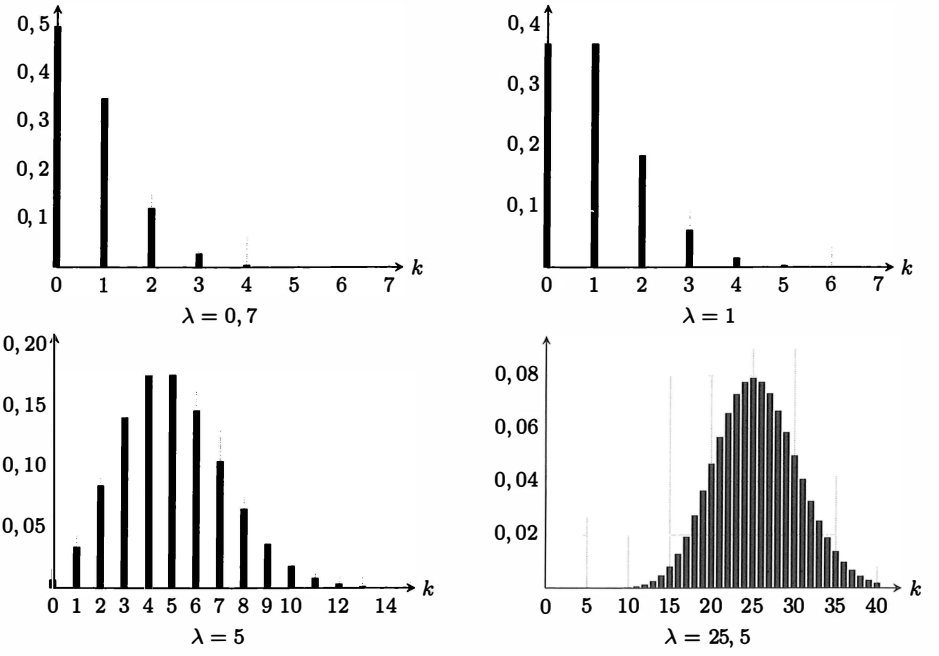


FIGURE 6.5 — *Histogrammes de la loi de Poisson pour différentes valeurs de λ. Si λ est entier, alors les valeurs $k = \lambda - 1$ et $k = \lambda$ sont des maxima de probabilité (des* modes*). Si λ est non entier, le seul mode est $\lfloor \lambda \rfloor$.*

■ **Théorème 6.21 (Limite d'épreuves de Bernoulli)** *Soit $(p_n)_{n \geqslant 1}$ une suite de réels tels que $0 < p_n < 1$ pour tout entier n, et vérifiant*

$$\lim_{n \to \infty} n p_n = \lambda.$$

Pour chaque n, on répète n fois une épreuve de Bernoulli de probabilité p_n, et on note X_n la variable aléatoire égale au nombre de succès obtenus, qui suit donc la loi binomiale $\mathscr{B}(n\,;\,p_n)$. Alors, pour tout $k \in \mathbb{N}$,

$$\lim_{n \to \infty} \mathbf{P}\{X_n = k\} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

(Plus tard, on dira que la suite $(X_n)_{n \geqslant 1}$ *converge en loi* vers la loi de Poisson de paramètre λ.)

Terminons ce paragraphe en ajoutant que les lois de Poisson entretiennent des liens très serrés avec les lois exponentielles[7].

(7). Qui sont les lois continues qui gouvernent, par exemple, les désintégrations radioactives que nous avons pris comme exemple pour modéliser la loi $\mathscr{P}(\lambda)$ au début du paragraphe : les intervalles de temps séparant deux désintégrations sont régis par une telle loi exponentielle, ce qui traduit une propriété de non vieillissement. Une loi de Poisson peut donc être définie comme celle du nombre d'événements durant un certain intervalle de temps, lorsque les intervalles entre ces événements suivent une loi exponentielle. Cette propriété sera utilisée, notamment, pour simuler informatiquement une loi de Poisson (§ 369) et pour définir les processus de Poisson (chapitre 30).

Tirages avec et sans remise. Loi hypergéométrique §81

La loi binomiale $\mathscr{B}(n\,;\,p)$ modélise la situation suivante : on considère une urne contenant des boules blanches et noires, les noires étant en proportion p. On effectue une succession de n tirages indépendants, le protocole étant, une fois que l'on a constaté la couleur de la boule tirée, de la remettre dans l'urne (et de mélanger pour garantir l'indépendance des tirages). On parle de **tirage avec remise**. Le nombre X de boules noires tirées (ou nombre de succès) suit la loi binomiale $\mathscr{B}(n\,;\,p)$.

Changeons maintenant le protocole : une fois que l'on a tiré une boule, on ne la remet pas dans l'urne, mais on la conserve par devers soi (**tirage sans remise**). On demande alors la loi de X, nombre de boules noires obtenues à l'issue des n tirages. Le nombre initial de boules noires est $p\mathrm{N}$, et celui de boules blanche $(1-p)\mathrm{N} = q\mathrm{N}$, en posant $q = 1-p$. À chaque boule retirée, la proportion de boules blanches et noire est modifiée. Supposons que le premier tirage amène une boule noire. Alors, il reste $p\mathrm{N}-1$ boules noires sur les $\mathrm{N}-1$ boules, la nouvelle proportion de boules noires est alors $p' = (p\mathrm{N}-1)/(\mathrm{N}-1)$. Si le premier tirage avait amené une boule blanche, la nouvelle proportion aurait été $p\mathrm{N}/(\mathrm{N}-1)$. On pourrait alors continuer le raisonnement en fonction du résultat du deuxième tirage, et ainsi de suite, en construisant un *arbre de probabilités*. Le lecteur concevra sans mal que l'on ne parviendrait à la loi de X qu'après des calculs longs et possiblement douteux.

Il est plus simple d'effectuer un raisonnement combinatoire pour obtenir la loi de X. On définit comme univers Ω l'ensemble des combinaisons de n boules parmi N, ce qui revient à dire que l'ordre des tirages n'est jamais pris en compte ; Ω est de cardinal

$$\operatorname{Card}\Omega = \binom{\mathrm{N}}{n}$$

et on le munit de la loi uniforme (tous les tirages sont équiprobables).

Dans un premier temps, déterminons les valeurs possibles prises par X. Tout d'abord, il ne saurait y avoir plus de n succès (nombre total de tirages), et pas plus de $p\mathrm{N}$ succès (nombre de boules noires dans l'urne). De plus, si n dépasse le nombre $q\mathrm{N}$ de boules blanches, il ne saurait y avoir moins de $n - q\mathrm{N}$ boules noires (cas où toutes les boules blanches ont été tirées). Tous les cas limites discutés pouvant manifestement être atteints, on a donc

$$\mathrm{X}(\Omega) = [\max(0, n-q\mathrm{N})\,;\,\min(n, p\mathrm{N})]\,.$$

Déterminons maintenant $\mathbf{P}\{\mathrm{X}=k\}$, lorsque k est un entier de $\mathrm{X}(\Omega)$. Un tirage amenant la valeur $\mathrm{X}=k$ est tel que k boules noires ont été tirées, ainsi que $n-k$ boules blanches. Le nombre de façons de choisir les boules noires est $\binom{\mathrm{N}p}{k}$, le nombre de façons de choisir les boules blanches est $\binom{\mathrm{N}q}{n-k}$. Chacun des choix étant affecté de la probabilité uniforme $1/\binom{\mathrm{N}}{n}$, on a donc

$$\mathbf{P}\{\mathrm{X}=k\} = \frac{\binom{\mathrm{N}p}{k}\binom{\mathrm{N}q}{n-k}}{\binom{\mathrm{N}}{n}}.$$

On dit alors que X suit la **loi hypergéométrique de paramètres N, n et p**, et on écrit « $\mathrm{X} \rightsquigarrow \mathscr{H}(\mathrm{N}, n, p)$ ». ⇗6.3

Remarque 6.22 On peut, alternativement, choisir pour univers l'ensemble Ω' de tous les tirages successifs (sans remise) de n boules dans une urne à N éléments ; on a dans ce cas

$$\operatorname{Card}\Omega' = \mathrm{N}(\mathrm{N}-1)(\mathrm{N}-2)\cdots(\mathrm{N}-n+1).$$

Chaque élément de Ω est en correspondance avec exactement $n!$ éléments de Ω'.

Remarque 6.23 (Formule de Vandermonde) De la relation $\sum_{k=0}^{n} \mathbf{P}\{X = k\} = 1$, on déduit la **formule de Vandermonde**

$$\sum_{k=0}^{n} \binom{Np}{k} \cdot \binom{Nq}{n-k} = \binom{N}{n}.$$

En notant $a = Np$ et $b = Nq$, cette relation se récrit sous la forme plus usuelle

$$\sum_{k=0}^{n} \binom{a}{k} \cdot \binom{b}{n-k} = \binom{a+b}{n}.$$

§ 82 Loi de l'image $\varphi(X)$ d'une variable aléatoire

Soit X une variable aléatoire discrète, d'image $X(\Omega) = \{x_k \,;\, k \in K\}$, et soit $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction, définie au moins sur $X(\Omega)$. On peut définir une nouvelle variable aléatoire Y par la formule

$$\forall \omega \in \Omega \qquad Y(\omega) := \varphi\big(X(\omega)\big).$$

On la note commodément $Y = \varphi(X)$. L'ensemble des valeurs prises par Y est au plus dénombrable et s'écrit $\{y_k \,;\, k \in K\}$, en posant $y_k := \varphi(x_k)$. La loi de Y se détermine, de manière théorique, en cherchant pour une valeur y_k donnée l'ensemble des antécédents de y_k par φ, ou au moins ceux qui sont également dans $X(\Omega)$ (les autres ne contribuent pas) :

$$\mathbf{P}\{Y = y_k\} = \sum_{\substack{i \in K \\ \varphi(x_i) = y_k}} \mathbf{P}\{X = x_i\} = \sum_{x \in \varphi^{-1}(\{y_k\})} \mathbf{P}\{X = x\}.$$

Cette loi peut être difficile à déterminer en pratique lorsque φ est fortement non injective.

Exemple 6.24 (Loi de X^2) Soit X une variable aléatoire de loi uniforme sur l'ensemble $X(\Omega) = \{-1, 0, 1, 2\}$, et posons $Y = X^2$. Alors $Y(\Omega) = \{0, 1, 4\}$ et

$$\mathbf{P}\{Y = 0\} = \mathbf{P}\{X = 0\} = \tfrac{1}{4} \qquad \mathbf{P}\{Y = 1\} = \mathbf{P}\{X = -1\} + \mathbf{P}\{X = 1\} = \tfrac{1}{2}$$
$$\mathbf{P}\{Y = 4\} = \mathbf{P}\{X = 2\} = \tfrac{1}{4}.$$

Lorsque φ est injective sur $X(\Omega)$, elle induit une bijection de $X(\Omega)$ sur son image $Y(\Omega)$ — bijection que l'on notera encore φ — et

$$\forall y \in Y(\Omega) \qquad \mathbf{P}\{Y = y\} = \mathbf{P}\{\varphi(X) = y\} = \mathbf{P}\big\{X = \varphi^{-1}(y)\big\}.$$

Les difficultés inhérentes à l'établissement de la loi de Y justifient l'intérêt d'un théorème du chapitre suivant, appelé *théorème de transfert*, et qui permet de calculer l'espérance de $\varphi(X)$ sans avoir à expliciter sa loi.

EXERCICES

♦ ***Exercice 6.1*** Soit $\lambda > 0$. Soit X une variable aléatoire suivant une loi de Poisson $\mathscr{P}(\lambda)$ de paramètre λ. Montrer que

$$\mathbf{P}\{\text{« X est impair »}\} < \mathbf{P}\{\text{« X est pair »}\}.$$

Étudier une propriété analogue lorsque X suit la loi géométrique de paramètre $p \in]0\,;1[$ (on pourra inspecter les histogrammes de la figure 6.4 page 136 pour deviner le résultat avant d'effectuer le calcul).

♦ ***Exercice 6.2 (Modes d'une loi de Poisson)*** Soit $\lambda > 0$. Soit X une variable aléatoire suivant la loi de Poisson de paramètre λ.

i) On suppose $\lambda \notin \mathbb{N}$. Montrer que $\mathbf{P}\{X = k\}$ est maximale en $k_0 = \lfloor \lambda \rfloor$. L'entier k_0 est l'*unique mode* de la loi $\mathscr{P}(\lambda)$.

ii) On suppose $\lambda \in \mathbb{N}^*$. Montrer que $\mathbf{P}\{X = \lambda - 1\} = \mathbf{P}\{X = \lambda\}$ est maximale. Les entiers $k_0 = \lambda - 1$ et $k_1 = \lambda$ sont les deux modes de $\mathscr{P}(\lambda)$.

♦ ***Exercice 6.3*** Soient N un entier, p un réel de $]0\,;1[$ tel que Np soit entier. On pose $q = 1-p$ et on note

$$h(N, n, p, k) = \frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}},$$

probabilité qu'une variable aléatoire X suivant la loi hypergéométrique $\mathscr{H}(N, n, p)$ prenne la valeur k. Montrer que, à k et n fixés, la quantité $h(N, n, p, k)$ admet une limite quand N tend vers l'infini. Reconnaître cette valeur et interpréter le résultat. *(Dans les calculs, on pourra dans un premiers temps considérer que Np et Nq sont toujours des entiers, ce qui est évidemment faux si p reste fixe et $N \to \infty$. Dans un second temps, on pourra remplacer la valeur de p par $p_N = \lfloor Np \rfloor / N$.)*

♦ ***Exercice 6.4 (Maximum d'un échantillon d'une loi uniforme)***
Le jour de l'examen de fin d'année, n élèves n'ont pas assez soigné la présentation de leur copie. Le correcteur, plutôt que de s'escrimer à lire d'infâmes brouillons, décide de noter au hasard, de manière indépendante, les n copies, en leur attribuant une note entière, au hasard, entre 0 et 20. On note X_n la variable aléatoire égale à la meilleure note du groupe.

i) Soit $k \in [\![0\,;20]\!]$. Calculer la probabilité que toutes les notes soient inférieures ou égales à k.

ii) (a) Calculer $\mathbf{P}\{X_n < k \,|\, X_n \leqslant k\}$.

(b) En déduire $\mathbf{P}\{X_n = k \,|\, X_n \leqslant k\}$.

iii) Calculer $\mathbf{P}\{X_n = k\}$ et déterminer sa limite quand $n \to \infty$. Interpréter.

♦ ***Exercice 6.5 (Lois sans vieillissement)*** Soit X une variable aléatoire à valeurs entières strictement positives, telle que pour tout $n \geqslant 1$, on ait $\mathbf{P}\{X = k\} \neq 0$. On suppose de plus que, pour tous $k, \ell \in \mathbb{N}$,

$$\mathbf{P}\{X > k + \ell \,|\, X > \ell\} = \mathbf{P}\{X > k\}.$$

Montrer que X suit une loi géométrique.

SOLUTIONS

♦ **Solution 6.1** Un calcul immédiat donne

$$\mathbf{P}\{\text{« X est pair »}\} = \sum_{n=0}^{\infty} \mathbf{P}\{X = 2n\} = \operatorname{ch}(\lambda)\, e^{-\lambda}$$

$$\mathbf{P}\{\text{« X est impair »}\} = \sum_{n=0}^{\infty} \mathbf{P}\{X = 2n+1\} = \operatorname{sh}(\lambda)\, e^{-\lambda}.$$

La différence des deux probabilités est $\mathbf{P}\{\text{« X est pair »}\} - \mathbf{P}\{\text{« X est impair »}\} = e^{-2\lambda} > 0$.

Si X suit la loi géométrique de paramètre $p \in]0\,;1[$, alors (en notant $q = 1-p$), la loi de X est donnée par $\mathbf{P}\{X = n\} = pq^{n-1}$ pour tout $n \geqslant 1$;

$$\mathbf{P}\{\text{« X est pair »}\} = \sum_{n=1}^{\infty} \mathbf{P}\{X = 2n\} = \sum_{n=1}^{\infty} p\, q^{2n-1} = \frac{pq}{(1-q^2)} = \frac{1-p}{2-p}$$

tandis que

$$\mathbf{P}\{\text{« X est impair »}\} = \sum_{n=0}^{\infty} \mathbf{P}\{X = 2n+1\} = \sum_{n=0}^{\infty} p\, q^{2n} = \frac{p}{1-q^2} = \frac{1}{2-p} = \frac{\mathbf{P}\{\text{« X est pair »}\}}{1-p},$$

donc cette fois-ci il y a plus de chances d'obtenir un résultat impair qu'un résultat pair.

♦ **Solution 6.3** On utilise les équivalents

$$\binom{Np}{k} \underset{N\to\infty}{\sim} \frac{(Np)^k}{k!} \qquad \text{et} \qquad \binom{Nq}{n-k} \underset{N\to\infty}{\sim} \frac{(Nq)^{n-k}}{(n-k)!}$$

pour obtenir que $h(N,n,p,k) \to b(n,p,k) := \binom{n}{k} p^k (1-p)^{n-k}$, probabilité qu'une variable aléatoire suivant la loi binomiale $\mathscr{B}(n\,;\,p)$ prenne la valeur k.

Si l'on remplace dans ces calculs p par $p_N := \lfloor Np \rfloor / N$, on a encore

$$\binom{Np_N}{k} \underset{N\to\infty}{\sim} \frac{(Np_N)^k}{k!} \underset{N\to\infty}{\sim} \frac{(Np)^k}{k!}$$

(et de même pour le second terme), et le résultat est inchangé. On dira que la loi hypergéométrique $\mathscr{H}(N,n,p_n)$ *converge en loi* vers la loi binomiale $\mathscr{B}(n\,;\,p)$.

♦ **Solution 6.4**

i) On a $\mathbf{P}\{X_1 \leqslant k, \cdots, X_n \leqslant k\} = \left(\frac{k+1}{21}\right)^n$.

ii) (a) On a $\mathbf{P}\{X_n < k \,|\, X_n \leqslant k\} = \dfrac{\mathbf{P}\{X_n < k, X_n \leqslant k\}}{\mathbf{P}\{X_n \leqslant k\}} = \dfrac{\mathbf{P}\{X_n \leqslant k-1\}}{\mathbf{P}\{X_n \leqslant k\}} = \left(\dfrac{k}{k+1}\right)^n$.

(b) On en déduit que $\mathbf{P}\{X_n = k \,|\, X_n \leqslant k\} = 1 - \mathbf{P}\{X_n < k \,|\, X_n \leqslant k\} = 1 - \left(\dfrac{k}{k+1}\right)^n$.

iii) Par définition, on a $\mathbf{P}\{X_n = k\} = \mathbf{P}\{X_n = k \,|\, X_n \leqslant k\} \cdot \mathbf{P}\{X_n \leqslant k\}$, ce qui donne donc

$$\mathbf{P}\{X_n = k\} = \left(\frac{k+1}{21}\right)^n \left[1 - \left(\frac{k}{k+1}\right)^n\right] = \left(\frac{k+1}{21}\right)^n - \left(\frac{k}{21}\right)^n.$$

On remarque que

$$\lim_{n\to\infty} \mathbf{P}\{X_n = k\} = \begin{cases} 0 & \text{si } k \leqslant 19, \\ 1 & \text{si } k = 20. \end{cases}$$

Si le nombre de copies est grand, on est presque sûr que la note maximale va être 20.

♦ **Solution 6.5** Notons $p := \mathbf{P}\{X = 1\}$ et $q := 1 - p = \mathbf{P}\{X > 1\}$. Pour tout $\ell \geqslant 1$, on a

$$q = \mathbf{P}\{X > \ell + 1 \,|\, X > \ell\} = \frac{\mathbf{P}\{X > \ell+1, X > \ell\}}{\mathbf{P}\{X > \ell\}} = \frac{\mathbf{P}\{X > \ell+1\}}{\mathbf{P}\{X > \ell\}}$$

ce qui montre que la suite de terme général $\mathbf{P}\{X > \ell\}$ est géométrique ; plus précisément, on a, pour tout $\ell \geqslant 1$, $\mathbf{P}\{X > \ell\} = q^\ell$. On en déduit que, pour tout $k \geqslant 1$,

$$\mathbf{P}\{X = k\} = \mathbf{P}\{X > k-1\} - \mathbf{P}\{X > k\} = q^{k-1} - q^k = q^{k-1}(1-q) = p(1-p)^{k-1}.$$

Ainsi, $X \rightsquigarrow \mathscr{G}(p)$.

Chapitre 7

Espérance d'une variable aléatoire discrète

Voici enfin les notions d'espérance, de moments, de variance, ainsi que les premières grandes inégalités (Bienaymé-Tchebychev et Cauchy-Schwarz) qui jouent un rôle central dans toute la théorie des probabilités. Dans le cas d'une variable discrète, ces notions ne présentent pas de difficulté théorique, et les diverses propriétés que nous pouvons démontrer rigoureusement dans ce cas particulier sont en fait des propriétés qui restent vraies en toute généralité, mais qui sont alors plus difficiles à démontrer.

Nous introduisons également les fonctions génératrices associées aux variables à valeurs entières, naturelles dans les problèmes de comptage.

Pour terminer, nous appliquons les outils présentés à un problème bien connu, celui du collectionneur de vignettes.

7.1 ESPÉRANCE

Espérance

§ 83

La notion d'*espérance mathématique* d'une variable aléatoire généralise celle de *moyenne pondérée* que même les moins matheux connaissent — qui, enfant, n'a pas calculé sa moyenne trimestrielle en tenant en compte des coefficients de chaque matière ? Soit X une variable aléatoire **simple**, c'est-à-dire ne prenant qu'un nombre fini de valeurs $x_1, \dots, x_n$ avec des probabilités respectives $p_k = \mathbf{P}\{\mathrm{X} = x_k\}$, où bien entendu $p_1 + p_2 + \cdots + p_n = 1$. La moyenne pondérée de ces valeurs, notée $\mathbf{E}(\mathrm{X})$ ou $\mathbf{E}\mathrm{X}$, est appelée **espérance (mathématique)** de X :

$$\mathbf{E}\mathrm{X} = \mathbf{E}(\mathrm{X}) = \sum_{k=1}^{n} x_k\, p_k.$$

Exemple 7.1 Considérons un univers Ω et une partition finie $\Omega = \mathrm{A} + \mathrm{B} + \mathrm{C}$ de cet univers. On suppose que X prend la valeur a sur A, la valeur b sur B et la valeur c sur C. On suppose de plus que $\mathbf{P}(\mathrm{A}) = 1/3$, $\mathbf{P}(\mathrm{B}) = 1/2$ et $\mathbf{P}(\mathrm{C}) = 1/6$. Alors

$$\mathbf{E}(\mathrm{X}) = \frac{a}{3} + \frac{b}{2} + \frac{c}{6}.$$

A	B	C
$\{\mathrm{X} = a\}$	$\{\mathrm{X} = b\}$	$\{\mathrm{X} = c\}$

Exemple 7.2 (Espérance d'une fonction indicatrice) Si A est un événement et $\mathbf{1}_{\mathrm{A}}$ la fonction indicatrice associée, alors $\mathbf{E}(\mathbf{1}_{\mathrm{A}}) = 0\,\mathbf{P}\{\mathbf{1}_{\mathrm{A}} = 0\} + 1\,\mathbf{P}\{\mathbf{1}_{\mathrm{A}} = 1\} = \mathbf{P}(\mathrm{A})$, puisque $\{\mathbf{1}_{\mathrm{A}} = 1\}$ et A représentent le même ensemble.

Cette notion se généralise aux variables discrètes prenant un nombre infini, mais dénombrable, de valeurs.

Définition 7.3 Soit X une variable aléatoire discrète, prenant un ensemble dénombrable de valeurs $\{x_k\ ;\ k \in \mathbb{N}\}$. Si la série $\sum |x_k|\ \mathbf{P}\{\mathrm{X} = x_k\}$ converge, on dit que X **admet une espérance** ou : **est d'espérance finie** ; dans ce cas, son **espérance (mathématique)**, notée $\mathbf{E}(\mathrm{X})$ ou $\mathbf{E}\mathrm{X}$, est la somme de cette série :

$$\mathbf{E}\mathrm{X} = \mathbf{E}(\mathrm{X}) = \sum_{k=0}^{\infty} x_k\,\mathbf{P}\{\mathrm{X} = x_k\}\,.$$

(On rappelle que toute série absolument convergente est convergente, cf. A10, page 670.)

Si la série $\sum x_k\,\mathbf{P}\{\mathrm{X} = x_k\}$ converge, sans toutefois être absolument convergente, alors sa somme n'a pas de signification solide : changer l'ordre des termes peut changer la valeur de la somme, voire la nature de la série[1]. C'est pourquoi la définition précédente est si restrictive : la convergence absolue de la série autorise la sommation dans n'importe quel ordre. Autrement dit, si l'on numérote les valeurs prises par X de deux manière différentes :

$$\mathrm{X}(\Omega) = \{x_k\ ;\ k \in \mathbb{N}\} = \{y_\ell\ ;\ \ell \in \mathbb{N}\}\,,$$

et si $\sum |x_k|\ \mathbf{P}\{\mathrm{X} = x_k\}$ converge, alors $\sum |y_\ell|\ \mathbf{P}\{\mathrm{Y} = y_\ell\}$ converge et

$$\mathbf{E}(\mathrm{X}) = \sum_{k=0}^{\infty} x_k\,\mathbf{P}\{\mathrm{X} = x_k\} = \sum_{\ell=0}^{\infty} y_\ell\,\mathbf{P}\{\mathrm{X} = y_\ell\}\,.$$

Notamment, si X admet une espérance, et si l'ensemble de ses valeurs $\mathrm{X}(\Omega) = \{x_k\ ;\ k \in \mathrm{K}\}$ est indicé par un ensemble K fini ou dénombrable, mais non nécessairement ordonné, on peut écrire sans qu'il y ait d'ambiguïté

$$\mathbf{E}(\mathrm{X}) = \sum_{k \in \mathrm{K}} x_k\,p_k,$$

et sans préciser davantage comment cette somme se calcule. Par exemple, si X prend ses valeurs dans $\mathbb{Z}$ et si l'on note $p_k = \mathbf{P}\{\mathrm{X} = k\}$, les procédés suivants sont tous valides :

$$\begin{aligned}\mathbf{E}(\mathrm{X}) &= \lim_{n\to\infty} \sum_{k=-n}^{n} k\,p_k = \left(\lim_{n\to\infty} \sum_{k=0}^{n} k\,p_k\right) + \left(\lim_{n\to\infty} \sum_{k=-n}^{-1} k\,p_k\right) \\ &= \left(\sum_{k\ \text{pair}} k\,p_k\right) + \left(\sum_{k\ \text{impair}} k\,p_k\right).\end{aligned}$$

(1). Cf. A12, page 671 et la discussion du paragraphe § 35.

Représentations de X comme combinaisons de fonctions indicatrices §84

Une variable aléatoire discrète X admet une écriture de la forme

$$X = \sum_{k \in K} x_k \, \mathbf{1}_{D_k},$$

où l'on a noté $D_k := \{X = x_k\}$. *Réciproquement*, on peut partir d'une telle écriture pour définir une variable aléatoire discrète. On se donne une partition $\mathfrak{A}$ de Ω, qu'on note $\mathfrak{A} = \{A_k \ ; \ k \in K\}$ et des réels $\{a_k \ ; \ k \in K\}$, pas nécessairement distincts deux à deux, et l'on définit une variable aléatoire en posant

$$X := \sum_{k \in K} a_k \, \mathbf{1}_{A_k}.$$

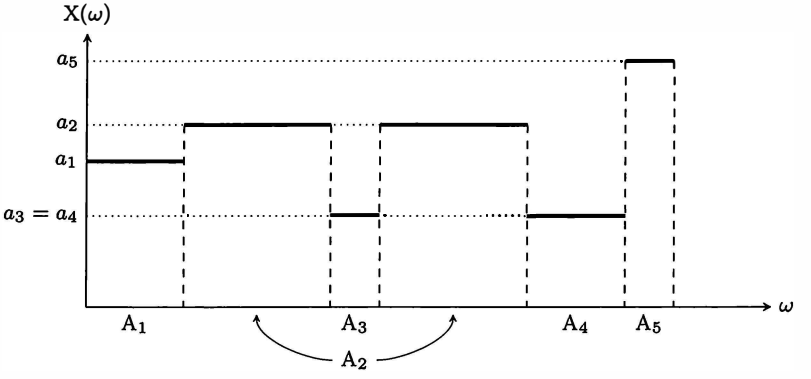


FIGURE 7.1 — *Sur cet exemple,* $A_1 = \{X = a_1\}$, $A_2 = \{X = a_2\}$, $A_5 = \{X = a_5\}$ *mais* $\{X = a_3\} = A_3 + A_4$.

Cette écriture n'est pas unique[2] mais le calcul de l'espérance ne dépend pas de la représentation choisie :

■ **Théorème 7.4** *Soit* X *une variable aléatoire discrète ; on suppose qu'elle admet une représentation de la forme*

$$X = \sum_i a_i \, \mathbf{1}_{A_i},$$

où les a_i *ne sont pas nécessairement distincts. Alors* X *admet une espérance si et seulement si* $\sum a_i \, \mathbf{P}(A_i)$ *converge absolument ; dans ce cas,*

$$\mathbf{E}(X) = \sum_i a_i \, \mathbf{P}(A_i).$$

DÉMONSTRATION : Supposons que X admette l'écriture $X = \sum_i a_i \, \mathbf{1}_{A_i}$, où l'indice i parcourt un ensemble I (fini ou dénombrable) et où les A_i forment une partition de Ω. L'ensemble $\{a_i \ ; \ i \in I\}$ étant fini ou dénombrable, on le ré-indice sous la forme $\{x_k \ ; \ k \in K\}$; la différence entre ces deux notations est que les x_k sont distincts deux à deux, alors que les a_i ne l'étaient pas *a priori*. On a alors, pour tout $k \in K$:

$$\sum_{i \ ; \ a_i = x_k} a_i \, \mathbf{P}(A_i) = x_k \sum_{i \ ; \ a_i = x_k} \mathbf{P}(A_i) = x_k \, \mathbf{P}\Big(\sum_{i \ ; \ a_i = x_k} A_i \Big) = x_k \, \mathbf{P}\{X = x_k\}$$

(2). Si $\Omega = A + B$ est une décomposition de Ω, la variable constante égale à a s'écrit aussi bien $X = a \, \mathbf{1}_\Omega$ que $X = a \, \mathbf{1}_A + a \, \mathbf{1}_B$. C'est pourquoi les A_k ne sont pas les D_k.

(les termes de cette somme étant tous de même signe). Si X admet une espérance, alors

$$\sum_{i\in I} a_i\,\mathbf{P}(A_i) = \sum_{k\in K}\Big(\sum_{i\,;\,a_i=x_k} a_i\,\mathbf{P}(A_i)\Big) = \sum_{k\in K} x_k\,\mathbf{P}\{X=x_k\} = \mathbf{E}(X), \qquad (*)$$

les réarrangements étant autorisés par le théorème de sommation par paquets A.14 page 674, puisque toutes les séries en jeu sont absolument convergentes.

Réciproquement, si $\sum a_i\,\mathbf{P}(A_i)$ converge absolument, le théorème de sommation par paquets montre que la famille $(a_i\,\mathbf{P}(A_i))_{i\in I}$ est sommable, donc que X admet une espérance et que la formule (*) reste valable. ∎

Exemple 7.5 (Cas fini : espérance d'une loi binomiale) Soit p un réel, $0 < p < 1$, et soit n un entier naturel non nul. On considère le nombre X de *pile* lors d'un lancer de n pièces identiques mais déséquilibrées, menant à *pile* avec une probabilité p. La variable aléatoire X suit alors la loi binomiale $\mathscr{B}(n\,;\,p)$; ne prenant qu'un nombre fini de valeurs, elle admet une espérance, et celle-ci vaut

$$\begin{aligned}\mathbf{E}(X) &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n}\frac{n!}{(k-1)!\,(n-k)!}\,p^k(1-p)^{n-k}\\ &= \sum_{k=1}^{n} n\,\frac{(n-1)!}{(k-1)!\,[(n-1)-(k-1)]!}\,p^k(1-p)^{n-k} = n\sum_{k=1}^{n}\binom{n-1}{k-1}p^k(1-p)^{n-k}\\ &= np\cdot\underbrace{\sum_{k=0}^{n-1}\binom{n-1}{k}p^k(1-p)^{n-1-k}}_{=(p+1-p)^{n-1}} = np.\end{aligned}$$

Remarque 7.6 (Variables aléatoires prenant une valeur infinie) Comme déjà vu, certaines variables aléatoires peuvent prendre la valeur $\pm\infty$; si l'événement $\{X = +\infty\}$ est de probabilité nulle, on conviendra que

$$(+\infty)\cdot\mathbf{P}\{X=+\infty\} = (+\infty)\cdot 0 = 0,$$

(et de même pour la valeur $-\infty$) et que cette valeur, aberrante mais négligeable, ne compte pas dans l'espérance. Une variable aléatoire discrète à valeurs dans $[-\infty\,;+\infty]$ et admettant une espérance ne prend des valeurs infinies qu'avec une probabilité nulle ou, exprimé différemment : *une variable aléatoire discrète admettant une espérance a presque sûrement une valeur finie.*

Exemple 7.7 (Cas dénombrable : loi géométrique) On effectue cette fois-ci une succession infinie de tirages à *pile* ou *face*, toujours avec une pièce déséquilibrée donnant *pile* avec une probabilité p ; on note X le rang du premier tirage donnant un *pile* : ce nombre peut être un entier, ou bien l'infini.

La probabilité que X prenne une valeur finie est

$$\mathbf{P}\{X<+\infty\} = \sum_{n=1}^{\infty} p\,(1-p)^{n-1} = \frac{p}{1-(1-p)} = 1,$$

ce qui montre que l'événement $\{X = +\infty\}$ est négligeable. L'espérance de X est alors[3]

$$\mathbf{E}(X) = \sum_{n=1}^{\infty} n\,p\,(1-p)^{n-1} = \frac{1}{p},$$

puisque la série définissant $\mathbf{E}(X)$ est bien (absolument) convergente.

■ **Proposition 7.8** *Toute variable aléatoire discrète et bornée admet une espérance.*

DÉMONSTRATION : Soit X une variable aléatoire discrète et bornée ; on peut l'écrire sous la forme

$$X = \sum_k x_k\,\mathbf{1}_{A_k},$$

où l'indice k parcourt un ensemble K au plus dénombrable. Supposons qu'il existe un réel M tel que $|x_k| \leqslant M$ pour tout k, alors la série positive $\sum |x_k|\,p_k$ converge, puisque ses
⇗7.5 sommes partielles sont toutes majorées par $\sum_{k\in K} M\,p_k = M$. ∎

(3). Le formulaire de l'annexe D récapitule le détail des calculs menant à l'équation (A.4) utilisée ici.

Espérances infinies § 85

On peut étendre la définition de l'espérance à certaines variables pour lesquelles la série $\sum x_k\, p_k$ diverge. Commençons par le cas d'une variable aléatoire *positive*, c'est-à-dire ne prenant ses valeurs que dans $\mathbb{R}^+$. Si la série $\sum x_k\, p_k$ diverge, la suite des ses sommes partielles tend vers $+\infty$ et la même divergence vers $+\infty$ sera observée si l'on modifie arbitrairement l'ordre de sommation. On peut alors, de façon cohérente, poser

$$\mathbf{E}(\mathrm{X}) := +\infty.$$

Ainsi, *l'espérance d'une variable aléatoire positive est une quantité toujours définie à condition de travailler dans* $\mathbb{R}\cup\{+\infty\}$. Il est toutefois convenu que la locution « X admet une espérance » implique que celle-ci est finie[4]. On écrit parfois « $\mathbf{E}(\mathrm{X}) < +\infty$ » comme synonyme de : « X admet une espérance (finie) ».

Dans le cas général d'une variable aléatoire à valeurs dans $\mathbb{R}$, décomposons X en sa **partie positive** et sa **partie négative** :

$$\mathrm{X} = \mathrm{X}^+ - \mathrm{X}^-$$

où X^+ et X^- sont les variables aléatoires *positives*

$$\mathrm{X}^+ := \max(0, \mathrm{X}) \qquad \text{et} \qquad \mathrm{X}^- := \max(0, -\mathrm{X})$$

comme représenté à la figure 7.2 ; cette décomposition permet d'écrire

$$|\mathrm{X}| = \mathrm{X}^+ + \mathrm{X}^-.$$

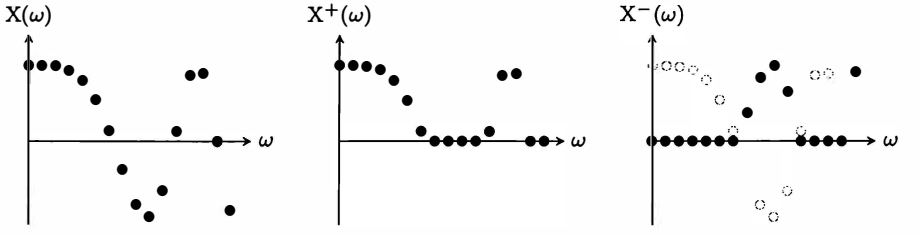


FIGURE 7.2 — *Parties positive et négative d'une fonction (l'ensemble de départ est représenté ici symboliquement comme un ensemble discret).*

Trois cas peuvent alors se produire.

- Si $\mathbf{E}(\mathrm{X}^+) = +\infty$ et si $\mathbf{E}(\mathrm{X}^-) = m^-$ est finie, alors on pose
$$\mathbf{E}(\mathrm{X}) := \mathbf{E}(\mathrm{X}^+) - \mathbf{E}(\mathrm{X}^-) = +\infty - m^- = +\infty.$$
- Si $\mathbf{E}(\mathrm{X}^+) = m^+$ est finie et si $\mathbf{E}(\mathrm{X}^-) = +\infty$, alors on pose
$$\mathbf{E}(\mathrm{X}) := \mathbf{E}(\mathrm{X}^+) - \mathbf{E}(\mathrm{X}^-) = m^+ - \infty = -\infty.$$
- Si $\mathbf{E}(\mathrm{X}^+) = \mathbf{E}(\mathrm{X}^-) = +\infty$, puisque la forme « $+\infty - \infty$ » est indéterminée, on ne peut pas définir $\mathbf{E}(\mathrm{X})$.

Enfin, des propriétés générales des séries[5], on déduit le résultat suivant qui résume notre étude :

(4). On dira alternativement que X est *intégrable*, cette terminologie s'expliquant naturellement dans le cadre de la théorie de la mesure (cf. chapitre 14).

(5). Voir proposition A.5 page 670.

■ **Proposition 7.9** *Soit* X *une variable aléatoire réelle ; alors il y a équivalence entre les énoncés :*

i) X *admet une espérance ;*

ii) $|X|$ *admet une espérance ;*

iii) X^+ *et* X^- *admettent une espérance.*

Dans le cas où ces conditions sont satisfaites, on a

$$\mathbf{E}(X) = \mathbf{E}(X^+) - \mathbf{E}(X^-)$$

et

$$\big|\mathbf{E}(X)\big| = \big|\mathbf{E}(X^+) - \mathbf{E}(X^-)\big| \leqslant \mathbf{E}(X^+) + \mathbf{E}(X^-) = \mathbf{E}|X|.$$

Remarque 7.10 (Signification de l'espérance) L'espérance mathématique d'une variable aléatoire véhicule une signification intuitive, celle d'une valeur « moyenne », ou de valeur « attendue » (que traduit bien le terme anglais : *expected value*) : si l'on répète l'expérience un grand nombre de fois et que, pour chaque expérience k, on note la valeur x_k prise par X, alors la valeur moyenne « devrait » être proche de $\mathbf{E}(X)$. L'emploi du conditionnel représente évidemment un flou inacceptable en mathématiques ; de plus, le cas des variables aléatoires à espérance infinie est moins clair. Ce sera la grande force de la théorie de probabilités que de donner un sens très précis à l'idée précédente, au moyen des divers théorèmes connus sous le nom de « lois des grands nombres ».

Bien sûr, l'espérance n'est pas la seule valeur numérique intéressante lorsque l'on étudie une variable aléatoire. Il est bien connu par exemple que

ex.7.1 99,9 % *de la population possède un nombre de jambes strictement supérieur à la moyenne.*

La valeur la plus probable (le « mode ») ou la valeur médiane (au-dessus de laquelle X prend ses valeurs avec une probabilité 1/2) sont parfois des paramètres plus pertinents.

§ 86 Linéarité

Une des propriétés les plus importantes de l'espérance est sa linéarité :

■ **Théorème 7.11** *Soient* X *et* Y *deux variables aléatoires réelles discrètes admettant une espérance. Soient* α *et* β *des réels. Alors* $\alpha X + \beta Y$ *admet une espérance et*

$$\mathbf{E}(\alpha X + \beta Y) = \alpha\,\mathbf{E}(X) + \beta\,\mathbf{E}(Y).$$

DÉMONSTRATION : Si une variable aléatoire X se décompose sous la forme $X = \sum x_k \mathbf{1}_{A_k}$, alors $\lambda X = \sum_k \lambda x_k \mathbf{1}_{A_k}$, ce qui prouve que λX est une variable aléatoire discrète admettant une espérance $\mathbf{E}(\lambda X) = \lambda\,\mathbf{E}(X)$.

Il ne reste donc plus qu'à démontrer que $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$. On suppose que X et Y se décomposent en

$$X = \sum_{i\in I} x_i \mathbf{1}_{A_i} \qquad \text{et} \qquad Y = \sum_{j\in J} y_j \mathbf{1}_{B_j},$$

où $\{A_i\ ;\ i \in I\}$ et $\{B_j\ ;\ j \in J\}$ sont des partitions, finies ou dénombrables, de Ω, et on note $Z = X + Y$. Posons, pour tous $i \in I$ et $j \in J$:

$$\Delta_{i,j} := A_i \cap B_j \qquad \text{et} \qquad z_{i,j} := x_i + y_j.$$

Alors $\{\Delta_{i,j}\ ;\ (i,j) \in I \times J\}$ est une partition, finie ou dénombrable, de Ω (certains éléments $\Delta_{i,j}$ peuvent être vides). Or, par construction, pour tout ω d'un $\Delta_{i,j}$ non vide, on a

$$Z(\omega) = X(\omega) + Y(\omega) = x_i + y_j = z_{i,j},$$

ce qui prouve que Z admet la décomposition

$$Z = \sum_{(i,j)\in I\times J} z_{i,j} \mathbf{1}_{\Delta_{i,j}}$$

Il ne reste plus qu'à établir la convergence absolue de la série double définissant l'espérance de Z, de terme général $(x_i + y_j)\,\mathbf{P}(\Delta_{i,j})$ (voir théorème A.12 page 673). Par simple inégalité triangulaire $|x_i + y_j| \leqslant |x_i| + |y_j|$, il suffit de démontrer la convergence des

deux séries doubles $\sum |x_i|\ \mathbf{P}(\Delta_{i,j})$ et $\sum |y_j|\ \mathbf{P}(\Delta_{i,j})$; par symétrie, il suffit d'ailleurs de prouver la convergence de la première seulement. Pour simplifier l'écriture, supposons $J = \mathbb{N}$ (cas dénombrable générique, le cas fini s'y ramène en posant $B_j = \varnothing$ pour une infinité d'indices j). Alors, pour tout $i \in I$, on a

$$\sum_{j=1}^{N} |x_i|\ \mathbf{P}(A_i \cap B_j) = |x_i|\,\mathbf{P}\Big(\sum_{j=1}^{N} A_i \cap B_j\Big) \leqslant |x_i|\ \mathbf{P}(A_i)$$

ce qui prouve que $\sum_j |x_i|\ \mathbf{P}(A_i \cap B_j)$ converge et que sa somme est majorée par $|x_i|\ \mathbf{P}(A_i)$; ce dernier terme étant le terme général d'une série convergente, la série double converge absolument et, d'après le théorème de Fubini, on peut la sommer dans l'ordre voulu. Notamment,

$$\begin{aligned}\mathbf{E}(X+Y) &= \sum_{i\in I}\sum_{j\in J}(x_i + y_j)\,\mathbf{P}(A_i \cap B_j) = \sum_{i\in I} x_i \sum_{j\in J}\mathbf{P}(A_i \cap B_j) + \sum_{j\in J} y_j \sum_{i\in I}\mathbf{P}(A_i \cap B_j)\\ &= \sum_{i\in I} x_i\,\mathbf{P}(A_i) + \sum_{j\in J} y_j\,\mathbf{P}(B_j) = \mathbf{E}(X) + \mathbf{E}(Y).\end{aligned}$$

∎

Application de la linéarité au calcul pratique d'une espérance

§ 87

Dans certaines situations, il est avantageux de décomposer une variable aléatoire sous la forme d'une somme de fonctions indicatrices :

$$X = \sum_i \alpha_i\, \mathbf{1}_{A_i},$$

où les A_i sont des événements, mais *pas forcément disjoints*. Sous réserve de convergence absolue, l'espérance de X est :

$$\mathbf{E}(X) = \sum_i \alpha_i\, \mathbf{E}(\mathbf{1}_{A_i}) = \sum_i \alpha_i\, \mathbf{P}(A_i).$$

Donnons trois exemples, le premier étant absolument élémentaire.

Exemple 7.12 (Espérance d'une loi binomiale) Une variable aléatoire X suivant la loi binomiale $\mathscr{B}(n\,;\,p)$ reflète le nombre de succès d'une suite de n tirages indépendants, chacun d'eux ayant une probabilité p d'amener un succès. Si l'on note A_k l'événement : « le k-ième tirage est un succès », on a donc

$$X = \sum_{k=1}^{n} \mathbf{1}_{A_k}$$

et par conséquent

$$\mathbf{E}(X) = \sum_{k=1}^{n} \mathbf{P}(A_k) = \sum_{k=1}^{n} p = np.$$

(On notera que l'hypothèse d'indépendance des tirages n'a pas été utilisée ; cette indépendance eût-elle été remise en question, la loi de X aurait été différente, mais pas son espérance.)

Exemple 7.13 (Le problème des chapeaux : nombre de points fixes d'une permutation) Le célèbre *problème des chapeaux*, ou *des conjoints*[6] est un problème classique de dénombrement. Sous sa forme habituelle, on considère n hommes qui, à l'issue d'une réunion, récupèrent tous leur chapeau au vestiaire. L'employé du vestiaire, qui a négligé de ranger les chapeaux, leur remet aléatoirement un chapeau chacun. On demande combien de chapeaux, en moyenne, seront remis à leur légitime propriétaire.

Sous forme mathématique, il s'agit simplement de calculer la valeur moyenne du nombre de points fixes des permutations φ de $[\![1\,;\,n]\!]$. L'univers Ω est choisi comme l'ensemble $\mathfrak{S}_n$ des permutations de $[\![1\,;\,n]\!]$; c'est un ensemble fini, de cardinal $n!$. On le munit de la tribu triviale

(6). En anglais, *matching problem*. On remplace les chapeaux par des épouses, et chaque homme choisit une femme au hasard pour entamer — en tout bien tout honneur — une danse.

$\mathfrak{T} = \mathfrak{P}(\Omega)$ et de la probabilité uniforme : $\mathbf{P}(\{\varphi\}) = 1/n!$ pour tout $\varphi \in \mathfrak{S}_n$. On définit enfin la variable aléatoire X donnant le nombre de points fixes :

$$\forall \varphi \in \Omega \qquad \mathrm{X}(\varphi) = \mathrm{Card}\,\{k\,;\,\varphi(k) = k\} = \sum_{k=1}^{n} \mathbf{1}_{\mathrm{A}_k}(\varphi)$$

où l'événement A_k est

$$\mathrm{A}_k = \{\varphi \in \mathfrak{S}_n\,;\,\varphi(k) = k\}.$$

Il est aisé de calculer $\mathbf{P}(\mathrm{A}_k)$: la valeur de $\varphi(k)$ peut être n'importe quel entier entre 1 et n, avec la même fréquence, donc $\mathbf{P}(\mathrm{A}_k) = 1/n$. Ainsi,

$$\mathbf{E}(\mathrm{X}) = \sum_{k=1}^{n} \frac{1}{n} = 1.$$

⏎7.3 Le nombre moyen de chapeaux remis à leur propriétaire légitime est donc égal à 1, indépendamment du nombre de participants.

Exemple 7.14 (Le problème du collectionneur) Une expérience consiste à choisir au hasard un entier entre 1 et n ; on effectue n expériences, indépendantes l'une de l'autre, et on note X le nombre de numéros jamais tirés. La loi de X est un peu délicate à obtenir, mais son espérance se calcule avec la même technique que précédemment. On note A_i l'événement

$$\mathrm{A}_i : \quad \text{« le numéro } i \text{ n'est jamais tiré »}.$$

On décompose $\mathrm{A}_i = \mathrm{A}_i^1 \cap \mathrm{A}_i^2 \cap \cdots \cap \mathrm{A}_i^n$, où A_i^j est l'événement « le numéro i n'est pas tiré à l'expérience j ». Chaque A_i^j est de probabilité $(n-1)/n$ et, par indépendance des expériences, on a

$$\mathbf{P}(\mathrm{A}_i) = \left(\frac{n-1}{n}\right)^n.$$

La variable X s'écrivant

$$\mathrm{X} = \sum_{i=1}^{n} \mathbf{1}_{\mathrm{A}_i},$$

son espérance est donc

$$\mathbf{E}(\mathrm{X}) = n\left(\frac{n-1}{n}\right)^n = \frac{(n-1)^n}{n^{n-1}} \underset{n\to\infty}{\sim} \frac{n}{\mathrm{e}}.$$

Ce calcul est un cas particulier du problème du collectionneur de vignettes, qui sera davantage développé dans la section 7.4 de ce chapitre.

§88 Autres propriétés de l'espérance

La proposition suivante énumère quelques propriétés simples, mais constamment utiles, de l'espérance.

■ **Proposition 7.15** *Soient* X *et* Y *des variables aléatoires discrètes.*

i) Si X *est presque sûrement égale à une constante* a, i.e. *si* $\mathbf{P}\{\mathrm{X} = a\} = 1$, *alors* X *admet une espérance et* $\mathbf{E}(\mathrm{X}) = a$. *Notamment,* $\mathbf{E}(\mathbf{E}(\mathrm{X})) = \mathbf{E}(\mathrm{X})$.

ii) Si $\mathrm{X} \geqslant 0$ *alors* $\mathbf{E}(\mathrm{X}) \geqslant 0$.

iii) Si $\mathrm{X} \geqslant 0$ *et si* $\mathbf{E}(\mathrm{X}) = 0$, *alors* $\mathrm{X} = 0$ *presque sûrement.*

iv) Si $\mathrm{X} \geqslant 0$ *presque sûrement, et si* $\mathbf{E}(\mathrm{X}) = 0$, *alors* $\mathrm{X} = 0$ *presque sûrement.*

v) Si $\mathrm{X} \leqslant \mathrm{Y}$ *et si* X *et* Y *admettent une espérance, alors* $\mathbf{E}(\mathrm{X}) \leqslant \mathbf{E}(\mathrm{Y})$.

vi) X *admet une espérance si et seulement si* $|\mathrm{X}|$ *en admet une, et dans ce cas* $|\mathbf{E}(\mathrm{X})| \leqslant \mathbf{E}(|\mathrm{X}|)$.

vii) Si A *est un événement, alors sa fonction indicatrice* $\mathbf{1}_{\mathrm{A}}$ *admet une espérance, et* $\mathbf{E}(\mathbf{1}_{\mathrm{A}}) = \mathbf{P}(\mathrm{A})$.

viii) Si Y *admet une espérance et si* $|\mathrm{X}| \leqslant |\mathrm{Y}|$, *alors* X *admet une espérance.*

DÉMONSTRATION : *i)* Toutes les autres valeurs x_k prises par X le sont avec une probabilité nulle, donc $\mathbf{E}(X) = a + \sum_k x_k \times 0 = a$.

ii) Immédiat.

iii) Notons $\{x_k \,;\, k \in K\}$ les valeurs prises par X, où les x_k sont des réels positifs. Alors

$$\mathbf{E}(X) = \sum x_k \, \mathbf{P}\{X = x_k\}$$

est une somme de termes positifs, et ne peut être nulle que si tous les termes sont nuls. Ainsi, pour tout $x_k > 0$, on a $\mathbf{P}\{X = x_k\} = 0$, ce qui prouve que $\mathbf{P}\{X = 0\} = 1$.

iv) La démonstration est similaire en notant que, dans la somme définissant l'espérance, on peut de toute façon omettre les valeurs négatives, de probabilité nulle.

v) Résulte de ce que $Y - X \geqslant 0$ et de la linéarité de l'espérance.

vi) Par définition de l'espérance, X et $|X|$ admettent simultanément, ou non, une espérance. L'inégalité triangulaire (A10) permet de conclure.

vii) Par définition de l'espérance.

viii) On applique *v)* et *vi)*. ∎

Intégration terme à terme

§ 89

Terminons pas un résultat fort utile d'échange de somme et d'espérance :

■ **Théorème 7.16 (Espérance d'une somme infinie de variables)** *Soit* $(X_n)_{n\geqslant 0}$ *une suite de variables aléatoires telle que* $\sum \mathbf{E}|X_n|$ *converge. Alors la série* $\sum X_n$ *converge (absolument) presque sûrement, et*

$$\mathbf{E}\left(\sum_{n=0}^{\infty} X_i\right) = \sum_{n=0}^{\infty} \mathbf{E}(X_i).$$

La première partie de la conclusion, ($\sum X_n$ converge presque sûrement) se démontre de manière élémentaire, voir par exemple l'exercice 19.6 page 412.

La démonstration de l'*intégration terme à terme*, en revanche, n'est pas accessible avec les moyens élémentaires développés jusqu'ici, car S peut très bien ne pas être une variable discrète : par exemple, si les X_n sont des variables de Bernoulli, la somme de la série $\sum X_n/2^n$ peut prendre toute valeur entre 0 et 1.

La démonstration de l'interversion somme/espérance sera exposée, dans une version généralisée aux variables aléatoire quelconques, dans le chapitre sur l'espérance (corollaire 14.24 page 316).

7.2 MOMENTS

§90 **Formule de transfert**

Soient X une variable aléatoire discrète et $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction, définie au moins sur $X(\Omega)$. Alors, la fonction $\varphi \circ X : \Omega \to \mathbb{R}$ est encore une variable aléatoire discrète, puisque l'image d'un ensemble dénombrable par une application quelconque est au plus dénombrable.

■ **Théorème 7.17 (Théorème de transfert)** *La variable aléatoire* $\varphi(X)$ *admet une espérance si et seulement si la série* $\sum \varphi(x_k)\, p_k$ *converge absolument. Si c'est le cas,*

$$\mathbf{E}\big[\varphi(X)\big] = \sum_{k \in K} \varphi(x_k)\, p_k.$$

DÉMONSTRATION : Supposons que la variable aléatoire X admette une représentation $X = \sum_k x_k \mathbf{1}_{A_k}$. Alors $\varphi(X)$ s'écrit $\varphi(X) = \sum_k \varphi(x_k) \mathbf{1}_{A_k}$ ce qui montre le résultat. ∎

Exemple 7.18 Soient $\lambda > 0$ et X une variable aléatoire suivant une loi de Poisson $\mathscr{P}(\lambda)$ de paramètre λ. On demande l'espérance de la variable aléatoire $Y = 1/(1+X)$.

Bien que la loi de Y soit simple à écrire, utilisons directement le théorème de transfert (la ▵7.2 série à termes positifs qui apparaît est absolument convergente) :

$$\mathbf{E}(Y) = \sum_{n=0}^{\infty} \frac{1}{1+n} e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{(n+1)!} = \frac{e^{-\lambda}}{\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{n!} = \frac{1 - e^{-\lambda}}{\lambda}.$$

§91 **Moments**

Pour tout entier n non nul, on définit le **moment d'ordre n** d'une variable aléatoire discrète X prenant les valeurs $\{x_k \,;\, k \in K\}$ comme l'espérance, si elle existe, de la variable aléatoire X^n :

$$m_n(X) := \mathbf{E}(X^n) = \sum_{k \in K} x_k^n \, \mathbf{P}\{X = x_k\}$$

sous réserve de convergence absolue de la série.

Remarque 7.19 Le moment d'ordre 0 d'une variable aléatoire existe toujours et vaut $\mathbf{E}(X^0) = \mathbf{E}(1) = 1$. Le moment d'ordre 1, s'il existe, est l'espérance de X.

■ **Théorème 7.20** *Soit* X *une variable aléatoire réelle discrète, admettant un moment d'ordre* $n \geqslant 2$. *Alors* X *admet des moments d'ordre* $1, 2, \ldots, n$.

DÉMONSTRATION : Soit i un entier, $1 \leqslant i \leqslant n$. Notons $\{x_k \,;\, k \in K\}$ l'ensemble des valeurs prises par X, posons $p_k = \mathbf{P}\{X = x_k\}$. Alors, pour tout $k \in \mathbb{K}$, on a

$$|x_k|^i \leqslant \begin{cases} 1 & \text{si } |x_k| \leqslant 1 \\ |x_k|^n & \text{si } |x_k| > 1, \end{cases}$$

et donc $|x_k|^i \, p_k \leqslant \big(1 + |x_k|^n\big)\, p_k$. Puisque les séries $\sum |x_k|^n \, p_k$ et $\sum 1\, p_k$ convergent, la ▵7.4 ▵7.11 série $\sum |x_k|^i \, p_k$ converge et X admet un moment d'ordre i. ∎

§92 **Moments centrés**

Soit X une variable aléatoire discrète, admettant une espérance. On dit que X admet un **moment centré d'ordre n** si la variable aléatoire centrée $X' := X - \mathbf{E}(X)$ admet un moment d'ordre n. Le moment centré d'ordre n de X est alors

$$m'_n(X) := \mathbf{E}(X'^n) = \mathbf{E}\big[(X - \mathbf{E}X)^n\big].$$

■ **Proposition 7.21** *Soit* X *une variable aléatoire discrète. Alors* X *admet un moment centré d'ordre n si et seulement si elle admet un moment d'ordre n.*

DÉMONSTRATION : Notons $m = \mathbf{E}(X)$. Supposons dans un premier temps que X admet un moment d'ordre n, et développons $(X - m)^n$:

$$(X - m)^n = \sum_{i=0}^{n} (-1)^{n-i} \binom{n}{i} X^i m^{n-i}.$$

Le théorème 7.20 montre que X admet des moments d'ordre 0 à n : l'expression de droite admet donc une espérance.

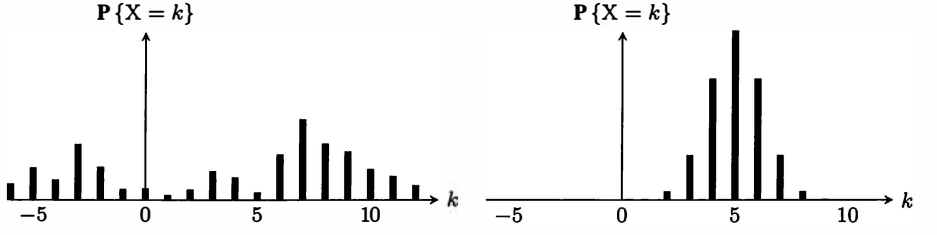
Réciproquement, si X admet un moment centré d'ordre n, c'est-à-dire si $X' = (X - m)$ admet des moments de tous ordres entre 1 et n, la relation

$$X^n = (X - m + m)^n = \sum_{i=0}^{n} \binom{n}{i} (X - m)^i m^{n-i}$$

montre que X admet un moment d'ordre n. ∎

Variance, écart-type § 93

La variance d'une variable aléatoire réelle discrète X est une valeur numérique permettant de quantifier la dispersion des valeurs possibles de X par rapport à sa moyenne (son espérance). Ci-dessous, les histogrammes de deux variables aléatoires à valeurs dans $\mathbb{Z}$, chacune de même espérance $\mathbf{E}(X) = 5$, mais dont la première est nettement plus dispersée que la seconde.



Pour quantifier cet étalement, on calcule tout d'abord $m = \mathbf{E}(X)$, espérance de X. La quantité $X - m$ est donc l'écart à la moyenne. On peut en calculer l'espérance, mais on obtient, par définition, zéro :

$$\mathbf{E}(X - m) = \mathbf{E}(X) - \mathbf{E}(m) = m - m = 0.$$

Autrement dit, la moyenne des écarts à la moyenne est nulle, les écarts positifs compensant précisément les écarts négatifs. Pour éviter ce phénomène, on peut par exemple calculer

$$\Delta(X) = \mathbf{E}\big(\,|X - m|\,\big),$$

mais c'est une quantité qui est difficile à manipuler. Il est beaucoup plus fructueux de considérer les écarts *quadratiques* à la moyenne et de calculer, si elle existe, la quantité $\mathbf{E}\big((X - m)^2\big)$, c'est-à-dire le moment centré d'ordre 2.

Définition 7.22 On dit qu'une variable aléatoire discrète **admet une variance** si elle admet un moment d'ordre 2. Dans ce cas, on définit sa **variance** comme son moment centré d'ordre 2 :

$$\mathbf{V}(X) := \mathbf{E}\big[(X - \mathbf{E}X)^2\big].$$

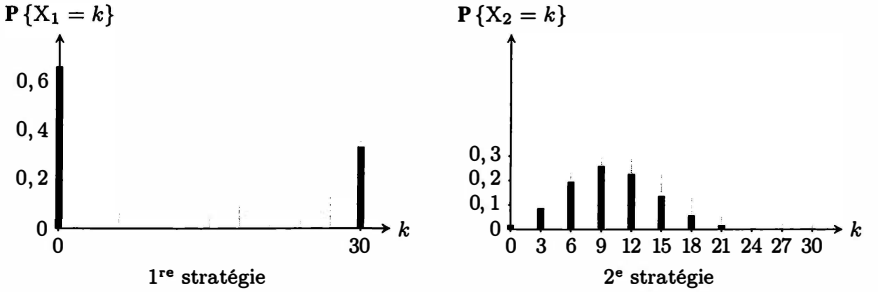
Exemple 7.23 (Jouer risqué ou jouer pépère ?) Raphaël de Valentin est désespéré : il n'a plus que 10 pièces d'or en poche et songe à en finir avec la vie. Il se rend à une maison de jeu et décide de jouer le restant de sa fortune à un jeu de hasard. La probabilité de gagner à ce jeu est $p = 1/3$, le gain étant de 3 fois la mise (la mise elle-même est perdue). Deux stratégies sont possibles :

- tout miser en une seule fois ;
- miser pièce par pièce.

Quelle que soit la stratégie choisie, Raphaël misera exactement ses 10 pièces et s'arrêtera ensuite. Notons X_1 sa fortune s'il joue selon la première stratégie, X_2 s'il joue selon la seconde.

- Dans le premier cas, X_1 vaut 30 avec une probabilité 1/3, et 0 avec une probabilité 2/3 ; la valeur moyenne de X_1 est $\mathbf{E}(X_1) = 10$. Sa variance est
$$\mathbf{V}(X_1) = \tfrac{1}{3}(30-10)^2 + \tfrac{2}{3}(0-10)^2 = 200.$$
- Dans le second cas, $X_2/3$ est égale au nombre de parties gagnées, et suit donc une loi binomiale $\mathscr{B}(10\,;\,1/3)$ de variance[7] $v = 10 \cdot \frac{1}{3} \cdot \frac{2}{3}$; l'espérance de X_2 est donc $\mathbf{E}(X_2) = 10 = \mathbf{E}(X_1)$, mais sa variance est
$$\mathbf{V}(X_2) = 9v = 20 < \mathbf{V}(X_1).$$

Voici les histogrammes associés à chacune des stratégies.



- La première permet d'espérer une somme rondelette en fin de partie, mais elle est « risquée » : avec une probabilité 2/3, Valentin se retrouve sans un sou ! C'est ce qui est traduit numériquement par la variance élevée de X_1.
- La seconde interdit pratiquement d'acquérir 30 pièces, mais assure que la somme en fin de partie va être « proche » de 10. C'est la stratégie « pépère » — peu de risques, peu de gains. Elle n'intéresse évidemment pas un joueur comme Raphaël[8].

■ **Proposition 7.24 (König-Huygens)** *Si* X *est une variable aléatoire admettant une variance, alors* X^2 *admet une espérance et*
$$\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2.$$

DÉMONSTRATION : Le fait que X^2 admette une espérance découle du théorème 7.21. On propose ensuite deux démonstrations de la formule. La première est élémentaire et a l'avantage de fonctionner pour des variables aléatoires plus générales (pas forcément discrètes), puisqu'elle n'utilise que la linéarité de l'espérance, acquise dès que tous les termes admettent une espérance. En notant $m = \mathbf{E}(X)$,
$$\mathbf{V}(X) = \mathbf{E}(X^2 - 2Xm + m^2) = \mathbf{E}(X^2) - 2m\,\mathbf{E}(X) + m^2 = \mathbf{E}(X^2) - 2m^2 + m^2 = \mathbf{E}(X^2) - m^2.$$
La seconde utilise la formule de transfert pour calculer explicitement la variance en fonction des valeurs x_k et $p_k = \mathbf{P}\{X = x_k\}$:
$$\begin{aligned}\mathbf{V}(X) &= \sum_{k\in K} (k-m)^2\, p_k = \sum_{k\in K} (k^2 - 2km + m^2)\, p_k \\ &= \sum_{k\in K} k^2 p_k - 2m \cdot \sum_{k\in K} k\, p_k + m^2 \sum_{k\in K} p_k = \mathbf{E}(X^2) - 2m^2 + m^2 = \mathbf{E}(X^2) - \mathbf{E}(X)^2.\end{aligned}$$
∎

[7]. Voir table des variances page 158.

[8]. En réalité, Raphaël perd tout mais ne se suicide pas, car quelqu'un lui donne une peau de chagrin, il rencontre une jeune fille... bref, la suite est un peu confuse et l'auteur a perdu une bonne occasion de présenter un joli calcul de variance.

Remarque 7.25 (Un point de terminologie) On dira, de manière équivalente :
— X admet une variance ;
— X admet un moment d'ordre 2 ;
— X^2 admet une espérance ;
— X est de carré sommable (ou intégrable), ou $X \in \mathcal{L}^2(\Omega)$.

De nombreux utilisateurs des probabilités sont intéressés, non par la variance d'une variable aléatoire, mais pas sa racine carrée, que l'on appelle **écart-type** de X :

$$\sigma_X := \sqrt{V(X)}.$$

Une des raisons est la suivante : la fonction X peut avoir une *dimension* (au sens physique : température, longueur, durée..., mais on pourrait également imaginer une unité monétaire). Dans ce cas, la variance de X a pour dimension le carré de celle de X, tandis que X et σ_X ont même dimension.

Remarque 7.26 (Écart-type d'une liste de données) On calcule parfois la moyenne ou l'écart-type d'une *liste de données* : le problème n'est plus de nature probabiliste, mais statistique. On considère donc une liste de n valeurs numériques $x_1, x_2, \ldots, x_n$. On définit leur moyenne par

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

(on n'a pas pondéré les données, mais il n'y aurait aucune difficulté à rajouter des coefficients de pondération) et leur écart-type par

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2.$$

Cet écart-type mesure, encore une fois, l'étalement des valeurs autour de la moyenne.

Prenons un exemple : dans l'organisation d'un concours faisant intervenir plusieurs matières, et donc plusieurs listes de données, la valeur de l'écart-type pour chaque matière est une donnée cruciale. On suppose pour simplifier que, dans ce concours, seules quatre matières sont évaluées — Maths, Physique, Informatique, Langue vivante, affectées chacune d'un coefficient : 5, 4, 4, et 2. Seule compte, pour le classement des candidats, la moyenne pondérée des notes de chaque matières. On demandera aux différents correcteurs, avant de rendre leurs listes de notes, de vérifier que leurs écarts-types sont égaux à une valeur fixée (par exemple 4). Que la moyenne de Maths soit différente de la moyenne de Chimie n'a aucune importance : rajouter une même constante à la note de Maths de tous les candidats ne modifiera pas leur classement. En revanche, si l'écart-type des Maths n'est que de 1 (tout le monde a eu entre 10 et 12) alors que celui de la Langue vivante est de 6 (les notes sont très étalées, entre 0 et 20), alors l'épreuve de Maths cesse d'être déterminante pour la sélection des candidats tandis que la note de Langue vivante fera la différence, malgré un coefficient plus faible.

Il est important de noter que, dans les exemples cités, nous n'avons pas fait intervenir d'aléa : les notes sont ce qu'elles sont, et on travaille sur des données fixées, les valeurs de la moyenne et de l'écart-type étant fournies par un simple algorithme de calcul.

On peut changer de point de vue et considérer la note d'un élève comme une variable aléatoire — nous vivons dans une réalité qui fixe ω, mais rien n'empêche, fictivement, de considérer que la collection des étudiants (réels ou virtuels) et des conditions extérieures (température de la pièce, qualité du repas de la cantine précédent l'examen, etc.) forme un ensemble Ω que l'on identifie au hasard. Dans ce cas, les valeurs $x_1, x_2, \ldots, x_n$ sont simplement des valeurs prises par une variable aléatoire X ou, mieux, par une suite $X_1, X_2, \ldots, X_n$ de variables aléatoires indépendantes et de même loi, qu'on appelle suite d'**échantillons** de X. La méthode de calcul de l'écart-type expliquée ci-dessus permet d'obtenir un **écart-type observé**, quantité statistique et non probabiliste. Cette quantité, puisqu'elle dépend de ω, est une variable aléatoire, qu'on appelle un *estimateur* de σ_X. Malheureusement, cet estimateur présente un défaut systématique : on montrera qu'une meilleure estimation de la variance de la variable aléatoire X est obtenu par le calcul de

$$V'(x_1, x_2, \ldots, x_n) = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2$$

qui, lui, est un estimateur sans biais (c'est-à-dire que l'espérance de V′ est exactement la variance de X[9]). C'est selon cette formule que les logiciels de calcul statistique (comme les tableurs Excel

(9). Voir la section 22.4 page 460.

ou OpenOffice, ou le logiciel R) calculent un écart-type ou une variance d'une liste de données : le préfacteur est bien $1/(n-1)$ et non $1/n$. Le lecteur est invité à se reporter à la section 22.4 pour plus de détails.

En statistiques, la présence d'un facteur $1/(n-1)$ au lieu de $1/n$ est souvent interprétée comme la perte d'un degré de liberté, puisque l'estimateur de la variance utilise *de facto* un estimateur de la moyenne.

§ 94 **Propriétés de la variance**

De la linéarité de l'espérance, on tire la propriété suivante :

■ **Proposition 7.27** *Soit* X *une variable aléatoire discrète admettant une variance. Soient* a *et* b *des réels. Alors* $aX+b$ *admet une variance,*

$$\mathbf{V}(aX+b) = a^2\,\mathbf{V}(X) \qquad \text{et} \qquad \sigma_{aX+b} = |a|\,\sigma_X.$$

DÉMONSTRATION : On donne ici une démonstration générale, qui restera valable pour les variables aléatoires non discrètes.

- Tout d'abord, pour tout réel b,

$$\mathbf{E}(X+b) = \mathbf{E}(X) + b \text{ donc } \mathbf{V}(X+b) = \mathbf{E}\big(X - \mathbf{E}(X)\big)^2 = \mathbf{V}(X).$$

- De plus, pour tout réel a, $\mathbf{E}(aX) = a\,\mathbf{E}(X)$ donc

$$\mathbf{V}(aX) = \mathbf{E}\big[(aX - a\,\mathbf{E}X)^2\big] = \mathbf{E}\big[a^2\,(X-\mathbf{E}X)^2\big] = a^2\,\mathbf{V}(X).$$

↗7.5
↗7.8 Ces deux propriétés permettent de conclure dans le cas général. ∎

§ 95 **Variable centrée réduite**

Si X est une variable aléatoire admettant une espérance, alors la variable

$$X' := X - \mathbf{E}(X)$$

est d'espérance nulle. Si de plus X admet une variance strictement positive, on définit la variable aléatoire

$$X^* := \frac{X - \mathbf{E}(X)}{\sigma_X},$$

qu'on appelle **variable aléatoire centrée réduite** associée à X. Son espérance est
↗7.7 nulle (variable **centrée**) et sa variance est égale à 1 (variable **réduite**).

§ 96 **Inégalité de Cauchy-Schwarz**

Voici la première irruption de l'inégalité de Cauchy-Schwarz, bien connue en géométrie et en analyse, dans le domaine des probabilités. Nous la retrouverons à maintes reprises.

■ **Théorème 7.28 (Cauchy-Schwarz)** *Soient* X *et* Y *deux variables aléatoires réelles discrètes admettant une variance. Alors* XY *admet une espérance et*

$$\mathbf{E}(XY)^2 \leqslant \mathbf{E}(X^2)\cdot\mathbf{E}(Y^2).$$

De plus, il n'y a égalité que si $X' = X - \mathbf{E}X$ *et* $Y' = Y - \mathbf{E}Y$ *sont linéairement dépendantes, de façon presque sûre, c'est-à-dire qu'il existe des scalaires* α *et* β*, non tous les deux nuls, tels que* $\alpha X + \beta Y = 0$ *presque sûrement, ou encore*

$$\mathbf{P}\{\alpha X + \beta Y = 0\} = 1.$$

DÉMONSTRATION : Montrons tout d'abord que, si X et Y admettent un moment d'ordre 2, alors XY admet une espérance. En notant $\mathfrak{D}_{X,Y} = \{A_k\,;\, k\in K\}$ la partition de Ω obtenue par intersections deux à deux des éléments de $\mathfrak{D}_X$ et de $\mathfrak{D}_Y$, on obtient une partition

de Ω permettant d'écrire les deux variables X et Y sous forme de combinaisons linéaires de fonctions indicatrices sur les mêmes parties[10] :

$$X = \sum_{k \in K} x_k \mathbf{1}_{A_k} \qquad \text{et} \qquad Y = \sum_{k \in K} y_k \mathbf{1}_{A_k}$$

(les valeurs x_k ne sont pas forcément distinctes, de même que les y_k.) On a alors

$$XY = \sum_{k \in K} x_k y_k \mathbf{1}_{A_k} \qquad X^2 = \sum_{k \in K} x_k^2 \mathbf{1}_{A_k} \qquad \text{et} \qquad Y^2 = \sum_{k \in K} y_k^2 \mathbf{1}_{A_k}.$$

Notons $p_k = \mathbf{P}(A_k)$; par hypothèse, les séries $\sum x_k^2 p_k$ et $\sum y_k^2 p_k$ convergent, et l'inégalité

$$|x_k y_k| \leqslant \frac{1}{2}(x_k^2 + y_k^2)$$

permet de conclure à la convergence absolue de la série $\sum x_k y_k p_k$, c'est-à-dire à l'existence de l'espérance de XY.

Quant à la démonstration de la formule proprement dite, elle est classique et semblable à celle de l'inégalité du même nom dans un espace préhilbertien.

Remarquons d'abord que, si Y est presque sûrement égale à 0, alors XY l'est également (il suffit de noter que $\{Y = 0\} \subset \{XY = 0\}$), donc $\mathbf{E}(XY) = \mathbf{E}(Y) = 0$ et l'inégalité est prouvée. Supposons désormais que Y n'est pas presque sûrement nulle ; alors $\mathbf{E}(Y^2) > 0$ d'après la proposition 7.15. On définit, pour tout réel λ, la quantité

$$Q(\lambda) = \mathbf{E}\big((X + \lambda Y)^2\big) = \mathbf{E}(X^2) + 2\lambda\,\mathbf{E}(XY) + \lambda^2\,\mathbf{E}(Y^2)$$

C'est un polynôme réel de degré exactement deux, ne prenant que des valeurs positives, son discriminant réduit $\Delta_\lambda = \mathbf{E}(XY)^2 - \mathbf{E}(X^2)\,\mathbf{E}(Y^2)$ est donc négatif, ce qui achève la démonstration de l'inégalité.

Enfin, le cas d'égalité entraîne la nullité de ce discriminant, ce qui est équivalent à l'existence d'une racine réelle λ^* du polynôme Q : on a alors

$$0 = Q(\lambda^*) = \mathbf{E}\big((X + \lambda^* Y)^2\big)$$

ce qui, toujours d'après la proposition 7.15, entraîne que $X + \lambda^* Y$ est presque sûrement nulle. ∎

Remarque 7.29 En appliquant cette relation aux variables aléatoires $|X|$ et $|Y|$, on a pour le même prix

$$\big(\mathbf{E}\,|XY|\big)^2 \leqslant \mathbf{E}\,(X^2) \cdot \mathbf{E}\,(Y^2)\,.$$

En spécialisant la variable Y et en choisissant $Y = 1$, on en déduit X admet une espérance et

$$\big|\mathbf{E}(X)\big| \leqslant \mathbf{E}\,|X| \leqslant \sqrt{\mathbf{E}(X^2)}.$$

Espérance et variance des lois classiques § 97

On dresse un tableau récapitulatif des lois classiques vues précédemment ; pour chacune d'entre elles, on donne son espérance et sa variance (table 7.1 page suivante). Les calculs aboutissant à ces valeurs sont tous élémentaires (sauf ceux concernant la loi hypergéométrique) et sont laissés en exercice au lecteur.

Inégalité de Bienaymé-Tchebychev § 98

Voici la première apparition d'une inégalité fondamentale de la théorie des probabilités[11], et dont l'interprétation est la suivante : si X est une variable aléatoire admettant une variance, alors elle ne peut pas « trop » s'écarter de sa moyenne. Plus précisément, l'écart-type mesure la tendance qu'a X de s'écarter de sa moyenne, et l'inégalité de Bienaymé-Tchebychev permet de quantifier cette tendance en majorant la probabilité que X diffère de sa moyenne de plus qu'une quantité donnée.

(10). Cette technique est mise en œuvre dans la démonstration du théorème 7.11.

(11). Nous la donnons (et la démontrons) ici pour une variable aléatoire discrète, mais elle est en fait valable quelle que soit la nature de X. Nous donnerons ainsi une seconde démonstration de ce théorème pour les variables à densité, ainsi qu'une troisième pour le cas général ; nous conseillons cependant au lecteur novice la lecture attentive (crayon à la main) de la preuve qui suit.

Nom	Désignation	$\mathbf{P}\{X = k\}$	$\mathbf{E}(X)$	$\mathbf{V}(X)$
Loi uniforme	$\mathscr{U}([1\,;n])$	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Loi uniforme	$\mathscr{U}([a\,;b])$	$\frac{1}{b-a+1}$	$\frac{a+b}{2}$	$\frac{(a-b)(a-b-2)}{12}$
Loi de Bernoulli	$\mathscr{B}(p)$	$\begin{cases} p & \text{si } k=1, \\ 1-p & \text{si } k=0 \end{cases}$	p	$p(1-p)$
Loi binomiale	$\mathscr{B}(n\,;\,p)$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Loi géométrique	$\mathscr{G}(p)$	$(1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Loi de Poisson	$\mathscr{P}(\lambda)$	$e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ
Loi hypergéométrique	$\mathscr{H}(N,n,p)$	$\frac{\binom{Np}{k}\binom{N(1-p)}{n-k}}{\binom{N}{n}}$	np	$np(1-p)\frac{N-n}{N-1}$

TABLE 7.1 — *Espérances et variances des lois discrètes usuelles.*

■ **Théorème 7.30 (Inégalité de Bienaymé-Tchebychev)** *Soit* X *une variable aléatoire réelle discrète admettant une variance. Notons* $m = \mathbf{E}(X)$ *l'espérance de* X, σ^2 *sa variance et* $X^* := (X-m)/\sigma$. *Alors, pour tout* $\varepsilon > 0$, *on a*

$$\mathbf{P}\{\,|X-m| \geqslant \varepsilon\} \leqslant \frac{\sigma^2}{\varepsilon^2},$$

ou encore

$$\mathbf{P}\{|X^*| \geqslant \varepsilon\} = \mathbf{P}\{\,|X-m| \geqslant \varepsilon\sigma\} \leqslant \frac{1}{\varepsilon^2}.$$

DÉMONSTRATION : Notons $\{x_k\,;\,k \in K\}$ l'ensemble des valeurs prises par X, K étant un ensemble fini ou dénombrable. On se fixe $\varepsilon > 0$, et on sépare l'ensemble K en deux parties disjointes : $K = K^+ + K^-$ où

$$K^- = \{k \in K\,;\,|x_k - m| < \varepsilon\} \qquad \text{et} \qquad K^+ = \{k \in K\,;\,|x_k - m| \geqslant \varepsilon\}.$$

Par définition, $\mathbf{P}\{\,|X-m| \geqslant \varepsilon\} = \sum_{k\in K^+} p_k$, donc

$$\begin{aligned} \varepsilon^2\,\mathbf{P}\{\,|X-m| \geqslant \varepsilon\} &= \sum_{k\in K^+} \varepsilon^2 p_k \\ &\leqslant \sum_{k\in K^+} |x_k - m|^2\, p_k \qquad \textit{car } |x_k-m|^2 \geqslant \varepsilon^2 \\ &\leqslant \sum_{k\in K} |x_k - m|^2\, p_k = \sigma^2 \end{aligned}$$

car on n'ajoute que des termes positifs ; ce qui achève la preuve. ∎

Remarque 7.31 Sur le même principe, on démontre l'inégalité de Markov :

■ **Théorème 7.32 (Inégalité de Markov)** *Soit* X *une variable aléatoire réelle discrète et positive, admettant une espérance ; pour tout* $\alpha > 0$,

$$\mathbf{P}\{X \geqslant \alpha\} \leqslant \frac{\mathbf{E}(X)}{\alpha}.$$

Exemple 7.33 À l'examen, plus de la moitié des copies s'est vue mettre une note supérieure ou égale à 10. Que dire de la moyenne ?

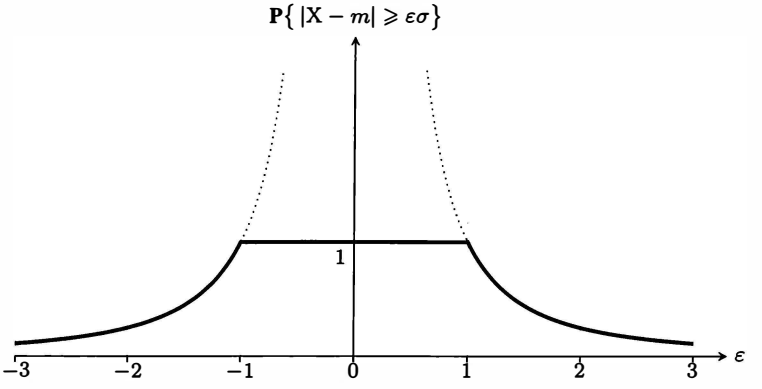


FIGURE 7.3 — *Probabilité maximale que la variable* X *se situe en dehors d'un intervalle* $[m - \varepsilon\sigma\,;\, m + \varepsilon\sigma]$ *pour différentes valeurs de* ε. *Pour les valeurs* ε *comprises entre* -1 *et* 1, *l'inégalité de Bienaymé-Tchebychev ne fournit aucun renseignement. La décroissance, en dehors de l'intervalle* $[-1\,;\, +1]$, *est relativement lente.*

Atouts et faiblesses de l'inégalite de Bienaymé-Tchebychev

§ 99

L'inégalité de Bienaymé-Tchebychev est remarquable à plus d'un titre. En premier lieu, elle est indépendante de la loi de X : l'estimation fournie est valable pour toutes les variables aléatoires discrètes admettant une variance. Plus généralement, on montrera qu'elle est valable pour toutes les variables aléatoires admettant une variance, qu'elles soient discrètes ou non. Le revers de la médaille est la médiocre efficacité numérique de cette estimation (cf. les exemples suivants, ainsi que l'exercice 11.2 page 254), qui n'offre qu'une décroissance « lente » en fonction de ε (voir figure 7.3 ainsi que la figure 22.3 page 453).

Une application de l'inégalité de Bienaymé-Tchebychev au problème de l'*aiguille de Buffon* est donnée au paragraphe § 382, page 534. ▵7.9 ▵7.10

Exemple 7.34 (Estimation d'un nombre suffisant de tirages) On lance n fois un dé équilibré à 6 faces ; le nombre *moyen* de tirages donnant un 6 est $n/6$. On demande comment choisir n pour que la probabilité d'obtenir un nombre de 6 compris entre 0 et $n/3$ soit supérieure à 0, 5, puis à 0, 9.

Notons Y_n le nombre de « 6 » obtenus à l'issue de n tirages. La condition cherchée peut se mettre sous la forme « symétrique » $\left|Y_n - \mathbf{E}(Y_n)\right| \leqslant \alpha_n$, et suggère l'utilisation de l'inégalité de Bienaymé-Tchebychev. Puisque Y suit la loi binomiale $\mathscr{B}(n\,;\, 1/6)$, son espérance vaut $\mathbf{E}(Y_n) = n/6$ et sa variance $\mathbf{V}(Y_n) = 5n/36$, ce qui mène à

$$\mathbf{P}\left\{\left|Y_n - \frac{n}{6}\right| \geqslant \frac{n}{6}\right\} \leqslant \frac{36\ \mathbf{V}(Y_n)}{n^2} = \frac{5}{n}.$$

Il *suffit* donc dans le premier cas de prendre n tel que $5/n \leqslant 1/2$, c'est-à-dire $n \geqslant 10$. Dans le second cas, il *suffit* de prendre n tel que $5/n \leqslant 0,1$, ou encore $n \geqslant 50$.

Insistons sur le fait que l'on a obtenu une condition *suffisante*, mais non nécessaire. On peut d'ailleurs, dans notre cas précis, calculer « à la main » la quantité $\mathbf{P}\{0 \leqslant Y_n \leqslant n/3\}$, puisque la loi de Y_n est connue :

$$p(n) := \mathbf{P}\left\{0 \leqslant Y_n \leqslant \frac{n}{3}\right\} = \frac{1}{6^n} \sum_{k=0}^{\lfloor n/3 \rfloor} \binom{n}{k} 5^{n-k}.$$

On obtient les valeurs suivantes (approchées à 10^{-10} près) :

n	$p(n)$	n	$p(n)$
1	0,8333333333	7	0,9042245370
2	0,6944444444	8	0,8651531064
3	0,9259259259	9	0,9519785078
4	0,8680555556	10	0,9302721574
5	0,8037551440	11	0,9044312642
6	0,9377143347	12	0,9636499779

d'où l'on conclut qu'en réalité, la première condition est toujours réalisée, et que la seconde l'est à partir de $n = 9$ (au lieu de 50). On notera les « irrégularités » dans le tableau précédent, dues au fait que la sommation se fait sur un domaine qui augmente d'une unité dès que n est un multiple de 3.

Si l'on cherchait une probabilité de 0,99, l'inégalité de Bienaymé-Tchebychev donnerait $n = 500$, tandis que le calcul exact de $p(n)$ donne $p(30) \approx 0,9932$, $p(31) \approx 0,9911$, $p(32) \approx 0,9884$ puis, à partir de $n = 33$, des valeurs $p(n) > 0,99$; ainsi, une grosse trentaine de tirages suffit pour assurer la condition cherchée avec une probabilité de 0,99.

7.3 FONCTIONS GÉNÉRATRICES

Les problèmes de comptage font apparaître naturellement des variables aléatoires à valeurs entières, pour lesquelles des méthodes spécifiques d'étude existent. Nous étudions ici les propriétés élémentaires des fonctions génératrices de telles variables aléatoires ; leur utilisation pratique sera exposée plus loin, au chapitre 8 pour la loi de la somme de deux variables aléatoires indépendantes, au chapitre 9 dans un problème d'attente de motif. Au chapitre 19, le lien entre fonctions génératrices et théorèmes de convergence sera étudié en détail.

Dans toute cette section, X est une variable aléatoire discrète prenant ses valeurs dans l'ensemble $\mathbb{N}$ des entiers positifs, et on note $p_n = \mathbf{P}\{X = n\}$ pour tout n entier.

§100 **Fonction génératrice**

Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On définit la **fonction génératrice** de X comme la somme de la série entière suivante :

$$G(t) = \sum_{n=0}^{\infty} p_n t^n.$$

Lorsque cette série converge absolument, la formule de transfert permet d'écrire

$$G(t) = \mathbf{E}(t^X).$$

Du fait des propriétés générales des séries entières, la fonction G jouit des propriétés élémentaires suivantes.

■ **Théorème 7.35** *Soit* X *une variable aléatoire discrète à valeurs dans* $\mathbb{N}$, *notons* G *sa fonction génératrice.*

i) Le rayon de convergence de la série entière $\sum p_n t^n$ *est au moins égal à* 1.

ii) G *est définie sur* $[-1\,;+1]$ *au moins, et de classe* $\mathscr{C}^\infty$ *sur* $]-1\,;+1[$ *au moins. On a de plus* $G(1) = 1$, $0 \leqslant G(t) \leqslant 1$ *pour tout* $t \in [0\,;1]$, *et* G *est absolument monotone sur* $[0\,;1]$ *: toutes ses dérivées successives sont positives. En particulier,* G *est croissante et convexe.*

iii) On peut récupérer les probabilités p_n au moyen des dérivées successives de G *en* 0 :

$$\forall n \in \mathbb{N} \qquad p_n = \mathbf{P}\{X = n\} = \frac{G^{(n)}(0)}{n!}.$$

iv) Notamment, la fonction génératrice caractérise *la loi discrète de* X *: deux variables aléatoires ayant même fonction génératrice ont même loi.*

DÉMONSTRATION : En $t = 1$, la série $\sum p_n t^n$ converge et a pour somme 1, donc le rayon de la série est supérieur ou égal à 1. De plus, pour tout entier k et tout réel $t \in [0\,;1]$,

$$G^{(k)}(t) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)\,p_n\,t^{n-k} \geqslant 0.$$

Les autres propriétés découlent des théorèmes généraux sur les séries entières (A39). ∎

Récupération des moments §101

Supposons que X admette une espérance, c'est-à-dire que la série positive $\sum n p_n$ converge. La série dérivée

$$G'(t) = \sum_{n=1}^{\infty} n\,p_n\,t^{n-1}$$

converge normalement, donc uniformément, sur $[0\,;1]$, ce qui prouve que G est de classe $\mathscr{C}^1$ sur $[0\,;1]$ et que la dérivée de G à gauche en 1 est

$$G'_g(1) = \sum_{n=1}^{\infty} n\,p_n = \mathbf{E}(X).$$

Réciproquement, si G admet une dérivée à gauche en 1, alors le théorème A.32 page 684 montre que $\sum n p_n$ converge et que

$$\mathbf{E}(X) = \sum_{n=0}^{\infty} n\,p_n = \lim_{t\to 1^-} G'(t) = G'_g(1),$$

la dernière étape provenant du théorème de dérivation au bord d'un intervalle.

Si X admet de plus un moment d'ordre 2, alors la série $\sum n^2 p_n$ converge, et donc la série $\sum n(n-1)\,p_n$ également. La série dérivée

$$G''(t) = \sum_{n=2}^{\infty} n(n-1)\,p_n\,t^n$$

converge normalement, et donc uniformément, sur $[0\,;1]$, donc G est de classe $\mathscr{C}^2$ sur $[0\,;1]$ et (en ajoutant un terme nul dans la somme)

$$\begin{aligned} G''(1) &= \sum_{n=1}^{\infty} n(n-1)\,p_n = \sum_{n=2}^{\infty} n^2\,p_n - \sum_{n=1}^{\infty} n\,p_n = \mathbf{E}\big(X(X-1)\big) \\ &= \mathbf{V}(X) + \mathbf{E}(X)^2 - \mathbf{E}(X). \end{aligned}$$

De même, le théorème A.32 permet de montrer la réciproque. En conclusion, on a prouvé le résultat suivant :

■ **Théorème 7.36 (Récupération des moments)**

i) X *admet une espérance si et seulement si* G *est dérivable à gauche en* 1, *et dans ce cas* $\mathbf{E}(X) = G'(1)$.

ii) X *admet une variance si et seulement si* G *est deux fois dérivable à gauche en* 1, *et dans ce cas* $\mathbf{V}(X) = G''(1) + G'(1) - G'(1)^2$.

iii) Plus généralement, X *admet un moment d'ordre* k *si et seulement si* G *est* k *fois dérivable à gauche en* 1, *et dans ce cas*

$$G^{(k)}(1) = \mathbf{E}\big(X(X-1)(X-2)\cdots(X-k+1)\big).$$

Remarque 7.37 (Nombres de Stirling) Les nombres de Stirling de première espèce S_n^k sont définis comme les coefficients de la décomposition de $X(X-1)(X-2)\cdots(X-n+1)$ dans la base canonique $1, X, X^2, \ldots, X^n$ (voir A56, page 691) ce qui permet d'écrire

$$G^{(n)}(1) = \sum_{k=0}^{n} S_n^k\, \mathbf{E}(X^k) = \sum_{k=0}^{n} S_n^k\, m_k.$$

La table de la page 692 permet donc d'écrire

$$G^{(2)}(1) = -m_1 + m_2 \qquad G^{(4)}(1) = -6\,m_1 + 11\,m_2 - 6\,m_3 + m_4.$$

Réciproquement, X^n se décompose dans la famille $\big(X, X(X-1), \ldots, X(X-1)\cdots(X-n+1)\big)$ avec pour coefficients les nombres de Stirling de seconde espèce, ce qui permet d'écrire

$$\mathbf{E}(X^n) = \sum_{n=0}^{\infty} s_n^k\, G^{(k)}(1).$$

▵7.13 La même table montre par exemple que
▵7.14

$$\mathbf{E}(X^4) = G'(1) + 7\,G''(1) + 6\,G'''(1) + G^{(4)}(1).$$

§ 102 **Fonctions génératrices des lois classiques**

Le lecteur est invité à calculer explicitement les fonctions génératrices des lois de Bernoulli, binomiales, géométriques et de Poisson ; on trouve les résultats résumés ci-dessous (exercice 7.14 page 168)

Loi	Désignation	$G(t)$	Domaine
Loi de Bernoulli	$\mathscr{B}(p)$	$1-p+pt$	$\mathbb{R}$
Loi binomiale	$\mathscr{B}(n,p)$	$(1-p+pt)^n$	$\mathbb{R}$
Loi géométrique	$\mathscr{G}_{\mathbb{N}^*}(p)$	$\dfrac{pt}{1-(1-p)t}$	$\left]-\dfrac{1}{1-p}\,;\,\dfrac{1}{1-p}\right[$
Loi de Poisson	$\mathscr{P}(\lambda)$	$e^{\lambda(t-1)}$	$\mathbb{R}$

TABLE 7.2 — *Fonctions génératrices des lois discrètes classiques.*

7.4 UN PROBLÈME CLASSIQUE : LE COLLECTIONNEUR

Position du problème § 103

Le problème du collectionneur de vignettes, connu dans le monde anglo-saxon sous le nom de *coupon collector's problem*, est le suivant. Rosencrantz achète un jour une tablette de chocolat et découvre, émerveillé, une image représentant Kolmogorov. Le marchand lui explique que chaque tablette contient une image, parmi les n de la collection « les grands probabilistes du XX^{e} siècle ». Rosencrantz décide de compléter sa collection. Sachant que les images sont réparties aléatoirement et uniformément dans les tablettes, combien Rosencrantz devra-t-il acheter de tablettes[12] ? Suivant la tradition, nous parlerons plutôt de temps d'attente, l'unité du « temps » discret correspondant à l'achat d'une nouvelle tablette.

Bien évidemment, si les premières images sont obtenues facilement[13], en revanche, les dernières images sont beaucoup plus difficiles à acquérir[14].

Ainsi, pour répondre à la question simple du temps d'attente avant de terminer la collection complète, la méthode la plus simple et la plus rapide est d'estimer les temps d'attente entre deux *nouvelles* acquisitions. On notera ainsi Θ_k le temps d'attente (nombre de tablettes à acheter) pour obtenir la k-ième image lorsque l'on en a déjà $k-1$. La loi de Θ_k est la loi géométrique de paramètre

$$p_k = \frac{n-k+1}{n}$$

puisque chaque nouvel achat offre une probabilité de succès égale au nombre d'images restant à collecter, à savoir $n-(k-1)$, sur le nombre total n d'images[15] ; l'espérance de Θ_k est donc $1/p_k$.

Enfin, le temps d'attente total étant $\mathrm{T}_n = \Theta_1 + \Theta_2 + \cdots + \Theta_n$, on en déduit par linéarité de l'espérance que

$$\mathbf{E}(\mathrm{T}_n) = \sum_{k=1}^{n} \mathbf{E}(\Theta_k) = \sum_{k=1}^{n} \frac{n}{n-k+1} = n\,\mathrm{H}_n,$$

où l'on a noté

$$\mathrm{H}_n = \sum_{k=1}^{n} \frac{1}{n-(k-1)} = \sum_{k=1}^{n} \frac{1}{k},$$

parfois appelé n-ième *nombre harmonique.*

Si par exemple la collection comprend $n = 20$ images, le nombre moyen de tablettes à acheter est

$$\mathbf{E}(\mathrm{T}_n) = \frac{279175675}{3879876} \approx 72.$$

Le raisonnement précédent permet également d'accéder à la variance de T, si l'on remarque que les variables $\Theta_1, \Theta_2, \ldots, \Theta_n$ sont indépendantes et en utilisant,

---

(12). Une façon alternative de voir ce problème : on se donne une urne contenant n boules numérotées, et on procède à une succession de tirages aléatoires avec remise. Combien de tirages sont-ils nécessaires pour avoir tiré, au moins une fois, chacune des boules ?

(13). Dès la première tablette, on obtient une première image ; à la seconde tablette, la probabilité d'augmenter sa collection est de $(n-1)/n$, puisqu'un doublon n'apparaît qu'avec la probabilité $1/n$.

(14). Une fois que l'on a $n-1$ images, la dernière n'est obtenue qu'avec une probabilité $1/n$, c'est-à-dire que le temps d'attente avant de l'obtenir est de n...

(15). Plus précisément : la loi de Θ_k *sachant* $\{\Theta_1 + \cdots + \Theta_{k-1} = i\}$ est égale à la loi géométrique de paramètre p_k, et ce pour toute valeur $i \geqslant k-1$. On en déduit (voir exercice 8.4 page 198) que Θ_k est indépendante de $\Theta_1 + \cdots + \Theta_{k-1}$ et que Θ_k suit la loi $\mathscr{G}(p_k)$.

par anticipation, l'additivité des variances de variables aléatoires indépendantes (théorème 8.21 page 186). On obtient

$$\mathbf{V}(\mathrm{T}_n) = \sum_{k=1}^{n} \mathbf{V}(\Theta_k) = \sum_{k=1}^{n} \frac{1-p_k}{p_k^2} = n^2\,\mathrm{R}_n - n\,\mathrm{H}_n,$$

où l'on a noté $\mathrm{R}_n = \sum_{k=1}^{n} 1/k^2$. On liste ci-dessous les valeurs de l'espérance et de l'écart-type de T_n pour quelques valeurs de n :

n	10	20	30	40	50	100	200
$\mathbf{E}(\mathrm{T}_n)$	29,3	72	120	171	224	519	1175
σ_{T_n}	11,2	23,8	36,5	49,2	62,0	126	254

Remarque 7.38 Dans la limite où n tend vers l'infini, on peut utiliser le développement asymptotique : $\mathrm{H}_n = \ln n + \gamma + \mathrm{o}(1)$, où $\gamma \approx 0,577$ est la constante d'Euler-Mascheroni (A7), ce qui donne

$$\mathbf{E}(\mathrm{T}_n) = n\,[\ln n + \gamma + \mathrm{o}(1)].$$

§ 104 **Loi du temps d'attente**

Il est possible de donner une formule explicite pour la loi de T_n. Pour cela, changeons de point de vue et notons N_k le nombre d'achats effectués avant d'obtenir l'image numéro k de la collection. Les variables $\mathrm{N}_1, \ldots, \mathrm{N}_n$ ont évidemment même loi, mais ne sont pas indépendantes[16]. On a alors

$$\mathrm{T}_n = \max(\mathrm{N}_1, \mathrm{N}_2, \ldots, \mathrm{N}_n),$$

et donc, pour tout entier $q \geqslant n$[17],

$$\mathbf{P}\{\mathrm{T}_n > q\} = \mathbf{P}\left(\bigcup_{k=1}^{n} \{\mathrm{N}_k > q\}\right).$$

Les événement $\{\mathrm{N}_k \geqslant q\}$ $(k = 1, \ldots, n)$ ne sont pas incompatibles, la probabilité de l'union n'est pas la somme des probabilités, et il faut se résoudre à utiliser la formule de Poincaré :

$$\mathbf{P}\{\mathrm{T}_n > q\} = \sum_{j=1}^{n} (-1)^{j+1} \left(\sum_{1 \leqslant i_1 < i_2 < \cdots < i_j \leqslant n} \mathbf{P}\{\mathrm{N}_{i_1} > q, \mathrm{N}_{i_2} > q, \ldots, \mathrm{N}_{i_j} > q\} \right). \tag{7.1}$$

Reste à interpréter l'événement $\{\mathrm{N}_{i_1} > q, \mathrm{N}_{i_2} > q, \ldots, \mathrm{N}_{i_j} > q\}$: il signifie que les images numérotées $i_1, \ldots, i_j$ n'ont pas été obtenues lors des q premiers achats, ou encore que, au cours de ces q premiers achats, on n'a obtenu que des images dont le numéro fait partie de l'ensemble $\{1, 2, \ldots, n\} \setminus \{i_1, \ldots, i_j\}$, qui contient $n - j$ indices. Les achats étant indépendants les uns des autres, on a donc

$$\mathbf{P}\{\mathrm{N}_{i_1} > q, \mathrm{N}_{i_2} > q, \ldots, \mathrm{N}_{i_j} > q\} = \left(\frac{n-j}{n}\right)^q.$$

(16). Par exemple, si $\mathrm{N}_3 = 1$, cela veut dire que la première image achetée était celle du probabiliste numéro 3, et donc on peut être sûr que $\mathrm{N}_1 > 1$.

(17). La collection ne pourra de toute façon pas être terminée avant au moins n achats !

Il ne reste plus qu'à terminer la sommation de l'équation (7.1) : il existe $\binom{n}{j}$ manières de choisir les indices $i_1, \dots, i_j$, donc

$$\mathbf{P}\{\mathrm{T}_n > q\} = \sum_{j=1}^{n} (-1)^{j+1} \binom{n}{j} \left(1 - \frac{j}{n}\right)^q. \tag{7.2}$$

La loi de T_n s'en déduit par soustraction : pour tout $q \geqslant n$,

$$\mathbf{P}\{\mathrm{T}_n = q\} = \mathbf{P}\{\mathrm{T}_n > q-1\} - \mathbf{P}\{\mathrm{T}_n > q\} = \sum_{j=1}^{n} (-1)^{j+1} \binom{n-1}{j-1} \left(1 - \frac{j}{n}\right)^{q-1}$$

Nous ne pouvons hélas guère faire mieux pour l'instant que cette formule. ⊿7.15

Une majoration de la probabilité de déviation § 105

En 1961, Paul Erdős et Alfred Rényi [31] ont apporté une contribution au problème du collectionneur[18]. Ils prouvèrent notamment que, pour tout réel x,

$$\mathbf{P}\{\mathrm{T}_n \leqslant n \ln n + xn\} \xrightarrow[n\to\infty]{} \exp\{-\mathrm{e}^{-x}\}.$$

Leur élégante preuve faisait appel aux fonctions caractéristiques (chapitre 17) ; nous nous contenterons ici d'une preuve plus élémentaire (mais plus longue également).

Fixons un réel x; on ne peut pas utiliser l'équation (7.2) en prenant $q = n \ln n + xn$, puisque cette valeur n'est pas entière. On note donc $\alpha_n = \lfloor n \ln n + xn \rfloor$, de telle sorte que

$$n \ln n + xn = \alpha_n + \mathrm{O}(1).$$

On reprend l'équation (7.2), en notant que

$$\begin{aligned} \mathbf{P}\{\mathrm{T}_n > n \ln n + xn\} &= \mathbf{P}\{\mathrm{T}_n > \alpha_n\} \\ &= \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \left(1 - \frac{k}{n}\right)^{\alpha_n}. \end{aligned} \tag{7.3}$$

On cherche maintenant à déterminer la limite, lorsque n tend vers l'infini, de cette quantité. Un simple développement du terme général de la somme mène à

$$\binom{n}{k} \left(1 - \frac{k}{n}\right)^{\alpha_n} = \binom{n}{k} \mathrm{e}^{-k \ln n - kx + \mathrm{O}(1/n)} \underset{n\to\infty}{\sim} \frac{\mathrm{e}^{-kx}}{k!}.$$

Sous réserve de pouvoir prendre « brutalement » la limite $n \to \infty$ dans l'équation (7.3), on obtient

$$\lim_{n\to\infty} \mathbf{P}\{\mathrm{T}_n > n \ln n + xn\} = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\mathrm{e}^{-kx}}{k!} = 1 - \exp\{-\mathrm{e}^{-x}\}$$

ce qui est identique au résultat de Erdős et Rényi.

Reste à justifier proprement ce passage à la limite. Pour cela, nous faisons appel à une version discrète du théorème de convergence dominée de Lebesgue :

■ **Lemme 7.39 (Convergence dominée discrète)** *Soient $v_n(k)$ $(n, k = 1, 2, \dots)$ des réels, tels que $v_n(k) \to v(k)$ quand n tend vers l'infini, pour tout k. On suppose de plus qu'il existe une suite positive $(w(k))_{k\geqslant 1}$ telle que $|v_n(k)| \leqslant w(k)$ pour tous k et $n \geqslant 1$, et telle que $\sum w(k)$ converge. Alors*

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} v_n(k) = \sum_{k=1}^{\infty} v(k).$$

Pour utiliser ce lemme, on écrit l'équation (7.3) sous la forme d'une somme infinie

$$\mathbf{P}\{\mathrm{T}_n > n \ln n + xn\} = \sum_{k=1}^{\infty} \mathbf{1}_{[1\,;\,n]}(k)\, (-1)^{k+1} \binom{n}{k} \left(1 - \frac{k}{n}\right)^{\alpha_n} = \sum_{k=1}^{\infty} v_n(k)$$

(18). Ou plutôt à un problème un peu plus général, dans lequel on désire avoir au moins m copies de chaque image.

et on majore les $v_n(k)$ en utilisant l'inégalité de convexité

$$\left(1-\frac{k}{n}\right)^{\alpha_n} \leqslant \mathrm{e}^{-k\alpha_n/n}$$

conjointement à l'inégalité

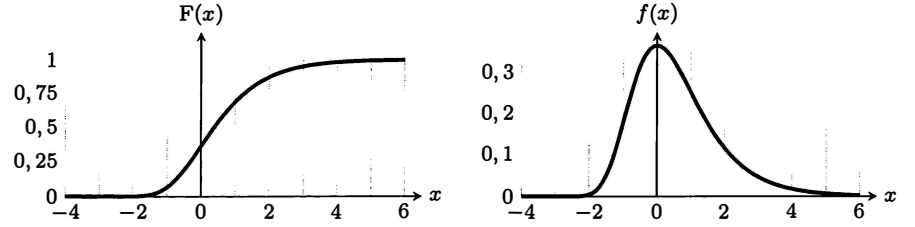
$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \leqslant \frac{n^k}{k!},$$

ce qui mène à

$$\left|v_n(k)\right| \leqslant \frac{n^k}{k!}\,\mathrm{e}^{-k\alpha_n/n} \leqslant \frac{n^k}{k!}\,\mathrm{e}^{-k(\ln n+x-1/n)} \leqslant \frac{n^k}{k!}\,\mathrm{e}^{-k(\ln n+x-1)} = \frac{\mathrm{e}^{-k(x-1)}}{k!} =: w_k.$$

La série $\sum w_k$ étant convergente, le théorème de convergence dominée s'applique, ce qui achève la preuve.

Remarque 7.40 (À propos de la loi de Gumbel) À la fonction de répartition $\mathrm{F}(x) = \exp(-\mathrm{e}^{-x})$ de la **loi de Gumbel** est associée la densité $f(x) = \mathrm{e}^{-x}\exp(-\mathrm{e}^{-x})$. Une variable admettant cette fonction pour densité admet des moments de tous ordres, a pour espérance $\mathbf{E}(\mathrm{X}) = \gamma \approx 0,577$ (constante d'Euler) et une variance valant $\mathbf{V}(\mathrm{X}) = \pi^2/6 \approx 1,645$.



EXERCICES

♦ ***Exercice 7.1 (Nombre de jambes)*** Expliquer la citation de la page 148.

♦ ***Exercice 7.2*** Soit $p \in]\,0\,;\,1\,[$. Soit X une variable aléatoire suivant la loi géométrique sur $\mathbb{N}^*$, de paramètre p. On pose $\mathrm{Y} = 1/\mathrm{X}$. Calculer $\mathbf{E}(\mathrm{Y})$.

♦ ***Exercice 7.3 (Dans l'ascenseur)*** Soit $n \in \mathbb{N}$. On suppose que n personnes montent dans l'ascenseur d'un immeuble de p étages. On suppose qu'elles se rendent chacune à un étage quelconque. On note X le nombre d'arrêts de l'ascenseur. Calculer $\mathbf{E}(\mathrm{X})$.

♦ ***Exercice 7.4 (Moments factoriels)*** Soit X une variable aléatoire suivant la loi de Poisson $\mathscr{P}(\lambda)$, où λ est un réel strictement positif. Calculer les **moments factoriels**

$$m'_n := \mathbf{E}\big(\mathrm{X}(\mathrm{X}-1)\cdots(\mathrm{X}-n+1)\big)$$

pour tout $n \geqslant 0$.

♦ ***Exercice 7.5 (Expression alternative de l'espérance et de la variance)*** Soit X une variable aléatoire à valeurs dans $\mathbb{N}$.

i) En remarquant que, pour tout $n \in \mathbb{N}$, on a $\sum_{k=0}^{n} k\,\mathbf{P}\{\mathrm{X}=k\} = \sum_{k=0}^{n-1} \mathbf{P}\{\mathrm{X}>k\} - n\,\mathbf{P}\{\mathrm{X}>n\}$, montrer que, si X admet une espérance, celle-ci peut s'exprimer sous la forme

$$\mathbf{E}(\mathrm{X}) = \sum_{n=0}^{\infty} \mathbf{P}\{\mathrm{X}>n\}\,.$$

ii) Montrer de même que, si X admet un moment d'ordre 2, alors

$$\mathbf{E}(X^2) = \sum_{n=0}^{\infty} (2n+1)\,\mathbf{P}\{X > n\}\,.$$

♦ ***Exercice 7.6 (Inégalité de Markov discrète)*** Soit Y une variable aléatoire discrète, positive, admettant une espérance. Pour tout $t > 0$, montrer que $t\,\mathbf{P}\{Y \geqslant t\} \leqslant \mathbf{E}(Y)$.

♦ ***Exercice 7.7 (Variable centrée réduite)*** Soit X une variable aléatoire discrète admettant une espérance m et un écart-type σ. On note $X^* = (X - m)/\sigma$ la variable centrée réduite issue de X.

Soient α, β des réels ; on pose $Y = \alpha X + \beta$. Exprimer la variable aléatoire centrée réduite Y^* en fonction de X^*.

♦ ***Exercice 7.8 (Variable quasi-certaine)*** On dit qu'une variable aléatoire réelle est *quasi-certaine* s'il existe un réel α tel que $\mathbf{P}\{X = \alpha\} = 1$. Montrer qu'il y a équivalence entre les énoncés

a. X admet une variance et $\mathbf{V}(X) = 0$;

b. X est quasi-certaine.

♦ ***Exercice 7.9*** Soit X une variable aléatoire suivant une loi de Poisson $\mathscr{P}(\lambda)$. Montrer que

$$\mathbf{P}\left\{X \leqslant \frac{\lambda}{2}\right\} \leqslant \frac{4}{\lambda} \qquad \text{et} \qquad \mathbf{P}\{X \geqslant 2\lambda\} \leqslant \frac{1}{\lambda}.$$

♦ ***Exercice 7.10*** En utilisant l'inégalité de Bienaymé-Tchebychev, montrer que, pour tout n entier, on a

$$\sum_{k=n+1}^{3n-1} \binom{4n}{k} \geqslant \frac{n-1}{n}\, 2^{4n}.$$

♦ ***Exercice 7.11 (Inégalités de Hölder, cas discret)*** Pour tout réel $p \geqslant 1$, notons $\mathcal{L}^p$ l'ensemble des variables aléatoires réelles discrètes sur un même espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ telles que $|X|^p$ admette une espérance.

Soient p et q deux réels supérieurs ou égaux à 1, et *conjugués*, c'est-à-dire que $1/p + 1/q = 1$. Soient $X \in \mathcal{L}^p$ et $Y \in \mathcal{L}^q$.

i) Montrer que, pour tous $a, b > 0$,

$$0 \leqslant ab \leqslant \frac{a^p}{p} + \frac{b^q}{q}.$$

ii) En déduire que $XY \in \mathcal{L}^1$ et que $\mathbf{E}|XY| \leqslant \left(\mathbf{E}|X^p|\right)^{1/p} \cdot \left(\mathbf{E}|Y^q|\right)^{1/q}$.

♦ ***Exercice 7.12 (Seconde fonction génératrice)*** L'exercice 7.5 page ci-contre montre que l'espérance d'une variable aléatoire X à valeurs dans $\mathbb{N}$ peut s'exprimer en fonction des événements $\{X > n\}$ pour $n \in \mathbb{N}$. Par extension, on définit la seconde fonction génératrice H de la variable aléatoire X comme la somme de la série entière

$$H(t) = \sum_{n=0}^{\infty} \mathbf{P}\{X > n\}\; t^n.$$

Montrer que H est définie (au moins) sur $]-1\,;\,+1[$ et que, pour tout $t \in]-1\,;\,+1[$,

$$H(t) = \frac{1 - G(t)}{1 - t}.$$

♦ ***Exercice 7.13*** Soit X une variable aléatoire suivant une loi de Poisson de paramètre $\lambda > 0$. On note $G_X(t) = \sum_{n=0}^{\infty} \mathbf{P}\{X = n\}\; t^n = e^{\lambda(t-1)}$ sa fonction génératrice.

i) Montrer, pour tout $t \geqslant 1$ et pour tout réel a, l'inégalité $\mathbf{P}\{X \geqslant a\} \leqslant \dfrac{G_X(t)}{t^a}$.

ii) En déduire que $\mathbf{P}\{X \geqslant 2\lambda\} \leqslant \left(\frac{e}{4}\right)^{\lambda}$. Comparer avec une majoration obtenue par l'inégalité de Bienaymé-Tchebychev.

♦ ***Exercice 7.14 (Fonctions génératrices classiques)***
Calculer les fonctions génératrices des lois classiques : binomiale, géométrique, de Poisson.

♦ ***Exercice 7.15 (Le capitaine pirate)*** *[Le capitaine pirate veut acheter un magnifique galion à Cutlass Liz.]*

— Il ne vous décevra pas, capitaine pirate, dit Cutlass Liz en lui tirant gentiment la barbe. Comment comptez-vous payer ? En doublons ou en trésors ?

Le capitaine pirate marqua une pause.

— Aaaar. Le problème, c'est que mes trésors ne sont pas disponibles en ce moment. Il va bien me falloir deux semaines pour les récupérer.

— Vous ne m'aviez pas dit que vous les conserviez à bord de votre bateau ?

— Oui. C'est-à-dire, oui... euh... j'ai environ une centaine de coffres, répondit le capitaine pirate en regardant pensivement ses pieds, mais ça implique que j'ai aussi une centaine de clés différentes. Pour des raisons évidentes de sécurité, je ne les ai pas étiquetées. Je n'ai donc aucun moyen de savoir quelle clé correspond à quel coffre.

Le capitaine pirate était assez content de cette explication, mais Cutlass Liz fronça les sourcils.

— Ouvrir un coffre n'est quand même pas si long ?

— Ah, c'est ce que vous croyez, hein ? Mais — ah — vous partez du principe que je respecte une sorte d'ordre pour ouvrir les coffres. Alors qu'en réalité, je ne fais que choisir les clés et les coffres au hasard, sans vérifier si j'ai déjà essayé telle clé ou tel coffre auparavant. Ça peut prendre des jours.

Gideon DEFOE. *Les pirates ! dans Une aventure avec les baleines.*

En admettant que Cutlass Liz se laisse berner par le capitaine pirate, et à raison d'une minute par essai (clé + coffre), combien de temps va-t-elle laisser au capitaine pour payer le galion ? On étudiera deux cas, selon que le capitaine pirate joue la sécurité face au sabre de Cutlass Liz (à chaque ouverture, il laisse la clé sur le coffre et ne touche plus à l'un ni à l'autre jusqu'à la fin), ou qu'au contraire il pousse la mauvaise foi jusqu'à réessayer un coffre *déjà* ouvert.

Indications

◊ *Indication 7.2 :* Utiliser la formule de transfert.

◊ *Indication 7.3 :* On pourra écrire X comme une somme de variables aléatoires de Bernoulli.

◊ *Indication 7.5 :* Dans la deuxième question, remarquer que $2n+1 = (n+1)^2 - n^2$.

◊ *Indication 7.6 :* Noter $Y(\Omega) = \{y_k \ ; \ k \in K\}$ et $K^+ = \{k \in K \ ; \ y_k \geqslant t\}$; exprimer $\mathbf{P}\{Y \geqslant t\}$ comme une somme sur K^+.

◊ *Indication 7.7 :* Séparer les cas selon le signe de α. Si $\alpha > 0$ vérifier que $Y^* = X^*$.

◊ *Indication 7.9 :* Commencer par remarquer que $\left\{X \leqslant \frac{\lambda}{2}\right\} \subset \left\{|X-\lambda| \geqslant \frac{\lambda}{2}\right\}$ et $\{X \geqslant 2\lambda\} \subset \{|X-\lambda| \geqslant \lambda\}$. Il suffit ensuite d'appliquer l'inégalité de Bienaymé-Tchebychev.

◊ *Indication 7.10 :* Considérer une variable de loi binomiale $\mathscr{B}\left(4n \ ; \ \frac{1}{2}\right)$.

◊ *Indication 7.11 :*

i) Utiliser la concavité de la fonction logarithme.

ii) Si $X = \sum x_k \mathbf{1}_{A_k}$ et $Y = \sum y_k \mathbf{1}_{A_k}$, appliquer l'inégalité précédente aus réels

$$a = x_k p_k^{1/p} / (\mathbf{E}|X|^p)^{1/p} \qquad \text{et} \qquad b = y_k p_k^{1/q} / (\mathbf{E}|Y|^p)^{1/p}.$$

◊ *Indication 7.15 :* Appeler T_k le temps (nombre d'essais) nécessaire pour ouvrir le k-ième coffre une fois que les $k-1$ premiers sont ouverts ; la loi de T_k est une des lois classique. En déduire l'espérance de $\Theta = T_1 + T_2 + \cdots + T_n$.

SOLUTIONS

♦ **Solution 7.1** Puisqu'il existe quelques unijambistes, le nombre moyen de jambes dans la population doit être quelque chose comme 1,99 ; l'écrasante majorité des gens a un nombre de jambes égal à 2, donc strictement supérieur à la moyenne.

♦ **Solution 7.2** La formule de transfert conduit à :

$$\mathbf{E}(Y) = \sum_{n=1}^{\infty} \frac{1}{n}\, p\,(1-p)^{n-1} = -\frac{p \ln p}{1-p},$$

la convergence absolue de la série étant assurée par comparaison avec une série géométrique de raison $(1-p)$.

♦ **Solution 7.3** Notons T_k la variable aléatoire valant 1 si au moins une personne s'arrête à l'étage k, et 0 sinon. La probabilité que T_k soit égale à 0 est donc celle que : aucune des n personnes ne s'arrête à l'étage k. Ainsi,

$$\mathbf{P}\{T_k = 0\} = \left(\frac{p-1}{p}\right)^n.$$

On en déduit que T_k suit la loi de Bernoulli de paramètre $q := 1 - (1-1/p)^n$. Enfin, on peut écrire $X = T_1 + T_2 + \cdots + T_p$. Bien entendu, les variables T_k ne sont pas indépendantes, mais cela n'a aucune importance dans le calcul de l'espérance : par linéarité, on a

$$\mathbf{E}(X) = \sum_{k=1}^{p} \mathbf{E}(T_k) = pq = p - p\left(1 - \frac{1}{p}\right)^n.$$

♦ **Solution 7.4** Pour tout $n \geqslant 0$,

$$m'_n = \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} k(k-1)\cdots(k-n+1) \cdot \frac{\lambda^k}{k!} \qquad \textit{formule de transfert}$$

$$= \mathrm{e}^{-\lambda} \sum_{k=n}^{\infty} \frac{\lambda^k}{(k-n)!} = \lambda^n \mathrm{e}^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^k}{k!} = \lambda^n \qquad \textit{changement d'indice.}$$

♦ **Solution 7.5**

i) Pour établir la première relation, on écrit que $\mathbf{P}\{X = k\} = \mathbf{P}\{X > k-1\} - \mathbf{P}\{X > k\}$; un changement d'indice permet de conclure par télescopage.
Il suffit ensuite de montrer que $n\,\mathbf{P}\{X > n\} \to 0$. Notons R_n le reste d'ordre n de la série positive convergente $\sum k\,\mathbf{P}\{X = k\}$. Pour tout $n \in \mathbb{N}$, on a

$$0 \leqslant n\,\mathbf{P}\{X > n\} = n \sum_{k=n+1}^{\infty} \mathbf{P}\{X = k\} \leqslant \sum_{k=n+1}^{\infty} k\,\mathbf{P}\{X = k\} = R_n$$

et puisque la suite des restes converge vers 0, la limite cherchée est prouvée ; la formule s'ensuit.
Elle se comprend intuitivement ainsi, en notant $p_n := \mathbf{P}\{X = n\}$:

$$\begin{aligned} \sum_{n=0}^{\infty} \mathbf{P}\{N > n\} = {} & p_1 + p_2 + p_3 + p_4 + \cdots \\ & \quad p_2 + p_3 + p_4 + \cdots \\ & \qquad\quad p_3 + p_4 + \cdots \\ & \qquad\qquad\quad p_4 + \cdots \\ = {} & p_1 + 2p_2 + 3p_3 + 4p_4 + \cdots = \mathbf{E}(X) \end{aligned}$$

ii) Puisque $2n+1 = (n+1)^2 - n^2$, on a, pour tout entier N,

$$\begin{aligned} \sum_{n=0}^{N} (2n+1)\,\mathbf{P}\{X > n\} &= \sum_{n=0}^{N} (n+1)^2\,\mathbf{P}\{X > n\} - \sum_{n=1}^{N} n^2\,\mathbf{P}\{X > n\} \\ &= \sum_{n=1}^{N+1} n^2\,\mathbf{P}\{X > n-1\} - \sum_{n=1}^{N} n^2\,\mathbf{P}\{X > n\} \quad \textit{(changement d'indice)} \\ &= \sum_{n=1}^{N} n^2 \Big(\mathbf{P}\{X > n-1\} - \mathbf{P}\{X > n\}\Big) + (N+1)^2\,\mathbf{P}\{X > N\} \\ &= \sum_{n=1}^{N} n^2\,\mathbf{P}\{X = n\} + (N+1)^2\,\mathbf{P}\{X > N\}. \end{aligned}$$

Enfin, la majoration

$$(N+1)^2\,\mathbf{P}\{X>N\} = (N+1)^2 \sum_{n=N+1}^{\infty} \mathbf{P}\{X=n\} \leqslant \sum_{n=N+1}^{\infty} n^2\,\mathbf{P}\{X=n\} \xrightarrow[N\to\infty]{} 0$$

achève la démonstration.

♦ **Solution 7.6** Notons $Y(\Omega) = \{y_k\ ;\ k \in K\}$. Certains de ces y_k sont plus grands que t, d'autres non. Notons K^+ l'ensemble des indices k tels que $y_k \geqslant t$. On a alors

$$\begin{aligned} t\,\mathbf{P}\{Y \geqslant t\} &= t \sum_{k\in K^+} \mathbf{P}\{Y=y_k\} \leqslant \sum_{k\in K^+} y_k\,\mathbf{P}\{Y=y_k\} && \text{car } y_k \geqslant t \\ &\leqslant \sum_{k\in K} y_k\,\mathbf{P}\{Y=y_k\} = \mathbf{E}(Y). && \text{car } y_k \geqslant 0 \end{aligned}$$

♦ **Solution 7.7** Si $\alpha > 0$, alors $Y^* = X^*$. Si $\alpha < 0$, alors $Y^* = -X^*$. Si $\alpha = 0$, il n'est pas possible de réduire Y.

♦ **Solution 7.8** On note $X(\Omega) = \{x_k\ ;\ k \in \mathbb{N}\}$ et $p_k = \mathbf{P}\{X = x_k\}$. Tout d'abord, si X est quasi-certaine, alors tous les p_k sont nuls à l'exception d'un seul d'entre eux, noté p_{k_0} et valant 1. L'espérance de X est alors x_{k_0} et sa variance vaut $\mathbf{V}(X) = \sum_k p_k\,(x_k - x_{k_0})^2 = 0$.

Réciproquement, supposons que X admet une variance nulle ; elle admet donc une espérance que l'on notera m. On en déduit que $\sum_{n=0}^{\infty} (x_k - m)^2\,p_k = 0$; par positivité des termes de la somme, on doit avoir $(x_k - m)^2\,p_k = 0$ pour tout entier n. Notamment, pour tout entier k tel que $p_k \neq 0$, on a $x_k = m$; X est donc quasi-certaine et $\mathbf{P}\{X = m\} = 1$.

Ce résultat se généralise aux variables autres que discrètes ; cf. exercice 14.4 page 330.

♦ **Solution 7.10** Soit n un entier, et considérons une variable X_n de loi binomiale $\mathscr{B}\left(4n\ ;\ \frac{1}{2}\right)$. Cette variable a pour espérance $\mathbf{E}(X_n) = 2n$ et pour variance $\sigma^2_{X_n} = n$, donc

$$\sum_{k=n+1}^{3n-1} \binom{4n}{k} \frac{1}{2^{4n}} = \mathbf{P}\{n < X_n < 3n\} = \mathbf{P}\{|X_n - 2n| < n\} = 1 - \mathbf{P}\{|X_n - 2n| \geqslant n\}$$

$$\geqslant 1 - \frac{\sigma^2_{X_n}}{n^2} = 1 - \frac{1}{n} = \frac{n-1}{n}$$

par l'inégalité de Bienaymé-Tchebychev.

♦ **Solution 7.11**

i) Par concavité de la fonction logarithme, on a

$$\ln\left(\frac{a^p}{p} + \frac{b^q}{q}\right) \geqslant \frac{\ln a^p}{p} + \frac{\ln b^q}{q} = \ln ab,$$

donc

$$0 \leqslant ab \leqslant \frac{a^p}{p} + \frac{b^q}{q}.$$

ii) On choisit une partition $\Omega = \sum A_k$ permettant de décomposer simultanément les variables *positives* $|X|$ et $|Y|$:

$$|X| = \sum_{k\in K} x_k\,\mathbf{1}_{A_k} \qquad |Y| = \sum_{k\in K} y_k\,\mathbf{1}_{A_k} \qquad |XY| = \sum_{k\in K} x_k\,y_k\,\mathbf{1}_{A_k}.$$

Notons enfin $p_k = \mathbf{P}(A_k)$: le but est maintenant de démontrer l'inégalité

$$\sum_{k\in K} x_k\,y_k\,p_k \leqslant \left(\sum_{k\in K} x_k^p\,p_k\right)^{1/p} \left(\sum_{k\in K} y_k^q\,p_k\right)^{1/q}.$$

Remarquons d'abord que, si l'un des facteurs du membre de droite vaut 0, c'est que l'une des variables est presque sûrement nulle, auquel cas l'inégalité est évidemment vérifiée. On suppose désormais que ces deux facteurs sont strictement positifs.

L'inégalité précédente, appliquée à $a = x_k\,p_k^{1/p}/(\mathbf{E}\,|X|^p)^{1/p}$ et $b = y_k\,p_k^{1/q}/(\mathbf{E}\,|Y|^p)^{1/p}$, donne

$$\frac{x_k\,y_k\,p_k}{\left(\mathbf{E}\,|X|^p\right)^{1/p}\left(\mathbf{E}\,|Y|^q\right)^{1/q}} \leqslant \frac{1}{p}\frac{x_k^p\,p_k}{\mathbf{E}\,|X|^p} + \frac{1}{q}\frac{y_k^q\,p_k}{\mathbf{E}\,|Y|^q}.$$

Le membre de droite est le terme général d'une série convergente puisque $X \in \mathcal{L}^p$ et $Y \in \mathcal{L}^q$; le membre de gauche est donc terme général d'une série convergente, donc $XY \in \mathcal{L}^1$ et, en sommant l'inégalité précédente, on obtient

$$\frac{\mathbf{E}\,|XY|}{(\mathbf{E}\,|X|^p)^{1/p}(\mathbf{E}\,|Y|^q)^{1/q}} \leqslant \frac{1}{p} + \frac{1}{q} = 1.$$

♦ **Solution 7.12** Notons $a_n = \mathbf{P}\{X = n\}$ et $r_n = \mathbf{P}\{X > n\}$ pour tout $n \in \mathbb{N}$. Alors la suite $(r_n)_{n\geqslant 0}$ est la suite des restes de la série $\sum a_n$, c'est-à-dire que, pour tout $n \in \mathbb{N}$,

$$r_n = \sum_{k=n+1}^{\infty} a_k \qquad \text{et} \qquad a_n = r_{n-1} - r_n,$$

cette dernière relation n'étant valable pour $n = 0$ qu'en définissant $r_{-1} := a_0 + r_0 = 1$. La suite $(r_n)_{n\geqslant 0}$ étant majorée ($0 \leqslant r_n \leqslant 1$), la série entière $\sum r_n t^n$ est de rayon au moins égal à 1, donc converge absolument sur $]-1\,;\,+1[$. De plus, pour tout $t \in]-1\,;\,+1[$, on a

$$(1-t)\,\mathrm{H}(t) = \sum_{n=0}^{\infty} r_n t^n - \sum_{n=0}^{\infty} r_n t^{n+1} = r_0 + \sum_{n=1}^{\infty} (r_n - r_{n-1})\,t^n = r_0 + a_0 - \mathrm{G}(t) = 1 - \mathrm{G}(t).$$

♦ **Solution 7.13**

i) Notons $N = \lceil a \rceil$ le plus petit entier supérieur à a (de sorte que $N - 1 < a \leqslant N$) ; alors

$$\begin{aligned} t^a\,\mathbf{P}\{X \geqslant a\} = t^a\,\mathbf{P}\{X \geqslant N\} &= t^a \sum_{n=N}^{\infty} \mathbf{P}\{X = n\} \\ &\leqslant \sum_{n=N}^{\infty} \mathbf{P}\{X = n\}\; t^n \qquad\qquad \textit{car } t \geqslant 1 \textit{ et } n \geqslant N \geqslant a \\ &\leqslant \sum_{n=0}^{\infty} \mathbf{P}\{X = n\}\; t^n = \mathrm{G}_X(t) \quad \textit{par positivité des termes.} \end{aligned}$$

ii) La valeur de t qui minimise l'expression

$$\frac{\mathrm{G}_X(t)}{t^{2\lambda}} = \frac{\mathrm{e}^{\lambda(t-1)}}{t^{2\lambda}} = \left(\frac{\mathrm{e}^{t-1}}{t^2}\right)^{\lambda},$$

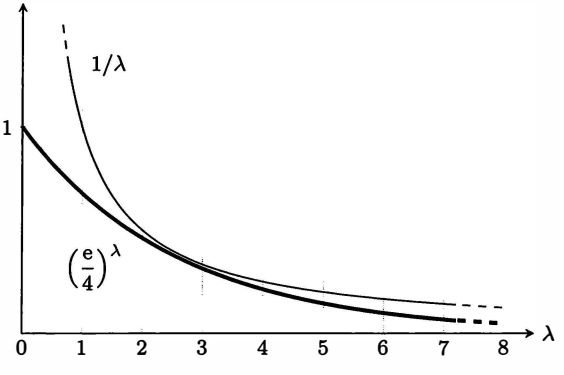
est $t = 2$; on applique l'inégalité précédente en ce point-là, on obtient l'inégalité voulue. L'inégalité de Bienaymé-Tchebychev mène, quant à elle, à

$$\mathbf{P}\{X \geqslant 2\lambda\} < \mathbf{P}\{\,|X - \lambda| \geqslant \lambda\} \leqslant \frac{1}{\lambda}$$

Pour $\lambda \in]\,0\,;\,1\,]$, elle ne nous apprend rien d'intéressant, alors que l'inégalité précédente, oui. De plus, on a

$$\forall \lambda > 0 \qquad \left(\frac{\mathrm{e}}{4}\right)^{\lambda} \leqslant \frac{1}{\lambda}$$

(ce que l'on montre par simple étude de la fonction $\varphi : \lambda \mapsto \lambda\,(\mathrm{e}/4)^{\lambda}$; c'est une fonction positive, admettant un maximum en $1/\ln(4/\mathrm{e}) \approx 2,59$, et ce maximum est inférieur à 1). Ainsi, l'approximation obtenue est meilleure que celle de Bienaymé-Tchebychev.
Si l'on trace les deux fonctions $\lambda \mapsto 1/\lambda$ et $\lambda \mapsto (\mathrm{e}/4)^{\lambda}$, on obtient les graphes ci-dessous :



♦ **Solution 7.14** Si $X \rightsquigarrow \mathscr{B}(n\,;\,p)$, alors $G(t) = \sum_{k=0}^{n} \binom{n}{k} (pt)^k (1-p)^{n-k} = (pt+1-p)^n$ d'après la formule du binôme de Newton.

Si $X \rightsquigarrow \mathscr{P}(\lambda)$, alors

$$G(t) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k t^k}{k!} = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}.$$

Enfin, si $X \rightsquigarrow \mathscr{G}(p)$, alors

$$G(t) = \sum_{k=1}^{\infty} p(1-p)^{k-1} t^k = pt \sum_{k=0}^{\infty} (1-p)^k t^k = \frac{pt}{1-(1-p)t}.$$

♦ **Solution 7.15** Notons $n = 100$ le nombre de coffres (et de clés), et traitons d'abord le cas où le Capitaine Pirate joue la sécurité. Tant qu'aucun coffre n'est ouvert, chaque essai (un coffre + une clé, choisis chacun aléatoirement), a une probabilité $p_1 = 1/n$ d'aboutir à l'ouverture : en effet, une fois choisi un coffre, la probabilité de choisir la clé correspondante[19] est $1/n$. Ainsi, T_1, qui est le temps d'attente du premier succès suit une loi géométrique de paramètre $p_1 = 1/n$, et $\mathbf{E}(T_1) = n$.

Une fois le premier coffre ouvert (il reste ouvert et sa clé reste dessus), on trouve dans la même situation, mais avec $n-1$ coffres et $n-1$ clés. Ainsi, T_2 suit une loi géométrique de paramètre $p_2 = 1/(n-1)$, et $\mathbf{E}(T_2) = n-1$.

De manière générale, $T_k \rightsquigarrow \mathscr{G}\left(\frac{1}{n-k+1}\right)$ et $\mathbf{E}(T_k) = n-k+1$.

Le temps total d'attente est $\Theta = T_1 + T_2 + \cdots + T_n$; par linéarité de l'espérance, le temps moyen d'attente en minutes est donc

$$\mathbf{E}(\Theta) = \sum_{k=1}^{n} \mathbf{E}(T_k) = n + (n-1) + \cdots + 2 + 1 = \frac{n(n+1)}{2} = 5\,050,$$

(résultat que n'eût pas désapprouvé le jeune Carl Friedrich), soit trois jours et demi pleins. Cutlass Liz risque de perdre patience avant la fin.

Si le capitaine pirate avait été d'encore plus mauvaise foi, et avait refermé chaque coffre vidé et remis la clé avec les autres, tirant ainsi à chaque fois un coffre parmi les 100 et une clé parmi les 100, le temps d'attente moyen aurait été plus long[20] :

$$\mathbf{E}(\Theta') = n^2 \sum_{k=1}^{n} \frac{1}{k} \approx 51\,874$$

soit un peu plus de 36 jours.

---

(19). On peut également raisonner ainsi : la probabilité d'ouvrir un coffre donné est $1/n^2$ par indépendance du choix du coffre et de la clé. Les événements « on ouvre le coffre numéro k » sont deux à deux disjoints ; la probabilité d'en ouvrir au moins 1 est la somme $\sum_{k=1}^{n} 1/n^2 = 1/n$.

(20). Il s'agirait alors d'une simple variante du problème du collectionneur, détaillé à la section 7.4.

Chapitre 8

Couples de variables aléatoires discrètes

L'exemple le plus simple de vecteur aléatoire est donné par un couple (X, Y) *de variables aléatoires discrètes. Étudier la loi d'un tel couple revient donc à déterminer les probabilités élémentaires notées indifféremment*

$$\mathbf{P}\{X = x, Y = y\}, \qquad \mathbf{P}\big(\{X = x\} \cap \{Y = y\}\big) \quad ou \quad \mathbf{P}\big\{(X, Y) = (x, y)\big\}.$$

- *Dans une première partie, nous relions la loi du couple aux* lois marginales, *qui sont les lois des seules variables* X *ou* Y *; nous étudions également la loi de variables de la forme* $\varphi(X, Y)$.
- *Nous nous concentrons ensuite sur le cas particulier des variables indépendantes, notamment sur l'espérance* $\mathbf{E}(XY)$ *de leur produit et sur la loi de leur somme* $X + Y$.
- *Nous enchaînons sur l'étude de la covariance d'un couple et sur la définition de son coefficient de corrélation.*
- *Enfin, nous abordons la notion d'espérance conditionnelle qui, dans le cas de variables discrètes, ne pose aucune difficulté théorique. Cette thématique sera reprise dans la suite (chapitre 23).*
- *Pour terminer le chapitre, on traite un exemple de sommation d'un nombre aléatoire de variables aléatoires, avec application à l'étude de l'extinction d'une espèce ou d'un nom de famille.*

Position du problème

§ 106

On se place dans un espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ et on suppose que l'on a deux variables aléatoires réelles discrètes, X et Y. Connaître la loi de X et la loi de Y ne permet pas, en général, de déterminer la probabilité que le couple (X, Y) prenne une valeur (x, y) donnée.

Donnons un exemple. On définit l'univers $\Omega := \{0, 1, 2\} \times \{0, 1, 2\}$, et on suppose que l'on connaît deux variables aléatoires X et Y de même loi

$$\mathbf{P}\{X = 0\} = \mathbf{P}\{Y = 0\} = 0,3$$
$$\mathbf{P}\{X = 1\} = \mathbf{P}\{Y = 1\} = 0,5$$
$$\mathbf{P}\{X = 2\} = \mathbf{P}\{Y = 2\} = 0,2$$

Cette configuration peut être obtenue de manières très différentes. Dans un premier cas, nous pouvons avoir par exemple les probabilités suivantes pour les valeurs prises simultanément par X et Y, résumées sous forme d'un tableau :

P	$\{Y=0\}$	$\{Y=1\}$	$\{Y=2\}$
$\{X=0\}$	0,3	0	0
$\{X=1\}$	0	0,5	0
$\{X=2\}$	0	0	0,2

Ici, les variables aléatoires X et Y sont égales (elles prennent toujours les mêmes valeurs). Mais on peut imaginer bien d'autres cas, par exemple

P	$\{Y=0\}$	$\{Y=1\}$	$\{Y=2\}$
$\{X=0\}$	0,1	0	0,2
$\{X=1\}$	0,1	0,4	0
$\{X=2\}$	0,1	0,1	0

La formule des probabilités totales permet, pour chaque ligne du tableau, d'interpréter la somme des coefficients comme une valeur $\mathbf{P}\{X=k\}$, c'est-à-dire de retrouver la loi de X seul (ce qu'on appelle la **loi marginale de X**) ; de même, la sommes des coefficients sur chaque colonne nous donne la loi de Y :

P	$\{Y=0\}$	$\{Y=1\}$	$\{Y=2\}$	loi de X
$\{X=0\}$	0,1	0	0,2	**0,3**
$\{X=1\}$	0,1	0,4	0	**0,5**
$\{X=2\}$	0,1	0,1	0	**0,2**
loi de Y	**0,3**	**0,5**	**0,2**	**1**

Cet exemple permet de voir le premier résultat important de l'étude de couples aléatoires : les lois de chaque variable aléatoire ne permettent pas de déterminer la loi du couple ; en revanche, la loi du couple permet de retrouver les lois marginales.

8.1 LOI D'UN COUPLE

§ 107 **Loi conjointe, lois marginales**

Soient X et Y deux variables aléatoires réelles discrètes. L'ensemble

$$(X,Y)(\Omega) = \Big\{\big(X(\omega),Y(\omega)\big)\ ;\ \omega\in\Omega\Big\}$$

étant inclus dans $X(\Omega)\times Y(\Omega)$, c'est une partie dénombrable de $\mathbb{R}^2$; d'après un raisonnement déjà fait, connaître la loi du couple (X, Y) (qu'on appelle également **loi conjointe**), c'est-à-dire la probabilité image

$$\begin{aligned}\mathbf{P}_{X,Y}:\ \mathfrak{P}(\mathbb{R}^2) &\longrightarrow \mathbb{R}^+\\ A &\longmapsto \mathbf{P}\big\{(X,Y)\in A\big\},\end{aligned}$$

revient en pratique à connaître les probabilités atomiques, c'est-à-dire les quantités

$$\mathbf{P}\{(X, Y) = (x_k, y_\ell)\} \qquad (x_k, y_\ell) \in X(\Omega) \times Y(\Omega).$$

Dans la suite, nous noterons

$$X(\Omega) = \{x_k \ ; \ k \in K\} \qquad Y(\Omega) = \{y_\ell \ ; \ \ell \in L\}$$

et

$$p_{k,\ell} = \mathbf{P}\{(X, Y) = (x_k, y_\ell)\} \qquad \text{ou} \qquad p_{k,\ell} = \mathbf{P}\{X = x_k, Y = y_\ell\}.$$

La formule des probabilités totales mène immédiatement à la relation

$$\sum_{k \in K} \sum_{\ell \in L} p_{k,\ell} = 1.$$

Si l'on note $p_{k,\cdot} = \mathbf{P}\{X = k\}$, on a de même

$$p_{k,\cdot} = \sum_{\ell \in L} \mathbf{P}\{(X, Y) = (x_k, y_\ell)\} = \sum_{\ell \in L} p_{k,\ell}.$$

La loi de X seule, caractérisée comme on l'a vu par les probabilités atomiques $p_{k,\cdot}$, est appelée **loi marginale en X** du couple (X, Y). Les variables X et Y sont appellées **marginales** du couple. De même, si l'on note $p_{\cdot,\ell} = \mathbf{P}\{Y = y_\ell\}$, on a

$$p_{\cdot,\ell} = \sum_{k \in K} \mathbf{P}\{(X, Y) = (k, \ell)\} = \sum_{k \in K} p_{k,\ell},$$

et la connaissance des $p_{\cdot,\ell}$ pour $\ell \in L$ caractérise la loi de Y, appelée **loi marginale en Y** du couple (X, Y).

On peut dresser un tableau résumant la situation, en faisant apparaître en *marge* du tableau les sommes des coefficients sur chaque ligne et sur chaque colonne :

$\mathbf{P}$	$\{Y = y_1\}$	$\{Y = y_2\}$	$\cdots$	$\{Y = y_j\}$	$\cdots$	loi de X
$\{X = x_1\}$	$p_{1,1}$	$p_{1,2}$	$\cdots$	$p_{1,j}$	$\cdots$	$\boldsymbol{p_{1,\cdot}}$
$\{X = x_2\}$	$p_{2,1}$	$p_{2,2}$	$\cdots$	$p_{2,j}$	$\cdots$	$\boldsymbol{p_{2,\cdot}}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$\{X = x_k\}$	$p_{i,1}$	$p_{i,2}$		$p_{i,j}$	$\cdots$	$\boldsymbol{p_{i,\cdot}}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
loi de Y	$\boldsymbol{p_{\cdot,1}}$	$\boldsymbol{p_{\cdot,2}}$	$\cdots$	$\boldsymbol{p_{\cdot,j}}$	$\cdots$	$\mathbf{1}$

Lois conditionnelles § 108

Soit (X, Y) un couple de variables aléatoires discrètes. Pour tout $y_\ell \in Y(\Omega)$ tel que l'événement $\{Y = y_\ell\}$ n'est pas négligeable, la **loi conditionnelle de X sachant $\{Y = y_\ell\}$** est la loi de X dans l'espace probabilisé $(\Omega, \mathcal{T}, \mathbf{P}_{\{Y=\ell\}})$, qui est caractérisée par les probabilités atomiques

$$\mathbf{P}\{X = x_k \mid Y = y_\ell\} = \frac{\mathbf{P}\{X = x_k, Y = y_\ell\}}{\mathbf{P}\{Y = y_\ell\}} = \frac{p_{k,\ell}}{p_{\cdot,\ell}}.$$

On définit de manière similaire la **loi conditionnelle de Y par rapport à X.**

Dans de nombreuses situations, on modélisera un couple (X, Y) en commençant par la loi de Y, puis en déterminant la loi de X conditionnée par celle de Y.

Exemple 8.1 (Naissance de garçons et de filles, 1/5) On désire modéliser la situation suivante : dans une famille, le nombre N d'enfants suit une loi de Poisson de paramètre $\lambda > 0$: $N \rightsquigarrow \mathscr{P}(\lambda)$. À chaque naissance, la probabilité que l'enfant soit une fille est p, et la probabilité que ce soit un garçon est $q = 1 - p$. *On demande la loi de* X *(nombre de filles) et celle de* Y *(nombre de garçons).*

Dans un premier temps, déterminons la loi conjointe du couple (N, X). L'ensemble des valeurs prises par ce couple étant inclus dans $\mathbb{N}^2$, on choisit un couple d'entiers (n, k) et on calcule

$$\mathbf{P}\{N = n, X = k\} = \mathbf{P}\{N = n\} \cdot \mathbf{P}\{X = k \mid N = n\}.$$

La modélisation de la propriété

« à chaque naissance, la probabilité que l'enfant soit une fille est p »

se fait ainsi : la loi de X selon la probabilité $\mathbf{P}_{\{N=n\}}$, c'est-à-dire sa loi conditionnée par l'événement $\{N = n\}$, est une loi binomiale $\mathscr{B}(n, p)$: on a donc

$$\mathbf{P}\{X = k \mid N = n\} = \binom{n}{k} p^k q^{n-k}.$$

On obtient donc la loi conjointe du couple (N, X) :

$$\forall (n,k) \in \mathbb{N}^2 \qquad \mathbf{P}\big\{(N, X) = (n, k)\big\} = e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{k} p^k q^{n-k}.$$

De cette loi conjointe, on tire la loi marginale de X, moyennant quelques lignes de calcul : pour tout entier k,

$$\begin{aligned}
\mathbf{P}\{X = k\} &= \sum_{n=0}^{\infty} \mathbf{P}\big\{(N, X) = (n, k)\big\} = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{k} p^k q^{n-k} \\
&= \sum_{n=k}^{\infty} e^{-\lambda} \frac{\lambda^n}{k!\,(n-k)!} p^k q^{n-k} \qquad \textit{les } k \textit{ premiers termes sont nuls} \\
&= \frac{e^{-\lambda} \lambda^k}{k!} p^k \sum_{n=0}^{\infty} \frac{\lambda^n q^n}{n!} = \frac{e^{-\lambda} \lambda^k}{k!} p^k e^{\lambda q} = \frac{e^{-\lambda p} (p\lambda)^k}{k!},
\end{aligned}$$

ce qui montre que X suit la loi de Poisson de paramètre λp. De même, la variable aléatoire Y suit la loi de Poisson $\mathscr{P}(\lambda q)$.

Au passage, on remarquera que, si le nombre moyen d'enfants par famille est $\hat{n} = \mathbf{E}(N) = \lambda$, le nombre moyen de filles par famille est $\mathbf{E}(X) = p\lambda = p\hat{n}$; résultat en tous points conforme à l'intuition.

(À suivre... → exemple 8.4)

§ 109 Somme de deux variables aléatoires à valeurs dans $\mathbb{Z}$

Si la loi de la somme de deux variables aléatoires discrètes quelconques est en général difficile à expliciter, en revanche le cas de variables à valeurs entières est plus simple. On suppose donc que $X(\Omega)$ et $Y(\Omega)$ sont inclus dans $\mathbb{Z}$; dans ce cas, la variable aléatoire $Z = X + Y$ est également à valeurs dans $\mathbb{Z}$. De plus, pour tout $n \in \mathbb{Z}$, on peut décomposer l'événement $\{Z = n\}$ en l'union disjointe

$$\{Z = n\} = \sum_{k \in \mathbb{Z}} \{X = k, Y = n - k\}$$

et par σ-additivité, on obtient le résultat :

$$\forall n \in \mathbb{N} \qquad \mathbf{P}\{X + Y = n\} = \sum_{k=-\infty}^{+\infty} \mathbf{P}\{X = k, Y = n - k\}.$$

§ 110 Théorème de transfert

Soient X et Y deux variables aléatoires réelles discrètes, et soit $\varphi : \mathbb{R}^2 \to \mathbb{R}$ une fonction quelconque (définie au moins sur le produit $X(\Omega) \times Y(\Omega)$). Alors la fonction composée $\varphi(X, Y)$ est une variable aléatoire discrète. On en déduit, en reprenant les résultats du paragraphe § 90 :

■ **Théorème 8.2 (Formule de transfert)** *La variable aléatoire* $\varphi(X,Y)$ *admet une espérance si et seulement si la série double* $\sum \varphi(x_k, y_\ell)\, p_{k,\ell}$ *converge absolument. Si c'est le cas,* ⏢8.6

$$\mathbf{E}\big[\varphi(X,Y)\big] = \sum_{\substack{k\in K\\ \ell\in L}} \varphi(x_k, y_\ell)\, p_{k,\ell}.$$

8.2 INDÉPENDANCE

Indépendance de deux variables aléatoires discrètes § 111

On rappelle que deux événements A et B d'un espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ sont dits indépendants si $\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B)$. C'est donc, ni plus, ni moins qu'une relation numérique entre les deux. Nous dirons de même que deux variables aléatoires discrètes X et Y, prenant des valeurs $\{x_k\ ;\ k \in K\}$ et $\{y_\ell\ ;\ \ell \in L\}$, sont **indépendantes** si le fait que X prenne une valeur quelconque est indépendant du fait que Y prenne une valeur quelconque et, plus précisément, si

$$\forall (k,\ell) \in K \times L \qquad \mathbf{P}\{X = x_k, Y = y_\ell\} = \mathbf{P}\{X = x_k\} \cdot \mathbf{P}\{Y = y_\ell\}.$$

Notons $p_i = \mathbf{P}\{X = x_i\}$ et $q_j = \mathbf{P}\{Y = y_j\}$; si X et Y sont indépendantes, la loi du couple se résume dans le tableau

P	$\{Y = y_1\}$	$\{Y = y_2\}$	$\cdots$	$\{Y = y_j\}$	$\cdots$	loi de X
$\{X = x_1\}$	p_1q_1	p_1q_2	$\cdots$	p_1q_j	$\cdots$	$\boldsymbol{p_1}$
$\{X = x_2\}$	p_2q_1	p_2q_2	$\cdots$	p_2q_j	$\cdots$	$\boldsymbol{p_2}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
$\{X = x_k\}$	p_iq_1	p_iq_2		p_iq_j	$\cdots$	$\boldsymbol{p_i}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
loi de Y	$\boldsymbol{q_1}$	$\boldsymbol{q_2}$	$\cdots$	$\boldsymbol{q_j}$	$\cdots$	**1**

Cette définition permet de montrer que « tous les événements liés à X sont indépendants de tous les événements liés à Y » :

■ **Proposition 8.3** *Si* X *et* Y *sont indépendantes, alors pour toutes parties* A *et* B *de* $\mathbb{R}$,

$$\mathbf{P}\{X \in A, Y \in B\} = \mathbf{P}\{X \in A\} \cdot \mathbf{P}\{Y \in B\}.$$

DÉMONSTRATION : Notons $K' = \{k \in K\ ;\ x_k \in A\}$ et $L' = \{\ell \in L\ ;\ y_\ell \in B\}$. L'événement $\{X \in A, Y \in B\}$ s'écrit sous la forme de l'union disjointe

$$\{X \in A, Y \in B\} = \sum_{\substack{k\in K'\\ \ell\in L'}} \{X = x_k, Y = y_\ell\},$$

donc

$$\mathbf{P}\{X \in A, Y \in B\} = \sum_{\substack{k\in K'\\ \ell\in L'}} \mathbf{P}\{X = x_k, Y = y_\ell\} = \sum_{\substack{k\in K'\\ \ell\in L'}} \mathbf{P}\{X = x_k\} \cdot \mathbf{P}\{Y = y_\ell\}$$

$$= \Big(\sum_{k\in K'} \mathbf{P}\{X = x_k\}\Big) \cdot \Big(\sum_{\ell\in L'} \mathbf{P}\{Y = y_\ell\}\Big) = \mathbf{P}\{X \in A\} \cdot \mathbf{P}\{X \in B\},$$

les manipulations sur les sommes étant toutes autorisées par convergence de chacune des séries positives concernées. ∎

Exemple 8.4 (Naissance de garçons et de filles, 2/5) Reprenons l'exemple 8.1 des naissances de garçons et de filles. On demande de déterminer si les variables aléatoires X et Y sont indépendantes l'une de l'autre ou non.

Rappelons que le nombre total N d'enfants suit la loi de Poisson $\mathscr{P}(\lambda)$ et que nous avons prouvé que $X \rightsquigarrow \mathscr{P}(p\lambda)$ et $Y \rightsquigarrow \mathscr{P}(q\lambda)$.

Soit (k,l) un couple d'entiers. Au lieu de considérer la valeur prise par le couple (X,Y), on peut de manière équivalente considérer celle prise par le couple (X,N), puisque

$$\{X = k, Y = \ell\} = \{X = k, X + Y = k + \ell\} = \{X = k, N = k + \ell\}.$$

On obtient donc

$$\begin{aligned}\mathbf{P}\{X = k, Y = \ell\} &= \mathbf{P}\{X = k, N = k + \ell\} = \mathbf{P}\{N = k + \ell\} \cdot \mathbf{P}\{X = k \,|\, N = k + \ell\} \\ &= e^{-\lambda} \frac{\lambda^{k+\ell}}{(k+\ell)!} \cdot \binom{k+\ell}{k} p^k (1-p)^\ell \\ &= e^{-p\lambda} \frac{(p\lambda)^k}{k!} e^{-q\lambda} \frac{(q\lambda)^\ell}{\ell!} = \mathbf{P}\{X = k\} \cdot \mathbf{P}\{Y = \ell\}.\end{aligned}$$

Conclusion : aussi étonnant que ça puisse paraître de prime abord, les variables X et Y sont effectivement indépendantes ! Il ne faut pas y chercher tête baissée de signification intuitive : si la variable aléatoire N avait suivi une *autre* loi qu'une loi de Poisson, il n'y aurait pas nécessairement eu indépendance[1] ; en fait, il est même classique de prouver que seule la loi de Poisson mène à
↫16.1 l'indépendance de X et Y.

(À suivre... → exemple 8.30)

§ 112 Indépendance mutuelle

On généralise la notion d'indépendance de deux variables aléatoires à une famille quelconque $(X_i)_{i\in I}$ de variables aléatoires (l'ensemble I des indices pouvant être fini ou infini) : ces variables sont **(mutuellement) indépendantes** si, pour toute famille $(A_i)_{i\in I}$ d'événements, les événements $\{X_i \in A_i\}$ sont mutuellement indépendants. Rappelons que cela signifie que, pour toute sous-famille *finie* $i_1, i_2, \ldots, i_n$ de N, les événements $\{X_{i_k} \in A_{i_k}\}$ sont indépendants :

$$\mathbf{P}\left(\bigcap_{k=1}^{n} \{X_{i_k} \in A_{i_k}\}\right) = \prod_{k=1}^{n} \mathbf{P}\{X_{i_k} \in A_{i_k}\}.$$

En pratique, on utilisera la caractérisation suivante, dans le cas d'un nombre fini de variables aléatoires :

■ **Proposition 8.5** *Les variables aléatoires $X_1, X_2, \ldots, X_n$ sont mutuellement indépendantes si et seulement si, pour tout $(x_1, x_2, \ldots, x_n) \in X_1(\Omega) \times \cdots \times X_n(\Omega)$, on a*

$$\mathbf{P}\{X_1 = x_1, \ldots, X_n = x_n\} = \mathbf{P}\{X_1 = x_1\} \cdots \mathbf{P}\{X_n = x_n\}.$$

DÉMONSTRATION : Si $X_1, X_2, \ldots, X_n$ sont indépendantes, il suffit de poser $A_k := \{x_k\}$ pour $k = 1, \ldots, n$ et d'appliquer la définition pour obtenir la formule.

Réciproquement, supposons que cette formule soit vraie pour tout choix de n-uplet $(x_1, x_2, \ldots, x_n)$, et fixons des parties $A_1, A_2, \ldots, A_n$ de $\mathbb{R}$. Pour chaque indice i de

---

(1). Prenons, pour loi de N, la loi constante $N = 2$, alors X et Y sont reliées par la relation $X + Y = 2$, ce qui leur interdit d'être indépendantes (si l'on sait qu'il y a deux filles, on est sûrs qu'il n'y a pas de garçon). Une preuve formelle consiste bien sûr en un contre-exemple numérique, par exemple

$$\mathbf{P}\{X = 0, Y = 0\} = 0 \qquad \text{mais} \qquad \mathbf{P}\{X = 0\} = q^2 > 0 \quad \text{et} \quad \mathbf{P}\{Y = 0\} = p^2 > 0.$$

$\{1, 2, \ldots, n\}$, décomposons la partie de A_i atteignable par la variable X_i sous la forme $A_i \cap X_i(\Omega) = \{x_{i,k} \ ; \ k \in K_i'\}$. Il suffit alors de remarquer que

$$\begin{aligned}
\mathbf{P}\{X_1 \in A_1, \ldots, X_n \in A_n\} &= \mathbf{P}\Big(\sum_{k_1} \cdots \sum_{k_n} \{X_1 = x_{1,k_1}, \ldots, X_n = x_{n,k_n}\} \Big) \\
&= \sum_{k_1} \cdots \sum_{k_n} \mathbf{P}\big(\{X_1 = x_{1,k_1}, \ldots, X_n = x_{n,k_n}\} \big) \\
&= \sum_{k_1} \cdots \sum_{k_n} \mathbf{P}\{X_1 = x_{1,k_1}\} \cdots \mathbf{P}\{X_n = x_{n,k_n}\} \\
&= \prod_{i=1}^{n} \sum_{k_i} \mathbf{P}\{X_i = x_{i,k_i}\} = \prod_{i=1}^{n} \mathbf{P}\{X_i \in A_i\}.
\end{aligned}$$

∎

Fonctions de variables indépendantes § 113

L'indépendance (mutuelle) de variables aléatoires permet de créer de nouvelles variables indépendantes. Si $X_1, X_2, \ldots, X_n$ sont des variables aléatoires discrètes et si $\varphi : \mathbb{R}^n \to \mathbb{R}$ est une application, convenons de noter $\varphi(X_1, X_2, \ldots, X_n)$ la variable aléatoire discrète[2]

$$\omega \longmapsto \varphi\big(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)\big).$$

■ **Théorème 8.6 (Fonctions de variables indépendantes)**

i) Soient X et Y deux variables aléatoires discrètes indépendantes. Soient φ et ψ deux applications de $\mathbb{R}$ dans $\mathbb{R}$. Alors les variables aléatoires $\varphi(X)$ et $\psi(Y)$ sont indépendantes.

ii) Plus généralement, si $X_1, X_2, \ldots, X_n$ sont des variables aléatoires indépendantes, si $\varphi_1, \varphi_2, \ldots, \varphi_n$ sont des applications de $\mathbb{R}$ dans $\mathbb{R}$, alors les variables $\varphi_1(X_1), \ldots, \varphi_n(X_n)$ sont indépendantes.

DÉMONSTRATION : Nous donnons ici une démonstration générale, ne faisant pas intervenir le caractère discret des variables aléatoires.

Traitons le cas de deux variables aléatoires X et Y indépendantes. Soient A et B deux événements de $\mathbb{R}$; alors

$$\{\varphi(X) \in A\} = \{X \in \varphi^{-1}(A)\} \qquad \text{et} \qquad \{\psi(Y) \in B\} = \{Y \in \psi^{-1}(B)\}.$$

Par indépendance de X et Y, on a donc

$$\begin{aligned}
\mathbf{P}\{\varphi(X) \in A, \psi(Y) \in B\} &= \mathbf{P}\{X \in \varphi^{-1}(A), Y \in \psi^{-1}(B)\} \\
&= \mathbf{P}\{X \in \varphi^{-1}(A)\} \cdot \mathbf{P}\{Y \in \psi^{-1}(B)\} \\
&= \mathbf{P}\{\varphi(X) \in A\} \cdot \mathbf{P}\{\psi(Y) \in B\}.
\end{aligned}$$

Ceci étant vrai pour tout choix de A et B, les variables aléatoires $\varphi(X)$ et $\psi(Y)$ sont indépendantes.

La démonstration dans le cas général est identique. ∎

Le procédé précédent se généralise encore :

■ **Théorème 8.7 (Lemme des coalitions)** *Si $X_1, X_2, \ldots, X_n$ sont des variables aléatoires indépendantes, si k est un entier compris entre 1 et n, et si $\varphi : \mathbb{R}^k \to \mathbb{R}$ et $\psi : \mathbb{R}^{n-k} \to \mathbb{R}$ sont des applications, alors $Y = \varphi(X_1, \ldots, X_k)$ et $Z = \psi(X_{k+1}, \ldots, X_n)$ sont indépendantes.*

(2). Elle est bien discrète car son image est l'image d'un ensemble discret par une application.

DÉMONSTRATION : Alors que le théorème précédent est une conséquence immédiate de la définition de l'indépendance de variables aléatoires, celui-ci se révèle étonnamment technique dans le cas général, et sa démonstration en est reportée à l'annexe B (lemme B.17 page 705).

Dans le cas particulier qui nous intéresse (les variables sont toutes discrètes), on peut toutefois s'en sortir avec des arguments élémentaires. Tout d'abord, d'après la proposition 8.5, il suffit de prouver que, si a et b sont des réels, alors $\mathbf{P}\{Y = a, Z = b\} = \mathbf{P}\{Y = a\}\,\mathbf{P}\{Z = b\}$. Remarquons que

$$\{Y = a\} = \{\varphi(X_1, \dots, X_k) = a\} = \Big\{(X_1, \dots, X_k) \in \varphi^{-1}\big(\{a\}\big)\Big\}.$$

L'ensemble $\varphi^{-1}\big(\{a\}\big)$ peut être fini ou infini, dénombrable ou indénombrable, ce qui empêche *a priori* de manipuler l'expression précédente. Cependant, on peut également écrire

$$\{Y = a\} = \Big\{(X_1, \dots, X_k) \in \varphi^{-1}\big(\{a\}\big) \cap \mathrm{Im}(X_1, \dots, X_k)\Big\},$$

et l'ensemble

$$A := \varphi^{-1}\big(\{a\}\big) \cap \mathrm{Im}(X_1, \dots, X_k)$$

a le bon goût d'être dénombrable, ce qui permet d'écrire $\{Y = a\}$ comme l'union dénombrable suivante :

$$\{Y = a\} = \bigcup_{(x_1,\dots,x_k)\in A} \{X_1 = x_1, \dots, X_k = x_k\}.$$

De même, on pose

$$B := \psi^{-1}\big(\{b\}\big) \cap \mathrm{Im}(X_{k+1}, \dots, X_n)$$

et on écrit

$$\{Z = b\} = \bigcup_{(x_{k+1},\dots,x_n)\in B} \{X_{k+1} = x_{k+1}, \dots, X_n = x_n\}.$$

Il ne reste plus qu'à écrire

$$\{Y = a, Z = b\} = \bigcup_{\substack{(x_1,\dots,x_k)\in A\\(x_{k+1},\dots,x_n)\in B}} \{X_1 = x_1, \dots, X_k = x_k, X_{k+1} = x_{k+1}, \dots, X_n = x_n\}$$

et, par σ-additivité et indépendance des variables,

$$\begin{aligned}
\mathbf{P}\{Y = a, Z = b\} &= \sum_{\substack{(x_1,\dots,x_k)\in A\\(x_{k+1},\dots,x_n)\in B}} \mathbf{P}\{X_1 = x_1, \dots, X_k = x_k, X_{k+1} = x_{k+1}, \dots, X_n = x_n\} \\
&= \sum_{\substack{(x_1,\dots,x_k)\in A\\(x_{k+1},\dots,x_n)\in B}} \mathbf{P}\{X_1 = x_1\} \cdots \mathbf{P}\{X_k = x_k\} \times \\
&\qquad\qquad \times \mathbf{P}\{X_{k+1} = x_{k+1}\} \cdots \mathbf{P}\{X_n = x_n\} \\
&= \Big(\sum_{(x_1,\dots,x_k)\in A} \mathbf{P}\{X_1 = x_1\} \cdots \mathbf{P}\{X_k = x_k\}\Big) \cdot \\
&\qquad\qquad \Big(\sum_{(x_{k+1},\dots,x_n)\in B} \mathbf{P}\{X_{k+1} = x_{k+1}\} \cdots \mathbf{P}\{X_n = x_n\}\Big) \\
&= \mathbf{P}\{Y = a\} \cdot \mathbf{P}\{Z = b\}.
\end{aligned}$$

∎

§ 114 **Un exemple : minimum et maximum de n variables aléatoires indépendantes**

Soient $X_1, X_2, \dots, X_n$ des variables aléatoires mutuellement indépendantes, de fonctions de répartition respectives $F_1, F_2, \dots, F_n$. On note

$$Y := \min(X_1, X_2, \dots, X_n) \qquad \text{et} \qquad Z := \max(X_1, X_2, \dots, X_n),$$

et on cherche la fonction de répartition de Y et de Z.

Soit x un réel. L'événement $\{Z \leqslant x\}$ est égal à l'événement

$$\{X_1 \leqslant x, X_2 \leqslant x, \dots, X_n \leqslant x\},$$

lequel est de probabilité $F_1(x)\,F_2(x)\cdots F_n(x)$ par indépendance des variables aléatoires. Ainsi,

$$F_Z = F_1 \cdot F_2 \cdots F_n.$$

En revanche, l'événement $\{Y \leqslant x\}$ est égal à l'union $\{X_1 \leqslant x\} \cup \{X_2 \leqslant x\} \cup \cdots \cup \{X_n \leqslant x\}$, ce qui ne permet pas d'obtenir une formule simple, les événements considérés n'étant pas disjoints. On considère alors plutôt l'événement contraire $\{Y > x\}$, égal à

$$\{X_1 > x, X_2 > x, \ldots, X_n > x\},$$

de probabilité $\big(1 - F_1(x)\big)\cdots\big(1 - F_n(x)\big)$. Ainsi, ⇙8.10

$$1 - F_Y = (1 - F_1)(1 - F_2)\cdots(1 - F_n).$$

Espérance d'un produit de variables aléatoires indépendantes §115

L'espérance est linéaire et $\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y)$ sans qu'il soit besoin d'hypothèse spéciale sur X et Y ; en revanche, la relation « $\mathbf{E}(XY) = \mathbf{E}(X)\,\mathbf{E}(Y)$ » est généralement fausse. Il est remarquable que ce soit toutefois vrai lorsque X et Y sont indépendantes. Nous donnons ici une première version de ce résultat général, dans le cadre des variables aléatoires discrètes. Une seconde démonstration sera donnée pour les variables à densité, et une troisième dans le cas général.

■ **Théorème 8.8** *Soient* X *et* Y *deux variables aléatoires réelles discrètes, indépendantes, admettant chacune une espérance. Alors le produit* XY *admet une espérance et*

$$\mathbf{E}(XY) = \mathbf{E}(X)\,\mathbf{E}(Y).$$

DÉMONSTRATION : Notons $X(\Omega) = \{x_k\ ;\ k \in K\}$ et $Y(\Omega) = \{y_\ell\ ;\ \ell \in L\}$, et posons

$$q_k = \mathbf{P}\{X = x_k\} \qquad r_\ell = \mathbf{P}\{Y = y_\ell\} \qquad \text{et} \qquad p_{k,\ell} = \mathbf{P}\{X = x_k, Y = y_\ell\} = q_k\, r_\ell,$$

cette dernière relation étant due à l'indépendance des variables aléatoires X et Y. Tout d'abord, nous voulons montrer que la série double $\sum_{k,\ell} x_k\, y_\ell\, p_{k,\ell}$ converge absolument. Pour cela, on fixe momentanément un indice k de K, et on considère la somme suivante[3] :

$$\sum_{\ell\in L} p_{k,\ell}\,|x_k\, y_\ell| = \sum_{\ell\in L} q_k\, r_\ell\,|x_k\, y_\ell| = q_k\,|x_k| \sum_{\ell\in L} |y_\ell|\, r_\ell = q_k\,|x_k|\,\mathbf{E}\big(|Y|\big) < +\infty.$$

De plus, la série $\sum_k q_k\,|x_k|\,\mathbf{E}\big(|Y|\big)$ converge, puisque X admet une espérance. Ainsi, d'après le théorème de Fubini A.12 page 673, la famille $\big\{x_k\, y_\ell\, p_{k,\ell}\ ;\ (k,\ell) \in K \times L\big\}$ est sommable (la série converge absolument), donc XY admet une espérance. Enfin, le même théorème de Fubini nous autorise à sommer dans l'ordre voulu :

$$\begin{aligned}\mathbf{E}(XY) &= \sum_{k\in K}\sum_{\ell\in L} p_{k,\ell}\, x_k\, y_\ell = \sum_{k\in K}\sum_{\ell\in L} x_k\, q_k\, y_\ell\, r_\ell \\ &= \Big(\sum_{k\in K} x_k\, q_k\Big)\Big(\sum_{\ell\in L} y_\ell\, r_\ell\Big) = \mathbf{E}(X)\cdot\mathbf{E}(Y).\end{aligned}$$

∎

(3). Correctement définie dans $\mathbb{R} \cup \{+\infty\}$, puisque composée uniquement de termes positifs.

§ 116 **Somme de variables aléatoires discrètes indépendantes — formule de convolution**
La formule générale donnant la loi d'une somme de deux variables aléatoires à valeurs entières prend une tournure plus intéressante dans le cas de variables indépendantes.

■ **Théorème 8.9** *Soient* X *et* Y *deux variables aléatoires à valeurs dans* $\mathbb{Z}$, *et indépendantes. Alors, pour tout* $n \in \mathbb{N}$,

$$\mathbf{P}\{X+Y=n\} = \sum_{k=-\infty}^{+\infty} \mathbf{P}\{X=k\} \cdot \mathbf{P}\{Y=n-k\}.$$

En particulier, si X *et* Y *sont à valeurs dans* $\mathbb{N}$,

$$\mathbf{P}\{X+Y=n\} = \sum_{k=0}^{n} \mathbf{P}\{X=k\} \cdot \mathbf{P}\{Y=n-k\}.$$

Cette dernière formule rappelle évidemment un produit de Cauchy ! On l'appelle **formule de convolution** ; si l'on note $p_n = \mathbf{P}\{X=n\}$, $q_n = \mathbf{P}\{Y=n\}$ et $r_n = \mathbf{P}\{X+Y=n\}$, elle s'écrit

$$r_n = \sum_{k\in\mathbb{Z}} p_k\, q_{n-k}$$

ou, symboliquement,

$$r = p * q.$$

Des propriétés générales du produit de deux séries entières (A41, page 684), on déduit donc le résultat fondamental suivant :

■ **Théorème 8.10 (Fonction génératrice d'une somme de v.a. indépendantes)** *Soient* X *et* Y *deux variables aléatoires à valeurs dans* $\mathbb{N}$, *et indépendantes. Alors la fonction génératrice de* $X+Y$ *est*

$$G_{X+Y} = G_X \cdot G_Y.$$

Remarque 8.11 On peut également démontrer cette formule de la manière suivante. On fixe $t \in]-1\,;+1[$, alors les variables aléatoires t^X et t^Y sont indépendantes, l'espérance de leur produit est égal au produit de leurs espérances :

$$G_{X+Y}(t) = \mathbf{E}(t^{X+Y}) = \mathbf{E}(t^X \cdot t^Y) = \mathbf{E}(t^X) \cdot \mathbf{E}(t^Y) = G_X(t) \cdot G_Y(t).$$

§ 117 **Loi binomiale comme somme de lois de Bernoulli**
Nous avions introduit la loi binomiale pour modéliser le nombre de succès dans une suite de tirages de Bernoulli indépendants (avant la formalisation de ce concept). Nous pouvons maintenant ériger ce résultat en théorème :

■ **Théorème 8.12** *Soit* p *un réel,* $0 < p < 1$. *Soit* n *un entier naturel, et soient* $X_1, X_2, \ldots, X_n$ *des variables aléatoires indépendantes de même loi de Bernoulli* $\mathscr{B}(p)$. *Alors la variable aléatoire* $S_n = X_1 + X_2 + \cdots + X_n$ *suit la loi binomiale* $\mathscr{B}(n\,;\,p)$.

DÉMONSTRATION : On peut procéder par récurrence sur n, en utilisant directement la formule du théorème 8.9 et la formule du triangle de Pascal. Le cas $n = 1$ est immédiat car $\mathscr{B}(1\,;\,p) = \mathscr{B}(p)$. Soit $n \geqslant 2$ un entier et supposons que S_{n-1} suive la loi binomiale

$\mathscr{B}(n-1\,;\,p)$. D'après le lemme des coalitions (8.7), $S_{n-1} = X_1 + X_2 + \cdots + X_{n-1}$ et X_n sont indépendantes, donc pour tout $k \in [\![1\,;\,n]\!]$,

$$\begin{aligned}\mathbf{P}\{S_n = k\} &= \mathbf{P}\{X = 0\}\,\mathbf{P}\{S_{n-1} = k\} + \mathbf{P}\{X = 1\}\,\mathbf{P}\{S_{n-1} = k-1\}\\ &= (1-p)\binom{n-1}{k} p^k (1-p)^{n-1-k} + p\binom{n-1}{k-1} p^{k-1}(1-p)^{n-k}\\ &= \left[\binom{n-1}{k} + \binom{n-1}{k-1}\right] p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k},\end{aligned}$$

le cas $k = 0$ pouvant être traité à part et menant au même résultat. ∎

Théorèmes de stabilité §118

La loi binomiale et la loi de Poisson partagent la propriété d'être stable par addition, sous réserve d'indépendance des variables :

■ **Théorème 8.13 (Stabilité de quelques lois)** *Soient* X *et* Y *deux variables aléatoires discrètes indépendantes.*

i) Si $X \rightsquigarrow \mathscr{B}(n\,;\,p)$ *et* $Y \rightsquigarrow \mathscr{B}(m\,;\,p)$, *alors* $X + Y \rightsquigarrow \mathscr{B}(n+m\,;\,p)$.

ii) Si $X \rightsquigarrow \mathscr{P}(\lambda)$ *et* $Y \rightsquigarrow \mathscr{P}(\mu)$, *alors* $X + Y \rightsquigarrow \mathscr{P}(\lambda+\mu)$.

DÉMONSTRATION : *i)* On peut modéliser X et Y comme des sommes de n et m variables de Bernoulli indépendantes, respectivement. Alors $X+Y$ est bien somme de $n+m$ variables aléatoires de Bernoulli de paramètre p, indépendantes.

Alternativement, on utilise la formule de convolution conjointement avec la **formule de Vandermonde** :

$$\sum_{k=0}^{n} \binom{n}{k}\binom{m}{p-k} = \binom{n+m}{p}.$$

Enfin, on peut utiliser les fonctions génératrices. En effet,

$$G_X(t) = (1-p+pt)^n \qquad G_Y(t) = (1-p+pt)^m$$

donc $G_{X+Y}(t) = (1-p+pt)^{m+n}$; on reconnaît la fonction génératrice de la loi binomiale $\mathscr{B}(n+m\,;\,p)$.

ii) Les fonctions génératrices de X et Y sont $G_X(t) = e^{\lambda(t-1)}$ et $G_Y(t) = e^{\mu(t-1)}$; la fonction génératrice de $X+Y$ est donc

$$G_{X+Y}(t) = e^{\lambda(t-1)} \cdot e^{\mu(t-1)} = e^{(\lambda+\mu)(t-1)},$$

caractéristique de la loi de Poisson $\mathscr{P}(\lambda+\mu)$.

∎

8.3 COVARIANCE

Covariance d'un couple discret §119

Lorsque X et Y sont des variables indépendantes, nous avons vu que $\mathbf{E}(XY) = \mathbf{E}(X)\,\mathbf{E}(Y)$; dans le cas général, cette relation n'est plus vraie. La *covariance* du couple (X, Y) est la mesure du défaut d'égalité :

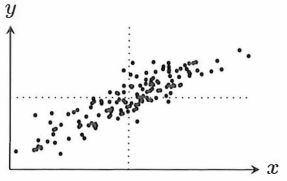
Définition 8.14 Soient X et Y deux variables aléatoires discrètes admettant une espérance, et telles que XY admet une espérance ; on dit alors que le couple (X, Y) **admet une covariance**. On note $X' := X - \mathbf{E}(X)$ et $Y' := Y - \mathbf{E}(Y)$ les variables centrées correspondantes. La **covariance** du couple (X, Y) est l'espérance du produit $X'Y'$:

$$\mathbf{Cov}(X,Y) = \mathbf{E}(X'Y') = \mathbf{E}\big[(X - \mathbf{E}X)(Y - \mathbf{E}Y)\big].$$

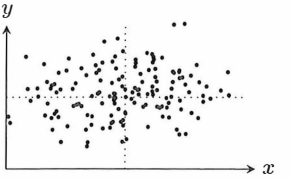
En développant, et par linéarité, on a également

$$\mathbf{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y).$$

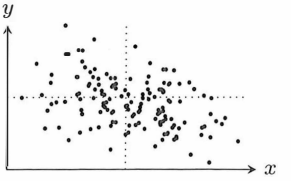
Exemple 8.15 (Interprétation du signe de la covariance)
Considérons une série de données (x_k, y_k), $k = 1, \ldots, n$; nous pouvons les modéliser comme les valeurs prises par un couple de variables aléatoires discrètes ; à chacune des valeurs prises, nous associons la probabilité uniforme $1/n$. Nous en représentons trois jeux différents, censés provenir de trois séries de mesures.



$\rho = 0,9$



$\rho = 0,1$



$\rho = -0,4$

(ρ est le *coefficient de corrélation*, proportionnel à la covariance — voir § 121.)

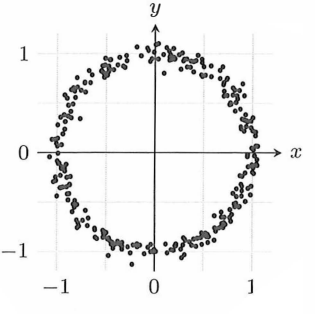
Sur chacun des graphes, nous avons représenté en pointillé des droites passant par l'abscisse $m_x = \mathbf{E}(X)$ et l'ordonnée $m_y = \mathbf{E}(Y)$.

Sur le premier graphe, de très nombreux points vérifient la double condition $x_k - m_x > 0$ et $y_k - m_y > 0$: ils sont dans le quart nord-est (NE) de l'image. De même, de très nombreux points vérifient $x_k - m_x < 0$ et $y_k - m_y < 0$: ils sont dans le quart sud-ouest (SO). Tous ces points (ici, 121 sur 150) contribuent donc pour une part *positive* au calcul de l'espérance de $(X - \mathbf{E}X)(Y - \mathbf{E}Y)$. Les 19 points restants contribuent négativement. Au total, la covariance de (X, Y) est fortement positive.

Sur le second graphe, la répartition des points dans chaque quart de plan est à peu près équitable et les diverses contributions se compensent peu ou prou. La covariance du couple est proche de zéro.

Enfin, sur le dernier graphe, il y a un peu plus de points dans les quarts NO et SE que dans les deux autres : la covariance est négative, mais plutôt petite.

Ainsi se dégage une interprétation générale (mais approximative) de la covariance : elle traduit la plus ou moins grande connaissance que l'on peut déduire de Y, connaissant la valeur de X. Dans le premier graphe, si l'on sait que x_k est grand, on en déduit que, probablement, y_k l'est également. Sur le second, au contraire, on ne peut pas déduire grand-chose sur y_k si l'on connaît la valeur de x_k. Ce point de vue est cependant trop simpliste : dans l'exemple ci-dessous, où à chaque point est associé une probabilité $1/n$, la covariance de (X, Y) est proche de 0 ; pourtant, la connaissance de X donne un renseignement fort sur Y, puisque $X^2 + Y^2$ est toujours « proche de 1 ».



$\rho \approx 0,007$

On dit que les variables aléatoires X et Y sont **décorrélées** si leur covariance est nulle. La première propriété à remarquer est fondamentale : *l'indépendance implique la décorrélation.*

■ **Théorème 8.16** *Si* X *et* Y *sont indépendantes, alors* $\mathbf{Cov}(X, Y) = 0$.

DÉMONSTRATION : Le théorème 8.8 montre que $\mathbf{E}(XY) = \mathbf{E}(X)\,\mathbf{E}(Y)$, ce qui est la définition ⇙8.1 de la nullité de la covariance du couple. ∎

La réciproque est *fausse* : la covariance d'un couple peut tout à fait être nulle sans que les marginales soient indépendantes.

Exemple 8.17 (Un exemple minimal) Considérons une variable aléatoire X de loi discrète uniforme sur l'ensemble $\{-1, 0, 1\}$, et définissons $Y := X^2$. La loi du couple (X, Y) et les lois marginales sont résumées dans le tableau ci-dessous :

P	$\{X=-1\}$	$\{X=0\}$	$\{X=1\}$	marginale de Y
$\{Y=0\}$	0	1/3	0	**1/3**
$\{Y=1\}$	1/3	0	1/3	**2/3**
marginale de X	**1/3**	**1/3**	**1/3**	**1**

Il est immédiat de vérifier que $\mathbf{E}(XY) = \mathbf{E}(X) = 0$, et que par conséquent $\mathbf{Cov}(X, Y) = 0$: les variables aléatoires sont décorrélées. En revanche, elles ne sont pas indépendantes, car il existe une « relation entre X et Y », qui permet par exemple d'écrire que

$$0 = \mathbf{P}\{X=0, Y=1\} \quad \neq \quad \mathbf{P}\{X=0\} \cdot \mathbf{P}\{Y=1\} = \tfrac{2}{9}.$$

■ **Théorème 8.18 (Propriétés de la covariance)** *Soient* X, Y, Z *et* T *des variables aléatoires admettant des espérances, et telles que les covariances ci-dessous sont toutes définies. Alors*

i) $\mathbf{Cov}(X, Y) = \mathbf{Cov}(Y, X)$;

ii) $\mathbf{Cov}(\alpha X + \beta, \gamma Y + \delta) = \alpha\gamma\ \mathbf{Cov}(X, Y)$;

iii) $\mathbf{Cov}(X+Y, Z+T) = \mathbf{Cov}(X, Z) + \mathbf{Cov}(X, T) + \mathbf{Cov}(Y, Z) + \mathbf{Cov}(Y, T)$;

iv) $\mathbf{V}(X) = \mathbf{Cov}(X, X)$;

v) $\left|\mathbf{Cov}(X, Y)\right| \leqslant \sigma_X \cdot \sigma_Y$.

DÉMONSTRATION : Toutes les propriétés sont immédiates, à l'exception peut-être de la dernière, qui provient de l'inégalité de Cauchy-Schwarz (§ 96) :

$$\left|\mathbf{Cov}(X, Y)\right|^2 = \left|\mathbf{E}(X'Y')\right|^2 \leqslant \mathbf{E}(X'^2)\,\mathbf{E}(Y'^2) = \mathbf{V}(X)\ \mathbf{V}(Y),$$

ce qui montre la formule voulue. ∎

Somme de variables aléatoires § 120

Le théorème suivant, complètement élémentaire, illustre le lien entre la covariance, concept probabiliste, et la géométrie des espaces euclidiens ou préhilbertiens.

■ **Théorème 8.19 (Variance d'une somme)** *Soient* X *et* Y *deux variables aléatoires admettant une variance ; alors* XY *admet une espérance,* X+Y *et* X−Y *admettent une variance et*

$$\mathbf{V}(X+Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\,\mathbf{Cov}(X, Y)$$
$$\mathbf{V}(X-Y) = \mathbf{V}(X) + \mathbf{V}(Y) - 2\,\mathbf{Cov}(X, Y).$$

De même, si $a, b \in \mathbb{R}$, *alors*

$$\mathbf{V}(aX + bY) = a^2\ \mathbf{V}(X) + b^2\ \mathbf{V}(Y) + 2ab\ \mathbf{Cov}(X, Y).$$

Plus généralement, si $X_1, X_2, \ldots, X_n$ *sont des variables aléatoires admettant des variances et des covariances. Alors*

$$\mathbf{V}(X_1 + \cdots + X_n) = \sum_{k=1}^{n} \mathbf{V}(X_k) + 2 \sum_{1 \leqslant i < j \leqslant n} \mathbf{Cov}(X_i, X_j)$$

DÉMONSTRATION : L'existence de $\mathbf{E}(XY)$ a été démontrée dans le théorème de l'inégalité de Cauchy-Schwarz (§ 96). Décomposons X et Y sur un système complet commun[4] :

$$X = \sum_{k \in K} x_k\, \mathbf{1}_{A_k} \qquad \text{et} \qquad Y = \sum_{k \in K} y_k\, \mathbf{1}_{A_k}.$$

(4). La technique a été mise en place dans la démonstration de l'inégalité de Cauchy-Schwarz 7.28 page 156.

On sait alors que $\sum x_k^2 p_k$ et $\sum y_k^2 p_k$ convergent. Le développement

$$(x_k \pm y_k)^2 = x_k^2 + y_k^2 \pm 2x_k y_k$$

et l'inégalité $|2x_k y_k| \leqslant x_k^2 + y_k^2$ permettent de conclure à la convergence de la série $\sum (x_k \pm y_k)^2 p_k$ et à l'existence du moment d'ordre 2 de $(X \pm Y)$. Bien, maintenant que l'on est assuré que tout le monde existe, on peut écrire

$$\mathbf{V}(X \pm Y) = \mathbf{E}\big((X' \pm Y')^2\big) = \mathbf{E}(X'^2 + Y'^2 \pm 2X'Y') = \mathbf{V}(X) + \mathbf{V}(Y) \pm 2\,\mathbf{Cov}(X, Y).$$

Enfin, en remarquant que $(aX)' = aX'$ et $(bX)' = bX'$, la deuxième formule s'ensuit, et se généralise immédiatement. ∎

Remarque 8.20 La covariance admet des propriétés analogues à celles d'un produit scalaire. L'écart-type fait office de norme (pour la variable centrée), ou plutôt de semi-norme : le fait que $\sigma_X = 0$ n'implique pas que $X' = X - \mathbf{E}X$ soit nulle, mais qu'elle soit presque sûrement nulle.

En conséquence du théorème précédent, on obtient la formule d'addition des variances des variables indépendantes !

■ **Corollaire 8.21 (Égalité de Bienaymé)** *Si* X *et* Y *sont deux variables aléatoires indépendantes et admettant un moment d'ordre* 2, *alors* $X + Y$ *admet un moment d'ordre* 2 *et*

$$\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y)$$

ou encore

$$\sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}.$$

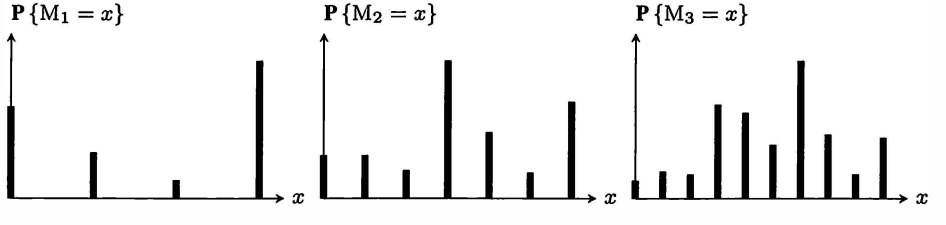
Plus généralement, si $X_1, X_2, \ldots, X_n$ *sont des variables aléatoires deux à deux indépendantes admettant un moment d'ordre* 2, *alors*

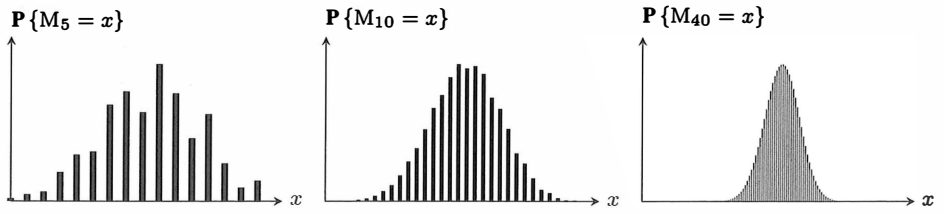
$$\mathbf{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbf{V}(X_i).$$

Remarque 8.22 Ce résultat sur les sommes de variables aléatoires deux à deux indépendantes, pour simple qu'il soit, est absolument fondamental et se trouve à l'origine des lois d'échelle du théorème central limite, du mouvement brownien, de la diffusion de la chaleur et de bien d'autres phénomènes !

Exemple 8.23 Soit X une variable aléatoire admettant une espérance, et $(X_n)_{n \geqslant 0}$ une suite de variables aléatoires indépendantes, de même loi que X. On pose $M_n = (X_1 + X_2 + \cdots + X_n)/n$. Alors l'espérance de M_n est $\mathbf{E}(M_n) = \mathbf{E}(X)$, sa variance $\mathbf{V}(M_n) = \mathbf{V}(X)/n$, et son écart-type $\sigma_{M_n} = \sigma_X/\sqrt{n}$. Cela signifie que M_n prend la même valeur moyenne que X, mais avec un écart moyen $\sigma_X/\sqrt{n}$ qui devient très petit quand n augmente. Ce phénomène est à l'origine des méthodes de Monte-Carlo pour l'évaluation de l'espérance d'une grandeur aléatoire ; il sera précisé d'un point de vue mathématique dans les théorèmes connus sous le nom de *loi (faible ou forte) des grands nombres*, ainsi que le *théorème limite central*.

Sur le diagramme ci-dessous, nous représentons l'histogramme d'une variable aléatoire X, ne prenant que les valeurs 0, 1, 2 et 3, ainsi que les histogrammes de quelques unes de ses moyennes M_n, obtenus par la formule de convolution (les échelles verticales ne sont pas respectées, mais les échelle horizontales le sont).





Le phénomène de réduction de la variance est bien visible ; on voit également les irrégularités de la distribution initiale se gommer peu à peu, le dernier histogramme interpolant, visuellement, une fonction gaussienne avec une grande précision (version locale du théorème limite central, voir chapitre 21).

Coefficient de corrélation

§ 121

Si (X, Y) est un couple de variables aléatoires discrètes admettant une covariance, et tel que $\mathbf{V}(X) > 0$ et $\mathbf{V}(Y) > 0$, on appelle **coefficient de corrélation** du couple le réel

$$\rho(X, Y) = \frac{\mathbf{Cov}(X, Y)}{\sigma_X \, \sigma_Y}.$$

D'après le théorème 8.18, ce réel est compris entre -1 et $+1$. Les cas extrêmes ne se produisent que si X et Y sont presque sûrement affinement liées, c'est-à-dire s'il existe des scalaires a et b et une partie $N \subset \Omega$ négligeable telle que

$$\forall \omega \in \Omega \setminus N \qquad X(\omega) = a\,Y(\omega) + b.$$

Plus précisément :

■ **Théorème 8.24** *Soit* (X, Y) *un couple de variables aléatoires admettant un coefficient de corrélation* $\rho = \rho(X, Y)$. *Alors* $-1 \leqslant \rho \leqslant 1$. *De plus,* $\rho = \pm 1$ *si et seulement si* $X' := X - \mathbf{E}(X)$ *et* $Y' := Y - \mathbf{E}(Y)$ *sont proportionnelles, et plus exactement si et seulement si*

$$\frac{Y - \mathbf{E}(Y)}{\sigma_Y} = \rho \, \frac{X - \mathbf{E}(X)}{\sigma_X}$$

presque sûrement. Si l'on note X^* *et* Y^* *les variables aléatoires centrées et réduites, on a donc* $\mathbf{P}\{X^* = Y^*\} = 1$ *si* $\rho = 1$, *et* $\mathbf{P}\{X^* = -Y^*\} = 1$ *si* $\rho = -1$.

DÉMONSTRATION : L'inégalité $-1 \leqslant \rho \leqslant 1$ est conséquence de l'inégalité de Cauchy-Schwarz. Si $|\rho| = 1$, on est dans le cas d'égalité $\mathbf{E}(X'Y') = \mathbf{E}(X')\,\mathbf{E}(Y')$, qui entraîne l'existence d'un scalaire λ tel que $Y' = \lambda X'$ presque sûrement. En évaluant la covariance du couple (X, Y), on obtient

$$\rho = \rho(X, Y) = \frac{\mathbf{Cov}(X, Y)}{\sigma_X \, \sigma_Y} = \frac{\mathbf{Cov}(X', Y')}{\sigma_X \, \sigma_Y} = \frac{\mathbf{Cov}(X', \lambda X')}{\sigma_X \, \sigma_Y} = \frac{\lambda \sigma_X^2}{\sigma_X \, \sigma_Y} = \lambda \, \frac{\sigma_X}{\sigma_Y},$$

œ qui entraîne le résultat annoncé. ∎

Remarque 8.25 Un coefficient de corrélation positif indique que Y est, globalement, en fonction croissante de X. Un coefficient négatif indique qu'il est, globalement, en fonction décroissante de X. Une valeur proche de ± 1 est l'indice d'un lien numérique entre les quantités X et Y.

Invariance par changement d'échelle de $\rho(X, Y)$

§ 122

Les *changements d'échelle* des variables aléatoires X et Y ne changent pas le coefficient de corrélation : si X est une longueur, on peut la mesurer en mètres, en pouces ou en furlongs (les conversions étant des transformations linéaires), si Y est une température, on peut la mesurer en Kelvin, en degrés Celsius ou en degrés

Fahrenheit (les conversions étant des transformations affines) — la valeur numérique du coefficient de corrélation restera inchangée. En effet, le théorème 8.18 implique que, si α, β, γ et δ sont des réels, avec $\alpha, \beta \neq 0$, alors

$$\rho(\alpha X + \gamma, \beta Y + \delta) = \frac{\mathbf{Cov}(\alpha X + \gamma, \beta Y + \delta)}{\sigma_{\alpha X+\gamma}\, \sigma_{\beta Y+\delta}} = \frac{\alpha\beta\, \mathbf{Cov}(X, Y)}{|\alpha|\, |\beta|\, \sigma_X\, \sigma_Y} = \operatorname{sgn}(\alpha\beta)\, \rho(X, Y),$$

où $\operatorname{sgn}(\alpha\beta)$ vaut $+1$ si $\alpha\beta > 0$ et -1 si $\alpha\beta < 0$. Les changement d'unités physiques se faisant toujours avec des coefficients de dilatation strictement positifs, la propriété est démontrée.

§ 123 **Covariance d'un couple de données statistiques**

À une série de données statistiques de la forme d'une liste de couples (x_k, y_k), $k = 1, \ldots, n$, on peut associer un couple (X, Y) de variables aléatoires, prenant chaque couple de valeurs avec la probabilité $p := 1/n$ (ou $mp = m/n$ pour une valeur du couple présente p fois dans la liste). Le **coefficient de corrélation** de cette liste de données est défini comme le coefficient de corrélation de (X, Y), on l'écrit sous forme symbolique :

$$\rho = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2} \cdot \sqrt{n\Sigma y^2 - (\Sigma y)^2}},$$

où l'on a noté brièvement « Σx » pour $\sum_{k=1}^{n} x_k$, « Σxy » pour $\sum_{k=1}^{n} x_k\, y_k$, etc.

La détermination de la **droite de régression linéaire** permettant d'approcher, au mieux, ces données, selon un modèle affine ($y = ax + b$), est exposée au chapitre 22, section 22.3.

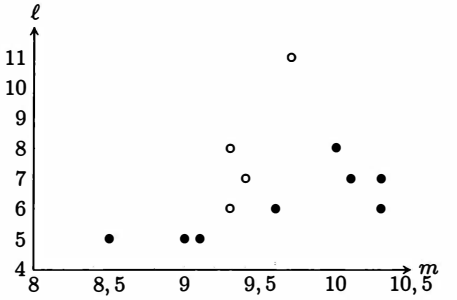
Remarque 8.26 (Limites de l'interprétation) Il convient d'être prudent dans l'utilisation du coefficient de corrélation. Une valeur $\rho = 0,8$ peut ou non signifier corrélation intéressante, selon les cas. Notamment, le nombre de données est crucial, comme on peut s'en convaincre avec l'exemple suivant, tiré de l'ouvrage de Jean MEEUS [75, chapitre 34]. On donne dans le tableau ci-dessous les magnitudes absolues[5] m de quelques astéroïdes, ainsi que le nombre ℓ de lettres de leur nom :

Nom	m	ℓ	Nom	m	ℓ
Niobe	8,5	5	Frigga	9,6	6
Feronia	10,3	7	Diana	9,1	5
Klytia	10,3	6	Eurynome	9,3	8
Galatea	10,1	7	Sappho	9,3	6
Eurydike	10,0	8	Terpsichore	9,7	11
Freia	9,0	5	Alkmene	9,4	7

Le coefficient de corrélation associé aux huit premiers astéroïdes (en noir sur la figure ci-après) est $\rho \approx 0,817$, ce qui est assez élevé. Pire, le coefficient associé aux cinq derniers est $\rho \approx 0,983$. Ceci n'est bien sûr qu'un accident, dû à la petitesse de l'échantillon. Il n'y a en réalité aucune corrélation entre la longueur du nom et la magnitude des astéroïdes. Si l'on prend l'échantillon de douze astéroïdes cités, le coefficient de corrélation devient $\rho \approx 0,4$ et, avec une cinquantaine d'astéroïdes[6], il n'est que de $\rho \approx -0,097$.

(5). C'est-à-dire une mesure logarithmique de leur luminosité à distance fixée.

(6). Nous reprenons le résultat de Jean Meeus, qui a utilisé les astéroïdes numérotés 51 à 100 dans les *Ephemerides of Minor Planets*, Leningrad, 1979.



8.4 ESPÉRANCE CONDITIONNELLE (CAS DISCRET)

Lois conditionnelles

§ 124

Soit (X, Y) un couple de variables aléatoires discrètes. On note $\{x_k \ ; \ k \in \mathrm{K}\}$ et $\{y_\ell \ ; \ \ell \in \mathrm{L}\}$ les valeurs prises par X et Y respectivement. Pour tout $k \in \mathrm{K}$ tel que $\mathbf{P}\{\mathrm{X} = x_k\}$ est non nulle, on définit la **loi de Y conditionnée par $\{\mathbf{X} = x_k\}$**, comme la loi de la restriction de Y à l'espace $\{\mathrm{X} = x_k\}$, muni de la probabilité $\mathbf{P}_{\{\mathrm{X}=x_k\}}$: pour tout événement B,

$$\mathbf{P}\{\mathrm{Y} \in \mathrm{B} \,|\, \mathrm{X} = x_k\} = \frac{\mathbf{P}\{\mathrm{Y} \in \mathrm{B}, \mathrm{X} = x_k\}}{\mathbf{P}\{\mathrm{X} = x_k\}}.$$

Cette loi sera notée $\mathbf{P}_{\mathrm{Y}}^{\{\mathrm{X}=x_k\}}$; elle est entièrement déterminée par les quantités ci-dessous, pour $\ell \in \mathrm{L}$:

$$\mathbf{P}\{\mathrm{Y} = y_\ell \,|\, \mathrm{X} = x_k\} = \frac{\mathbf{P}\{\mathrm{X} = x_k, \mathrm{Y} = y_\ell\}}{\mathbf{P}\{\mathrm{X} = x_k\}} = \frac{p_{k,\ell}}{p_{k,\cdot}}.$$

Bien sûr, lorsque $\mathbf{P}\{\mathrm{X} = x_k\}$ est nulle, on ne peut pas définir de manière cohérente la quantité $\mathbf{P}\{\mathrm{Y} = y_\ell \,|\, \mathrm{X} = x_k\}$.

Remarque 8.27 (Lois conditionnelles et reconstruction de la loi de couple) La connaissance de la loi de X et de la loi conditionnelle de Y par rapport à chacun des événements $\{\mathrm{X} = x_k\}$ tels que $\mathbf{P}\{\mathrm{X} = x_k\} > 0$ est suffisante pour reconstituer entièrement la loi du couple (X, Y). Soit (x_k, y_ℓ) un couple de valeurs prises par (X, Y).

- **1er cas : $\mathbf{P}\{\mathbf{X} = x_k\} \neq 0$.** Alors $\mathbf{P}\{\mathrm{X} = x_k, \mathrm{Y} = y_\ell\} = \mathbf{P}\{\mathrm{Y} = y_\ell \,|\, \mathrm{X} = x_k\} \cdot \mathbf{P}\{\mathrm{X} = x_k\}$.
- **2e cas : $\mathbf{P}\{\mathbf{X} = x_k\} = 0$.** L'événement $\{\mathrm{X} = x_k, \mathrm{Y} = y_\ell\}$ étant inclus dans $\{\mathrm{X} = x_k\}$, il est donc de probabilité nulle.

Dans tous les cas, on a pu obtenir $\mathbf{P}\{\mathrm{X} = x_k, \mathrm{Y} = y_\ell\}$. À noter que l'on peut étendre la définition de la loi conditionnelle de la façon suivante :

$$\mathbf{P}\{\mathrm{Y} = y_\ell \,|\, \mathrm{X} = x_k\} = \begin{cases} \dfrac{\mathbf{P}\{\mathrm{X} = x_k, \mathrm{Y} = y_\ell\}}{\mathbf{P}\{\mathrm{X} = x_k\}} & \text{si } \mathbf{P}\{\mathrm{X} = x_k\} \neq 0 \\ \text{une valeur arbitraire} & \text{si } \mathbf{P}\{\mathrm{X} = x_k\} = 0. \end{cases}$$

Avec cette définition étendue, la formule $\mathbf{P}\{\mathrm{X} = x_k, \mathrm{Y} = y_\ell\} = \mathbf{P}\{\mathrm{Y} = y_\ell \,|\, \mathrm{X} = x_k\} \cdot \mathbf{P}\{\mathrm{X} = x_k\}$ fournit la bonne loi du couple dans tous les cas. Pour des raisons de cohérence, il vaut mieux d'ailleurs choisir une famille $\{\alpha_\ell \ ; \ \ell \in \mathrm{L}\}$ de réels positifs, sommable et de somme 1, et poser

$$\mathbf{P}\{\mathrm{Y} = y_\ell \,|\, \mathrm{X} = x_k\} := \begin{cases} \dfrac{\mathbf{P}\{\mathrm{X} = x_k, \mathrm{Y} = y_\ell\}}{\mathbf{P}\{\mathrm{X} = x_k\}} & \text{si } \mathbf{P}\{\mathrm{X} = x_k\} \neq 0 \\ \alpha_\ell & \text{si } \mathbf{P}\{\mathrm{X} = x_k\} = 0, \end{cases}$$

afin que $\mathbf{P}_{\mathrm{Y}}^{\{\mathrm{X}=\ell\}}$ puisse décemment mériter le nom de « loi ».

§ 125 **Espérance conditionnelle, première approche**

Soit x_k une valeur telle que $\mathbf{P}\{X = x_k\} > 0$. On peut maintenant définir, si elle existe, l'**espérance de Y conditionnée par $\{X = x_k\}$** comme l'espérance de la variable Y selon la loi $\mathbf{P}_{\{X=x_k\}}$:

$$\mathbf{E}(Y \mid X = x_k) = \mathbf{E}_{\{X=x_k\}}(Y) = \sum_{\ell \in L} y_\ell \, \mathbf{P}\{Y = y_\ell \mid X = x_k\}.$$

■ **Proposition 8.28** *Si* Y *admet une espérance et si* $\mathbf{P}\{X = x_k\} > 0$, *alors* Y *admet une espérance conditionnelle* $\mathbf{E}(Y \mid X = x_k)$.

DÉMONSTRATION : La série définissant l'espérance conditionnelle de Y est $\sum_\ell y_\ell p_{k,\ell}/p_k$. Or, pour tout ℓ, on a $0 \leqslant p_{k,\ell} \leqslant p_\ell$ et la série $\sum_\ell y_\ell p_\ell$ converge absolument, donc la série $\sum_\ell y_\ell p_{k,\ell}$ également. ∎

Lorsque $\mathbf{P}\{X = x_k\} = 0$, on peut décider d'attribuer une valeur arbitraire à $\mathbf{E}(Y \mid X = x_k)$; de cette façon, on peut écrire une formule reliant l'espérance de Y et les espérances conditionnées par tous les événements du type $\{X = x_k\}$:

■ **Théorème 8.29 (Formule de l'espérance totale)** *Si* Y *admet une espérance, alors*

$$\mathbf{E}(Y) = \sum_{k \in K} \mathbf{P}\{X = x_k\} \cdot \mathbf{E}_{\{X=x_k\}}(Y).$$

DÉMONSTRATION : Il ne s'agit que d'intervertir deux sommes, ce que le théorème de Fubini autorise en vertu de la convergence absolue de toutes les séries concernées.

Notons $K^* := \{k \in K \, ; \, p_{k,\cdot} > 0\}$; alors

$$\sum_{k \in K} \mathbf{P}\{X = x_k\} \cdot \mathbf{E}_{\{X=x_k\}}(Y) = \sum_{k \in K^*} \mathbf{P}\{X = x_k\} \cdot \mathbf{E}_{\{X=x_k\}}(Y)$$
$$= \sum_{k \in K^*} p_{k,\cdot} \sum_{\ell \in L} y_\ell \frac{p_{k,\ell}}{p_{k,\cdot}} = \sum_{\ell \in L} y_\ell \sum_{k \in K^*} p_{k,\ell}.$$

Or, la dernière somme peut être étendue à l'ensemble K tout entier (puisque, si $p_{k,\cdot} = 0$, alors $p_{k,\ell} = 0$ pour tout $\ell \in L$), donc

$$\sum_{k \in K} \mathbf{P}\{X = x_k\} \cdot \mathbf{E}_{\{X=x_k\}}(Y) = \sum_{\ell \in L} y_\ell \sum_{k \in K} p_{k,\ell} = \sum_{\ell \in L} y_\ell \, p_{\cdot,\ell} = \mathbf{E}(Y).$$

∎

Exemple 8.30 (Naissance de garçons et de filles, 3/5) On revient une fois encore sur le modèle des naissances de garçons et de filles présenté dans les exemples 8.1 et 8.4. On désire de nouveau calculer le nombre moyen de naissance de filles, en utilisant les espérances conditionnelles. L'espérance conditionnelle du nombre de filles, sachant que le nombre total d'enfants est n, est aisée à calculer car c'est l'espérance d'une loi binomiale $\mathscr{B}(n \, ; \, p)$:

$$\mathbf{E}_{\{N=n\}}(X) = \sum_{k=0}^{\infty} k \, \mathbf{P}_{\{N=n\}}\{X = k\} = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = np.$$

On en déduit l'espérance totale

$$\mathbf{E}(X) = \sum_{n=0}^{\infty} \mathbf{E}_{\{N=n\}}(X) \cdot \mathbf{P}\{N = n\} = \sum_{n=0}^{\infty} np \frac{e^{-\lambda} \lambda^n}{n!} = p e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} = p\lambda.$$

(À suivre... → exemple 8.31)

Espérance conditionnelle, seconde approche § 126

Heuristiquement, on conditionne la variable aléatoire Y par l'événement $\{X = x_k\}$ lorsque l'on a effectué une mesure de X et que le résultat $X(\omega)$ est égal à x_k. Si l'on n'a pas d'autre renseignement, notre connaissance de Ω se limite donc à « $\omega \in \{X = x_k\}$ », et la valeur moyenne attendue de Y est $\mathbf{E}(Y \mid X = x_k)$.

Pour toute valeur de ω, une fois que l'on a mesuré $x := X(\omega)$, on peut estimer (ne serait-ce que théoriquement) la quantité $\mathbf{E}(Y \mid X = x)$. L'espérance conditionnelle de Y devient alors une fonction de ω, c'est-à-dire une variable aléatoire. Cette variable s'appelle **espérance conditionnelle de Y sachant X** et s'écrit, formellement,

$$\mathbf{E}(Y \,\|\, X) := \sum_{k \in K} \mathbf{E}(Y \mid X = x_k)\, \mathbf{1}_{\{X = x_k\}}.$$

Il est crucial de bien garder en tête le fait que $\mathbf{E}(Y \,\|\, X)$ est bien une *variable aléatoire* discrète, et non un simple nombre.

Exemple 8.31 (Naissance de garçons et de filles, 4/5) Nous avons calculé, dans l'exemple 8.31, que

$$\mathbf{E}(X \mid N = n) = np$$

pour tout n entier. Puisque l'ensemble des $\{N = n\}$ (n parcourant $\mathbb{N}$) est une partition de Ω, on en déduit que l'espérance conditionnelle de X sachant N est

$$\mathbf{E}(X \,\|\, N)(\omega) = \begin{cases} \mathbf{E}(X \mid N = 0) = 0 & \text{si } N(\omega) = 0, \\ \mathbf{E}(X \mid N = 1) = p & \text{si } N(\omega) = 1, \\ \mathbf{E}(X \mid N = 2) = 2p & \text{si } N(\omega) = 2, \\ \quad\vdots & \quad\vdots \\ \mathbf{E}(X \mid N = n) = np & \text{si } N(\omega) = n, \\ \quad\vdots & \quad\vdots \end{cases}$$

ou, en termes plus concis,

$$\mathbf{E}(X \,\|\, N) = p\, N.$$

(Suite et fin... → exemple 8.38)

Lorsque l'événement $\{X = x_k\}$ est de probabilité nulle, on attribue une valeur arbitraire à $\mathbf{E}(Y \,\|\, X)$, de sorte que l'on peut écrire

$$\mathbf{E}(Y \,\|\, X)(\omega) = \begin{cases} \mathbf{E}(Y \mid X = x_k) & \text{si } X(\omega) = x_k \text{ et } \mathbf{P}\{X = x_k\} > 0, \\ \text{une valeur arbitraire} & \text{si } X(\omega) = x_k \text{ et } \mathbf{P}\{X = x_k\} = 0. \end{cases}$$

On ne prend évidemment que rarement la peine de préciser quelle valeur arbitraire on choisit lorsque $\mathbf{P}\{X = x_k\} = 0$; à la place, on préfère dire que l'espérance conditionnelle est définie *à une égalité presque sûre près*, puisque la portion de Ω sur laquelle $\mathbf{E}(Y \,\|\, X)$ n'est pas bien définie est négligeable.

La formule de l'espérance totale prend alors la forme suivante :

■ **Théorème 8.32** *Si* Y *admet une espérance, alors* $\mathbf{E}(Y \,\|\, X)$ *est une variable aléatoire bien définie et admet une espérance ; de plus,*

$$\mathbf{E}\big(\mathbf{E}(Y \,\|\, X)\big) = \mathbf{E}(Y).$$

Enfin, par construction, on peut vérifier aisément le résultat suivant, qui sera généralisé aux variables quelconques (chapitre 23) :

■ **Théorème 8.33** *Soient* X *et* Y *deux variables aléatoires discrètes, telles que* Y *admet une espérance. Si* X *et* Y *sont indépendantes, alors* $\mathbf{E}(Y \,\|\, X) = \mathbf{E}(Y)$, *c'est-à-dire que c'est une fonction (presque sûrement) constante.* ◬8.7

Remarque 8.34 (Espérance conditionnelle et tribu engendrée par X *(*))*
La variable $\mathbf{E}(Y \| X)$ est « mesurable par rapport à la tribu $\sigma(X)$ », c'est-à-dire que, pour tout borélien B, l'ensemble $\mathbf{E}(Y \| X) \in B$ est de la forme $\{X \in C\}$, où C est encore un borélien (voir le chapitre 23). C'est une autre façon de dire que la fonction $\mathbf{E}(Y \| X)$ est constante sur tous les $\{X = x_k\}$. C'est une propriété fondamentale qui permettra de généraliser la notion d'espérance conditionnelle au cas où X est à valeurs continues, et non pas discrètes, et de contourner la difficulté apparaissant lorsque $\mathbf{P}\{X = x\} = 0$ pour tout réel x.

8.5 SOMMATION D'UN NOMBRE ALÉATOIRE DE VARIABLES

Au milieu des renseignements précis et des opinions très-sensées de MM. Benoiston de Châteauneuf, Galton et autres statisticiens, je n'ai pas rencontré la réflexion bien importante qu'ils auraient dû faire de l'extinction inévitable *des noms de famille. Évidemment tous les noms doivent s'éteindre [...]. Un mathématicien pourrait calculer comment la réduction des noms ou titres aurait lieu, d'après la probabilité des naissances toutes féminines ou toutes masculines ou mélangées, et la probabilité d'absence de naissances dans un couple quelconque.*

Alphonse DE CANDOLLE, Histoire des Sciences et des Savants, *1873*[7].

§ 127 **Contexte historique et mathématique**

Le statisticien Francis Galton se posait, en 1873, la question de la pérennité d'un nom de famille dans une société où ce nom se transmet uniquement par le père. N'ayant pas réussi à résoudre ce problème, il soumit sa question à l'*Educational Times*[8] :

> *Problem 4001: A large nation, of whom we will only concern ourselves with adult males, N in number, and who each bear separate surnames colonise a district. Their law of population is such that, in each generation, a_0 per cent of the adult males have no male children who reach adult life; a1 have one such male child; a_2 have two; and so on up to a_5 who have five. Find (1) what proportion of their surnames will have become extinct after r generations; and (2) how many instances there will be of the surname being held by m persons.*[9]

N'ayant reçu que des réponses incomplètes ou fausses, il demanda à son ami, le révérend Henry William Watson, qui le premier posa le problème de façon correcte. Sa réponse hélas fut erronée : il conclut que la probabilité d'extinction était toujours égale à 1[10].

La résolution de ce problème passant par la sommation d'un nombre aléatoire de variables aléatoires, nous commençons par une l'étude d'un problème plus simple.

---

(7). Cité par Francis Galton dans l'introduction à l'article du révérend Watson : *On the probability of extinction on families*, paru dans le *Journal of the Royal Anthropological Institute.* Alphonse de Candolle était un botaniste franco-suisse.

(8). L'utilisation d'outils mathématiques, et notamment de probabilités, dans le but d'étudier la dynamique d'une population, est exposée dans le remarquable petit ouvrage : Nicolas BACAËR, *Histoires de mathématiques et de populations* [3].

(9). Problème 4001 : Une grande nation, parmi laquelle nous ne considérerons que les hommes adultes, au nombre de N, et qui portent des noms de familles différents, colonise un district. La loi de leur population est telle que, à chaque génération, a_0 pourcent des hommes n'ont aucun descendant mâle atteignant l'âge adulte ; a_1 en ont un ; a_2 en ont deux ; et ainsi de suite jusqu'à a_5 qui en ont cinq. Trouver (1) quelle proportion de noms de familles sera éteinte au bout de r générations ; et (2) combien de fois un même nom de famille sera porté par m personnes.

(10). En réalité, Jules Bienaymé avait déjà trouvé la réponse correcte en 1845, mais l'avait publiée dans une revue peu connue. Par ailleurs, sa note était quelque peu lapidaire et ne contenait pas la démonstration complète de ses calculs.

Identités de Wald

§ 128

On se donne une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires discrètes, *identiquement distribuées*, c'est-à-dire de même loi, et N une variable aléatoire à valeurs entières positives ; on suppose que N et les X_k sont mutuellement indépendantes. On définit alors une nouvelle variable

$$S_N = X_1 + X_2 + \cdots + X_N,$$

c'est-à-dire que S_N est une somme des X_k, dont le nombre de termes est lui même une variable aléatoire : pour tout $\omega \in \Omega$, on a

$$S_N(\omega) = \sum_{k=1}^{N(\omega)} X_k(\omega).$$

(On convient que, si $N(\omega) = 0$, alors $S_N(\omega) = 0$.) On se pose la question de calculer l'espérance de S_N. Le théorème suivant donne une réponse somme toute conforme à l'intuition : s'il y a en moyenne n variables aléatoires dans la somme, et si chacune est de valeur moyenne m, alors la valeur moyenne de S_N est mn.

■ **Théorème 8.35 (Identité de Wald)** *Si les variables aléatoires X_k admettent une espérance, si N admet une espérance, alors*

$$\mathbf{E}(S_N) = \mathbf{E}(X_1 + X_2 + \cdots + X_N) = \mathbf{E}(N) \cdot \mathbf{E}(X).$$

Si les X_k admettent un moment d'ordre 2, alors

$$\mathbf{E}\left([S_N - N\,\mathbf{E}X]^2 \right) = \mathbf{E}(N) \cdot \mathbf{V}(X).$$

DÉMONSTRATION : L'ensemble des valeurs possiblement prises par S_N est dénombrable, car inclus dans une union dénombrable d'ensemble dénombrables. On le notera $\{s_k \ ; \ k \in \mathbb{N}\}$.

La famille des $\{N = n\}$, pour n parcourant $\mathbb{N}$, est un système quasi-complet d'événements, donc l'événement $\{S_N = s_k\}$ se décompose en l'union disjointe

$$\{S_N = s_k\} = \sum_{n=0}^{\infty} \{S_N = s_k, N = n\}.$$

L'intersection des événements $\{S_N = s_k\}$ et de $\{N = n\}$ est évidemment la même que l'intersection des événements $\{X_1 + X_2 + \cdots + X_n = s_k\}$ et $\{N = n\}$, ce qui permet d'écrire

$$\{S_N = s_k\} = \sum_{n=0}^{\infty} \{X_1 + X_2 + \cdots + X_n = s_k, N = n\}$$

et donc

$$\mathbf{P}\{S_N = s_k\} = \sum_{n=0}^{\infty} \mathbf{P}\{X_1 + X_2 + \cdots + X_n = s_k, N = n\}.$$

Enfin, puisque les variables $X_1 + X_2 + \cdots + X_n$ et N sont indépendantes (ce qui n'est évidemment pas le cas de $X_1 + X_2 + \cdots + X_N$ et N...), on obtient

$$\mathbf{P}\{S_N = s_k\} = \sum_{n=0}^{\infty} \mathbf{P}\{X_1 + X_2 + \cdots + X_n = s_k\} \cdot \mathbf{P}\{N = n\}.$$

Calculons maintenant l'espérance de S_N (la justification de la convergence absolue et la permutation des sommes sera faite à la fin) :

$$\begin{aligned}
\mathbf{E}(S_N) &= \sum_{k=0}^{\infty} s_k \sum_{n=0}^{\infty} \mathbf{P}\{X_1 + X_2 + \cdots + X_n = s_k\} \cdot \mathbf{P}\{N = n\} \\
&= \sum_{n=0}^{\infty} \mathbf{P}\{N = n\} \sum_{k=0}^{\infty} s_k \, \mathbf{P}\{X_1 + X_2 + \cdots + X_n = s_k\} \qquad \textit{(Fubini)} \\
&= \sum_{n=0}^{\infty} \mathbf{P}\{N = n\} \, \mathbf{E}(X_1 + X_2 + \cdots + X_n)
\end{aligned}$$

Par linéarité et équidistribution des variables, on a

$$\mathbf{E}(X_1 + X_2 + \cdots + X_n) = \mathbf{E}(X_1) + \mathbf{E}(X_2) + \cdots + \mathbf{E}(X_n) = n\,\mathbf{E}(X)$$

donc

$$\mathbf{E}(S_N) = \sum_{n=0}^{\infty} n\,\mathbf{P}\{N = n\}\ \mathbf{E}(X) = \mathbf{E}(N)\,\mathbf{E}(X).$$

Nous pouvons maintenant justifier la sommabilité de la famille mise en jeu dans la double somme par l'existence des espérances de X et de N, ce qui conclut la preuve de la première formule.

Pour la seconde, un raisonnement semblable montre que

$$\mathbf{E}\big([S_N - N\,\mathbf{E}X]^2\big) = \sum_{n=0}^{\infty} \mathbf{P}\{N = n\}\ \mathbf{E}\big((S_n - n\,\mathbf{E}X)^2\big) = \sum_{n=0}^{\infty} \mathbf{P}\{N = n\}\ \mathbf{V}(S_n)$$
$$= \sum_{n=0}^{\infty} \mathbf{P}\{N = n\}\ n\ \mathbf{V}(X) = \mathbf{E}(N) \cdot \mathbf{V}(X).$$

∎

Remarque 8.36 Le fait que $\mathbf{E}(S_N) = \mathbf{E}(N)\,\mathbf{E}(X)$ peut se démontrer de manière plus générale, lorsque X est une variable aléatoire quelconque admettant une espérance, et pas forcément une variable aléatoire discrète.

Cette propriété est encore vraie si N n'est pas indépendante des X_n, mais si l'événement $\{N = n\}$ est indépendant de $X_{n+1}, X_{n+2}, \ldots$ (on dit alors que N est un *temps d'arrêt*). Cette généralisation sera démontrée au paragraphe § 252, chapitre 16.

Dans le cas où les variables X_n prennent des valeurs entières, il existe une relation simple entre les fonctions génératrices des variables X, N et S :

■ **Théorème 8.37** *On suppose que les variables aléatoires X_k, toujours indépendantes et identiquement distribuées, suivent une loi à valeurs dans $\mathbb{N}$. Notons G_N la fonction génératrice de N, G_X celle, commune, des X_i, et G_S celle de S_N. Alors $G_S = G_N \circ G_X$, c'est-à-dire que*

$$\forall t \in [-1\,;+1] \qquad G_S(t) = G_N\big(G_X(t)\big).$$

DÉMONSTRATION : Pour tout entier n, notons $S_n = \sum_{k=1}^{n} X_k$. Par indépendance des X_k, on a $G_{S_n} = G_X^n$. La probabilité que S_N soit égale à un entier n est

$$\mathbf{P}\{S_N = n\} = \sum_{k=0}^{\infty} \mathbf{P}\{S_N = n, N = k\} = \sum_{k=0}^{\infty} \mathbf{P}\{S_k = n, N = k\}$$
$$= \sum_{k=0}^{\infty} \mathbf{P}\{S_k = n\} \cdot \mathbf{P}\{N = k\}\,,$$

cette dernière égalité provenant de l'indépendance des X_i par rapport à N. Alors, pour tout réel $t \in\ ]-1\,;+1[$,

$$G_S(t) = \sum_{n=0}^{\infty} \mathbf{P}\{S_N = n\}\ t^n = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} \mathbf{P}\{S_k = n\} \cdot \mathbf{P}\{N = k\}\ t^n \qquad \textit{(Fubini)}$$
$$= \sum_{k=0}^{\infty} \mathbf{P}\{N = k\} \cdot \underbrace{\sum_{n=0}^{\infty} \mathbf{P}\{S_k = n\}\ t^n}_{G_{S_k}(t)} = \sum_{k=0}^{\infty} \mathbf{P}\{N = k\}\ \big(G_X(t)\big)^k = G_N\big(G_X(t)\big).$$

∎

⁂8.8

Exemple 8.38 (Naissance de garçons et de filles, 5/5) Nous avons maintenant des outils plus performants pour reprendre le problème des naissances de garçons et de filles. On rappelle

que, dans notre modèle, une famille engendre N enfants, où N suit une loi de Poisson de paramètre $\lambda > 0$; le nombre de filles est une variable aléatoire pouvant s'écrire

$$X = B_1 + B_2 + \cdots + B_N,$$

c'est-à-dire

$$\forall \omega \in \Omega \qquad X(\omega) = \sum_{k=1}^{N(\omega)} B_k,$$

où les variables aléatoires B_k sont indépendantes et de même loi de Bernoulli $\mathscr{B}(p)$, et où il est donc entendu que $X = 0$ si $N = 0$. D'après le théorème précédent, la fonction génératrice de X est $G_X = G_N \circ G_B$; or $G_B(t) = (1-p) + pt$, $G_N(u) = e^{\lambda(u-1)}$ et donc

$$G_X(t) = e^{\lambda p(t-1)},$$

ce qui est la fonction génératrice d'une loi de Poisson de paramètre λp. Puisque la fonction génératrice caractérise la loi (théorème 7.35 page 160), on en déduit que X suit la loi de Poisson de paramètre λp.

Processus de Galton-Watson

§ 129

On étudie la dynamique d'un modèle simple de population issue d'un seul individu[11]. Le temps est discrétisé et représenté par une variable n à valeurs entières. À l'instant initial $n = 0$, le nombre total d'individus est donc $Z_0 = 1$. À chaque nouvel instant n, chacun des individus donne naissance à un certain nombre de descendants, et meurt. On suppose que la loi du nombre de descendants est la même pour chaque individu, et que ce nombre est indépendant de la descendance des autres individus.

Pour formaliser ceci, on se donne une variable aléatoire X de loi discrète sur $\mathbb{N}$:

$$\mathbf{P}\{X = k\} = p_k,$$

une suite double (X_i^n) (où $n \geqslant 0$ et $i \geqslant 1$) de variables aléatoires indépendantes et de même loi que X et on écrit que, à la génération n, s'il y a Z_n individus, chacun des individus, labellé par un indice i variant entre 1 et Z_n, va engendrer X_i^n descendants. À la génération suivante, le nombre d'individus de la population est

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_i^n.$$

(On notera que, si $Z_n = 0$, alors nécessairement $Z_{n+1} = 0$, comme il se doit. On remarquera également que, en fonction de la réalisation ω du hasard, certaines des variables X_i^n ne servent pas.) C'est ce que l'on appelle un **processus de branchement**, ou **processus de Galton-Watson**.

On suppose que les X_i^n admettent un moment d'ordre 2, et on note

$$m = \mathbf{E}(X) \qquad \sigma = \sigma_X.$$

Posons

$$G(t) = \mathbf{E}(t^X) = \sum_{k=0}^{\infty} p_k t^k$$

(11). Biologiquement, cela sonne étrange si l'on songe à la reproduction sexuée. Dans le problème soulevé par Francis Galton, il s'agissait de la postérité d'un nom de famille ; l'individu unique représente, selon les conventions de l'Angleterre victorienne en ce qui concerne la transmission du nom, un homme — les femmes ne jouant aucun rôle dans la transmission des noms à cette époque (ces règles se sont quelque peu assouplies en Europe depuis le XIX^{e} siècle et le modèle présenté ne peut plus décrire correctement le processus d'extinction des noms). Ce modèle semble cependant plus adapté à l'étude de la descendance d'une bactérie, d'une épidémie, ou du nombre de neutrons dans un réacteur nucléaire, les descendants étant alors les neutrons libérés lors de la collision avec un noyau atomique.

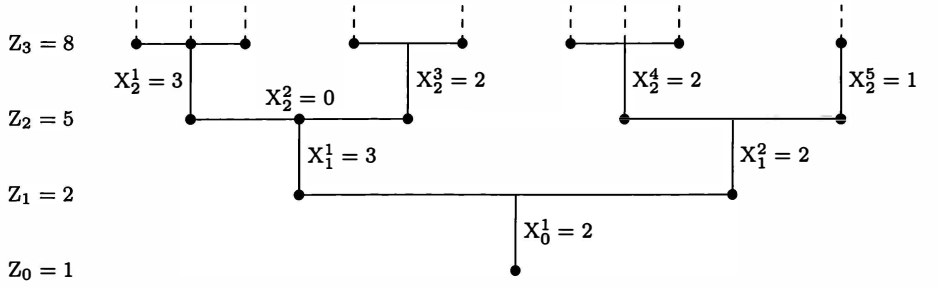


FIGURE 8.1 — *Arbre de la descendance de Z_0, pour une réalisation particulière du hasard. À la deuxième génération, un des individus meurt sans laisser de descendance.*

la fonction génératrice des X_i^n. Enfin, on note x_n la probabilité que la population soit éteinte à la génération n : $x_n = \mathbf{P}\{Z_n = 0\}$.

On remarquera que l'événement « M : la population finit par s'éteindre » est l'union croissante des événements $\{Z_n = 0\}$ pour $n \geqslant 0$; d'après l'axiome 3′, on a donc

$$\mathbf{P}(M) = \mathbf{P}\Big(\bigcup_{n=0}^{\infty} \{Z_n = 0\}\Big) = \lim_{n\to\infty} x_n.$$

§130 **Probabilité d'extinction**

On peut déterminer la probabilité d'extinction en étudiant le comportement de la suite $(x_n)_{n\geqslant 0}$. Le théorème 8.37 permet de calculer la loi de Z_n :

■ **Théorème 8.39** *La fonction génératrice G_n de Z_n vérifie la relation de récurrence*

$$G_{n+1} = G \circ G_n.$$

Puisque $G_0 = \mathrm{Id}$ par construction, on en déduit $G_1 = G$ et

$$G_n = \underbrace{G \circ G \circ \cdots \circ G}_{n \text{ fois}}.$$

Puisque $x_n = \mathbf{P}\{Z_n = 0\} = G_n(0)$, on en déduit que

■ **Proposition 8.40 (Steffensen[12], 1930)** *La suite $(x_n)_{n\geqslant 0}$ est caractérisée par sa valeur initiale $x_0 = 0$ et la relation de récurrence*

$$x_{n+1} = G(x_n).$$

Notamment, la suite $(x_n)_{n\geqslant 0}$ converge, et sa limite α est égale à la plus petite racine de l'équation $G(x) = x$ sur $[0\,;1]$.

DÉMONSTRATION : La fonction G est croissante, $G(0) \geqslant 0$ et $G(1) = 1$: ainsi, G laisse stable l'intervalle $[0\,;1]$. Notons α la plus petite racine sur $[0\,;1]$ de l'équation $G(x) = x$.

Si $G(0) = 0$, alors $\alpha = 0$ et la suite $(x_n)_{n\geqslant 0}$ est constante égale à 0 ; il s'agit du cas où un individu donne toujours naissance à au moins un descendant, ce qui fait que l'extinction ne saurait avoir lieu.

Si $G(0) > 0$, alors l'intervalle $[0\,;\alpha]$ est stable par G (voir figure 8.2 page ci-contre), et $G(x) \geqslant x$ pour tout $x \in]0\,;\alpha[$. Cela prouve que : la suite $(x_n)_{n\geqslant 0}$ est à valeurs dans $[0\,;\alpha]$, est croissante et converge ; par continuité, elle ne peut converger que vers un point fixe de G, c'est-à-dire vers α. ∎

(12). Johan Frederik Steffensen, professeur de mathématiques à l'université de Copenhague.

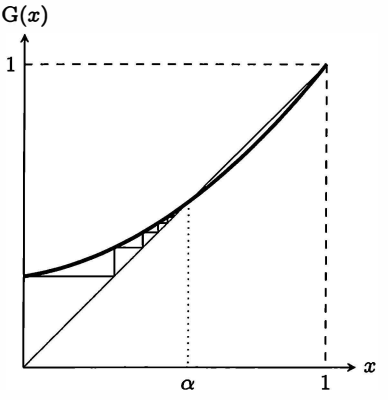


FIGURE 8.2 — *Fonction* G *génératrice de* X. *Le réel* α *est la plus petite racine de l'équation* $G(x) = x$. *On construit également les premiers termes de la suite définie par* $x_0 = 0$ *et* $x_{n+1} = G(x_n)$.

La fonction G est croissante et convexe, et vérifie de plus $G(1) = 1$. Par conséquent, la valeur de α est fortement reliée au comportement de la fonction G au voisinage de 1. Notamment :

- si $m = G'(1) < 1$, alors $\alpha = 1$; de plus, la suite $(x_n)_{n\geqslant 0}$ converge vers 1 à une vitesse exponentiellement rapide. L'extinction est certaine. Le système est dit **sous-critique**.
- Si $m = G'(1) > 1$, alors $0 < \alpha < 1$. L'extinction de la population n'est pas certaine. Le système est dit **sur-critique**.
- si $m = G'(1) = 1$, le système est dit **critique** ; il faut encore séparer deux cas :
  - la fonction G est égale à l'identité, ce qui revient à dire que $p_1 = 1$ et $p_k = 0$ pour tout $k \neq 1$; autrement dit, le modèle est déterministe et chaque individu donne naissance à exactement 1 descendant. Dans ce cas $\alpha = 0$ et l'extinction n'a jamais lieu ;
  - la fonction G n'est pas égale à l'identité, il existe des indices $k \geqslant 2$ tels que $p_k > 0$; G est alors strictement convexe et $\alpha = 1$: l'extinction est encore certaine.

  On pourra vérifier que la distinction entre ces deux cas peut se faire ainsi : $\mathbf{V}(X) = 0$ dans le premier cas, $\mathbf{V}(X) > 0$ dans le second.

Remarque 8.41 Si le révérend Watson avait bien établi que α était solution de l'équation $G(x) = x$, il crut pouvoir en déduire que $\alpha = 1$ puisque 1 est toujours point fixe de G.

Remarque 8.42 Le seul cas où la probabilité d'extinction est nulle est lorsque $p_0 = 0$; s'il y a une chance, aussi petite soit-elle, qu'un individu n'ait aucune descendance, alors la probabilité d'extinction est non nulle (même dans le cas sur-critique).

Évolution de la population

§ 131

On s'intéresse ici à la valeur moyenne de la population à la n-ième génération, c'est-à-dire à $\mathbf{E}(Z_n)$. Puisque $\mathbf{E}(Z_{n+1}) = G'_{n+1}(1) = (G\circ G_n)'(1) = G'(1)\cdot G_n(1) = m\,G_n(1)$, on déduit que

$$\forall n \in \mathbb{N} \qquad \mathbf{E}(Z_n) = m^n.$$

Ainsi :

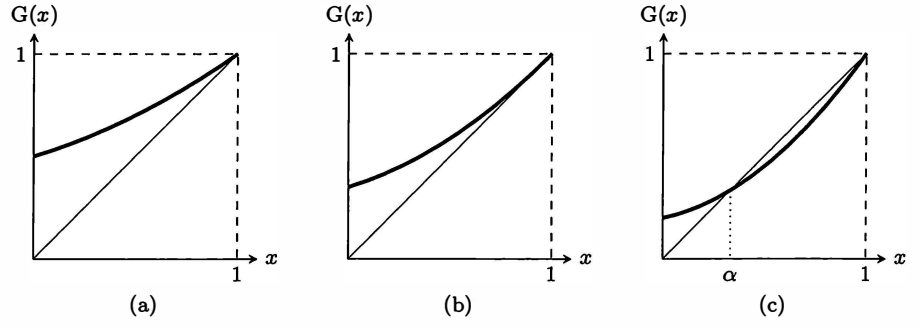


FIGURE 8.3 — *Fonction génératrice de* X, *dans les cas : (a) sur-critique, (b) critique et (c) sous-critique.*

— Dans le cas sous-critique $m < 1$, la population moyenne tend rapidement vers 0.
— Dans le cas critique $m = 1$, la population moyenne reste constante.
— Dans le cas sur-critique, la population moyenne tend rapidement vers $+\infty$.

Remarque 8.43 (Explosion de la population et extinction) On remarquera que, dans le cas sur-critique, si la population moyenne « explose » exponentiellement vite, sa probabilité d'extinction n'est cependant pas nulle. Ce n'est pas contradictoire : une proportion α de Ω mène à une extinction (plus ou moins rapide), tandis que les autres mènent à un accroissement rapide de la population. En prenant en compte toutes ces éventualités, l'effectif moyen de la population augmente exponentiellement vite.

De fait, il est plus intéressant de calculer la population moyenne *sachant* qu'il n'y a pas extinction ; il est possible de montrer que $\mathbf{E}(\mathrm{Z}_{n+1} \,\|\, \mathrm{Z}_n) = m\,\mathrm{Z}_n$, ce qui permet de conclure, en utilisant des théorèmes de martingales (théorème 23.32 page 486) que la suite de variables Z_n/m^n converge, sur un ensemble Ω' de probabilité 1, vers une certaine variable aléatoire L (d'espérance unité) ; de plus, L est non nulle sur un ensemble qui coïncide presque sûrement avec l'ensemble des ω pour lesquels il n'y a pas extinction (cf. par exemple Williams [100] ou Delmas & Jourdain [22]).

EXERCICES

♦ ***Exercice 8.1 (Indépendance et décorrélation chez Bernoulli)***
Soient X et Y de variables aléatoires de lois de Bernoulli respectives $\mathscr{B}(p)$ et $\mathscr{B}(q)$. Montrer qu'elles sont indépendantes si et seulement si elles sont décorrélées.

♦ ***Exercice 8.2 (Fonction génératrice d'une combinaison linéaire)*** Soit n un entier ; on considère des variables aléatoires $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$ indépendantes, à valeurs dans $\mathbb{N}$, de fonctions génératrices respectives $\mathrm{G}_1, \mathrm{G}_2, \ldots, \mathrm{G}_n$. Calculer la fonction génératrice de $\mathrm{Y} = \alpha_1\mathrm{X}_1 + \alpha_2\mathrm{X}_2 + \cdots + \alpha_n\mathrm{X}_n$.

♦ ***Exercice 8.3*** On considère une succession (infinie) de tirages à *pile* ou *face*, avec une pièce donnant *pile* avec une probabilité p $(0 < p < 1)$. On note N le nombre de tirages nécessaires pour avoir le premier *pile* ; et l'on effectue alors N nouveaux tirages de pièce. Montrer que cette expérience permet d'obtenir en moyenne un *pile* :

i) en explicitant la loi de X ;
ii) en utilisant les espérances conditionnelles.

♦ ***Exercice 8.4*** Soient X et Y deux variables aléatoires discrètes, prenant respectivement les valeurs $\{x_k \,;\, k \in \mathrm{K}\}$ et $\{y_\ell \,;\, \ell \in \mathrm{L}\}$. On suppose que, pour tout y_ℓ tel que $\mathbf{P}\{\mathrm{Y} = y_\ell\} > 0$, la loi de X sachant $\{\mathrm{Y} = y_\ell\}$ est une même loi $\mathscr{L}$. Vérifier que X et Y sont indépendants et que la loi de X est $\mathscr{L}$.

♦ ***Exercice 8.5 (Piper des dés)*** Voici une application amusante des fonctions génératrices : *quelle que soit la façon dont on pipe deux dés, la somme des points de ces dés ne peut pas suivre une loi uniforme sur* $\{2, 3, \dots, 12\}$.

i) On considère le polynôme $P(t) = \sum_{k=0}^{10} t^k$. Peut-on décomposer ce polynôme sous forme d'un produit de deux polynômes réels de degré inférieur ou égal 5 ?

ii) On lance deux dés à six faces. Notons S la variable aléatoire égale à la somme des deux faces. Montrer qu'il n'est pas possible de truquer les deux dés (pas nécessairement de la même manière) de telle sorte que S suive une loi uniforme.

♦ ***Exercice 8.6*** Soient X et Y deux variables aléatoires réelles indépendantes qui suivent chacune une loi géométrique de paramètre $p \in]0\,;\,1[$. On définit $Z = X/Y$. Quelles sont les valeurs prises par Z ? Déterminer sa loi et son espérance.

♦ ***Exercice 8.7 (Propriétés de l'espérance conditionnelle)*** Soient X, Y, Y′ des variables aléatoires discrètes. On suppose que Y, Y′ admettent une espérance. Soient λ, μ des réels. Vérifier les égalités suivantes (valables presque sûrement) :

i) linéarité : $\mathbf{E}(\lambda Y + \mu Y' \,\|\, X) = \lambda\,\mathbf{E}(Y \,\|\, X) + \mu\,\mathbf{E}(Y' \,\|\, X)$;

ii) $\mathbf{E}(Y \,\|\, Y) = \mathbf{E}(Y)$.

♦ ***Exercice 8.8*** Redémontrer le théorème 8.37 sans utiliser la formule des probabilités totales, mais en utilisant l'espérance conditionnelle et en remarquant que

$$G_S(t) = \mathbf{E}(t^{S_N}) = \mathbf{E}\big(\mathbf{E}(t^{S_N} \,\|\, N)\big).$$

♦ ***Exercice 8.9 (Une bébé-chaîne de Markov)*** Un joueur joue avec deux pièqces B_0 et B_1 de la façon suivante : si au tirage k il obtient *pile*, il change de pièce pour le tirage suivant, sinon il garde la même pièce. On suppose que la pièce B_1 a une probabilité $p = 0,35$ de donner pile, et la pièce B_0, une probabilité $q = 0,4$. On définit la variable aléatoire X_k égale à 1 si l'on joue avec la pièce B_1 au tirage k, égale à 0 sinon. On notera

$$u_k := \mathbf{P}\{X_k = 1\}, \qquad v_k := \mathbf{P}\{X_k = 0\} \qquad \text{et} \qquad H_k := \begin{pmatrix} u_k \\ v_k \end{pmatrix}.$$

i) Montrer que $H_{k+1} = M\,H_k$ où M est une matrice 2×2 que l'on déterminera. En déduire H_k en fonction de H_0.

ii) Diagonaliser M. Montrer que $(H_k)_{k \geqslant 1}$ converge vers un vecteur indépendant des conditions initiales. En déduire que la loi de X_k converge vers une loi que l'on précisera.

♦ ***Exercice 8.10 (Maximum de deux lois géométriques)***
Soient X et Y deux variables aléatoires indépendantes suivant une même loi géométrique $\mathscr{G}(p)$. On note $Z = \max(X, Y)$. Déterminer la loi de Z et calculer son espérance.

Indications

◊ *Indication 8.1 :* On rappelle (théorème 5.5 page 112) que deux événements A et B sont indépendants si et seulement si A et B^c le sont.

◊ *Indication 8.2 :* On remarquera que $t^Y = t^{\alpha_1 X_1} \cdot t^{\alpha_2 X_2} \cdots t^{\alpha_n X_n}$.

SOLUTIONS

♦ **Solution 8.1** On sait déjà que, si elles sont indépendantes, elles sont décorrélées. Supposons donc qu'elle sont décorrélées : $\mathbf{Cov}(X, Y) = 0$. Notamment, $\mathbf{E}(XY) = \mathbf{E}(X)\,\mathbf{E}(Y)$. Or $\mathbf{E}(X) = \mathbf{P}\{X = 1\}$, $\mathbf{E}(Y) = \mathbf{P}\{Y = 1\}$ et, surtout, $\mathbf{E}(XY) = \mathbf{P}\{XY = 1\} = \mathbf{P}\{X = 1, Y = 1\}$. On a bien $\mathbf{P}\{X = 1, Y = 1\} = \mathbf{P}\{X = 1\}\,\mathbf{P}\{Y = 1\}$: les événements $\{X = 1\}$ et $\{Y = 1\}$ sont indépendants.

On en déduit successivement, par utilisation de l'indication, que sont indépendants :

- $\{X = 1\}$ et $\{Y = 0\}$;
- $\{X = 0\}$ et $\{Y = 0\}$;
- $\{X = 0\}$ et $\{Y = 1\}$.

Conclusion : les événements $\{X = i\}$ et $\{Y = j\}$ sont indépendants, pour tout couple (i, j) d'entiers de $\{0, 1\}$, c'est-à-dire que X et Y sont indépendantes.

♦ **Solution 8.2** Par indépendance,

$$\begin{aligned} G_Y(t) &= \mathbf{E}(t^{\alpha_1 X_1} \cdot t^{\alpha_2 X_2} \cdots t^{\alpha_n X_n}) = \mathbf{E}(t^{\alpha_1 X_1}) \cdot \mathbf{E}(t^{\alpha_2 X_2}) \cdots \mathbf{E}(t^{\alpha_n X_n}) \\ &= G_1(t^{\alpha_1}) \cdot G_2(t^{\alpha_2}) \cdots G_n(t^{\alpha_n}). \end{aligned}$$

♦ **Solution 8.3** Tout d'abord, N est une variable aléatoire suivant la loi géométrique $\mathscr{G}(p)$ de paramètre p, c'est-à-dire que $\mathbf{P}\{N = n\} = p\,q^{n-1}$ pour tout $n \in \mathbb{N}^*$ (en notant $q := 1 - p$). On note X le nombre de *pile* obtenus au cours des N tirages. La loi de X, conditionnée par un événement $\{N = n\}$, est une loi binomiale de paramètres n et p :

$$X \rightsquigarrow_{\{N=n\}} \mathscr{B}(n\,;\,p) \qquad \text{ou encore} \qquad \mathbf{P}\{X = k \,|\, N = n\} = \binom{n}{k} p^k\, q^{n-k}.$$

i) Dans une première méthode, on explicite la loi de X. La formule des probabilités totales, appliquée au système quasi-complet d'événements $\big\{\{N = n\}\ ;\ n \in \mathbb{N}^*\big\}$, permet d'écrire que, pour tout $k \in \mathbb{N}$,

$$\begin{aligned} \mathbf{P}\{X = k\} &= \sum_{n=k}^{\infty} \mathbf{P}\{X = k \,|\, N = n\} \cdot \mathbf{P}\{N = n\} \\ &= \sum_{n=k}^{\infty} \binom{n}{k} p^{k+1}\, q^{2n-k-1} = p^{k+1}\, q^{k-1} \sum_{n=k}^{\infty} \binom{n}{k} q^{2n-2k}. \end{aligned}$$

Une des formules du paragraphe A55, page 691 donne alors

$$\mathbf{P}\{X = k\} = \frac{p^{k+1}\, q^{k-1}}{(1-q^2)^{k+1}} = \frac{q^{k-1}}{(1+q)^{k+1}}.$$

On en tire alors l'espérance de X, grâce à la formule (A.4) :

$$\mathbf{E}(X) = \sum_{k=0}^{\infty} k\,\mathbf{P}\{X = k\} = \frac{1}{q(1+q)} \sum_{k=1}^{\infty} k \left(\frac{q}{1+q}\right)^k = 1$$

ii) L'espérance de X vaut (dans $\mathbb{R}^+ \cup \{+\infty\}$, tous les calculs suivants sont valides puisque l'on ne somme que des termes positifs)

$$\mathbf{E}(X) = \sum_{k=0}^{\infty} k\,\mathbf{P}\{X = k\} = \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} k\,\mathbf{P}\{X = k | N = n\} \cdot \mathbf{P}\{N = n\}.$$

On applique alors le théorème de Fubini pour intervertir les sommes :

$$\begin{aligned} \mathbf{E}(X) &= \sum_{n=1}^{\infty} \sum_{k=0}^{n} k\,\mathbf{P}\{X = k | N = n\} \cdot \mathbf{P}\{N = n\} \\ &= \sum_{n=1}^{\infty} \left(\sum_{k=0}^{n} k\,\mathbf{P}\{X = k | N = n\} \right) \cdot \mathbf{P}\{N = n\}. \end{aligned}$$

Dans le terme entre parenthèses, on reconnaît l'espérance d'une loi binomiale $\mathscr{B}(n\,;\,p)$, ce qui est conforme à la formule de l'espérance conditionnelle :

$$\mathbf{E}(X) = \sum_{n=1}^{\infty} \mathbf{E}_{\{N=n\}}(X) \cdot \mathbf{P}\{N = n\} = \sum_{n=1}^{\infty} (np) \cdot p\, q^{n-1} = 1.$$

♦ **Solution 8.4** Notons $\mathrm{L}^* := \{\ell \in \mathrm{L}\ ;\ \mathbf{P}\{\mathrm{Y} = y_\ell\} > 0\}$. Par hypothèse, pour tout $k \in \mathrm{K}$, il existe un réel p_k tel que : $\mathbf{P}\{\mathrm{X} = x_k \,|\, \mathrm{Y} = y_\ell\} = p_k$ pour tout $\ell \in \mathrm{L}^*$. La formule des probabilités totales, utilisée avec le système quasi-complet d'événements $\big(\{\mathrm{Y} = y_\ell\}\big)_{\ell \in \mathrm{L}^*}$, donne

$$\mathbf{P}\{\mathrm{X} = x_k\} = \sum_{\ell \in \mathrm{L}^*} \{\mathrm{Y} = y_\ell\}\, p_k = p_k.$$

Notamment, pour tous $k \in \mathrm{K}$ et tout $\ell \in \mathrm{L}$, on a

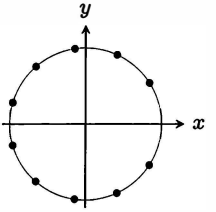
$$\mathbf{P}\{\mathrm{X} = x_k, \mathrm{Y} = y_\ell\} = p_k\, \mathbf{P}\{\mathrm{Y} = y_\ell\} = \mathbf{P}\{\mathrm{X} = x_k\}\, \mathbf{P}\{\mathrm{Y} = y_\ell\},$$

cette formule découlant de l'hypothèse si $\ell \in \mathrm{L}^*$, et étant immédiate si $\mathbf{P}\{\mathrm{Y} = y_\ell\} = 0$. Ainsi, X et Y sont indépendantes, et $\mathrm{X} \rightsquigarrow \mathscr{L}$.

♦ **Solution 8.5** On commence par écrire que

$$1 + t + t^2 + \cdots + t^{10} = \frac{1 - t^{11}}{1 - t} = \prod_{k=1}^{10} (t - \omega^k),$$

où l'on a défini $\omega = \mathrm{e}^{2\mathrm{i}\pi/11}$. Supposons que l'on factorise P en un produit $\mathrm{P} = \mathrm{QR}$ de deux polynômes réels de degré 5. Alors les racines de Q et celles de R forment une partition de l'ensemble $\{\omega, \omega^2, \ldots, \omega^{10}\}$ figuré ci-dessous ; notamment, aucun des deux polynômes ne peut avoir ses racines deux à deux conjuguées.



La fonction génératrice associée à un dé à six faces est un polynôme de degré inférieur ou égal à 6. Supposons que l'on puisse piper les dés comme indiqué, et notons $\mathrm{G_X}$ et $\mathrm{G_Y}$ les fonctions génératrices de ces dés : ce sont des polynômes de terme constant nul. La fonction génératrice de la somme des deux dés (supposés indépendants) serait

$$\mathrm{G_{X+Y}}(t) = \mathrm{G_X}(t)\, \mathrm{G_Y}(t) = \frac{1}{11}\big[t^2 + t^3 + \cdots + t^{12}\big] = \frac{t^2}{11}\big[1 + t + \cdots + t^{10}\big],$$

ce qui est en contradiction avec le résultat précédent.

♦ **Solution 8.6** Les lois de X et Y sont données par $\mathbf{P}\{\mathrm{X} = k\} = \mathbf{P}\{\mathrm{Y} = k\} = p\, q^{k-1}$ en notant $q := 1 - p$; la variable Z peut prendre toutes les valeurs rationnelles positives : $\mathrm{Z}(\Omega) = \mathbb{Q}_+^*$.

Pour déterminer la loi de Z, considérons $r = n/k$ un rationnel irréductible. La variable aléatoire Z prend la valeur r si et seulement si X prend une valeur ℓn et Y une valeur $k\ell$, pour un entier ℓ quelconque. D'après la formule des probabilités totales et par indépendance,

$$\begin{aligned}\mathbf{P}\{\mathrm{Z} = r\} &= \sum_{\ell=1}^{\infty} \mathbf{P}\{\mathrm{X} = \ell n\} \cdot \mathbf{P}\{\mathrm{Y} = \ell k\} = \sum_{\ell=1}^{\infty} q^{\ell n - 1}\, q^{\ell k - 1}\, p^2 \\ &= \frac{p^2}{q^2} \sum_{\ell=1}^{\infty} \left(q^{n+k}\right)^\ell = \frac{p^2}{q^2}\, \frac{q^{n+k}}{1 - q^{n+k}}.\end{aligned}$$

Pour calculer l'espérance de Z, le plus simple est de *ne pas utiliser sa loi*, mais de tirer profit de la formule de transfert. Les séries considérées étant positives, les sommes suivantes sont bien définies dans $\mathbb{R}^+$ et leur convergence est justifiée par les derniers calculs :

$$\mathbf{E}(\mathrm{Z}) = \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \frac{k}{\ell}\, \mathbf{P}\{\mathrm{X} = k, \mathrm{Y} = \ell\} = \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \frac{k}{\ell}\, q^{k+\ell-2}\, p^2.$$

Les relations $\displaystyle\sum_{k=1}^{\infty} k\, q^k = \frac{q}{p^2}$ et $\displaystyle\sum_{\ell=1}^{\infty} \frac{q^\ell}{\ell} = -\ln p$ (équation (A.5)) permettent de conclure à

$$\mathbf{E}(\mathrm{Z}) = -\frac{\ln p}{1 - p}.$$

♦ **Solution 8.7** La première propriété vient de la linéarité de l'espérance : pour toute valeur x_k telle que $\mathbf{P}\{X = x_k\} > 0$,

$$\mathbf{E}(\lambda Y + \mu Y' \,|\, X = x_k) = \lambda \mathbf{E}(Y \,\|\, X = x_k) + \mu \mathbf{E}(Y' \,\|\, X = x_k).$$

La définition du paragraphe § 126 permet alors de conclure.

Pour la seconde propriété, si l'on note $Y(\Omega) = \{y_\ell \,;\, \ell \in L\}$, alors pour tout ω on a

$$\mathbf{E}(Y \,\|\, Y)(\omega) = \begin{cases} \mathbf{E}(Y \,|\, Y = y_\ell) & \text{si } Y(\omega) = y_\ell \text{ et } \mathbf{P}\{Y = y_\ell\} > 0 \\ \text{une valeur arbitraire} & \text{si } Y(\omega) = y_\ell \text{ et } \mathbf{P}\{Y = y_\ell\} = 0. \end{cases}$$

Or, par définition,

$$\mathbf{E}(Y \,|\, Y = y_\ell) = \sum_{k \in L} y_k \, \mathbf{P}\{Y = y_k \,|\, Y = y_\ell\} = y_\ell,$$

puisque

$$\mathbf{P}\{Y = y_k \,|\, Y = y_\ell\} = \begin{cases} 1 & \text{si } y_k = y_\ell \\ 0 & \text{sinon.} \end{cases}$$

Ainsi, $\mathbf{E}(Y \,\|\, Y)$ est égale à Y sauf sur un ensemble de probabilité nulle (l'union des $\{Y = y_\ell\}$, l'indice ℓ parcourant l'ensemble des indices tels que $\mathbf{P}\{Y = y_\ell\} = 0$).

♦ **Solution 8.8** On a, pour tout $t \in [-1\,;+1]$:

$$\begin{aligned} G_S(t) = \mathbf{E}(t^{S_N}) = \mathbf{E}\big(\mathbf{E}(t^{S_N} \,\|\, N)\big) &= \sum_{n=0}^{\infty} \mathbf{P}\{N = n\}\ \mathbf{E}(t^{S_N} \,|\, N = n) \\ &= \sum_{n=0}^{\infty} \mathbf{P}\{N = n\}\ \mathbf{E}(t^{X_1 + X_2 + \cdots + X_n}) \\ &= \sum_{n=0}^{\infty} \mathbf{P}\{N = n\}\ \mathbf{E}(t^{X_1}) \cdot \mathbf{E}(t^{X_2}) \cdots \mathbf{E}(t^{X_n}) \qquad \textit{par indépendance} \\ &= \sum_{n=0}^{\infty} \mathbf{P}\{N = n\}\ G_X(t)^n = G_N\big(G_X(t)\big). \end{aligned}$$

♦ **Solution 8.9**

i) Avec la formule des probabilités totales, on trouve que $M = \begin{pmatrix} 1-p & q \\ p & 1-q \end{pmatrix}$.

ii) La matrice M se diagonalise en $M = PDP^{-1}$ où

$$P = \begin{pmatrix} q & 1 \\ p & -1 \end{pmatrix} \qquad D = \begin{pmatrix} 1 & \\ & 1-p-q \end{pmatrix} \qquad \text{et} \qquad P^{-1} = \frac{1}{p+q} \begin{pmatrix} 1 & 1 \\ p & -q \end{pmatrix}.$$

Puisque $|1 - p - q| < 1$, on en déduit que

$$\lim_{n \to \infty} M^n = P \begin{pmatrix} 1 & \\ & 0 \end{pmatrix} P^{-1} = \frac{1}{p+q} \begin{pmatrix} q & q \\ p & p \end{pmatrix}$$

et donc que

$$\lim_{k \to \infty} H_k = \frac{1}{p+q} \begin{pmatrix} q \\ p \end{pmatrix},$$

indépendamment des valeurs u_0 et v_0 (on a simplement utilisé la relation $u_0 + v_0 = 1$). Ainsi, la loi limite de $(H_k)_{k \geqslant 1}$ est une loi de Bernoulli de paramètre $q/(p+q) \approx 0,5333$.

♦ **Solution 8.10** Pour tout $k \geqslant 1$, on a $\mathbf{P}\{X = k\} = p\,q^{k-1}$. Par indépendance, on a

$$\mathbf{P}\{Z \leqslant k\} = \mathbf{P}\{X \leqslant k, Y \leqslant k\} = \mathbf{P}\{X \leqslant k\} \cdot \mathbf{P}\{Y \leqslant k\} = \left(\sum_{i=1}^{k} p\,q^{i-1} \right)^2 = (1 - q^k)^2,$$

d'où l'on déduit par différence

$$\mathbf{P}\{Z = k\} = (1 - q^k)^2 - (1 - q^{k-1})^2 = 2p\,q^{k-1} - q^{2k-2}(1 - q^2).$$

La formule $\sum_{n=1}^{\infty} n\,x^n = x/(1-x)^2$ permet alors de conclure, après quelques lignes de calcul, que

$$\mathbf{E}(Z) = \frac{3 - 2p}{p(2-p)}.$$

Chapitre 9

Interlude : le jeu de pile ou face

Dans ce chapitre, nous explorons quelques propriétés élémentaires du jeu de pile *ou* face. *Après avoir défini un espace probabilisé permettant de modéliser une succession infinie de tirages, on s'intéresse à la loi de la longueur de séquences homogènes, puis à un jeu consistant à parier sur la sortie d'une séquence particulière (PPF) avant une autre (FPP). Cela nous permet d'aborder la théorie des événements régénératifs. Enfin, nous terminons par un parallèle entre les suites de tirages à* pile *ou* face *et les probabilités continues,* via *le théorème de Borel sur les nombres normaux, parallèle qui mènera à deux conclusions :*

- *une suite de variables de Bernoulli indépendantes ou un réel de* $[0\,;1[$, *c'est (presque) pareil ;*
- *il est possible de prouver (avec une certitude absolue, et non pas presque sûrement) l'existence de nombres possédant certaines propriétés, en utilisant l'attirail des probabilités, et ce bien qu'aucune des propriétés calculées n'ait un quelconque rapport avec le hasard.*

9.1 FORMALISATION

Espace probabilisé associé à une partie infinie de *pile* ou *face* §132

Par convention, les lettres P et F représentent l'issue d'un tirage à *pile* ou *face* ; on les identifiera conventionnellement avec les valeurs numériques $P = 1$ et $F = 0$, ou avec des notions de succès (P = succès, F = échec).

On modélise l'expérience consistant à tirer une infinité de fois à *pile* ou *face* par l'espace $\Omega = \{P,F\}^{\mathbb{N}^*}$, que l'on munit d'une tribu $\mathfrak{T}$ et d'une probabilité $\mathbf{P}$ définie sur cette tribu. Un élément ω de Ω est donc une suite

$$\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots) \qquad \omega_k \in \{P,F\}.$$

Avant de déterminer quelle tribu choisir, on s'accorde d'abord sur les « événements » dont on exige, *a minima*, qu'ils soient dans $\mathfrak{T}$. Par exemple, on aimerait pouvoir parler de l'ensemble des parties telles que le troisième tirage est *pile* et le cinquième *face*.

Si n est un entier non nul quelconque, $\mathrm{I} := (i_1, i_2, \ldots, i_n)$ un n-uplet d'indices tels que $i_1 < i_2 < \cdots < i_n$ et $\mathrm{R} = \{r_{i_1}, r_{i_2}, \ldots, r_{i_n}\}$ un n-uplet de $\{\mathrm{P}, \mathrm{F}\}$, convenons d'appeler **cylindre élémentaire** la partie

$$\mathrm{C_R} = \{\omega \in \Omega \,;\, \omega_{i_1} = r_{i_1}, \ldots, \omega_{i_n} = r_{i_n}\}$$

composée des suites dont n valeurs particulières sont spécifiées par R. Les cylindres élémentaires sont des événements souhaitables, ainsi que toute union finie de tels événements. On note $\mathfrak{C}$ l'algèbre engendrée par tous les cylindres élémentaires. Ses éléments sont appelés **cylindres**, et ce sont les parties C de Ω vérifiant : il existe un entier minimal r (le **rang** de C) et une partie $\mathrm{H} \subset \{\mathrm{P}, \mathrm{F}\}^r$ tels que

$$\mathrm{C} = \{\omega \in \Omega \,;\, (\omega_1, \omega_2, \ldots, \omega_r) \in \mathrm{H}\}.$$

Exemple 9.1 L'union des deux cylindres élémentaires

$$\mathrm{C}_1 = \{\omega \,;\, \omega_1 = \mathrm{P} \text{ et } \omega_2 = \mathrm{F}\} \qquad \text{et} \qquad \mathrm{C}_2 = \{\omega \,;\, \omega_1 = \mathrm{F} \text{ et } \omega_2 = \mathrm{P}\}$$

n'est pas un cylindre élémentaire, mais c'est bien un cylindre, de rang 2. Si l'on note $\mathrm{H} = \{(\mathrm{P}, \mathrm{F}),\ (\mathrm{F}, \mathrm{P})\}$, alors $\mathrm{C}_1 \cup \mathrm{C}_2 = \{\omega \,;\, (\omega_1, \omega_2) \in \mathrm{H}\}$.

On choisit alors comme tribu des événements la tribu $\mathfrak{T}$ engendrée par $\mathfrak{C}$:

$$\mathfrak{T} := \sigma(\mathfrak{C}).$$

On munit enfin $(\Omega, \mathfrak{T})$ d'une probabilité en la définissant tout d'abord pour des cylindres élémentaire ; notons $\pi(\mathrm{R})$ le nombre d'éléments r_i égaux à « P » et $\varphi(\mathrm{R})$ le nombre de ceux égaux à « F », et posons

$$\mathbf{P}_0(\mathrm{C_R}) = p^{\pi(\mathrm{R})} \cdot (1-p)^{\varphi(\mathrm{R})}. \tag{9.1}$$

La fonction $\mathbf{P}_0$ s'étend clairement à $\mathfrak{C}$, puisque chaque cylindre est une union disjointe finie de cylindres élémentaires[1].

La classe $\mathfrak{C}$ est une algèbre (elle est stable par union finie et passage au complémentaire) et $\mathbf{P}_0$ est encore additive sur $\mathfrak{C}$. Elle est en fait σ-additive, ce qui est un peu délicat à prouver[2]. Le théorème de Carathéodory permet alors d'étendre,
↛9.1 de manière unique, l'application $\mathbf{P}$ à toute la tribu $\mathfrak{T}$ engendrée par $\mathfrak{C}$.

§ 133 **Quelques notations**

Dans toute la suite, nous utiliserons les notations suivantes, en convenant que $\mathrm{P} = 1$ et $\mathrm{F} = 0$:

- X_k est l'application k-ième coordonnée : $\mathrm{X}_k : \omega \mapsto \omega_k$. C'est une variable aléatoire. Les variables X_k sont, par construction de l'espace $(\Omega, \mathfrak{T}, \mathbf{P})$, indépendantes.
- P_n est l'ensemble des réalisations retournant *pile* au n-ième tirage :

$$\mathrm{P}_n = \{\omega \in \Omega \,;\, \omega_n = 1\} = \{\mathrm{X}_n = 1\}\,.$$

De même, on définit $\mathrm{F}_n = \mathrm{P}_n^{\mathrm{c}} = \{\mathrm{X}_n = 0\}$. On écrira parfois $\mathrm{P}_i\mathrm{F}_j\mathrm{F}_k$ pour désigner l'intersection

$$\mathrm{P}_i \cap \mathrm{F}_j \cap \mathrm{F}_k = \{\omega \in \Omega \,;\, \mathrm{X}_i(\omega) = 1,\ \mathrm{X}_j(\omega) = 0,\ \mathrm{X}_k(\omega) = 0\}\,.$$

(1). Il faut bien sûr vérifier que cette définition ne dépend pas de la manière de décomposer un cylindre en cylindres élémentaires.

(2). La démonstration est reportée en annexe, A71, page 702.

- On note $S_n = \text{Card}\{k \in \mathbb{N}^* \,;\, k \leqslant n \text{ et } \omega_k = 1\} = X_1 + X_2 + \cdots + X_n$, nombre de *pile* sortis au moment du n-ième tirage.

Remarque 9.2 Le jeu de *pile* ou *face* modélise en fait les situations typiques suivantes.

— Une expérience menant à un succès avec la probabilité p et à un échec avec la probabilité $1-p$ est répétée, de manière indépendante, une infinité (dénombrable) de fois. Cela revient à considérer une suite d'événements $(A_n)_{n\geqslant 1}$ indépendants les uns des autres, et à poser $X_n = \mathbf{1}_{A_n}$, fonction indicatrice de A_n. La variable aléatoire $S_n = X_1 + X_2 + \cdots + X_n$ compte le nombre de succès au bout de la n-ième expérience.

— Un mobile effectue une *marche au hasard* : sur une échelle de temps discrète, il fait un pas sur la droite avec la probabilité p et un pas sur la gauche avec la probabilité $1-p$. Sa position R_n est la somme de variables indépendantes : $R_n = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n$, où ε_k est une variable aléatoire prenant les valeurs 1 avec la probabilité p et -1 avec la probabilité $1-p$.

Le cas symétrique $p = 1/2$ est particulièrement intéressant. On appelle **suite de Rademacher** une suite $(\varepsilon_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, identiquement distribuées, de loi

$$\mathbf{P}\{\varepsilon_n = 1\} = \mathbf{P}\{\varepsilon_n = -1\} = \frac{1}{2}.$$

Avec une pièce équilibrée, les variables aléatoires $\varepsilon_n = 2X_n - 1$ sont des variables de Rademacher indépendantes.

9.2 LE MOTIF DANS LE TAPIS

Longueur moyenne de la deuxième séquence homogène

§ 134

Dans une suite infinie $(X_n)_{n\geqslant 1}$ de tirages à *pile* ou *face*, on demande quelle est la longueur moyenne de la première et de la deuxième séquence *homogène*. Dans le tirage suivant :

$$(X_n(\omega))_{n\in\mathbb{N}^*} = (\underbrace{P\,P\,P\,P}_{L_1}\,\overbrace{F\,F\,F\,F\,F}^{L_2}P\,P\,P\,F\,F\,F\,P\,P\cdots)$$

la première séquence homogène a pour longueur $L_1(\omega) = 4$; la deuxième séquence homogène a pour longueur $L_2(\omega) = 5$. Comme toujours, les variables aléatoires sont supposées indépendantes, identiquement distribuées selon une loi de Bernoulli

$$\mathbf{P}\{X_k = 1\} = p \qquad \mathbf{P}\{X_k = 0\} = 1-p$$

(en convenant que « 1 » désigne *pile* et que « 0 » désigne *face*.) On cherche maintenant à calculer $\mathbf{E}(L_1)$ et $\mathbf{E}(L_2)$ en fonction de p.

Plutôt que d'expliciter entièrement les lois de L_1 et de L_2 (ce qui d'ailleurs n'est pas très difficile), prenons un petit raccourci en utilisant les espérances conditionnées par les événements $F_1 = \{X_1 = 0\}$ et $P_1 = \{X_1 = 1\}$. Sous la probabilité $\mathbf{P}_{P_1} = \mathbf{P}_{\{X_1=1\}}$,

$$\begin{aligned}\mathbf{P}_{P_1}\{L_1 = n\} &= \frac{\mathbf{P}\{X_1 = 1, X_1 = X_2 = \cdots = X_n = 1, X_{n+1} = 0\}}{\mathbf{P}\{X_1 = 1\}} \\ &= \mathbf{P}\{X_2 = \cdots = X_n = 1, X_{n+1} = 0\}, \qquad \textit{par indépendance} \\ &= p^{n-1}(1-p),\end{aligned}$$

ce qui montre que L_1 suit une loi géométrique de paramètre $1-p$, et a pour espérance

$$\mathbf{E}_{P_1}(L_1) = \frac{1}{1-p}.$$

De même, sous la probabilité $\mathbf{P}_{F_1} = \mathbf{P}_{\{X_1=0\}}$, on a

$$\mathbf{P}_{F_1}\{L_1 = n\} = \mathbf{P}\{X_2 = \cdots = X_n = 0, X_{n+1} = 1\} = (1-p)^{n-1}p$$

ce qui montre que L_1 suit une loi géométrique de paramètre p, et a pour espérance

$$\mathbf{E}_{F_1}(L_1) = \frac{1}{p}.$$

La formule des espérances totales (théorème 8.29) montre que

$$\mathbf{E}(L_1) = \mathbf{E}_{F_1}(L_1)\cdot\mathbf{P}(F_1) + \mathbf{E}_{P_1}(L_1)\cdot\mathbf{P}(P_1) = \frac{(1-p)}{p} + \frac{p}{1-p} = 2 + \frac{(2p-1)^2}{p(1-p)}.$$

Cette quantité est toujours supérieure ou égale à 2, et ne vaut 2 que si $p = 1/2$.

Un calcul semblable montre que :

— sous la probabilité conditionnée par F_1, la variable aléatoire L_2 suit une loi géométrique de paramètre $1-p$;

— sous la probabilité conditionnée par P_1, la variable aléatoire L_2 suit une loi géométrique de paramètre p.

Ainsi,

$$\mathbf{E}(L_2) = \mathbf{E}_{F_1}(L_2)\cdot\mathbf{P}(F_1) + \mathbf{E}_{P_1}(L_2)\cdot\mathbf{P}(P_1) = \frac{1-p}{1-p} + \frac{p}{p} = 2.$$

Cette quantité est indépendante de p.

Remarque 9.3 Heuristiquement, si $p > 1/2$, c'est-à-dire si le *pile* est plus fréquent que le *face*, la première série est souvent formée de *pile* (et donc longue) et rarement composée de *face* (et donc courte). Au contraire, la deuxième série est souvent formée de *face*, et donc courte ; elle est rarement composée de *pile*, et donc longue. Quantitativement, il y a une compensation exacte entre les phénomènes.

§ 135 **Vaut-il mieux parier FPP ou PPF ?**

Rosencrantz décide de compliquer un peu les règles de leur partie de *pile* ou *face*, supposée être jouée avec une pièce équilibrée. Dans cette nouvelle règle, Rosencrantz gagne la partie si la séquence « FPP » sort avant la séquence « PPF », et Guildenstern gagne si la séquence « PPF » sort en premier. Guildenstern s'apprête à accepter de jouer mais — vu que c'est Rosencrantz qui a proposé cette règle — il décide d'étudier sérieusement la question :

le jeu est-il équilibré ?

« Bien sûr, que le jeu est équilibré, s'empresse d'affirmer Rosencrantz, et la preuve en est que le temps moyen pour obtenir, pour la première fois, la séquence FPP est de 8, et que le temps moyen pour obtenir, pour la première fois, la séquence PPF est également de 8. Que te faut-il de plus ? » Guildenstern effectue quelques calculs, qui confirment ces valeurs[3].

Cependant, l'empressement de Rosencrantz lui paraît suspect. Guildenstern décide de noter g_n la probabilité de l'événement

$$G_n : \begin{cases} \text{la séquence « PPF » sort au } n\text{-ième tirage,} \\ \textit{et} \text{ la séquence « FPP » n'est jamais sortie avant.} \end{cases}$$

(3). Le raisonnement de Rosencrantz est expliqué aux paragraphes suivants.

Bien entendu, $g_1 = g_2 = 0$ et $g_3 = 1/2^3$, puisque seul le tirage

P	P	F	*	*	···

(G_3)

amène un gain de Guildenstern au bout du troisième tirage. Pour le calcul de g_4, on sait que les tirages gagnants sont nécessairement de la forme

*	P	P	F	*	···

mais si le premier tirage est un *face*, alors la séquence « FPP » sera sortie au troisième tirage. Ainsi, g_4 est la probabilité de

P	P	P	F	*	···

(G_4)

et vaut $g_4 = 1/2^4 = 1/16$. De la même façon, $g_5 = 1/2^5 = 1/32$, probabilité des seuls tirages de la forme

P	P	P	P	F	*	···

(G_5)

Au final, on trouve que

$$\begin{cases} g_1 = g_2 = 0 \\ g_n = 1/2^n \quad (n \geqslant 3). \end{cases}$$

Les événements G_n étant disjoints, la probabilité de gain de Guildenstern est

$$\mathbf{P}(G) = \mathbf{P}\left(\bigcup_{n=3}^{\infty} G_n\right) = \sum_{n=3}^{\infty} \frac{1}{2^n} = \frac{1}{4}.$$

La probabilité de gain de Rosencrantz est donc de 3/4... si du moins l'on montre que, avec une probabilité 1, un des deux joueurs sera réputé vainqueur. Pour cela, notons que les événements $A_n = P_{n-2} \cap P_{n-1} \cap F_n$, de probabilité 1/8 chacun, ne sont pas indépendants ; en revanche les événements A_{3n} le sont ! L'événement « il n'y a pas de gagnant » est inclus dans l'intersection des complémentaires des A_{3n} :

$$\{\text{il n'y a pas de gagnant}\} \subset \bigcap_{n=0}^{N} A_{3n}^c$$

donc $\mathbf{P}\{\text{il n'y a pas de gagnant}\} \leqslant (7/8)^{3N}$ pour tout N. *Conclusion* : Guildenstern gagne avec une probabilité 1/4, Rosencrantz avec une probabilité 3/4 (et la partie ne finit jamais avec une probabilité nulle).

Occurrence de motifs, événements régénératifs §136

Il est temps d'établir la loi du temps d'apparition des motifs « PPF » et « FPP », afin de justifier l'affirmation de Rosencrantz.

Mais d'abord, un petit détour théorique. On s'intéresse à un événement $\mathscr{E}$ pouvant se produire à un ou plusieurs tirages de la suite $(X_n)_{n\geqslant 1}$, comme par exemple l'occurrence du motif « PPF ». Ce n'est pas un événement au sens usuel du terme, mais plutôt une propriété que peut avoir, ou non, une chaîne finie $X_1, \ldots, X_n$ (ici : se terminer par la séquence voulue). Dans la chaîne ci-dessous, l'événement $\mathscr{E}$, marqué d'une étoile (*), se produit aux tirages numéros 5 et 9 :

F_1	P_2	P_3	P_4	F_5	F_6	P_7	P_8	F_9	···
				*				*	

Définition 9.4 Un événement $\mathscr{E}$, pouvant se produire à un tirage quelconque de la suite $(X_n)_{n\geqslant 0}$, est dit **régénératif** si les distances entre les occurrences successives de $\mathscr{E}$ sont des variables aléatoires mutuellement indépendantes et de même loi.

Cela signifie que, lorsque $\mathscr{E}$ se produit, les pendules sont remises à zéro et que l'on oublie complètement le passé. Si l'événement $\mathscr{E}$ se produit au tirage n, alors la probabilité (conditionnelle) qu'il se produise au tirage $n+k$ est égale à la probabilité (absolue) qu'il se produise au tirage k.

Par exemple, l'événement « le tirage donne *face* » est régénératif. De même, « le tirage donne PPF » est régénératif : lorsque le dernier *face* a eu lieu, il faut attendre de nouveau qu'au moins trois nouveaux tirages soient effectués, et toute se passe comme si le passé était oublié.

En revanche, l'événement « le tirage donne PPP » n'est *pas* régénératif. Si l'on note a_n la probabilité qu'il se produise au tirage n, alors $a_1 = a_2 = 0$ tandis que $a_3 = 1/8$. Maintenant, supposons qu'il se soit produit au n-ième tirage. Alors la probabilité qu'il se produise au tirage $n+1$ est égale à $1/2$, puisqu'il suffit d'un nouveau *pile* pour qu'il ait lieu :

$\cdots$	P_{n-2}	P_{n-1}	P_n	P_{n+1}	$\cdots$
			*	*	

Dans toute la suite, on adopte les notations suivantes : on note a_n la probabilité que l'événement $\mathscr{E}$ ait lieu au n-ième tirage — et on dira plutôt « **au site n** » pour éviter les confusions — et b_n la probabilité qu'il ait lieu *pour la première fois* au site n.

$$\begin{cases} a_n = \mathbf{P}\{\mathscr{E} \text{ a lieu au site } n\} \\ b_n = \mathbf{P}\{\mathscr{E} \text{ a lieu pour la première fois au site } n\}. \end{cases} \tag{9.2}$$

Pour des raisons qui apparaîtront plus tard, on pose $a_0 = 1$ et $b_0 = 0$ (sans qu'il y ait d'interprétation spécifique à ces valeurs). Enfin, on note

$$\mathrm{A}(x) = \sum_{n=0}^{\infty} a_n x^n \qquad \mathrm{B}(x) = \sum_{n=0}^{\infty} b_n x^n \qquad b = \sum_{n=0}^{\infty} b_n = \mathrm{B}(1). \tag{9.3}$$

On remarquera que A n'est pas la fonction génératrice d'une loi de probabilité, les scalaires a_n ne définissant pas la loi d'une variable aléatoire (la somme de ces scalaires peut valoir n'importe quoi, et est souvent infinie). Cependant, puisque ses coefficients sont tous positifs et majorés par 1, le rayon de la série $\sum a_n x^n$ est supérieur ou égal à 1, et A est définie sur $[\,0\,;\,1\,[$ au moins.

On note T la variable aléatoire exprimant l'instant de la première occurrence de l'événement $\mathscr{E}$ (cet instant pouvant être $+\infty$ si l'événement $\mathscr{E}$ n'a pas lieu). Grâce à notre hypothèse sur a_0, T représente également la distance entre deux occurrences successives de l'événement $\mathscr{E}$ (en sous-entendant que $\mathscr{E}$ a lieu au site $n = 0$). La loi de T est donc $\mathbf{P}\{\mathrm{T} = n\} = b_n$.

La quantité $1-b$ est la probabilité que $\mathscr{E}$ n'ait jamais lieu : $1-b = \mathbf{P}\{\mathrm{T} = +\infty\}$. Si $b = 1$, on dit que l'événement est **récurrent**, si $b < 1$, on dit qu'il est **transient** (l'explication en sera donnée à la fin du paragraphe).

■ **Proposition 9.5** *Si $\mathscr{E}$ est régénératif, les suites $a = (a_n)_{n\geqslant 0}$ et $b = (b_n)_{n\geqslant 0}$ sont liées par la relation de convolution $a_n = (a * b)_n$ pour tout $n \geqslant 1$, c'est-à-dire*

$$\forall n \in \mathbb{N}^* \qquad a_n = \sum_{k=0}^{n} b_k\, a_{n-k}. \tag{9.4}$$

DÉMONSTRATION : Soit $n \geqslant 1$. La probabilité que l'événement $\mathscr{E}$, supposé régénératif, ait lieu pour la première fois à un instant $k < n$ et de nouveau à l'instant n est $b_k\, a_{n-k}$. La probabilité qu'il ait lieu pour la première fois à l'instant n est b_n, que l'on réécrit plutôt

$b_n\, a_0$. Ces différents cas sont mutuellement exclusifs et épuisent les possibilités que $\mathscr{E}$ se produise à l'instant n ; la formule des probabilités totales montre que

$$a_n = b_1\, a_{n-1} + b_2\, a_{n-2} + \cdots + b_n\, a_0,$$

ce qui est bien l'équation proposée (en tenant compte du fait que $b_0 = 0$). ∎

■ **Théorème 9.6** *Pour tout* $x \in [0\,;1[$,

$$\mathrm{A}(x) = \frac{1}{1-\mathrm{B}(x)} \qquad \textit{et} \qquad \mathrm{B}(x) = 1 - \frac{1}{\mathrm{A}(x)}. \tag{9.5}$$

DÉMONSTRATION : Multiplions chacune des équations (9.4) par x^n et sommons pour n variant entre 1 et l'infini. À gauche, on trouve $\mathrm{A}(x) - a_0$, c'est-à-dire $\mathrm{A}(x) - 1$. À droite, on reconnaît le produit de Cauchy $\mathrm{A}(x)\,\mathrm{B}(x)$. Ainsi,

$$\forall x \in [0\,;1[\qquad \mathrm{A}(x) - 1 = \mathrm{A}(x)\,\mathrm{B}(x),$$

ce qui mène à la formule proposée. ∎

Il est parfois plus facile de calculer $\mathrm{A}(x)$, d'où l'on déduit $\mathrm{B}(x)$ par la relation précédente ; une fois que l'on a B, on peut en déduire le temps moyen d'attente par

$$\mathbf{E}(\mathrm{T}) = \sum_{n=0}^{\infty} n\,\mathbf{P}\{\mathrm{T} = n\} = \sum_{n=1}^{\infty} n\, b_n = \mathrm{B}'(1).$$

(Cette relation, valable dans $\mathbb{R} \cup \{+\infty\}$, est justifiée par la positivité des termes ; de manière générale, B est dérivable en 1 si et seulement si $\sum n\, b_n$ converge.)

Remarque 9.7 (Un théorème sur le temps d'attente) Comme on le verra dans les deux paragraphes suivants, obtenir $\mathbf{E}(\mathrm{T})$ par le calcul de A, puis de B, est parfois fastidieux. Il existe un théorème remarquable, dont la version la plus générale, présentée ici, est due à P. ERDŐS, W. FELLER et H. POLLARD, qui relie directement $\mathbf{E}(\mathrm{T})$ à la suite $(a_n)_{n\geqslant 0}$.

■ **Théorème 9.8 (Erdős, Feller, Pollard, 1949)** *On suppose que* $\mathscr{E}$ *est récurrent et que, de plus, il n'est pas périodique, c'est-à-dire qu'il n'existe pas d'entier* $q \geqslant 2$ *tel que* $\mathscr{E}$ *ne puisse se réaliser* que *aux rangs* q, $2q$, $3q$ *etc., ce qui revient à dire que* $\mathrm{PGCD}\{n \geqslant 1\,;\, a_n \neq 0\} = 1$. *Dans ce cas, la suite* $(a_n)_{n\geqslant 0}$ *converge et*

$$\lim_{n\to\infty} a_n = \frac{1}{\mathbf{E}(\mathrm{T})}.$$

La preuve de ce théorème est assez technique et ne fait appel qu'à des calculs sur les séries entières, sans contenu probabiliste. Le lecteur intéressé en trouvera la démonstration dans le lumineux ouvrage d'un de auteurs précités [33, XIII.11].

Remarque 9.9 (Événements transients) Supposons que $b < 1$; alors la formule (9.5) montre que $\mathrm{A}(1) = \sum_{n=0}^{\infty} a_n < +\infty$, ce qui implique à son tour que l'événement $\mathscr{E}$ n'a lieu, presque sûrement, qu'un nombre fini de fois (c'est le lemme de Borel-Cantelli, dans le sens facile).

Temps d'attente d'un des motifs PPF et FPP

§ 137

Revenons à nos moutons. On s'intéresse à l'événement $\mathscr{E}$, occurrence du motif « PPF ». Comme déjà dit, cet événement est naturellement régénératif, et la probabilité qu'il se produise au site n est $a_n = 1/2^3 = 1/8$ si $n \geqslant 3$, tandis que $a_1 = a_2 = 0$. On en déduit la fonction A sur $[0\,;1[$:

$$\mathrm{A}(x) = 1 + \sum_{n=3}^{\infty} \frac{x^n}{2^3} = 1 + \frac{1}{8}\frac{x^3}{1-x} = \frac{x^3 - 8x + 8}{8(1-x)},$$

puis, par la relation (9.5),

$$\mathrm{B}(x) = \frac{x^3}{x^3 - 8x + 8}.$$

On remarque d'abord que $\mathrm{B}(1) = 1$: *l'événement $\mathscr{E}$ est récurrent.* Par simple dérivation,

$$\mathrm{B}'(x) = \frac{8x^2(3-2x)}{(x^3-8x+8)^2}$$

et donc

$$\mathbf{E}(\mathrm{T}) = \mathrm{B}'(1) = 8.$$

Le même raisonnement peut être effectué pour le motif « FPP », lui aussi régénératif, et caractérisé par les mêmes coefficients a_n et b_n.

■ **Proposition 9.10** *Le temps d'attente moyen de chacun des deux motifs* FPP *et* PPF *est le même et vaut* 8.

Remarque 9.11 Le lecteur essaiera de se convaincre en quoi cela n'est *pas* contradictoire avec le raisonnement de Guildenstern, qui assurait la victoire avec une probabilité de 3/4 à l'apparition du motif « FPP » avant celle de « PPF ». Un tableau de ce qui se passe pendant les quatre premiers tirages peut être utile. Pour chacune des $2^4 = 16$ possibilités, on note la position éventuelle d'apparition des chacun des motifs gagnants pour Rosencrantz (FPP) et Guildenstern (PPF).

	R	G		R	G		R	G		R	G
PPPP			PFPP	4		FPPP	3		FFPP	4	
PPPF		4	PFPF			FPPF	3	4	FFPF		
PPFP		3	PFFP			FPFP			FFFP		
PPFF		3	PFFF			FPFF			FFFP		

La loi du temps d'attente du motif choisi est la même pour chacun des joueurs (sur 16 tirages, il apparaît deux fois en position 3, et deux fois en position 4) ; Rosencrantz est déclaré vainqueur 4 fois, tandis que Guildenstern ne l'est que 3 fois.

§ 138 **Temps d'attente du motif PPP**

Les deux motifs précédemment étudiés sont « naturellement » régénératifs. Le cas du motif « PPP » est un peu différent puisque, comme déjà expliqué, il est non régénératif. L'idée est de ne considérer que des séquences *non-recouvrantes* : on dira ainsi que l'événement $\mathscr{E}'$ se produit, dans la séquence ci-dessous, aux sites 5, 10 et 13, mais pas aux sites 6, 11, 12 ou 14 :

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	···
P	F	P	P	P	P	F	P	P	P	P	P	P	P	0	P	0	···
				*					*			*					

Avec cette définition, $\mathscr{E}'$ est régénératif. On notera $\mathscr{E}$ l'événement non régénératif lié à la simple occurrence du motif « PPP », sans restriction de recouvrement. Le point essentiel de cette démarche est que la première séquence sans recouvrement est également la première séquence tout court ; ainsi, le temps d'attente de $\mathscr{E}$ est égal à celui de $\mathscr{E}'$ — que l'on peut calculer avec les outils développés dans le paragraphe précédent.

Soit $n \geqslant 3$. On considère l'événement A_n : « les tirages $n-2, n-1, n$ sont des *pile* », qui est de probabilité 1/8. Dans le cadre de cet événement, $\mathscr{E}'$ a nécessairement lieu à un, et un seul, des emplacements $n-2$, $n-1$ et n :

— il peut avoir lieu en $n-2$, et il est suivi de « PP » ;
— il peut avoir lieu en $n-1$, et il est suivi de « P » ;
— il peut avoir lieu en n.

La formule des probabilités totales permet d'écrire, en tenant compte de l'indépendance des tirages et de la propriété d'absence de mémoire liée au caractère régénératif de $\mathscr{E}'$:

$$\tfrac{1}{8} = a_n + \tfrac{1}{2}\,a_{n-1} + \tfrac{1}{4}\,a_{n-2}$$

Multiplions cette équation par x^n et sommons de 3 à l'infini :

$$\frac{x^3}{8(1-x)} = \big[\mathrm{A}(x)-1\big]\left(1+\frac{x}{2}+\frac{x^2}{4}\right) = \big[\mathrm{A}(x)-1\big]\frac{1-x^3/8}{1-x/2},$$

ce qui mène à

$$\mathrm{A}(x) = \frac{1-x+x^4/16}{(1-x)(1-x^3/8)} \qquad \text{et} \qquad \mathrm{B}(x) = -\frac{x^3}{x^3+2x^2+4x-8}.$$

On en déduit que $\mathrm{B}(1) = 1$ et donc que $\mathscr{E}'$ est récurrent. Enfin, le temps moyen d'attente de $\mathscr{E}'$ est $\mathrm{B}'(1) = 14$, et celui de $\mathscr{E}$ également.

Conclusion : *le temps moyen d'attente de la première séquence de trois* pile *dans une succession infinie de tirages à* pile *ou* face *est de* 14.

Cette technique se généralise à l'attente d'une séquence de longueur K quelconque, ou d'autres motifs qui ne sont pas « naturellement » régénératifs. ⇗9.3 ⇗9.2

9.3 THÉORÈMES LIMITES

Une certitude : la moyenne § 139

Les quelques théorèmes déjà rencontrés dans les chapitres précédents permettent déjà d'énoncer un résultat particulièrement important, la *loi des grands nombres*, sous une forme que l'on appelle *faible*. On rappelle que $\mathrm{M}_n = (\mathrm{X}_1+\mathrm{X}_2+\cdots+\mathrm{X}_n)/n$ et on note p la probabilité que la pièce utilisée sorte *pile*, de sorte que X_k suit une loi de Bernoulli $\mathscr{B}(p)$ pour tout k.

■ **Théorème 9.12 (Loi faible des grands nombres)**
La suite $(\mathrm{M}_n)_{n\geqslant 1}$ *des moyennes converge vers* p, *au sens où, pour tout* $\varepsilon > 0$,

$$\mathbf{P}\big\{\,|\mathrm{M}_n - p| > \varepsilon\,\big\} \xrightarrow[n\to\infty]{} 0.$$

Avant de démontrer ce théorème, mettons tout de suite en garde contre des interprétations fausses de cet énoncé. Par exemple, il ne dit pas que pour une réalisation ω donnée du hasard, la suite de terme général $\mathrm{M}_n(\omega)$ va converger vers p (c'est faux), ni même que ce résultat sera vrai pour presque tout ω (c'est vrai, mais ce n'est pas ce que dit le théorème présent).

DÉMONSTRATION : Les variables X_k étant indépendantes, la variance de $\mathrm{X}_1+\mathrm{X}_2+\cdots+\mathrm{X}_n$ est la somme de leurs variances, soit $np(1-p)$, et la variance de M_n est donc $p(1-p)/n$. Par ailleurs, l'espérance de M_n est égale à p par linéarité. L'inégalité de Bienaymé-Tchebychev s'écrit

$$\mathbf{P}\,\{|\mathrm{M}_n - p| > \varepsilon\} \leqslant \frac{p(1-p)}{n\varepsilon} \xrightarrow[n\to\infty]{} 0. \qquad \blacksquare$$

Démonstration très simple, comme on le voit ! En contrepartie, le contenu du théorème n'est pas vraiment satisfaisant. La probabilité des ensembles $\big\{\,|\mathrm{M}_n - p| > \varepsilon\big\}$ tend certes vers 0, mais rien n'est dit sur ces ensembles eux-mêmes — forment-ils par exemple une suite décroissante ? Et comment libérer ε, pour obtenir la mesure de l'événement « la suite $(\mathrm{M}_n)_{n\geqslant 1}$ tend vers p » ?

La réponse à cette dernière question se trouve dans la version suivante de la *loi forte des grands nombres.*

■ Théorème 9.13 (Loi forte des grands nombres)
L'ensemble $\{M_n \to p\}$ *des* ω *tels que la suite* $\big(M_n(\omega)\big)_{n\geqslant 1}$ *converge vers* p *est un événement de* $\mathfrak{T}$*, et sa probabilité vaut* 1 *:*

$$\mathbf{P}\left\{\lim_{n\to\infty} M_n \to p\right\} = 1.$$

DÉMONSTRATION : Le fait que l'ensemble $A = \left\{\lim_{n\to\infty} M_n = p\right\}$ soit un événement de $\mathfrak{T}$ est traité dans l'exercice 9.1. Nous conseillons au lecteur de le résoudre maintenant si ce n'est déjà fait.

On pose $Y_n := X_n - p$ pour tout entier n, afin de travailler avec des variables centrées ; on note $T_n := Y_1 + Y_2 + \cdots + Y_n$ les sommes partielles et $Z_n := T_n/n = M_n - p$ les moyennes successives des Y_k. On doit donc montrer que $A = \{Z_n \to 0\}$ est de probabilité 1.

En suivant une idée de Cantelli, notons $\sigma^2 := \mathbf{E}(Y_1^2)$ et $\xi^4 := \mathbf{E}(Y_1^4)$. On développe T_n^4 et on obtient par linéarité

$$\mathbf{E}(T_n^4) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\ell=1}^{n}\mathbf{E}(Y_i\,Y_j\,Y_k\,Y_\ell).$$

On doit maintenant examiner les différents termes, selon que les indices sont égaux ou non. Si un indice diffère des trois autres, par exemple l'indice i, on a

$$\mathbf{E}(Y_i\,Y_j\,Y_k\,Y_\ell) = \mathbf{E}(Y_i)\,\mathbf{E}(Y_j\,Y_k\,Y_\ell) = 0$$

par indépendance et parce que Y_i est centrée. Ainsi, la plupart des termes disparaissent, et ne restent que deux types de termes : ceux de la forme $\mathbf{E}(Y_i^4) = \xi^4$, au nombre de n, et ceux de la forme $\mathbf{E}(Y_i^2\,Y_j^2) = \mathbf{E}(Y_i^2)\,\mathbf{E}(Y_j)^2 = \sigma^4$, au nombre de $3n(n-1)$ (3 possibilités de regrouper les quatre indices en deux paires, et $n(n-1)$ valeurs possibles pour les deux indices supposés distincts). Ainsi,

$$\mathbf{E}(T_n^4) = n\,\xi^4 + 3n(n-1)\,\sigma^4$$

ce qui prouve que

$$\mathbf{E}(Z_n^4) = \frac{\mathbf{E}(T_n^4)}{n^4} = O\left(\frac{1}{n^2}\right).$$

Notamment : *la série* $\sum \mathbf{E}(Z_n^4)$ *est convergente.*

Soit $\varepsilon > 0$. D'après l'inégalité de Markov (exercice 7.6),

$$\mathbf{P}\{\,|Z_n| > \varepsilon\} = \mathbf{P}\{\,|Z_n|^4 > \varepsilon^4\} \leqslant \frac{\mathbf{E}(Z_n^4)}{\varepsilon^4},$$

donc la série $\sum \mathbf{P}\{\,|Z_n| > \varepsilon\}$ est également convergente. D'après le premier lemme de Borel-Cantelli, l'événement $\{\,|Z_n| > \varepsilon\ ;\ \text{i.s.}\}$ est de probabilité nulle.

Il reste à conclure en montrant l'équivalence entre les assertions « $\mathbf{P}\{\,|Z_n| > \varepsilon\ ;\ \text{i.s.}\} = 0$ pour tout $\varepsilon > 0$ » et « $\mathbf{P}\{Z_n \to 0\} = 1$ ». Le plus simple est de décrire le complémentaire de $\{Z_n \to 0\}$; en effet, dire que $Z_n(\omega)$ ne converge pas vers 0 revient à dire qu'il existe un réel $\varepsilon > 0$ et une infinité d'indices n tels que $|Z_n(\omega)| > \varepsilon$:

$$\{Z_n \to 0\}^c = \bigcup_{\varepsilon>0}\{\,|Z_n| > \varepsilon\ ;\ \text{i.s.}\}.$$

On peut d'ailleurs limiter cette union aux ε de la forme $1/2^m$ pour m entier :

$$\{Z_n \to 0\}^c = \bigcup_{m=0}^{\infty}\{\,|Z_n| > 2^{-m}\ ;\ \text{i.s.}\}.$$

Or, dans le membre de droite, tous les ensembles sont de probabilité nulle ; leur union dénombrable est encore de probabilité nulle, donc $\{Z_n \to 0\}$ est de probabilité 1. ∎

Retour à l'équilibre et marches aléatoires § 140

La partie de *pile* ou *face* sert à modéliser une marche aléatoire (équilibrée ou non). À chaque tirage, un marcheur[4] effectue un pas sur la droite si la pièce donne *pile*, et un pas sur la gauche si elle donne *face*. La quantité

$$\varepsilon_n = 2X_n - 1$$

est le déplacement effectué juste après le tirage n, et la position du marcheur peut être vue comme un élément de la droite discrète $\mathbb{Z}$, égal à

$$T_n = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n.$$

On s'intéresse aux retours possibles du marcheur à l'origine[5]. On dira que $\mathscr{E}$ se produit au tirage n si $T_n = 0$. L'événement $\mathscr{E}$ est alors régénératif : une fois le marcheur à l'origine, la loi de ses trajets futurs étant indépendante du passé, elle est la même qu'au début de la marche.

La probabilité de revenir à l'équilibre à un tirage de rang impair est nulle ; dans le cas pair,

$$a_{2n} = \mathbf{P}\{T_{2n} = 0\} = \binom{2n}{n} p^n q^n \qquad (q = 1 - p).$$

Le lecteur féru de développements en série entière aura peut être reconnu le développement de

$$\frac{1}{\sqrt{1+t}} = \sum_{n=0}^{\infty} (-1)^n \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n\, n!}\, t^n = \sum_{n=0}^{\infty} \frac{(-1)^n}{4^n} \binom{2n}{n} t^n,$$

ce qui permet d'écrire[6]

$$A(x) = \frac{1}{\sqrt{1 - 4pq\,x^2}}.$$

En suivant les techniques introduites précédemment, on obtient

$$B(x) = 1 - \sqrt{1 - 4pq\,x^2}.$$

Cela permet, par une nouvelle application de la formule du binôme généralisé avec $a = 1/2$, de déterminer si on le désire tous les coefficients b_n. On tire également de cette formule la valeur de b : en notant $q := 1 - p$, on a

$$\begin{aligned} b = B(1) &= 1 - \sqrt{1 - 4p(1-p)} = 1 - \sqrt{(p+q)^2 - 4pq} \\ &= 1 - \sqrt{(p-q)^2} = 1 - |p - q|\,. \end{aligned}$$

---

(4). Dont les mauvaises langues disent généralement qu'il est alcoolique.

(5). T_n peut être également vu comme la gain total d'un des joueurs qui parie sur *pile* à chaque tirage, et qui gagne ou perd un euro à chaque fois. On parlera alors de *retour à l'équilibre*.

(6). On peut procéder de manière plus méthodique en utilisant la formule du binôme généralisé (A54). On commence par transformer l'expression des coefficients en

$$a_{2n} = \binom{-\frac{1}{2}}{n} (-4pq)^n,$$

ce qui permet d'écrire

$$A(x) = \sum_{n=0}^{\infty} \binom{-\frac{1}{2}}{n} (-4pq\,x^2)^n.$$

On peut alors utiliser la formule du binôme généralisé avec $a = -1/2$ et obtenir le résultat annoncé.

Celle-ci ne vaut 1 que si $p = q$, c'est-à-dire pour $p = 1/2$ seulement. On a donc montré :

■ **Théorème 9.14 (Retour à l'équilibre)** *Le retour à l'équilibre est un événement récurrent si et seulement si le jeu est équilibré ($p = 1/2$).*

Remarque 9.15 C'est dans le cas équilibré, et dans ce cas seulement, qu'un retour à l'équilibre est certain ; et même ainsi, le temps moyen d'attente avant le premier retour à l'équilibre est infini : $\mathbf{E}(\mathrm{T}) = \mathrm{B}'(1) = +\infty$. Nous verrons, au chapitre 33, que le signe de T_n a une grande tendance à la stabilité, ce qui explique qu'il faille attendre « longtemps » avant le premier retour à l'équilibre.

La différence entre événement transient et récurrent est encore plus nette à travers le théorème suivant :

■ **Théorème 9.16 (Récurrence du retour à l'équilibre)** *On considère une marche aléatoire telle que la probabilité de faire un pas à droite est p et celle de faire un pas à gauche est $q = 1 - p$.*

i) Si $p = 1/2$, alors il y a presque sûrement une infinité de retours à l'équilibre.

ii) Si $p \neq 1/2$, alors presque sûrement, il n'y aura qu'un nombre fini de retours à l'équilibre.

DÉMONSTRATION : Le cas déséquilibré a déjà été traité (différemment) à la remarque 5.19 page 118. On peut cependant le traiter de manière élémentaire. La quantité b est la probabilité que l'événement $\mathscr{E}$ (« il y a retour à l'équilibre ») ait lieu à un tirage ou à un autre. Sous l'hypothèse qu'il a lieu, la probabilité qu'il ait lieu une seconde fois est encore b, c'est-à-dire que la probabilité qu'il ait lieu au moins deux fois est b^2. De même, la probabilité qu'il ait lieu au moins n fois est b^n, et la probabilité qu'il ait lieu une infinité de fois est nulle.

Considérons le cas équilibré. La probabilité que l'événement $\mathscr{E}$ ait lieu est 1. Ainsi, en notant Ω_1 l'événement « il y a retour à l'équilibre », on a $\mathbf{P}(\Omega_1) = 1$. Maintenant, puisque $\mathscr{E}$ est régénératif, la probabilité conditionnelle que $\mathscr{E}$ ait lieu une deuxième fois sachant qu'il a eu lieu une première fois est encore 1 : on note de même $\Omega_2 \subset \Omega_1$ l'événement « il y un second retour à l'équilibre ». La relation $\mathbf{P}(\Omega_2 \,|\, \Omega_1) = 1$ entraîne $\mathbf{P}(\Omega_2) = \mathbf{P}(\Omega_2 \,|\, \Omega_1)\,\mathbf{P}(\Omega_1) = 1$.

On construit ainsi de proche en proche une suite décroissante Ω_n d'événements, tous de probabilité 1, Ω_n étant l'événement « il y a n retours à l'équilibre ». Il y a alors une infinité de retours à l'équilibre sur l'intersection $\mathrm{A} = \bigcap \Omega_n$ de tous ces événements, qui est encore de probabilité 1. ∎

Remarque 9.17 La démonstration précédente s'applique à tout événement régénératif. S'il est récurrent, il va se produire une infinité de fois avec une probabilité 1 (ce qui explique le choix du nom « récurrent »). S'il est transient, il ne se produira presque sûrement qu'un nombre fini de fois.

Remarque 9.18 Au paragraphe § 308, page 421, on présente d'autres propriétés des marches aléatoires équilibrées et de l'apparition de motifs, en utilisant des techniques différentes (essentiellement les lemmes de Borel-Cantelli et la *loi du tout ou rien de Kolmogorov*).

9.4 LES NOMBRES NORMAUX EXISTENT !

Le jeu de pile *ou* face, *dont le principe est si simple, possède un très grand caractère de généralité et conduit, lorsqu'on l'étudie en détail, aux mathématiques les plus élevées.*

Émile Borel, Calcul des probabilités, *1958.*

Une suite de variables de Bernoulli indépendantes est un réel de $[0;1[$ (et vice-versa) §141

Non seulement l'espace $\{P,F\}^{\mathbb{N}^*}$ est-il non dénombrable, mais il existe de plus une manière naturelle de le mettre en correspondance avec un ensemble « continu », à savoir $[0;1[$. À chaque réel x de $[0;1[$, on peut associer son développement dyadique propre bien connu :

$$x = \sum_{n=1}^{\infty} \frac{\mathscr{X}_n(x)}{2^n},$$

où l'on a posé

$$\mathscr{X}_n(x) = \lfloor 2^n x \rfloor - 2 \lfloor 2^{n-1} x \rfloor . \qquad (*)$$

La suite des fonctions $\mathscr{X}_n$ permet donc d'associer, à chaque réel x, une unique suite à valeurs dans $\{0,1\}$. Toutes les suites de $\{0,1\}^{\mathbb{N}^*}$ ne sont pas représentées : on sait que certains nombres admettent deux développements dyadiques, par exemple

$$\frac{1}{2} = \overline{0,100000\cdots} = \overline{0,01111111\cdots}$$

mais le second développement est le développement *impropre*. Or, si l'on injecte $x = 1/2$ dans la formule $(*)$, on obtient le développement propre $0,100000\ldots$ et non le développement impropre. En revanche, toutes les suites *propres* à valeurs dans $\{0,1\}$, c'est-à-dire les suites qui ne sont pas stationnaires en 1, sont images d'un réel, et plus précisément : si $(\omega_n)_{n\geqslant 1}$ est une suite propre, alors c'est l'image de

$$x = \sum_{n=1}^{\infty} \frac{\omega_n}{2^n}.$$

Ainsi, on a une correspondance bijective entre $[0;1[$ est l'ensemble Ω' des suites propres (non stationnaires en 1)

$$\varphi\colon\ [0;1[\longrightarrow \Omega' \qquad \text{et} \qquad \varphi^{-1}\colon\ \Omega' \longrightarrow [0;1[$$
$$x \longmapsto \big(\mathscr{X}_n(x)\big)_{n\in\mathbb{N}} \qquad (\omega_n)_{n\geqslant 1} \longmapsto \sum_{n=1}^{\infty} \frac{\omega_n}{2^n}.$$

Il est aisé de voir que l'ensemble $N := \Omega - \Omega'$ des suite impropres est dénombrable. Notons B_k l'ensemble des suites stationnaires en 1 à partir de l'indice k : c'est un ensemble fini de cardinal 2^{k-1}. L'ensemble N est l'union (dénombrable) des B_k pour k parcourant $\mathbb{N}^*$, il est donc dénombrable. Chaque singleton de Ω étant un événement de probabilité nulle[7] et par σ-additivité, on en déduit que N est un événement et que $\mathbf{P}(N) = 0$, d'où $\mathbf{P}(\Omega') = 1$.

La probabilité $\mathbf{P}$ sur Ω' se transpose alors naturellement sur $[0;1[$, la probabilité image est la mesure μ de Lebesgue[8]. Pour cela, il suffit de vérifier que ces deux

(7). Voir la discussion du paragraphe § 13.
(8). Voir paragraphe § 52.

suites propres	réels
Ω'	$[0\,;1[$
$\mathfrak{T}$	$\mathfrak{B}([0\,;1[)$
cylindres	intervalles dyadiques
$\mathbf{P}$	μ
$(\omega_1,\omega_2,\dots)=\omega$	$x=\sum_{n=1}^{\infty}\frac{\mathscr{X}_n(x)}{2^n}$
ω_n	$\mathscr{X}_n(x)$
X_n	$\mathscr{X}_n$
ε_n	r_n
$\mathbf{E}(\mathrm{X})$	$\int_0^1 \mathscr{X}(t)\,\mathrm{d}t$
$\mathbf{P}(\mathrm{A})$	$\int_0^1 \mathbf{1}_{\mathscr{A}}(t)\,\mathrm{d}t, \quad \mathscr{A}=\varphi^{-1}(\mathrm{A})$

TABLE 9.1 — *Dictionnaire entre le langage des suites de variables de Bernoulli et celui des réels de* $[0\,;1[$.

mesures (probabilité image et Lebesgue) coïncident sur la classe des cylindres, et même qu'elles coïncident sur les cylindres élémentaires. Nous laissons le lecteur s'en convaincre à travers un exemple élémentaire : l'image de l'événement $\mathrm{P}_2\mathrm{F}_3\mathrm{P}_5$ est l'ensemble des x tels que le développement décimal de x est de la forme

$$x=\overline{0,*10*1****\cdots}$$

et c'est $\mathrm{A}=\left[\frac{9}{32}\,;\frac{11}{32}\right[\cup\left[\frac{25}{32}\,;\frac{27}{32}\right[$, de largeur $\mu(\mathrm{A})=1/8=2^{-3}$.

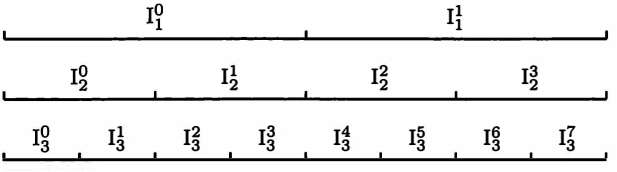
Les lois marginales et les lois jointes des X_n et des $\mathscr{X}_n$ sont les mêmes ; notamment, les variables $\mathscr{X}_n$ sont indépendantes[(9)].

§ 142 **Propriétés des fonctions de Rademacher**

Il est utile de décrire les variations des fonctions X_n, à n fixé, lorsque x parcourt $[0\,;1[$. On introduit d'abord les intervalles dyadiques

$$\mathrm{I}_n^k=\left[\frac{k}{2^n}\,;\frac{k+1}{2^n}\right[\qquad 0\leqslant k\leqslant 2^n-1,$$

l'entier n étant appelé le **rang** des intervalles I_n^k.



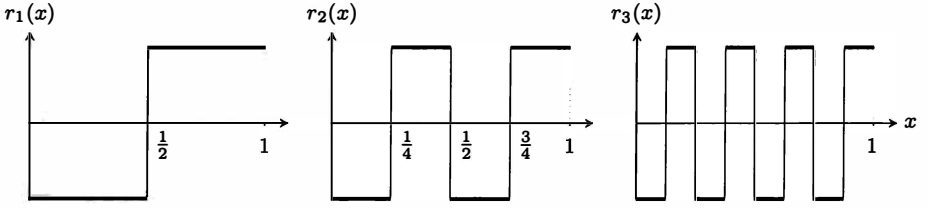
Il est plus pratique de définir les **fonctions de Rademacher**

$$r_n(x)=2\,\mathrm{X}_n(x)-1=\begin{cases}1 & \text{si } \mathrm{X}_n(x)=1\\ -1 & \text{si } \mathrm{X}_n(x)=0,\end{cases}$$

(9). Bien entendu, dans cette dernière affirmation, on peut gommer entièrement le contenu probabiliste, et transformer l'affirmation $\mathbf{P}\{\mathscr{X}_i=1,\mathscr{X}_j=0\}=\mathbf{P}\{\mathscr{X}_i=1\}\,\mathbf{P}\{\mathscr{X}_j=0\}$ en

$$\int_0^1 \mathbf{1}_{\mathscr{X}_i^{-1}(1)}(t)\,\mathbf{1}_{\mathscr{X}_j^{-1}(0)}(t)\,\mathrm{d}t=\int_0^1 \mathbf{1}_{\mathscr{X}_i^{-1}(1)}(t)\,\mathrm{d}t\int_0^1 \mathbf{1}_{\mathscr{X}_j^{-1}(0)}(t)\,\mathrm{d}t.$$

de sorte que l'allure des fonctions r_n est :



On remarquera notamment que la fonction r_n est constante sur chaque intervalle dyadique de rang supérieur ou égal à n.

La proposition suivante, bien qu'introduite en termes probabilistes, ne traite en réalité que d'intégrales de fonctions :

■ **Proposition 9.19** *Les variables r_n sont* centrées *et* orthogonales *(ou :* décorrélées*) deux à deux :*

i) Pour tout $n \geqslant 1$, r_n est d'intégrale nulle : $\int_0^1 r_n = 0$.

ii) Pour tous indices distincts k et ℓ, $\int_0^1 r_k r_\ell = 0$.

iii) De même, les intégrales $\int_0^1 r_{k_1} r_{k_2} \cdots r_{k_n}$ sont nulles lorsqu'au moins un des indices est distinct de tous les autres.

DÉMONSTRATION : Ces propriétés découlent de celles des ε_n et de leur correspondance avec les r_n : $\mathbf{E}(\varepsilon_n) = 0$, $\mathbf{E}(\varepsilon_k \varepsilon_\ell) = 0$ etc. On peut les vérifier directement.

i) r_n vaut -1 sur la moitié des intervalles dyadiques d'ordre n, et 1 sur l'autre.

ii) Supposons $k < \ell$. Sur chaque intervalle dyadique de rang $\ell - 1$, r_k est constante car $k \leqslant \ell - 1$; quant à r_ℓ, elle vaut -1 sur la moitié gauche de cet intervalle et $+1$ sur la moitié droite. Ainsi, $\int_{J_{\ell-1}^i} r_k r_\ell = 0$ pour $i = 0, \ldots, 2^{\ell-1} - 1$ et donc $\int_0^1 r_k r_\ell = 0$.

iii) Il est évident que les indices présents deux fois n'interviennent pas, car $X_k^2 = 1$ pour tout k. On peut donc supposer que les indices sont en fait tous distincts et qu'il en reste au moins deux. Notons k le plus grand de ces indices, alors sur un intervalle d'ordre $k - 1$, les autres variables sont constantes tandis que r_k prend les valeurs -1 et $+1$ sur chaque moitié de l'intervalle.

∎

Remarque 9.20 (Les r_n sont indépendantes) De même, on peut démontrer l'indépendance deux à deux des r_n, c'est-à-dire la propriété suivante (lorsque $k < \ell$)

$$\mathbf{P}\{r_k = 1, r_\ell = 1\} = \mathbf{P}\{r_k = 1, r_\ell = -1\} = \mathbf{P}\{r_k = -1, r_\ell = 1\} = \mathbf{P}\{r_k = -1, r_\ell = -1\} = \frac{1}{4},$$

ce qu'un examen des valeurs de r_k et r_ℓ sur les intervalles dyadiques d'ordre $\ell - 1$ confirme aussitôt.

On peut même démontrer leur indépendance mutuelle. Pour cela, le lecteur se convaincra qu'il suffit en pratique de prouver que

$$\mathbf{P}\{r_{i_1} = 1, r_{i_2} = 1, \ldots, r_{i_n} = 1\} = \mathbf{P}\{r_{i_1} = 1\}\, \mathbf{P}\{r_{i_2} = 1\} \cdots \mathbf{P}\{r_{i_n} = 1\} = \frac{1}{2^n}$$

pour tout n-uplet $(i_1, i_2, \ldots, i_n)$ d'indices distincts, et pour tout entier n. Là encore, l'examen des valeurs de chaque variable sur les intervalles dyadiques convenables permet de vérifier (mais c'est un peu fastidieux) ce résultat. En langage probabiliste :

la suite $(r_n)_{n \geqslant 1}$ est une suite de variables de Rademacher mutuellement indépendantes.

§ 143 **Le théorème des nombres normaux de Borel**

Convenons d'appeler **simplement normal en base 2** un nombre x vérifiant la propriété suivante : la proportion de 0 et de 1 dans le développement dyadique de x est égale à $1/2$, c'est-à-dire

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} r_k(x) = 0$$

ou encore

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathrm{X}_k(x) = \frac{1}{2}.$$

■ **Théorème 9.21 (Borel, 1909)** *Presque tout réel de* $[0\,;1[$ *est simplement normal en base* 2.

DÉMONSTRATION : Cette propriété n'est rien d'autre que la transposition de la *loi forte des grands nombres* (théorème 9.13 page 212). La démonstration proposée est encore applicable ici, modulo quelques ajustements, le rôle des variables Y_n étant ici joué par les r_n, l'espérance d'une fonction r étant définie par $\mathbf{E}(r) = \int_0^1 r$, et la probabilité d'un événement A étant $\int_0^1 \mathbf{1}_{\mathrm{A}}(t)\,\mathrm{d}t$. Dans l'expression $\mathbf{E}(r_i\, r_j\, r_k\, r_\ell) = \int_0^1 r_i\, r_j\, r_k\, r_\ell$, seule la décorrélation des variables est en fait utile pour montrer que ne restent que les intégrales du type $\int_0^1 r_n^4 = \xi^4 = 1$ et $\int_0^1 r_i^2\, r_j^2 = \sigma^4 = 1$. ∎

§ 144 **Existence de nombres complètement normaux**

Convenons d'appeler **nombre normal en base 2** un nombre dont le développement dyadique fait apparaître, non seulement la même proportion de 0 que de 1, mais également la même proportion de tout motif (comme « 011001 ») que d'un autre de même longueur. Ainsi, dans un nombre normal, un motif donné de n caractères aura une fréquence relative de $1/2^n$.

■ **Théorème 9.22** *Presque tout réel de* $[0\,;1[$ *est normal en base* 2.

DÉMONSTRATION : Examinons pour commencer le cas d'un motif « $a_1\, a_2$ » de longueur 2, avec a_1, a_2 éléments de $\{0,1\}$. On définit les fonctions (variables aléatoires) Y_n par $\mathrm{Y}_n(x) = 1$ si $\mathrm{X}_n(x) = a_1$ et $\mathrm{X}_{n+1}(x) = a_2$, et $\mathrm{Y}_n(x) = 0$ sinon. Ce sont des variables de même loi : $\mathbf{P}\{\mathrm{Y}_n = 1\} = 1/4$ pour tout n. Elles ne sont pas indépendantes mais, d'après le lemme des coalitions, les variables d'indice impair sont mutuellement indépendantes ; de même pour les variables d'indice pair.

Ces hypothèses sont suffisantes pour que l'on puisse appliquer la loi forte des grands nombres 9.13 : avec une probabilité 1,

$$\frac{\mathrm{Y}_1 + \mathrm{Y}_3 + \cdots + \mathrm{Y}_{2n-1}}{n} \to \frac{1}{4} \qquad \text{et} \qquad \frac{\mathrm{Y}_2 + \mathrm{Y}_4 + \cdots + \mathrm{Y}_{2n}}{n} \to \frac{1}{4}.$$

L'intersection de deux événements presque certains étant presque certain, on en déduit qu'avec une probabilité 1 les deux limites précédentes sont vérifiées, et donc

$$\frac{\mathrm{Y}_1 + \mathrm{Y}_2 + \cdots + \mathrm{Y}_{2n}}{2n} \to \frac{1}{4}.$$

Enfin, puisque $2n \sim 2n+1$ et que Y_{2n+1} est à valeurs dans $\{0,1\}$ donc bornée,

$$\frac{\mathrm{Y}_1 + \mathrm{Y}_2 + \cdots + \mathrm{Y}_{2n+1}}{2n+1} \to \frac{1}{4}$$

avec une probabilité 1 également. Cela suffit pour conclure. ∎

Terminons maintenant en convenant d'appeler **nombre simplement normal en base b** un réel $x \in [0\,;1[$ dont le développement b-adique $x = \sum a_n/b^n$ est tel que les chiffres $\{0,1,\ldots,b-1\}$ sont représentés avec la même fréquence relative $1/b$. On généralise de même la notion de nombre normal en base b (toutes les séquences

de même longueur n dans le développement en base b ont même fréquence $1/b^n$). Sans surprise, les démonstrations précédentes s'adaptent aisément et on généralise le théorème 9.22 en : *presque tout réel de* $[0;1[$ *est normal en base b.*

On appelle alors **nombre complètement normal** un nombre qui est normal dans toute base $b \geqslant 2$. Notons Ω_p l'ensemble des nombres normaux en base p, ensemble de probabilité 1. L'ensemble des nombres complètement normaux est

$$\mathcal{N} = \bigcap_{p=2}^{\infty} \Omega_p$$

intersection dénombrable d'ensembles de probabilité 1, donc lui-même de probabilité 1. Ainsi :

■ **Théorème 9.23** *Presque tout réel de* $[0;1[$ *est complètement normal.*

Un corollaire trivial, mais important, de ce théorème est que $\mathcal{N}$ est non vide ! Cela n'a l'air de rien mais... à l'heure actuelle, si l'on sait que $\mathcal{N}$ remplit presque tout $[0;1[$, *on ne connaît explicitement aucun nombre complètement normal !*

Remarque 9.24 (Quelques nombres) Un nombre rationnel n'est jamais complètement normal en base b : le développement en base b d'un nombre rationnel est en effet périodique (par exemple, 1/7 s'écrit, en base 10, sous la forme 0, 142857142857... avec une succession infinie de séquences « 142857 »). Un nombre moins régulier, comme π ou e, ou même $\sqrt{2}$, est peut-être normal en base b ? C'est possible, mais non prouvé — ni infirmé.

Le premier nombre connu complètement normal en base 10 est le *nombre de Champernowne* (1933), obtenu en écrivant successivement tous les entiers, puis en les concaténant :

$$0,12345678910111213141516 17\ldots$$

ou encore celui obtenu en concaténant la suite des nombres premiers (Erdős et Copeland, 1945) :

$$0,23571113171923293137\ldots$$

On perçoit ce que ces nombres ont d'artificiel — comme l'était le premier nombre transcendant fabriqué par Liouville au prix de grands efforts. Le parallèle est d'ailleurs intéressant : l'ensemble des nombres algébriques étant dénombrable, l'ensemble des nombres transcendants de $[0;1[$ est lui aussi un borélien de mesure (de probabilité) égal à 1 — c'est donc lui aussi un ensemble qui « remplit » presque entièrement $[0;1[$, mais dont les éléments sont restés cachés longtemps.

EXERCICES

♦ ***Exercice 9.1*** Les événements suivants sont-ils dans la tribu $\mathcal{T} = \sigma(\mathcal{C})$ engendrée par les cylindres de $\Omega = \{\mathrm{P}, \mathrm{F}\}^{\mathbb{N}^*}$?

i) L'ensemble des suites dont tous les termes pairs sont « P ».

ii) Un singleton $\{\omega\}$, où $\omega = (\omega_n)_{n \geqslant 1}$ est fixé.

iii) L'ensemble des suites contenant autant de « P » que de « F » en proportion, c'est-à-dire telles que

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{Card}\{k \leqslant n \,;\, \omega_k = \mathrm{P}\} = \lim_{n\to\infty} \frac{1}{n} \operatorname{Card}\{k \leqslant n \,;\, \omega_k = \mathrm{F}\} = \frac{1}{2}.$$

♦ ***Exercice 9.2 (Temps d'attente du motif PFP)*** Dans une séquence infinie de tirages à *pile* ou *face*, quel est le temps moyen d'apparition du motif « PFP » ? On commencera par déterminer si l'événement $\mathscr{E}$ associé est régénératif ou non.

♦ ***Exercice 9.3 (Temps d'attente d'une séquence de* pile*)*** Soit $\mathrm{K} \geqslant 1$ un entier. On s'intéresse aux séquences non recouvrantes de K *pile*, c'est-à-dire que l'on définit de proche en proche le fait que $\mathscr{E}$ se produit au tirage n : les K derniers tirages, numérotés $n-\mathrm{K}+1, n-\mathrm{K}+2, \ldots, n$, sont des *pile* et $\mathscr{E}$ n'a pas eu lieu aux tirages $n-\mathrm{K}+1, n-\mathrm{K}+2, \ldots, n-1$.

Pour garder un peu plus de généralité, on note p la probabilité que la pièce tombe sur *pile* à chacun des tirages (indépendants), et $q := 1-p$ la probabilité qu'elle tombe sur *face*.

i) Quelle est la probabilité d'obtenir une séquence de K *pile* aux tirages $(n-\mathrm{K}+1), \ldots, n-2, n-1, n$?

ii) En utilisant la formule des probabilités totales et en remarquant que, si les tirages indexés $n-\mathrm{K}+1$ à n sont des *pile*, l'événement $\mathscr{E}$ a lieu à un et un seul de ces indices, trouver une relation entre $a_{n-\mathrm{K}+1}, \ldots, a_n$ et p.

iii) En déduire les expressions de $\mathrm{A}(x)$ et de $\mathrm{B}(x)$.

iv) Calculer le temps moyen d'attente de $\mathscr{E}$.

♦ ***Exercice 9.4 (Un ensemble de Cantor de mesure nulle)*** Appelons *pair* un réel $x \in [0\,;1[$ dont le développement décimal propre $x = \sum_{k=1}^{n} d_n\, 10^{-n}$ ne contient que des entiers pairs. Notons $\mathscr{P}$ l'ensemble des réels pairs de $[0\,;1[$. Montrer que $\mathscr{P}$ est un borélien de mesure nulle.

Indications

◊ *Indication 9.1 :* Pour la troisième question, écrire que l'événement cherché est de la forme

$$\mathrm{A} = \left\{\omega \in \Omega \,;\, \forall \epsilon > 0 \quad \exists \mathrm{N} \in \mathbb{N}^* \quad \forall n \geqslant \mathrm{N} \qquad \left|\mathrm{S}_n(\omega)/n - \tfrac{1}{2}\right| \leqslant \epsilon\right\}.$$

◊ *Indication 9.4 :* On pourra traiter le problème directement en utilisant les fonctions d_n : $x \mapsto \lfloor 10^n x \rfloor - 10 \lfloor 10^{n-1} x \rfloor$ et les intervalles $\left[k\,10^{-n}\,;\,(k+1)\,10^{-n}\right[$; ou bien en utilisant la correspondance avec une suite de tirages indépendants avec des dés à 10 faces.

SOLUTIONS

♦ **Solution 9.1**

i) L'événement $P_{2n} = \{X_{2n} = \text{« P »}\}$ est dans la classe $\mathfrak{C}$ des cylindres, et ce pour tout n entier. L'intersection dénombrable des P_{2n} est donc dans $\mathfrak{T} = \sigma(\mathfrak{C})$.

ii) Si $\omega = (\omega_n)_{n\geqslant 1}$ est spécifié, on peut alors écrire

$$\{\omega\} = \bigcap_{n=1}^{\infty} \{X_n = \omega_n\},$$

intersection dénombrable de cylindres, donc dans $\mathfrak{T}$.

iii) Notons A l'ensemble en question. On convient que *pile* vaut 1, on note $S_n = X_1 + \cdots + X_n$, alors

$$A = \left\{\omega \in \Omega\,;\ \forall \varepsilon > 0 \quad \exists N \in \mathbb{N}^* \quad \forall n \geqslant N \qquad \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| \leqslant \varepsilon\right\}.$$

Pour tout n et tout $\varepsilon > 0$, posons

$$A_n(\varepsilon) = \left\{\omega \in \Omega\,;\ \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| \leqslant \varepsilon\right\}.$$

C'est un cylindre, donc un élément de $\mathfrak{C}$, puisqu'il est entièrement décrit par une fonction portant sur les tirages $X_1, X_2, \ldots, X_n$. Alors

$$\left\{\omega \in \Omega\,;\ \forall n \geqslant N \quad \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| \leqslant \varepsilon\right\} = \bigcap_{n=N}^{\infty} A_n(\varepsilon)$$

est élément de la tribu $\mathfrak{T}$. De même,

$$A(\varepsilon) := \left\{\omega \in \Omega\,;\ \exists N \in \mathbb{N}^* \quad \forall n \geqslant N \quad \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| \leqslant \varepsilon\right\} = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} A_n(\varepsilon) \quad \in \mathfrak{T}.$$

Enfin, écrire $A = \bigcap_{\varepsilon>0} A(\varepsilon)$ ne permet pas de conclure (c'est une intersection non dénombrable). En revanche, écrire

$$A = \bigcap_{p=0}^{\infty} A(2^{-p})$$

permet de conclure que A est encore élément de la tribu $\mathfrak{T}$.

♦ **Solution 9.2** L'événement $\mathscr{E}$ défini par l'occurrence de ce motif n'est pas naturellement régénératif, puisque la suite « PFPFP $\cdots$ » fait que $\mathscr{E}$ se produit aux tirages 3 et 5, alors qu'il ne peut pas se produire au tirage 2.

On s'intéresse donc uniquement aux occurrences non recouvrantes de ce motif. La probabilité d'avoir le motif PFP aux sites $n-2$, $n-1$ et n est $1/8$. Lorsque ceci arrive, l'événement $\mathscr{E}$ peut s'être produit aux sites n (sans condition) ou $n-2$ (et alors il y a eu en plus les tirages F_{n-1} et P_n), ces deux possibilités étant mutuellement exclusives, et l'une des deux étant nécessairement vraie. La formule des probabilités totales permet alors d'écrire

$$\frac{1}{8} = a_n + \frac{1}{4}\, a_{n-2}.$$

On multiplie par x^n et on somme pour n entre 3 et l'infini, ce qui donne

$$\frac{1}{8}\,\frac{x^3}{1-x} = [A(x) - 1]\left(1 + \frac{x^2}{4}\right)$$

puis, après quelques lignes de calcul, la valeur $B(x) = x^3/(8 - 8x + 3x^3 - 2x^4)$ et finalement

$$\mathbf{E}(T) = B'(1) = 10.$$

On peut trouver cette valeur plus rapidement en invoquant le théorème de Erdős, Feller et Pollard : la suite $(a_n)_{n\geqslant 1}$ converge vers une limite ℓ, et celle-ci vérifie donc $\frac{1}{8} = \ell + \frac{1}{4}\,\ell$, donc $\ell = 1/10$, et $\mathbf{E}(T) = 1/\ell = 10$.

Enfin, une troisième manière de calculer cette valeur sera exposée dans le chapitre sur les chaînes de Markov, paragraphe § 409 page 572.

♦ **Solution 9.3**

i) La probabilité est bien entendu p^{K}.

ii) $p^{\mathrm{K}} = a_n + a_{n-1}p + a_{n-2}p^2 + \cdots + a_{n-\mathrm{K}+1}p^{\mathrm{K}-1} = \sum_{i=0}^{\mathrm{K}-1} a_{n-i}\,p^i$.

iii) En multipliant par x^n et en sommant pour n variant entre K et l'infini :

$$\sum_{n=\mathrm{K}}^{\infty} p^{\mathrm{K}}\,x^n = \left(1 + xp + x^2p^2 + \cdots + x^{\mathrm{K}-1}p^{\mathrm{K}-1}\right) \sum_{n=\mathrm{K}}^{\infty} a_n\,x^n.$$

Comme de plus $a_1 = \cdots = a_{\mathrm{K}-1} = 0$ et $a_0 = 1$, $\dfrac{p^{\mathrm{K}}x^{\mathrm{K}}}{1-x} = \left[\mathrm{A}(x) - 1\right]\dfrac{1-p^{\mathrm{K}}x^{\mathrm{K}}}{1-px}$ donc

$$\mathrm{A}(x) = \frac{1 - x + qp^{\mathrm{K}}\,x^{\mathrm{K}+1}}{(1-x)(1-p^{\mathrm{K}}x^{\mathrm{K}})}. \qquad \text{et} \qquad \mathrm{B}(x) = \frac{p^{\mathrm{K}}(1-px)\,x^{\mathrm{K}}}{1 - x + qp^{\mathrm{K}}\,x^{\mathrm{K}+1}}.$$

iv) $\mathbf{E}(\mathrm{T}) = \mathrm{B}'(1) = \dfrac{1-p^{\mathrm{K}}}{qp^{\mathrm{K}}}$ après quelques lignes de calcul ou utilisation du théorème d'Erdős, Feller et Pollard.

Conseils de lecture

Sur le temps d'attente d'un motif et sur les événements régénératifs, l'indispensable référence est

- William Feller, *An introduction to probability theory and its applications*, Volume I [33].

Chapitre 10

Variables aléatoires à densité

En l'absence (temporaire) d'une théorie de la mesure et, notamment, de la mesure de Lebesgue sur $\mathbb{R}$ et sur $\mathbb{R}^n$ ainsi que du cortège de théorèmes d'intégration qui lui est associée, les trois chapitres qui suivent, et qui traitent tous de variables à densité, présentent un certain nombre de résultats non démontrés, mais justes (!) ; l'accent est mis sur la pratique. La théorie justifiant les méthodes proposées fait l'objet des chapitres 13 à 15.

Dans ce chapitre,

- *nous définissons les variables à densité (qu'on appelle aussi* variables absolument continues*) ;*
- *nous présentons les lois classiques (uniforme, exponentielle et normale, notamment) ;*
- *enfin, nous établissons les formules utiles lors d'un changement de variable aléatoire* $Y = \varphi(X)$.

À propos de l'intégrabilité des fonctions § 145

Nous rappelons un résultat essentiel de la théorie de l'intégration : si f est une fonction à valeurs *positives* et possédant la régularité nécessaire[1], alors on peut définir l'intégrale de f sur $\mathbb{R}$:

— si f est intégrable, la quantité $\int_{-\infty}^{+\infty} f$ est finie ;

— si f n'est pas intégrable, $\int_{-\infty}^{+\infty} f = +\infty$.

Nous utiliserons librement cette propriété, notamment lors de la preuve de l'intégrabilité d'une fonction.

Si f est une fonction à valeurs réelles (pas forcément positives), la situation est un tout petit peu plus compliquée ; l'intégrabilité de f est équivalente à la relation

$$\int_{-\infty}^{+\infty} |f| < +\infty.$$

Notamment, la quantité $\int_{-\infty}^{+\infty} f$ peut être définie[2] sans que la fonction f soit intégrable.

Enfin, les fonctions considérées dans ce chapitre auront toujours, même si ce n'est pas précisé, la régularité voulue pour les intégrer, au moins localement.

---

(1). Selon le niveau d'abstraction et de raffinement cherché, f peut être continue par morceaux, ou mesurable et localement intégrable.

(2). Au sens de la limite de $\int_A^B f$ quand les bornes A et B tendent respectivement vers $-\infty$ et $+\infty$.

10.1 VARIABLES ALÉATOIRES À DENSITÉ

Dans tout ce chapitre, les variables aléatoires considérées sont définies sur le même espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$, et à valeurs réelles.

§ 146 **Variables continues**

Soit X une variable aléatoire réelle. On dit que X est une **variable aléatoire continue** si sa fonction de répartition, définie sur $\mathbb{R}$ par

$$\mathrm{F}(x) = \mathbf{P}\{\mathrm{X} \leqslant x\},$$

est une fonction continue.

On sait déjà que, si X une variable aléatoire réelle, sa fonction de répartition F vérifie les propriétés suivantes :

i) F est croissante et continue à droite ;

ii) $\lim_{x\to-\infty} \mathrm{F}(x) = 0$ et $\lim_{x\to+\infty} \mathrm{F}(x) = 1$.

■ **Proposition 10.1** *Soit* X *une variable aléatoire, alors elle est continue si et seulement si pour tout* $a \in \mathbb{R}$, $\mathbf{P}\{\mathrm{X} = a\} = 0$[3].

DÉMONSTRATION : Soit $a \in \mathbb{R}$. On peut écrire le singleton $\{a\}$ comme l'intersection d'une famille décroissante d'intervalles :

$$\{a\} = \bigcap_{n=1}^{\infty}]\, a - \epsilon_n \,;\, a\,],$$

où $(\epsilon_n)_{n\geqslant 1}$ est une suite quelconque de réels décroissant vers 0. L'axiome de σ-additivité, sous sa forme de l'axiome 3″ (page 86), permet d'écrire que

$$\mathbf{P}\{\mathrm{X} = a\} = \mathbf{P}\{\mathrm{X} \in \{a\}\} = \lim_{n\to+\infty} \mathbf{P}\{\mathrm{X} \in]\, a - \epsilon_n \,;\, a\,]\} = \lim_{n\to\infty} \mathrm{F}(a) - \mathrm{F}(a - \epsilon_n). \quad (*)$$

Supposons maintenant que X soit continue, alors F est continue et la relation $(*)$ montre que $\mathbf{P}\{\mathrm{X} = a\} = 0$. Réciproquement, si $\mathbf{P}\{\mathrm{X} = a\} = 0$, alors la relation $(*)$, valable pour toute suite $(\epsilon_n)_{n\geqslant 1}$ décroissant vers 0, montre que F est continue à gauche en a, et donc continue en a puisqu'elle est déjà continue à droite. ∎

■ **Corollaire 10.2** *Si* X *est continue, alors pour tous* $a, b \in \mathbb{R}$ *tels que* $a \leqslant b$,

$$\mathbf{P}\{a < \mathrm{X} \leqslant b\} = \mathbf{P}\{a \leqslant \mathrm{X} \leqslant b\} = \mathbf{P}\{a < \mathrm{X} < b\} = \mathbf{P}\{a \leqslant \mathrm{X} < b\} = \mathrm{F}(b) - \mathrm{F}(a).$$

§ 147 **Variables absolument continues, ou à densité**

Parmi les variables continues, certaines offrent une encore plus grande régularité : celles dont la fonction de répartition est de classe $\mathscr{C}^1$. Une telle fonction vérifie alors

$$\forall x \in \mathbb{R} \qquad \mathrm{F}(x) = \int_{-\infty}^{x} \mathrm{F}'(t)\,\mathrm{d}t.$$

La définition suivante généralise cette propriété.

Définition 10.3 Une variable aléatoire réelle X est dite **absolument continue** s'il existe une fonction $f : \mathbb{R} \to \mathbb{R}$ positive telle que, pour tout $x \in \mathbb{R}$, on ait

$$\mathrm{F}(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t. \qquad (10.1)$$

Une telle fonction f est appelée **densité de probabilité** de la variable X. On dit également que X **admet une densité**.

---

(3). On dit que la loi de X est une probabilité **diffuse**.

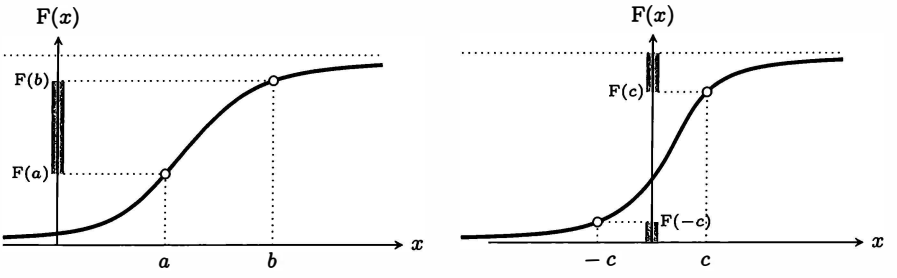


FIGURE 10.1 — *Fonction de répartition d'une variable continue. On y fait apparaître, en traits épais, les quantités* $\mathbf{P}\{a < X \leqslant b\} = F(b) - F(a)$ *et* $\mathbf{P}\{|X| > c\} = 1 - F(c) + F(-c) - 0$.

Remarque 10.4 (Régularité de la densité, point de vue pratique) Afin de pouvoir définir l'intégrale $\int_{-\infty}^{x} f$, il est nécessaire que f ait une certaine régularité. Selon le niveau d'abstraction que l'on est prêt à accepter, on demandera que f soit

— continue sur $\mathbb{R}$, éventuellement privé d'un nombre fini $x_1, x_2, \ldots, x_n$ de points (ce qui permet essentiellement de construire une intégrale de type « Riemann ») ;

— mesurable au sens de Lebesgue, et localement intégrable (ce qui permet la construction d'une intégrale de Lebesgue).

Dans tout le reste de ce chapitre, nous supposerons implicitement que les densités rencontrées ont toutes le type voulu de régularité.

Remarque 10.5 (Sur l'unicité de la densité) Supposons que X soit une variable aléatoire absolument continue, sa fonction de répartition F est dite *absolument continue*, c'est-à-dire qu'elle s'écrit sous la forme $F : x \mapsto \int_{-\infty}^{x} f$. En tout point de continuité de f, la fonction F est alors dérivable, et de dérivée $F'(x) = f(x)$.

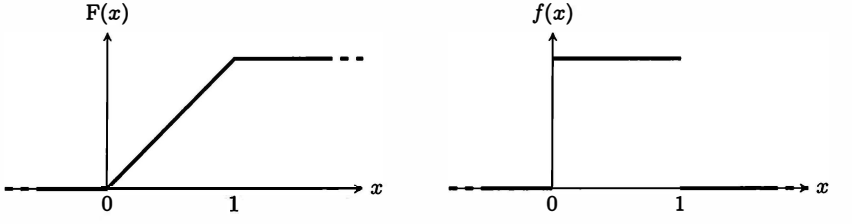
La fonction de densité n'est pas définie de manière unique (on peut, par exemple, modifier f en des points isolés, sans modifier l'intégrale de f sur un intervalle). Cependant, f est « essentiellement » unique : si f et g sont deux fonctions de densité de X, alors elles sont presque partout égales (et, même, presque partout égales à F'). Aussi, l'usage admet de parler de *la* densité de X, plutôt que d'*une* densité de X.

Enfin, signalons que, dans la majorité des cas pratiques, la fonction F sera dérivable sur $\mathbb{R}$ sauf, éventuellement, en un nombre fini $x_1, x_2, \ldots, x_n$ de points ; auquel cas la fonction f sera évidemment définie comme égale à F' sur $\mathbb{R}$ privé de ces points.

Exemple 10.6 (Loi uniforme sur $[0\,;1]$*)* Considérons une variable aléatoire $X : \Omega \to \mathbb{R}$ dont la fonction de répartition est

$$F(x) = \begin{cases} 1 & \text{si } x \geqslant 1 \\ x & \text{si } x \in [0\,;1] \\ 0 & \text{si } x < 0. \end{cases}$$

Elle admet alors pour densité $f(x) = \mathbf{1}_{[0\,;1]}(x)$.



Cette variable permet de modéliser de nombreuse situations ; par exemple, l'instant d'arrivée d'un véhicule lorsque l'on suppose

— qu'il arrive entre les instants $t = 0$ et $t = 1$, et que

— la probabilité d'arriver dans un intervalle de temps donné $I \subset [0\,;1]$ ne dépend que de la longueur de cet intervalle, et pas de sa position temporelle.

Lors de l'étape de modélisation, l'absence de connaissance sur la loi d'arrivée du véhicule, en dehors de : « c'est entre 0 et 1 », incite parfois à utiliser une loi uniforme de manière abusive.

L'existence d'une densité de probabilité permet de calculer la probabilité que X prenne une valeur dans une partie A de $\mathbb{R}$:

■ **Proposition 10.7 (Densité et calcul de probabilités)**
Soit X *une variable aléatoire admettant une densité* f. *Alors* X *est une variable continue et on a les règles de calcul suivantes :*

i) pour tout $x \in \mathbb{R}$,

$$\mathbf{P}\{X < x\} = \mathbf{P}\{X \leqslant x\} = \int_{-\infty}^{x} f(t)\,dt.$$

ii) Si a *et* b *sont des réels tels que* $a \leqslant b$, *et si l'on note* I *l'un des intervalles* $[a\,;b]$, $[a\,;b[$, $]a\,;b]$ *ou* $]a\,;b[$, *alors*

$$\mathbf{P}\{X \in I\} = \int_{a}^{b} f(t)\,dt.$$

iii) En particulier, pour tout $a \in \mathbb{R}$, $\mathbf{P}\{X = a\} = 0$.

iv) De manière plus générale, si A *est un événement de* $\mathbb{R}$ *(c'est-à-dire un borélien), alors*

$$\mathbf{P}\{X \in A\} = \int_{A} f(x)\,dx.$$

DÉMONSTRATION : Si X admet une densité, alors sa fonction de répartition vérifie la relation (10.1), donc est continue. Les trois premiers points découlent de cette propriété et de la proposition 10.1. Quant à la dernière propriété, elle est vraie pour tout intervalle ; la propriété de σ-additivité permet de montrer qu'elle est vraie pour toute union dénombrable d'intervalles : si $A = \sum A_n$ est réunion d'intervalles deux à deux disjoints, alors

$$\mathbf{P}\{X \in A\} = \mathbf{P}\left\{X \in \sum_{n=1}^{\infty} A_n\right\} = \sum_{n=1}^{\infty} \int_{A_n} f(x)\,dx = \int_{\sum A_n} f(x)\,dx = \int_{A} f(x)\,dx.$$

Des techniques standard, dont on peut omettre le détail en première lecture[4], permettent d'étendre cette propriété à l'ensemble des parties boréliennes de $\mathbb{R}$. ∎

⇘10.1

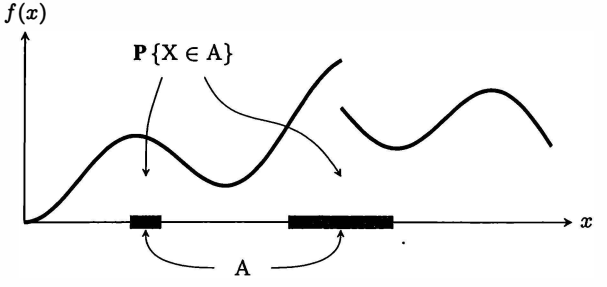


FIGURE 10.2 — *La densité de probabilité* f *d'une variable aléatoire permet de calculer* $\mathbf{P}\{X \in A\}$ *pour tout événement* A.

(4). Voir page 704, et le corollaire B.22.

Propriétés de la densité de probabilité

§ 148

Par définition de la fonction de densité d'une variable aléatoire X, la fonction f est positive[5] et d'intégrale unité :

$$\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = 1.$$

Réciproquement, toute fonction $f : \mathbb{R} \to \mathbb{R}$ mesurable[6], positive et d'intégrale unité, peut être vue comme la densité de probabilité d'une certaine variable aléatoire.

La démonstration se fait en deux étapes :
- à partir de f, on construit une fonction F par simple intégration : $\mathrm{F}(x) = \int_{-\infty}^{x} f$. La fonction F obtenue est croissante, continue et presque partout dérivable.
- on construit explicitement un espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ et une variable aléatoire X à partir de la fonction F, selon la méthode exposée au théorème C.1 page 711.

Remarque 10.8 (Ensemble image et densité de probabilité) Signalons que X prend, essentiellement, ses valeurs sur l'ensemble $\mathrm{A} = \{x \in \mathbb{R} \,;\, f(x) > 0\}$ des points en lesquels f est non nulle. « Essentiellement » parce que $\mathrm{X}(\Omega)$ peut différer de A. Il peut comprendre des points supplémentaires, par exemple être égal à $\mathrm{A} \cup \mathrm{N}$, où N est un ensemble négligeable pour la probabilité $\mathbf{P}_\mathrm{X}$ (c'est-à-dire qu'il vérifie $\mathbf{P}\{\mathrm{X} \in \mathrm{N}\} = 0$). Il peut aussi ne pas inclure certains points de A, par exemple être égal à $\mathrm{A} \setminus \mathrm{N}'$, où encore une fois N' est négligeable pour la probabilité $\mathbf{P}_\mathrm{X}$.

Changement d'échelle

§ 149

Supposons que X admette une densité de probabilité f. Si α est un réel non nul ($\alpha \neq 0$) et si β est un réel quelconque, alors la variable aléatoire $\mathrm{Y} = \alpha\mathrm{X} + \beta$ admet une densité de probabilité g reliée à f par

$$g(y) = \frac{1}{|\alpha|} f\left(\frac{y-\beta}{\alpha}\right).$$

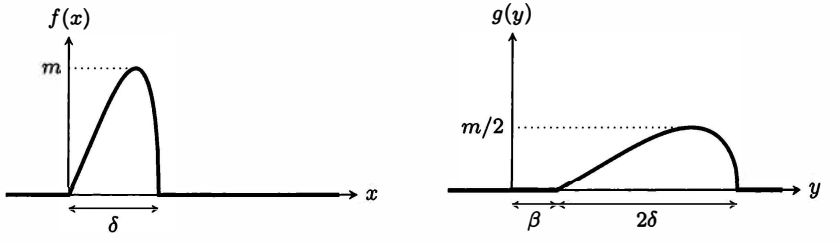


FIGURE 10.3 — *Densité d'une variable aléatoire* X *et de la variable* $\mathrm{Y} = 2\mathrm{X} + \beta$. *: le graphe est dilaté d'un facteur* 2 *en largeur, et diminué d'un facteur* 2 *en hauteur, puis translaté de la quantité* β.

DÉMONSTRATION : Notons F et G les fonctions de répartition de f et g respectivement. Supposons dans un premier temps que $\alpha > 0$. Alors, pour tout $y \in \mathbb{R}$, on a

$$\mathrm{G}(y) = \mathbf{P}\{\mathrm{Y} \leqslant y\} = \mathbf{P}\{\alpha\mathrm{X} + \beta \leqslant y\} = \mathbf{P}\left\{\mathrm{X} \leqslant \frac{y-\beta}{\alpha}\right\} = \mathrm{F}\left(\frac{y-\beta}{\alpha}\right).$$

(5). Ou, du moins, puisqu'une petite liberté nous est laissée, on peut la choisir positive sans restriction.

(6). Si l'on ne veut pas s'encombrer de considérations issues de la théorie de la mesure, on peut remplacer cette condition par : continue sauf en un nombre raisonnable de points, cf. la remarque 10.4 ci-avant.

Cela montre que Y admet une densité donnée par $g(y) = \mathrm{G}'(y) = \dfrac{1}{\alpha} f\left(\dfrac{y-\beta}{\alpha}\right)$.

Si α est strictement négatif, il faut être plus prudent avec les inégalités ; on trouve

$$\mathrm{G}(y) = \mathbf{P}\{\mathrm{Y} \leqslant y\} = \mathbf{P}\{\alpha\mathrm{X} + \beta \leqslant y\} = \mathbf{P}\left\{\mathrm{X} \geqslant \frac{y-\beta}{\alpha}\right\} = 1 - \mathrm{F}\left(\frac{y-\beta}{\alpha}\right),$$

donc Y admet pour densité $g(y) = \mathrm{G}'(y) = -\dfrac{1}{\alpha} f\left(\dfrac{y-\beta}{\alpha}\right) = \dfrac{1}{|\alpha|} f\left(\dfrac{y-\beta}{\alpha}\right)$. ∎

§150 Médiane

On appelle **médiane** d'une variable aléatoire X de fonction de répartition F tout réel a tel que

$$\mathrm{F}(a) \geqslant \frac{1}{2} \qquad \text{et} \qquad \mathrm{F}(a^-) \leqslant \frac{1}{2}.$$

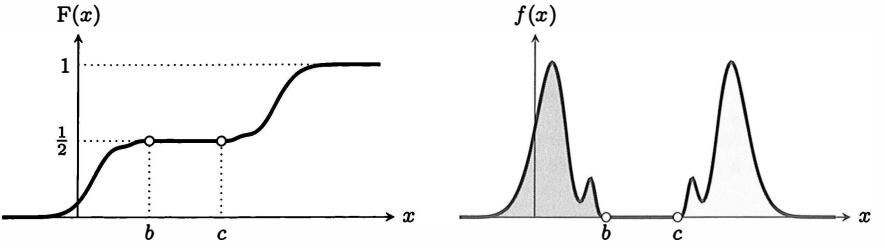


FIGURE 10.4 — *Fonction de répartition et densité d'une variable aléatoire admettant plusieurs médianes : tous les réels du segment* $[b\,;\,c]$ *sont médianes de* X.

Si F est continue, alors a est médiane si et seulement si il est solution de l'équation $\mathrm{F}(x) = \frac{1}{2}$; si, de plus, F est strictement croissante, cette solution est unique. C'est le cas notamment si X admet une densité strictement positive. La médiane est alors l'unique valeur telle que : il y a autant de chances que X soit inférieur à cette valeur, que supérieur.

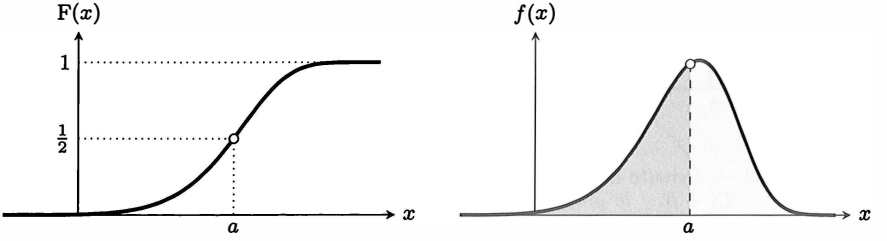


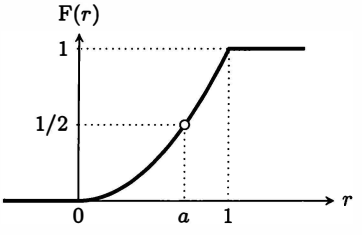
FIGURE 10.5 — *Fonction de répartition et densité d'une variable aléatoire admettant une unique médiane a.*

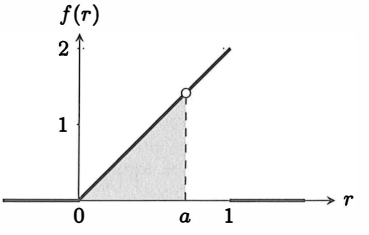
Exemple 10.9 (Tir à l'arc) On modélise, de manière extrêmement schématique, le point d'impact d'une flèche sur une cible circulaire, comme un vecteur aléatoire de répartition uniforme sur le disque unité $\mathscr{D}(0\ ;\ 1)$. On note R la variable aléatoire égale à la distance du point d'impact au centre. La couronne comprise entre les cercles de rayons r et $r + \mathrm{d}r$ étant de surface $2\pi r\,\mathrm{d}r$, on impose une relation de proportionnalité $\mathbf{P}\{r \leqslant \mathrm{X} \leqslant r + \mathrm{d}r\} \propto r\,\mathrm{d}r$ en modélisant l'expérience par une variable aléatoire de densité $f(r) = \alpha r$ pour tout $r \in [0\,;1]$, $f(r) = 0$ ailleurs. La normalisation $\int_{-\infty}^{+\infty} f(r)\,\mathrm{d}r = 1$ conduit à poser $\alpha = 2$.

En conclusion, la loi de R est donnée par

$$\mathrm{F}(r) = \begin{cases} 0 & \text{si } r < 0, \\ r^2 & \text{si } 0 \leqslant r \leqslant 1, \\ 1 & \text{si } r > 1. \end{cases}$$

La fonction F est bien continue, et même absolument continue. L'unique médiane de X est la solution de $\mathrm{F}(a) = 1/2$, c'est-à-dire $a = \sqrt{2}$.



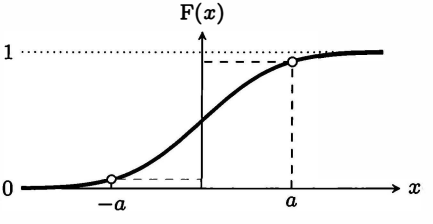


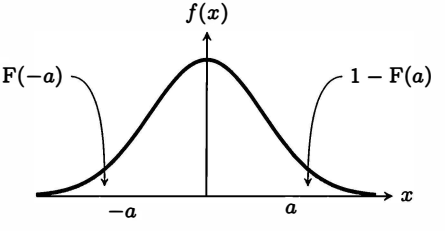
Variables symétriques § 151

Une variable aléatoire X, admettant une densité f, est dite **symétrique** si f est paire. Puisque

$$\mathrm{F}(-a) = \int_{-\infty}^{-a} f(t)\,\mathrm{d}t = \int_{a}^{+\infty} f(-u)\,\mathrm{d}u = 1 - \int_{-\infty}^{a} f(-u)\,\mathrm{d}u$$

la parité de X implique que $\mathrm{F}(-a) = 1 - \mathrm{F}(a)$ pour tout réel a. Réciproquement, si cette propriété est vérifiée, alors f vérifie $f(-a) = f(a)$ pour (presque tout) réel a.





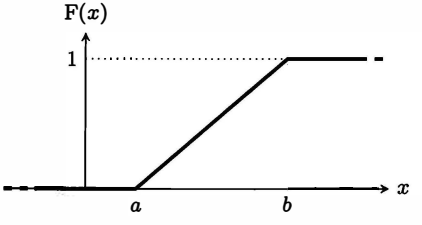
10.2 LOIS CLASSIQUES

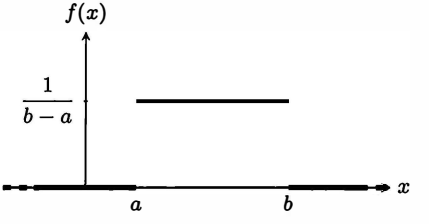
Loi uniforme § 152

Soit $[a\,;\,b]$ un segment de $\mathbb{R}$. Une variable aléatoire X suit la **loi uniforme** sur $[a\,;\,b]$ si elle admet pour fonction de répartition

$$\mathrm{F}(x) = \begin{cases} 1 & \text{si } x \geqslant b \\ \frac{x-a}{b-a} & \text{si } x \in [a\,;\,b] \\ 0 & \text{si } x < a. \end{cases}$$

Elle admet une densité $f(x) = \frac{1}{b-a}\,\mathbf{1}_{[a\,;\,b]}(x)$, et on note « $\mathrm{X} \rightsquigarrow \mathscr{U}\big([a\,;\,b]\big)$ ».



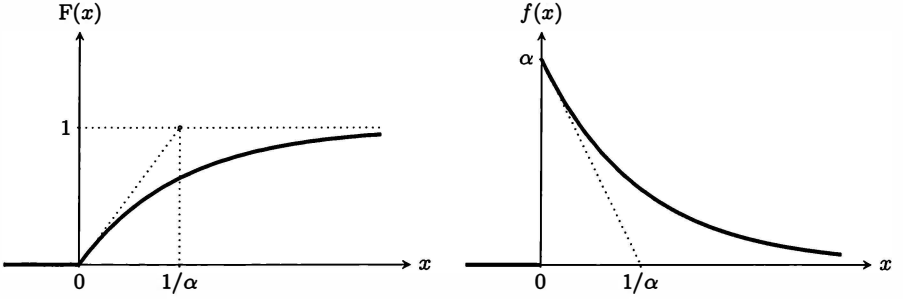


Remarque 10.10 Les lois uniformes sur $[a\,;\,b]$, sur $[a\,;\,b[$, sur $]a\,;\,b]$ ou sur $]a\,;\,b[$ sont identiques. Lorsque l'on prend la peine de préciser, par exemple, que $X \rightsquigarrow \mathscr{U}([a\,;\,b[)$, cela sous-entend en général que X ne prend jamais la valeur b et donc que $X(\Omega) = [a\,;\,b[$ — ce qui est plus fort que de dire que $\mathbf{P}\{X = b\} = 0$.

§153 **Loi exponentielle**

Nous avons introduit (exemple 2.10 page 62) un raisonnement motivant la définition suivante. Une variable aléatoire X suit la **loi exponentielle de paramètre α** (où $\alpha > 0$), et on note « $X \rightsquigarrow \mathscr{E}(\alpha)$ », si elle admet pour densité la fonction f, nulle sur $]-\infty\,;\,0[$, et valant $f(x) = \alpha\,e^{-\alpha x}$ pour tout $x \geqslant 0$. Sa fonction de répartition
▲10.2 est alors égale à

$$F(x) = \begin{cases} 0 & \text{si } x \leqslant 0, \\ 1 - e^{-\alpha x} & \text{si } x \geqslant 0. \end{cases}$$



La médiane d'une telle variable est l'unique réel a tel que $F(a) = 1/2$, c'est-à-dire $a = (\ln 2)/\alpha$; en physique nucléaire, on l'appelle *demi-vie* de l'élément dont la désintégration est modélisée par X.

Pour rester dans le cadre de la radioactivité de l'exemple 2.10, l'hypothèse que la désintégration d'un atome entre les instants t et $t + \delta t$ ne dépende que de t se traduit comme le **non vieillissement**, ou l'**absence de mémoire**, de l'atome. Autrement dit, un atome d'un élément radioactif qui vient de se former dans une supernova, ou un même atome âgé de treize milliards d'années, se comportent de façon similaire, et ont autant de chances de se désintégrer dans la minute qui vient. Cette propriété se traduit ainsi :

■ **Proposition 10.11 (Non vieillissement / absence de mémoire)**
Soit X *une variable aléatoire de loi exponentielle* $X \rightsquigarrow \mathscr{E}(\alpha)$. *Pour tous* $t > s > 0$,

$$\mathbf{P}\{X > t \,|\, X > s\} = \mathbf{P}\{X > t - s\}.$$

DÉMONSTRATION : Par définition d'une probabilité conditionnelle,

$$\mathbf{P}\{X > t \,|\, X > s\} = \frac{\mathbf{P}\{X > t, X > s\}}{\mathbf{P}\{X > s\}} = \frac{\mathbf{P}\{X > t\}}{\mathbf{P}\{X > s\}}$$

puisque l'événement $\{X > t\}$ est inclus dans l'événement $\{X > s\}$. Enfin, puisque s, t et $t - s$ sont positifs, $\mathbf{P}\{X > t\} = 1 - F(t) = e^{-\alpha t}$, $\mathbf{P}\{X > s\} = e^{-\alpha s}$ et $\mathbf{P}\{X > t - s\} = e^{-\alpha(t-s)}$, ce qui achève la démonstration. ∎

Cette propriété d'absence de mémoire est vérifiée, dans le cas discret, par les lois géométriques et par elles seules. Dans le cas absolument continu, seules les lois exponentielles la vérifient :

■ **Proposition 10.12** *Soit* X *une variable aléatoire à valeurs positives, et continue. On suppose de plus que* $F(x) < 1$ *pour tout* $x \geqslant 0$. *Si* X *est sans mémoire, c'est-à-dire si*

$$\forall t \geqslant s > 0 \qquad \mathbf{P}\{X > t \mid X > s\} = \mathbf{P}\{X > t - s\},$$

alors X *suit une loi exponentielle.*

DÉMONSTRATION : Notons $G(x) = 1 - F(x) = \mathbf{P}\{X > x\}$, quantité non nulle pour tout $x > 0$ par hypothèse. Le fait que X soit sans mémoire se traduit par $\mathbf{P}\{X > t+s\} = \mathbf{P}\{X > t\} \cdot \mathbf{P}\{X > s\}$, ou encore $G(t+s) = G(t)\,G(s)$, pour tous $s, t > 0$. Cette relation fonctionnelle, et la continuité de la fonction G, impliquent que G est de la forme $G(x) = e^{-\alpha x}$ (voir exercice 10.5) ; puisque $\lim_{x \to +\infty} G(x) = 0$, on en déduit que $\alpha > 0$. ∎

Remarque 10.13 (Lois exponentielles et lois de Poisson) Les lois exponentielles entretiennent avec les lois de Poisson des liens très étroits. Si des événements se produisent à des instants aléatoires

$$T_1 < T_2 < \cdots < T_n < \cdots$$

de telle manière que le nombre d'événements se produisant sur un intervalle de temps $[t\,;\, t+\alpha]$ suit une loi de Poisson de paramètre $\lambda\alpha$, alors l'intervalle entre deux événements successifs T_n et T_{n+1} suit une loi exponentielle de paramètre λ (voir exercice 10.6). ⚠10.6

Réciproquement, si l'on définit $T_0 = 0$ et $T_{n+1} = T_n + \Delta_n$, où les Δ_n sont indépendantes et suivent des lois exponentielles de même paramètre λ, le nombre de variables prenant leur valeur dans l'intervalle $[0\,;\, \lambda\alpha]$ suit une loi de Poisson de paramètre α (voir le paragraphe § 369, page 514).

Loi normale centrée réduite § 154

La **loi gaussienne**, ou de **loi de Laplace-Gauss**(7), ou encore **loi normale**(8), est une des lois les plus fondamentales de la théorie des probabilités, notamment parce que c'est une loi limite universelle et qu'elle modélise de façon extrêmement précise une grande variété de phénomènes d'origines très diverses ; c'est pourquoi nous nous attarderons un peu sur ses caractéristiques numériques.

On dit qu'une variable aléatoire X suit la **loi normale centrée réduite** si elle admet pour densité la fonction définie sur $\mathbb{R}$ par

$$n(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

On écrit alors « $\mathbf{X} \rightsquigarrow \mathcal{N}(0\,;\,1)$ ».

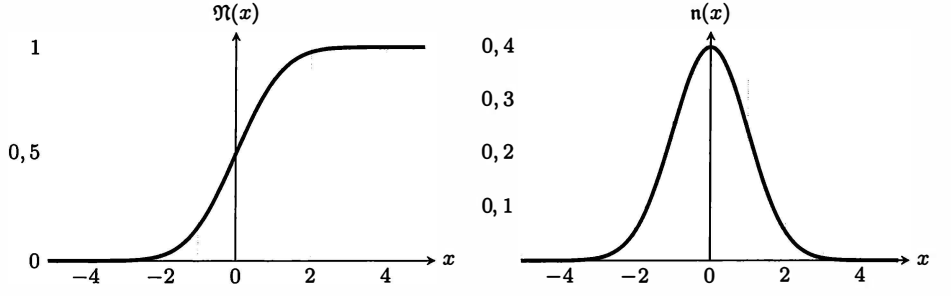


FIGURE 10.6 — *Fonction de répartition* $\mathfrak{N}(x)$ *et densité de probabilité* $n(x)$ *de la loi normale centrée réduite* $\mathcal{N}(0\,;\,1)$. *On notera l'écrasement très rapide de la densité, quasiment nulle au delà de* $|x| = 4$.

(7). Laplace discutait déjà de variables gaussiennes dans les années 1780, bien avant Gauss, et énonçait une version du théorème limite central en 1812.

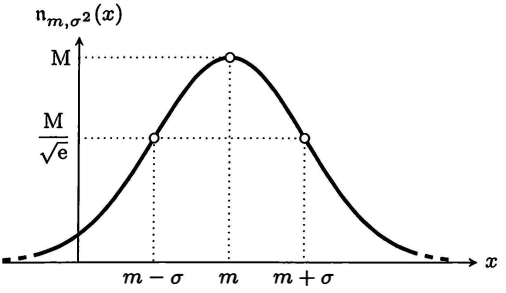
(8). Les physiciens parlent également de **distribution maxwellienne**, car le facteur de Boltzmann associé à l'énergie cinétique $E = \frac{1}{2}mv^2$ mène à un facteur statistique $e^{-mv^2/2k_BT}$.

§155 **Loi normale**

De même, on dit que X suit la **loi normale de paramètres m et σ^2** (où $\sigma > 0$) si elle admet pour densité

$$\mathfrak{n}_{m,\sigma^2}(x) = \frac{1}{\sigma}\,\mathfrak{n}\left(\frac{x-m}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}\,\exp\left(-\frac{(x-m)^2}{2\,\sigma^2}\right).$$

On écrit alors « $\mathbf{X} \rightsquigarrow \mathcal{N}\left(m\,;\,\sigma^2\right)$ [9] ». Ces paramètres m et σ^2 seront prochainement interprétés comme l'espérance et la variance de X ; le premier est la valeur autour de laquelle la fonction de densité est symétrique, la seconde indique une largeur typique à hauteur $1/\sqrt{\mathrm{e}}$ par rapport au maximum $\mathrm{M} = 1/\sigma\sqrt{2\pi}$:



Dans tout l'ouvrage, on notera $\mathfrak{N}$ la fonction de répartition de la loi normale centrée réduite :

$$\mathfrak{N}(x) = \int_{-\infty}^{x} \mathfrak{n}(t)\,\mathrm{d}t = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\,\mathrm{d}t.$$

Cette fonction ne s'exprime pas en terme de fonctions usuelles, aussi a-t-elle été tabulée très tôt à l'usage des statisticiens et différents usagers des mathématiques (voir page 720). L'usage de telles tables tend à disparaître, car la plupart des logiciels de calcul formel peuvent donner la valeur de $\mathfrak{N}(x)$ [10]. Leur usage occasionnel, notamment à l'occasion des exercices, permet de se familiariser avec les ordres de grandeur et de retenir quelques valeurs usuelles.

Remarque 10.14 Si X est une variable aléatoire réelle, alors il est équivalent de dire que X suit la loi $\mathcal{N}\left(m\,;\,\sigma^2\right)$ ou que $\mathrm{X}^* := (\mathrm{X} - m)/\sigma$ suit la loi normale centré réduite $\mathcal{N}\,(0\,;\,1)$. En effet, si l'on note respectivement F et F* leurs fonctions de répartition et si l'on suppose que X suit la

[9]. Deux usages sont en concurrence en ce qui concerne le choix du second paramètre : certains préfèrent prendre l'écart-type σ et noter $\mathcal{N}\,(m\,;\,\sigma)$ la loi que nous notons $\mathcal{N}\left(m\,;\,\sigma^2\right)$. Pour des raisons de cohérence avec les notations usuelles concernant les vecteurs gaussiens (lois normales à valeurs dans $\mathbb{R}^d$), nous emploierons plutôt la variance comme second paramètre.

[10]. Des logiciels de calcul formel ou numérique (Maxima, Maple/Mathematica, Matlab/Scilab, GNUPLOT,...) fournissent plutôt la **fonction d'erreur** « erf », définie par

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \mathrm{e}^{-t^2}\,\mathrm{d}t,$$

et liée à la fonction $\mathfrak{N}$ par

$$\mathfrak{N}(x) = \tfrac{1}{2}\left(1 + \mathrm{erf}(x/\sqrt{2})\right) \qquad \mathrm{erf}(x) = 2\,\mathfrak{N}\left(\sqrt{2}\,x\right) - 1.$$

loi $\mathcal{N}(m\,;\,\sigma^2)$, alors pour tout réel x, on a

$$\mathrm{F}^*(x) = \mathbf{P}\{\mathrm{X}^* \leqslant x\} = \mathbf{P}\{\mathrm{X} \leqslant \sigma x + m\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\sigma x + m} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right) \mathrm{d}t$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y = \mathfrak{N}(x)$$

par changement de variable $t = \sigma y + m$, ce qui montre que X suit la loi normale centrée réduite. La réciproque se montre de la même façon.

La parité de la densité implique immédiatement le résultat suivant :

■ **Proposition 10.15** *La loi normale centrée réduite est symétrique : pour tout x,*

$$\mathfrak{N}(-x) = 1 - \mathfrak{N}(x).$$

Le théorème suivant donne quelques valeurs numériques caractéristiques d'une loi normale, illustrées à la figure 10.7.

■ **Proposition 10.16 (Répartition de la loi normale centrée réduite)** *Si la variable aléatoire* X *suit la loi $\mathcal{N}(m\,;\,\sigma^2)$, alors $(\mathrm{X} - m)/\sigma$ suit la loi $\mathcal{N}(0\,;\,1)$ et*

$$\mathbf{P}\{|\mathrm{X} - m| \leqslant \varepsilon\sigma\} = \mathfrak{N}(\varepsilon) - \mathfrak{N}(-\varepsilon) = 2\,\mathfrak{N}(\varepsilon) - 1$$

pour tout $\varepsilon > 0$. Les valeurs numériques suivantes sont utiles :

$$\mathbf{P}\{|\mathrm{X} - m| \leqslant 0{,}674\,\sigma\} \approx 50\,\% \qquad \mathbf{P}\{|\mathrm{X} - m| \leqslant \sigma\} \approx 68{,}27\,\%$$

$$\mathbf{P}\{|\mathrm{X} - m| \leqslant 1{,}960\,\sigma\} \approx 95\,\% \qquad \mathbf{P}\{|\mathrm{X} - m| \leqslant 2\sigma\} \approx 95{,}45\,\%$$

$$\mathbf{P}\{|\mathrm{X} - m| \leqslant 2{,}576\,\sigma\} \approx 99\,\% \qquad \mathbf{P}\{|\mathrm{X} - m| \leqslant 3\sigma\} \approx 99{,}73\,\%$$

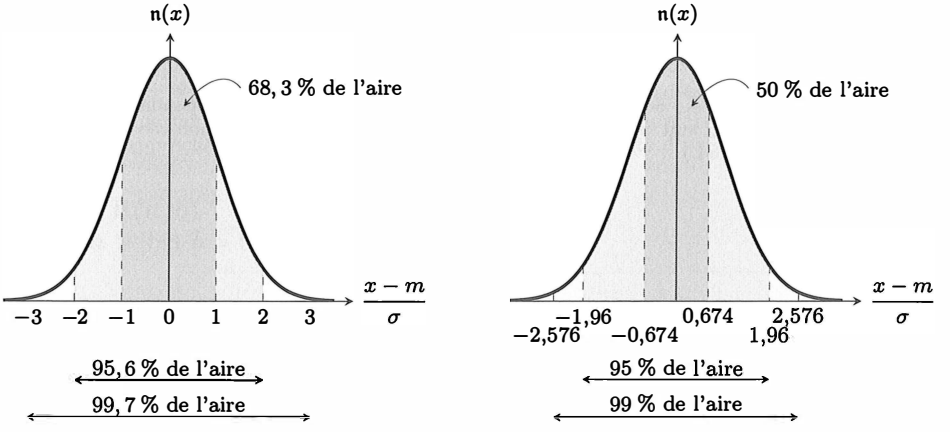


FIGURE 10.7 — *Densité d'une loi normale centrée $\mathcal{N}(m\,;\,\sigma^2)$. Les zones grisées sont les domaines $\{|\mathrm{X} - m| \leqslant \varepsilon\sigma\}$, dont on indique la probabilité pour différentes valeurs de ε.*

§ 156 **Loi de Cauchy**

Une variable aléatoire réelle X suit la **loi de Cauchy** de paramètres (m, a), où m est un réel quelconque et $a > 0$, si elle admet pour densité la fonction

$$f(x) = \frac{a}{\pi} \frac{1}{a^2 + (x-m)^2}.$$

Dans ce cas, sa fonction de répartition est

$$\mathrm{F}(x) = \frac{a}{\pi} \int_{-\infty}^{x} \frac{1}{a^2 + (x-m)^2} = \frac{1}{2} + \frac{1}{\pi} \operatorname{Arc\,tan}\left(\frac{x-m}{a}\right).$$

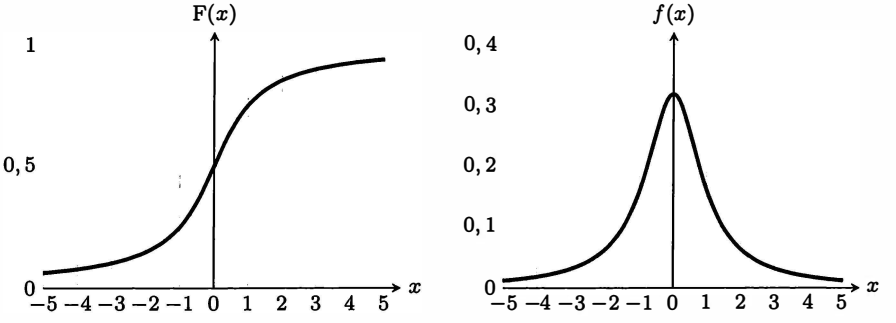


FIGURE 10.8 — *Fonction de répartition et densité de probabilité de la loi de Cauchy* $\mathscr{C}(0\,;\,1)$.

La loi de Cauchy de paramètres $(0,1)$, de densité $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$, est la loi que suit la tangente d'un angle Θ suivant une loi uniforme sur $\left]-\frac{\pi}{2}\,;+\frac{\pi}{2}\right[$. En effet, si $\Theta \rightsquigarrow \mathscr{U}\left(\left]-\frac{\pi}{2}\,;+\frac{\pi}{2}\right[\right)$ et si l'on pose $\mathrm{T} = \tan\Theta$, alors pour tout réel t, on a

$$\mathrm{F_T}(t) = \mathbf{P}\{\mathrm{T} \leqslant t\} = \mathbf{P}\{\tan\Theta \leqslant t\} = \mathbf{P}\{\Theta \leqslant \operatorname{Arc\,tan} t\} = \frac{1}{\pi}\left(\operatorname{Arc\,tan} t + \frac{\pi}{2}\right).$$

Exemple 10.17 Si X_1 et X_2 sont deux variables aléatoires gaussiennes indépendantes, alors la variable aléatoire $\mathrm{X}_1/\mathrm{X}_2$ suit une loi de Cauchy. (Cette propriété sera démontrée dans le cha-
⏢10.4 pitre 12, paragraphe § 184.)

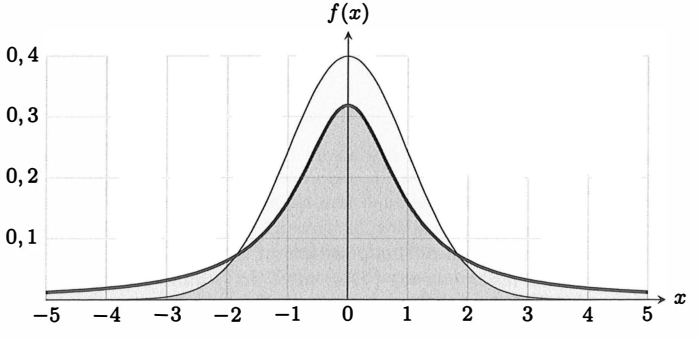


FIGURE 10.9 — *Comparaison d'une loi normale centrée réduite (en trait fin) et d'une loi de Cauchy* $\mathscr{C}(0\,;\,1)$ *(en gras). La densité gaussienne décroît beaucoup plus rapidement que celle de la loi de Cauchy.*

Loi Γ (Gamma) § 157

Une variable aléatoire X suit la **loi $\Gamma(\alpha, \beta)$** (avec $\alpha > 0$ et $\beta > 0$) si elle admet une densité f nulle sur $\mathbb{R}^-$ et définie, pour tout $x > 0$, par

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}.$$

Le facteur de normalisation est défini avec la fonction Γ d'Euler (voir, en appendice, le paragraphe A21, page 675).

Si $\alpha = n$ est entier, c'est la loi suivie par la somme de n variables aléatoires indépendantes de même loi exponentielle $\mathscr{E}(\beta)$. ⚓12.10

Loi γ § 158

Une variable aléatoire suit la **loi γ de paramètre $\nu > 0$** si elle admet pour densité la fonction nulle sur $\mathbb{R}^-$, et définie sur $\mathbb{R}^{*+}$ par

$$f(x) = \frac{x^{\nu-1} e^{-x}}{\Gamma(\nu)}.$$

La loi $\gamma(\nu)$ est donc égale à la loi $\Gamma(\nu, 1)$. Si $\nu = n$ est un entier, c'est la loi suivie par la somme de n variables aléatoires indépendantes de même loi $\mathscr{E}(1)$. ⚓12.10

Loi du χ^2 § 159

Soit p un entier naturel non nul. On dit qu'une variable aléatoire X suit la **loi du khi-deux à p degrés de liberté**, et on écrit « $X \rightsquigarrow \chi^2_p$ », si elle admet pour densité la fonction nulle sur $\mathbb{R}^-$ et définie, pour tout $x > 0$, par

$$f(x) = \frac{1}{2^{p/2}\,\Gamma(p/2)} x^{p/2-1} e^{-x/2}.$$

La loi χ^2_p est donc égale à la loi $\Gamma\left(\frac{p}{2}, \frac{1}{2}\right)$.

Remarque 10.18 L'importance des lois du χ^2 vient du résultat suivant. Il est aisé de démontrer que, si $N \rightsquigarrow \mathcal{N}(0\,;\,1)$, alors $X = N^2$ suit la loi χ^2_1 (voir l'exemple 10.21 page 239) ; de manière plus générale, si $X = N_1^2 + N_2^2 + \cdots + N_p^2$ est somme des carrés de p variables aléatoires indépendantes de même loi normale $\mathcal{N}(0\,;\,1)$, alors $X \rightsquigarrow \chi^2_p$. Les lois du χ^2 apparaissent naturellement, mais de manière approchée, lorsque l'on cherche à tester l'homogénéité d'une série de mesures.

Loi bêta B(a, b) § 160

Soient $a > 0$ et $b > 0$ des réels. On pose

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1}\, dt.$$

Si a et b sont entiers non nuls, une simple récurrence permet d'établir que

$$B(a, b) = \frac{(a-1)!\,(b-1)!}{(a+b-1)!}.$$

D'une manière plus générale (a et b pouvant être des réels non entiers), le nombre $B(a, b)$ vaut

$$B(a, b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)},$$

où Γ est la fonction d'Euler (A21).

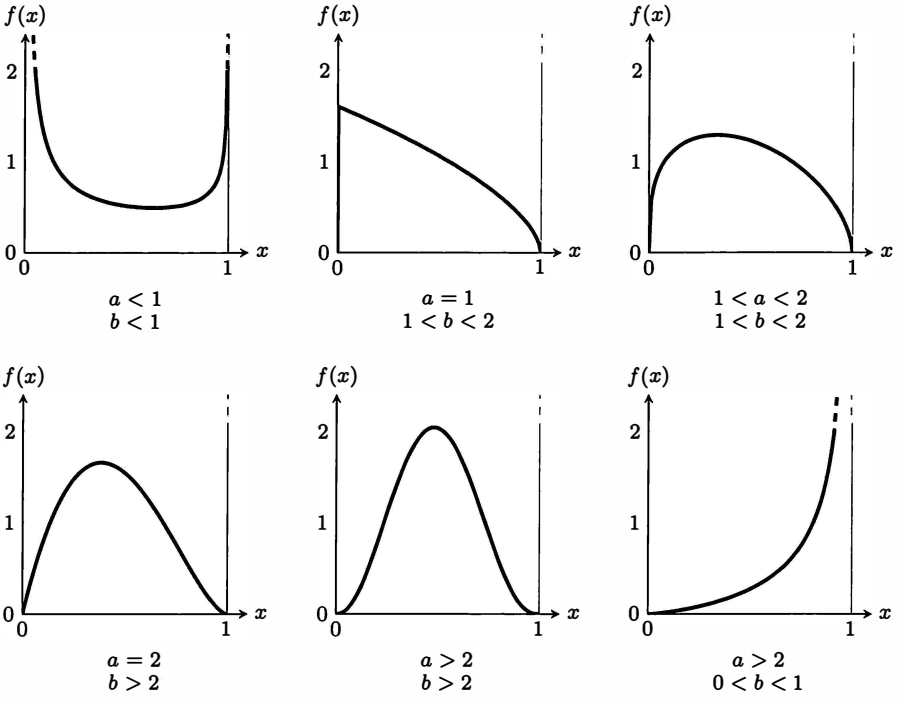


FIGURE 10.10 — *Quelques exemples d'allure de loi* $\mathrm{B}(a,b)$ *selon les valeurs des paramètres* a *et* b.

On dit qu'une variable aléatoire X suit la **loi bêta de paramètres a et b**, et on écrit « $\mathrm{X} \rightsquigarrow \mathrm{B}(a,b)$ », si elle admet pour densité la fonction

$$f(x) = \begin{cases} \dfrac{x^{a-1}\,(1-x)^{b-1}}{\mathrm{B}(a,b)} & \text{si } 0 \leqslant x \leqslant 1, \\ 0 & \text{sinon.} \end{cases}$$

Exemple 10.19 (Min et max de lois uniformes) Soient $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$ des variables aléatoires
↳12.9 indépendantes, de même loi uniforme sur $[0\,;1]$; alors

$$\mathrm{I} := \min(\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n) \qquad \text{et} \qquad \mathrm{M} := \max(\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n)$$

suivent respectivement les lois bêta $\mathrm{B}(1,n)$ et $\mathrm{B}(n,1)$. (La méthode générale est au § 185).

Les lois bêta, de par la diversité de l'allure de leurs densités, permettent de modéliser de nombreuses situations où l'on a affaire à des variables aléatoires à valeurs dans $[0\,;1]$. Selon la valeur de a, la fonction de densité a l'allure suivante au voisinage 0 :

$0 < a < 1$	$a = 1$	$1 < a < 2$	$a = 2$	$2 < a$
divergence	limite > 0	limite nulle tangente verticale	limite nulle tangente diagonale	limite nulle tangente horizontale

La densité a la même allure au voisinage de 1 selon la valeur du paramètre b (voir figure 10.10).

10.3 CHANGEMENT DE VARIABLE ALÉATOIRE

Soit X une variable aléatoire, et soit φ une fonction à valeurs réelles, que l'on suppose définie sur $X(\Omega)$, ce qui permet de considérer la variable aléatoire $\varphi \circ X$, que l'on note plus volontiers $\varphi(X)$[11]. Le problème général abordé ici est de déterminer la loi de la variable aléatoire $\varphi(X)$, connaissant celle de X.

Si l'on note F la fonction de répartition de X et G celle de Y, on a par définition, et pour tout $y \in \mathbb{R}$:

$$G(y) = \mathbf{P}\{Y \leqslant y\} = \mathbf{P}\big\{\varphi(X) \leqslant y\big\}.$$

On répond donc à la question en déterminant, pour tout y, l'ensemble $\big\{\varphi(X) \leqslant y\big\}$, c'est-à-dire l'ensemble des ω vérifiant $\varphi\big(X(\omega)\big) \leqslant y$. Selon la fonction φ choisie, la résolution de cette inégalité peut être plus ou moins aisée. Nous abordons deux cas simples parmi les plus fréquents, dans lesquels X admet une densité f et φ est une fonction de classe $\mathscr{C}^1$.

Loi de $\varphi(X)$ lorsque φ est injective § 161

Soit A un intervalle tel que $X(\Omega) \subset A$, et supposons $\varphi : A \to \mathbb{R}$ continue, de classe $\mathscr{C}^1$ sauf éventuellement en un nombre fini de points, et injective ; notamment, φ est strictement monotone.

Dans un premier temps, supposons que φ est strictement croissante. Notons $B = \varphi(A)$ l'image de φ ; on peut écrire, pour tout $y \in B$

$$G(y) = \mathbf{P}\big\{\varphi(X) \leqslant y\big\} = \mathbf{P}\big\{X \leqslant \varphi^{-1}(y)\big\} = F\big(\varphi^{-1}(y)\big),$$

où $\varphi^{-1} : B \to A$ est la bijection réciproque de φ. Les théorèmes usuels prouvent que φ^{-1} est continue, et donc que G est continue. De plus, si $y = \varphi(x)$ est un réel tel que $\varphi'(x) \neq 0$, la fonction φ^{-1} est dérivable en y, donc G l'est également et de dérivée

$$g(y) = G'(y) = \frac{F'\big(\varphi^{-1}(y)\big)}{\varphi'\big(\varphi^{-1}(y)\big)} = \frac{f(x)}{\varphi'(x)}.$$

La fonction g, définie presque partout, est alors une densité de $\varphi(X)$.

Exemple 10.20 (Loi de X^3) On suppose que X admet une densité f sur $\mathbb{R}$, et l'on pose $Y = X^3$. L'application $\varphi : x \mapsto x^3$ est bijective de $\mathbb{R}$ sur $\mathbb{R}$ et sa dérivée ne s'annule qu'en 0. Pour tout $y \in \mathbb{R}$, on a

$$G(y) = F(y^{1/3}),$$

donc G est continue sur $\mathbb{R}$ et dérivable en tout $y \neq 0$, de dérivée

$$g(y) = G'(y) = \frac{f(y^{1/3})}{3\,(y^{1/3})^2} = \frac{f(y^{1/3})}{3\,y^{2/3}}.$$

On peut alors vérifier explicitement que g est une densité de Y, c'est-à-dire que g est intégrable et que

$$G(y) = \int_{-\infty}^{y} g'(s)\,ds$$

pour tout réel y.

Supposons maintenant que φ soit une bijection décroissante. On peut écrire, pour tout réel y :

$$G(y) = \mathbf{P}\big\{\varphi(X) \leqslant y\big\} = \mathbf{P}\big\{X \geqslant \varphi^{-1}(y)\big\} = 1 - F\big(\varphi^{-1}(y)\big),$$

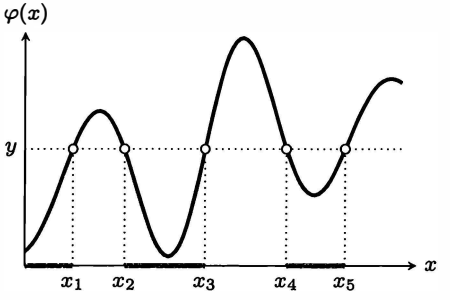
(11). Supposer que $\varphi(X)$ est une variable aléatoire, c'est supposer qu'elle est mesurable ; il suffit que φ soit elle-même mesurable pour que ce soit le cas. En pratique, toutes les fonctions usuelles conviendront.

ce qui montre que G est continue, et dérivable en tout $y \neq 0$, de dérivée

$$g(y) = \mathrm{G}'(y) = -\frac{\mathrm{F}'\big(\varphi^{-1}(y)\big)}{\varphi'\big(\varphi^{-1}(y)\big)} = \frac{f(x)}{|\varphi'(x)|}.$$

§ 162 **Loi de $\varphi(\mathrm{X})$: cas général**

Supposons toujours φ de classe $\mathscr{C}^1$, mais de variations quelconques. Soit y un réel ; on suppose que l'équation $\varphi(x) = y$ admet un nombre fini de solutions $x_1, x_2, \ldots, x_n$ et que, de plus, $\varphi'(x_k) \neq 0$ pour chacune de ces solutions. C'est une situation finalement assez fréquente : il suffit de supposer que φ est $\mathscr{C}^1$ et strictement monotone par morceaux. Pour fixer les idées, imaginons la situation suivante :



L'ensemble des x tels que $\varphi(x) \leqslant y$ est, dans ce cas,

$$\big\{x \in \mathbb{R}\,;\ \varphi(x) \leqslant y\big\} =]-\infty\,;\,x_1] \cup [x_2\,;\,x_3] \cup [x_4\,;\,x_5]\,,$$

ce qui donne

$$\big\{\varphi(\mathrm{X}) \leqslant y\big\} = \big\{\mathrm{X} \in]-\infty\,;\,x_1] \cup [x_2\,;\,x_3] \cup [x_4\,;\,x_5]\big\}$$

et donc

$$\begin{aligned}\mathrm{G}(y) &= \mathbf{P}\big\{\mathrm{X} \in]-\infty\,;\,x_1]\big\} + \mathbf{P}\big\{\mathrm{X} \in [x_2\,;\,x_3]\big\} + \mathbf{P}\big\{\mathrm{X} \in [x_4\,;\,x_5]\big\} \\ &= \mathrm{F}(x_1) - \mathrm{F}(x_2) + \mathrm{F}(x_3) - \mathrm{F}(x_4) + \mathrm{F}(x_5) \\ &= \sum_k \mathrm{F}(x_k) \cdot \operatorname{sgn}\big(\varphi'(x_k)\big).\end{aligned}$$

Le lecteur se convaincra aisément que cette formule est bien générale, à un détail près : si $\varphi(x) \leqslant 0$ sur le dernier intervalle $[x_n\,;\,+\infty[$, alors le dernier terme de la somme précédente est $\mathbf{P}\{\mathrm{X} \in [x_n\,;\,+\infty[\} = 1 - \mathrm{F}(x_n)$, et

$$\mathrm{G}(y) = \sum_k \mathrm{F}(x_k) \cdot \operatorname{sgn}\big(\varphi'(x_k)\big) + 1.$$

Dans un cas comme dans l'autre, la dérivation de cette formule mène à

$$g(y) = \sum_{x_k\,;\,\varphi(x_k)=y} \frac{f(x_k)}{|\varphi'(x_k)|}.$$

⚓10.7
⚓10.8

Exemple 10.21 (Loi de X^2) Soit X une variable aléatoire de densité f et posons $Y = X^2$. Alors la densité g de Y vaut

$$g(y) = \begin{cases} 0 & \text{si } y < 0, \\ \dfrac{f(\sqrt{y}) + f(-\sqrt{y})}{2\sqrt{y}} & \text{si } y > 0 \end{cases}$$

et n'est pas définie en $y = 0$.

Si X suit une loi normale $\mathcal{N}(0\,;\,1)$, alors $Y = X^2$ admet pour densité

$$g(y) = \begin{cases} 0 & \text{si } y < 0, \\ \dfrac{e^{-y/2}}{\sqrt{2\pi}\sqrt{y}} & \text{si } y > 0, \end{cases}$$

et Y suit la *loi du khi-deux à un degré de liberté.*

Simuler une loi à densité

§ 163

Soit X une variable aléatoire admettant une densité strictement positive sur un intervalle I, et nulle ailleurs; sa fonction de répartition F est donc continue, strictement croissante sur I, nulle « à gauche » de I et égale à 1 « à droite » de I (ces expressions pouvant ne concerner que les seuls points $-\infty$ ou $+\infty$ si I est non borné). Quitte à remplacer I par l'intervalle $\overline{\mathrm{I}}$ de $\overline{\mathbb{R}}$, égal à I auquel on a ajouté les points $-\infty$ et $+\infty$ si nécessaire, l'application F induit donc une bijection de $\overline{\mathrm{I}}$ dans $[0\,;\,1]$. On note G la réciproque de cette bijection.

■ **Proposition 10.22** *Sous les hypothèses précédentes, la variable aléatoire* $F(X)$ *suit la loi uniforme* $\mathcal{U}([0\,;\,1])$. *Réciproquement, si* U *suit la loi uniforme* $\mathcal{U}([0\,;\,1])$, *alors* $G(U)$ *admet* F *pour fonction de répartition.*

DÉMONSTRATION : Tout d'abord, l'ensemble des valeurs prises par $F(X)$ est inclus dans $[0\,;\,1]$. Pour tout $y \in]0\,;\,1[$, il existe un unique $x = G(y) \in I$ tel que $y = F(x)$; on a alors, par stricte croissance de F :

$$\mathbf{P}\{F(X) \leqslant y\} = \mathbf{P}\{X \leqslant x\} = F(x) = y.$$

Cela suffit pour prouver que la variable $F(X)$ suit la loi uniforme $\mathcal{U}([0\,;\,1])$.

Réciproquement, si U est uniforme sur $[0\,;\,1]$, alors $G(U)$ prend ses valeurs dans I et

$$\mathbf{P}\{G(U) \leqslant x\} = \mathbf{P}\{U \leqslant F(x)\} = F(x). \qquad ∎$$

Remarque 10.23 Dans le cas général, si F n'établit pas de bijection, il n'existe pas de réciproque de F, mais on peut construite un « pseudo-inverse » de F (voir A29, page 679) ; la méthode est détaillée au paragraphe § 370, page 515.

Munis de cette proposition, on peut simuler numériquement une variable aléatoire de fonction de répartition F en partant d'une variable aléatoire U de loi uniforme $\mathcal{U}([0\,;\,1])$ et en calculant $X = G(U)$.

Exemple 10.24 (Loi exponentielle) Simulons informatiquement une variable aléatoire X de loi exponentielle $\mathcal{E}(\alpha)$, où $\alpha > 0$ est un réel fixé. La fonction de répartition F de la loi $\mathcal{E}(\alpha)$ vérifie $F(x) = 0$ si $x \leqslant 0$ et $F(x) = 1 - e^{-\alpha x}$ si $x \geqslant 0$. Ainsi, F établit une bijection de $[0\,;+\infty[$ dans $[0\,;\,1[$, dont la réciproque est

$$G : y \longmapsto -\frac{1}{a}\ln(1-y).$$

Si U suit la loi $\mathcal{U}([0\,;\,1])$, alors $(-1/a)\ln(1-U)$ suit la loi $\mathcal{E}(\alpha)$[12]. On peut donc simuler X par le biais du programme suivant, où la variable x est construite, dans chaque langage, pour suivre une loi uniforme sur $[0\,;\,1[$[13] :

---

(12). La variable $1 - U$ suit également la loi uniforme sur $[0\,;\,1]$, ce qui fait que l'on peut tout aussi bien poser $X = (-1/a)\ln U$

(13). Avec Maple, la fonction `rand()` retourne un entier aléatoire entre 0 et $10^{12} - 1$; la multiplication par 10^{-12} le transforme en un réel de $[0\,;\,1[$.

```
E := proc(a)
  local x,y ;
  x := rand()*10^(-12) ;
  y := -log(1-x)/a ;
  return(y) ;
end proc
```

Code Maple

```
function y = E(a)
  x = rand()
  y = -log(1-x)/a
endfunction
```

Code Scilab

```
from random import random
from math import log

def exponentielle(a):
    y = -log(1 - random()) / a
    return y
```

Code Python

Des simulations de la loi $\mathscr{E}(1)$ sont présentées à la figure 20.2 page 429.

La simulation de lois quelconques est davantage détaillée dans le chapitre 25.

EXERCICES

♦ ***Exercice 10.1*** Soit X une variable aléatoire continue, de fonction de répartition F. On pose $X^+ = \max(X, 0)$, $X_- = \min(X, 0)$ et $X^- = -X_- = \max(-X, 0)$. Déterminer les fonctions de répartition F^+, F_- et F^- de X^+, X_- et X^-. Faire un graphe pour chacune de ces fonctions.

♦ ***Exercice 10.2*** Soit X une variable aléatoire suivant la loi exponentielle de paramètre 1 : $X \rightsquigarrow \mathscr{E}(1)$. Soit $\lambda > 0$; montrer que la variable aléatoire $Y = \lambda X$ suit une loi exponentielle dont on précisera le paramètre.

♦ ***Exercice 10.3 (Deux variables de même loi)*** Soit X une variable aléatoire admettant pour densité $f(x) = \dfrac{1}{\ln(2)} \cdot \dfrac{1}{1+x}$ si $x \in]0\,;1[$ et $f(x) = 0$ sinon. Déterminer la loi de $Y = \dfrac{1}{X} - \left\lfloor \dfrac{1}{X} \right\rfloor$.

♦ ***Exercice 10.4*** Soit X une variable aléatoire suivant la loi de Cauchy $\mathscr{C}(0\,;\,1)$. Montrer que les variables

$$Y = \frac{1}{X} \qquad \text{et} \qquad Z = \frac{1+X}{1-X}$$

suivent encore une loi de Cauchy $\mathscr{C}(0\,;\,1)$.

♦ ***Exercice 10.5 (Lois sans vieillissement)*** Soit $G : \mathbb{R}^+ \to \mathbb{R}$ une fonction *continue* et non nulle, vérifiant de plus $G(r+s) = G(r)\,G(s)$ pour tous $r, s \geqslant 0$. Montrer qu'il existe un réel α tel que $G(x) = e^{-\alpha x}$.

♦ ***Exercice 10.6 (Loi de Poisson et loi exponentielle)*** On considère un événement[(14)] qui se produit à des instants aléatoires $0 \leqslant T_1 < T_2 < T_3 < \ldots$, tels que : pour tout intervalle de temps $[t\,;t+a]$, le nombre d'événements qui se produisent entre t et $t+a$ suit une loi de Poisson de paramètre λa. Quelle est la loi suivie par T_1 ? Par T_n ? (On calculera la fonction de répartition et la densité de T_n, et on comparera au résultat de l'exercice 12.10.)

♦ ***Exercice 10.7*** Soit A une variable aléatoire de loi exponentielle de paramètre $\alpha > 0$. On note Q le polynôme $Q(X) = X^2 + 2X(A-2) + 2A - 4$. Quelle est la probabilité que Q ait ses racines réelles ?

♦ ***Exercice 10.8*** Soit X une variable aléatoire suivant une loi exponentielle $\mathscr{E}(\alpha)$. Calculer une densité de $Y = \exp(X)$.

♦ ***Exercice 10.9 (Loi de l'Arc sinus)*** On dit qu'une variable aléatoire X suit la **loi de l'Arc sinus** $\mathscr{A}$ si elle est à valeurs dans $]-1\,;+1[$ et admet pour densité

$$f(x) = \frac{1}{\pi} \frac{1}{\sqrt{1-x^2}}.$$

i) Vérifier que f est bien une densité de probabilité et expliciter la fonction de répartition associée.

ii) Soit Θ une variable aléatoire suivant la loi uniforme sur $]0\,;2\pi[$. Montrer que les variables aléatoires $X = \cos\Theta$ et $Y = \sin\Theta$ suivent la loi $\mathscr{A}$ de l'Arc sinus. Montrer que les variables aléatoires $\cos 2\Theta$ et $\sin 2\Theta$ suivent la loi $\mathscr{A}$. En déduire que les variables aléatoires $X^2 - Y^2$ et $2XY$ suivent également la loi $\mathscr{A}$.

iii) Soit C une variable aléatoire suivant la loi de Cauchy $\mathscr{C}(0\,;\,1)$. Montrer que la variable aléatoire $A = \dfrac{1-C^2}{1+C^2}$ suit la loi de l'Arc sinus.

---

(14). Au sens usuel du terme. Typiquement, on peut considérer que l'instant T_k est celui de la k-ième désintégration dans une quantité macroscopique de matière radioactive, où l'instant d'arrivée de la k-ième voiture à un péage autoroutier.

Indications

◊ *Indication 10.5 :* Commencer par montrer que $G(0) \neq 0$, puis que $G(1) > 0$. Poser ensuite $\alpha = -\ln G(1)$ et montrer successivement que $G(x) = e^{-\alpha x}$: pour les entiers, pour les rationnels, et enfin pour tous les réels, en utilisant la continuité de G et la densité des rationnels.

◊ *Indication 10.6 :* Si l'on note N_t le nombre d'événements qui se sont produits dans l'intervalle $[0\,;t]$, interpréter $\{T_1 > t\}$ en fonction de N.

◊ *Indication 10.7 :* Calculer la probabilité que le discriminant soit positif.

◊ *Indication 10.9 : ii)* : revoir les figures du paragraphe § 25 page 58.

SOLUTIONS

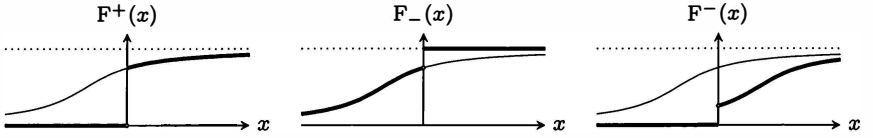
♦ **Solution 10.1** Par définition,

$$F^+(x) = \mathbf{P}\{X^+ \leqslant x\} = \begin{cases} \mathbf{P}\{X \leqslant x\} = F(x) & \text{si } x \geqslant 0, \\ 0 & \text{si } x < 0, \end{cases}$$

$$F_-(x) = \mathbf{P}\{X_- \leqslant x\} = \begin{cases} \mathbf{P}\{X \leqslant x\} = 1 & \text{si } x \geqslant 0, \\ F(x) & \text{si } x < 0, \end{cases}$$

$$F^-(x) = \mathbf{P}\{X^- \leqslant x\} = \begin{cases} \mathbf{P}\{X \geqslant -x\} = 1 - F(-x) & \text{si } x \geqslant 0, \\ 0 & \text{si } x < 0, \end{cases}$$

Voici les graphes correspondants avec, en trait mince, la fonction de répartition F de X.



♦ **Solution 10.2** La fonction de répartition de X est nulle pour $x < 0$, et vaut, pour $x \geqslant 0$, $F_X(x) = \mathbf{P}\{X \leqslant x\} = 1 - e^{-x}$. Pour tout $y > 0$, on a $F_Y(y) = \mathbf{P}\{Y \leqslant y\} = \mathbf{P}\{X \leqslant y/\lambda\} = 1 - e^{-y/\lambda}$ ce qui prouve que $Y \rightsquigarrow \mathscr{E}(1/\lambda)$.

♦ **Solution 10.3** On vérifie que $Y(\Omega) = [0\,;1[$ et que, pour tout $y \in [0\,;1[$, on a

$$F_Y(y) := \mathbf{P}\{Y \leqslant y\} = \sum_{n=1}^{\infty} \mathbf{P}\left\{n \leqslant \frac{1}{X} \leqslant n+y\right\} = \sum_{n=1}^{\infty} \mathbf{P}\left\{\frac{1}{n+y} \leqslant X \leqslant \frac{1}{n}\right\}$$

$$= \frac{1}{\ln 2} \sum_{n=1}^{\infty} \left[\ln\left(\frac{n+1}{n}\right) - \ln\left(\frac{n+y+1}{n+y}\right)\right].$$

Pour calculer cette somme télescopique, on passe par les sommes partielles :

$$\sum_{n=1}^{N} [\ln(n+1) - \ln(n) - \ln(n+y+1) + \ln(n+y)] = \ln\left(\frac{N+1}{N+y+1}\right) + \ln(1+y)$$

$$\xrightarrow[N\to+\infty]{} \ln(1+y)$$

Ainsi, pour tout $y \in [0\,;1[$:

$$F_Y(y) = \frac{\ln(1+y)}{\ln(2)}$$

On en déduit que Y aussi admet f comme densité de probabilité : X *et* Y *suivent la même loi.*

♦ **Solution 10.4** Si U est uniforme sur $]-\pi/2\,;+\pi/2[$, alors $T = \tan U$ suit la loi de Cauchy $\mathscr{C}(0\,;1)$, et il en est de même si U suit une loi uniforme sur n'importe quel intervalle de largeur π. Les relations

$$\frac{1}{\tan x} = \tan\left(\frac{\pi}{2} - x\right) \qquad \text{et} \qquad \frac{1+\tan x}{1-\tan x} = \tan\left(\frac{\pi}{4} + x\right)$$

permettent de conclure.

Bien évidemment, on peut également procéder plus pédestrement en calculant les fonctions de répartition de Y et de Z. Traitons simplement le premier cas ici. La variable Y prend ses valeurs dans $\mathbb{R}$ avec une probabilité 1, puisque $\mathbf{P}\{X = 0\} = 0$. On cherche à évaluer $\mathbf{P}\{Y \leqslant y\}$. Séparons les cas selon le signe de y, en notant F la fonction de répartition de X :

— si $y > 0$, alors (voir figure 10.11) :

$$\begin{aligned}\mathbf{P}\{Y \leqslant y\} &= \mathbf{P}\left\{\frac{1}{X} \leqslant y\right\} = \mathbf{P}\{X \leqslant 0\} + \mathbf{P}\left\{X \geqslant \frac{1}{y}\right\} \\ &= \frac{1}{2} + 1 - \left(\frac{1}{2} + \frac{1}{\pi}\operatorname{Arc\,tan}\frac{1}{y}\right) = 1 - \frac{1}{\pi}\left(\frac{\pi}{2} - \operatorname{Arc\,tan} y\right) \\ &= \frac{1}{2} + \frac{1}{\pi}\operatorname{Arc\,tan} y,\end{aligned}$$

ce qui est le résultat attendu ;

— de même, si $y < 0$,

$$\begin{aligned}\mathbf{P}\{Y \leqslant y\} &= \mathbf{P}\left\{\frac{1}{X} \leqslant y\right\} = \mathbf{P}\left\{\frac{1}{y} \leqslant X < 0\right\} \\ &= F(0) - F\left(\frac{1}{y}\right) = \frac{1}{2} - \frac{1}{2} - \frac{1}{\pi}\operatorname{Arc\,tan}\frac{1}{y} \\ &= -\frac{1}{\pi}\left(-\frac{\pi}{2} - \operatorname{Arc\,tan} y\right) = \frac{1}{2} + \frac{1}{\pi}\operatorname{Arc\,tan} y\end{aligned}$$

comme il se doit. (On a utilisé la formule $\operatorname{Arc\,tan} y + \operatorname{Arc\,tan}(1/y) = \operatorname{sgn}(y) \cdot \pi/2$.)

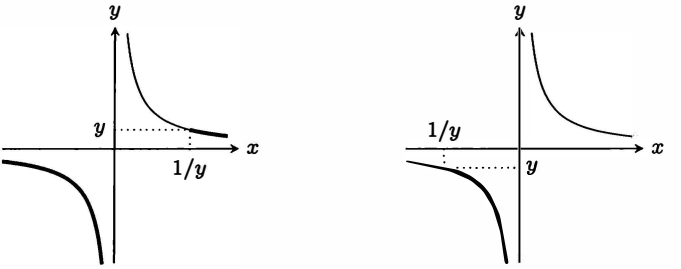


FIGURE 10.11 — *Détermination de* $\{Y \leqslant y\}$ *lorsque* $y > 0$ (à gauche) *et lorsque* $y < 0$ (à droite).

♦ **Solution 10.5** Prenons $r = s = 0$, on obtient $G(0) = G(0)^2$, et donc $G(0) \in \{0, 1\}$. Si $G(0) = 0$, alors la relation fonctionnelle montre que $G(x) = G(x)\,G(0) = 0$ pour tout $x \geqslant 0$, et donc que G est la fonction nulle, ce que l'on a supposé être faux. Ainsi, $G(0) = 1$.

Ensuite, $G(1) = G(1/2)^2 \geqslant 0$. Si l'on suppose que $G(1) = 0$, alors

$$0 = G(1) = G\left(\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}\right) = G\left(\frac{1}{n}\right)^n$$

par simple récurrence sur n, et donc $G(1/n) = 0$ pour tout $n \in \mathbb{N}^*$; par continuité de G en 0, on obtient $G(0) = 0$ et une contradiction : ainsi, $G(1) > 0$; on note alors $\alpha := -\ln G(1)$.

Une autre récurrence montre que $G(p) = G(1)^p = e^{-\alpha p}$, puis $G(1/q) = G(1)^{1/q}$ et $G(p/q) = G(1)^{p/q} = e^{-\alpha p/q}$ pour tous $p \in \mathbb{N}$ et $q \in \mathbb{N}^*$. Ainsi, $G(r) = e^{-\alpha r}$ pour tout rationnel $r \geqslant 0$. Par densité de $\mathbb{Q}^+$ dans $\mathbb{R}^+$, si x est un réel, on peut l'écrire comme limite d'une suite $(r_n)_{n\geqslant 0}$ de rationnels. La continuité de G en x permet alors d'écrire que

$$G(x) = \lim_{n\to\infty} G(r_n) = \lim_{n\to\infty} e^{-\alpha r_n} = e^{\alpha x},$$

ce qui achève la preuve.

♦ **Solution 10.6** Dire que $T_1 > t$ équivaut à dire qu'à l'instant t, il n'y a eu aucun événement ; ainsi, $\{T_1 > t\} = \{N_t = 0\}$. La fonction de répartition de T_1 se calcule donc comme

$$F_1(t) = 1 - \mathbf{P}\{T_1 > t\} = 1 - e^{-\lambda t}.$$

Ainsi, T_1 suit une loi géométrique $\mathscr{E}(\lambda)$.

De même, $\{T_2 > t\} = \{N_t \leqslant 1\}$ donc

$$F_2(t) = 1 - \mathbf{P}\{T_2 > t\} = 1 - e^{-\lambda t} - \lambda t\, e^{-\lambda t}$$

et, de manière générale,

$$F_n(t) = 1 - \mathbf{P}\{T_n > t\} = \{N_t \leqslant n-1\} = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$

Si l'on dérive, les termes s'annulent deux à deux et ne reste que

$$f_n(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t},$$

ce qui est la loi densité de la somme de n variables aléatoires indépendantes de même loi $\mathscr{E}(t)$ (exercice 12.10). Cela vient de la relation profonde entre loi de Poisson et lois exponentielles. Si l'on fixe l'origine de l'expérience juste après la premier événement, on voit que $T_2 - T_1$ suit la même loi que T_1. De même, on peut montrer que les $T_k - T_{k-1}$ sont indépendants et de même loi $\mathscr{E}(\lambda)$. Ces liens sont utilisés pour modéliser informatiquement une loi de Poisson (§ 369) ; ils sont approfondis dans le chapitre 30.

♦ **Solution 10.7** La probabilité cherchée est celle que le discriminant soit positif. Le discriminant (réduit) vaut $\Delta' = A^2 - 6A + 8$. Étudions le signe de l'expression $\Delta(a) = a^2 - 6a + 8$. Son discriminant réduit étant $\delta' = 1$, ses racines sont $a' = 4$ et $a'' = 2$, ce qui montre que $\{\Delta \geqslant 0\} = \{A \notin [2;4]\}$, ce qui arrive avec la probabilité $p = \mathbf{P}\{A \leqslant 2\} + \mathbf{P}\{A \geqslant 4\} = 1 - e^{-2\alpha} + e^{-4\alpha}$.

♦ **Solution 10.8** $X(\Omega) = [0; +\infty[$ donc $Y(\Omega) = \exp([0; +\infty[) = [1; +\infty[$. La fonction de répartition de Y est donnée, pour tout $y \geqslant 1$, par

$$F_Y(y) = \mathbf{P}\{Y \leqslant y\} = \mathbf{P}\{X \leqslant \ln y\} = 1 - \frac{1}{y^\alpha}.$$

Ainsi, Y admet pour densité

$$f(y) = \begin{cases} \alpha y^{-(\alpha+1)} & \text{si } y \geqslant 1 \\ 0 & \text{sinon.} \end{cases}$$

Chapitre 11

Espérance d'une variable aléatoire à densité

Par analogie avec le cas discret, nous donnons ici une définition de l'espérance d'une variable aléatoire à densité. La plupart des propriétés de l'espérance — linéarité, formule de transfert — sont provisoirement admises. Elles seront toutes justifiées dans le chapitre 14, qui requiert une approche théorique plus élaborée. Ici, l'accent est mis sur l'utilisation pratique de ces propriétés.

11.1 ESPÉRANCE ET MOMENTS

Espérance § 164

Comme nous l'avons vu dans le chapitre 2, la notion d'espérance généralise, pour une variable aléatoire quelconque, la notion de moyenne d'une variable aléatoire finie. Reprenons un point de vue « heuristique[1] ». Soit X une variable aléatoire admettant une densité f. La probabilité $\mathrm{d}p$ que la variable X soit comprise entre x et $x + \mathrm{d}x$ est $\mathrm{d}p \approx f(x)\,\mathrm{d}x$; en sommant les valeurs de X pondérées par ces probabilités, on définit la valeur moyenne de X comme

$$\int_{-\infty}^{+\infty} x\, f(x)\, \mathrm{d}x.$$

Les contraintes vues au paragraphe § 33 étant *a fortiori* encore valables, on est amenés à poser plus précisément la définition suivante :

Définition 11.1 Soit X une variable aléatoire réelle, admettant une densité f. On dit que **X admet une espérance** si la fonction $x \mapsto |x|\ f(x)$ est intégrable, ce qui revient à dire que les deux intégrales

$$\int_{-\infty}^{0} x\, f(x)\, \mathrm{d}x \qquad \text{et} \qquad \int_{0}^{+\infty} x\, f(x)\, \mathrm{d}x$$

(1). C'est-à-dire le point de vue du physicien, ou du mathématicien de l'époque de ténèbres précédant la théorie de la mesure et son cortège de tribus, mais durant laquelle il fallait tout de même avancer.

sont convergentes. Si c'est le cas, l'**espérance de X** est

$$\mathbf{E}(X) := \int_{-\infty}^{+\infty} x\, f(x)\, dx.$$

Exemple 11.2 Soit X une variable aléatoire suivant une loi exponentielle de paramètre $\alpha > 0$. La densité de f étant nulle sur $\mathbb{R}^-$ et la fonction $x \mapsto \alpha x e^{-\alpha x}$ étant intégrable sur $\mathbb{R}^+$, X admet une espérance et

$$\mathbf{E}(X) = \int_0^{+\infty} \alpha x\, e^{-\alpha x}\, dx = \frac{1}{\alpha}.$$

On admettra la linéarité de l'espérance :

■ **Théorème 11.3** *Si* X *et* Y *sont des variables aléatoires à densité admettant une espérance, si* $\lambda X + \mu Y$ *admet également une densité, alors* $\lambda X + \mu Y$ *admet une espérance et*

$$\mathbf{E}(\lambda X + \mu Y) = \lambda\, \mathbf{E}(X) + \mu\, \mathbf{E}(Y).$$

Cette propriété, pourtant très simple, se révèle étonnamment délicate à démontrer proprement dans le cadre du point de vue adopté ici, qui est de définir l'espérance *via* la fonction de densité. Une première démonstration heuristique en sera donnée au paragraphe § 178 ; dans le chapitre 14, on (re)définira l'espérance d'une variable aléatoire comme une intégrale selon une mesure de probabilité, point de vue certes plus abstrait, mais qui permettra de prouver la linéarité de l'espérance de manière très générale (il ne sera plus nécessaire de supposer que $\lambda X + \mu Y$ admet une densité, par exemple).

§ 165 **Formule de transfert**

Soient X une variable aléatoire admettant une densité f, et $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction réelle, définie au moins sur $X(\Omega)$. La fonction composée $\varphi \circ X : \Omega \to \mathbb{R}$ est généralement notée $\varphi(X)$; à tout ω elle associe $\varphi\big(X(\omega)\big)$. C'est une variable aléatoire[(2)], mais qui n'a pas de raison d'admettre une densité ; sa loi est généralement compliquée à expliciter (à part quelques cas simples, voir la section 10.3). Son espérance est toutefois beaucoup plus facile à calculer, grâce à la formule de transfert, qui généralise la formule équivalente pour les variables aléatoires discrètes.

■ **Théorème 11.4 (Théorème de transfert)** *La variable aléatoire* $\varphi(X)$ *admet une espérance si et seulement si* φf *est intégrable. Si c'est le cas,*

$$\mathbf{E}\big[\varphi(X)\big] = \int_{-\infty}^{+\infty} \varphi(x)\, f(x)\, dx.$$

Nous admettons momentanément ce théorème, dont une démonstration requiert un minimum de théorie de l'intégration selon une mesure (voir chapitre 14, paragraphe § 217). Dans des cas simples (φ est une bijection strictement croissante de $\mathbb{R}$ dans $\mathbb{R}$), la formule de transfert se démontre aisément (exercice 11.4).

⏢11.3 ⏢11.5

(2). Comme déjà dit : il suffit que φ soit une fonction mesurable pour que ce soit le cas. En effet, par définition d'une fonction mesurable, pour tout borélien B de $\mathbb{R}$, $A := \varphi^{-1}(B)$ est un borélien, et

$$\{\varphi(X) \in B\} = (\varphi \circ X)^{-1}(B) = X^{-1}(A) = \{X \in A\}$$

est un borélien. Le cas des fonctions φ non mesurables n'est pas abordé ici.

Moments

§ 166

Soit X une variable aléatoire admettant une densité f. Pour tout entier k, on dit que **X admet un moment d'ordre k** si la variable aléatoire X^k admet une espérance ; c'est le cas si et seulement si $x \mapsto |x|^k f(x)$ est intégrable. Dans ce cas, le **moment d'ordre k de X** est

$$m_k(X) := \mathbf{E}(X^k) = \int_{-\infty}^{+\infty} x^k f(x)\,dx.$$

Remarque 11.5 Le moment d'ordre 0 d'une variable aléatoire existe toujours et vaut $\mathbf{E}(X^0) = \mathbf{E}(1) = 1$. ✦ Le moment d'ordre 1, s'il existe, est l'espérance de X.

■ **Théorème 11.6** *Soit* X *une variable aléatoire réelle à densité, et soit* $k \geqslant 2$. *Si* X *admet un moment d'ordre* $k \geqslant 2$, *alors* X *admet des moments d'ordre* $1, 2, \ldots, k$.

DÉMONSTRATION : Notons f une densité de X ; on suppose donc $x \mapsto |x|^k f(x)$ est intégrable. Soit ℓ un entier compris entre 1 et $k-1$. De l'inégalité

$$\forall x \in [-1\,;+1] \qquad 0 \leqslant |x|^\ell f(x) \leqslant f(x)$$

et de l'intégrabilité de f sur $[-1\,;+1]$, on déduit que $x \mapsto |x|^\ell f(x)$ est intégrable sur $[-1\,;+1]$. De l'inégalité

$$\forall x \in]-\infty\,;-1] \cup [+1\,;+\infty[\qquad 0 \leqslant |x|^\ell f(x) \leqslant |x|^k f(x)$$

on déduit l'intégrabilité de $x \mapsto |x|^\ell f(x)$ sur $\mathbb{R} \setminus [-1\,;+1]$. Ainsi, $x \mapsto |x|^\ell f(x)$ est intégrable sur $\mathbb{R}$ et X admet un moment d'ordre ℓ. ∎

Remarque 11.7 (Nullité des moments impairs d'une variable symétrique)
Soit X une variable aléatoire symétrique, c'est-à-dire admettant une densité f paire. Si X admet un moment d'ordre $2k+1$ impair, alors ce moment, intégrale d'une fonction impaire et intégrable, est nul : $\mathbf{E}(X^{2k+1}) = 0$. Bien évidemment, si X n'admet pas de moment d'ordre $2k+1$, on ne peut rien conclure.

Moments centrés

§ 167

Soit X une variable aléatoire à densité, admettant une espérance. On dit que X admet un **moment centré d'ordre k** si la variable aléatoire centrée $X' := X - \mathbf{E}(X)$ admet un moment d'ordre k. Le moment centré d'ordre k de X est alors

$$\mu_k(X) := m_k(X') = \mathbf{E}(X'^k) = \mathbf{E}\big[(X - \mathbf{E}X)^k\big].$$

■ **Proposition 11.8** *Soit* X *une variable aléatoire à densité. Alors* X *admet un moment centré d'ordre* k *si et seulement si elle admet un moment d'ordre* k.

DÉMONSTRATION : Notons $m = \mathbf{E}(X)$. Supposons dans un premier temps que X admet un moment d'ordre k, et développons $(X-m)^k$:

$$(X-m)^k = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} X^i m^{k-i}.$$

Le théorème 11.6 montre que X admet des moments d'ordre 0 à k ; par linéarité, l'expression de droite admet donc une espérance.

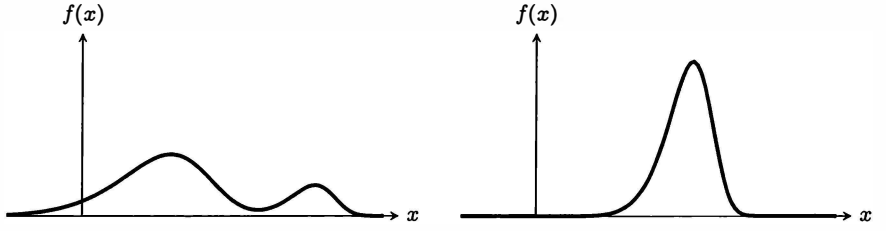
Réciproquement, si X admet un moment centré d'ordre k, cela implique que $(X-m)$ admet des moments de tous ordres entre 1 et k, et donc que

$$X^k = (X - m + m)^k = \sum_{i=0}^{k} \binom{k}{i} (X-m)^i m^{k-i}$$

admet une espérance. ∎

§ 168 **Variance**

Si la loi d'une variable aléatoire X n'est déjà qu'une vision partielle de X en tant qu'application (on ne s'intéresse qu'à la distribution de ses valeurs), la connaissance unique de son espérance est évidemment encore plus réducteur : des distributions très différentes peuvent avoir la même valeur moyenne. Sur les figures ci-dessous, l'étalement des valeurs possibles de X est manifeste sur le graphe de gauche, tandis que celui de droite présente un pic étroit.



Cette différence qualitative peut se quantifier en calculant les *écarts à la moyenne* $m := \mathbf{E}(\mathrm{X})$. Comme dans le cas des variables aléatoires discrètes (§ 93), la moyenne des écarts à la moyenne est, par définition, nulle[3] et la quantité intéressante est la moyenne des écarts *quadratiques* à la moyenne, c'est-à-dire l'espérance de $(\mathrm{X}-m)^2$.

On dira qu'une variable aléatoire réelle X, admettant une densité f, **admet une variance** si elle admet un moment d'ordre 2, c'est-à-dire si $x \mapsto x^2 f(x)$ est intégrable. Dans ce cas, la **variance** de X est son moment centré d'ordre 2 :

$$\mathbf{V}(\mathrm{X}) := \mu_2(\mathrm{X}) = \mathbf{E}\big([\mathrm{X} - \mathbf{E}(\mathrm{X})]^2\big).$$

On dira parfois, de manière équivalente, que X **est de carré intégrable**. Si X admet une variance, son **écart-type** est la racine carrée de sa variance :

$$\sigma_{\mathrm{X}} = \sqrt{\mathbf{V}(\mathrm{X})}.$$

De manière analogue au cas discret (puisque finalement il ne s'agit que d'utiliser la linéarité de l'espérance), la variance d'une variable aléatoire s'exprime plus commodément sous la forme suivante :

■ **Proposition 11.9 (König-Huygens)** *Si* X *est une variable aléatoire à densité admettant une variance, alors*

$$\mathbf{V}(\mathrm{X}) = \mathbf{E}(\mathrm{X}^2) - \mathbf{E}(\mathrm{X})^2.$$

Enfin, par changement d'échelle, on a les règles de calcul suivantes :

■ **Proposition 11.10** *Soit* X *une variable aléatoire à densité, admettant une variance. Soient a et b des réels. Alors*

$$\mathbf{V}(a\mathrm{X} + b) = a^2\, \mathbf{V}(\mathrm{X}) \qquad \textit{et} \qquad \sigma_{a\mathrm{X}+b} = |a|\, \sigma_{\mathrm{X}}.$$

La démonstration a déjà été vue dans le chapitre 7 et ne fait pas intervenir la nature (discrète ou absolument continue) de la variable aléatoire.

---

(3). Si l'on note $m = \mathbf{E}(\mathrm{X}) = \int x\, f(x)\, \mathrm{d}x$, alors

$$\mathbf{E}(\mathrm{X} - m) = \int_{-\infty}^{+\infty} (x - m)\, f(x)\, \mathrm{d}x = \int_{-\infty}^{+\infty} x\, f(x)\, \mathrm{d}x - m \int_{-\infty}^{+\infty} f(x)\, \mathrm{d}x = m - m = 0.$$

Remarque 11.11 (Stricte positivité de la variance) Remarquons bien si X est absolument continue et admet une variance, alors $\mathbf{V}(X) > 0$; en effet, on peut écrire, en notant $m = \mathbf{E}(X)$,

$$\mathbf{V}(X) = \mathbf{E}[(X-m)^2] = \int_{-\infty}^{+\infty} (x-m)^2 f(x)\,dx.$$

Cette intégrale ne peut être nulle que si f est presque partout nulle, ce qui est exclu.

Inégalité de Cauchy-Schwarz

§ 169

Nous redonnons l'inégalité de Cauchy-Schwarz, déjà vue au chapitre 7[4].

■ **Théorème 11.12 (Cauchy-Schwarz)** *Soient* X *et* Y *deux variables aléatoires à densité admettant une variance. Alors* XY *admet une espérance et*

$$\mathbf{E}(XY)^2 \leqslant \mathbf{E}(X^2) \cdot \mathbf{E}(Y^2).$$

Nous admettrons le résultat suivant, déjà démontré dans le cas de variables discrètes, mais que l'on ne peut plus traiter dans le cadre des variables à densité[5] :

■ **Proposition 11.13 (Cas d'égalité)** *Il n'y a égalité dans l'inégalité précédente que si les variables centrées* $X' := X - \mathbf{E}X$ *et* $Y' := Y - \mathbf{E}Y$ *sont linéairement dépendantes, de façon presque sûre. Puisque ni* X′ *ni* Y′ *n'est nulle, cela équivaut à : il existe un scalaires* λ *tels que* $Y' = \lambda X'$ *presque sûrement, ou encore*

$$\mathbf{P}\{Y' = \lambda X'\} = 1.$$

Remarque 11.14 En appliquant l'inégalité de Cauchy-Schwarz aux variables aléatoires |X| et |Y|, on a également

$$(\mathbf{E}|XY|)^2 \leqslant \mathbf{E}|X^2| \cdot \mathbf{E}|Y^2|.$$

Notamment, en choisissant $Y = 1$, on en déduit que, si X admet une variance, alors X admet une espérance et

$$|\mathbf{E}(X)| \leqslant \mathbf{E}|X| \leqslant \sqrt{\mathbf{E}(X^2)}.$$

Espérance et variance des lois classiques

§ 170

On résume dans la table 11.1 page suivante les valeurs des espérances et variances des lois usuelles (quand elles sont définies). Le calcul est à chaque fois élémentaire et est laissé en exercice au lecteur.

Remarque 11.15 (La loi de Cauchy n'a pas d'espérance) Il peut sembler étonnant d'affirmer qu'une variable aléatoire X suivant une loi de Cauchy $\mathscr{C}(m\,;\,a)$ n'admet pas d'espérance. Pour simplifier, prenons $m = 0$ et $a = 1$; la fonction de densité f est alors paire, et il peut être tentant de définir l'espérance de X dans le sens suivant :

$$\lim_{M\to+\infty} \frac{1}{\pi} \int_{-M}^{M} \frac{x}{1+x^2}\,dx,$$

ce qui laisserait penser que 0 est un bon candidat pour l'espérance de X, au moins dans un sens élargi ; après tout, la probabilité que X prenne ses valeurs entre x et $x + dx$ est la même que la probabilité qu'elle prenne ses valeurs entre $-x$ et $-(x + dx)$. Comme déjà vu (§ 34), il faut s'attendre à une mauvaise concordance entre la valeur putative « 0 » et son interprétation en terme de « moyenne statistique ». La figure ci-après montre, pour trois réalisations ω du hasard, les moyennes successives $S_n = (X_1 + \cdots + X_n)/n$, où les X_k ($k \geqslant 1$) sont des variables aléatoires

(4). La démonstration que nous en avions donnée étant générale, elle s'applique encore ici, à condition d'admettre momentanément que si X et Y admettent une variance, alors XY admet une espérance. La démonstration la plus générale de cette inégalité est présentée au chapitre 14, paragraphe § 219.

(5). Dans la démonstration, on prouve que $Y - \lambda X$ a une variance nulle ; ce n'est donc pas une variable à densité.

Loi	Désignation	X(Ω)	Densité $f(x)$	$\mathbf{E}(X)$	$\mathbf{V}(X)$
uniforme	$\mathscr{U}([a\,;b])$	$[a\,;b]$	$\dfrac{1}{b-a}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
normale	$\mathscr{N}(m\,;\sigma^2)$	$\mathbb{R}$	$\dfrac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi}\,\sigma}$	m	σ^2
exponentielle	$\mathscr{E}(\alpha)$	$\mathbb{R}^+$	$\alpha\,e^{-\alpha x}$	$\dfrac{1}{\alpha}$	$\dfrac{1}{\alpha^2}$
de Cauchy	$\mathscr{C}(m\,;a)$	$\mathbb{R}$	$\dfrac{a}{\pi}\dfrac{1}{a^2+(x-m)^2}$	/	/
Gamma	$\Gamma(\alpha,\beta)$	$\mathbb{R}_+^*$	$\dfrac{\beta^\alpha}{\Gamma(\alpha)}\,x^{\alpha-1}\,e^{-\beta x}$	$\dfrac{\alpha}{\beta}$	$\dfrac{\alpha}{\beta^2}$
gamma	$\gamma(\nu)$	$\mathbb{R}_+^*$	$\dfrac{1}{\Gamma(\nu)}\,x^{\nu-1}\,e^{-x}$	ν	ν
khi-deux	χ^2_p	$\mathbb{R}_+^*$	$\dfrac{x^{p/2-1}\,e^{-x/2}}{2^{p/2}\,\Gamma(p/2)}$	p	$2p$
Bêta	$B(a,b)$	$]0\,;1[$	$\dfrac{x^{a-1}\,(1-x)^{b-1}}{B(a,b)}$	$\dfrac{a}{a+b}$	$\dfrac{ab}{(a+b)^2\,(a+b+1)}$

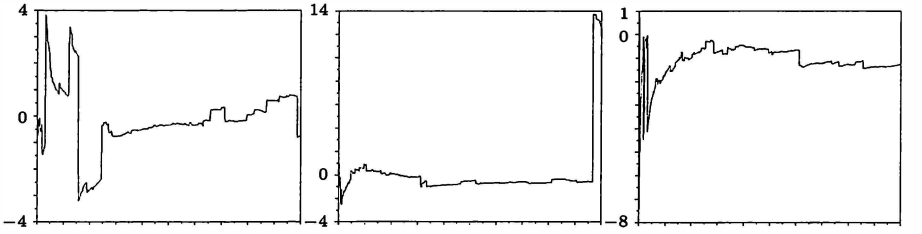
TABLE 11.1 — *Espérances et variances des lois classiques à densité.*

indépendantes de même loi de Cauchy que X ; ces moyennes sont respectivement calculées pour $n = 1000$, 10 000 et 100 000 tirages aléatoires d'une loi de Cauchy, et obtenues par le code suivant (langage Scilab[6]) :

```
function y=cauchy()
  x=rand()
  y=tan(%pi*(x-1/2))
endfunction
```

```
function moyennes(n)
  Y=[] ;  S=0
  for i=1:n
    y=cauchy() ; S=S+y ; Y=[Y,S/i]
  end
  clf ; plot(1:n,Y)
endfunction
```

Simulation de $\mathscr{C}(0\,;1)$ *Affichages des moyennes*



$n = 1000$ $n = 10\,000$ $n = 100\,000$

On notera la « mauvaise convergence » apparente de la suite de ces moyennes : les accidents (la variable X prend une valeur gigantesque) ne sont pas assez rares et modifient la moyenne observée de manière erratique ; ce phénomène est d'ailleurs confirmé par un théorème précis, assurant qu'avec une probabilité 1, la suite de ces moyennes passera, une infinité de fois, au-dessus de n'importe quel réel $M \geqslant 0$ donné, et en dessous de n'importe quel réel $M' \leqslant 0$ donné (même si les temps d'attente pour voir arriver ces événements sont astronomiquement grands).

(6). Nous n'avons pas essayé de tirer profit des fonctionnalités particulières de Scilab ; ce code est transposable, *mutatis mutandis*, dans n'importe quel autre langage.

Cette mauvaise convergence s'explique, dans le cas particulier des lois de Cauchy $\mathscr{C}(0\,;\,a)$, par le fait que la loi de S_n/n est la même que celle de X (voir exercice 17.4). En d'autres termes, *effectuer un seul tirage selon la loi de* X, *ou effectuer un moyenne sur un nombre arbitraire de tirages, donne exactement les mêmes renseignements sur cette loi !*

Les paramètres m et a de la loi de Cauchy $\mathscr{C}(m\,;\,a)$ peuvent toutefois recevoir une interprétation. Le réel m est l'abscisse du centre de symétrie de la fonction de répartition, et celle de l'axe de symétrie de la densité : c'est donc la *médiane* de la loi ou encore, la fonction de répartition étant continue strictement croissante, l'unique réel vérifiant

$$\mathbf{P}\{X \leqslant m\} = \frac{1}{2}.$$

Quant à a, il mesure la *dispersion* de la loi (dispersion qui est trop grande pour que la variance existe, puisque déjà l'espérance n'existe pas) ; plus précisément, $m + a$ est le **troisième quartile** de la loi, c'est-à-dire le réel vérifiant

$$\mathbf{P}\{X \leqslant m + a\} = \frac{3}{4}.$$

Le réel $m - a$ est le **premier quartile** : $\mathbf{P}\{X \leqslant m - a\} = 1/4$. L'**écart inter-quartile** vaut $2a$; cette quantité, souvent utilisée en statistique descriptive, offre une mesure de la dispersion même quand la variance n'est pas définie (voir figure 11.1).

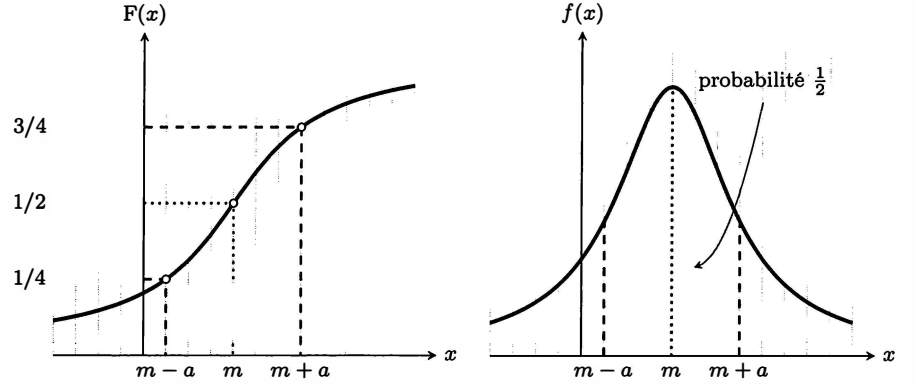


FIGURE 11.1 — *La médiane et les quartiles d'une loi de Cauchy* $\mathscr{C}(m\,;\,a)$.

Remarque 11.16 (Durée de vie et demi-vie) Revenons à l'exemple 2.10 page 62 de la désintégration d'un atome radioactif. Nous avons vu que, si un tel atome ne s'est pas désintégré à l'instant $t = 0$, alors le temps écoulé avant sa désintégration ultime peut être modélisé par une variable aléatoire X de loi exponentielle $\mathscr{E}(\alpha)$.

L'espérance de X porte alors le nom de **durée de vie** de l'atome :

$$\tau = \mathbf{E}(X) = \int_0^{+\infty} t\, f(t)\, dt = \alpha \int_0^{+\infty} t\, e^{-\alpha t}\, dt = \frac{1}{\alpha} \qquad \text{(durée de vie)}.$$

Pour un physicien, c'est bien sûr le réel τ qui sert à caractériser la loi de X et sa densité $f(t) = e^{-t/\tau}/\tau$.

La médiane, c'est-à-dire l'unique instant θ vérifiant $F(\theta) = 1/2$, est appelée **demi-vie** ; elle vérifie $e^{-\alpha\theta} = 1/2$, ou encore

$$\theta = \frac{\ln 2}{\alpha} = \tau \ln 2 \qquad \text{(demi-vie)}.$$

Sur un échantillon de N atomes ($N \gg 1$), la demi-vie est la durée durant laquelle, en moyenne, la moitié des atomes se désintègre.

§ 171 **Variables centrées réduites**

Si X est une variable aléatoire à densité admettant une espérance, alors $X - \mathbf{E}(X)$ est une variable aléatoire **centrée**, c'est-à-dire d'espérance nulle. Si X admet un moment d'ordre 2, alors la variable aléatoire

$$X^* := \frac{X - \mathbf{E}(X)}{\sigma_X} = \frac{X - m}{\sigma}$$

est centrée et **réduite** (d'espérance nulle et d'écart-type égale à 1). Si f est la densité de X, alors celle de X^* est

$$f^*(x) = \sigma \cdot f\big(\sigma(x + m)\big).$$

§ 172 **Autres caractéristiques numériques d'une variable aléatoire à densité**

Les moments d'ordre 3 et 4 d'une variable aléatoire absolument continue, s'ils existent, permettent de définir deux nouvelles quantités qui rendent comptent de la forme globale de f. Le **coefficient d'asymétrie** est la quantité

$$\gamma_1 := \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3},$$

où $\mu_3 = \mathbf{E}(X - \mathbf{E}X)^3$ est le moment centré d'ordre trois de X, et $\mu_2 = \sigma^2$ celui d'ordre deux, c'est-à-dire sa variance. Si la distribution de f est plus allongée à droite de $\mathbf{E}(X)$ qu'elle ne l'est à gauche, ce coefficient est positif ; il est négatif dans le cas contraire (voir figure 11.2).

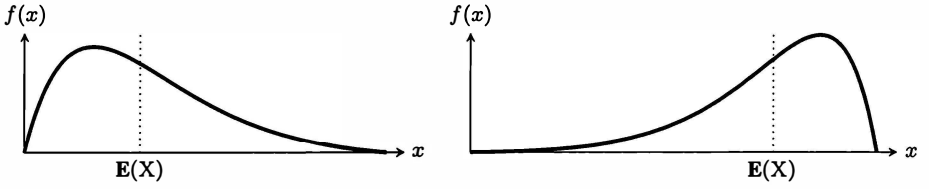


FIGURE 11.2 — À gauche : *coefficient d'asymétrie positif.* À droite : *coefficient d'asymétrie négatif.*

Le **coefficient d'aplatissement**(7) est

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}.$$

On le normalise parfois en rajoutant la constante -3 :

$$\gamma_2 := \beta_2 - 3 = \frac{\mu_4}{\sigma^4} - 3.$$

De cette manière, le coefficient d'aplatissement normalisé d'une loi gaussienne est nul. Si $\gamma_2 > 0$, cela signifie (en gros) que le graphe de f est « plus pointu » que celui de la densité gaussienne ; si $\gamma_2 < 0$, il est plus plat(8).

Des estimations de ces deux paramètres servent de base à certains tests statistiques servant à déterminer si une série de données peut être correctement modélisée par une loi gaussienne.

(7). En « anglais » : *kurtosis.*

(8). Pour cette raison, on aurait mieux fait d'appeler γ le coefficient de « pointicité », et non d'applatissement...

11.2 INÉGALITÉS DE MARKOV ET DE BIENAYMÉ-TCHEBYCHEV

Inégalité de Markov

§ 173

L'inégalité de Markov quantifie la propension d'une variable aléatoire positive à rester concentrée autour de 0, et constitue une première étape simple vers la démonstration de l'inégalité de Bienaymé-Tchebychev.

■ **Lemme 11.17 (Inégalité de Markov)** *Soit* Y *une variable aléatoire* positive *admettant une espérance, et soit* $t > 0$ *; alors*

$$\mathbf{P}\{Y \geqslant t\} \leqslant \frac{1}{t}\,\mathbf{E}(Y).$$

DÉMONSTRATION : Donnons une démonstration dans le cas où Y admet une densité f. Tout d'abord,

$$\mathbf{P}\{Y \geqslant t\} = \mathbf{P}\{Y > t\} = \int_t^{+\infty} f(u)\,du$$

et donc

$$t\,\mathbf{P}\{Y \geqslant t\} = \int_t^{+\infty} t\,f(u)\,du \leqslant \int_t^{+\infty} u\,f(u)\,du.$$

Par positivité des valeurs de Y, on a donc

$$t\,\mathbf{P}\{Y \geqslant t\} \leqslant \int_t^{+\infty} u\,f(u)\,du \leqslant \int_{-\infty}^{+\infty} u\,f(u)\,du = \mathbf{E}(Y).$$

Une démonstration plus générale sera donnée page 326. ▮ ⇙11.2

Inégalité de Bienaymé-Tchebychev

§ 174

Cette fameuse inégalité[9] montre qu'une variable aléatoire admettant un moment d'ordre 2 ne peut pas s'éloigner trop de sa valeur moyenne.

■ **Théorème 11.18 (Inégalité de Bienaymé-Tchebychev)** *Soit* X *une variable aléatoire réelle à densité, admettant une variance. Notons* $m = \mathbf{E}(X)$ *l'espérance de* X, σ^2 *sa variance et* $X^* := (X - m)/\sigma$. *Alors, pour tout* $\varepsilon > 0$, *on a*

$$\mathbf{P}\{\,|X - m| \geqslant \varepsilon\} \leqslant \frac{\sigma^2}{\varepsilon^2}.$$

Cette inégalité est plus souvent utilisée sous la forme

$$\mathbf{P}\{\,|X^*| \geqslant \varepsilon\} = \mathbf{P}\{\,|X - m| \geqslant \varepsilon\sigma\} \leqslant \frac{1}{\varepsilon^2}.$$

DÉMONSTRATION : Pour tout $\varepsilon > 0$,

$$\mathbf{P}\{\,|X - m| \geqslant \varepsilon\} = \mathbf{P}\{\,|X - m|^2 \geqslant \varepsilon^2\} \leqslant \frac{1}{\varepsilon^2}\,\mathbf{E}(\,|X - m|^2\,) = \frac{\sigma^2}{\varepsilon^2}.$$

d'après l'inégalité de Markov appliquée à la variable positive $|X - \mathbf{E}X|^2$. ▮

Exemple 11.19 Si X est une variable aléatoire de moyenne m et d'écart-type σ, la probabilité que le résultat $X(\omega)$ d'une expérience diffère de m de plus de 4σ, est certainement inférieure à 1/16 (c'est-à-dire 6,25 %), et ce, indépendamment du détail de la loi de distribution de X. Une valeur $\varepsilon < 1$ dans la seconde inégalité n'apporte aucune renseignement. (Voir la figure 7.3 page 159.)

(9). Connue, ailleurs qu'en France, sous le nom d'inégalité de Tchebychev (*Chebychev's inequality*).

EXERCICES

♦ ***Exercice 11.1*** Soit X une variable aléatoire admettant une densité f continue sur $\mathbb{R}$.

i) Caractériser la valeur du réel a qui minimise l'expression

$$\varphi(a) = \mathbf{E}\big[(X-a)^2\big] = \int_{-\infty}^{+\infty} (t-a)^2 f(t)\,dt.$$

ii) Caractériser la valeur des réels a qui minimisent l'expression

$$\psi(a) = \mathbf{E}\big[|X-a|\big] = \int_{-\infty}^{+\infty} |t-a|\ f(t)\,dt.$$

♦ ***Exercice 11.2 (Testons Bienaymé-Tchebychev)*** Soit X une variable aléatoire suivant une loi gaussienne $\mathcal{N}(m\,;\,\sigma^2)$. Calculer $\mathbf{P}\left\{\frac{|X-m|}{\sigma} \geqslant 2\right\}$ et comparer avec l'estimation fournie par l'inégalité de Bienaymé-Tchebychev[10].

♦ ***Exercice 11.3*** Soit $\alpha > 0$. Soit U une variable aléatoire telle que $U \rightsquigarrow \mathcal{E}(\alpha)$. On pose $X = \cos U$ et $Y = \sin U$.

i) Calculer $\mathbf{P}\left\{Y \geqslant \frac{1}{2}\right\}$.

ii) Calculer $\mathbf{E}(X)$ et $\mathbf{E}(Y)$.

♦ ***Exercice 11.4 (Formule de transfert, cas particulier)*** On suppose que $\varphi : \mathbb{R} \to \mathbb{R}$ est une bijection strictement monotone. Soit X une variable aléatoire admettant une densité f. Déterminer la loi de $\varphi(X)$, et en déduire que

$$\mathbf{E}\big(\varphi(X)\big) = \int_{-\infty}^{+\infty} \varphi(x)\, f(x)\,dx.$$

♦ ***Exercice 11.5*** Soit X une variable aléatoire de la loi uniforme sur $[0\,;\,1]$.

i) Déterminer la loi et l'espérance de $Y = \min(X, 1-X)$, ainsi que de $Z = \max(X, 1-X)$.

ii) Calculer, si elles existent, les espérances de Y/Z et de Z/Y.

♦ ***Exercice 11.6 (Moments de la loi normale)*** Soit X une variable aléatoire suivant la loi normale centrée réduite $\mathcal{N}(0\,;\,1)$. Calculer $\mathbf{E}(X^k)$ pour tout entier naturel k.

♦ ***Exercice 11.7*** Soit X une variable aléatoire suivant la loi normale $\mathcal{N}(m\,;\,\sigma^2)$. Calculer l'espérance de $Y := X^2$.

Indications

◊ *Indication 11.6 :* Justifier la nullité des moments d'ordre impair ; puis montrer par récurrence que $\mathbf{E}(X^{2n}) = (2n)!/2^n\, n!$.

◊ *Indication 11.7 :* Ne pas se lancer tête baissée dans un calcul intégral !

---

[10]. On pourra trouver, page 453, les graphes comparés des estimations de $\mathbf{P}\{|X-m| \geqslant x\sigma\}$ obtenues, soit par l'inégalité de Bienaymé-Tchebychev, soit en valeur exacte, en fonction de x.

SOLUTIONS

♦ **Solution 11.1**

i) En développant le facteur $(t-a)^2$ dans la première intégrale, on aboutit à une expression dans laquelle on reconnaît

$$\varphi(a) = a^2 - 2a\,\mathbf{E}(\mathrm{X}) + \mathbf{E}(\mathrm{X}^2),$$

polynôme de degré 2, minimal en $a = \mathbf{E}(\mathrm{X})$.

ii) On pose décompose tout d'abord l'intégrale en deux parties : de $-\infty$ à a puis de a à $+\infty$; on trouve alors

$$\psi(a) = a\int_{-\infty}^{a} f - a\int_{a}^{+\infty} f - \int_{-\infty}^{a} t\,f(t)\,\mathrm{d}t + \int_{a}^{+\infty} t\,f(t)\,\mathrm{d}t,$$

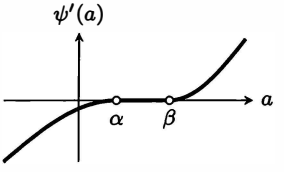
ce qui permet de prouver que ψ est dérivable et de dérivée

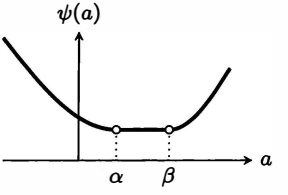
$$\psi'(a) = \int_{-\infty}^{a} f - \int_{a}^{+\infty} f.$$

Cette dérivée est croissante, comme somme de deux fonctions croissantes, tend vers -1 en $-\infty$ et vers $+1$ en $+\infty$; elle s'annule donc en tout point a tel que

$$\int_{a}^{+\infty} f = \int_{-\infty}^{a} f,$$

c'est-à-dire en toute médiane de f. L'ensemble M des médianes, c'est-à-dire des points d'annulation de ψ', est un segment $[\,\alpha\,;\,\beta\,]$. La fonction ψ' est strictement négative sur $[\,-\infty\,;\,\alpha\,[$ et strictement positive sur $]\,\beta\,;\,+\infty\,]$: cela prouve que ψ est minimale en tout point de M.



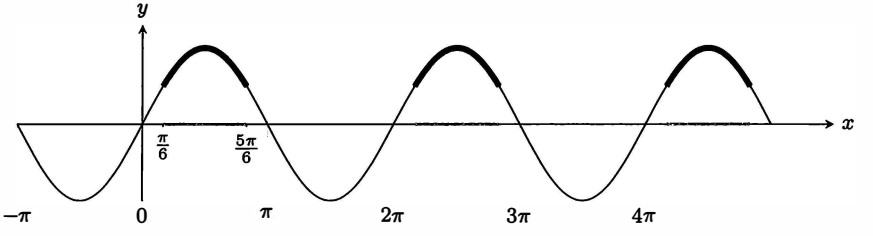


♦ **Solution 11.2** La probabilité demandée vaut exactement $p = 2 - 2\mathfrak{N}(2) \approx 0,046$; une majoration de cette quantité par l'inégalité de Bienaymé-Tchebychev donne $p \leqslant 1/4 = 0,25$, ce qui est plutôt grossier.

♦ **Solution 11.3**

i) Un simple graphe permet de voir que

$$\mathbf{P}\left\{\mathrm{Y} \geqslant \frac{1}{2}\right\} = \int_{\mathrm{A}} \alpha\,\mathrm{e}^{-\alpha u}\,\mathrm{d}u \qquad \text{où} \qquad \mathrm{A} = \bigcup_{k=0}^{\infty}\left[\frac{\pi}{6} + 2k\pi\,;\,\frac{5\pi}{6} + 2k\pi\right].$$



Après intégration et sommation d'une série géométrique, on trouve :

$$\mathbf{P}\left\{\mathrm{Y} \geqslant \frac{1}{2}\right\} = \mathrm{e}^{-\alpha\pi/6}\,\frac{1 - \mathrm{e}^{-2\pi\alpha/3}}{1 - \mathrm{e}^{-2\pi\alpha}}.$$

ii) Notons

$$\mathrm{I} = \int_0^{+\infty} \mathrm{e}^{-\alpha u} \cos u \, \mathrm{d}u \qquad \text{et} \qquad \mathrm{J} = \int_0^{+\infty} \mathrm{e}^{-\alpha u} \sin u \, \mathrm{d}u$$

Deux intégrations par parties amènent les relations $\mathrm{I} = \dfrac{1}{\alpha}(1 - \mathrm{J})$ et $\mathrm{J} = \dfrac{\mathrm{I}}{\alpha}$ d'où l'on déduit $\mathrm{I} = \dfrac{\alpha}{1+\alpha^2}$ et $\mathrm{J} = \dfrac{1}{1+\alpha^2}$. Puisque ces intégrales convergent absolument, la formule de transfert nous donne

$$\mathbf{E}(\mathrm{X}) = \alpha \mathrm{I} = \frac{\alpha^2}{1+\alpha^2} \qquad \text{et} \qquad \mathbf{E}(\mathrm{Y}) = \alpha \mathrm{J} = \frac{\alpha}{1+\alpha^2}$$

♦ **Solution 11.4** La fonction φ est strictement monotone. Supposons, pour fixer les idées, qu'elle soit strictement croissante. Une densité de $\mathrm{Y} = \varphi(\mathrm{X})$ est alors

$$g(y) = \frac{f\big(\varphi^{-1}(y)\big)}{\varphi'\big(\varphi^{-1}(y)\big)},$$

ce qui donne, en utilisant le changement de variable $y = \varphi(x)$, $\mathrm{d}y = \varphi'(x)\,\mathrm{d}x$:

$$\mathbf{E}(\mathrm{Y}) = \int_{-\infty}^{+\infty} y\, g(y)\, \mathrm{d}y = \int_{-\infty}^{+\infty} y \frac{f\big(\varphi^{-1}(y)\big)}{\varphi'\big(\varphi^{-1}(y)\big)}\, \mathrm{d}y = \int_{-\infty}^{+\infty} \varphi(x)\, f(x)\, \mathrm{d}x.$$

♦ **Solution 11.5**

i) Notons $\varphi(x) = \min(x, 1-x)$, alors $\varphi\big([\,0\,;\,1\,]\big) = \big[\,0\,;\,\frac{1}{2}\,\big]$ donc $\mathrm{Y}(\Omega) = \big[\,0\,;\,\frac{1}{2}\,\big]$. Pour tout $y \in \big[\,0\,;\,\frac{1}{2}\,\big]$, calculons

$$\mathbf{P}\{\mathrm{Y} > y\} = \mathbf{P}\{\mathrm{X} > y, 1 - \mathrm{X} > y\} = \mathbf{P}\{y < \mathrm{X} < 1 - y\} = 1 - 2y = 1 - \mathrm{F}_{\mathrm{Y}}(y),$$

donc $\mathrm{F}_{\mathrm{Y}}(y) = 2y$. Ainsi, Y suit la loi uniforme sur $\big[\,0\,;\,\frac{1}{2}\,\big]$.

De même, on pose $\psi(x) = \max(x, 1-x)$ et on vérifie que $\mathrm{Z}(\Omega) = \big[\,\frac{1}{2}\,;\,1\,\big]$ et que Z suit la loi uniforme sur $\big[\,\frac{1}{2}\,;\,1\,\big]$.

ii) Puisque φ/ψ est continue sur $[\,0\,;\,1\,]$, Y/Z admet une espérance qui est donnée par

$$\mathbf{E}\left(\frac{\mathrm{Y}}{\mathrm{Z}}\right) = \int_0^1 \frac{\min(x, 1-x)}{\max(x, 1-x)}\, \mathrm{d}x = \int_0^{1/2} \frac{x}{1-x} + \int_{1/2}^1 \frac{1-x}{x}\, \mathrm{d}x = 2\ln(2) - 1.$$

En revanche, la fonction $y \mapsto \dfrac{1-y}{y}$ n'est pas intégrable sur $\big]\,0\,;\,\frac{1}{2}\,\big]$, donc Z/Y n'admet pas d'espérance.

♦ **Solution 11.6** Si k est impair, alors par intégrabilité et imparité, $\mathbf{E}(\mathrm{X}^k) = 0$.

Supposons k pair et écrivons-le sous la forme $k = 2n$. Une intégration par parties mène à

$$\mathbf{E}(\mathrm{X}^{2n}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2n}\, \mathrm{e}^{-x^2/2}\, \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \underbrace{x^{2n-1}}_{u(x)} \underbrace{x\, \mathrm{e}^{-x^2/2}}_{v'(x)}\, \mathrm{d}x$$

$$= (2n-1) \int_{-\infty}^{+\infty} x^{2n-2}\, \mathrm{e}^{-x^2/2}\, \mathrm{d}x = (2n-1)\, \mathbf{E}(\mathrm{X}^{2(n-1)})$$

ce qui, par récurrence, fournit le résultat

$$\mathrm{I}_n = 1 \cdot 3 \cdot 5 \cdots (2n-1) = \frac{(2n)!}{2^n\, n!}.$$

♦ **Solution 11.7** Le plus simple est de remarquer que $\mathbf{E}(\mathrm{Y}) = \mathbf{E}(\mathrm{X}^2) = \mathbf{E}(\mathrm{X})^2 + \mathbf{V}(\mathrm{X}) = m^2 + \sigma^2$. ☺

Chapitre 12

Couples de variables à densité

Dans tout ce chapitre, nous considérons des couples (X, Y) *de variables aléatoires, ce qui revient à considérer une variable aléatoire* $\mathbf{X} = (X, Y)$ *à valeurs dans* $\mathbb{R}^2$. *La généralisation à des n-uplets* $(X_1, X_2, \ldots, X_n)$ *de variables aléatoires (ou à des variables aléatoires* $\mathbf{X}$ *à valeurs dans* $\mathbb{R}^n$*) est presque toujours immédiate.*

Les propriétés déjà vues pour les couples discrets sont reprises ici, et quelques cas particuliers sont mis en avant (comme la somme de variables normales indépendantes). Une approche heuristique des lois conditionnelles est proposée, avec un exemple détaillé, ainsi qu'un nouvel aperçu de l'espérance conditionnelle. Là encore, le but de ce chapitre est de faire manipuler les notions ; aussi le chapitre s'achève-t-il sur une quinzaine d'exercices, aisés pour la plupart, dont nous conseillons fortement l'étude avant de passer à la partie suivante, consacrée à des aspects plus théoriques.

12.1 LOI D'UN COUPLE

Loi d'un couple, lois marginales §175

La **fonction de répartition (conjointe)** du couple (X, Y) est la fonction $F : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$F(x, y) = \mathbf{P}\{X \leqslant x, Y \leqslant y\}$$

Elle représente la probabilité que le couple (X, Y) prenne ses valeurs dans le domaine $\Delta(x, y)$ du plan représenté sur la figure 12.1 page suivante.

La fonction F est croissante par rapport à chacune de ses variables, donc admet une limite quand $x \to \pm\infty$ ou $y \to \pm\infty$. Nous noterons $F(x, +\infty)$ pour $\lim_{y \to +\infty} F(x, y)$ et ainsi de suite.

■ **Proposition 12.1** $F(+\infty, +\infty) = 1$ *et* $F(x, -\infty) = F(-\infty, y) = 0$ *pour tous* $x, y \in \mathbb{R}$.

DÉMONSTRATION : Pour *toutes* suites croissantes $(x_n)_{n \geqslant 0}$ et $(y_n)_{n \geqslant 0}$ de limites $+\infty$, on peut écrire $\mathbb{R}^2$ comme l'union croissante

$$\mathbb{R}^2 = \bigcup_{n=0}^{\infty} \{(X, Y) \in \Delta(x_n, y_n)\}.$$

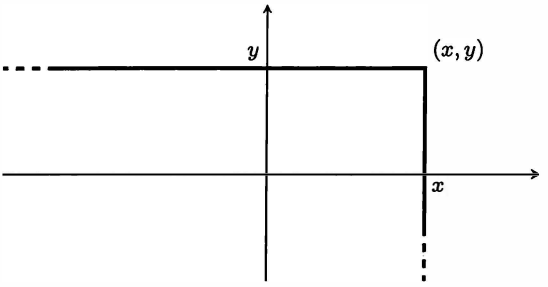


FIGURE 12.1 — *Le domaine* $\Delta(x,y)$.

L'application de l'axiome 3′ permet d'écrire que

$$1 = \mathbf{P}\left(\lim_n \uparrow \{(\mathrm{X},\mathrm{Y}) \in \Delta(x_n,y_n)\}\right) = \lim_{n\to\infty} \mathrm{F}(x_n,y_n).$$

De même, si $(x_n)_{n\geqslant 0}$ et $(y_n)_{n\geqslant 0}$ sont des suites décroissantes de limite $-\infty$, l'ensemble vide est égal à l'intersection décroissante

$$\varnothing = \bigcap_{n=0}^{\infty} \{(\mathrm{X},\mathrm{Y}) \in \Delta(x,y_n)\} = \bigcap_{n=0}^{\infty} \{(\mathrm{X},\mathrm{Y}) \in \Delta(x_n,y)\}.$$

L'axiome 3″ permet de conclure. ∎

Comme dans le cas d'une variable réelle, la fonction de répartition d'un couple caractérise entièrement sa loi[1]. Les variables X et Y sont appelées **marginales** du couple (X, Y). Leurs lois sont caractérisées par leurs fonctions de répartition $\mathrm{F_X}$ et $\mathrm{F_Y}$, elles-mêmes liées à la fonction de répartition du couple par

$$\mathrm{F_X}(x) = \lim_{y\to+\infty} \mathrm{F}(x,y) \qquad \text{et} \qquad \mathrm{F_Y}(y) = \lim_{x\to+\infty} \mathrm{F}(x,y)$$

La première formule découle du fait que, pour toute suite $(y_n)_{n\geqslant 0}$ croissante et de limite $+\infty$

$$\{\mathrm{X} \leqslant x\} = \bigcup_{n=1}^{\infty} \{\mathrm{X} \leqslant x, \mathrm{Y} \leqslant y_n\},$$

la seconde s'obtient par un argument symétrique. Les lois ainsi obtenues sont les **lois marginales** en X et en Y du couple.

§ 176 **Couple admettant une densité**

Un couple (X, Y) de variables aléatoires est **absolument continu**, ou **admet une densité**, s'il existe une fonction $f : \mathbb{R}^2 \to \mathbb{R}$ positive, appelée **densité de probabilité (conjointe)** telle que

$$\forall x,y \in \mathbb{R} \qquad \mathrm{F}(x,y) = \iint_{\Delta(x,y)} f(s,t)\,\mathrm{d}s\,\mathrm{d}t = \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,t)\,\mathrm{d}s\,\mathrm{d}t.$$

Une telle fonction f vérifie donc la propriété

$$\iint_{\mathbb{R}^2} f(s,t)\,\mathrm{d}s\,\mathrm{d}t = 1.$$

(1). Ce point est momentanément admis, sa démonstration requérant des outils plus sophistiqués, cf. A74.

Réciproquement, si l'on se donne une fonction $f : \mathbb{R}^2 \to \mathbb{R}$ positive et d'intégrale 1, alors on admet qu'il existe un couple (X, Y) de variables aléatoires dont la fonction de répartition est la fonction F définie par $F(x,y) = \iint_{\Delta(x,y)} f$ (une démonstration sera donnée dans l'annexe C).

Si (X, Y) admet une densité, sa densité de probabilité (conjointe) f peut être obtenue par dérivation de la fonction de répartition F en presque tout couple (x, y) :

$$f(x,y) = \frac{\partial^2 F}{\partial x\, \partial y}(x,y).$$

De plus, on obtient la règle de calcul suivante :

■ **Proposition 12.2** *Soient $a < b$, $c < d$ des réels. Alors*

$$\begin{aligned} \mathbf{P}\{a \leqslant X \leqslant b,\ c \leqslant Y \leqslant d\} &= F(b,d) - F(a,d) - F(b,c) + F(a,c) \\ &= \iint_R f(s,t)\, ds\, dt \end{aligned}$$

où R *est le rectangle* $[a\,;\,b] \times [c\,;\,d]$.

On admettra que ce résultat se généralise à tout domaine « raisonnable » A du plan[(2)] :

■ **Théorème 12.3 (Densité et calcul de probabilités)** *Soit* A *une partie du plan ; alors*

$$\mathbf{P}\{(X,Y) \in A\} = \iint_A f(s,t)\, ds\, dt.$$

Notamment,

$$F(x,y) = \iint_{\Delta(x,y)} f(s,t)\, ds\, dt$$

Cette formule ne se justifie proprement qu'avec un minimum de théorie de la mesure, et sera donc démontrée ultérieurement. Donnons tout de suite un exemple d'application de cette formule. ◁12.2

Exemple 12.4 On définit la fonction f par

$$f(x,y) = \begin{cases} 4xy\, e^{-x^2-y^2} & \text{si } x \geqslant 0 \text{ et } y \geqslant 0, \\ 0 & \text{sinon.} \end{cases}$$

C'est une fonction positive vérifiant $\iint f = 1$; on suppose que c'est la densité d'un couple (X, Y). *On demande les lois marginales de ce couple, ainsi que la loi des variables aléatoires* $Z = X^2 + Y^2$ *et* $T = Y/X$.

On obtient la fonction de répartition de X par

$$F_X(x) = F(x, +\infty) = \int_{-\infty}^{x} \int_{-\infty}^{+\infty} f(s,t)\, dt\, ds = \begin{cases} 0 & \text{si } x < 0, \\ 1 - e^{-x^2} & \text{sinon.} \end{cases}$$

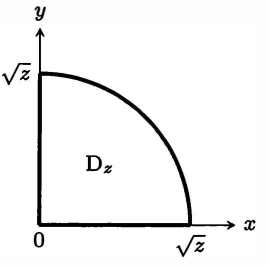
La situation étant symétrique en x et y, la loi de Y est donnée par $F_Y = F_X$.

L'ensemble des valeurs prises par Z est $\mathbb{R}^+$ et, pour tout $z \geqslant 0$:

$$F_Z(z) = \mathbf{P}\{X^2 + Y^2 \leqslant z\} = \iint_{D_z} f(x,y)\, dx\, dy,$$

où D_z est le quart de disque de rayon $\sqrt{z}$:

(2). « Raisonnable » veut dire ici « mesurable » ; puisque l'on se place dans $\mathbb{R}^2$, A doit être un élément de la tribu $\mathcal{B}(\mathbb{R}^2)$ des boréliens du plan.



En passant en coordonnées polaires, on obtient

$$F_Z(z) = \int_0^{\pi/2} \int_0^{\sqrt{z}} 4\rho^3 \cos\theta \sin\theta \, e^{-\rho^2} \, d\rho \, d\theta.$$

Les intégrales se séparent ; d'une part, $\int_0^{\pi/2} 4\cos\theta \sin\theta \, d\theta = 2\int_0^{\pi/2} \sin(2\theta)\, d\theta = 2$ et (après intégration par parties) :

$$\int_0^{\sqrt{z}} \rho^3 e^{-\rho^2} \, d\rho = \frac{1}{2}\int_0^{\sqrt{z}} 2\rho e^{-\rho^2} \, d\rho - \frac{1}{2}\Big[\rho^2 e^{-\rho^2}\Big]_0^{\sqrt{z}} = \frac{1}{2}\Big[1 - (1+z)e^{-z}\Big].$$

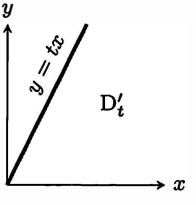
Finalement, on obtient la fonction de répartition et sa densité associée :

$$F_Z(z) = \begin{cases} 1 - (1+z)e^{-z} & \text{si } z \geqslant 0, \\ 0 & \text{sinon.} \end{cases} \qquad \text{et} \qquad f_Z(z) = \begin{cases} z e^{-z} & \text{si } z \geqslant 0 \\ 0 & \text{sinon.} \end{cases}$$

De même, $T = Y/X$ peut prendre toutes valeurs positives, et est définie presque sûrement, l'événement $\{X = 0\}$ étant négligeable. Pour tout $t \in \mathbb{R}^+$, on a

$$F_T(t) = \mathbf{P}\left\{\frac{Y}{X} \leqslant t\right\} = \mathbf{P}\{Y \leqslant tX\} = \iint_{D'_t} f(x,y)\, dx\, dy = 1 - \frac{1}{1+t^2},$$

où l'on a noté D'_t le domaine suivant :

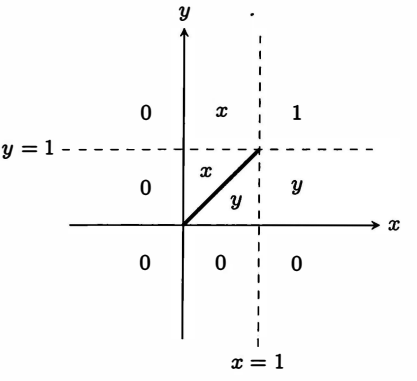


Pour conclure, la fonction de répartition de T est

$$F_T(t) = \begin{cases} 1 - \dfrac{1}{1+t^2} & \text{si } t \geqslant 0, \\ 0 & \text{sinon.} \end{cases}$$

Contrairement au cas de deux variables aléatoires discrètes dont le couple était nécessairement discret, avoir deux variables aléatoires à densité ne suffit pas à conclure que le couple admet une densité !

Exemple 12.5 Considérons une variable aléatoire X de loi uniforme sur $[0\,;\,1]$, et définissons la variable Y par $Y = X$. Alors le couple (X, Y) n'admet pas de densité. En effet, la fonction de répartition de ce couple prend les valeurs résumées sur le graphe ci-dessous :



La dérivée croisée $\partial^2 F/\partial x\,\partial y$ est donc nulle en tous points où elle est définie. La seule manière de définir une « densité » pour ce couple serait, au sens des distributions, par une densité linéique de Dirac localisée sur le segment $0 \leqslant x = y \leqslant 1$, représenté en gras sur la figure. ⇗12.3 ⇗12.4

Densités marginales §177

Le théorème de Fubini montre que, si f est une densité du couple (X, Y), alors la fonction $y \mapsto f(x, y)$ est intégrable pour presque tout $x \in \mathbb{R}$. Posons alors

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)\,dy. \tag{12.1}$$

Toujours d'après le théorème de Fubini,

$$\begin{aligned}
\mathbf{P}\{X \leqslant x\} &= \lim_{y\to+\infty} \mathbf{P}\{X \leqslant x, Y \leqslant y\} = \lim_{y\to+\infty} \int_{-\infty}^{y} \left(\int_{-\infty}^{x} f(s,t)\,ds\right) dt \\
&= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{x} f(s,t)\,ds\right) dt = \int_{-\infty}^{x} \left(\int_{-\infty}^{+\infty} f(s,t)\,dt\right) ds \\
F_X(x) &= \int_{-\infty}^{x} f_X(s)\,ds
\end{aligned}$$

ce qui montre que la variable aléatoire X admet une densité f_X, et que celle-ci est définie, pour presque tout x, par la formule (12.1). On l'appelle **densité marginale** de X.

De même, Y admet une densité marginale donnée (pour presque tout y) par

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)\,dx.$$

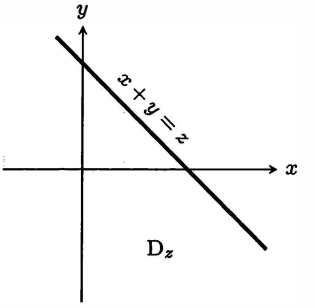
Somme de deux variables aléatoires à densité §178

Tout comme dans le paragraphe §109, on cherche à déterminer la densité de la somme de deux variables aléatoires à densité. Malheureusement, une telle somme n'est pas *a priori* une variable aléatoire à densité. Un exemple tout simple nous en convainc : on choisit une variable aléatoire X de loi normale $\mathcal{N}(0\,;\,1)$ et on pose $Y = -X$. La somme de ces variables aléatoires est évidemment la variable égale à 0, qui n'admet pas de densité.

La situation est plus favorable si le couple admet une densité conjointe f. On peut alors trouver une densité de $S = X + Y$ en fonction de f. Pour tout $z \in \mathbb{R}$,

$$\mathbf{P}\{S \leqslant z\} = \iint_{D_z} f(x, y)\,dx\,dy,$$

où $D_z = \{(x,y) \in \mathbb{R}^2 \,;\, x+y \leqslant z\}$ est le domaine représenté ci-dessous :



On a alors, par changement de variable ($x = t-s, y = s$) de jacobien unité :

$$\begin{aligned}
\mathbf{P}\{S \leqslant z\} &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-y} f(x,y)\,dx \right) dy \\
&= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z} f(t-s,s)\,dt \right) ds \quad \textit{(chg}^{t}\textit{ de variable)} \\
&= \int_{-\infty}^{z} \left(\int_{-\infty}^{+\infty} f(t-s,s)\,ds \right) dt, \qquad \textit{(Fubini)}
\end{aligned}$$

ce qui montre qu'une densité de S est donnée, pour presque tout réel z, par

$$f_S(z) = \int_{-\infty}^{+\infty} f(z-s,s)\,ds.$$

Remarque 12.6 Pour certaines valeurs de z, la formule précédente peut diverger. En pratique, si f est raisonnablement continue, la divergence n'interviendra qu'en un nombre fini de valeurs.

On peut en déduire l'espérance de S, en utilisant le théorème de Fubini et un changement de variable :

$$\begin{aligned}
\mathbf{E}(S) &= \int_{-\infty}^{+\infty} z\, f_S(z)\,dz = \int_{-\infty}^{+\infty} z \int_{-\infty}^{+\infty} f(z-s,s)\,ds\,dz \\
&= \iint_{\mathbb{R}^2} (u+s)\, f(u,s)\,ds\,du \\
&= \iint_{\mathbb{R}^2} u\, f(u,s)\,ds\,du + \iint_{\mathbb{R}^2} s\, f(u,s)\,ds\,du \\
&= \int_{-\infty}^{+\infty} u\, f_X(u)\,du + \int_{-\infty}^{+\infty} s\, f_Y(s)\,ds = \mathbf{E}(X) + \mathbf{E}(Y).
\end{aligned}$$

Dans le cas général, on *admet*[3] que l'espérance de la somme de deux variables aléatoires est la somme de leurs espérances :

■ **Théorème 12.7 (Linéarité de l'espérance)** *Soient* X *et* Y *deux variables aléatoires absolument continues admettant une espérance, telles que* $C := \alpha X + \beta Y$ *est absolument continue; alors* C *admet une espérance et*

$$\mathbf{E}(\alpha X + \beta Y) = \alpha\, \mathbf{E}(X) + \beta\, \mathbf{E}(Y).$$

(3). Encore une fois, une théorie proprement fondée de l'intégration abstraite sur un espace probabilisé permet de prouver ce résultat ; l'idée étant de sommer, non pas sur les valeurs x prises par les variables aléatoires, mais sur les éléments ω de Ω ; voir le chapitre 14.

Théorème de transfert § 179

Soit (X, Y) un couple absolument continu, de densité f. Si $\varphi : \mathbb{R}^2 \to \mathbb{R}$ est une fonction, définie au moins sur l'ensemble des valeurs prises par (X, Y), alors $\varphi(X, Y)$ est une variable aléatoire. Nous admettons momentanément le résultat suivant, analogue du théorème 8.2 page 177 :

■ **Théorème 12.8** *La variable aléatoire $\varphi(X, Y)$ admet une espérance si et seulement si la fonction $(x, y) \mapsto \varphi(x, y)\, f(x, y)$ est intégrable sur $\mathbb{R}^2$. Dans ce cas,*

$$\mathbf{E}\big(\varphi(X, Y)\big) = \iint_{\mathbb{R}^2} \varphi(x, y)\, f(x, y)\, dx\, dy.$$

12.2 INDÉPENDANCE

Indépendance de deux variables aléatoires à densité § 180

Soient X et Y deux variables à densité ; on dit qu'elles sont **indépendantes** si, pour tous événements A et B de $\mathbb{R}$, les événements $\{X \in A\}$ et $\{Y \in B\}$ sont indépendants, c'est-à-dire si

$$\mathbf{P}\big\{(X, Y) \in A \times B\big\} = \mathbf{P}\{X \in A\} \cdot \mathbf{P}\{Y \in B\}.$$

En fait, pour prouver l'indépendance de deux variables, il n'est pas nécessaire de traiter *tous* les événements A et B : la classe des événements de la forme $]-\infty\,;\,x]$ suffit[4]. On admettra donc que

■ **Théorème 12.9** *Deux variables* X *et* Y *sont indépendantes si et seulement si*

$$\forall x, y \in \mathbb{R} \qquad \mathbf{P}\{X \leqslant x, Y \leqslant y\} = \mathbf{P}\{X \leqslant x\} \cdot \mathbf{P}\{Y \leqslant y\},$$

ce qui s'écrit encore

$$\forall x, y \in \mathbb{R} \qquad F(x, y) = F_X(x) \cdot F_Y(y).$$

Si g et h sont des applications, notons $g \otimes h$ la fonction $(x, y) \mapsto g(x)\, h(y)$; on l'appelle **produit direct** (ou **produit tensoriel**) de g et h. Avec ces notations, on a donc

■ **Théorème 12.10** X *et* Y *sont indépendantes si et seulement si la fonction de répartition du couple est le produit tensoriel des fonctions de répartition marginales :* $F = F_X \otimes F_Y$.

Par simple dérivation partielle, on obtient :

■ **Théorème 12.11** X *et* Y *sont indépendantes si et seulement si la densité conjointe du couple est (presque partout) le produit tensoriel des densités marginales :*

$$X \text{ et } Y \text{ indépendantes} \iff f = f_X \otimes f_Y.$$

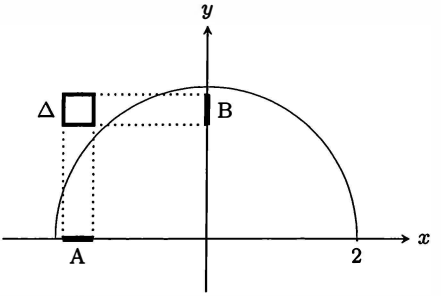
(4). Cette propriété capitale se démontre avec le lemme de classe monotone ; voir A74 et notamment le corollaire B.21 page 706.

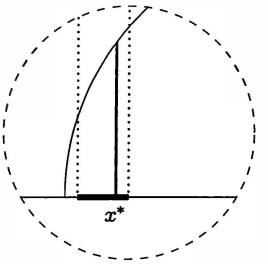
Remarque 12.12 (Un critère graphique de non indépendance) Si (X, Y) est un couple absolument continu de densité conjointe f, et si X et Y sont indépendantes, alors leur densité se factorise $f = f_X \otimes f_Y$. Si l'on dessine dans le plan les zones où f est strictement positive, ces zones ont alors la forme de rectangles (éventuellement semi-infinis ou infinis) ; *toute autre forme exclut la possibilité que* X *et* Y *soient indépendantes.*

Par exemple, supposons que la densité conjointe du couple (X, Y) soit donnée par

$$f(x,y) = \begin{cases} 1/2\pi & \text{si } y \geqslant 0 \text{ et } x^2+y^2 \leqslant 4 \\ 0 & \text{sinon.} \end{cases}$$

Avant même de déterminer les lois marginales, on peut tracer le graphe suivant :





La zone gris clair est le domaine de non-nullité de f ; $\Delta = A \times B$ est un rectangle de $\mathbb{R}^2$ vérifiant

$$\mathbf{P}\{(X,Y) \in \Delta\} = \iint_\Delta f = 0.$$

La projection A de Δ sur l'axe des abscisses est telle que la densité marginale f_X est *strictement positive* sur A (la valeur $f_X(x^*)$ est obtenue par l'intégrale de $f(x^*,\cdot)$ sur un segment de $\mathbb{R}^2$ sur lequel $f(x^*,\cdot)$ est strictement positive, voir détail de l'image à droite). Ainsi, $\mathbf{P}\{X \in A\} > 0$. De même, $\mathbf{P}\{Y \in B\} > 0$. On a donc au total

$$\underbrace{\mathbf{P}\{(X,Y) \in A \times B\}}_{=0} \neq \underbrace{\mathbf{P}\{X \in A\}}_{>0} \cdot \underbrace{\mathbf{P}\{Y \in B\}}_{>0},$$

ce qui prouve que X et Y ne sont pas indépendantes.

§ 181 Indépendance mutuelle

On généralise la notion d'indépendance de deux variables aléatoires à une famille (finie ou non) de variables aléatoires ; la définition est identique à celle que l'on a déjà rencontrée pour des variables aléatoires discrètes.

Des variables $X_1, X_2, \ldots, X_n$ sont **mutuellement indépendantes** si, pour tous événements $A_1, A_2, \ldots, A_n$ de $\mathbb{R}$, on a

$$\mathbf{P}\left(\bigcap_{k=1}^{n} \{X_k \in A_k\}\right) = \prod_{k=1}^{n} \mathbf{P}\{X_k \in A_k\}$$

Des variables aléatoires $(X_i)_{i\in I}$ (en nombre infini, dénombrable ou non) sont **mutuellement indépendantes** si, pour toute famille $(A_j)_{j\in J}$ *finie* d'événements de $\mathbb{R}$, on a

$$\mathbf{P}\left(\bigcap_{j\in J} \{X_j \in A_j\}\right) = \prod_{j\in J} \mathbf{P}\{X_j \in A_j\}$$

En pratique, il suffit de vérifier que : pour toute famille finie $(x_j)_{j\in J}$ de réels, on a

$$\mathbf{P}\left(\bigcap_{j\in J} \{X_j \leqslant x_j\}\right) = \prod_{j\in J} \mathbf{P}\{X_j \leqslant x_j\}.$$

Enfin, nous rappelons des résultats, déjà vus pour les variables aléatoires discrètes, qui sont vrais dans le cas général :

■ **Théorème 12.13 (Fonctions de variables indépendantes)**

i) Soient X *et* Y *deux variables aléatoires indépendantes. Soient* φ *et* ψ *deux applications de* $\mathbb{R}$ *dans* $\mathbb{R}$. *Alors les variables aléatoires* $\varphi(X)$ *et* $\psi(Y)$ *sont indépendantes.*

ii) Plus généralement, si $X_1, X_2, \dots, X_n$ *sont des variables aléatoires indépendantes, si* $\varphi_1, \varphi_2, \dots, \varphi_n$ *sont des applications de* $\mathbb{R}$ *dans* $\mathbb{R}$, *alors les variables* $\varphi_1(X_1), \dots, \varphi_n(X_n)$ *sont indépendantes.*

DÉMONSTRATION : La démonstration donnée page 179 ne fait pas intervenir le caractère discret des variables aléatoires concernées ; elle reste valable ici. ∎

Quant au lemme des coalitions, déjà rencontré page 179, sa démonstration est nettement plus ardue dans le cadre des variables à densité ; il sera démontré proprement au chapitre B (théorème B.17 page 705). Nous l'admettons donc ici.

■ **Théorème 12.14 (Lemme des coalitions)**
Si $X_1, X_2, \dots, X_n$ *sont des variables aléatoires indépendantes, si* k *est un entier compris entre* 1 *et* n, *et si* $\varphi : \mathbb{R}^k \to \mathbb{R}$ *et* $\psi : \mathbb{R}^{n-k} \to \mathbb{R}$ *sont des applications, alors* $Y = \varphi(X_1, \dots, X_k)$ *et* $Z = \psi(X_{k+1}, \dots, X_n)$ *sont indépendantes.*

Somme de deux variables aléatoires indépendantes § 182

Soient X et Y deux variables aléatoires indépendantes, et supposons qu'elles admettent des densités respectives f_X et f_Y ; notons F_X et F_Y leurs fonctions de répartition. Alors le couple (X, Y) a pour fonction de répartition $F = F_X \otimes F_Y$, et admet une densité $f = f_X \otimes f_Y$. On peut reprendre le calcul du § 178 : toujours en notant $S = X + Y$, on a pour tout z réel

$$\begin{aligned} \mathbf{P}\{S \leqslant z\} &= \int_{-\infty}^{z} \left(\int_{-\infty}^{+\infty} f(t-s, s)\, ds \right) dt \\ &= \int_{-\infty}^{z} \left(\int_{-\infty}^{+\infty} f_X(t-s) \cdot f_Y(s)\, ds \right) dt. \end{aligned}$$

On définit le **produit de convolution** $f_X * f_Y$ par

$$(f_X * f_Y)(t) = \int_{-\infty}^{+\infty} f_X(t-s)\, f_Y(s)\, ds$$

en tout réel t en lequel $f_X(t - \cdot)\, f_Y(\cdot)$ est intégrable sur $\mathbb{R}$, ce qui est le cas pour presque tout réel t (voir A45, page 686) ; cela montre que X+Y admet pour densité la fonction $f_X * f_Y$.

■ **Théorème 12.15 (de convolution)** *Si* X *et* Y *sont deux variables aléatoires indépendantes admettant une densité, alors* $S = X + Y$ *admet pour densité leur produit de convolution* $f_S = f_X * f_Y$.

Remarque 12.16 (Point de vue heuristique) Bien avant l'invention du calcul intégral moderne, cette formule était connue par analogie avec le cas discret. La quantité $f_S(t)\, dt$ représente en effet ≏12.10
la probabilité que $S = X + Y$ soit égale à t (à une incertitude dt près). Il suffit pour la calculer de sommer selon toutes les possibilités de valeurs prises par Y ; lorsque Y prend une valeur y, X prend nécessairement la valeur $t - y$, et par indépendance, cela se fait avec la probabilité $f_X(t-y)\, f_Y(y)$. Par un raisonnement de type « formule des probabilités totales », il ne reste plus qu'à sommer ces diverses contributions, ce qui mène naturellement au résultat voulu.

Remarque 12.17 Ce résultat sera généralisé au paragraphe § 248 dans un cadre abstrait.

§ 183 **Somme de lois normales indépendantes**

Le résultat suivant montre que la somme de deux variables aléatoires normales indépendantes est encore normale. Les paramètres sont aisés à retenir : les moyennes s'ajoutent (linéarité de l'espérance), ainsi que les variances (égalité de Bienaymé).

■ **Théorème 12.18** *Soient* X_1 *et* X_2 *deux variables normales indépendantes de lois respectives* $\mathcal{N}(m_1\,;\,\sigma_1^2)$ *et* $\mathcal{N}(m_2\,;\,\sigma_2^2)$. *Alors leur somme* X_1+X_2 *suit une loi normale :*

$$X_1+X_2 \rightsquigarrow \mathcal{N}(m_1+m_2\,;\,\sigma_1^2+\sigma_2^2)\,.$$

De même, avec des notations semblables, $X_1+\cdots+X_n$ *suit une loi normale de paramètres* $m_1+\cdots+m_n$ *et* $\sigma_1^2+\cdots+\sigma_n^2$.

DÉMONSTRATION : La démonstration la plus simple et la plus élégante fait intervenir les fonctions caractéristiques[5]. Cependant, une démonstration élémentaire — mais un tantinet fastidieuse — est possible. Tout d'abord, quitte à remplacer X_1 par X_1-m_1 et X_2 par X_2-m_2, on peut supposer $m_1=m_2=0$. Les densités de X_1 et X_2 sont respectivement $f_1(x)=(1/\sqrt{2\pi}\,\sigma_1)\,\mathrm{e}^{-x^2/2\sigma_1^2}$ et $f_2(x)=(1/\sqrt{2\pi}\,\sigma_2)\,\mathrm{e}^{-x^2/2\sigma_2^2}$. Posons $\sigma=\sqrt{\sigma_1^2+\sigma_2^2}$. Le théorème de convolution nous dit que X_1+X_2 admet la densité $f=f_1*f_2$, c'est-à-dire

$$\begin{aligned} f(x) &= \frac{1}{2\pi\,\sigma_1\,\sigma_2}\int_{-\infty}^{+\infty} \mathrm{e}^{-t^2/2\sigma_1^2}\,\mathrm{e}^{-(x-t)^2/2\sigma_2^2}\,\mathrm{d}t \\ &= \frac{1}{2\pi\,\sigma_1\,\sigma_2}\int_{-\infty}^{+\infty} \exp\left\{-\left(\frac{\sigma^2}{2\,\sigma_1^2\sigma_2^2}\,t^2-\frac{x\sigma_1^2}{\sigma_1^2\sigma_2^2}\,t+\frac{x^2}{2\sigma_2^2}\right)\right\}\mathrm{d}t. \end{aligned}$$

Mettons le trinôme en t sous sa forme canonique :

$$f(x)=\frac{1}{2\pi\,\sigma_1\,\sigma_2}\int_{-\infty}^{+\infty}\exp\left\{-\left(\frac{\sigma^2}{2\,\sigma_1^2\sigma_2^2}\left(t-\frac{x\,\sigma_1^2}{\sigma^2}\right)^2-\frac{\sigma_1^2x^2}{2\,\sigma_2^2\,\sigma^2}+\frac{x^2}{2\sigma_2^2}\right)\right\}\mathrm{d}t.$$

Le changement de variable $s=t-x\sigma_1^2/\sigma^2$ mène à

$$f(x)=\frac{1}{2\pi\,\sigma_1\,\sigma_2}\exp\left\{-\frac{\sigma_1^2x^2}{2\,\sigma_2^2\,\sigma^2}+\frac{x^2}{2\,\sigma^2}\right\}\int_{-\infty}^{+\infty}\exp\left\{-\frac{\sigma^2}{2\,\sigma_1^2\sigma_2^2}\,s^2\right\}\mathrm{d}s.$$

Quelques simplifications dans la première exponentielle montrent qu'elle vaut $\mathrm{e}^{-x^2/2\sigma^2}$; on conclut en utilisant la relation $\int_{-\infty}^{+\infty}\mathrm{e}^{-\alpha^2s^2/2}\,\mathrm{d}s=\sqrt{2\pi}/\alpha$:

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\,\mathrm{e}^{-x^2/2\sigma^2},$$

ce qui est le résultat désiré pour prouver que $X_1+X_2\rightsquigarrow\mathcal{N}(0\,;\,\sigma^2)$. ∎

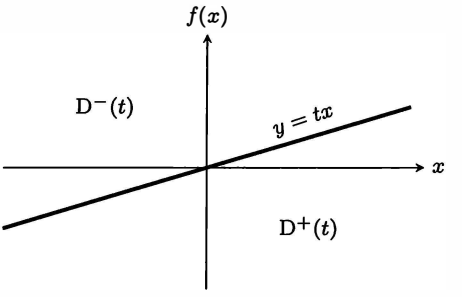
§ 184 **Un exemple : quotient de deux lois normales**

Soient X et Y deux variables aléatoires *indépendantes* suivant la même loi normale centrée réduite $\mathcal{N}(0\,;\,1)$. On cherche la loi de la variable $T=X/Y$. Tout d'abord, il convient de noter que T est définie partout sauf sur l'ensemble $\{Y=0\}$, qui est de probabilité nulle. Soit t un réel. Notons

$$D(t)=\left\{(x,y)\in\mathbb{R}^2\,;\,\frac{y}{x}\leqslant t\right\}.$$

Cet ensemble, représenté ci-après en gris, est l'union $D(t)=D^+(t)\cup D^-(t)$ de deux parties presque disjointes (elles n'ont en commun qu'un point, de probabilité nulle).

(5). Voir exemple 17.11 page 361.



Par symétrie, on peut intégrer uniquement sur $D(t)^+$ et multiplier le résultat par deux :

$$\mathbf{P}\left\{\frac{Y}{X} \leqslant t\right\} = 2 \iint_{D(t)^+} n(x)\, n(y)\, dx\, dy.$$

Enfin, un passage aux coordonnées polaires conduit à

$$F_T(t) = \mathbf{P}\left\{\frac{Y}{X} \leqslant t\right\} = \frac{1}{\pi} \int_0^{+\infty} \int_{-\infty}^{tx} e^{-(x^2+y^2)/2}\, dx\, dy$$
$$= \frac{1}{\pi} \int_0^{+\infty} \int_{-\pi/2}^{\operatorname{Arc\,tan}(t)} r\, e^{-r^2/2}\, d\theta\, dr = \frac{1}{\pi}\left[\operatorname{Arc\,tan}(t) + \frac{\pi}{2}\right].$$

On reconnaît dans le membre de droite la fonction de répartition d'une loi de Cauchy $\mathscr{C}(0\,;\,1)$.

■ **Théorème 12.19** *Si* X *et* Y *sont deux variables aléatoires indépendantes et de même loi normale centrée réduite* $\mathscr{N}(0\,;\,1)$, *alors* X/Y *et* Y/X *suivent la loi de Cauchy* $\mathscr{C}(0\,;\,1)$.

Maximum et minimum de N variables aléatoires indépendantes § 185

Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes, de même loi absolument continue ; notons F leur fonction de répartition, f leur densité et

$$I := \min(X_1, X_2, \ldots, X_n) \qquad \text{et} \qquad M := \max(X_1, X_2, \ldots, X_n).$$

On se propose de déterminer les lois de I et M au moyen de leurs fonctions de répartition F_* et F^*. Pour la variable M, on remarque d'abord que le maximum de n nombres est inférieur à x si et seulement si chacun de ces nombres est inférieur à x :

$$F^*(x) = \{M \leqslant x\} = \bigcap_{k=1}^{n} \{X_k \leqslant x\}.$$

Par indépendance mutuelle,

$$\mathbf{P}\{M \leqslant x\} = \prod_{k=1}^{n} \mathbf{P}\{X_k \leqslant x\} = F(x)^n,$$

de quoi l'on déduit que M est une variable aléatoire absolument continue, admettant pour densité

$$f^*(x) = n\, f(x)\, F(x)^{n-1}.$$

De même,

$$\{I > x\} = \bigcap_{k=1}^{n} \{X_k > x\}$$

et, par indépendance mutuelle

$$\mathrm{F}_*(x) = \mathbf{P}\{\mathrm{I} \leqslant x\} = 1 - \mathbf{P}\{\mathrm{I} > x\} = 1 - \prod_{k=1}^{n} \mathbf{P}\{\mathrm{X}_k > x\} = 1 - \big[1 - \mathrm{F}(x)\big]^n,$$

ce qui montre que I est absolument continue et admet pour densité

$$f_*(x) = n\, f(x)\, \big[1 - \mathrm{F}(x)\big]^{n-1}.$$

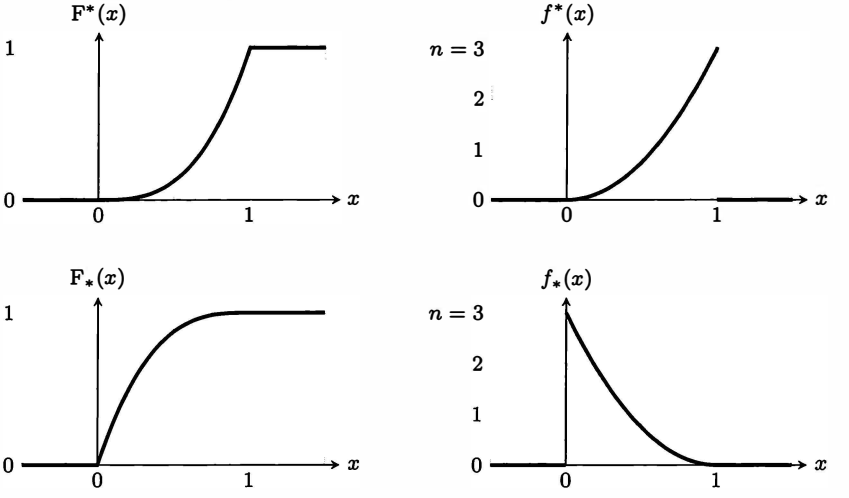
Exemple 12.20 (Maximum et minimum de n lois uniformes)
On suppose que $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$ sont indépendantes et suivent la même loi uniforme $\mathscr{U}([0\,;1])$. Alors M et I prennent leurs valeurs dans $[0\,;1]$ et, pour tout $x \in [0\,;1]$, on a

$$\mathrm{F}^*(x) = x^n \qquad f^*(x) = n\, x^{n-1}$$

tandis que

$$\mathrm{F}_*(x) = 1 - (1-x)^n \qquad f_*(x) = n(1-x)^{n-1}.$$

L'allure des fonctions de répartition et des densités de M et I est donnée ci-dessous, pour $n = 3$.



△12.8
△12.9

§ 186 **Espérance du produit de variables indépendantes**

Soient X et Y deux variables aléatoires de densités respectives f_{X} et f_{Y}, et admettant chacune une espérance. On suppose de plus que X et Y sont indépendantes. Une densité conjointe du couple est donc $f = f_{\mathrm{X}} \otimes f_{\mathrm{Y}}$. On peut alors écrire

$$\iint xy\, f(x,y)\, \mathrm{d}x\, \mathrm{d}y = \iint xy\, f_{\mathrm{X}}(x)\, f_{\mathrm{Y}}(y)\, \mathrm{d}x\, \mathrm{d}y = \int x\, f_{\mathrm{X}}(x)\, \mathrm{d}x \cdot \int y\, f_{\mathrm{Y}}(y)\, \mathrm{d}y$$

en vertu du théorème de Fubini, puisque les fonctions $x \mapsto \big|x\, f_{\mathrm{X}}(x)\big|$ et $y \mapsto \big|y\, f_{\mathrm{Y}}(y)\big|$ sont intégrables. Cela prouve le théorème suivant :

■ **Théorème 12.21** *Si* X *et* Y *sont des variables aléatoires absolument continues et indépendantes, et admettant chacune une espérance, alors* XY *admet une espérance et*

$$\mathbf{E}(\mathrm{XY}) = \mathbf{E}(\mathrm{X}) \cdot \mathbf{E}(\mathrm{Y}).$$

Remarque 12.22 On comparera utilement ce résultat (et sa démonstration) à sa version discrète 8.8 page 181. En réalité, ce résultat est toujours vrai, quelle que soit la nature des variables aléatoires X et Y (cf. théorème 16.19 page 347).

12.3 COVARIANCE

Covariance d'un couple à densité § 187

Soit (X, Y) un couple absolument continu de densité conjointe f. On suppose que X, Y ainsi que XY admettent chacune une espérance[6] ; on définit alors la **covariance** du couple (X, Y) par

$$\mathbf{Cov}(X,Y) := \mathbf{E}(X'Y') = \mathbf{E}\big[(X-\mathbf{E}X)(Y-\mathbf{E}Y)\big],$$

en notant toujours $X' = X - \mathbf{E}X$ et $Y' = Y - \mathbf{E}Y$ les variables centrées. De par les propriétés générales de l'espérance[7], la covariance de (X, Y) s'écrit également

$$\mathbf{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y)$$

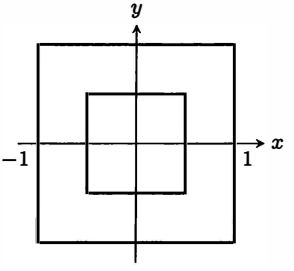
ou encore, en développant la formule :

$$\mathbf{Cov}(X,Y) = \iint xy\, f(x,y)\,dx\,dy - \left(\iint x\, f(x,y)\,dx\,dy\right)\left(\iint y\, f(x,y)\,dx\,dy\right).$$

Les propriétés de la covariance d'un couple à densité sont essentiellement les mêmes que celles d'un couple discret, passons-les rapidement en revue. ⇙12.5

■ **Théorème 12.23 (Indépendance et décorrélation)** *Soient* X *et* Y *deux variables aléatoires à densité indépendantes ; alors elles sont décorrélées :* $\mathbf{Cov}(X,Y) = 0$.

Remarque 12.24 Comme dans le cas discret, la réciproque est fausse : deux variables aléatoires peuvent être décorrélées sans être indépendantes. Nous pouvons construire un exemple de couple absolument continu vérifiant cette propriété de la manière suivante, en choisissant pour densité $f(x,y) = 1/3$ dans le domaine grisé ci-dessous et $f(x,y) = 0$ partout ailleurs :



À titre d'exercice, le lecteur pourra vérifier (sans effectuer de calcul mais par des arguments de parité) que $\mathbf{E}(XY) = \mathbf{E}(X) = \mathbf{E}(Y) = 0$, et donc que $\mathbf{Cov}(X,Y) = 0$; mais que les variables ne sont pas indépendantes, puisque par exemple

$$\mathbf{P}\{|X| \leqslant \tfrac{1}{2}, |Y| \leqslant \tfrac{1}{2}\} = 0 \quad \neq \quad \mathbf{P}\{|X| \leqslant \tfrac{1}{2}\} \cdot \mathbf{P}\{|Y| \leqslant \tfrac{1}{2}\} = \tfrac{1}{3} \cdot \tfrac{1}{3} = \tfrac{1}{9}.$$

■ **Théorème 12.25 (Propriétés de la covariance)** *Soient* X, Y, Z *et* T *des variables aléatoires à densité admettant des espérances, et telles que les covariances ci-dessous sont toutes définies. Alors*

---

(6). C'est par exemple le cas si X et Y admettent des moments d'ordre deux, en vertu de l'inégalité de Cauchy-Schwarz.

(7). La linéarité, mais également le fait que $\mathbf{E}(\mathbf{E}X) = \mathbf{E}(X)$, propriété que nous ne pouvons démontrer dans le cadre des variables aléatoires à densité, puisque $\mathbf{E}(X)$ est une variable aléatoire certaine, donc discrète. Nous faisons donc usage des propriétés générales qui seront exposées dans le chapitre 14 et qui justifient, par exemple, la formule de transfert qui permet également de démontrer la formule donnée.

i) $\mathbf{Cov}(X,Y) = \mathbf{Cov}(Y,X)$ *(symétrie);*

ii) $\mathbf{Cov}(\alpha X+\beta,\gamma Y+\delta) = \alpha\gamma\,\mathbf{Cov}(X,Y)$;

iii) $\mathbf{Cov}(X+Y,Z+T) = \mathbf{Cov}(X,Z)+\mathbf{Cov}(X,T)+\mathbf{Cov}(Y,Z)+\mathbf{Cov}(Y,T)$;

iv) $\mathbf{V}(X) = \mathbf{Cov}(X,X)$;

v) $\left|\mathbf{Cov}(X,Y)\right| \leqslant \sigma_X \cdot \sigma_Y$. *(Cauchy-Schwarz)*

■ **Théorème 12.26 (Variance d'une somme)** *Soient* X *et* Y *deux variables aléatoires admettant une variance et une covariance. Alors*

$$\mathbf{V}(X+Y) = \mathbf{V}(X)+\mathbf{V}(Y)+2\,\mathbf{Cov}(X,Y).$$

$$\mathbf{V}(X-Y) = \mathbf{V}(X)+\mathbf{V}(Y)-2\,\mathbf{Cov}(X,Y).$$

Si $a,b\in\mathbb{R}$, *alors*

$$\mathbf{V}(aX+bY) = a^2\,\mathbf{V}(X)+b^2\,\mathbf{V}(Y)+2ab\,\mathbf{Cov}(X,Y).$$

Plus généralement, si $X_1,X_2,\dots,X_n$ *sont des variables aléatoires admettant des variances et des covariances. Alors*

$$\mathbf{V}(X_1+\cdots+X_n) = \sum_{k=1}^{n}\mathbf{V}(X_k)+2\sum_{1\leqslant i<j\leqslant n}\mathbf{Cov}(X_i,X_j)$$

■ **Corollaire 12.27 (Égalité de Bienaymé)** *Si* X *et* Y *sont deux variables aléatoires absolument continues, indépendantes et admettant un moment d'ordre* 2 ; *alors* $X+Y$ *est absolument continue, admet un moment d'ordre* 2 *et*

$$\mathbf{V}(X+Y) = \mathbf{V}(X)+\mathbf{V}(Y) \qquad \text{et} \qquad \sigma_{X+Y} = \sqrt{\sigma_X^2+\sigma_Y^2}.$$

De même si $X_1,X_2,\dots,X_n$ *sont des variables aléatoires absolument continues indépendantes,*

$$\mathbf{V}\left(\sum_{i=1}^{n}X_i\right) = \sum_{i=1}^{n}\mathbf{V}(X_i).$$

§ 188 **Coefficient de corrélation**

Si (X,Y) est un couple absolument continu de variables aléatoires, admettant une covariance, alors X et Y admettent une densité ; nous avons vu dans la remarque 11.11 page 249 que cela implique $\mathbf{V}(X) > 0$ et $\mathbf{V}(Y) > 0$; on appelle **coefficient de corrélation** du couple le réel

$$\rho(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sigma_X\,\sigma_Y}.$$

Comme dans le cas discret (la démonstration étant identique et reposant sur l'inégalité de Cauchy-Schwarz), on a le résultat suivant :

■ **Théorème 12.28** *Soit* (X,Y) *un couple de variables aléatoires admettant un coefficient de corrélation* $\rho = \rho(X,Y)$. *Alors* $-1 \leqslant \rho \leqslant 1$. *De plus,* $\rho = \pm 1$ *si et seulement si* $X-\mathbf{E}(X)$ *et* $Y-\mathbf{E}(Y)$ *sont proportionnelles, et plus exactement si et seulement si*

$$\frac{Y-\mathbf{E}(Y)}{\sigma_Y} = \rho\,\frac{X-\mathbf{E}(X)}{\sigma_X}$$

presque sûrement. Si l'on note X^* *et* Y^* *les variables aléatoires centrées et réduites, on a donc* $\mathbf{P}\{X^*=Y^*\}=1$ *si* $\rho=1$, *et* $\mathbf{P}\{X^*=-Y^*\}=1$ *si* $\rho=-1$.

12.4 LOIS CONDITIONNELLES

Lois conditionnelles § 189

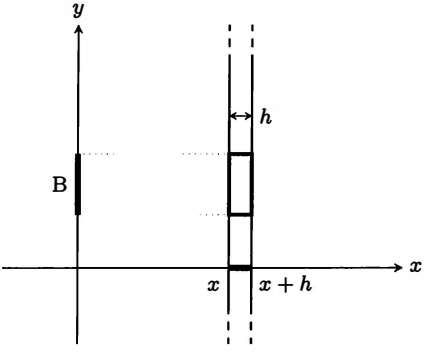
Si (X, Y) est un couple absolument continu, les probabilités de la forme $\mathbf{P}\{X = x\}$ sont toutes nulles, puisque le domaine $\Delta = \{(x,t)\ ;\ t \in \mathbb{R}\}$, qui est une droite du plan, est de mesure nulle. Ainsi, on ne peut pas, comme on l'avait fait pour des variables aléatoires discrètes, définir de loi conditionnelle de Y sachant $\{X = x\}$ avec une formule de la forme

$$\mathbf{P}\{Y \in B \,|\, X = x\} = \frac{\mathbf{P}\{Y \in B, X = x\}}{\mathbf{P}\{X = x\}},$$

dont le numérateur comme le dénominateur est nul. Une approche heuristique permet de construire une méthode pour définir le membre de gauche. L'idée est d'élargir le domaine de X, puis de passer à la limite ; on considère donc, pour tout $h > 0$, la quantité

$$\varphi(h) = \frac{\mathbf{P}\{Y \in B, x \leqslant X \leqslant x+h\}}{\mathbf{P}\{x \leqslant X \leqslant X+h\}}.$$

Le numérateur est égal à l'intégrale de f sur $[x\,;\,x+h] \times B$, le dénominateur est l'intégrale de f sur $[x\,;\,x+h] \times \mathbb{R}$.



On reconnaît dans le dénominateur une intégrale sur la densité marginale f_X de X :

$$\varphi(h) = \frac{\displaystyle\int_x^{x+h} \left(\int_B f(s,t)\,dt \right) ds}{\displaystyle\int_x^{x+h} \int_{-\infty}^{+\infty} f(s,t)\,dt\,ds} = \frac{\displaystyle\int_x^{x+h} \left(\int_B f(s,t)\,dt \right) ds}{\displaystyle\int_x^{x+h} f_X(s)\,ds}$$

Si tout se passe bien, on doit pouvoir effectuer un développement limité au premier ordre en h :

$$\varphi(h) \approx \frac{h \int_B f(x,t)\,dt + \cdots}{h\, f_X(x) + \cdots}$$

et donc, en faisant tendre h vers 0, on peut imaginer obtenir

$$\mathbf{P}\{Y \in B \,|\, X = x\} \stackrel{?}{=} \lim_{h \to 0+} \varphi(h) \stackrel{?}{=} \frac{1}{f_X(x)} \int_B f(x,t)\,dt.$$

Notamment, si l'on choisit $B =]-\infty\,;\,y]$, on accéderait ainsi à la fonction de répartition de Y sachant $\{X = x\}$:

$$\mathbf{P}\{Y \leqslant y \,|\, X = x\} \stackrel{?}{=} \frac{1}{f_X(x)} \int_{-\infty}^{y} f(x,t)\,dt$$

et, par dérivation, à la densité de probabilité de Y sachant $\{X = x\}$, qu'on notera $f^Y_{\{X=x\}}(y)$ ou, plus succinctement et lorsqu'aucune confusion n'est possible, $f_Y(y \mid x)$:

$$f_Y(y \mid x) = f^Y_{\{X=x\}}(y) = \frac{f(x,y)}{f_X(x)}.$$

Évidemment, nous n'avons rien contrôlé dans les diverses opérations (passages à la limite, développements, dérivations,...) et il est difficile, en l'absence d'hypothèses fortes ou fastidieuses à énoncer, de justifier rigoureusement les raisonnements précédents. L'idée est plutôt de *partir du résultat heuristique*, c'est-à-dire de poser la définition suivante :

$$f_Y(y \mid x) = f^Y_{\{X=x\}}(y) := \begin{cases} \dfrac{f(x,y)}{f_X(x)} & \text{si } f_X(x) \neq 0 \\ \textit{une valeur arbitraire positive} & \text{si } f_X(x) = 0. \end{cases}$$

Remarque 12.29 La *valeur arbitraire* ne devrait pas poser de problème : en anticipant un peu, la formule $\mathbf{P}(A \cap B) = \mathbf{P}(B \mid A) \cdot \mathbf{P}(A)$ devrait se généraliser, pour les densités, en $f(x,y) = f_Y(y \mid x) \cdot f_X(x)$, c'est-à-dire que la quantité arbitraire est multipliée par zéro.

Afin que $f_Y(\cdot \mid x)$ ait tout de même les propriétés d'une densité, on restreint (légèrement) l'arbitraire de la manière suivante :

Définition 12.30 Soit (X, Y) un couple absolument continu, de densité f et de densités marginales f_X et f_Y. Soit ρ une densité arbitraire sur $\mathbb{R}$. On définit, pour tout $(x, y) \in \mathbb{R}^2$, *une* **densité conditionnelle de Y par rapport à X** par

$$f_Y(y \mid x) = f^Y_{\{X=x\}}(y) := \begin{cases} \dfrac{f(x,y)}{f_X(x)} & \text{si } f_X(x) > 0 \\ \rho(y) & \text{si } f_X(x) = 0. \end{cases}$$

Avec cette définition, la propriété suivante est immédiate.

■ **Proposition 12.31** *Pour tout $x \in \mathbb{R}$, l'application $f^Y_{\{X=x\}}$ est une densité de probabilité.*

De plus, pour tout $x \in \mathbb{R}$, on a, pour presque tout $y \in \mathbb{R}$ la relation

$$f(x,y) = f_X(x)\, f_Y(y \mid x),$$

qui est l'équivalent pour les densités de la formule bien connue $\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B \mid A)$. En effet, lorsque $f_X(x) > 0$, cette relation est vraie pour tout y par construction. Si x est un réel tel que $f_X(x) = 0$, alors $\int_{-\infty}^{+\infty} f(x,y)\,dy = 0$, ce qui veut dire que $f(x,y) = 0$ pour presque tout y, et donc que la relation est vraie pour presque tout y.

Remarque 12.32 Cette construction se généralise (de manière plus rigoureuse) en introduisant la notion de *noyau de probabilité* et dans la théorie générale des *espérances conditionnelles* ; voir § 350.

Les densités conditionnelles sont souvent utilisées dans le sens réciproque. Si X admet une densité f_X et si, pour presque tout[8] $x \in \mathbb{R}$, on peut déterminer une

(8). « Presque tout » est à prendre ici au sens de la loi $\mathbf{P}_X$ de X, c'est-à-dire que la propriété est vraie sauf sur un ensemble $A \subset \mathbb{R}$ tel que $\mathbf{P}_X(A) = \mathbf{P}\{X \in A\} = 0$.

densité conditionnelle $f_Y(\cdot \mid x)$ de Y sachant $\{X = x\}$, alors le couple (X, Y) admet une densité valant, pour presque tout[9] couple $(x, y) \in \mathbb{R}^2$:

$$f(x, y) = f_X(x) \cdot f_Y(y \mid x).$$

C'est le point de vue qui sera exposé dans le paragraphe § 190.

Remarquons enfin que, si l'on symétrise la dernière relation sous la forme

$$f(x, y) = f_X(x) \cdot f_Y(y \mid x) = f_Y(y) \cdot f_X(x \mid y),$$

on obtient l'égalité

$$f_Y(y \mid x) = \frac{f_Y(y) \cdot f_X(x \mid y)}{f_X(x)}.$$

Or, on peut développer $f_X(x)$ comme « combinaison convexe » de la densité conditionnelle $f_X(x \mid \cdot)$:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, t)\, dt = \int_{-\infty}^{+\infty} f_Y(t) \cdot f_X(x \mid t)\, dt,$$

ce qui fait que l'équation précédente s'écrit

$$f_Y(y \mid x) = \frac{f_Y(y) \cdot f_X(x \mid y)}{\int_{-\infty}^{+\infty} f_Y(t) \cdot f_X(x \mid t)\, dt},$$

relation que l'on peut interpréter comme une **généralisation continue de la formule de Bayes** discrète qui, pour mémoire, a la forme

$$\mathbf{P}(B_i \mid A) = \frac{\mathbf{P}(B_i) \cdot \mathbf{P}(A \mid B_i)}{\sum_k \mathbf{P}(B_k) \cdot \mathbf{P}(A \mid B_k)}.$$

Remarque 12.33 (Lois a priori *et* a posteriori*)* La loi marginale de Y dans le couple (X, Y) est parfois appelée **loi a priori** de Y, tandis que la loi conditionnée par $\{X = x\}$ est appelée **loi a posteriori**.

Cette terminologie vient de l'interprétation suivante : la densité marginale f_Y représente l'état de connaissance de la variable Y lorsqu'*aucune* mesure de X n'a été effectuée. Au contraire, la densité conditionnelle $f_Y(\cdot \mid x)$ prend en compte le résultat d'une observation « $X = x$ ». ▵12.14

Bien entendu, lorsque les variables X et Y sont indépendantes, les densités conditionnelles sont égales aux densités marginales :

■ **Théorème 12.34 (Densités conditionnelles et indépendance)**
Soit (X, Y) *un couple de variables aléatoires admettant une densité ; on suppose* X *et* Y *indépendantes. Soit* x *un réel tel que* $f_X(x) > 0$*, alors*

$$f_Y(y \mid x) = f_Y(y)$$

pour tout $y \in \mathbb{R}$*. De même, si* y *vérifie* $f_Y(x) > 0$*, alors* $f_X(x \mid y) = f_X(x)$ *pour tout* $x \in \mathbb{R}$*.*

DÉMONSTRATION : Il suffit de remarquer que l'indépendance des variables permet d'écrire la densité conjointe comme le produit tensoriel des densités marginales :

$$\forall x, y \in \mathbb{R} \qquad f(x, y) = f_X(x) \cdot f_Y(y).$$

L'application de la définition d'une densité conditionnelle permet de conclure. ∎

[9]. Cette fois-ci, le « presque tout » est relatif à la loi $\mathbf{P}_{(X,Y)}$ du couple.

§ 190 **Exemple de variable conditionnée par une autre**

Dans de nombreuses situations, on ne cherche pas à calculer la densité conditionnelle à partir de la densité conjointe. Au contraire, c'est la densité conditionnelle qui est la quantité « naturelle », c'est-à-dire celle que l'on postule dans l'étape de modélisation d'une situation, et dont découle notamment la densité conjointe.

On choisit au hasard X dans $[0;1]$, puis Y au hasard dans $[0;X]$. On demande la loi et l'espérance de Y, ainsi que le coefficient de corrélation $\rho(X,Y)$ du couple.

La loi conditionnée de Y sachant $\{X=x\}$ est la loi uniforme sur $[0;x]$; sa densité est donc, pour tout $x \in]0;1]$:

$$f_Y(y \mid x) = f^Y_{\{X=x\}}(y) = \begin{cases} \frac{1}{x} & \text{si } 0 \leqslant y \leqslant x, \\ 0 & \text{sinon.} \end{cases}$$

Or, $f^Y_{\{X=x\}}(y) = \dfrac{f(x,y)}{f_X(x)}$, et $f_X(x) = 1$ pour tout $x \in [0;1]$. On en déduit que

$$f(x,y) = \begin{cases} \frac{1}{x} & \text{si } 0 \leqslant y \leqslant x \leqslant 1 \\ 0 & \text{sinon.} \end{cases}$$

Il ne nous reste plus qu'à calculer la densité marginale : pour tout $y \in]0;1]$,

$$f_Y(y) = \int_y^1 \frac{1}{x}\,dx = -\ln(y).$$

On en déduit (toutes les intégrales étant bien définies)

$$\mathbf{E}(Y) = -\int_0^1 y\ln(y)\,dy = \int_0^1 \frac{y^2}{2}\cdot\frac{1}{y}\,dy - \left[\frac{y^2}{2}\ln(y)\right]_0^1 = \frac{1}{4}.$$

On aurait pu également calculer cette espérance par la formule de transfert en utilisant la densité conjointe :

$$\mathbf{E}(Y) = \iint_{\mathbb{R}^2} y\cdot f(x,y)\,dx\,dy = \int_0^1 \left(\int_0^x \frac{y}{x}\,dy\right) dx = \int_0^1 \frac{x}{2}\,dx = \frac{1}{4}.$$

Maintenant, toujours par la formule de transfert,

$$\mathbf{E}(XY) = \iint_{\mathbb{R}^2} xy\cdot f(x,y)\,dx\,dy = \int_0^1 \left(\int_0^x y\,dy\right) dx = \int_0^1 \frac{x^2}{2}\,dx = \frac{1}{6},$$

tandis que $\mathbf{E}(X) = \frac{1}{2}$ et $\mathbf{E}(Y) = \frac{1}{4}$, ce qui permet de conclure que

$$\mathbf{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y) = \frac{1}{6} - \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{24}.$$

Il ne reste plus qu'à calculer les variances : X étant uniformément distribuée sur $[0;1]$, sa variance vaut $\mathbf{V}(X) = \frac{1}{12}$. Quant à celle de Y, on l'obtient en calculant d'abord

$$\mathbf{E}(Y^2) = \int_0^1 \left(\int_0^x \frac{y^2}{x}\,dy\right) dx = \int_0^1 \frac{x^2}{3}\,dx = \frac{1}{9},$$

d'où $\mathbf{V}(Y) = \mathbf{E}(Y^2) - \mathbf{E}(Y)^2 = \dfrac{1}{9} - \dfrac{1}{16} = \dfrac{7}{144}$. Pour finir, le coefficient de corrélation vaut

$$\rho(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sqrt{\mathbf{V}(X)}\sqrt{\mathbf{V}(Y)}} = \sqrt{\frac{3}{7}}.$$

≏12.12

§ 191 **Un exemple de loi mixte (discret, continu) : temps d'attente du métro**

Afin d'illustrer encore une fois le fait que les probabilités conditionnelles sont souvent le point de départ d'une modélisation, nous donnons ici un exemple un peu plus compliqué d'un couple dont une des variables est discrète, l'autre à densité.

La situation à modéliser est la suivante : lorsque l'on descend sur le quai du métro, on devine rapidement, au vu du nombre de personnes déjà à attendre sur le quai, si le précédent métro est passé il y a peu ou au contraire si cela fait déjà un certain temps.

Pour quantifier cela, nous faisons quelques suppositions :

— le métro passe à intervalles réguliers de durée $\theta = 10$ minutes ;
— au bout d'un temps t après le dernier passage du métro, le nombre N de personnes à attendre suit une loi de Poisson de paramètre λt.

Cette deuxième condition est typique de la modélisation d'un flux aléatoire, mais de densité constante, de personnes[10]. Elle implique notamment que le nombre moyen de voyageurs à attendre à l'instant t est λt. Pour nos applications numériques, on prendra $\lambda = 3$, de telle façon que le nombre moyen de voyageurs à chaque passage du métro soit $a = \lambda\theta = 30$.

On note T la variable aléatoire donnant l'instant d'arrivée de l'observateur sur le quai, l'instant 0 étant pris au passage précédent du métro. La variable T suit donc une loi uniforme sur l'intervalle $[0\,;\,\theta[$. Enfin, au moment où l'observateur arrive, le nombre N de passagers qui attendent sur le quai est supposée avoir la valeur particulière n_0.

On cherche maintenant à modéliser la loi du couple (N, T). *Ce qui est connu* est donc la loi conditionnelle de N sachant $\{T = t\}$, pour toutes les valeurs de $t \in [0\,;\,\theta[$:

$$\mathbf{P}\{N = n \mid T = t\} = e^{-\lambda t}\,\frac{(\lambda t)^n}{n!}.$$

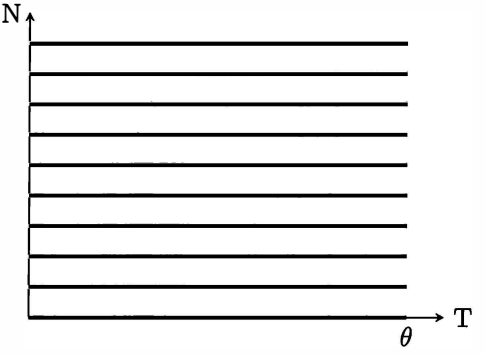
Ce que nous cherchons est la loi de T sachant notre observation $N = n_0$, mais comme les quantités $\mathbf{P}\{T = t \mid N = n\}$ sont bien évidemment nulles, on cherchera plutôt

$$\mathbf{P}\{T \leqslant t \mid N = n\}.$$

D'une certaine manière, il doit exister un moyen d'exprimer cette quantité en fonction de la loi connue — on reconnaît un *renversement des causes*, comme dans la formule de Bayes. Ici, la difficulté est que l'événement $\{T = t\}$ est de probabilité nulle, et que l'on ne peut pas aveuglément utiliser les mêmes techniques que Bayes. La stratégie est la suivante :

— on détermine la loi conjointe du couple ;
— on en déduit la loi marginale de N ;
— on en déduit enfin la loi de T conditionnée par l'événement $\{N = n_0\}$.

Première étape, nous devons déterminer la loi conjointe du couple. Heuristiquement, les valeurs possibles de (N, T) se répartissent ainsi dans le plan :



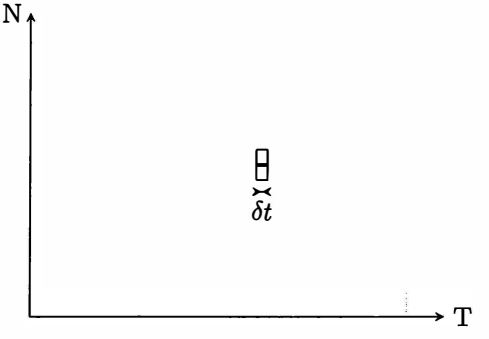
et ne forment évidemment pas une densité[11]. Donnons-nous un petit intervalle

(10). Voir § 80 et le chapitre 30.

(11). Sauf pour les physiciens qui seront heureux d'écrire que la densité du couple est une somme de masses linéiques de Dirac...

de temps $[t\,;\,t+\delta t]$; on peut alors faire l'approximation

$$\begin{aligned}\mathbf{P}\{\mathrm{N}=n, t\leqslant \mathrm{T}\leqslant t+\delta t\} &= \mathbf{P}\{\mathrm{N}=n\,|\,t\leqslant \mathrm{T}\leqslant t+\delta t\}\cdot\mathbf{P}\{t\leqslant \mathrm{T}\leqslant t+\delta t\}\\ &\approx \mathbf{P}\{\mathrm{N}=n\,|\,\mathrm{T}=t\}\cdot\mathbf{P}\{t\leqslant \mathrm{T}\leqslant t+\delta t\}\ \textit{(approché)}\\ &= \mathbf{P}\{\mathrm{N}=n\,|\,\mathrm{T}=t\}\cdot\frac{\delta t}{\theta}.\ \textit{(loi marginale uniforme)}\end{aligned}$$



On est donc amenés, en sommant ces relations et en passant à la limite $\delta t\to 0$, à *définir* la loi du couple (N, T) par la formule

$$\mathbf{P}\{\mathrm{N}=n, \mathrm{T}\leqslant t\} = \frac{1}{\theta}\int_0^t \mathbf{P}\{\mathrm{N}=n\,|\,\mathrm{T}=s\}\,\mathrm{d}s$$

c'est-à-dire en somment les quantités $\mathbf{P}\{\mathrm{N}=n\,|\,\mathrm{T}=s\}$ avec le « poids » uniforme $\mathrm{d}s/\theta$, ce qui donne finalement

$$\mathbf{P}\{\mathrm{N}=n, \mathrm{T}\leqslant t\} = \frac{1}{\theta}\int_0^t \mathrm{e}^{-\lambda s}\frac{(\lambda s)^n}{n!}\,\mathrm{d}s = \frac{1}{a\,n!}\int_0^{\lambda t} u^n\,\mathrm{e}^{-u}\,\mathrm{d}u.$$

(On rappelle que $a=\lambda\theta$.)

Remarque 12.35 Avec cette formule, on vérifie d'abord que la loi marginale de T est bien la loi uniforme dont on est parti :

$$\begin{aligned}\mathbf{P}\{\mathrm{T}\leqslant t\} &= \sum_{n=0}^{\infty}\mathbf{P}\{\mathrm{N}=n, \mathrm{T}\leqslant t\} = \sum_{n=0}^{\infty}\frac{1}{\theta}\int_0^t \mathbf{P}\{\mathrm{N}=n\,|\,\mathrm{T}=s\}\,\mathrm{d}s\\ &= \frac{1}{\theta}\int_0^t \underbrace{\sum_{n=0}^{\infty}\mathbf{P}\{\mathrm{N}=n\,|\,\mathrm{T}=s\}}_{=1}\,\mathrm{d}s = \frac{t}{\theta},\end{aligned}$$

l'interversion somme-intégrale étant justifiée par la positivité des quantités mises en jeu.

On peut maintenant déterminer la loi marginale de N :

$$\mathbf{P}\{\mathrm{N}=n\} = \mathbf{P}\{\mathrm{N}=n, \mathrm{T}\leqslant\theta\} = \frac{1}{a\,n!}\int_0^a u^n\,\mathrm{e}^{-u}\,\mathrm{d}u$$

ainsi que la loi de T conditionnée par $\{\mathrm{N}=n\}$:

$$\mathrm{F}_n(t) := \mathbf{P}\{\mathrm{T}\leqslant t\,|\,\mathrm{N}=n\} = \frac{\mathbf{P}\{\mathrm{N}=n, \mathrm{T}\leqslant t\}}{\mathbf{P}\{\mathrm{N}=n\}} = \frac{\int_0^{\lambda t} u^n\,\mathrm{e}^{-u}\,\mathrm{d}u}{\int_0^a u^n\,\mathrm{e}^{-u}\,\mathrm{d}u}.$$

Ainsi, dans l'espace $(\Omega, \mathcal{T}, \mathbf{P}_{\{N=n\}})$, la variable aléatoire T admet une densité

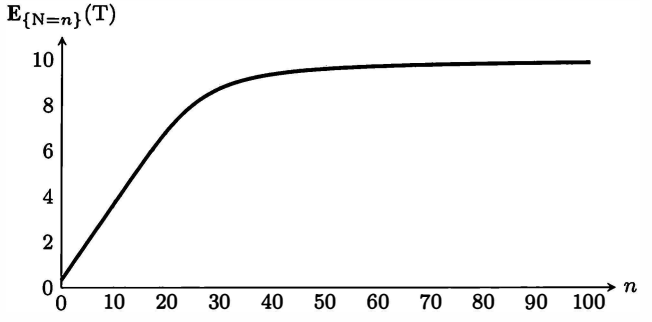
$$f_n(t) = F'_n(t) = \frac{\lambda\,(\lambda t)^n\,e^{-\lambda t}}{\int_0^a u^n\,e^{-u}\,du}$$

et admet donc pour espérance[12]

$$\mathbf{E}_{\{N=n\}}(T) = \int_0^\theta t\,f_n(t)\,dt = \frac{\theta}{a}\cdot\frac{\int_0^a u^{n+1}\,e^{-u}\,du}{\int_0^a u^n\,e^{-u}\,du},$$

ce qui répond à notre question initiale (quand le métro a-t-il pu passer, et donc combien de temps va-t-on attendre ?).

En fonction du nombre n, on trace ci-dessous la valeur de $\mathbf{E}_{\{N=n\}}(T)$, pour les valeurs $\theta = 10$ et $a = 30$; pour des valeurs n très grandes, ce nombre est très proche de $\theta = 10$ (l'attente sera courte !). Par exemple, si l'on observe une trentaine de voyageurs trépignant sur le quai, on peut s'attendre à ce que le métro passe rapidement : $\mathbf{E}(T \mid n = 30) \approx 8,7$.



Espérances conditionnelles § 192

Nous avons déjà défini la notion d'espérance conditionnelle pour des variables discrètes ; nous pouvons faire de même pour des variables à densité, à condition d'admettre que « tout se passe bien » au niveau du calcul intégral. On note

$$\mathbf{E}(Y \mid X = x) := \int_{-\infty}^{+\infty} y\,f_{\{X=x\}}(y)\,dy$$

que l'on appelle **espérance de Y conditionnée par $\{X = x\}$** ; cette quantité dépend évidemment de x, c'est-à-dire que la fonction $c : x \mapsto \mathbf{E}(Y \mid X = x)$ est une fonction de $\mathbb{R}$ dans $\mathbb{R}$. On peut maintenant définir une fonction sur l'espace Ω par

$$\omega \longmapsto \mathbf{E}\big(Y \mid X = X(\omega)\big).$$

Cette fonction[13] n'est rien d'autre que $c \circ X$:

$$c \circ X(\omega) = \mathbf{E}\big(Y \mid X = X(\omega)\big).$$

---

(12). On peut également calculer l'espérance de T conditionnée par $\{N = n\}$ par la formule (valable pour une variable aléatoire positive) : $\mathbf{E}(X) = \int_0^{+\infty} \big(1 - F(x)\big)\,dx = \int_0^{+\infty} \mathbf{P}\{X \geqslant x\}\,dx$ (voir § 201). On obtient évidemment (après une intégration par parties ou bien par l'utilisation du théorème de Fubini) le même résultat.

(13). Définie de manière arbitraire quand le réel $x = X(\omega)$ vérifie $f_X(x) = 0$...

C'est maintenant une variable aléatoire[14] (une application de $\Omega \to \mathbb{R}$), on l'appelle **espérance conditionnelle de Y sachant X** et on la note $\mathbf{E}(Y \,\|\, X)$.

■ **Théorème 12.36** *Soit* (X, Y) *un couple absolument continu. On suppose de plus que* Y *admet une espérance. Alors* Y *admet une espérance conditionnelle sachant* X*, et celle-ci admet une espérance. De plus,*

$$\mathbf{E}\big(\mathbf{E}(Y \,\|\, X)\big) = \mathbf{E}(Y).$$

DÉMONSTRATION : On ne démontre ici ce résultat que dans le cas où l'on peut écrire, pour tout couple (x, y) de réels, $f(x, y) = f_X(x)\, f_Y(y \,|\, x)$. On peut alors écrire l'espérance de Y sous la forme

$$\mathbf{E}(Y) = \iint_{\mathbb{R}^2} y\, f(x, y)\, dx\, dy = \iint_{\mathbb{R}^2} y\, f_X(x)\, f_Y(y \,|\, x)\, dx\, dy.$$

La convergence absolue de cette intégrale permet d'appliquer le théorème de Fubini :

$$\begin{aligned}
\mathbf{E}(Y) &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} y f_Y(y \,|\, x)\, dy \right) f_X(x)\, dx \\
&= \int_{-\infty}^{+\infty} \mathbf{E}(Y \,|\, X = x)\, f_X(x)\, dx \qquad (*) \\
&= \mathbf{E}\big(c(X)\big) = \mathbf{E}\big(\mathbf{E}(Y \,\|\, X)\big) \qquad \textit{(formule de transfert)}
\end{aligned}$$

ce qui montre le résultat. ∎

On remarquera que la relation $(*)$ s'interprète encore comme la formule de l'espérance totale vue au théorème 8.29 page 190 dans le cas discret.

EXERCICES

♦ ***Exercice 12.1 (Un gâteau avec deux gros œufs)*** Solveig veut faire un gâteau au chocolat ; la recette précise qu'il faut « deux gros œufs (100 g) ». Or, il ne lui reste plus que deux œufs, chacun venant d'une boîte différente : dans la première, les œufs sont calibrés « gros » (ils font entre 50 g et 60 g, avec une répartition uniforme) et dans la seconde « moyens » (entre 40 et 45 g avec répartition uniforme). Quelle est la densité de probabilité associée à la masse totale d'œuf ? Quels sont la moyenne et l'écart-type de cette masse ? Quelle est la probabilité que la masse des œufs dépasse les 100 grammes ?

♦ ***Exercice 12.2 (Un couple presque normal)*** Soit Z une variable de loi normale $\mathcal{N}(0\,;\,1)$. Soit U une variable aléatoire, indépendante de Z, telle que $\mathbf{P}\{U = -1\} = \mathbf{P}\{U = 1\} = \frac{1}{2}$. On pose $X = U|Z|$ et $Y = Z$. Montrer que X suit la loi normale centrée réduite. Les variables aléatoires X et Y sont-elles indépendantes ? Le couple (X, Y) admet-il une densité ?

♦ ***Exercice 12.3*** Soit (X, Y) un couple de fonction de répartition $F(x, y) = \mathfrak{N}\big(\min(x, y)\big)$. On note D la droite du plan, d'équation $y = x$. Vérifier que F est une fonction continue sur $\mathbb{R}^2$, de classe $\mathscr{C}^2$ sur $\mathbb{R}^2 \setminus D$. Montrer que F n'admet pas de densité. Construire un couple (X, Y) dont F soit la fonction de répartition.

♦ ***Exercice 12.4 (La baguette brisée)*** Voici un très ancien problème de probabilité qui se résout essentiellement de manière géométrique, ce qui fait une grande part de son charme.

On coupe une baguette de longueur 1 en deux endroits représentés par deux variables aléatoires indépendantes X et Y de loi uniforme sur $[0\,;\,1]$. On note A, B et C les longueurs des 3 morceaux ainsi formés de gauche à droite.

i) Déterminer la loi de A.

ii) Montrer que C a la même loi que A. Déterminer l'espérance de A et de C.

---

(14). Nous ne sous soucions pas encore de questions de mesurabilité des fonctions pour les appeler « variables aléatoires ». Une théorie mathématique plus rigoureuse de l'espérance conditionnelle est développée au chapitre 23.

iii) Déterminer la loi de B.

iv) Calculer la probabilité que le morceau du milieu soit le plus court des trois.

♦ ***Exercice 12.5 (Loi de*** $X|X|$ ***quand*** X ***est normale)*** Soit X une variable suivant la loi normale centrée réduite. On pose $Y = X|X|$. Donner la loi de Y et son espérance. Calculer la covariance du couple (X, Y).

♦ ***Exercice 12.6*** Soient X et Y deux variables aléatoires indépendantes de loi normale centrée réduite. Déterminer la loi de $X^2 + Y^2$.

♦ ***Exercice 12.7 (Le rendez-vous)*** Rosencrantz et Guildenstern se fixent un rendez-vous dans une taverne, entre 0h et 1h. On fera la supposition que chacun arrive, indépendamment de l'autre, à un instant aléatoire suivant une loi uniforme sur $[0\,;\,1]$.

i) Calculer le temps moyen d'attente de la première personne arrivée sur les lieux.

ii) Notons X et Y les heures d'arrivée respectives de la première et de la seconde personne. Déterminer les lois de X et Y ainsi que le coefficient de corrélation $\rho(X, Y)$.

iii) En fait, Guildenstern attend au maximum 30 minutes avant de repartir s'il n'a pas vu son compère ; quant à Rosencrantz, encore moins patient, il n'attend que 15 minutes tout au plus. Quelle est la probabilité qu'ils se rencontrent ?

♦ ***Exercice 12.8 (Minimum et maximum de*** N ***lois exponentielles indépendantes)*** Soit $\alpha > 0$ un réel ; soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes, de même loi exponentielle de paramètre α. On note $U = \min(X_1, X_2, \ldots, X_n)$ et $M = \max(X_1, X_2, \ldots, X_n)$. Déterminer les lois de U et M, ainsi que leurs espérances. Donner un équivalent de $\mathbf{E}(M)$ quand n tend vers l'infini.

♦ ***Exercice 12.9 (Lois uniformes et loi bêta)*** Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes de même loi uniforme sur $[0\,;\,1]$. Montrer que

$$I := \min(X_1, X_2, \ldots, X_n) \qquad \text{et} \qquad M := \max(X_1, X_2, \ldots, X_n)$$

suivent des lois bêta : $I \rightsquigarrow B(1, n)$ et $M \rightsquigarrow B(n, 1)$.

♦ ***Exercice 12.10 (Somme de lois exponentielles et loi Gamma)*** On considère des variables aléatoires indépendantes $X_1, X_2, \ldots, X_n$, de même loi exponentielle de paramètre $\lambda > 0$. Expliciter la loi de $S_n = X_1 + \cdots + X_n$.

♦ ***Exercice 12.11 (Le paradoxe des grenouilles)*** Une grenouille effectue des sauts (toujours vers l'avant) de longueur comprise entre 0 et 1 mètre, avec une répartition uniforme sur $[0\,;\,1]$. Calculer la distance moyenne parcourue en deux sauts. Calculer ensuite le nombre moyen de sauts nécessaires pour que la grenouille atteigne ou dépasse 1 mètre. Commenter.

♦ ***Exercice 12.12 (Variable définie par une loi conditionnelle)*** Soit X une variable aléatoire admettant une densité f_X définie par

$$f_X(x) = \begin{cases} x\,e^{-x} & \text{si } x > 0, \\ 0 & \text{sinon.} \end{cases}$$

On considère une variable aléatoire Y telle que, pour tout $x > 0$, la loi conditionnelle de Y sachant $\{X = x\}$ est la loi uniforme sur $[0\,;\,x]$.

i) Déterminer la densité du couple (X, Y). En déduire la loi de Y.

ii) Calculer $\mathbf{Cov}(X, Y)$

iii) Déterminer la loi de $X - Y$ et montrer que Y et $X - Y$ sont indépendantes.

iv) Retrouver la valeur de $\mathbf{Cov}(X, Y)$.

v) Déterminer la loi de Y/X.

♦ ***Exercice 12.13*** $X_1, X_2, \ldots, X_n$ sont des variables aléatoires indépendantes suivant chacune une loi exponentielle de paramètre λ. On souhaite prouver le résultat (évident par symétrie) : la probabilité que X_1 soit la plus petite des variables $X_1, X_2, \ldots, X_n$ est $1/n$.

i) Déterminer la loi de $U_n = \min(X_2, \ldots, X_n)$.

ii) En déduire la loi de $U_n - X_1$. Conclure.

♦ ***Exercice 12.14 (Max et min de trois lois uniformes)*** Soient U, V et W trois variables aléatoires indépendantes, de même loi uniforme sur $[0\,;1]$. On pose

$$X := \max(U, V, W) \qquad \text{et} \qquad Y := \min(U, V, W).$$

Déterminer la loi du couple (X, Y), puis la densité conditionnelle de X sachant $\{Y = y\}$ (pour $0 \leqslant y \leqslant 1$). Calculer également la covariance du couple (X, Y) et son coefficient de corrélation. Le signe était-il prévisible ?

♦ ***Exercice 12.15 (Densité d'une somme et somme de densités)*** Soit X et Y deux variables aléatoires indépendantes, admettant des densités respectives f et g. On sait qu'une densité de la somme est $f * g$. À quelle variable aléatoire aléatoire correspond alors la densité $(f+g)/2$? Ou, plus généralement, $\lambda f + (1-\lambda)g$ lorsque $0 < \lambda < 1$?

♦ ***Exercice 12.16 (Loi exponentielle bilatérale)*** On note $f(t) = \frac{1}{2}\,\mathrm{e}^{-|t|}$ pour tout $t \in \mathbb{R}$.

i) Montrer que f est une densité de probabilité. Soit X une variable aléatoire admettant f pour densité. Calculer l'espérance et l'écart-type de X.

ii) Déterminer une densité g de la variable aléatoire $Z := X - Y$, où X et Y sont deux variables aléatoires indépendantes admettant f pour densité.

Indications

◊ *Indication 12.2 :* Si l'on note D le domaine formé du plan privé des deux diagonales principales $y = \pm x$, vérifier que $\mathbf{P}\{(X, Y) \in D\} = 0$ et en déduire que (X, Y) n'admet pas de densité.

◊ *Indication 12.10 :* On calculera par convolution la densité f_n de S_n pour $n = 2, 3, 4$; on montrera par récurrence la formule qui semble se dégager.

◊ *Indication 12.11 :* Si l'on note S_n la longueur totale parcourue à l'issue du n-ième saut et N le nombre de saut nécessaires pour dépasser un mètre, alors $\{N > n\} = \{S_n \leqslant 1\}$. L'espérance de N se calcule ensuite (exercice 7.5) par $\mathbf{E}(N) = \sum_{n=0}^{\infty} \mathbf{P}\{N > n\}$.

◊ *Indication 12.13 :* Déterminer des densités de U_n et de $-X_1$, et obtenir celle de $T = U_n - X_1$ par convolution.

◊ *Indication 12.14 :* Calculer $\mathbf{P}\{X \leqslant x, Y > y\}$ pour $0 \leqslant y \leqslant x \leqslant 1$ pour en déduire la fonction de répartition de (X, Y).

◊ *Indication 12.16 :* On vérifiera d'abord que $-Y$ admet également f pour densité.

SOLUTIONS

♦ **Solution 12.1** Les densités de probabilité de la masse des œufs n° 1 et n° 2 sont respectivement

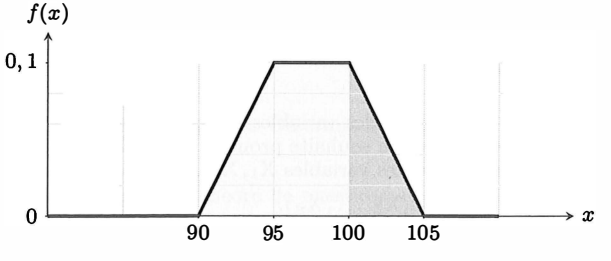
$$f_1(x) = \begin{cases} 0 & \text{si } x \notin [50\,;60]\,, \\ \dfrac{1}{10} & \text{si } x \in [50\,;60]\,, \end{cases} \qquad \text{et} \qquad f_2(x) = \begin{cases} 0 & \text{si } x \notin [40\,;45]\,, \\ \dfrac{1}{5} & \text{si } x \in [40\,;45]\,. \end{cases}$$

La moyenne de la masse de l'œuf 1 est bien sûr 55 g, la moyenne du second est 42,5 g, la moyenne de la masse totale est donc 97,5 g.

La densité de probabilité de la masse d'œuf est la convolution

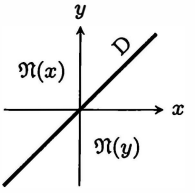
$$f(x) = f_1 * f_2(x) = \int_{-\infty}^{+\infty} f_1(t)\, f_2(x-t)\,\mathrm{d}t,$$

c'est-à-dire que le graphe de f est donné par la figure suivante :



On peut alors calculer l'espérance, qui vaut $\mathrm{E} = \int_0^{+\infty} x\, f(x)\, \mathrm{d}x = 97,5$ comme il se doit. Les variances de deux variables indépendantes s'ajoutent, donc $\sigma^2 = \sigma_1^2 + \sigma_2^2$. Or $\sigma_1^2 = 25/3$ (calcul immédiat) et $\sigma_2^2 = 25/12$. On obtient donc $\sigma^2 = 125/12$ et $\sigma = 5\sqrt{5}/2\sqrt{3}$. Enfin, la probabilité que la masse dépasse les 100 g est égale à $\int_{100}^{105} f(x)\, \mathrm{d}x = 1/4$ (représentée en gris foncé sur la figure ci-dessus).

♦ **Solution 12.3** Graphiquement, la fonction F prend les valeurs suivantes :



La dérivée partielle croisée $\partial^2 \mathrm{F}/\partial x\, \partial y$ de F est identiquement nulle sur $\mathbb{R}^2 \setminus \mathrm{D}$, et D est de mesure nulle ; ainsi, F n'est pas absolument continue. La fonction F peut en revanche être vue comme la fonction de répartition d'un couple (X, Y) où X suit une loi normale $\mathscr{N}(0\,;\,1)$ et $\mathrm{Y} = \mathrm{X}$.

♦ **Solution 12.4** Puisque X et Y suivent des lois uniformes sur $[0\,;\,1]$ et sont indépendantes, le couple (X, Y) suit une loi uniforme sur le carré $\mathscr{C} = [0\,;\,1] \times [0\,;\,1]$. Ainsi, pour tout domaine $\mathrm{A} \subset \mathscr{C}$, la probabilité $\mathbf{P}\{(\mathrm{X}, \mathrm{Y}) \in \mathrm{A}\}$ que (X, Y) prenne sa valeur dans A peut s'identifier à l'aire de A. Cela nous autorisera une résolution presqu'entièrement géométrique de cet exercice classique.

i) La variable aléatoire A vaut $\mathrm{A} = \inf(\mathrm{X}, \mathrm{Y})$ et elle prend ses valeurs dans $\mathrm{A}(\Omega) = [0\,;\,1]$. De plus, pour tout $a \in [0\,;\,1]$, le domaine $\{\mathrm{A} \leqslant a\}$ a pour image, dans le plan, le domaine

$$\Delta_a = \big\{(x, y) \in [0\,;\,1]^2 \;;\; x \leqslant a \quad \text{ou} \quad y \leqslant a\big\},$$

domaine représenté sur la figure 12.2, et d'aire

$$\mathbf{P}\{\mathrm{A} \leqslant a\} = \mathrm{Aire}(\Delta_a) = 1 - (1-a)^2 = a(2-a).$$

Une densité de A sur $[0\,;\,1]$ est donc $a \mapsto 2 - 2a$.

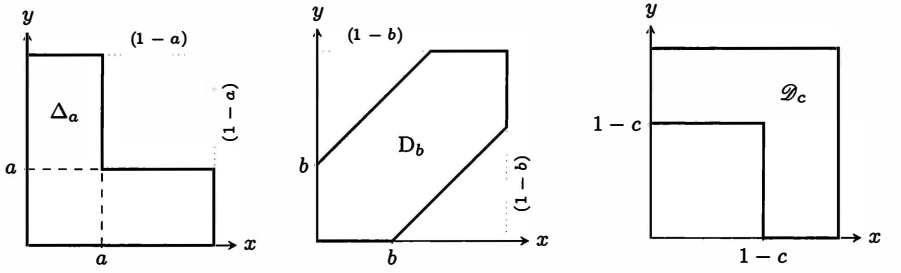


FIGURE 12.2 — (À gauche) : *le domaine* Δ_a, *image de* $\{\mathrm{A} \leqslant a\}$. (Au milieu) : *le domaine* D_b *image de* $\{\mathrm{B} \leqslant b\}$. (À droite) : *le domaine* $\mathscr{D}_c$ *image de* $\{\mathrm{C} \leqslant c\}$.

Les lecteurs à qui la géométrie ne parle pas peuvent évidemment préférer trouver ce résultat analytiquement en écrivant que

$$\mathbf{P}\{\mathrm{A} \leqslant a\} = 1 - \mathbf{P}\{\mathrm{A} > a\} = 1 - \mathbf{P}\{\mathrm{X} > a, \mathrm{Y} > a\} = 1 - (1-a)^2 = a(2-a).$$

ii) Le résultat est immédiat par symétrie du problème. On peut également représenter le domaine $\{\mathrm{C} \leqslant c\}$ par un domaine du plan. Puisque $\mathrm{C} = 1 - \sup(\mathrm{X}, \mathrm{Y})$, il y a donc équivalence entre $\mathrm{C} \leqslant c$ et $\sup(\mathrm{X}, \mathrm{Y}) \geqslant 1 - c$; le domaine image de $\{\mathrm{C} \leqslant c\}$ est donc

$$\mathscr{D}_c = \big\{(x, y) \in [0\,;\,1]^2 \;;\; x \geqslant 1 - c \text{ ou } y \geqslant 1 - c\big\}$$

(figure 12.2, à droite) : $\mathbf{P}\{\mathrm{C} \leqslant c\} = \mathrm{Aire}(\mathscr{D}_c) = 1 - (1-c)^2 = c(2-c)$. La variable C a la même fonction de répartition que A, et donc la même loi. Enfin, on calcule directement puis $\mathbf{E}(\mathrm{C}) = \mathbf{E}(\mathrm{A}) = \int_0^1 a(2-2a)\, \mathrm{d}a = \frac{1}{3}$.

iii) Par construction, $B = |X - Y|$, donc $B(\Omega) = [0\,;1]$. Pour tout $b \in [0\,;1]$, le domaine image de $\{B \leqslant b\}$ est

$$D_b = \{(x,y) \in [0\,;1]^2 \ ; \ |x-y| \leqslant b\},$$

domaine représenté au milieu de la figure 12.2. L'aire de ce domaine est encore $1-(1-b)^2$ (en rapprochant les deux triangles blancs, on forme un carré de côté $1-b$), donc :

$$\mathbf{P}\{B \leqslant b\} = \text{Aire}(D_b) = 1-(1-b)^2 = b(2-b).$$

On en déduit que $\mathbf{E}(B) = \frac{1}{3}$ également[15].

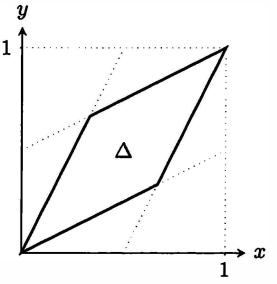
iv) Le morceau du milieu est le plus court si et seulement si $B \leqslant A$ et $B \leqslant C$. La première condition $B \leqslant A$ se traduit, sur le plan, par la double inégalité $|y-x| \leqslant x$ et $|y-x| \leqslant y$, ou encore

$$0 \leqslant y \leqslant 2x \qquad \text{et} \qquad 0 \leqslant x \leqslant 2y. \tag{$*$}$$

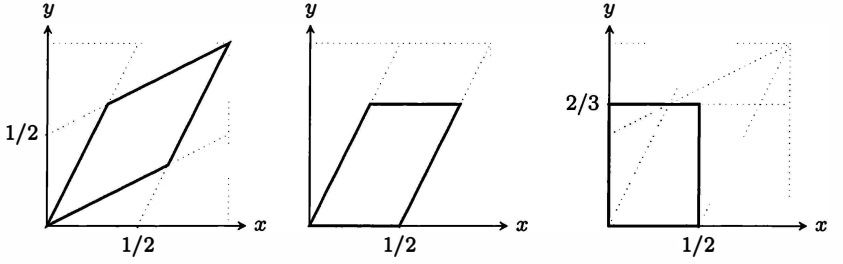
La seconde condition $B \leqslant C$, se traduit par $|y-x| \leqslant 1-x$ et $|y-x| \leqslant 1-y$, ou encore

$$2x-1 \leqslant y \leqslant 1 \qquad \text{et} \qquad 2y-1 \leqslant x \leqslant 1. \tag{$**$}$$

On définit le domaine Δ comme l'ensemble des points (x,y) de $[0\,;1]^2$ vérifiant les quatre conditions $(*)$ et $(**)$, et que l'on peut représenter ainsi :



La probabilité cherchée est $p = \iint_\Delta dx\,dy = \text{Aire}(\Delta)$. Un peu de géométrie élémentaire permet en effet de transformer, sans en changer l'aire, le domaine Δ en un rectangle de largeur $1/2$ et de hauteur $2/3$, donc d'aire $1/3$, d'où l'on déduit que $p = 1/3$.



♦ **Solution 12.5** Soit y un réel.

— Si $y \geqslant 0$, alors $\mathbf{P}\{X|X| \leqslant y\} = \mathbf{P}\{X \leqslant \sqrt{y}\} = \mathfrak{N}(\sqrt{y})$;

— si au contraire $y < 0$, $\mathbf{P}\{X|X| \leqslant y\} = \mathbf{P}\{X \leqslant -\sqrt{-y}\} = \mathfrak{N}(-\sqrt{-y})$ car la réciproque de $x \mapsto x|x|$ est la fonction

$$x \longmapsto \begin{cases} \sqrt{x} & \text{si } x \geqslant 0 \\ -\sqrt{-x} & \text{si } x < 0. \end{cases}$$

Par la formule de transfert, et puisqu'une fonction intégrable impaire est d'intégrale nulle,

$$\mathbf{E}(Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x|x|\, e^{-x^2/2}\, dx = 0.$$

---

[15]. On aurait pu montrer directement ce résultat à partir des relations $A + B + C = 1$ et $\mathbf{E}(A) = \mathbf{E}(C) = \frac{1}{3}$.

Enfin,

$$\mathbf{E}(XY) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 |x| \, e^{-x^2/2} \, dx = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} x^3 e^{-x^2/2} \, dx$$
$$= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} 2x \, e^{-x^2/2} \, dx = \frac{4}{\sqrt{2\pi}},$$

d'où $\mathbf{Cov}(X, Y) = \dfrac{2\sqrt{2}}{\sqrt{\pi}}$.

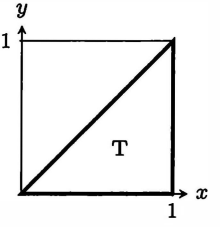
♦ **Solution 12.6** Le domaine des valeurs prises par $Z = X^2 + Y^2$ est $Z(\Omega) = \mathbb{R}^+$. Soit $z \geqslant 0$ et notons $\mathscr{D}(z)$ le disque de rayon $\sqrt{z}$; alors

$$\mathbf{P}\{X^2 + Y^2 \leqslant z\} = \iint_{\mathscr{D}(z)} n(x)\, n(y) \, dx \, dy = \frac{1}{2\pi} \int_0^{2\pi} \left(\int_0^{\sqrt{z}} r e^{-r^2/2} \, dr \right) d\theta$$
$$= \left[-e^{-r^2/2} \right]_0^{\sqrt{z}} = 1 - e^{-z/2},$$

ce qui prouve que $Z = X^2 + Y^2$ suit la loi exponentielle $\mathscr{E}(1/2)$.

♦ **Solution 12.7** Notons R et G les instants d'arrivée de Rosencrantz et de Guildenstern respectivement. Alors, R et G suivent toutes deux la loi uniforme $\mathscr{U}([0\,;\,1])$.

i) Le temps d'attente du premier arrivé est $T = |R - G|$. Notons D le carré $[0\,;\,1]^2$ et Δ le triangle sous la première diagonale.



Alors

$$\mathbf{E}(T) = \iint_D |x - y| \, dx \, dy = 2 \iint_\Delta (x - y) \, dx \, dy$$
$$= 2 \int_0^1 \left(\int_0^x (x - y) \, dy \right) dx = 2 \int_0^1 \left(x^2 - \frac{x^2}{2} \right) dx = \frac{1}{3},$$

c'est-à-dire 20 minutes.

ii) On a $X = \inf(R, G)$. Alors $X(\Omega) = [0\,;\,1]$ et, pour tout $x \in [0\,;\,1]$, on a $1 - F_X(x) = \mathbf{P}\{X \geqslant x\} = (1 - x)^2$, et donc $F_X(x) = 2x - x^2$. On en déduit une densité $f(x) = 2 - 2x$ sur $[0\,;\,1]$ et l'espérance $\mathbf{E}(X) = \int_0^1 x\, f(x) \, dx = \frac{1}{3}$. (Cf. exercice 12.4.)

De même, $Y = \sup(R, G)$ et pour tout $y \in [0\,;\,1]$, on a $F_Y(y) = y^2$, ce qui conduit (après calcul de la densité) à $\mathbf{E}(Y) = \frac{2}{3}$. Enfin, puisque $X + Y = R + G$, on a d'une part

$$\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\,\mathbf{Cov}(X, Y)$$

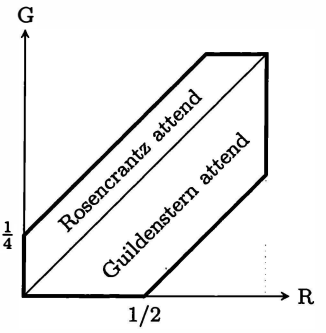
et d'autre part

$$\mathbf{V}(X + Y) = \mathbf{V}(R + G) = \mathbf{V}(R) + \mathbf{V}(G) = \frac{1}{6}.$$

Comme par ailleurs les variances de X et Y se calculent directement, $\mathbf{V}(X) = \mathbf{V}(Y) = \frac{1}{18}$, on en déduit $\mathbf{Cov}(X, Y) = \frac{1}{36}$, puis enfin le coefficient de corrélation $\rho(X, Y) = \frac{1}{2}$.

iii) La probabilité cherchée est $p = \iint_\Delta dx \, dy$, où Δ est le domaine suivant :

$$\Delta = \{(r, g) \in [0\,;\,1]^2 \ ; \ r - g \leqslant \tfrac{1}{2} \quad \text{et} \quad g - r \leqslant \tfrac{1}{4}\}.$$



Un peu de géométrie montre alors que $p = 1 - \frac{1}{2}\left(\frac{1}{2}\right)^2 - \frac{1}{2}\left(\frac{3}{4}\right)^2 = 1 - \frac{1}{8} - \frac{9}{32} = \frac{19}{32}$.
Conclusion : Rosencrantz et Guildenstern se rencontrent avec la probabilité $p = 19/32$.

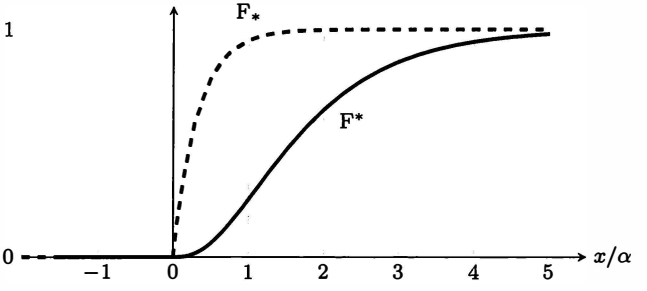
♦ **Solution 12.8** Notons F la fonction de répartition de la loi $\mathscr{E}(\alpha)$: $F(x) = 1 - e^{-\alpha x}$ pour tout $x \geqslant 0$, et $G = 1 - F$. Les variables aléatoires U et M prennent leurs valeurs dans $\mathbb{R}^+$. Pour tout $x \geqslant 0$, on a par indépendance

$$F^*(x) = \mathbf{P}\{M \leqslant x\} = \prod_{k=1}^{n} (1 - e^{-\alpha x}) = (1 - e^{-\alpha x})^n$$

tandis que

$$F_*(x) = \mathbf{P}\{U \leqslant x\} = 1 - \prod_{k=1}^{n} G(x) = 1 - e^{-n\alpha x},$$

ce qui montre que U suit une loi exponentielle de paramètre $n\alpha$, et a pour espérance $\mathbf{E}(U) = 1/n\alpha$. Ci-dessous, les fonctions de répartition F^* et F_* pour une valeur $n = 3$.



La variable M admet une densité $f^*(x) = n\alpha e^{-\alpha x}(1 - e^{-\alpha x})^{n-1}$ pour $x \geqslant 0$; puisque M est positive, elle admet une espérance dans $\mathbb{R} \cup \{+\infty\}$ et celle-ci vaut

$$\mathbf{E}(M) = \int_0^{+\infty} x\, f^*(x)\, dx = \int_0^{+\infty} x\, n\alpha\, e^{-\alpha x}(1 - e^{-\alpha x})^{n-1}\, dx.$$

Le changement de variable $t = e^{-\alpha x}$ donne

$$\mathbf{E}(M) = -\frac{n}{\alpha}\int_0^1 \ln t\,(1-t)^{n-1}\, dt = -\frac{n}{\alpha}\int_0^1 \ln(1-t)\, t^{n-1}\, dt.$$

On intègre par parties (on choisit d'intégrer $-n\,t^{n-1}$ en $1 - t^n$), cela donne

$$\mathbf{E}(M) = \frac{1}{\alpha}\int_0^1 \frac{1-t^n}{1-t}\, dt = \frac{1}{\alpha}\int_0^1 (1 + t + t^2 + \cdots + t^{n-1})\, dt = \frac{1}{\alpha}\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right).$$

Notamment, $\mathbf{E}(M) \underset{n\to\infty}{\sim} (\ln n)/\alpha$.

♦ **Solution 12.9** Voir paragraphe § 185.

♦ **Solution 12.10** Notons f_n la densité de S_n et f celle de X. Alors $f_n = f * \cdots * f$. On calcule cette quantité par récurrence, en calculant d'abord $f_2 = f * f$, puis $f_3 = f_2 * f$ et $f_4 = f_3 * f$: chacune de ces fonctions est nulle sur $\mathbb{R}^-$, et vaut, pour tout $x \geqslant 0$:

$$f_2(x) = \lambda^2\, x\, e^{-\lambda x}, \qquad f_3(x) = \lambda^3\, \frac{x^2}{2}\, e^{-\lambda x} \qquad \text{et} \qquad f_4(x) = \lambda^4\, \frac{x^3}{6}\, e^{-\lambda x}.$$

On est à même de rédiger une récurrence et de prouver que

$$f_n(x) = \begin{cases} 0 & \text{si } x \leqslant 0, \\ \lambda^n \dfrac{x^{n-1}}{(n-1)!} \mathrm{e}^{-\lambda x}, \end{cases}$$

On reconnaît une loi $\Gamma(n, \lambda)$.

♦ **Solution 12.11** Notons X_k la longueur du k-ième saut effectué par la grenouille, et S_n la longueur totale parcourue à l'issue du n-ième saut : $S_n = X_1 + X_2 + \cdots + X_n$.

Par simple linéarité, $\mathbf{E}(S_2) = \mathbf{E}(X_1) + \mathbf{E}(X_2) = \frac{1}{2} + \frac{1}{2} = 1$: la longueur moyenne parcourue au bout de deux sauts est de 1 mètre.

Notons N le nombre de saut nécessaires pour dépasser un mètre, on a donc $\{N \leqslant n\} = \{S_n \geqslant 1\}$ et $\{N > n\} = \{S_n < 1\}$. En utilisant le résultat de l'exercice 7.5, il ne reste donc plus qu'à calculer la somme $\mathbf{E}(N) = \sum_{n=0}^{\infty} \mathbf{P}\{N > n\} = \sum_{n=0}^{\infty} F_n(1)$, où F_n est la fonction de répartition de S_n.

Les sauts successifs de la grenouille sont modélisés par des variables aléatoires indépendantes de même loi uniforme $\mathscr{U}([0\,;1])$; ainsi, S_n est une variable aléatoire absolument continue et, si l'on note f_n la densité de S_n et f la densité uniforme sur $[0\,;1]$, on a $f_{n+1} = f_n * f$ pour tout $n \in \mathbb{N}$, et donc $f_n = f^{*n} = f * f * \cdots * f$. Son support est inclus dans $[0\,;n]$.

Pour tout $x \in [0\,;1]$, la fonction de répartition de S_n vaut

$$F_n(x) = \mathbf{P}\{S_n \leqslant x\} = \mathbf{P}\{S_n < x\} = \int_0^x f^{*n}(t)\,\mathrm{d}t$$
$$= \int_0^x f^{*(n-1)} * f(t)\,\mathrm{d}t = \int_0^x \int_0^t f^{*(n-1)}(s)\, f(t-s)\,\mathrm{d}s\,\mathrm{d}t.$$

Dans l'intégrale précédente, la quantité $t-s$ est toujours comprise entre 0 et 1, donc $f(t-s) = 1$. On trouve donc finalement

$$F_n(x) = \int_0^x F_{n-1}(t)\,\mathrm{d}t$$

pour tout $x \in [0\,;1]$ et tout $n \in \mathbb{N}$. On peut calculer les premiers termes :

$$F_1(x) = \int_0^x 1\,\mathrm{d}t = x \qquad F_2(x) = \int_0^x t\,\mathrm{d}t = \frac{x^2}{2} \qquad F_3(x) = \int_0^x \frac{t^2}{2}\,\mathrm{d}t = \frac{x^3}{6}$$

et une récurrence immédiate confirme que $F_n(x) = x^n/n!$ pour tout $n \in \mathbb{N}$. Finalement, on obtient

$$\mathbf{E}(N) = \sum_{n=0}^{\infty} F_n(1) = \sum_{n=0}^{\infty} \frac{1}{n!} = \mathrm{e} \approx 2,71828...$$

On remarquera que ce nombre n'est pas celui qu'un raisonnement naïf laisserait croire (puisque chaque saut est de longueur moyenne $\mathbf{E}(X_k) = 1/2$, on pouvait croire que le nombre moyen de sauts nécessaires est 2).

Si le résultat final ne saute pas aux yeux, un raisonnement simple permet cependant de montrer que $\mathbf{E}(N) > 2$. En effet, le nombre de sauts nécessaires ne peut être égal à 1 que si le premier saut mène à une longueur égale à 1 mètre, ce qui n'a lieu qu'avec une probabilité nulle. Pour toutes les autres réalisations du hasard, le nombre de sauts nécessaires sera au moins égal à 2 et, dans la moitié des cas[16], au moins égale à 3.

♦ **Solution 12.12**

i) Une densité du couple est la fonction f définie par $f(x,y) = \mathrm{e}^{-x}$ si $x \geqslant y \geqslant 0$, et $f(x,y) = 0$ sinon. On en déduit par intégration que

$$F(x,y) = \begin{cases} 1 - \mathrm{e}^{-y} - y\,\mathrm{e}^{-x} & \text{si } 0 \leqslant y \leqslant x, \\ 1 - (1+x)\,\mathrm{e}^{-x} & \text{si } 0 \leqslant x \leqslant y, \\ 0 & \text{sinon.} \end{cases}$$

La loi marginale de Y est obtenue en faisant tendre x vers $+\infty$ dans la fonction de répartition conjointe ; on obtient finalement que Y suit la loi exponentielle $\mathscr{E}(1)$.

(16). Le lecteur trouvera sans peine pourquoi.

ii) On calcule que

$$\mathbf{E}(XY) = \int_0^{+\infty} \left(\int_0^x xy\, e^{-x}\, dy \right) dx = \int_0^{+\infty} \frac{x^3}{2} e^{-x}\, dx = 3.$$

Or $\mathbf{E}(X) = \int_0^{+\infty} x^2 e^{-x}\, dx = 2$ et $\mathbf{E}(Y) = 1$ donc $\mathbf{Cov}(X, Y) = 3 - 2 = 1$.

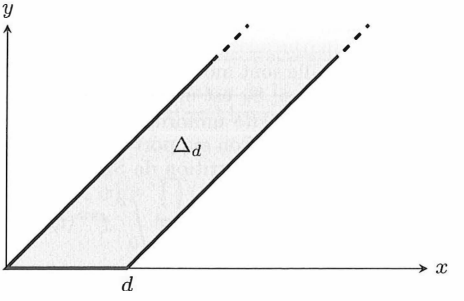
iii) Tout d'abord, $(X - Y)(\Omega) = [0\,;\, +\infty[$. Soit $d \in [0\,;\, +\infty[$. Alors

$$\mathbf{P}\{X - Y \leqslant d\} = \iint_{\Delta_d} e^{-x}\, dx\, dy,$$

où l'on a posé

$$\Delta_d := \{(x, y) \in \mathbb{R}^2 \ ;\ 0 \leqslant y \leqslant x \text{ et } x \leqslant y + d\},$$

ensemble représenté ci-dessous :



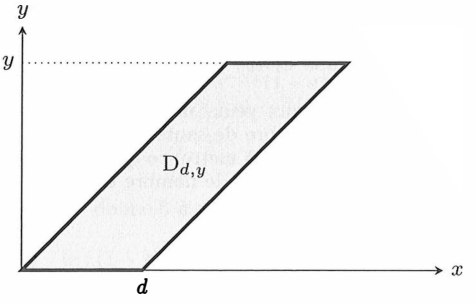
On trouve alors

$$F_{X-Y}(d) = \int_0^{+\infty} \left(\int_y^{y+d} e^{-x}\, dx \right) dy = \int_0^{+\infty} (e^{-y} - e^{-y-d})\, dy = 1 - e^{-d},$$

ce qui montre que $X - Y$ suit la loi $\mathscr{E}(1)$.
On trouve enfin que, pour tout $(d, y) \in \mathbb{R}^{+2}$, on a

$$\begin{aligned} \mathbf{P}(\{X - Y \leqslant d\} \cap \{Y \leqslant y\}) &= \iint_{D_{d,y}} e^{-x}\, dx = \int_0^y (1 - e^{-d})\, e^{-t}\, dt \\ &= (1 - e^{-d})(1 - e^{-y}) = F_{X-Y}(d) \cdot F_Y(y). \end{aligned}$$

où l'on a noté $D_{d,y}$ le domaine représenté ci-dessous :



On en déduit que Y et $X - Y$ sont indépendantes.

iv) Par indépendance, on a $0 = \mathbf{Cov}(X - Y, Y) = \mathbf{Cov}(X, Y) - \mathbf{Cov}(Y, Y)$ ce qui permet de retrouver que $\mathbf{Cov}(X, Y) = 1$.

v) Enfin, $(Y/X)(\Omega) = [0\,;\, 1]$ et, pour tout $q \in [0\,;\, 1]$, on a

$$\mathbf{P}\{Y/X \leqslant q\} = \int_0^{+\infty} \left(\int_0^{qx} e^{-x}\, dy \right) dx = \int_0^{+\infty} qx\, e^{-x}\, dx = q.$$

Conclusion : la variable Y/X suit la loi uniforme $\mathscr{U}([0\,;\, 1])$.

♦ **Solution 12.13** Toutes les variables aléatoires X_k admettent pour densité $f(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}+}(x)$ et pour fonction de répartition $F(x) = 1 - e^{-\lambda x}$ pour $x \geqslant 0$, $F(x) = 0$ sinon.

i) L'ensemble des valeurs prises par U_n est $U_n(\Omega) = [0\,;+\infty[$ et, pour tout $x > 0$,

$$\begin{aligned}\mathbf{P}\{U_n \leqslant x\} &= 1 - \mathbf{P}\{U_n > x\} = 1 - \mathbf{P}\big(\{X_2 > x\} \cap \cdots \cap \{X_n > x\}\big)\\ &= 1 - \big(\mathbf{P}\{X_2 > x\}\big)^{n-1} = 1 - e^{-(n-1)\lambda x},\end{aligned}$$

ce qui prouve que $\mathscr{U}_n \rightsquigarrow \mathscr{E}\big((n-1)\lambda\big)$.

ii) Posons $T = U_n - X_1$. Puisque U_n et X_1 sont indépendantes, on obtient une fonction de densité de T_n par la formule de convolution ; pour cela, on a besoin d'une densité de $-X_1$, qui est la symétrisée de celle de X_1 :

$$f_{-X_1}(x) = f(-x) = \begin{cases}\lambda e^{\lambda x} & \text{si } x < 0,\\ 0 & \text{sinon.}\end{cases}$$

On en déduit par convolution que T admet pour densité

$$f_T(t) = \int_{-\infty}^{+\infty} f_{U_n}(u)\, f_{-X_1}(t-u)\,du.$$

Dans cette équation, on doit avoir, pour obtenir la non-nullité de l'intégrande, les deux conditions $u \geqslant 0$ et $u \geqslant t$.

1er cas : $t \geqslant 0$. Alors

$$f_T(t) = \int_t^{+\infty} (n-1)\,\lambda^2\, e^{\lambda t}\, e^{-n\lambda u}\,du = \frac{(n-1)\,\lambda}{n}\, e^{-(n-1)\lambda t}.$$

2er cas : $t \leqslant 0$. Alors

$$f_T(t) = \int_0^{+\infty} (n-1)\lambda^2\, e^{\lambda t}\, e^{-n\lambda u}\,du = \frac{(n-1)\,\lambda}{n}\, e^{\lambda t}.$$

(On vérifie que $\int_{-\infty}^{+\infty} f_T = 1$ comme pour toute fonction de densité bien élevée.) Enfin, on en déduit que

$$\mathbf{P}\{U_n \geqslant X_1\} = \mathbf{P}\{T \geqslant 0\} = \int_0^{+\infty} f_T(t)\,dt = \frac{1}{n}.$$

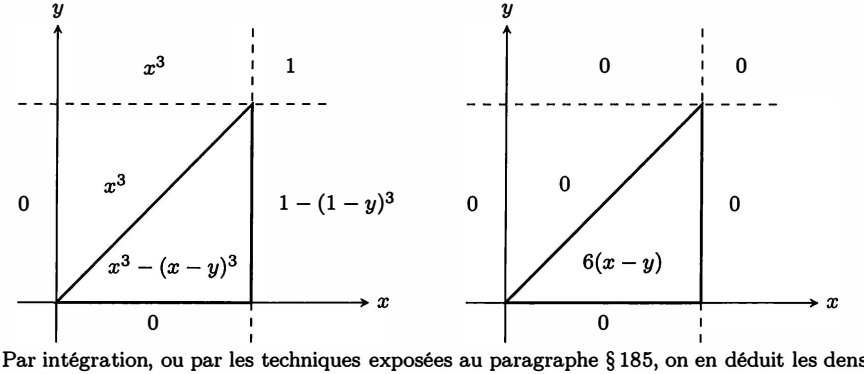
♦ **Solution 12.14** Soient $0 \leqslant y \leqslant x \leqslant 1$; on a, par indépendance,

$$\begin{aligned}\mathbf{P}\{X \leqslant x, Y > y\} &= \mathbf{P}\big\{U, V, W \in]\,y\,;\,x\,]\,\big\}\\ &= \mathbf{P}\{y < U \leqslant x\}\ \mathbf{P}\{y < V \leqslant x\}\ \mathbf{P}\{y < W \leqslant x\}\\ &= (x-y)^3.\end{aligned}$$

Puisque l'on peut décomposer $\{X \leqslant x\} = \{X \leqslant x, Y \leqslant y\} + \{X \leqslant x, Y > y\}$, on en déduit la valeur $F(x,y)$ de la fonction de répartition du couple (X, Y) pour tous $0 \leqslant y \leqslant x \leqslant 1$:

$$F(x,y) = \mathbf{P}\{X \leqslant x, Y \leqslant y\} = \mathbf{P}\{X \leqslant x\} - \mathbf{P}\{X \leqslant x, Y > y\} = x^3 - (x-y)^3.$$

Si $0 \leqslant x < y \leqslant 1$, alors $\mathbf{P}\{X \leqslant x, Y \leqslant y\} = \mathbf{P}\{X \leqslant x\} = x^3$. On obtient donc le graphe suivant pour les valeurs de la fonction de répartition F *(à gauche)* et de la densité f *(à droite)* :



Par intégration, ou par les techniques exposées au paragraphe § 185, on en déduit les densités marginales

$$\forall (x,y) \in [0\,;1]^2 \qquad f_X(x) = 3x^2 \qquad \text{et} \qquad f_Y(y) = 3(1-y)^2$$

qui nous permettent de calculer d'une part la densité de X conditionnellement à $Y = y$:

$$f_X(x \mid y) = f_X(x \mid Y = y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} \frac{2(x-y)}{(1-y)^2} & \text{si } 0 \leqslant y \leqslant x < 1 \\ 0 & \text{sinon.} \end{cases}$$

et d'autre part les espérances marginales

$$\mathbf{E}(X) = \frac{3}{4} \qquad \text{et} \qquad \mathbf{E}(Y) = \frac{1}{4}.$$

Enfin,

$$\begin{aligned} \mathbf{Cov}(X,Y) &= \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y) = \iint xy\, f(x,y)\,dx\,dy - \mathbf{E}(X)\,\mathbf{E}(Y) \\ &= 6\int_0^1 \left(\int_0^x xy(x-y)\,dy \right) dx - \frac{3}{4}\cdot\frac{1}{4} = \frac{1}{5} - \frac{3}{16} = \frac{1}{80}. \end{aligned}$$

On calcule enfin $\sigma_X = \sigma_Y = \sqrt{3}/4\sqrt{5}$ et $\rho(X,Y) = \frac{1}{3}$.

Que le coefficient de corrélation soit positif est cohérent avec l'idée que « plus Y est grand, plus X l'est ».

♦ **Solution 12.15** On considère une variable aléatoire B suivant la loi de Bernoulli de paramètre λ : $B \rightsquigarrow \mathscr{B}(\lambda)$. On suppose cette variable indépendante de X et Y. On définit la variable aléatoire Z par

$$Z(\omega) = \begin{cases} X(\omega) & \text{si } B(\omega) = 1, \\ Y(\omega) & \text{si } B(\omega) = 0. \end{cases}$$

La formule des probabilités totales montre que Z a pour fonction de répartition $\lambda F + (1-\lambda)G$, et donc pour densité $\lambda f + (1-\lambda)g$.

♦ **Solution 12.16**

i) Un simple calcul mène à $\mathbf{E}(X) = 0$ et $\mathbf{V}(X) = 2$, donc $\sigma_X = \sqrt{2}$.

ii) La fonction f étant paire, $-Y$ admet f comme densité. Ainsi, une densité de Z est donnée par la formule de convolution : $g = f * f$, c'est-à-dire que, pour tout $x \in \mathbb{R}$, on a

$$g(x) = \frac{1}{4}\int_{-\infty}^{+\infty} e^{-|t|}\, e^{-|x-t|}\,dt.$$

•**1er cas : $x \geqslant 0$.** Alors

$$g(x) = \frac{1}{4}\int_{-\infty}^{0} e^{t}\, e^{t-x}\,dt + \frac{1}{4}\int_0^x e^{-t}\, e^{t-x}\,dt + \frac{1}{4}\int_x^{+\infty} e^{-t}\, e^{x-t}\,dt = \frac{1}{4}(x+1)\,e^{-x}.$$

•**2e cas : $x \leqslant 0$.** Avec des techniques semblables, on trouve $g(x) = \frac{1}{4}(1-x)\,e^{x}$.

Conclusion : une densité de Z est $g : x \mapsto \frac{1}{4}\big(1+|x|\big)\,e^{-|x|}$.

Troisième partie

Plus formellement...

Chapitre 13

Variables aléatoires : cas général

Certaines variables aléatoires ne sont ni discrètes, ni absolument continues. Si la fonction de répartition de certaines de ces variables peut s'écrire comme une combinaison linéaire d'une fonction de répartition discrète et d'une fonction de répartition absolument continue, le cas général mérite une étude à part entière.

Nous terminons ce court chapitre en établissant un théorème d'une grande importance : toute variable aléatoire positive est limite (au sens de la convergence simple) d'une suite croissante de variables aléatoires simples (étagées). Ce théorème permettra de construire la notion générale d'espérance d'une variable aléatoire, ce qui sera l'objet du chapitre suivant.

Dans tout ce chapitre, $(\Omega, \mathfrak{T}, \mathbf{P})$ est un espace probabilisé ; en l'absence de précision, toute les variables aléatoires rencontrées sont définies sur cet espace, et à valeurs réelles.

13.1 MESURABILITÉ

Définition d'une variable aléatoire § 193

Soit $X : \Omega \to \mathbb{R}$ une application à valeurs réelles. Afin d'étudier la probabilité image induite par une fonction $X : \Omega \to \mathbb{R}$, c'est-à-dire de calculer la probabilité $\mathbf{P}\{X \in B\}$ pour que X prenne ses valeurs dans un borélien B de $\mathbb{R}$, il est nécessaire que les ensembles $\{X \in B\}$ soient tous éléments de la tribu $\mathfrak{T}$.

Définition 13.1 Soit $(\Omega, \mathfrak{T})$ un espace probabilisable. Une **variable aléatoire réelle** est une application $X : \Omega \to \mathbb{R}$ telle que, pour tout borélien B de $\mathbb{R}$, l'événement $\{X \in B\}$ est un élément de la tribu $\mathfrak{T}$. On dit également que X est une application **mesurable** de l'espace mesurable $(\Omega, \mathfrak{T})$ dans l'espace mesurable $(\mathbb{R}, \mathfrak{B})$.

On notera :

- $\mathcal{L}^0(\Omega, \mathfrak{T})$ l'espace des applications mesurables de $(\Omega, \mathfrak{T})$ dans $(\mathbb{R}, \mathfrak{B})$;
- $\mathcal{L}^0_+(\Omega, \mathfrak{T})$ l'espace des applications mesurables positives, et

- $\mathcal{L}^0_b(\Omega, \mathfrak{T})$ celui des applications mesurables bornées.

Si aucune confusion n'est possible, on abrégera en $\mathcal{L}^0(\Omega)$, $\mathcal{L}^0_+(\Omega)$, $\mathcal{L}^0_b(\Omega)$ — voire en $\mathcal{L}^0$, $\mathcal{L}^0_+$ et $\mathcal{L}^0_b$.

On peut généraliser cette définition aux fonctions à valeurs dans $\mathbb{C}$, dans $\mathbb{R}^n$, ou même à valeurs dans un espace mesurable $(F, \mathfrak{F})$, c'est-à-dire un espace F muni d'une tribu $\mathfrak{F}$ (voir le paragraphe A69, page 700).

Si $(E, \mathfrak{E})$ et $(F, \mathfrak{F})$ sont des espaces mesurables, une application $f : E \to F$ est dite **$\mathfrak{E}$-mesurable** si $f^{-1}(B) \in \mathfrak{E}$ pour tout $B \in \mathfrak{F}$. Lorsque l'espace d'arrivée F est la droite $\mathbb{R}$, nous le munirons *toujours* de la tribu $\mathfrak{B}$ des boréliens ; de même, si $F = \mathbb{R}^n$, il sera muni de la tribu $\mathfrak{B}^n$ (A66).

§ 194 **Calcul des images réciproques**

Nous donnons ici quelques résultats simples, tous vérifiables à titre d'exercice. Soit $X : E \to F$ une application. Pour toutes parties $B_1, B_2, \dots$ de F, on a les propriétés suivantes (que l'on retranscrit à droite avec les notations probabilistes habituelles, qui montrent ainsi leur souplesse d'utilisation) :

$$X^{-1}\Big(\bigcup_n B_n\Big) = \bigcup_n X^{-1}(B_n) \qquad \Big\{X \in \bigcup_n B_n\Big\} = \bigcup_n \{X \in B_n\}$$

$$X^{-1}\Big(\bigcap_n B_n\Big) = \bigcap_n X^{-1}(B_n) \qquad \Big\{X \in \bigcap_n B_n\Big\} = \bigcap_n \{X \in B_n\}$$

$$X^{-1}(B^c) = \big(X^{-1}(B)\big)^c \qquad \{X \notin B\} = \{X \in B\}^c$$

(Notez que les unions et intersections peuvent être dénombrables ou non.) De plus, si les B_n sont disjoints deux à deux, il en est de même des ensembles $\{X \in B_n\}$, ce qui permet d'écrire la règle supplémentaire

$$X^{-1}\Big(\sum_n B_n\Big) = \sum_n X^{-1}(B_n) \qquad \Big\{X \in \sum_n B_n\Big\} = \sum_n \{X \in B_n\}.$$

On retiendra que :

« X^{-1} *commute avec* $\bigcup$, $\bigcap$, $\sum$ *et passage au complémentaire* ».

§ 195 **Quelques résultats sur la mesurabilité**

Les résultats de ce paragraphe sont utiles pour l'étude générale des fonctions mesurables ; le lecteur pressé peut se contenter, en première lecture, d'en parcourir les énoncés. Le premier de ces résultats montre que, de même que la composée de deux fonctions continues est continue, la composée de deux fonctions mesurables est mesurable.

■ **Proposition 13.2 (Un peu de stabilité)** *Si* X *est une application mesurable de* $(E, \mathfrak{E})$ *dans* $(F, \mathfrak{F})$ *et* Y *une applications mesurable de* $(F, \mathfrak{F})$ *dans* $(G, \mathfrak{G})$, *alors* $Y \circ X$ *est mesurable de* $(E, \mathfrak{E})$ *dans* $(G, \mathfrak{G})$.

DÉMONSTRATION : Pour toute partie $C \in \mathfrak{G}$, la partie $B := Y^{-1}(C)$ est dans $\mathfrak{F}$, donc la partie $A := X^{-1}(B)$ est dans $\mathfrak{E}$. Il suffit alors de remarquer que $A = (Y \circ X)^{-1}(C)$, propriété que l'on démontre aisément avec les notations standard, mais encore plus aisément avec les notations probabilistes :

$$(Y \circ X)^{-1}(C) = \{Y(X) \in C\} = \{X \in Y^{-1}(C)\} = \{X \in B\} = X^{-1}(B) = A. \qquad \blacksquare$$

Le second résultat, fondamental, stipule que, pour montrer qu'une application X de $(E, \mathfrak{E})$ dans $(F, \mathfrak{F})$ est mesurable, on peut se limiter à l'étude des ensembles $\{X \in A\}$ lorsque A parcourt un ensemble beaucoup plus restreint que la tribu $\mathfrak{F}$ tout entière :

■ **Théorème 13.3** *Soient* $(E, \mathfrak{E})$ *et* $(F, \mathfrak{F})$ *deux espaces mesurables. Soit* $\mathfrak{A}$ *un ensemble de parties de* F *qui engendre* $\mathfrak{F}$*, c'est-à-dire tel que* $\sigma(\mathfrak{A}) = \mathfrak{F}$*. Alors* X *est mesurable si et seulement si* $\{X \in A\} \in \mathfrak{E}$ *pour tout* $A \in \mathfrak{A}$*.*

DÉMONSTRATION : Le sens direct est immédiat.

Réciproquement, supposons que $\{X \in A\} \in \mathfrak{E}$ pour tout $A \in \mathfrak{A}$. Considérons alors l'ensemble $\mathfrak{H}$ des parties A de F telles que $\{X \in A\}$ est dans $\mathfrak{E}$. Il est aisé de constater que $\varnothing \in \mathfrak{H}$, que $\mathfrak{H}$ est stable par union dénombrable (en vertu des relations du § 194) et que $\mathfrak{H}$ est stable par passage au complémentaire ; par conséquent, $\mathfrak{H}$ est une tribu. Or, l'hypothèse faite sur $\mathfrak{A}$ s'écrit $\mathfrak{A} \subset \mathfrak{H}$; la proposition 3.16 permet d'en déduire $\mathfrak{F} \subset \sigma(\mathfrak{H}) = \mathfrak{H}$, ce qui est la transcription de la propriété « X est mesurable ». ∎

Nous donnons deux corollaires de ce résultat ; nous reprenons pour l'occasion les notations usuelles : espace Ω, tribu $\mathfrak{T}$.

■ **Corollaire 13.4 (Critère pratique de mesurabilité)** *Soit* X *une application sur un espace* $(\Omega, \mathfrak{T})$*. Alors* X *est une variable aléatoire si et seulement si les événements* $\{X \leqslant x\}$ *sont dans* $\mathfrak{T}$ *pour tout* $x \in \mathbb{R}$*. Cette propriété est encore vraie si l'on se limite à l'ensemble (dénombrable !) des* x *rationnels.*

DÉMONSTRATION : Notons $\mathfrak{A}$ l'ensemble des intervalles $]-\infty ; x]$ pour x parcourant $\mathbb{R}$. Alors $\mathfrak{A}$ engendre la tribu $\mathfrak{B}(\mathbb{R})$ des boréliens et on applique la proposition 13.3.

De même, l'ensemble $\mathfrak{C}$ des intervalles $]-\infty ; x]$ pour x parcourant $\mathbb{Q}$ engendre la tribu des boréliens. ∎

■ **Corollaire 13.5** *Les applications continues de* $\mathbb{R}^n$ *dans* $\mathbb{R}$ *sont mesurables (pour les tribus boréliennes).*

DÉMONSTRATION : Rappelons qu'une application $f : \mathbb{R}^n \to \mathbb{R}$ est continue si et seulement si l'image réciproque de tout ouvert de $\mathbb{R}$ est un ouvert de $\mathbb{R}^n$. Soit donc f une application continue. Notons $\mathfrak{O}$ la classe des ouverts de $\mathbb{R}$. Pour tout $U \in \mathfrak{O}$, $f^{-1}(U)$ est un ouvert de $\mathbb{R}^n$, et donc un borélien. Puisqu'enfin la tribu $\sigma(\mathfrak{O})$ engendrée par les ouverts n'est autre que $\mathfrak{B}(\mathbb{R})$, le théorème 13.3 permet de conclure. ∎

■ **Corollaire 13.6 (Encore de la stabilité)** *Soient* X *et* Y *deux applications mesurables de* $(E, \mathfrak{E})$ *et à valeurs réelles. Alors* $X + Y$, XY, $\max(X, Y)$, $\min(X, Y)$ *sont mesurables. Si* $\varphi : \mathbb{R} \to \mathbb{R}$ *est mesurable, alors* $\varphi(X)$ *est mesurable.*

DÉMONSTRATION : Les applications qui à (x, y) associent respectivement $x + y$, xy, $\max(x, y)$ et $\min(x, y)$ sont continues, donc mesurables. La composée d'applications mesurables est encore mesurable. ∎

Désormais, nous utiliserons donc, sans rappeler à chaque fois le théorème utilisé, des propriétés du style : si X et Y sont des variables aléatoires, alors $\min(X, Y)$ est une variable aléatoire.

■ **Corollaire 13.7 (Toujours de la stabilité)** *Soit* $(X_n)_{n \geqslant 0}$ *une suite de variables aléatoires réelles, qui converge simplement vers* $X : \Omega \to \mathbb{R}$*. Alors* X *est mesurable : c'est encore une variable aléatoire.*

Plus généralement, $S = \sup_n X_n$, $I = \inf_n X_n$, $X_* = \liminf_n X_n$ *et* $X^* = \limsup_n X_n$ *sont des variables aléatoires à valeurs dans* $[-\infty ; +\infty]$*.*

DÉMONSTRATION : Pour tout réel x, on a

$$\{S \leqslant x\} = \bigcap_{n=0}^{\infty} \{X_n \leqslant x\}$$

et chaque $\{X_n \leqslant x\}$ étant dans la tribu $\mathfrak{T}$, leur intersection l'est également. De plus, on peut écrire

$$\{S = -\infty\} = \bigcap_{n=0}^{\infty} \{S \leqslant -n\},$$

ce qui prouve que $\{S = -\infty\}$ est un élément de $\mathfrak{T}$. Un raisonnement similaire prouve que $\{S = +\infty\}$ est également dans $\mathfrak{T}$.

On procède de même pour I en notant que les $\{X_n > x\}$ sont également éléments de $\mathfrak{T}$ par passage au complémentaire. Le corollaire 13.4 permet de conclure que S et I sont des variables aléatoires.

Traitons maintenant X^*. On pose $Y_n = \sup_{k \geqslant n} X_k$ pour tout $n \in \mathbb{N}$: ce sont des variables aléatoires d'après ce qui précède, et $X^* = \inf Y_n$ est donc une variable aléatoire. On fait de même pour $X_* = \sup Z_n$, où $Z_n = \inf_{k \geqslant n} X_k$.

Pour conclure, si $X_n \to X$, alors $X = \limsup X_n = \liminf X_n$ est encore une variable aléatoire. ∎

13.2 FONCTION DE RÉPARTITION

§196 **Loi d'une variable aléatoire**

Soit X une variable aléatoire sur $(\Omega, \mathfrak{T}, \mathbf{P})$. On appelle **loi de X** l'application $\mathbf{P}_X$ définie sur la tribu $\mathfrak{B}$ des boréliens par

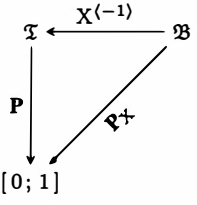
$$\begin{aligned} \mathbf{P}_X : \ \mathfrak{B} &\longrightarrow [0\,;1] \\ B &\longmapsto \mathbf{P}\{X \in B\} = \mathbf{P}\big(X^{-1}(B)\big). \end{aligned}$$

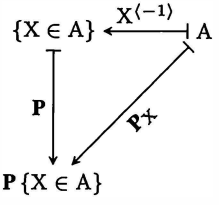
On l'appelle également **probabilité image** de **P** par X.

Remarque 13.8 (Définition plus formelle) L'application $X : \Omega \to \mathbb{R}$ permet de définir une application, que l'on notera $X^{\langle -1 \rangle}$, qui va de l'ensemble des boréliens de $\mathbb{R}$ dans la tribu $\mathfrak{T}$:

$$\begin{aligned} X^{\langle -1 \rangle} : \ \mathfrak{B} &\longrightarrow \mathfrak{T}, \\ A &\longmapsto X^{-1}(A) = \{X \in A\}. \end{aligned}$$

Avec cette définition, la probabilité image de X est $\mathbf{P}_X := \mathbf{P} \circ X^{\langle -1 \rangle}$; on peut la représenter comme sur le diagramme suivant :





■ **Proposition 13.9** *Si* X *est une variable aléatoire, alors sa loi* $\mathbf{P}_X$ *est une mesure de probabilité sur* $\mathbb{R}$.

DÉMONSTRATION : Vérifions explicitement les axiomes d'une mesure de probabilité.

- $\mathbf{P}_X$ est évidemment à valeurs positives ;
- $\mathbf{P}_X(\mathbb{R}) = \mathbf{P}\{X \in \mathbb{R}\} = \mathbf{P}(\Omega) = 1$;
- si $(B_n)_{n \geqslant 0}$ est une suite disjointe de boréliens, alors les événements $\{X \in B_n\}$ sont deux à deux disjoints et donc

$$\mathbf{P}_X\Big(\sum_{n=0}^{\infty} B_n\Big) = \mathbf{P}\Big\{X \in \sum_{n=0}^{\infty} B_n\Big\} = \mathbf{P}\Big(\sum_{n=0}^{\infty} \{X \in B_n\}\Big) \overset{*}{=} \sum_{n=0}^{\infty} \mathbf{P}\{X \in B_n\} = \sum_{n=0}^{\infty} \mathbf{P}_X(B_n).$$

l'avant-dernière égalité (*) venant de la σ-additivité de la mesure de probabilité **P**. ∎

§197 **Fonction de répartition**

Si X est une variable aléatoire réelle, on définit sa **fonction de répartition** F par la formule, valable pour tout réel x :

$$F(x) := \mathbf{P}\{X \leqslant x\}.$$

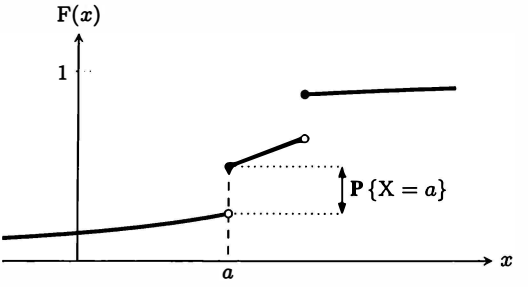


FIGURE 13.1 — *Une fonction de répartition est croissante et continue à droite. En un point de discontinuité x, la hauteur du saut de discontinuité est égal à la probabilité que la variable prenne la valeur x.*

Comme nous l'avons déjà évoqué dans les chapitres précédents, la fonction de répartition contient absolument toutes les informations relatives à la loi de X :

■ **Théorème 13.10** *La fonction de répartition caractérise la loi : si* X *et* Y *sont deux variables aléatoires de même fonction de répartition, alors elles ont même loi.*

Ce résultat est démontré dans un chapitre à part (corollaire B.19 page 706) car il nécessite un petit détour technique par le lemme de classe monotone. On peut n'en retenir pour l'instant que ceci : on peut, en principe, calculer les probabilités $\mathbf{P}\{X \in A\}$ pour tout borélien A de $\mathbb{R}$ avec les seules valeurs de la fonction F. En pratique, un tel calcul est impossible si A n'est pas une partie « simple » de $\mathbb{R}$, mais pour les exemples courants, il n'y a aucune difficulté :

$$\begin{aligned}
\mathbf{P}\{a < X \leqslant b\} &= F(b) - F(a),\\
\mathbf{P}\{X = a\} &= F(a) - F(a^-),\\
\mathbf{P}\{a \leqslant X \leqslant b\} &= F(b) - F(a^-),\\
\mathbf{P}\Big\{X \in \sum_k]\, a_k\, ;\, b_k\,]\Big\} &= \sum_k F(b_k) - F(a_k).
\end{aligned}$$

On rappelle pour mémoire les principales propriétés d'une fonction de répartition, déjà démontrées au début du chapitre 6.

■ **Théorème 13.11 (Propriétés de la fonction de répartition)** *Soit* X *une variable aléatoire de fonction de répartition* F. *Alors*

i) F *est croissante sur* $\mathbb{R}$.

ii) F *est continue à droite (mais pas forcément à gauche) ;*

iii) Pour tout x réel, $\mathbf{P}\{X = x\} = F(x) - F(x^-)$.

iv) $\lim_{x\to+\infty} F(x) = 1$ *et* $\lim_{x\to-\infty} F(x) = 0$.

Remarque 13.12 L'usage, dans certains ouvrages, est de définir la fonction de répartition par une inégalité stricte : $F(x) := \mathbf{P}\{X < x\}$. Dans ce cas, F est continue à gauche et $\mathbf{P}\{X = x\} = F(x^+) - F(x)$. C'est ainsi qu'est rédigé l'ouvrage fondateur de Kolmogorov [62]. Cette convention est désormais largement minoritaire.

§ 198 Variables discrètes, continues, mixtes

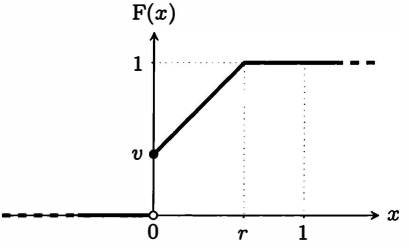
Nous avons abondamment rencontré des variables aléatoires de deux types : variables aléatoires discrètes, et variables aléatoires absolument continues. Les premières sont caractérisées par le fait que l'ensemble $X(\Omega)$ des valeurs prises par X est dénombrable[1], et leurs fonctions de répartitions sont des fonctions étagées, présentant en chaque x_k de $X(\Omega)$ un saut de discontinuité. Les secondes ont une fonction de répartition presque partout dérivable de la forme $F(x) = \int_{-\infty}^{x} f(t)\,dt$, où f est une fonction positive d'intégrale égale à 1.

Voici deux exemples *mixtes*, combinant les caractères discret et à densité.

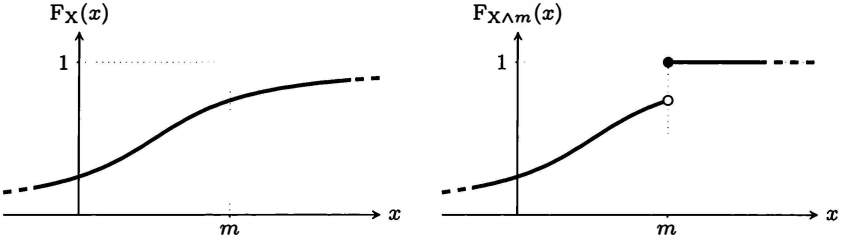
Exemple 13.13 (Temps d'attente au feu tricolore) On modélise le temps d'attente à un feu de circulation. Un tel feu est supposé être au rouge pendant une durée r, puis au vert pendant une durée v, avant de recommencer le cycle. Quitte à changer les unités de mesure, on peut supposer que $r + v = 1$. On suppose que, lorsqu'une voiture arrive à ce feu, celui-ci est dans un état quelconque : l'instant d'arrivée dans le cycle suit une loi uniforme sur $[0\,;1]$. Si l'automobiliste arrive pendant la période « rouge », il ne démarre qu'à la fin de cette période. Bien sûr, s'il arrive pendant que le feu est au vert, il ne s'arrête pas. On note X le temps d'attente au feu.

- La probabilité que l'automobiliste attende un temps strictement négatif est évidemment nulle. En aucun cas, il n'attendra plus que r, donc $F(r) = \mathbf{P}\{X \leqslant r\} = 1$.
- La probabilité que l'automobiliste n'ait pas à attendre (c'est-à-dire $X = 0$) est égale à la probabilité qu'il arrive pendant que le feu est au vert : $F(0) = v$.
- Enfin, la probabilité que la voiture ait à attendre un temps inférieur ou égal à $x \in [0\,;r]$ est donnée (en notant R l'événement « le feu est rouge » et V « le feu est vert ») par

$$F(x) = \mathbf{P}\{X \leqslant x\} = \mathbf{P}(R) \cdot \mathbf{P}_R\{X \leqslant x\} + \mathbf{P}(V) \cdot \mathbf{P}_V\{X \leqslant x\} = r \cdot \frac{x}{r} + v \cdot 1 = x + v.$$



Exemple 13.14 (Variables tronquées) Soit X une variable aléatoire continue. Pour tout réel m, on définit la **variable tronquée** $Y = \min(X, m)$, que l'on note également $Y = X \wedge m$. Les fonctions de répartition de X et de Y ont alors les allures génériques suivantes :



♦ ***Exercice 13.1*** On peut également tronquer X en ramenant à zéro les valeurs trop grandes, c'est-à-dire en posant $Z = X^{(m)} := X\mathbf{1}_{|X| \leqslant m}$. Quelle serait alors l'allure du graphe de la fonction de répartition de Z ?

---

(1). On peut alors (sans obligation) choisir un modèle où l'univers Ω est lui-même dénombrable — du moins jusqu'à ce que la fantaisie nous prenne de considérer, par exemple, une suite $(X_n)_{n \geqslant 0}$ de variables aléatoires discrètes mais indépendantes, qui nécessitera *ipso facto* l'introduction d'un univers non dénombrable.

La fonction de répartition d'une variable aléatoire mixte peut s'écrire comme la somme

$$F = F_d + F_c$$

d'une fonction étagée (« discrète ») F_d et d'une fonction continue F_c. Pour obtenir cette décomposition, on note $\{a_k\ ;\ k \in K\}$ l'ensemble des points de discontinuité de F. Cet ensemble est au plus dénombrable[2]. S'il est vide, on pose $F_d = 0$ et $F_c = F$. Sinon, on note s_k le saut de discontinuité en a_k :

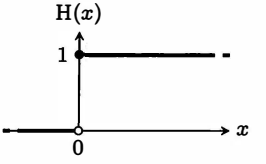
$$s_k = F(a_k) - F(a_k^-) > 0$$

et $s = \sum_k s_k$ la somme de tous les sauts : $0 < s \leqslant 1$. Enfin, on pose

$$F_d(x) = \sum_{k\ ;\ a_k \leqslant x} s_k = \sum_k s_k \mathbf{1}_{]-\infty\ ;\ x]}(a_k) = \sum_k s_k\, H(x - a_k),$$

où H est la fonction de Heaviside

$$H(x) = \mathbf{1}_{[0\ ;\ +\infty[}(x) = \begin{cases} 1 & \text{si } x \geqslant 0, \\ 0 & \text{si } x < 0. \end{cases}$$



Par construction, la fonction $F_c := F - F_d$ ne possède aucun saut de discontinuité ; comme elle est croissante, (exercice 13.2) elle est continue. Si elle a le bon goût d'être de la forme $F_c(x) = \int_{-\infty}^{x} f$, alors elle est absolument continue et F est un mélange de fonction discrète et de fonction à densité.

Remarque 13.15 Bien entendu, cela ne veut absolument pas dire que l'on puisse décomposer la variable aléatoire X elle-même en $X = X_d + X_{ac}$, où X_d serait une variable aléatoire discrète et X_{ac} une variable aléatoire à densité.

Décomposition de Lebesgue d'une fonction de répartition § 199

Dans la construction précédente, la fonction $F_c := F - F_d$ est continue, mais elle n'est pas *a priori* absolument continue. Étant croissante, elle est presque partout dérivable, mais l'intégrale de sa dérivée F_c' ne redonne pas forcément la fonction initiale[3] ; dans ce cas, il n'existe pas de fonction positive f telle que

$$F_c(x) = \int_{-\infty}^{x} f(t)\,dt,$$

puisqu'une telle fonction serait nécessairement égale presque partout à F_c'. On extrait alors de la fonction F_c sa partie absolument continue, en intégrant F_c' :

$$F_{ac}(x) = \int_{-\infty}^{x} F_c'(t)\,dt,$$

de telle sorte que F_{ac} est presque partout dérivable et $F_{ac}' = F_c'$. La partie

$$F_s := F_c - F_{ac}$$

est encore croissante d'après le théorème A.21 page 677, qui montre l'inégalité $\int_{-\infty}^{x} F_c' \leqslant F_c(x)$. On l'appelle **partie singulière** de F_c (une fonction est dite **singulière** si elle est continue, de dérivée presque partout nulle).

(2). Voir A25, page 677

(3). Des résultats utiles sur les fonctions croissantes sont résumés dans l'appendice A, paragraphe A26, page 677. Ces résultats, très classiques, sont démontrés dans de nombreux ouvrages d'intégration. Nous conseillons notamment au lecteur l'excellent ouvrage de Billingsley [9].

■ **Théorème 13.16 (Théorème de décomposition de Lebesgue)**
Soit F *une fonction de répartition. Alors il existe trois fonctions croissantes* F_d *(étagée),* F_{ac} *(absolument continue) et* F_s *(singulière) telles que* $F = F_d + F_{ac} + F_s$. *Cette décomposition est unique.*

Remarque 13.17 (À quoi ressemble une fonction singulière non constante ? (∗)*)*
L'exemple le plus connu de fonction singulière est l'escalier de Cantor (voir A28), qui peut paraître quelque peu artificiel dans un cadre probabiliste[4]. Un exemple plus naturel est le suivant. On considère une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, de même loi de Bernoulli $\mathscr{B}(p)$, où p est un réel, $0 < p < 1$. On pose $p_0 := 1 - p = \mathbf{P}\{X_n = 0\}$ et $p_1 := p = \mathbf{P}\{X_n = 1\}$, et on définit la variable aléatoire

$$X := \sum_{n=1}^{\infty} \frac{X_n}{2^n}$$

dont on note F la fonction de répartition. On peut vérifier (exercice 13.4) que, si $p \neq 1/2$, alors F est une fonction continue, strictement croissante, mais que sa dérivée est presque partout nulle.
⚠13.4 Chaque valeur $X(\omega)$ est un réel $[0\,;1]$ dont un développement dyadique est donné par la suite $(X_1, X_2, \dots)$. Or, la loi de $(X_2, X_3, \dots)$ est la même que la loi de $(X_1, X_2, \dots)$, et la probabilité que X_1 soit égal à 1 est p ; on en déduit que, pour tout $x \leqslant 1/2$ (correspondant à une valeur de X pour laquelle on a nécessairement $X_1 = 0$), on a

$$\begin{aligned} F(x) &= \mathbf{P}\left\{\sum_{n=1}^{\infty} 2^{-n} X_n \leqslant x\right\} = \mathbf{P}\left\{X_1 = 0, \sum_{n=2}^{\infty} 2^{-n} X_n \leqslant x\right\} \\ &= \mathbf{P}\{X_1 = 0\}\, \mathbf{P}\left\{\sum_{n=1}^{\infty} 2^{-n} X_n \leqslant 2x\right\} = (1-p)\, F(2x). \end{aligned}$$

De même, on a, si $x > 1/2$,

$$F(x) = (1-p) + p\, F(2x-1).$$

Ces deux relations entraînent l'autosimilarité de la structure du graphe de F : la partie contenue entre 0 et 1/2 est semblable, à une homothétie de rapport 1/2 et une affinité orthogonale près, au graphe complet. De même pour la partie entre 1/2 et 1.

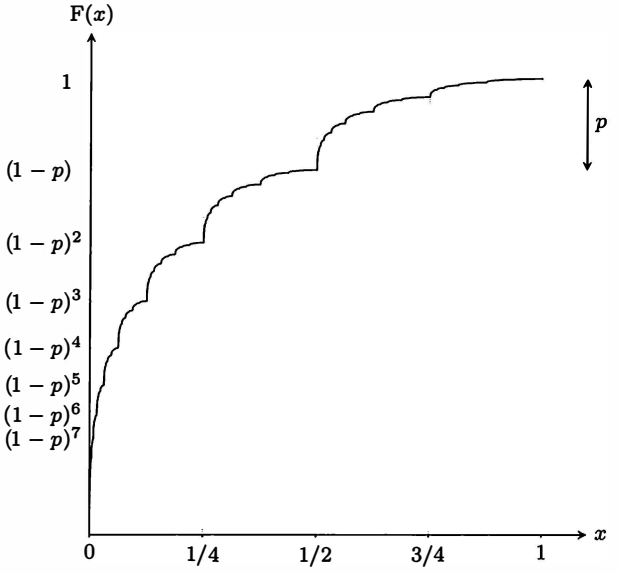


FIGURE 13.2 — *La fonction de répartition singulière* F *de la variable aléatoire* $X = \sum_{n=1}^{\infty} X_n/2^n$, *lorsque les* X_n *sont indépendantes et suivent la loi de Bernoulli* $\mathscr{B}(p)$, *avec ici* $p = 0,2$.

(4). L'ensemble de Cantor, ou un de ses cousins, se retrouve malgré tout dans la Nature, même s'il faut chercher un peu, par exemple dans l'effet Hall quantique.

Remarque 13.18 Hâtons-nous d'ajouter que, dans la majorité des cas « pratiques », c'est-à-dire issus de la modélisation d'une situation et non d'une problématique théorique, la fonction de répartition d'une variable aléatoire se décompose uniquement en ses deux parties discrète et absolument continue (ce qui justifie leur étude préalablement à ce chapitre). Sauf exception, on pourra toujours imaginer que $F = F_d + F_{ac}$.

Densité au sens des distributions

§ 200

Il est tentant de « dériver » une fonction de répartition de la forme $F = F_d + F_{ac}$. La partie absolument continue est presque partout dérivable, et l'on sait dériver une fonction discrète au sens des distributions : si

$$F_d(x) = \sum_{k \in K} s_k \, H(x - x_k),$$

(l'ensemble K étant au plus dénombrable), sa dérivée au sens des distributions est

$$F'_d = \sum_{k \in K} s_k \, \delta_{x_k}$$

où δ_{x_k} est la distribution définie par $\langle \delta_{x_k}, \varphi \rangle = \varphi(x_k)$ pour toute fonction-test φ. Cette procédure a de nombreux avantages : intuitivement, une densité se modélise très bien par un pic de Dirac, et les physiciens apprécient ce formalisme. Par ailleurs, les calculs d'espérance, de moments,... faisant intervenir la densité restent formellement corrects : en fait, tous calculs de la forme

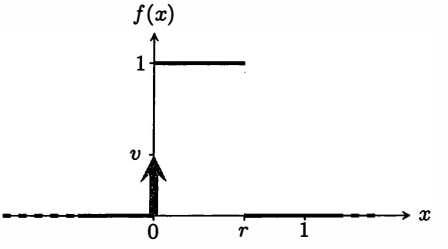
$$\int_{-\infty}^{+\infty} \varphi(x) \, f(x) \, dx$$

où φ est une fonction continue en tous les x_k. Enfin, il est aisé de retrouver la fonction F par « intégration[5] » de f, ce qui fait que toute l'information sur la loi de X est effectivement contenue dans sa densité au sens des distributions.

Exemple 13.19 (Temps d'attente au feu tricolore) Reprenons l'exemple 13.13. La densité de X au sens des distributions est

$$f(x) = v \, \delta(x) + \mathbf{1}_{[0\,;\,r]}(x).$$

On peut la représenter ainsi :



Ce point de vue « distributions » peut être remplacé par la formalisme des intégrales de Stieltjes, plus courant dans les ouvrages de probabilités ; nous n'évoquerons plus le point de vue des distributions que dans un but d'illustration.

---

(5). Plus précisément, dériver une fonction au sens des distributions, puis l'intégrer à partir de $-\infty$, mène à une fonction G presque partout égale à F. On pourrait objecter ceci : obtenir des fonctions de répartition *presque partout égales* n'est absolument pas suffisant : si $G(x) - G(x^-)$ n'est pas égal à $F(x) - F(x^-)$, on a perdu l'information « probabilité de $\{X = x\}$ ». En réalité, si l'on impose à G d'être continue à droite, alors elle ne peut être égale qu'à F, et l'ambiguïté est levée. On le démontre simplement ainsi : si x est un réel, on peut trouver une suite décroissante $(x_n)_{n \geqslant 0}$ de limite x, telle que $F(x_n) = G(x_n)$ pour tout entier n, et on passe à la limite par continuité à droite pour obtenir $F(x) = G(x)$.

§ 201 **Une définition de l'espérance n'utilisant que la fonction de répartition (*)**

Si X est une variable absolument continue admettant une espérance, une intégration par parties permet d'exprimer l'espérance en n'utilisant que la fonction de répartition. Toutefois, un calcul brutal n'aboutit pas, et il convient de séparer l'intégrale en deux au préalable :

$$\begin{aligned}\int_A^B x\,f(x)\,dx &= \int_A^0 x\,f(x)\,dx + \int_0^B x\,f(x)\,dx \\ &= -\int_A^0 F(x)\,dx + \Big[x\,F(x)\Big]_A^0 + \int_0^B \big(1-F(x)\big)\,dx - \Big[x\,\big(1-F(x)\big)\Big]_0^B. \qquad (13.1)\end{aligned}$$

Pour les termes de bord, remarquons que, si $A < 0 < B$,

$$0 \leqslant B\,\big(1-F(B)\big) = B\int_B^{+\infty} f(x)\,dx \leqslant \int_B^{+\infty} x\,f(x)\,dx$$

et cette quantité tend vers 0 puisque $x \mapsto x\,f(x)$ est intégrable. De même, le terme $A\,F(A)$ tend vers 0 lorsque A tend vers $-\infty$. En passant à la limite dans la relation (13.1), on obtient :

$$\mathbf{E}(X) = -\int_{-\infty}^0 F(x)\,dx + \int_0^{+\infty} \big(1-F(x)\big)\,dx, \qquad (13.2)$$

expression qui ne fait plus appel à la densité.

Réciproquement, supposons que les intégrales précédentes existent. Alors F intégrable sur $\mathbb{R}^-$, et de même $1-F$ est intégrable sur $\mathbb{R}^+$. On en déduit que $x\,F(x)$ tend vers 0 en $-\infty$ et que et $x\big(1-F(x)\big)$ tend vers 0 en $+\infty$[6]. La relation (13.1) permet alors de conclure que X admet une espérance et que celle-ci est donnée par (13.2).

En résumé, on a montré le résultat suivant :

■ **Proposition 13.21** *Soit* X *une variable aléatoire absolument continue, de fonction de répartition* F *; alors* X *admet une espérance si et seulement si* F *est intégrable sur* $]-\infty\,;\,0]$ *et* $1-F$ *est intégrable sur* $[\,0\,;+\infty\,[$. *Si c'est le cas, alors*

$$\mathbf{E}(X) = -\int_{-\infty}^0 F(x)\,dx + \int_0^{+\infty} \big(1-F(x)\big)\,dx,$$

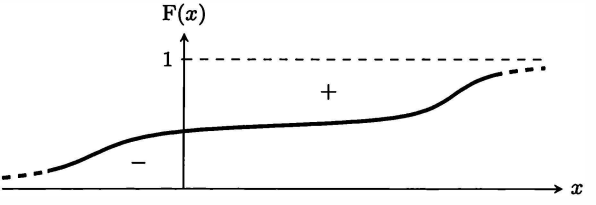


FIGURE 13.3 — *En gris, l'espérance d'une variable aléatoire comme intégrale de sa fonction de répartition. (Le domaine de gauche compte négativement.) Comparer à la figure 14.2 page 321.*

Si X est une variable aléatoire ne prenant qu'un nombre fini de valeurs, la fonction F est constante par morceaux, et le lecteur vérifiera sans peine que la formule (13.2) reste encore valable. De même si X est une variable aléatoire discrète, ou mixte.

---

(6). Cela découle du lemme suivant :

■ **Lemme 13.20** *Si* g *est une fonction décroissante et intégrable sur* $\mathbb{R}^+$, *alors* $\lim_{x\to+\infty} x\,g(x) = 0$.

La démonstration est élémentaire : tout d'abord, g admet une limite nulle en $+\infty$. Par décroissance, $0 \leqslant x\,g(x) \leqslant 2\int_{x/2}^x g(t)\,dt$, et cette dernière quantité tend vers 0 par intégrabilité de g.

Nous pouvons alors utiliser cette relation comme *définition* de l'espérance d'une variable aléatoire quelconque, la condition d'existence étant que les intégrales convergent. Cette définition unifie les précédentes et les généralise[7].

Définition 13.22 (Définition alternative de l'espérance) Soit X une variable aléatoire réelle, de fonction de répartition F. On dit que X **admet une espérance** si F est intégrable sur $\mathbb{R}^-$ et si $1 - F$ l'est sur $\mathbb{R}^+$; dans ce cas, l'**espérance de X** est ◁13.3

$$\mathbf{E}(X) := -\int_{-\infty}^{0} F(x)\,dx + \int_{0}^{+\infty} (1 - F(x))\,dx.$$

Remarque 13.23 Cette approche, si elle permet d'unifier les points de vue discret et continu, reste trop limitée du point de vue théorique : elle ne permet pas, par exemple, de montrer la linéarité de l'espérance. De même, les calculs des moments, la formule de transfert, etc., se déduisent mal d'un tel formalisme. Il est bien plus rentable d'intégrer, non sur les valeurs prises par X (ensemble d'arrivée), mais sur l'ensemble de départ Ω lui-même ; c'est le but du chapitre 14.

L'espérance et les notations de l'intégrale de Stieltjes (∗) § 202

Une autre façon d'unifier les points de vue discret et absolument continu est d'utiliser le formalisme de l'intégrale de Stieltjes.

Dans le cas d'une variable discrète, notons $h(x) = \mathbf{P}\{X = x\}$ pour chacune des valeurs x prises par X (elles sont dénombrables) ; l'espérance de X est

$$\mathbf{E}(X) = \sum_x x\,h(x).$$

Si au contraire X est absolument continue et admet une densité f, on a

$$\mathbf{E}(X) = \int_{-\infty}^{+\infty} x\,f(x)\,dx.$$

On réunit ces deux cas en notant

$$\mathbf{E}(X) = \int_{-\infty}^{+\infty} x\,dF(x),$$

où $dF(x) = f(x)\,dx$ dans le cas absolument continu, et $dF(x) = F(x) - F(x^-) = h(x)$ dans le cas discret.

Une construction plus précise de l'intégrale précédente, appelée **intégrale de Stieltjes**[8] est donnée dans le chapitre suivant ; le lecteur peut se contenter de savoir que, si F est la somme d'une fonction absolument continue et de fonctions de Heaviside

$$F(x) = G(x) + \sum_k \alpha_k\,H(x - a_k),$$

alors l'intégrale de φ selon la mesure de Stieltjes dF vaut

$$\int \varphi\,dF = \int_{-\infty}^{+\infty} \varphi(x)\,dF(x) = \int_{-\infty}^{+\infty} \varphi(x)\,G'(x)\,dx + \sum_k \alpha_k\,\varphi(a_k).$$

Pour des variables aléatoires quelconques, on écrit

$$\mathbf{E}(X) = \int_{-\infty}^{+\infty} x\,dF(x) \qquad \mathbf{E}(X^k) = \int_{-\infty}^{+\infty} x^k\,dF(x)$$

---

(7). La seule contrainte technique étant de donner un sens à l'intégrale ; là encore, le lecteur choisira le degré d'abstraction qu'il préfère : on demandera que F soit continue par morceaux pour une intégrale de Riemann, mesurable pour une intégrale de Lebesgue, etc. Dans le premier cas, les intégrales en jeu sont bien définies si l'ensemble des valeurs prises par X n'admet pas de point d'accumulation — par exemple, si X est à valeurs entières.

(8). On dit également, pour des raisons historiques : intégrale de Riemann-Stieltjes, ou de Lebesgue-Stieltjes, selon la classe de fonctions que l'on intègre, le premier cas permettant de passer à la limite dans des généralisations de sommes de Riemann, le second cas requérant un peu de théorie de la mesure. Nous renvoyons le lecteur à Burk [15] pour un exposé historique et mathématique des diverses intégrales ayant été construites au cours des siècles.

et, pour des variables indépendantes, la formule de convolution

$$\mathrm{F}_{\mathrm{X}+\mathrm{Y}}(z) = \int_{-\infty}^{+\infty} \mathrm{F}_{\mathrm{X}}(z-y)\,\mathrm{dF}_{\mathrm{Y}}(y)$$

Certains ouvrages font un grand usage des notations précédentes, notamment celui de William Feller ; rien que pour le plaisir de lire ce joyau, il vaut la peine de se familiariser avec elles. Nous n'insisterons pas plus sur la construction historique de l'intégrale de Stieltjes ; dans le chapitre suivant, nous présentons la construction de l'intégrale selon la mesure de probabilité image, notée $\mathbf{P}_{\mathrm{X}}(\mathrm{d}x)$. L'intégrale de Stieltjes est exactement l'intégrale selon cette mesure :

$$\int_{-\infty}^{+\infty} \varphi(x)\,\mathrm{dF}(x) = \int_{-\infty}^{+\infty} \varphi(x)\,\mathbf{P}_{\mathrm{X}}(\mathrm{d}x).$$

13.3 LIMITE SIMPLE DE VARIABLES ALÉATOIRES

Il est fréquent de pouvoir définir une notion, ou démontrer aisément une propriété, pour une variable aléatoire *simple*, c'est-à-dire ne prenant qu'un nombre fini de valeurs : c'est le cas le plus simple où n'interviennent que des sommes en nombre fini. Des techniques permettent alors de transférer ces notions ou ces propriétés à des classes plus larges de variables aléatoires : pour cela, on montre d'abord que l'on peut approcher une variable aléatoire X quelconque par une suite $(\mathrm{X}_n)_{n\geqslant 0}$ de variables aléatoires simples. Pour des raisons techniques, nous restreignons ce résultat au cas des variables *positives* ; en contrepartie, nous obtiendrons la *croissance* de la suite des approximations simples.

Définition 13.24 Une variable aléatoire $\mathrm{X} : \Omega \to \mathbb{R}$ est dite **simple** (ou **étagée**[9]) s'il existe un entier $n \geqslant 0$, une partition $\Omega = \mathrm{A}_1 + \mathrm{A}_2 + \cdots + \mathrm{A}_n$ et des réels $x_1, x_2, \ldots, x_n$ tels que

$$\mathrm{X} = \sum_{k=1}^{n} x_k\,\mathbf{1}_{\mathrm{A}_k}.$$

■ **Théorème 13.25** *Soit* X *est une variable aléatoire* positive *; alors il existe une suite croissante* $(\mathrm{X}_n)_{n\geqslant 0}$ *de variables aléatoires simples, qui converge simplement vers* X.

Rappelons que, par « suite croissante qui converge simplement vers X », nous entendons une suite $(\mathrm{X}_n)_{n\geqslant 0}$ de variables aléatoires telle que, pour tout $\omega \in \Omega$, la suite de terme général $\mathrm{X}_n(\omega)$ est croissante et converge vers $\mathrm{X}(\omega)$.

DÉMONSTRATION : Une telle suite est facile à construire. L'idée est de subdiviser $\mathbb{R}^+$ en intervalles égaux de largeur ε_n « petite », par exemple $\varepsilon_n = 1/2^n$. On approche ensuite $\mathrm{X}(\omega)$ par la valeur $k\,\varepsilon_n$ la plus proche inférieurement, c'est-à-dire par $\lfloor 2^n\,\mathrm{X}_n(\omega)\rfloor/2^n$. Si l'on procède ainsi, cependant, on risque de se trouver en face d'une infinité de valeurs distinctes, pour peu que X ne soit pas bornée. On commence donc par tronquer X, en posant

$$\mathrm{Y}_n := \mathrm{X} \wedge n = \min(\mathrm{X}, n).$$

On pose ensuite

$$\mathrm{X}_n = \frac{\lfloor 2^n\,\mathrm{Y}_n\rfloor}{2^n}.$$

(9). Ou encore *finie*, mais c'est une moins bonne idée ; on désigne plutôt par « variable aléatoire finie » une variable ne prenant que des valeurs finies, et jamais $\pm\infty$.

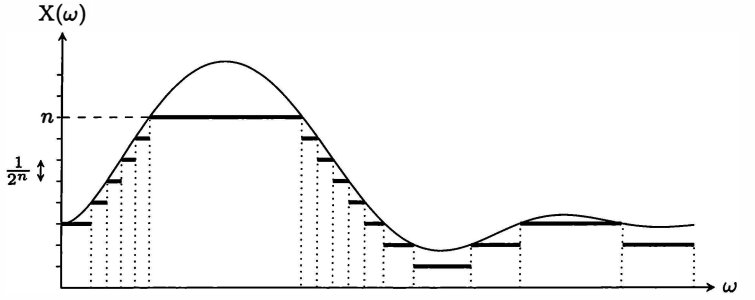


FIGURE 13.4 — *Construction de la variable* X_n.

Les applications $X_n : \Omega \to \mathbb{R}$ sont, par construction, étagées, et prennent leurs valeurs dans $\{k\,2^{-n}\ ;\ k = 0, \ldots, n\,2^n\}$. Il ne reste plus qu'à démontrer que chaque X_n est bien une variable aléatoire. Pour cela, il suffit de noter que, pour $k = 0, \ldots, n\,2^n - 1$, l'événement $\{X_n = k\,2^{-n}\} = \{k\,2^{-n} \leqslant X < (k+1)\,2^{-n}\}$ est bien dans $\mathfrak{T}$, et pour $k = n\,2^n$, l'événement $\{X_n = n\} = \{X \geqslant n\}$ est encore dans $\mathfrak{T}$.

Ainsi, la suite $(X_n)_{n \geqslant 0}$ est une bien une suite de variables aléatoires et, par construction, elle est croissante (voir figure ci-dessous) et de limite simple égale à X . ∎

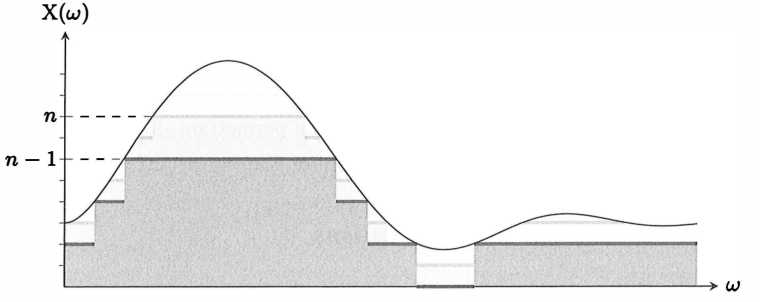


FIGURE 13.5 — *Croissance de la suite* $(X_k)_{k \geqslant 0}$. *En gris foncé, la variable* X_{n-1} ; *en gris clair, la variable* X_n. *En tout point* ω *de* Ω, *on a l'inégalité* $X_{n-1}(\omega) \leqslant X_n(\omega)$.

EXERCICES

♦ ***Exercice 13.1*** Voir en bas de la page 296.

♦ ***Exercice 13.2 (Croissance de la partie continue)*** Dans la décomposition $F = F_d + F_c$, montrer que la partie continue est croissante.

♦ ***Exercice 13.3 (Un critère d'intégrabilité)*** Soit X une variable aléatoire à valeurs positives, dont on note F la fonction de répartition. On rappelle que X admet une espérance si et seulement si $1 - F$ est intégrable sur $[0; +\infty[$. Montrer qu'il y a équivalence entre les énoncés

i) X admet une espérance ;

ii) pour tout $\alpha > 0$, la série $\sum [1 - F(n\alpha)] = \sum \mathbf{P}\{X > n\alpha\}$ converge.

Lorsque X admet une espérance, exprimer $\mathbf{E}(X)$ comme limite d'une expression faisant intervenir $\sum \mathbf{P}\{X > n\alpha\}$.

♦ ***Exercice 13.4 (Une fonction singulière)*** On reprend la fonction F de l'exemple 13.17. On note

$$p_1 := p = \mathbf{P}\{X_1 = 1\} \qquad \text{et} \qquad p_0 := 1 - p = \mathbf{P}\{X_1 = 0\}.$$

On suppose que $p \neq 1/2$ (sinon, on a $F(x) = x$ pour tout x de $[0; 1]$ et l'exercice perd singulièrement d'intérêt). Montrer que :

i) F est continue ;

ii) F est strictement croissante ;

iii) si $p < 1/2$, F n'est pas dérivable en 0 ; si $p > 1/2$, F n'est pas dérivable en 1.

iv) F n'est dérivable en aucun point dyadique ;

v) $F'(x) = 0$ en tout point x où F est dérivable.

Vérifier que l'on peut calculer explicitement la valeur de F en chaque point dyadique ; cela suffit-il pour connaître F sur $[0; 1]$? Déterminer un algorithme permettant de tracer le graphe de F à l'ordinateur pour différentes valeurs de p.

Indications

◊ *Indication 13.3 :* Utiliser la décroissance de $1 - F$ et effectuer une comparaison série/intégrale.

◊ *Indication 13.4 :*

i) Montrer que $F((k+1)\,2^{-n}) - F(k\,2^{-n})$ est un produit de la forme $p_{\varepsilon_1} p_{\varepsilon_2} \cdots p_{\varepsilon_n}$, où les ε_k sont éléments de $\{0, 1\}$. En déduire que $\mathbf{P}\{X = x\} = 0$ pour tout $x \in [0; 1]$.

ii) Si $0 \leqslant a < b \leqslant 1$, considérer un intervalle de la forme $[k\,2^{-n}; (k+1)\,2^{-n}]$ inclus dans $[a; b]$.

iii) Si $p > 1/2$, utiliser la relation $F(x) = p\,F(2x)$ pour tout $x \in [0; 1/2]$. Si $p < 1/2$, utiliser $F(x) = p + (1-p)\,F(2x-1)$ pour $x > 1/2$.

iv) Utiliser les mêmes relations puis démontrer la formule voulue par récurrence.

v) Inclure x dans une suite $I_n^{k_n}$ d'intervalles de la forme

$$I_n^{k_n} = [\alpha_n; \beta_n] := \left[\frac{k_n}{2^n}; \frac{k_n + 1}{2^n} \right[,$$

puis supposer que $\frac{F(\beta_n) - F(\alpha_n)}{\beta_n - \alpha_n} \to F'(x) \neq 0$. Remarquer que la quantité précédente s'écrit sous la forme $p_{\varepsilon_1} p_{\varepsilon_2} \cdots p_{\varepsilon_n}$ où $\varepsilon_k \in \{0, 1\}$, et conclure à une contradiction.

SOLUTIONS

♦ **Solution 13.3** Par décroissance de $1 - F$, on a pour tout entier n :

$$\alpha\big[1 - F((n+1)\alpha)\big] \leqslant \int_{n\alpha}^{(n+1)\alpha} [1 - F(t)]\, dt \leqslant \alpha\big[1 - F(n\alpha)\big]$$

ce qui prouve l'équivalence entre la convergence de la série et l'intégrabilité de la fonction positive $1 - F$. Dans le cas de la convergence, en sommant ces inégalités on obtient

$$\alpha \sum_{n=1}^{\infty} \big[1 - F(n\alpha)\big] \leqslant \mathbf{E}(X) \leqslant \alpha \sum_{n=0}^{\infty} \big[1 - F(n\alpha)\big]$$

donc, puisque $\alpha[1 - F(0)] \to 0$ quand $\alpha \to 0$,

$$\mathbf{E}(X) = \lim_{\alpha \to 0+} \alpha \sum_{n=0}^{\infty} \mathbf{P}\{X > n\alpha\}\,.$$

♦ **Solution 13.4**

i) Soient $n \geqslant 1$ un entier, et k un entier compris entre 0 et $2^n - 1$. L'ensemble

$$\{k\,2^{-n} \leqslant X < (k+1)2^{-n}\}$$

est égal, à un ensemble négligeable près, à un ensemble de la forme $\{X_1 = \varepsilon_1, X_2 = \varepsilon_2, \ldots, X_n = \varepsilon_n\}$, où les ε_k sont des éléments de $\{0, 1\}$, et plus précisément les coefficients du développement dyadique de $k\,2^{-n}$. Il est donc de probabilité $p_{\varepsilon_1} p_{\varepsilon_2} \cdots p_{\varepsilon_n}$, inférieure à α^n où $\alpha = \max(p_0, p_1)$.

Soit maintenant un réel x de $[0; 1]$. Pour tout entier n, on choisit k tel que $k\,2^{-n} \leqslant x < (k+1)2^{-n}$, ce qui montre que $\mathbf{P}\{X = x\} \leqslant \{k\,2^{-n} \leqslant X < (k+1)2^{-n}\} \leqslant \alpha^n$. Ainsi, $\mathbf{P}\{X = x\} = 0$. On en déduit que F est continue sur $[0; 1]$.

ii) Soient a et b deux réels tels que $a < b$. On choisit un entier n suffisamment grand pour que $2^{-n} \leqslant (b-a)/2$, ce qui permet de trouver ensuite un entier k tel que

$$\big[k\,2^{-n}\,;\,(k+1)\,2^{-n}\big] \subset [a\,;\,b].$$

On a alors

$$F(b) - F(a) \geqslant F((k+1)\,2^{-n}) - F(k\,2^{-n}) = p_{\varepsilon_1} p_{\varepsilon_2} \cdots p_{\varepsilon_n} > 0.$$

Ainsi, la fonction F est strictement croissante.

iii) Considérons le cas où $p_1 = p < 1/2$ et donc $p_0 > 1/2$. Pour tout $x \in [0; 1/2]$, on a établi que $F(x) = p_0\,F(2x)$; notamment, pour tout entier $n \geqslant 1$, on a

$$F(2^{-n}) = p_0\,F(2^{-n+1}) = p_0^2\,F(2^{-n+2}) = \cdots = p_0^{n-1}F(2^{-1}) = p_0^n\,F(1) = p_0^n.$$

Le taux d'accroissement

$$\frac{F(2^{-n}) - F(0)}{2^{-n}} = (2p_0)^n$$

tend vers $+\infty$ quand $n \to \infty$, ce qui prouve que F n'est pas dérivable en 0.

De manière symétrique, si $p > 1/2$, on utilise la relation $F(x) = 1 - p + p\,F(2x-1)$ qui se récrit plus simplement $G(y) = p\,G(2y)$ en posant $G(y) := 1 - F(1-y)$. On montre de la même manière que G n'est pas dérivable en 0, donc que F ne l'est pas en 1.

iv) Les relations précédentes permettent d'atteindre tous les points dyadiques.

v) Soit x un point en lequel F est dérivable, et supposons que $F'(x) \neq 0$. Pour tout $n \geqslant 1$, il existe un entier k_n tel que $x \in \big[k_n\,2^{-n}\,;\,(k_n+1)\,2^{-n}\big]$. On note $\alpha_n := k_n\,2^{-n}$ et $\beta_n = \alpha_n + 2^{-n}$. Alors $\alpha_n \uparrow x$ et $\beta_n \downarrow x$. Posons enfin $a_n := \big[F(\beta_n) - F(\alpha_n)\big]$. D'une part, a_n est de la forme

$$a_n = p_{\varepsilon_1} p_{\varepsilon_2} \cdots p_{\varepsilon_n}$$

comme déjà vu, et $a_{n+1} = p_{\varepsilon_1} p_{\varepsilon_2} \cdots p_{\varepsilon_n} p_{\varepsilon_{n+1}}$.

D'autre part, $a_n/2^{-n}$ est le taux d'accroissement entre α_n et β_n, donc $a_n/2^{-n} \to F'(x)$. Notamment, $p_{\varepsilon_{n+1}} = a_{n+1}/a_n \to 1/2$, ce qui est impossible puisque ni p_0 ni p_1 n'est égal à $1/2$.

Pour tracer le graphe de F, on utile la relation

$$\mathrm{F}(x) = \begin{cases} (1-p)\,\mathrm{F}(2x) & \text{si } x \leqslant 1/2, \\ 1-p+p\,\mathrm{F}(2x-1) & \text{si } x > 1/2 \end{cases}$$

et la valeur de F en 0 et 1, qui permettent de calculer F successivement :

— en 0, 1/2, 1 ;
— en 0, 1/4, 1/2, 3/4, 1 :
— puis, par récurrence, en tous les $k/2^n$, pour tout $k \in [0\,;\,2^n]$ et pour des valeurs croissantes de n.

Par continuité, la fonction F est entièrement déterminée par ses valeurs sur l'ensemble dense des dyadiques.

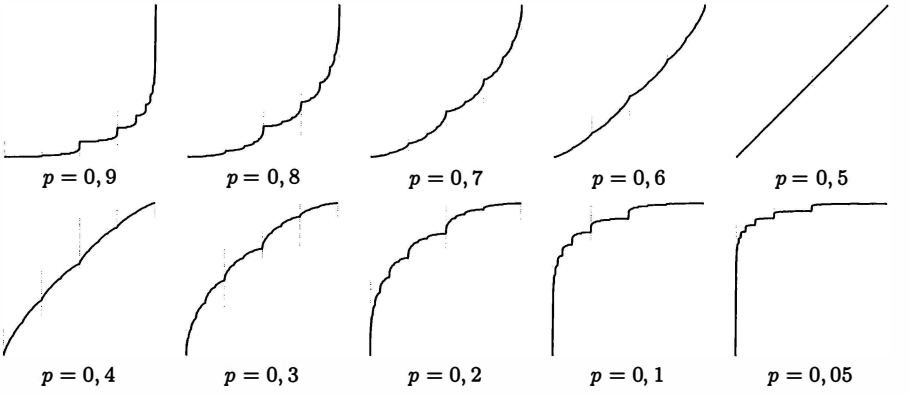


FIGURE 13.6 — *Les fonctions de répartition* F *pour différentes valeurs de* p.

Chapitre 14

L'espérance comme intégration sur une mesure de probabilité

Un des (nombreux) avantages de l'utilisation de la théorie abstraite de l'intégrale de Lebesgue en probabilités, initiée notamment par Fréchet et Kolmogorov, est de pouvoir écrire l'espérance d'une variable aléatoire $X : \Omega \to \mathbb{R}$, *non en prenant pour objet de base l'ensemble d'arrivée (la droite* $\mathbb{R}$*), comme dans les formules*

$$\mathbf{E}(X) = \sum_{n=0}^{\infty} x_n \, \mathbf{P}\{X = x_n\}, \qquad ou \qquad \mathbf{E}(X) = \int_{-\infty}^{+\infty} x \, f(x) \, dx$$

déjà vues pour des variables aléatoires respectivement discrètes et absolument continues, mais au contraire en partant de l'espace Ω *lui-même, selon une formule qui prendra l'allure suivante :*

$$\mathbf{E}(X) = \int_{\Omega} X(\omega) \, d\mathbf{P}(\omega) \qquad ou \qquad \mathbf{E}(X) = \int_{\Omega} X \, d\mathbf{P}.$$

(L'intégrale est calculée sur l'ensemble Ω *grâce à la mesure de probabilité, notée* $d\mathbf{P}(\omega)$*, selon une procédure standard qui sera brièvement présentée.) Ce changement radical de point de vue permet notamment de prouver des résultats « simples », comme la linéarité de l'espérance, qui sont quasiment indémontrables avec les formules précédentes*[1].

Dans tout ce chapitre, les variables aléatoires sont définies sur un même espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$.

(1). On peut certes prouver la linéarité de l'espérance dans le cas de variables discrètes, mais au prix de notations bien lourdes. Quant au cas continu, il a été admis.

14.1 CONSTRUCTION DE L'ESPÉRANCE COMME INTÉGRALE

La stratégie pour définir, de manière la plus générale possible, l'espérance d'une variable aléatoire, est semblable à une des méthodes permettant de définir l'intégrale d'une fonction de la variable réelle : on commence par le cas où X ne prend qu'un nombre fini de valeurs, puis on généralise, par un processus de limite, aux variables positives, aux variables réelles, et enfin aux variables complexes.

La construction qui va suivre permet de démontrer naturellement différentes propriétés fondamentales de l'espérance (positivité, linéarité), et offre en prime les grands théorèmes du calcul intégral : théorème de convergence monotone, théorème de convergence dominée.

§ 203 **Espérance d'une variable aléatoire simple**

Soit X une variable aléatoire réelle et **simple**, c'est-à-dire ne prenant qu'un nombre fini de valeurs ; on peut l'écrire

$$X = \sum_{k=1}^{n} \alpha_k \mathbf{1}_{A_k}$$

où $A_1, A_2, \ldots, A_n$ sont des éléments de la tribu $\mathfrak{T}$. On définit alors l'**espérance** de X comme le nombre, noté $\mathbf{E}(X)$ ou $\mathbf{E}X$:

$$\mathbf{E}(X) := \sum_{k=1}^{n} \alpha_k \mathbf{P}(A_k),$$

et que l'on écrit, selon les goûts et les ouvrages :

$$\mathbf{E}(X) = \int_\Omega X \, d\mathbf{P} = \int_\Omega X(\omega) \, d\mathbf{P}(\omega) = \int_\Omega X(\omega) \, \mathbf{P}(d\omega).$$

Remarque 14.1 Comme déjà démontré (théorème 7.4 page 145), la formule précédente ne dépend pas de la représentation choisie pour X : si X s'écrit des deux manières différentes suivantes :

$$X = \sum_{k=1}^{n} \alpha_k \mathbf{1}_{A_k} \qquad \text{et} \qquad X = \sum_{\ell=1}^{p} \beta_\ell \mathbf{1}_{B_\ell}$$

où $\{A_k \, ; \, k = 1, \ldots, n\}$ et $\{B_\ell \, ; \, \ell = 1, \ldots, p\}$ sont des partitions de Ω, alors $\sum_{k=1}^{n} \alpha_k \mathbf{P}(A_k) = \sum_{\ell=1}^{p} \beta_\ell \mathbf{P}(B_\ell)$.

La définition proposée est évidemment compatible avec la définition 7.3 page 144 pour une variable discrète et possède déjà toutes les propriétés fondamentales :

- **Linéarité** Si X et Y sont des variables aléatoires simples, alors $\alpha X + \beta Y$ est encore une variable aléatoire simple et $\mathbf{E}(\alpha X + \beta Y) = \alpha \mathbf{E}(X) + \beta \mathbf{E}(Y)$. La démonstration est évidente en choisissant une représentation de X et de Y utilisant la même partition : si $X = \sum \alpha_k \mathbf{1}_{A_k}$ et $Y = \sum \beta_\ell \mathbf{1}_{B_\ell}$, on partitionne Ω en les $\Delta_{k,\ell} := A_k \cap B_\ell$.
- **Croissance** Si X et Y sont des variables aléatoires simples telles que $X \leqslant Y$, alors $\mathbf{E}(X) \leqslant \mathbf{E}(Y)$. Notamment, l'espérance d'une variable aléatoire positive est toujours positive : si $X \geqslant 0$, alors $\mathbf{E}(X) \geqslant 0$. On le démontre de même en décomposant X et Y selon la même partition des $\Delta_{k,\ell}$.

Espérance d'une variable aléatoire positive § 204

Soit X une variable aléatoire réelle positive. On définit l'**espérance** de X comme la borne supérieure des $\mathbf{E}(Y)$, pour Y parcourant l'ensemble des variables aléatoires simples positives inférieures à X.

$$\mathbf{E}(X) := \sup_{\substack{0 \leqslant Y \leqslant X \\ Y \text{ simple}}} \mathbf{E}(Y).$$

Cette borne supérieure est toujours définie dans $\mathbb{R}^+ \cup \{+\infty\}$. Si $\mathbf{E}(X) < +\infty$, on dit que **X admet une espérance (finie)**, ou encore que **X est intégrable**.

On peut étendre cette définition aux variables à valeurs dans $[0; +\infty]$, en utilisant les règles de multiplication par $+\infty$:

$$\alpha \cdot (+\infty) = +\infty \quad \text{si } \alpha > 0 \qquad \text{et} \qquad 0 \cdot (+\infty) = 0.$$

On remarquera que, si $\mathbf{P}\{X = +\infty\} > 0$, alors $\mathbf{E}(X) = +\infty$:

■ **Proposition 14.2** *Si* $X \geqslant 0$ *et* $\mathbf{E}(X) < +\infty$, *alors* X *est presque sûrement finie.*

Le paragraphe suivant est un peu technique. Il montre que l'on peut exprimer $\mathbf{E}(X)$ non seulement comme une borne supérieure, mais aussi comme une *limite* de suite, ce qui est beaucoup plus utile en pratique.

Espérance comme limite croissante d'une suite de variables simples (*) § 205

Tout d'abord, on note que, si $(Y_n)_{n \geqslant 0}$ est une suite croissante de variables aléatoires positives convergeant simplement vers X (voir 13.3), alors $0 \leqslant Y_n \leqslant X$ pour tout $n \in \mathbb{N}$ et $\mathbf{E}(Y_n) \leqslant \mathbf{E}(X)$ par définition de l'espérance de X. Par croissance de l'espérance, la suite de terme général $\mathbf{E}(Y_n)$ est croissante ; étant majorée, elle converge, et sa limite est inférieure ou égale à $\mathbf{E}(X)$. On a donc montré le premier lemme :

■ **Lemme 14.3** *Si* $Y_n \uparrow X$, *la suite de terme général* $\mathbf{E}(Y_n)$ *converge et* $\lim_{n \to \infty} \mathbf{E}(Y_n) \leqslant \mathbf{E}(X)$.

Les deux lemmes suivants permettront de montrer qu'il y a en fait égalité.

■ **Lemme 14.4** *Soit* X *une variable aléatoire positive. Soit* $(Y_n)_{n \geqslant 0}$ *une suite croissante de variables aléatoires simples et positives, telle que* $Y_n \uparrow X$. *Pour toute variable aléatoire simple* Z *telle que* $Z \leqslant X$, *on a* $\lim_{n \to \infty} \mathbf{E}(Y_n) \geqslant \mathbf{E}(Z)$.

DÉMONSTRATION : Fixons momentanément un réel $\varepsilon > 0$. Posons, pour tout $n \in \mathbb{N}$,

$$A_n = \{Z - \varepsilon \leqslant Y_n\}.$$

La suite $(A_n)_{n \geqslant 0}$ est croissante et d'union Ω : en effet, pour tout $\omega \in \Omega$, il existe un rang N à partir duquel $Y_n(\omega) \geqslant Z(\omega) - \varepsilon$. Par ailleurs,

$$Y_n \geqslant (Z - \varepsilon)\, \mathbf{1}_{A_n}$$

(l'inégalité est vraie pour tout $\omega \in A_n$ par définition, et elle est encore vraie si ω est dans le complémentaire de A_n par positivité de Y_n). Évaluons les espérances des deux membres (ce sont des variables aléatoires simples à chaque fois) :

$$\mathbf{E}(Y_n) \geqslant \mathbf{E}(Z\, \mathbf{1}_{A_n}) - \varepsilon\, \mathbf{P}(A_n)$$

Afin de minorer $\mathbf{E}(Z\, \mathbf{1}_{A_n})$, on décompose $\mathbf{1}_{A_n} = 1 - \mathbf{1}_{A_n^c}$, et on note $\|Z\| = \max_{\omega \in \Omega} Z(\omega)$ (rappelons que Z ne prend qu'un nombre fini de valeurs) ; on obtient

$$\mathbf{E}(Y_n) \geqslant \mathbf{E}(Z) - \mathbf{E}(Z\, \mathbf{1}_{A_n^c}) - \varepsilon\, \mathbf{P}(A_n) \geqslant \mathbf{E}(Z) - \|Z\|\, \mathbf{P}(A_n^c) - \varepsilon.$$

Enfin, la suite de terme général A_n^c décroît et est d'intersection vide ($A_n^c \downarrow \varnothing$), donc $\mathbf{P}(A_n^c) \downarrow 0$. Pour n suffisamment grand, on a donc $\mathbf{E}(Y_n) \geqslant \mathbf{E}(Z) - 2\varepsilon$.

Ceci étant vrai pour tout $\varepsilon > 0$, la propriété est démontrée. ∎

■ **Lemme 14.5** *Si* $(X_n)_{n\geqslant 0}$ *et* $(Z_n)_{n\geqslant 0}$ *sont des suites de variables aléatoires simples telles que* $Y_n \uparrow X$ *et* $Z_n \uparrow X$, *alors*

$$\lim_{n\to\infty} \mathbf{E}(Y_n) = \lim_{n\to\infty} \mathbf{E}(Z_n).$$

DÉMONSTRATION : Pour tout entier k, on a l'inégalité $X \geqslant Z_k$; le lemme précédent permet de conclure que $\lim_{n\to\infty} \mathbf{E}(Y_n) \geqslant \mathbf{E}(Z_k)$. En passant à la limite dans le membre de droite, on obtient

$$\lim_{n\to\infty} \mathbf{E}(Y_n) \geqslant \lim_{k\to\infty} \mathbf{E}(Z_k).$$

Un raisonnement symétrique permet de conclure à l'égalité entre ces deux limites. ∎

Voici donc où l'on en est arrivé : toute suite de variables aléatoires simples, croissante et de limite X, admet une même espérance limite. Montrons que cette limite commune ℓ est alors nécessairement $\mathbf{E}(X)$.

Par l'absurde, supposons que $\ell \neq \mathbf{E}(X)$. D'après le lemme 14.3, on a en fait $\ell < \mathbf{E}(X)$; par définition d'une borne supérieure, il existe donc une variable aléatoire Y simple telle que $Y \leqslant X$ et $\ell < \mathbf{E}(Y)$. La variable aléatoire $X - Y$ étant positive, on l'approche par une suite croissante $(\Delta_n)_{n\geqslant 0}$ de variables aléatoires simples positives. La suite de terme général $Z_n = Y + \Delta_n$ est donc croissante, de limite X, et vérifie $Y \leqslant Z_n$; en prenant l'espérance et en passant à la limite, on a alors $\ell < \mathbf{E}(Y) \leqslant \ell$ et une contradiction.

Résumons le résultat obtenu :

■ **Théorème 14.6 (L'espérance comme limite)** *Soit* X *une variable aléatoire positive. Pour toute suite* $(Y_n)_{n\geqslant 0}$ *croissante de variables aléatoires simples positives telle que* $Y_n \uparrow X$,

$$\mathbf{E}(X) = \lim_{n\to\infty} \mathbf{E}(Y_n).$$

Ce résultat est valable que $\mathbf{E}(X)$ *soit finie ou infinie.*

La croissance et la linéarité de l'espérance des variables simples permet de déduire, en utilisant le théorème précédent, les propriétés suivantes pour des variables positives :

■ **Proposition 14.7** *Soient* X *et* Y *des variables positives.*

- ***Croissance*** *: Si* $0 \leqslant X \leqslant Y$, *alors* $0 \leqslant \mathbf{E}(X) \leqslant \mathbf{E}(Y)$.
  Ce résultat est encore valable dans $[0\,;+\infty]$.
- ***Domination*** *: Si* Y *est intégrable et si* $0 \leqslant X \leqslant Y$, *alors* X *est intégrable.*
- ***Additivité*** *:* $\mathbf{E}(X+Y) = \mathbf{E}(X) + \mathbf{E}(Y)$.
- *Si* $\alpha \geqslant 0$, *alors* $\mathbf{E}(\alpha X) = \alpha\,\mathbf{E}(X)$.

DÉMONSTRATION : *i)* Soient X et Y telles que $0 \leqslant X \leqslant Y$. On construit une suite $(X_n)_{n\geqslant 0}$ de variables aléatoires simples, positives, telle que $X_n \uparrow X$; notamment, $X_n \leqslant Y$ pour tout entier n, donc $\mathbf{E}(X_n) \leqslant \mathbf{E}(Y)$ par définition de l'espérance de Y. En passant à la limite, on obtient $\mathbf{E}(X) \leqslant \mathbf{E}(Y)$.

ii) Si $\mathbf{E}(Y) < +\infty$, il est en évidemment de même pour $\mathbf{E}(X)$ d'après le résultat précédent.

iii) On construit deux suites de variables simples, positives, telles que $X_n \uparrow X$ et $Y_n \uparrow Y$. Alors les variables $X_n + Y_n$ sont simples, positives et $X_n + Y_n \uparrow X + Y$. On en déduit le résultat.

iv) De même, $\alpha X_n \uparrow \alpha X$ et $\mathbf{E}(\alpha X_n) = \alpha\,\mathbf{E}(X_n)$.

∎

§ 206 **Espérance d'une variable aléatoire réelle**

Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle. On note (figure 14.1)

$$X^+ := \max(X, 0) \qquad \text{et} \qquad X^- := \max(-X, 0),$$

que l'on appelle respectivement **partie positive** et **partie négative** de X. On a alors

$$X = X^+ - X^- \qquad \text{et} \qquad |X| = X^+ + X^-.$$

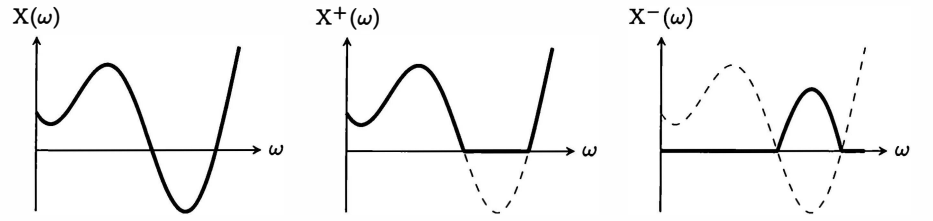


FIGURE 14.1 — *Parties positive et négative d'une variable aléatoire.*

Si $\mathbf{E}(X^+)$ et $\mathbf{E}(X^-)$ sont toutes deux finies, on dit que **X admet une espérance finie**, ou que **X est intégrable**, et on pose

$$\mathbf{E}(X) := \mathbf{E}(X^+) - \mathbf{E}(X^-).$$

On vérifie aisément que $0 \leqslant X^+ \leqslant |X|$ et $0 \leqslant X^- \leqslant |X|$. Ainsi, l'intégrabilité de $|X|$ implique celles de X^+ et de X^-. Réciproquement, l'intégrabilité de X^+ et de X^- implique que $|X| = X^+ + X^-$ est intégrable.

■ **Proposition 14.8** *Une variable aléatoire* X *est intégrable si et seulement si sa valeur absolue l'est, c'est-à-dire si* $\mathbf{E}|X| < +\infty$. *Dans ce cas,* $\mathbf{E}|X| = \mathbf{E}(X^+) + \mathbf{E}(X^-)$.

Pour cette raison, on note souvent « $\mathbf{E}|X| < +\infty$ » pour dire « X est intégrable ».

Remarque 14.9 (Espérance définie... mais infinie) On dit parfois que **l'espérance de X est définie** si l'une au moins des espérances de X^+ et de X^- est finie. Dans ce cas, on définit, dans $\overline{\mathbb{R}} = [-\infty\,;+\infty]$, l'**espérance** de X comme

$$\mathbf{E}(X) := \mathbf{E}(X^+) - \mathbf{E}(X^-).$$

— Si $\mathbf{E}(X^+) < +\infty$ et $\mathbf{E}(X^-) = +\infty$, alors $\mathbf{E}(X) = -\infty$.
— Si au contraire $\mathbf{E}(X^+) = +\infty$ et $\mathbf{E}(X^-) < +\infty$, alors $\mathbf{E}(X) = +\infty$.
— Le cas où $\mathbf{E}(X^+) = \mathbf{E}(X^-) = +\infty$ est le seul à être litigieux et l'espérance de X n'est pas définie.

Espérance d'une variable aléatoire complexe § 207

De la même façon que l'on a défini l'espérance d'une variable aléatoire réelle en la décomposant en ses parties positive et négative, on définit l'espérance d'une variable aléatoire complexe en la séparant en ses parties réelle et imaginaire. Dans une telle décomposition $X = Y + iZ$, les applications $Y = \mathscr{R}e(X)$ et $Z = \mathscr{I}m(X)$ sont des variables aléatoires réelles[(2)] ; si Y et Z admettent une espérance, on pose

$$\mathbf{E}(X) := \mathbf{E}(Y) + i\,\mathbf{E}(Z).$$

Les inégalités

$$|Y| \leqslant |X| \qquad |Z| \leqslant |X| \qquad \text{et} \qquad |X| \leqslant |Y| + |Z|$$

montrent que X est intégrable si et seulement si Y et Z le sont.

Propriétés de l'espérance ; espace $\mathcal{L}^1$ § 208

On définit l'espace $\mathcal{L}^1$ (ou $\mathcal{L}^1(\Omega, \mathfrak{T}, \mathbf{P})$ pour éviter toute ambiguïté) comme l'ensemble des variables aléatoires (réelles, en l'absence de toute autre précision) intégrables, c'est-à-dire vérifiant $\mathbf{E}|X| < +\infty$.

(2). Les problèmes de mesurabilité sont détaillés au paragraphe § 253, page 357.

■ **Théorème 14.10 (Linéarité de l'espérance)** *L'espace $\mathcal{L}^1$ des variables aléatoires intégrables est un espace vectoriel. De plus, l'espérance est une application linéaire sur l'espace $\mathcal{L}^1$:*

$$\mathbf{E}(X+Y) = \mathbf{E}(X) + \mathbf{E}(Y) \qquad \mathbf{E}(\alpha X) = \alpha\,\mathbf{E}(X)$$

pour toutes variables aléatoires $X, Y \in \mathcal{L}^1$ et tout $\alpha \in \mathbb{R}$.

DÉMONSTRATION : Soient X et Y des variables intégrables ; on sait déjà que X^+, X^-, Y^+ et Y^- sont positives intégrables. De plus, les inégalités

$$0 \leqslant (X+Y)^+ \leqslant X^+ + Y^+ \qquad \text{et} \qquad 0 \leqslant (X+Y)^- \leqslant X^- + Y^-$$

montrent que les variables aléatoires $(X+Y)^+$ et $(X+Y)^-$ sont intégrables, donc $X+Y$ l'est également. De même, $(\alpha X)^{\pm} \leqslant |\alpha|\,(X^+ + X^-)$, donc $(\alpha X)^+$ et $(\alpha X)^-$ sont toutes deux intégrables. Cela suffit pour prouver que $\mathcal{L}^1$ est un espace vectoriel.

Si α est un réel positif, on a plus précisément $(\alpha X)^+ = \alpha X^+$ et $(\alpha X)^- = \alpha X^-$, ce qui prouve que $\mathbf{E}(\alpha X) = \alpha\,\mathbf{E}(X)$; tandis que si α est négatif, $(\alpha X)^+ = -\alpha X^-$ et $(\alpha X)^- = -\alpha X^+$ et on obtient de nouveau $\mathbf{E}(\alpha X) = \alpha\,\mathbf{E}(X)$.

Pour l'additivité, soient X et Y deux variables aléatoires intégrables. On a, par définition,

$$(X+Y)^+ - (X+Y)^- = X + Y = X^+ - X^- + Y^+ - Y^-$$

ce que l'on réorganise en

$$(X+Y)^+ + X^- + Y^- = (X+Y)^- + X^+ + Y^+.$$

Dans chacun des membres de cette équation, on n'a que des variables aléatoires positives et intégrables ; on prend leurs espérances, puisque la propriété d'additivité est déjà établie pour les variables positives :

$$\mathbf{E}(X+Y)^+ + \mathbf{E}(X^-) + \mathbf{E}(Y^-) = \mathbf{E}(X+Y)^- + \mathbf{E}(X^+) + \mathbf{E}(Y^+).$$

Il ne reste plus qu'à réorganiser de nouveau :

$$\mathbf{E}(X+Y)^+ - \mathbf{E}(X+Y)^- = \mathbf{E}(X^+) - \mathbf{E}(X^-) + \mathbf{E}(Y^+) - \mathbf{E}(Y^-)$$

c'est-à-dire, par définition, $\mathbf{E}(X+Y) = \mathbf{E}(X) + \mathbf{E}(Y)$. ∎

■ **Théorème 14.11 (Autres propriétés)** *Soient X et Y deux variables aléatoires.*

i) ***Positivité.*** *Si $X \geqslant 0$, alors $\mathbf{E}(X) \geqslant 0$.*

ii) ***Croissance.*** *Si X et Y admettent une espérance finie* ou infinie, *et si $X \leqslant Y$, alors $\mathbf{E}(X) \leqslant \mathbf{E}(Y)$.*

iii) ***Inégalité triangulaire.*** *Si X est intégrable, alors $|\mathbf{E}X| \leqslant \mathbf{E}\,|X|$.*

iv) ***Domination.*** *Si Y est intégrable et si $|X| \leqslant |Y|$, alors X est intégrable.*

v) ***Variables bornées.*** *Si X est bornée, alors elle est intégrable.*

vi) Si $X = 0$ presque sûrement, alors $\mathbf{E}(X) = 0$.

vii) Si $X \geqslant 0$ et $\mathbf{E}(X) = 0$, alors $X = 0$ presque sûrement.

viii) Si $\mathbf{E}\,|X| = 0$, alors $X = 0$ presque sûrement.

ix) Si $X = Y$ p.s. et si X est intégrable, alors Y est intégrable et $\mathbf{E}(Y) = \mathbf{E}(X)$.

DÉMONSTRATION : *i)* Le résultat est immédiat par construction si X est étagée. Dans le cas général, on prend la borne supérieure sur un ensemble de nombres positifs.

ii) Résulte de la linéarité de l'espérance et du fait que $Y - X \geqslant 0$ implique $\mathbf{E}(Y-X) \geqslant 0$.

iii) On peut le démontrer directement :

$$|\mathbf{E}X| = \left|\mathbf{E}X^+ - \mathbf{E}X^-\right| \leqslant \left|\mathbf{E}X^+\right| + \left|\mathbf{E}X^-\right| = \mathbf{E}X^+ + \mathbf{E}X^- = \mathbf{E}\,|X|\,,$$

ou bien utiliser la propriété de croissance, en remarquant que $-|X| \leqslant X \leqslant |X|$ implique $-\mathbf{E}\,|X| \leqslant \mathbf{E}X \leqslant \mathbf{E}\,|X|$ et donc le résultat voulu.

iv) Si Y est intégrable, il en est de même de $|Y|$, donc de $|X|$, donc de X.

v) Une fonction constante est intégrable[3] ; si $|X| \leqslant M$, alors X est intégrable.

vi) Si X est une variable simple, écrivons-la $X = \sum_{k=1}^{n} \alpha_k \mathbf{1}_{A_k}$; pour chaque indice k, si $\alpha_k \neq 0$ alors $\mathbf{P}(A_k) = 0$, ce qui donne immédiatement que $\mathbf{E}(X) = 0$.
Si maintenant X est une variable positive, alors toute variable Y simple telle que $0 \leqslant Y \leqslant X$ est presque sûrement nulle, donc vérifie $\mathbf{E}(Y) = 0$; en passant à la borne supérieure, on obtient $\mathbf{E}(X) = 0$.
Enfin, dans le cas général, $|X|$ est presque sûrement nulle, donc $\mathbf{E}|X| = 0$, donc par inégalité triangulaire $\mathbf{E}(X) = 0$.

vii) Notons $A := \{X > 0\}$ et montrons que $\mathbf{P}(A) = 0$. On a $A = \lim_n \uparrow A_n$, où $A_n = \{X \geqslant 2^{-n}\}$; de l'inégalité

$$0 \leqslant 2^{-n} \mathbf{1}_{A_n} \leqslant X \mathbf{1}_{A_n} \leqslant X,$$

on tire

$$0 \leqslant 2^{-n} \mathbf{P}(A_n) \leqslant \mathbf{E}(X \mathbf{1}_{A_n}) \leqslant \mathbf{E}(X) = 0,$$

ce qui mène à $\mathbf{P}(A_n) = 0$ pour tout entier n. Par l'axiome 3′, on en déduit que $\mathbf{P}(A) = \lim \mathbf{P}(A_n) = 0$.

viii) Si $\mathbf{E}|X| = 0$, alors $|X| = 0$ presque sûrement et $X = 0$ presque sûrement.

ix) On a $X - Y = 0$ presque sûrement donc $X - Y$ est intégrable et $\mathbf{E}(X - Y) = 0$.

∎

Remarque 14.12 (Sur les variables non intégrables) La dernière propriété est encore vraie si l'on ne suppose plus X intégrable, mais seulement que X admet une espérance dans $[-\infty ; +\infty]$. Dans ce cas, Y admet également une espérance et $\mathbf{E}(X) = \mathbf{E}(Y)$.

Intégration sur une partie

§ 209

Soit $A \in \mathfrak{T}$ un événement. La variable aléatoire $\mathbf{1}_A$ est alors mesurable, donc $X \mathbf{1}_A$ l'est également. De plus, si X est intégrable, alors la relation $|X \mathbf{1}_A| \leqslant |X|$ prouve que $X \mathbf{1}_A$ est intégrable. On définit l'**intégrale de X sur A** par

$$\int_A X \, d\mathbf{P} := \int_\Omega X \cdot \mathbf{1}_A \, d\mathbf{P} = \mathbf{E}(X \mathbf{1}_A).$$

Espaces $\mathcal{L}^1$ et $\mathbf{L}^1$

§ 210

On peut munir l'ensemble des variables aléatoires sur Ω de la relation d'équivalence « égalité presque sûre » : $X \equiv Y$ si et seulement si $\mathbf{P}\{X = Y\} = 1$.

Si $X \equiv Y$, la propriété *ix*) du théorème 14.11 montre que $\mathbf{E}(X) = \mathbf{E}(Y)$: l'espérance est une fonction constante sur chacune des classes d'équivalence de la relation « $\equiv$ ». Notons $L^1(\Omega, \mathfrak{T}, \mathbf{P})$ — ou L^1, s'il n'y a pas d'ambiguïté — l'ensemble de ces classes. La structure d'espace vectoriel de $\mathcal{L}^1$ étant clairement compatible avec la relation d'équivalence, elle passe également au quotient, et l'espace L^1 est donc un espace vectoriel. De plus, l'application $\mathbf{E}$ induit sur cet espace une application, toujours linéaire, que l'on note également $\mathbf{E}$: si $\widetilde{X}$ est la classe d'une variable aléatoire X, $\mathbf{E}(\widetilde{X}) := \mathbf{E}(X)$.

À vrai dire, il est fréquent (et véniel) d'identifier une variable aléatoire avec sa classe d'équivalence, ou de parler d'une classe comme si c'était une variable aléatoire.

(3). Cette propriété fondamentale dérive bien sûr du fait qu'une probabilité est une mesure (positive) *finie*.

§211 **Moments d'ordre supérieur**

Soit X une variable aléatoire. Pour tout entier k, on dit que **X admet un moment d'ordre k** si X^k admet une espérance ; dans ce cas, le **moment d'ordre k de X** est

$$m_k(X) := \mathbf{E}(X^k) = \int_\Omega X^k \, d\mathbf{P}.$$

■ **Théorème 14.13** *Soit* X *une variable aléatoire réelle admettant un moment d'ordre* $k \geqslant 2$. *Alors* X *admet des moments d'ordre* $1, 2, \ldots, k$.

DÉMONSTRATION : Notons $A = \{|X| \leqslant 1\}$. Soit $\ell \in \{1, 2, \ldots, k\}$. La variable $X^\ell \mathbf{1}_A$ étant bornée, elle est intégrable. De plus,

$$|X|^\ell \, \mathbf{1}_{A^c} \leqslant |X|^k \, \mathbf{1}_{A^c} \leqslant |X|^k .$$

L'intégrabilité de $|X|^k$ entraîne celle de $|X|^\ell \, \mathbf{1}_{A^c}$. On en déduit que $X^\ell = X^\ell \mathbf{1}_A + X^\ell \mathbf{1}_{A^c}$ est intégrable. ∎

Cette propriété, conjointement à des inégalités de Hölder

$$\forall n \in [\![0\,;\, k]\!] \qquad |X|^n \cdot |Y|^{k-n} \leqslant \frac{|X|^k}{k/n} + \frac{|Y|^k}{(k-n)/n},$$

montre que si X et Y admettent chacune un moment d'ordre k, alors $X^n Y^{k-n}$ est inégrable, donc toute combinaison linéaire de X et Y admet un moment d'ordre k. On note $\mathcal{L}^k$ l'espace des variables aléatoires admettant un moment d'ordre k.

Le **moment centré d'ordre k** de X, lorsqu'il est défini, est

$$\mu_k(X) := \mathbf{E}\big[(X - \mathbf{E}X)^k\big] = \int_\Omega (X - \mathbf{E}X)^k \, d\mathbf{P}.$$

■ **Proposition 14.14** *Une variable aléatoire* X *admet un moment d'ordre k si et seulement si elle admet un moment centré d'ordre k.*

DÉMONSTRATION : La démonstration donnée dans le cas discret (page 153) est recyclable ici. ∎

Bien entendu, la **variance** d'une variable aléatoire de carré intégrable est son moment centré d'ordre 2 :

$$\mathbf{V}(X) := \mu_2(X) = \mathbf{E}\big((X - \mathbf{E}X)^2\big)$$

et son **écart-type** est la racine carrée de sa variance :

$$\sigma_X := \sqrt{\mathbf{V}(X)}.$$

La linéarité de l'intégrale nous permet d'écrire immédiatement la **formule de König-Huygens** :

$$\begin{aligned}\mathbf{V}(X) &= \mathbf{E}\big(X^2 - 2X\,\mathbf{E}X + (\mathbf{E}X)^2\big) = \mathbf{E}(X^2) - 2\,(\mathbf{E}X)^2 + (\mathbf{E}X)^2 \\ &= \mathbf{E}(X^2) - (\mathbf{E}X)^2.\end{aligned}$$

Comme dans le cas de variables discrètes, on a le résultat suivant :

■ **Théorème 14.15 (Cas d'une variance nulle)** *Soit* X *une variable aléatoire, que l'on suppose de carré intégrable. Alors il y a équivalence entre les énoncés*

i) $\mathbf{V}(X) = 0$;

ii) X *est presque sûrement constante.*

DÉMONSTRATION : Si X est presque sûrement constante, alors $X - \mathbf{E}X = 0$ presque sûrement et donc $\mathbf{V}(X) = 0$. Réciproquement, si $\mathbf{V}(X) = \mathbf{E}\,|X - \mathbf{E}X|^2 = 0$, alors $|X - \mathbf{E}X|^2$ est presque sûrement nulle (théorème 14.11), donc $X = \mathbf{E}(X)$ presque sûrement. ∎

14.2 THÉORÈMES DE CONVERGENCE

On donne ici les trois principaux théorèmes permettant de « permuter limite et intégrale », c'est-à-dire permettant de calculer l'espérance de la limite d'une suite de variables aléatoires.

Théorème de convergence monotone § 212

Le premier, et le plus important, est le théorème de convergence monotone, dont découlent tous les autres. Pour ne pas ralentir le rythme de cette section, sa démonstration est reportée en annexe de ce chapitre, page 329.

■ **Théorème 14.16 (Convergence monotone)** *Soit* $(X_n)_{n\geqslant 1}$ *une suite de variables aléatoires* positives *telle que* $0 \leqslant X_n \uparrow X$ *presque sûrement ; alors* $\mathbf{E}(X_n) \uparrow \mathbf{E}(X)$.

■ **Corollaire 14.17** *Soit* Y *une variable aléatoire telle que* $\mathbf{E}(Y) > -\infty$[4] *; soit* $(X_n)_{n\geqslant 1}$ *une suite de variables aléatoires telle que* $Y \leqslant X_n \uparrow X$ *presque sûrement ; alors* $\mathbf{E}(X_n) \uparrow \mathbf{E}(X)$.

DÉMONSTRATION : Poser $Z_n = X_n - Y$ et appliquer le théorème précédent. ∎

■ **Corollaire 14.18 (Séries positives)** *Si* $X_n \geqslant 0$ *pour tout* $n \in \mathbb{N}$, *alors*

$$\mathbf{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbf{E}(X_n),$$

cette égalité ayant lieu dans $[0\,;+\infty]$.

Lemme de Fatou (*) § 213

Le lemme de Fatou est un outil permettant, notamment, de montrer l'existence de $\lim \mathbf{E}(X_n)$. Il est une conséquence directe du théorème de convergence monotone.

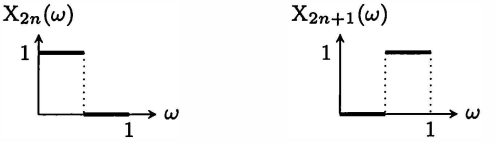
■ **Lemme 14.19 (Lemme de Fatou)** *Si* $X_n \geqslant 0$ *pour tout entier* n[5], *alors*

$$\mathbf{E}(\liminf X_n) \leqslant \liminf \mathbf{E}(X_n).$$

Si $\mathbf{E}|Y| < +\infty$ *et* $|X_n| \leqslant Y$ *pour tout* $n \in \mathbb{N}$, *alors*

$$\mathbf{E}(\liminf X_n) \leqslant \liminf \mathbf{E}(X_n) \leqslant \limsup \mathbf{E}(X_n) \leqslant \mathbf{E}(\limsup X_n).$$

Exemple 14.20 (... et moyen mnémotechnique) Pour ne pas se tromper dans le sens de l'inégalité (qui est plus petit, la lim inf de l'espérance ou l'espérance de la lim inf ?), le plus simple est de se souvenir de ce petit exemple, qui donne au passage un cas d'inégalité stricte. Sur l'espace $\Omega = [0\,;1]$ muni de la mesure de Lebesgue sur les boréliens, on définit les variables X_{2n} et X_{2n+1} par



Alors $\liminf X_n = 0$ tandis que $\mathbf{E}(X_n) = 1/2$ pour tout $n \in \mathbb{N}$, donc $\liminf \mathbf{E}(X_n) = 1/2$.

On utilise souvent la forme plus faible suivante :

---

(4). C'est-à-dire, on le rappelle, telle que $\mathbf{E}(Y^-) < +\infty$, l'espérance de Y^+ pouvant être finie ou infinie.

(5). Ou, plus généralement, si $\mathbf{E}(Y) > -\infty$ et si $X_n \geqslant Y$.

■ **Corollaire 14.21 (Fatou en pratique)** *Si* $X_n \geqslant 0$ *et si* $X_n \to X$ *presque sûrement, alors*

$$0 \leqslant \mathbf{E}(X) \leqslant \liminf \mathbf{E}(X_n).$$

DÉMONSTRATION DU LEMME 14.19 : Montrons la première inégalité, en supposant $X_n \geqslant 0$ pour tout entier n. Soit k un entier ; notons $Z_k = \inf_{n \geqslant k} X_n$, alors $X_n \geqslant Z_k$ et $\mathbf{E}(X_n) \geqslant \mathbf{E}(Z_k)$ pour tout $n \geqslant k$, donc

$$\mathbf{E}(Z_k) \leqslant \inf_{n \geqslant k} \mathbf{E}(X_n).$$

Or, $\liminf X_n = \lim_k \uparrow Z_k$; par application du théorème de convergence monotone,

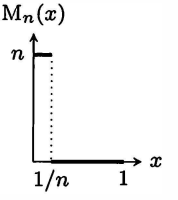
$$\mathbf{E}(\liminf X_n) = \lim_k \uparrow \mathbf{E}(Z_k) \leqslant \lim_k \uparrow \inf_{n \geqslant k} \mathbf{E}(X_n) = \liminf_n \mathbf{E}(X_n).$$

Dans le cas général, il suffit d'appliquer ce résultat à $Y - X_n$ pour obtenir les autres inégalités. ∎

Remarque 14.22 (Sur l'inégalité de Fatou et les processus de branchement)
Même lorsque $X_n \to X$ presque sûrement, le lemme de Fatou ne nous garantit qu'une *inégalité*. Dans l'exemple déjà rencontré du processus de Galton-Watson (page 195), notons $M_n = Z_n/m^n$; des théorèmes de martingales permettent de montrer que la suite $(M_n)_{n \geqslant 1}$ converge presque sûrement vers une variable aléatoire M. Plaçons-nous dans le cas critique $m = 1$. Alors on peut montrer que, presque sûrement, $M = 0$; notamment, son espérance est nulle : $\mathbf{E}(M) = 0$. Or, $\mathbf{E}(M_n) = 1$ pour tout entier n, ce qui fait que l'on a bien

$$0 = \mathbf{E}(M) = \mathbf{E}(\lim M_n) < \liminf \mathbf{E}(M_n).$$

Ce qui se passe ici c'est que, bien que M_n s'annule sur un ensemble de plus en plus grand lorsque n augmente, sur l'ensemble (réduit) où elle ne s'annule pas, elle est très grande (c'est obligatoire pour garantir que l'espérance de M_n vaut toujours 1). Un équivalent analytique simple de cet exemple est la suite des fonctions M_n d'allure ci-contre.
Si l'on pouvait dominer tous les M_n par une même variable aléatoire intégrable, ceci n'arriverait pas ; ce qui nous entraîne naturellement au dernier grand théorème, le célébrissime théorème de convergence dominée.



§ 214 Théorème de convergence dominée

Un résultat des plus populaires, le théorème de convergence dominée, permet de traiter des variables de signe quelconque.

■ **Théorème 14.23 (Convergence dominée)** *Soit* Y *une variable aléatoire positive intégrable. Soient* X *et* $X_1, X_2, \ldots, X_n, \ldots$ *des variables aléatoires ; on suppose que* $|X_n| \leqslant Y$ *pour tout entier* n, *et que* $X_n \to X$ *presque sûrement. Alors* X *est intégrable et* $(X_n)_{n \geqslant 1}$ *converge vers* X *dans* L^1, *c'est-à-dire* $\mathbf{E}|X_n - X| \to 0$. *Notamment,*

$$\lim_{n \to \infty} \mathbf{E}(X_n) = \mathbf{E}(X).$$

DÉMONSTRATION : On a $|X_n - X| \leqslant 2Y$ pour tout entier n, et $\mathbf{E}(2Y) < +\infty$; le lemme de Fatou donne

$$\limsup \mathbf{E}|X_n - X| \leqslant \mathbf{E}(\limsup |X_n - X|) = \mathbf{E}(0) = 0,$$

ce qui revient à dire que $\mathbf{E}|X_n - X| \to 0$. ∎

■ **Corollaire 14.24 (Intégration terme à terme)** *Si* $(X_n)_{n \geqslant 1}$ *est une suite de variables aléatoires réelles, et si la série positive* $\sum \mathbf{E}|X_n|$ *converge, alors la série* $\sum X_n$ *converge presque sûrement. De plus,*

$$\mathbf{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbf{E}(X_n).$$

DÉMONSTRATION : Notons $S := \sum_{n=1}^{\infty} |X_n|$, variable à valeurs dans $[0; +\infty]$. Le théorème de convergence monotone, et plus précisément son corollaire 14.18, montre que $\mathbf{E}(S) < +\infty$, et donc que S est presque sûrement finie (proposition 14.2). Notamment, la série $\sum X_n$ converge absolument presque sûrement, donc converge presque sûrement (une démonstration élémentaire de cette propriété est présentée dans l'exercice 19.6 page 412). De plus, pour tout $n \in \mathbb{N}$, on a ⇗19.6

$$\left| \sum_{k=1}^{n} X_k \right| \leqslant \sum_{k=1}^{n} |X_k| \leqslant S.$$

L'intégrabilité de S permet d'utiliser le théorème de convergence dominée et de conclure à la formule proposée. ∎

Le tableau ci-dessous résume ces résultats, l'hypothèse de convergence $X_n \to X$ étant toujours à prendre au sens « presque sûr ».

Théorèmes de convergence

Nom	Hypothèses	Résultat
Convergence monotone	$0 \leqslant X_n \uparrow X$	$\mathbf{E}(X_n) \uparrow \mathbf{E}(X)$
Séries positives	$X_n \geqslant 0$	$\mathbf{E}\Big(\sum_{n=1}^{\infty} X_n\Big) = \sum_{n=1}^{\infty} \mathbf{E}(X_n)$
Lemme de Fatou	$0 \leqslant X_n \to X$	$\mathbf{E}(X) \leqslant \liminf \mathbf{E}(X_n)$
Convergence dominée	$\lvert X_n\rvert \leqslant Y$, $\mathbf{E}(Y) < +\infty$, $X_n \to X$	$\mathbf{E}\lvert X_n - X\rvert \to 0$ et $\mathbf{E}(X_n) \to \mathbf{E}(X)$
Intég. terme à terme	$\sum \mathbf{E}\lvert X_n\rvert$ converge	$\mathbf{E}\Big(\sum_{n=1}^{\infty} X_n\Big) = \sum_{n=1}^{\infty} \mathbf{E}(X_n)$

La Cavalerie Légère à l'œuvre

§ 215

Le théorème de convergence monotone est souvent utilisé au sein d'une procédure standard permettant de prouver une certaine propriété *linéaire* sur un espace de variables.

- On montre que cette propriété est vraie pour les fonctions indicatrices ;
- par linéarité, on en déduit qu'elle est vraie pour toute variable simple ;
- si X est une variable *positive*, on l'approche (théorème 13.25) par une suite $(X_n)_{n\geqslant 1}$ de variables simples telle que $X_n \uparrow X$, et on applique le théorème de convergence monotone ;
- dans le cas général, on décompose X en $X^+ - X^-$ et on conclut.

It is a profoundly erroneous truism, repeated by all copy-books and by eminent people when they are making speeches, that we should cultivate the habit of thinking about what we are doing. The precise opposite is the case. Civilization advances by extending the numbers of important operations which we can perform without thinking about them. Operations of thought are like cavalry charges in battle – they are strictly limited in number, they require fresh horses, and must only be made at decisive moments.(6)

Alfred North WHITEHEAD

(6). Il existe un lieu commun totalement erroné, repris dans tous les manuels et dans les discours de gens éminents, selon lequel on devrait cultiver l'usage de réfléchir à ce que l'on fait. En réalité, c'est exactement l'opposé. La civilisation progresse en multipliant le nombre d'opérations fondamentales que l'on peut accomplir sans y penser. Les processus de pensée sont comme les charges de cavalerie dans une bataille : elles sont en nombre strictement limité, nécessitent des chevaux frais et ne doivent être lancées qu'à des moments décisifs.

14.3 LE THÉORÈME DE TRANSFERT ET SES APPLICATIONS

§ 216 **Le théorème général**

On a déjà vu que, pour calculer l'espérance d'une variable aléatoire, on avait deux moyens à notre disposition. Le premier est de construire une intégrale abstraite utilisant la mesure de probabilité $\mathbf{P}$ sur l'ensemble de départ Ω :

$$\mathbf{E}(\mathrm{X}) = \int_{\Omega} \mathrm{X}\,\mathrm{d}\mathbf{P}.$$

Le second est d'intégrer, sur l'ensemble d'arrivée $\mathbb{R}$, avec la mesure de probabilité image $\mathrm{d}\mathbf{P}_{\mathrm{X}}$ que l'on note parfois $\mathbf{P}_{\mathrm{X}}(\mathrm{d}x)$ ou $\mathrm{dF}(x)$, notamment pour préciser le nom de la variable réelle d'intégration :

$$\mathbf{E}(\mathrm{X}) = \int_{-\infty}^{+\infty} x\,\mathbf{P}_{\mathrm{X}}(\mathrm{d}x).$$

Le théorème suivant permet de généraliser ce double point de vue pour l'image $\varphi(\mathrm{X})$ d'une variable aléatoire par une application φ. On peut le voir comme une généralisation d'un théorème de changement de variable. Il traduit également le fait que l'espérance d'une variable aléatoire ne dépend que de sa loi (la répartition de ses valeurs) et pas de la manière dont telle valeur est affectée à tel ou tel ω particulier.

■ **Théorème 14.25 (Théorème de transfert)** *Soit $(\Omega, \mathcal{T}, \mathbf{P})$ un espace probabilisé. Soient* X *une variable aléatoire sur cet espace, de loi $\mathbf{P}_{\mathrm{X}}$, et $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction mesurable. Alors $\varphi(\mathrm{X}) = \varphi \circ \mathrm{X}$ est intégrable par rapport à $\mathbf{P}$ si et seulement si φ l'est par rapport à $\mathbf{P}_{\mathrm{X}}$; si c'est le cas, l'espérance de $\varphi(\mathrm{X})$ s'écrit sous les deux formes suivantes :*

$$\mathbf{E}\big[\varphi(\mathrm{X})\big] = \int_{\Omega} \varphi(\mathrm{X})\,\mathrm{d}\mathbf{P} = \int_{-\infty}^{+\infty} \varphi(x)\,\mathbf{P}_{\mathrm{X}}(\mathrm{d}x).$$

Ce théorème permet donc de transférer un procédé d'intégration (sur Ω) en un autre procédé d'intégration (sur $\mathbb{R}$).

Remarque 14.26 Dans le cas d'une variable aléatoire admettant une densité f, on montrera dans le paragraphe suivant que

$$\mathbf{P}_{\mathrm{X}}(\mathrm{d}x) = \mathrm{dF}(x) = f(x)\,\mathrm{d}x,$$

ce qui redonne la formule de transfert du théorème 11.4 page 246. De même, dans le cas d'une variable aléatoire discrète prenant des valeurs x_k ($k \in \mathrm{K}$) avec des probabilités respectives p_k, on a

$$\mathrm{d}\mathbf{P}_{\mathrm{X}} = \mathrm{dF} = \sum_{k \in \mathrm{K}} p_k\,\delta_{x_k}$$

où δ_{x_k} est la masse de Dirac en x_k, et la formule

$$\int_{\mathbb{R}} \varphi(x)\,\mathbf{P}_{\mathrm{X}}(\mathrm{d}x) = \sum_{k \in \mathrm{K}} \varphi(x_k)\,p_k$$

redonne bien la formule de transfert du théorème 7.17 page 152.

DÉMONSTRATION DU THÉORÈME DE TRANSFERT : On met en œuvre la méthode de la Cavalerie Légère du paragraphe § 215 et on démontre ce théorème en trois étapes.

• *1re étape : cas où φ est simple positive.* Supposons que φ est une fonction simple de la forme $\varphi = \sum_{k=1}^{n} \alpha_k \mathbf{1}_{A_k}$, où $\{A_k\ ;\ k = 1, \dots, n\}$ est une partition de $\mathbb{R}$; alors la variable aléatoire $\varphi(X)$ est également simple :

$$\varphi(X) = \sum_{k=1}^{n} \alpha_k \mathbf{1}_{A_k}(X) = \sum_{k=1}^{n} \alpha_k \mathbf{1}_{\{X \in A_k\}}.$$

Puisque, pour tout borélien A, on a

$$\int_\Omega \mathbf{1}_A(X)\, d\mathbf{P} = \mathbf{E}\big(\mathbf{1}_A(X)\big) = \mathbf{P}\{X \in A\} = \mathbf{P}_X(A) = \int_{-\infty}^{+\infty} \mathbf{1}_A(x)\, \mathbf{P}_X(dx),$$

la linéarité de l'espérance et de l'intégrale permettent d'écrire que

$$\mathbf{E}[\varphi(X)] = \sum_{k=1}^{n} \alpha_k \mathbf{E}\big(\mathbf{1}_{A_k}(X)\big) = \int_{-\infty}^{+\infty} \underbrace{\sum_{k=1}^{n} \alpha_k \mathbf{1}_{A_k}(x)}_{=\varphi(x)}\, \mathbf{P}_X(dx) = \int_{-\infty}^{+\infty} \varphi(x)\, \mathbf{P}_X(dx).$$

• *2e étape : cas où φ est positive.* On approche la fonction φ par une suite croissante $(\varphi_n)_{n \geqslant 1}$ de fonctions simples, positives, de limite simple égale à φ (grâce au théorème 13.25 — démontré pour des variables aléatoires mais évidemment valable pour des fonctions mesurables). Alors la suite de terme général $\varphi_n(X)$ est une suite croissante de variables simples, de limite égale à $\varphi(X)$, ce qui permet d'écrire que

$$\mathbf{E}[\varphi(X)] = \lim_{n \to \infty} \mathbf{E}[\varphi_n(X)].$$

Puisque chaque $\varphi_n(X)$ est simple, le résultat de la première étape donne

$$\begin{aligned} \mathbf{E}[\varphi(X)] &= \lim_{n \to \infty} \int_{-\infty}^{+\infty} \varphi_n(x)\, \mathbf{P}_X(dx) \\ &= \int_{-\infty}^{+\infty} \lim_{n \to \infty} \varphi_n(x)\, \mathbf{P}_X(dx) \quad \textit{par croissance monotone} \\ &= \int_{-\infty}^{+\infty} \varphi(x)\, \mathbf{P}_X(dx) \end{aligned}$$

• *3e étape : cas général.* Si φ est une fonction mesurable quelconque, on la décompose en ses parties positive et négative : $\varphi = \varphi^+ - \varphi^-$. Par linéarité de l'espérance, on obtient

$$\begin{aligned} \mathbf{E}[\varphi(X)] = \mathbf{E}[\varphi^+(X)] - \mathbf{E}[\varphi^-(X)] &= \int_{-\infty}^{+\infty} \varphi^+(x)\, \mathbf{P}_X(dx) - \int_{-\infty}^{+\infty} \varphi^-(x)\, \mathbf{P}_X(dx) \\ &= \int_{-\infty}^{+\infty} \varphi(x)\, \mathbf{P}_X(dx), \end{aligned}$$

ce qui achève la preuve du théorème de transfert. ∎

Cas des variables à densité § 217

Nous pouvons maintenant traiter plus rigoureusement la notion de variable à densité, et justifier le théorème de transfert tel que nous l'avons vu dans le chapitre 11.

Soit X une variable aléatoire admettant une densité f ; cela signifie que sa mesure image $d\mathbf{P}_X$ vérifie, pour tout réel a,

$$\mathbf{P}_X\big(]-\infty\,;\,a]\big) = \mathbf{P}\{X \leqslant a\} = \int_{-\infty}^{a} f(x)\, dx.$$

Cette propriété entraîne classiquement la formule

$$\mathbf{P}_X(A) = \mathbf{P}\{X \in A\} = \int_A f(x)\, dx$$

lorsque A est un intervalle, puis lorsque A est un ouvert (un ouvert étant union dénombrable d'ouverts disjoints) puis, par le théorème de classe monotone, lorsque

A est un borélien[7]. Si l'on compare avec l'expression de la même quantité obtenue par le théorème de transfert :

$$\mathbf{P}_X(A) = \mathbf{P}\{X \in A\} = \mathbf{E}\big(\mathbf{1}_A(X)\big) = \int_{-\infty}^{+\infty} \mathbf{1}_A(x)\, \mathbf{P}_X(dx) = \int_A \mathbf{P}_X(dx),$$

on voit que le passage d'une formule à l'autre se fait en écrivant formellement $\mathbf{P}_X(dx) = f(x)\,dx$, ou encore, puisque $\mathbf{P}(dx)$ se note parfois $dF(x)$:

$$dF(x) = f(x)\,dx.$$

Le théorème qui suit justifie l'emploi brutal de la formule précédente :

■ **Théorème 14.27 (« $dF(x) = f(x)\,dx$ »)** *Si* X *admet pour densité* f*, alors pour toute fonction mesurable positive* φ*,*

$$\int_{-\infty}^{+\infty} \varphi(x)\, \mathbf{P}_X(dx) = \int_{-\infty}^{+\infty} \varphi(x)\, f(x)\,dx,$$

cette égalité étant valable dans $\mathbb{R} \cup \{+\infty\}$.

Si φ *est une fonction mesurable et de signe quelconque, elle est intégrable par rapport à* $\mathbf{P}_X$ *si et seulement si* φf *est intégrable ; dans ce cas, on a encore*

$$\int_{-\infty}^{+\infty} \varphi(x)\, \mathbf{P}_X(dx) = \int_{-\infty}^{+\infty} \varphi(x)\, f(x)\,dx.$$

DÉMONSTRATION : De nouveau, faisons appel à la Cavalerie Légère. La formule est vraie pour toute fonction indicatrice $\varphi = \mathbf{1}_A$, où $A \in \mathcal{B}$ est un borélien, car

$$\int_{-\infty}^{+\infty} \mathbf{1}_A(x)\, \mathbf{P}_X(dx) = \mathbf{E}(\mathbf{1}_A(X)) = \mathbf{P}_X(A) = \int_A f(x)\,dx.$$

Par linéarité, elle est donc vraie pour toute fonction simple. Enfin, si φ est une fonction positive, on l'écrit comme limite croissante d'une suite $(\varphi_n)_{n\geqslant 0}$ de fonctions simples pour obtenir le résultat voulu.

Dans le cas d'une fonction de signe quelconque, il suffit d'appliquer la formule précédente à $|\varphi|$, ce qui montre la condition nécessaire et suffisante d'intégrabilité. Dans le cas d'intégrabilité, on applique la même formule séparément à φ^+ et φ^- pour conclure. ∎

■ **Corollaire 14.28 (Théorème de transfert, cas à densité)**
Soit X *une variable aléatoire absolument continue, de densité* f*. Alors, pour toute fonction* φ *mesurable,* $\varphi(X)$ *est intégrable si et seulement si* φf *est intégrable sur* $\mathbb{R}$*, auquel cas on a*

$$\mathbf{E}\big[\varphi(X)\big] = \int_{-\infty}^{+\infty} \varphi(x)\, f(x)\,dx.$$

Remarque 14.29 (Expression de l'espérance) Nous avions, à la fin du chapitre précédent (§ 201), signalé que l'espérance d'une variable aléatoire pouvait se définir comme

$$\mathbf{E}(X) = -\int_{-\infty}^{0} F(x)\,dx + \int_0^{+\infty} \big(1 - F(x)\big)\,dx,$$

ce qui s'interprète graphiquement comme ci-dessous sur la figure de gauche (les deux domaines grisés contribuent respectivement en négatif et en positif).

(7). Voir A74 et le théorème B.22 page 706.

Avec la formule de transfert, elle s'écrit

$$\mathbf{E}(X) = \int_{-\infty}^{+\infty} x \, dF(x),$$

dont l'interprétation graphique est donnée sur la figure de droite.

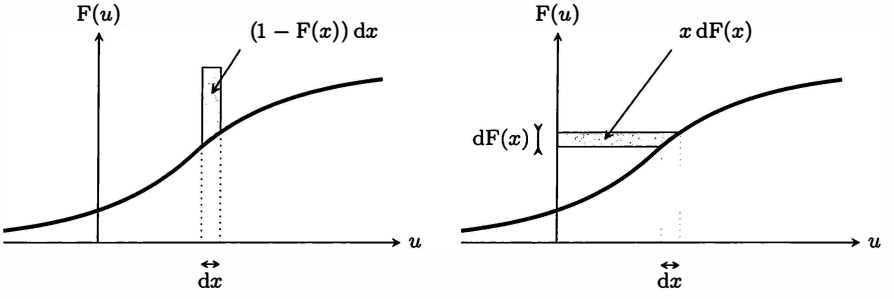


FIGURE 14.2 — *Deux manières d'intégrer selon* F *pour obtenir l'espérance d'une variable aléatoire. Comparer à la figure 13.3 page 300.*

14.4 ESPACE L^2 ; INTERPRÉTATIONS GÉOMÉTRIQUES

Espaces $\mathcal{L}^2$ et L^2 § 218

On définit l'espace $\mathcal{L}^2(\Omega, \mathfrak{T}, \mathbf{P})$ — ou $\mathcal{L}^2$ s'il n'y a pas d'ambiguïté possible — comme l'ensemble des variables aléatoires réelles sur $(\Omega, \mathfrak{T}, \mathbf{P})$ qui sont de carré intégrable, c'est-à-dire qui admettent un moment d'ordre 2.

De la même manière que nous avons défini l'espace L^1 page 313, on définit l'espace L^2 comme celui des classes d'équivalences de $\mathcal{L}^2$ pour la relation d'équivalence d'égalité presque sûre : si X et Y sont des variables aléatoires de $\mathcal{L}^2$ et si $X = Y$ p.s., alors elles sont dans la même classe (on dit que X et Y sont des **représentants** de cette classe). Par abus, nous écrirons souvent « $X \in L^2$ » pour dire que X est une variable aléatoire de carré intégrable, mais que la propriété envisagée ne dépend en réalité que de la classe de X et pas du choix d'un représentant.

Inégalité de Cauchy-Schwarz § 219

On ne s'en lasse pas, voici une nouvelle fois l'inégalité de Cauchy-Schwarz dans le cadre abstrait de l'espace $\mathcal{L}^2$.

■ **Théorème 14.30 (Inégalité de Cauchy-Schwarz)** *Soient* X *et* Y *deux variables de* $\mathcal{L}^2$ *; alors* XY *est intégrable et*

$$\left|\mathbf{E}(XY)\right|^2 \leqslant \mathbf{E}(X^2)\,\mathbf{E}(Y^2).$$

Il n'y a égalité que si X *et* Y *sont presque sûrement liées.*

DÉMONSTRATION : L'inégalité arithmético-géométrique permet de montrer la première propriété :

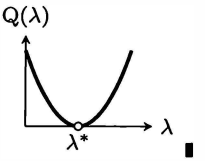
$$|XY| \leqslant \frac{1}{2}(X^2 + Y^2)$$

et le membre de droite est intégrable, donc celui de gauche également. De plus, pour tout réel λ, la variable $(X + \lambda Y)^2 = X^2 + 2\lambda XY + \lambda^2 Y^2$ est intégrable, et

$$0 \leqslant Q(\lambda) := \mathbf{E}\big((X + \lambda Y)^2\big) = \mathbf{E}(X^2) + 2\lambda\,\mathbf{E}(XY) + \lambda^2\,\mathbf{E}(Y^2).$$

Si $\mathbf{E}(Y^2) = 0$, alors $Y = 0$ presque sûrement et l'inégalité cherchée se réduit à $0 = 0$, ce qui est un cas d'égalité ; de plus, on a $Y = 0 \cdot X$ presque sûrement.

Dans le cas contraire, le membre de droite de l'expression précédente est un trinôme $Q(\lambda)$ du second degré et de signe constant ; son discriminant est donc négatif : $4\big[\mathbf{E}(XY)^2 - \mathbf{E}(X^2)\,\mathbf{E}(Y^2)\big] \leqslant 0$, ce qui achève la preuve de l'inégalité. Le cas d'égalité est celui où le discriminant s'annule ; dans ce cas, le trinôme admet une racine double λ^*, en laquelle on peut écrire que $0 = \mathbf{E}\big((X + \lambda^* Y)^2\big)$, ce qui montre que $X + \lambda^* Y$ est presque sûrement nulle. ∎



On sait déjà que $\mathcal{L}^2 \subset \mathcal{L}^1$ et donc que $L^2 \subset L^1$: d'après le théorème 14.13 toute variable admettant un moment d'ordre deux admet un moment d'ordre un[8]. On déduit du théorème précédent la positivité de la variance :

■ **Corollaire 14.31** *Si* $X \in \mathcal{L}^2$, *alors* $\mathbf{E}(X)^2 \leqslant \mathbf{E}(X^2)$. *Notamment,* $\mathbf{V}(X) \geqslant 0$.

DÉMONSTRATION : Prendre $Y = 1$ dans l'inégalité de Cauchy-Schwarz. ∎

§220 **Covariance**

Soient X et Y deux variables de carré intégrable. Alors $X' := X - \mathbf{E}X$ et $Y' := Y - \mathbf{E}Y$ sont également de carré intégrable, donc $X'Y'$ est intégrable par l'inégalité de Cauchy-Schwarz. On définit leur **covariance** comme

$$\mathbf{Cov}(X, Y) = \mathbf{E}\big[(X - \mathbf{E}X)(Y - \mathbf{E}Y)\big] = \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y).$$

Les propriétés suivantes découlent de la linéarité de l'espérance et de l'inégalité de Cauchy-Schwarz :

■ **Théorème 14.32 (Propriétés de la covariance)** *Soient* X, Y, Z *et* T *des variables aléatoires de carré intégrable. Alors*

i) $\mathbf{Cov}(X, Y) = \mathbf{Cov}(Y, X)$ *(symétrie);*

ii) $\mathbf{Cov}(\alpha X + \beta, \gamma Y + \delta) = \alpha\gamma\ \mathbf{Cov}(X, Y)$;

iii) $\mathbf{Cov}(X, C^{\text{te}}) = 0$;

iv) $\mathbf{Cov}(X + Y, Z + T) = \mathbf{Cov}(X, Z) + \mathbf{Cov}(X, T) + \mathbf{Cov}(Y, Z) + \mathbf{Cov}(Y, T)$;

v) $\mathbf{V}(X) = \mathbf{Cov}(X, X)$;

vi) $\big|\,\mathbf{Cov}(X, Y)\big| \leqslant \sigma_X \cdot \sigma_Y$. *(Cauchy-Schwarz)*

vii) $\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\,\mathbf{Cov}(X, Y)$.

viii) $\mathbf{V}(aX + bY) = a^2\ \mathbf{V}(X) + b^2\ \mathbf{V}(Y) + 2ab\ \mathbf{Cov}(X, Y)$.

Plus généralement, si $X_1, X_2, \ldots, X_n$ *sont des variables aléatoires admettant des variances et des covariances, alors*

$$\mathbf{V}(X_1 + \cdots + X_n) = \sum_{k=1}^{n} \mathbf{V}(X_k) + 2 \sum_{1 \leqslant i < j \leqslant n} \mathbf{Cov}(X_i, X_j)$$

(8). On peut également définir les espaces L^p des variables admettant un moment d'ordre p ; le théorème 14.13 montre que l'on a les inclusions

$$\cdots \subset L^{p+1} \subset L^p \subset L^{p-1} \subset \cdots \subset L^2 \subset L^1.$$

Structure de l'espace L^2 § 221

Les espaces $\mathcal{L}^2$ et L^2 sont évidemment stables par multiplication par un scalaire. Si $X, Y \in \mathcal{L}^2$, alors $(X+Y)^2 = X^2+2XY+Y^2$ est intégrable d'après le théorème 14.30, donc $X + Y$ est également dans $\mathcal{L}^2$. On en déduit que

■ **Théorème 14.33** *Les espaces* $\mathcal{L}^2$ *et* L^2 *ont une structure d'espace vectoriel.*

La structure de L^2 est considérablement enrichie par la remarque suivante :

■ **Proposition 14.34** *L'application* $\langle\cdot|\cdot\rangle$ *définie par* $\langle X|Y\rangle := \mathbf{E}(XY)$ *est un produit scalaire sur l'espace* L^2 *des classes de variables aléatoires réelles de carré intégrable. Cet espace a donc une structure d'espace préhilbertien.*

L'application en question est bien définie sur L^2, au sens où l'espérance calculée ne dépend pas des représentants choisis pour les classes d'équivalence de X et de Y : si $X' = X$ p.s. et si $Y' = Y$ p.s., alors $\mathbf{E}(X'Y') = \mathbf{E}(XY)$. Cette application est bilinéaire et symétrique et si X est un élément de $\mathcal{L}^2$ tel que $\langle X|X\rangle = 0$, alors $\mathbf{E}(X^2) = 0$ donc X est presque sûrement nulle, c'est-à-dire que « $X = 0$ » du point de vue des classes.

On en déduit classiquement que, si l'on pose

$$\|X\|_2 = \sqrt{\mathbf{E}(X^2)}$$

pour tout $X \in L^2$, l'application $\|\cdot\|_2$ est une norme — le seul point à démontrer est l'inégalité triangulaire, qui découle de l'inégalité de Cauchy-Schwarz :

$$\|X+Y\|_2^2 = \|X\|_2^2 + 2\langle X|Y\rangle + \|Y\|_2^2 \leqslant \|X\|_2^2 + 2\|X\|_2\|Y\|_2 + \|Y\|_2^2 = \big(\|X\|_2 + \|Y\|_2\big)^2.$$

■ **Théorème 14.35 (Riesz-Fisher)** *L'espace* L^2*, muni de son produit scalaire*

$$\langle X|Y\rangle = \mathbf{E}(XY),$$

est complet. C'est donc un espace de Hilbert.

La démonstration de ce théorème est un peu technique, mais est intéressante car elle donne un exemple de recherche de la limite d'une suite de variables aléatoires.

DÉMONSTRATION : Commençons par rappeler le résultat classique :

■ **Lemme 14.36** *Soit* $(X_n)_{n\geqslant 1}$ *une suite de variables aléatoires* positives *telle que* $\sum \mathbf{E}X_n$ *converge. Alors* $\sum X_n$ *converge presque sûrement.*

DÉMONSTRATION DU LEMME : Notons $S = \sum_{n=1}^{\infty} X_n$, qui est une variable aléatoire à valeurs dans $[0\,;+\infty]$; l'égalité du corollaire 14.18, qui s'écrit $\mathbf{E}(S) = \sum_{n=1}^{\infty} \mathbf{E}X_n$, montre que S est presque sûrement finie, ou encore que la série $\sum X_n$ converge presque sûrement. ▮

Soit $(X_n)_{n\geqslant 1}$ une suite de Cauchy de L^2. On peut donc construire par récurrence une suite $(k_n)_{n\geqslant 1}$ strictement croissante d'entiers, telle que

$$\forall p, q \geqslant k_n \qquad \|X_p - X_q\|_2 \leqslant \frac{1}{2^n}.$$

L'inégalité de Cauchy-Schwarz permet d'écrire que

$$\forall n \geqslant 1 \qquad \mathbf{E}|X_{k_{n+1}} - X_{k_n}| \leqslant \|X_{k_{n+1}} - X_{k_n}\|_2 \leqslant \frac{1}{2^n},$$

ce qui fait que la série $\sum \mathbf{E}|X_{k_{n+1}} - X_{k_n}|$ converge. D'après le lemme précédent, la série positive $\sum|X_{k_{n+1}} - X_{k_n}|$ converge presque sûrement et, par complétude de $\mathbb{R}$, la série

$\sum(X_{k_{n+1}} - X_{k_n})$ converge presque sûrement elle aussi (A3) : en d'autres termes, la suite $(X_{k_n})_{n\geqslant 1}$ converge presque sûrement. On ne peut pas définir « $\lim X_n$ » autrement que « presque sûrement » ; on a donc recours à l'astuce suivante : on définit

$$X := \limsup_{n\to\infty} X_{k_n},$$

ce qui fait que X est bien une variable aléatoire et que $X(\omega) = \lim X_{k_n}(\omega)$ pour tout ω d'un ensemble de probabilité 1.

Il reste à montrer que X est bien limite, dans L^2, de la suite $(X_n)_{n\geqslant 1}$, et pas seulement d'une de ses sous-suites. Fixons momentanément un entier n. Pour tout $p \geqslant k_n$ et pour tout $i \geqslant n$, on a

$$\|X_p - X_{k_i}\|_2^2 = \mathbf{E}|X_p - X_{k_i}|^2 \leqslant \frac{1}{2^{2n}}.$$

Faisons maintenant tendre i vers l'infini et appliquons le lemme de Fatou : puisque $|X_p - X_{k_i}|^2 \geqslant 0$ et $|X_p - X_{k_i}|^2 \xrightarrow[i\to\infty]{} |X_p - X|^2$ presque sûrement, on conclut que $\mathbf{E}|X_p - X|^2 \leqslant \liminf_i \mathbf{E}|X_p - X_{k_i}|^2 \leqslant 1/2^{2n}$; ceci est vrai pour tout entier $p \geqslant k_n$.

Cela prouve d'abord que $X = (X - X_p) + X_p$ est bien dans L^2. Par ailleurs, si l'on note ψ le « pseudo-inverse » de $n \mapsto k_n$, défini par $\psi(p) = \max\{n \in \mathbb{N} ;\ k_n \leqslant p\}$, alors $\psi(p)$ tend vers l'infini quand $p \to \infty$. L'inégalité précédente s'écrit :

$$\forall p \in \mathbb{N}^* \qquad \mathbf{E}|X_p - X|^2 \leqslant \frac{1}{2^{2\psi(p)}}$$

ce qui montre que $(X_p)_{p\geqslant 1}$ converge bien vers X dans L^2 et achève la démonstration. ∎

Remarque 14.37 De manière plus générale, les espaces L^p ($p \geqslant 1$) sont complets pour la norme

$$\|X\|_p = \left(\mathbf{E}|X|^p\right)^{1/p}.$$

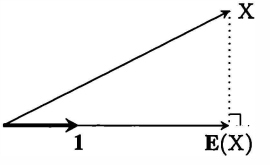
La démonstration est la même que la précédente ; on remplace simplement l'inégalité de Cauchy-Schwarz $\mathbf{E}|X| \leqslant \|X\|_2$ par l'inégalité plus générale de Lyapounov $\|X\|_1 \leqslant \|X\|_p$. Nous n'utiliserons pas ces propriétés ici.

§ 222 **Interprétations géométriques**

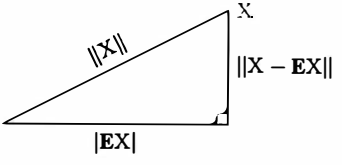
Si l'on identifie un réel à la classe des variables aléatoires presque sûrement égales à ce réel, alors pour toute variable $X \in L^2$,

$$\mathbf{E}(X) = \frac{\langle X|\mathbf{1}\rangle}{\langle \mathbf{1}|\mathbf{1}\rangle}\,\mathbf{1}$$

où l'on a noté $\mathbf{1}$ la (classe de la) variable aléatoire constante égale à 1. Ainsi, *l'espérance de* X *est la projection orthogonale de* X *sur la droite vectorielle* Vect($\mathbf{1}$). C'est la valeur qui approche le mieux X au sens de la norme $\|\cdot\|_2$.



Appliquons le théorème de Pythagore au triangle rectangle précédent, dont on représente ici les longueurs :



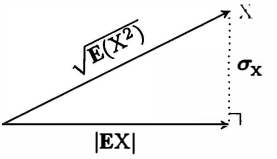
On obtient

$$\mathbf{E}(X^2) = \|X\|^2 = \mathbf{E}(X)^2 + \mathbf{E}(X - \mathbf{E}X)^2,$$

que l'on écrit habituellement sous la forme

$$\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2.$$

Ainsi, *la formule de König-Huygens est une expression du théorème de Pythagore dans une projection orthogonale.* La figure précédente est donc plus agréable sous l'aspect suivant :



Puisque la projection orthogonale d'un vecteur sur un sous-espace vectoriel minimise la distance, nous pouvons généraliser le résultat de l'exercice 11.1 page 254. Soit $X \in L^2$ et notons $m = \mathbf{E}(X)$. Pour tout réel a, calculons

$$\begin{aligned}\varphi(a) &= \mathbf{E}\big[(X-a)^2\big] = \mathbf{E}\big[(X-m+m-a)^2\big] \\ &= \mathbf{E}\big[(X-m)^2\big] + 2(m-a)\,\mathbf{E}(X-m) + (m-a)^2 \\ &= \mathbf{V}(X) + (m-a)^2,\end{aligned}$$

ce qui prouve que la valeur $a = m$ est l'unique valeur qui minimise φ. Le lecteur tracera sans peine un graphe donnant l'interprétation géométrique de cette égalité.

■ **Proposition 14.38** *Soit $X \in L^2$ une variable aléatoire de carré intégrable, et notons σ^2 sa variance. Alors, pour tout $a \in \mathbb{R}$, on a*

$$\mathbf{E}\big[(X-a)^2\big] \geqslant \sigma^2,$$

avec égalité si et seulement si $a = \mathbf{E}(X)$.

Enfin, soient X et Y deux variables aléatoires de L^2 décorrélées, c'est-à-dire vérifiant $\mathbf{Cov}(X,Y) = 0$; la relation $\mathbf{E}\big[(X-\mathbf{E}X)(Y-\mathbf{E}Y)\big] = 0$ s'interprète géométriquement comme l'orthogonalité de $X' = X - \mathbf{E}X$ et de $Y' = Y - \mathbf{E}Y$. *La décorrélation de deux variables n'est rien d'autre que l'orthogonalité des variables centrées correspondantes.*

Meilleur estimateur quadratique d'une variable aléatoire § 223

Soit $(X_1, X_2, \ldots, X_n)$ une famille de variables aléatoires de l'espace L^2, deux à deux orthogonales, et de variance unité :

$$\langle X_i | X_j \rangle = \mathbf{E}(X_i X_j) = \delta_{i,j}.$$

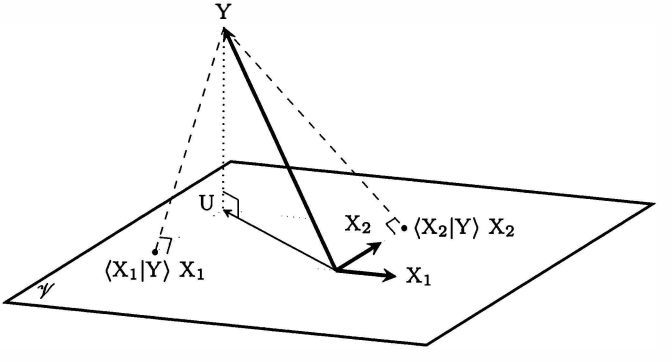
Une telle famille est dite **orthonormée**. (On rappelle que, en partant d'une famille libre d'un espace préhilbertien, on peut construire une famille orthonormée au moyen du procédé de Gram-Schmidt .)

Soit $Y \in L^2$ une variable aléatoire. On cherche à déterminer, dans l'espace

$$\mathscr{V} = \mathrm{Vect}\,(X_1, X_2, \ldots, X_n),$$

quelle est la variable U qui approche le mieux Y dans L^2, c'est-à-dire celle pour laquelle $\|U - Y\|^2$ est minimal. C'est un problème de géométrie bien connu dans les espaces préhilbertiens ; la solution, unique, est donnée par la projection orthogonale U de Y sur le sous-espace vectoriel $\mathscr{V}$. On peut exprimer U dans la base orthonormée $(X_1, X_2, \ldots, X_n)$ de $\mathscr{V}$:

$$U = \sum_{k=1}^{n} \langle X_k | Y \rangle\, X_k.$$



Ce résultat se généralise. Si $\mathscr{V}$ est un sous-espace vectoriel complet[9] de L^2, et si $(X_n)_{n\geqslant 1}$

(9). En pratique, il suffit donc qu'il soit fermé.

est une famille orthonormée telle que $\mathcal{Y}$ soit égal à l'adhérence de $\mathrm{Vect}\{X_n \ ; \ n \geqslant 1\}$, alors $(X_n)_{n\geqslant 1}$ est une base hilbertienne de $\mathcal{Y}$ et la meilleure approximation de Y dans $\mathcal{Y}$ est son projeté orthogonal

$$\pi(Y) = \sum_{n=1}^{\infty} \langle X_n | Y \rangle \, X_n,$$

la série précédente étant convergente dans L^2, c'est-à-dire que

$$\lim_{n\to\infty} \left\| Y - \sum_{k=1}^{n} \langle X_k | Y \rangle \, X_k \right\|^2 = 0.$$

14.5 INÉGALITÉS

§224 **Inégalité de Markov**

Tout comme ses avatars que sont les inégalités de Bienaymé-Tchebychev, l'inégalité de Markov est une inégalité à caractère *universel* — elle est valable pour toute variable aléatoire positive — et ne donne généralement que des majorations grossières.

■ **Théorème 14.39 (Inégalité de Markov)** *Soit* Y *une variable aléatoire positive et intégrable. Alors, pour tout* $t > 0$,

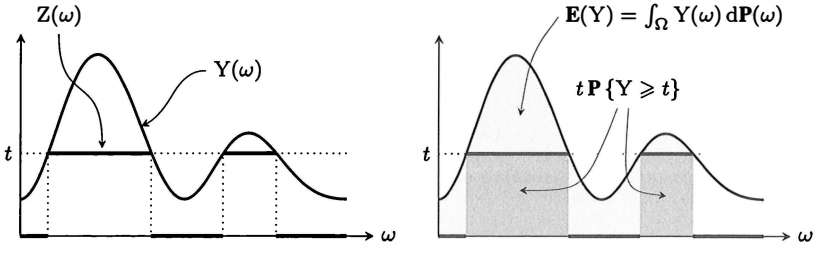
$$\mathbf{P}\{Y \geqslant t\} \leqslant \frac{\mathbf{E}(Y)}{t}.$$

DÉMONSTRATION : Donnons deux démonstrations de cette inégalité, sans supplément de prix.

Notons Z la variable aléatoire simple définie par $Z(\omega) = t$ si $Y(\omega) \geqslant t$, et $Z(\omega) = 0$ sinon. Alors on a l'inégalité de fonctions : $0 \leqslant Z \leqslant Y$. De plus, Z ne prenant que les valeurs 0 et t, son espérance vaut

$$\mathbf{E}(Z) = t\,\mathbf{P}\{Z = t\} = t\,\mathbf{P}\{Y \geqslant t\}.$$

En évaluant l'espérance dans chaque membre de l'inégalité $Z \leqslant Y$ et en utilisant la propriété de croissance de l'espérance, on obtient le résultat voulu, qui est illustré ci-dessous.



Le but de la seconde démonstration (qui, au fond, ne diffère en rien de la première) est de faire manipuler les notations acquises dans ce chapitre. Soit $t > 0$. On a, par croissance (*) puis positivité (**) :

$$t\,\mathbf{P}\{Y \geqslant t\} = \int_{\{Y\geqslant t\}} t \, d\mathbf{P} \underset{(*)}{\leqslant} \int_{\{Y\geqslant t\}} Y \, d\mathbf{P} \underset{(**)}{\leqslant} \int_{\Omega} Y \, d\mathbf{P} = \mathbf{E}(Y). \qquad \blacksquare$$

On utilise parfois l'inégalité plus faible $\mathbf{P}\{Y > t\} \leqslant \mathbf{E}(Y)/t$.

Inégalités de Bienaymé-Tchebychev §225

Comme déjà vu dans le cadre des variables aléatoires absolument continues, l'inégalité de Markov permet de démontrer la fameuse inégalité de Bienaymé-Tchebychev.

■ **Théorème 14.40 (Inégalité de Bienaymé-Tchebychev)** *Soit* X *une variable aléatoire telle que* X^2 *soit intégrable, et* $\varepsilon > 0$. *Alors*

$$\mathbf{P}\{\,|X - \mathbf{E}X| \geqslant \varepsilon\} \leqslant \frac{\sigma_X^2}{\varepsilon^2}.$$

DÉMONSTRATION : Il suffit d'appliquer l'inégalité de Markov à la variable positive $|X - \mathbf{E}X|^2$, en remarquant que $\{\,|X - \mathbf{E}X| \geqslant \varepsilon\} = \{\,|X - \mathbf{E}X|^2 \geqslant \varepsilon^2\}$. ∎

Si l'on applique l'inégalité de Markov à $|X|^k$, on obtient une autre majoration, qui peut être plus précise que la précédente, et qui est valable dès que X admet un moment d'ordre k.

■ **Théorème 14.41 (Inégalité de Bienaymé-Tchebychev d'ordre k)** *Si* k *est un entier,* X *une variable aléatoire telle que* X^k *soit intégrable, et* $\varepsilon > 0$,

$$\mathbf{P}\{\,|X| \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}\,|X|^k}{\varepsilon^k}.$$

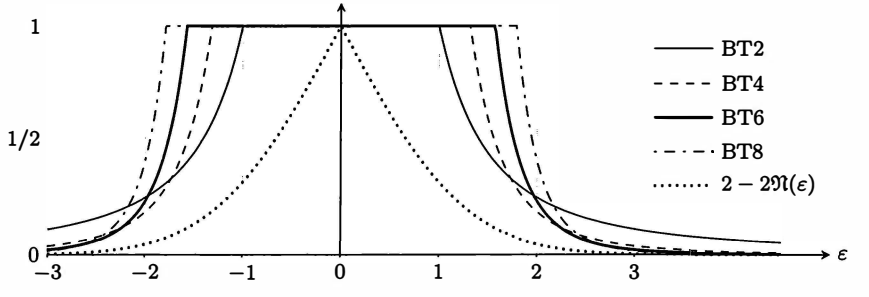
Cette inégalité est bien sûr plus profitable lorsqu'on l'applique à une variable centrée :

$$\mathbf{P}\{\,|X - \mathbf{E}X| \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}\,|X - \mathbf{E}X|^k}{\varepsilon^k}.$$

Remarque 14.42 (Comparaisons numériques) On donne ci-dessous les valeurs (connues et tabulées) de $\mathbf{P}\{|X| > \varepsilon\}$ lorsque X suit la loi normale $\mathcal{N}(0\,;\,1)$, ainsi que les majorations obtenues par les inégalités de Bienaymé-Tchebychev d'ordre 2, 4, 6 et 8. Lorsque ces majorations sont inutilisables (le membre de droite étant supérieur ou égal à 1), elles apparaissent entre crochets.

ε	$\mathbf{P}\{\,\lvert X\rvert \geqslant \varepsilon\}$	BT2	BT4	BT6	BT8
	$2(1 - \mathfrak{N}(\varepsilon))$	$1/\varepsilon^2$	$3/\varepsilon^4$	$15/\varepsilon^6$	$105/\varepsilon^8$
1	0,3173	1	[3]	[15]	[105]
2	0,0455	0,25	0,1875	0,2344	0,4102
3	0,0027	0,111	0,037	0,0205	0,016
4	0,000063	0,0625	0,01172	0,0037	0,0016

Sur le graphe ci-dessous, on trace les courbes correspondant à la valeur exacte de $\mathbf{P}\{|X| > \varepsilon\}$, ainsi que les majorations par les inégalités de Bienaymé-Tchebychev d'ordre 2 à 8. Pour des grandes valeurs de ε, les meilleurs résultats sont obtenus pour un ordre élevé, mais restent très grossiers comparés au résultat exact.



§226 **Inégalité de Bernstein**

Les inégalités précédentes se généralisent à d'autres fonctions que les fonctions puissances : en effet, $\{X \geqslant \varepsilon\} \subset \{\varphi(X) \geqslant \varphi(\varepsilon)\}$ dès que φ est une fonction croissante. Si φ est de plus positive, on peut alors appliquer l'inégalité de Markov et obtenir le résultat suivant.

■ **Théorème 14.43 (Inégalités de Bernstein)** *Soit $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction croissante et à valeurs strictement positives. Si X est une variable aléatoire telle que $\varphi(X)$ soit intégrable,*

$$\mathbf{P}\{X \geqslant \varepsilon\} \leqslant \frac{\mathbf{E}[\varphi(X)]}{\varphi(\varepsilon)}.$$

§227 **Inégalité de Jensen (convexité)**

Si φ est une fonction convexe définie sur un intervalle I et si $p_1, p_2, \ldots, p_n$ sont des réels positifs de somme 1, alors pour tous réels $x_1, x_2, \ldots, x_n$ de I, on a l'inégalité de convexité classique

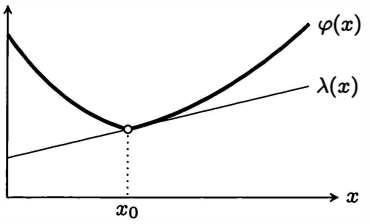
$$\varphi\left(\sum_{k=1}^{n} p_k x_k\right) \leqslant \sum_{k=1}^{n} p_k \varphi(x_k).$$

Cette inégalité admet une généralisation continue (A34), que l'on peut d'ailleurs interpréter en termes de densité de probabilité. En voici une version probabiliste.

■ **Théorème 14.44 (Inégalité de Jensen)** *Soit $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction mesurable convexe. Si X est une variable aléatoire intégrable,*

$$\varphi(\mathbf{E}X) \leqslant \mathbf{E}\varphi(X).$$

DÉMONSTRATION : Si x_0 est un réel, il existe une fonction affine λ telle que $\lambda(x_0) = \varphi(x_0)$ et $\lambda \leqslant \varphi$. En effet, φ étant convexe, elle admet en x_0 des demi-tangentes à gauche comme à droite, et il suffit de choisir pour fonction λ la fonction affine associée à l'une des ces demi-tangentes. On a donc, pour tout réel x, l'inégalité $\varphi(x) \geqslant \lambda(x) = \alpha(x - x_0) + \varphi(x_0)$.



Fixons la valeur $x_0 := \mathbf{E}(X)$, et prenons α comme précédemment. Alors, pour tout $\omega \in \Omega$, en posant $x = X(\omega)$ l'inégalité précédente devient

$$\varphi(X) \geqslant \alpha(X - \mathbf{E}X) + \varphi(\mathbf{E}X).$$

Le membre de droite admet une espérance, donc le membre de gauche en admet également une dans $\mathbb{R} \cup \{+\infty\}$[10], et en prenant l'espérance on obtient $\mathbf{E}\varphi(X) \geqslant \varphi(\mathbf{E}X)$. ∎

⇗14.1

(10). Dans cette phrase, on ne dit rien de plus que : toute variable positive admet une espérance dans $[0\,;+\infty]$.

14.6 ANNEXE : LE THÉORÈME DE CONVERGENCE MONOTONE

On donne ici une démonstration élémentaire du théorème de convergence monotone : *si* $(X_n)_{n\geqslant 1}$ *est une suite croissante de variables aléatoires positives telle que* $X_n \uparrow X$ *presque sûrement, alors* $\mathbf{E}(X_n) \uparrow \mathbf{E}(X)$.

La propriété de croissance de l'espérance (proposition 14.7) montre que la suite de terme général $\mathbf{E}(X_n)$ est croissante, et majorée (dans $\mathbb{R}^+ \cup \{+\infty\}$) par $\mathbf{E}(X)$: elle converge et

$$\lim_{n\to\infty} \mathbf{E}(X_n) \leqslant \mathbf{E}(X). \qquad (*)$$

Il reste à montrer l'inégalité réciproque.

Pour chaque entier n, on construit une suite $\left(X_n^{(k)}\right)_k$ de variables simples (théorème 13.25) telle que $X_n^{(k)} \uparrow X_n$ quand $k \to \infty$. On pose alors, pour tout entier k :

$$Y^{(k)} = \max\left(X_1^{(k)}, X_2^{(k)}, \ldots, X_k^{(k)}\right).$$

- Première remarque : *la suite de terme général* $Y^{(k)}$ *est croissante.* En effet, pour tout entier $k \geqslant 2$, on a

  $$Y^{(k-1)} = \max\left(X_1^{(k-1)}, X_2^{(k-1)}, \ldots, X_{k-1}^{(k-1)}\right) \leqslant \max\left(X_1^{(k)}, X_2^{(k)}, \ldots, X_{k-1}^{(k)}, X_k^{(k)}\right) = Y^{(k)}.$$

  Cette suite admet donc une limite simple — une variable à valeurs dans $[0\,;+\infty]$ — que l'on notera Y.
- Seconde remarque : on a, pour tout entier k,

  $$Y^{(k)} = \max\left(X_1^{(k)}, X_2^{(k)}, \ldots, X_k^{(k)}\right) \leqslant \max(X_1, X_2, \ldots, X_k) = X_k.$$

On en déduit, pour tous entiers n et k tels que $1 \leqslant n \leqslant k$, la double inégalité

$$X_n^{(k)} \leqslant Y^{(k)} \leqslant X_k.$$

À n fixé, on peut passer à la limite $k \to \infty$ et obtenir

$$X_n \leqslant Y \leqslant X,$$

puis prendre la limite $n \to \infty$ qui donne directement $X = Y$.

Ainsi, la suite $(Y^{(k)})_{k\geqslant 1}$ est une suite croissante de variables simples, qui converge vers X. D'après le théorème 14.6, on a donc $\mathbf{E}(X) = \lim \mathbf{E}(Y^{(k)})$. Puisqu'enfin $Y^{(k)} \leqslant X_k$ pour tout entier k, on en déduit

$$\mathbf{E}(X) \leqslant \lim_{k\to\infty} \mathbf{E}(X_k). \qquad (**)$$

Les deux équations (*) et (**) permettent de conclure.

EXERCICES

♦ ***Exercice 14.1 (Inégalité de Lyapounov)*** Soit X une variable aléatoire telle que X^n soit intégrable. Montrer que, pour tous réels $0 < r < s \leqslant n$, on a

$$\left(\mathbf{E}\,|X|^r\right)^{1/r} \leqslant \left(\mathbf{E}\,|X|^s\right)^{1/s}.$$

Noter que l'on peut omettre la restriction $s \leqslant n$ si l'on accepte de travailler dans $\mathbb{R}^+ \cup \{+\infty\}$.

♦ ***Exercice 14.2*** Quelles sont les variables aléatoires qui vérifient $\mathbf{E}(X^2) = \mathbf{E}(X)^2$? Celles qui vérifient $\mathbf{E}(1/X) = 1/\mathbf{E}(X)$? (On supposera que X est positive.)

♦ ***Exercice 14.3 (Espérance de la valeur maximale)*** Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires positives, suivant toutes la même loi, et intégrables ; elles ne sont pas supposées indépendantes. On note $M_n = \max(X_1, X_2, \ldots, X_n)$. Le but de l'exercice est de montrer que $\mathbf{E}(M_n) = o(n)$.

i) Montrer que, pour tout réel x,

$$g_n(x) := \frac{1}{n}\mathbf{P}\{M_n \geqslant x\} \leqslant \mathbf{P}\{X_1 \geqslant x\}.$$

ii) Conclure.

♦ ***Exercice 14.4*** Soit X une variable aléatoire admettant un moment d'ordre 2 ; on note m l'espérance de X. Montrer que $\mathbf{V}(X) = 0$ si et seulement si $X = m$ presque sûrement.

Indications

◊ *Indication 14.1 :* Poser $q = s/r$, $Y = |X|^r$ et appliquer l'inégalité de Jensen à la fonction $\varphi(x) = |x|^q$.

◊ *Indication 14.2 :* Pour la première, on peut supposer $X \in L^2$ et considérer la variance. Pour la seconde, $X \in L^1$ et $1/X \in L^1$... puis Cauchy-Schwarz à la rescousse.

◊ *Indication 14.3 :* On utilisera la propriété suivante : si Y est une variable aléatoire positive de fonction de répartition F, alors Y admet une espérance si et seulement si $1 - F$ est intégrable sur $[0\,;+\infty[$, et si c'est le cas, alors $\mathbf{E}(Y) = \int_0^{+\infty} 1 - F$ (voir paragraphe § 237, page 336).

◊ *Indication 14.4 :* Utiliser l'inégalité de Bienaymé-Tchebychev.

SOLUTIONS

♦ **Solution 14.1** On pose $q = s/r$; on a $q \geqslant 1$. La fonction $\varphi : x \mapsto |x|^q$ est convexe ; l'inégalité de Jensen s'écrit, pour la variable aléatoire $Y = |X|^r$: $\varphi(\mathbf{E}Y) \leqslant \mathbf{E}\varphi(Y)$, soit $(\mathbf{E}Y)^q \leqslant \mathbf{E}(Y^q)$, c'est-à-dire $(\mathbf{E}|X|^r)^{s/r} \leqslant \mathbf{E}|X|^s$, ce qui est le résultat attendu.

♦ **Solution 14.2** Si $\mathbf{E}(X^2) = \mathbf{E}(X)^2$ et si ces quantités sont finies, alors X est de variance nulle, donc X est presque sûrement constante ; la réciproque est immédiate.

Soit X une variable aléatoire positive de L^1 telle que $1/X \in L^1$ (on a alors $X > 0$ p.s.), et vérifiant $\mathbf{E}(1/X) = 1/\mathbf{E}(X)$. Ces quantités étant toutes deux finies, $\mathbf{E}(X)\,\mathbf{E}(1/X) = 1 = \mathbf{E}(1)^2$, ou encore $\mathbf{E}(Y^2)\,\mathbf{E}(Z^2) = \mathbf{E}(YZ)^2$ en posant $Y = \sqrt{X}$ et $Z = 1/\sqrt{X}$: on est donc dans le cas d'égalité de Cauchy-Schwarz. On en déduit que X et $1/X$ sont presque sûrement liées, c'est-à-dire qu'il existe une constante a telle que $X = a/X$, ou encore $X^2 = a$ presque sûrement. Ainsi, X est une variable ne prenant que les valeurs $\sqrt{a}$ et $-\sqrt{a}$. Notons $p := \mathbf{P}\{X = \sqrt{a}\}$ et $1 - p = \mathbf{P}\{X = -\sqrt{a}\}$. Alors $\mathbf{E}(X) = (2p-1)\sqrt{a}$ tandis que $\mathbf{E}(1/X) = (2p-1)/\sqrt{a}$. La condition $\mathbf{E}(X)\,\mathbf{E}(1/X) = 1$ implique $(2p-1)^2 = 1$, c'est-à-dire $p = 0$ ou $p = 1$. Conclusion : la variable aléatoire X est presque sûrement constante. La réciproque est immédiate.

♦ **Solution 14.3**

i) L'événement $\{M_n \geqslant x\}$ a lieu si et seulement si au moins l'un des X_k est supérieur ou égal à x. On en déduit que

$$\mathbf{P}\{M_n \geqslant x\} = \mathbf{P}\left(\bigcup_{k=1}^{n} \{X_k \geqslant x\}\right) \leqslant \sum_{k=1}^{n} \mathbf{P}\{X_k \geqslant n\} = n\,\mathbf{P}\{X_1 \geqslant n\}.$$

ii) Puisque M_n est une variable positive, on peut écrire

$$\mathbf{E}(M_n) = \int_0^{+\infty} \mathbf{P}\{M_n \geqslant x\}\,\mathrm{d}x.$$

On a donc

$$\frac{1}{n}\mathbf{E}(M_n) = \int_0^{+\infty} g_n(x)\,\mathrm{d}x$$

et l'on a dominé dans la première question la suite $(g_n)_{n\geqslant 1}$ par la fonction $\psi : 1 - F$ qui est intégrable. Comme de plus $g_n(x) \leqslant 1/n$ pour tout $x \geqslant 0$, la suite $(g_n)_{n\geqslant 1}$ converge simplement vers 0 et le théorème de convergence dominée (pour les fonctions) permet de conclure que $\mathbf{E}(M_n)/n \to 0$, ou encore $\mathbf{E}(M_n) = o(n)$.

♦ **Solution 14.4** Si $X = m$ presque sûrement, alors $X = \mathbf{E}(X)$ presque sûrement, ce qui entraîne $(X - \mathbf{E}X)^2 = 0$ presque sûrement et $\mathbf{V}(X) = 0$.

Réciproquement, supposons $\mathbf{V}(X) = 0$. Par l'inégalité de Bienaymé-Tchebychev, on a, pour tout $n \in \mathbb{N}$, $\mathbf{P}\{|X - \mathbf{E}X| \geqslant 1/n\} = 0$ et, par l'axiome de continuité monotone décroissante 3'', on obtient

$$\mathbf{P}\{|X - \mathbf{E}X| > 0\} = \lim_{n\to\infty} \downarrow \mathbf{P}\{|X - \mathbf{E}X| \geqslant 1/n\} = 0.$$

(On peut également invoquer le théorème 14.11 page 312.)

Chapitre 15

Vecteurs aléatoires

La généralisation de couples de variables aléatoires est la notion de vecteurs aléatoires, ou encore variables aléatoires à valeurs dans $\mathbb{R}^n$. Bien que l'on puisse travailler dans un espace vectoriel euclidien de dimension n quelconque, nous supposerons toujours que celui-ci est égal à $\mathbb{R}^n$, et nous identifierons donc un vecteur $\mathbf{X}$ à un n-uplet $(X_1, X_2, \dots, X_n)$ de variables aléatoires.

Après une rapide étude de la manière dont on munit $\mathbb{R}^n$ d'une structure d'espace probabilisé, nous introduisons la notion de matrice de covariance d'un vecteur aléatoire, puis le théorème de Fubini permettant de calculer des intégrales multiples en se ramenant à n intégrales simples, sous des hypothèses simples. La dernière partie du chapitre concerne les changements de variables aléatoires — permettant, par exemple, d'étudier la loi d'un couple (U, V), image d'un autre couple (X, Y) de loi connue par une application $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$.

Notations

§ 228

Dans tout ce chapitre, l'espace $E = \mathbb{R}^n$ est muni de sa base canonique

$$\mathcal{E} = (e_1, e_2, \dots, e_n)$$

et identifié à l'espace $\mathfrak{M}_{n,1}(\mathbb{R})$ des matrices colonnes ; le contexte permet de trancher. Il est muni de son produit scalaire canonique défini par

$$\langle x | y \rangle = {}^t x\, y = \sum_{k=1}^{n} x_k\, y_k$$

en écrivant (abusivement)

$$x = \sum_{k=1}^{n} x_k\, e_k = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Rappelons qu'alors, pour tous vecteurs x et y, et pour toute matrice $A \in \mathfrak{M}_n(\mathbb{R})$, on a

$$\langle x | Ay \rangle = \langle {}^t A x | y \rangle. \tag{15.1}$$

15.1 PROBABILITÉS SUR $\mathbb{R}^N$

§229 **Boréliens de $\mathbb{R}^n$**

On appelle **tribu des boréliens de $\mathbb{R}^n$** la tribu, notée $\mathfrak{B}(\mathbb{R}^n)$ ou $\mathfrak{B}^n$, engendrée par les ouverts de $\mathbb{R}^n$. Comme dans le cas des boréliens de $\mathbb{R}$, on peut montrer (A66) que cette tribu est engendrée, indifféremment :

— par les « rectangles » de la forme

$$\Delta(a_1, a_2, \ldots, a_n) =]-\infty\,;\, a_1] \times]-\infty\,;\, a_2] \times \cdots \times]-\infty\,;\, a_n] \;;$$

— par les rectangles $I_1 \times I_2 \times \cdots \times I_n$, où les I_k sont des intervalles de $\mathbb{R}$;

— par les rectangles boréliens $B_1 \times B_2 \times \cdots \times B_n$, où les B_k sont des boréliens de $\mathbb{R}$.

§230 **Probabilité produit**

Soient $(\Omega_1, \mathfrak{T}_1, \mathbf{P}_1)$ et $(\Omega_2, \mathfrak{T}_2, \mathbf{P}_2)$ deux espaces probabilisés. On note $\mathfrak{T}_1 \otimes \mathfrak{T}_2$ la tribu engendrée par l'ensemble des « rectangles » de la forme $B_1 \times B_2$, où $B_1 \in \mathfrak{T}_1$ et $B_2 \in \mathfrak{T}_2$. Sur cette tribu, appelée **tribu produit**, on peut définir une probabilité, notée $\mathbf{P}_1 \otimes \mathbf{P}_2$, appelée **probabilité produit**, en demandant simplement que

$$\mathbf{P}_1 \otimes \mathbf{P}_2(B_1 \times B_2) = \mathbf{P}_1(B_1)\,\mathbf{P}_2(B_2) \tag{15.2}$$

pour tous les événements B_1 et B_2. On l'étend par additivité sur l'algèbre engendrée par les rectangles, on vérifie qu'elle est σ-additive sur cette algèbre et on peut appliquer le théorème de Carathéodory. Ainsi, la relation (15.2) suffit à caractériser entièrement la probabilité $\mathbf{P}_1 \otimes \mathbf{P}_2$. Cette construction se généralise au produit $\mathbf{P}_1 \otimes \mathbf{P}_2 \otimes \cdots \otimes \mathbf{P}_n$ de n probabilités.

Le produit de deux espaces probabilisés permet de modéliser le résultat de deux expériences indépendantes.

Remarque 15.1 On construit de la même façon la mesure de Lebesgue sur les boréliens de $\mathbb{R}^n$ à partir de la mesure de Lebesgue sur $\mathbb{R}$

§231 **Densités**

On dit qu'une probabilité $\mathbf{P}$ sur $(\mathbb{R}^n, \mathfrak{B}^n)$ admet une densité $f : \mathbb{R}^n \to \mathbb{R}$ si f est une fonction mesurable (on dit aussi : **borélienne**) vérifiant

$$\mathbf{P}(A) = \int \cdots \int_A f(\boldsymbol{x})\, \mathrm{d}^n\boldsymbol{x} = \int \cdots \int f(\boldsymbol{x})\, \mathbf{1}_A(\boldsymbol{x})\, \mathrm{d}^n\boldsymbol{x}$$

pour tout borélien $A \in \mathfrak{B}^n$. Cette densité f est uniquement déterminée à un ensemble de mesure nulle près.

Exemple 15.2 Si $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ sont des scalaires strictement positifs, la fonction f définie par

$$f(x_1, x_2, \ldots, x_n) = \prod_{k=1}^{n} \frac{\mathrm{e}^{-x_k^2/2\lambda_k}}{\sqrt{2\pi\,\lambda_k}}$$

définit une probabilité à densité.

15.2 VECTEURS ALÉATOIRES

On appelle **vecteur aléatoire** à valeurs dans $\mathbb{R}^n$ toute application $\mathbf{X} : \Omega \to \mathbb{R}^n$ mesurable (pour la tribu $\mathfrak{B}^n$ à l'arrivée et la tribu $\mathfrak{T}$ au départ). Si l'on note les composantes de $\mathbf{X}$ sous la forme

$$\mathbf{X} = (X_1, X_2, \ldots, X_n)\,,$$

alors les applications $X_1, X_2, \ldots, X_n$ sont des variables aléatoires réelles, appelées **marginales de X**. (Le fait que ce soient des variables aléatoires découle de ce que $X_k = \pi_k \circ \mathbf{X}$, où π_k est la k-ième forme coordonnée de $\mathbb{R}^n$, application continue et donc mesurable.)

Espérance § 232

Soit $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ un vecteur aléatoire. Si chacune des variables aléatoires X_k admet une espérance $m_k = \mathbf{E}(X_k)$, on définit l'**espérance** de $\mathbf{X}$ comme le vecteur $\boldsymbol{m}$ de $\mathbb{R}^n$ suivant :

$$\boldsymbol{m} = \mathbf{E}(\mathbf{X}) = \begin{pmatrix} \mathbf{E}(X_1) \\ \mathbf{E}(X_2) \\ \vdots \\ \mathbf{E}(X_n) \end{pmatrix} = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{pmatrix}.$$

L'espérance est obtenue par le même processus d'intégration, coordonnée par coordonnée, que pour une variable aléatoire réelle. On notera ainsi

$$\mathbf{E}(\mathbf{X}) = \int_\Omega \mathbf{X}\, d\mathbf{P}.$$

Matrice de covariance § 233

Pour tout couple (i, j) de $[\![1\,;\,n]\!]^2$, notons

$$C_{i,j} = \mathbf{Cov}(X_i, X_j) = \mathbf{Cov}\left(\langle e_i|\mathbf{X}\rangle, \langle e_j|\mathbf{X}\rangle\right).$$

La matrice $C^{\mathbf{X}} = (C_{ij})_{i,j}$ est appelée **matrice de covariance** de $\mathbf{X}$. Ses coefficients diagonaux sont des variances $C_{ii} = \mathbf{V}(X_i)$:

$$C^{\mathbf{X}} = \begin{pmatrix} \mathbf{V}(X_1) & \mathbf{Cov}(X_1, X_2) & & \mathbf{Cov}(X_1, X_n) \\ \mathbf{Cov}(X_2, X_1) & \mathbf{V}(X_2) & & \vdots \\ & & & \mathbf{Cov}(X_{n-1}, X_n) \\ \mathbf{Cov}(X_n, X_1) & & \mathbf{Cov}(X_n, X_{n-1}) & \mathbf{V}(X_n). \end{pmatrix}.$$

Si l'on note σ_i l'écart-type de X_i et $\rho_{ij} = \mathbf{Cov}(X_i, X_j)/\sigma_i\,\sigma_j$ le coefficient de corrélation de (X_i, X_j), cette matrice s'écrit également

$$C^{\mathbf{X}} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\,\sigma_1\sigma_2 & & \rho_{1n}\,\sigma_1\sigma_n \\ \rho_{21}\,\sigma_2\sigma_1 & \sigma_2^2 & & \vdots \\ & & & \rho_{n-1,n}\,\sigma_{n-1}\sigma_n \\ \rho_{n1}\,\sigma_n\sigma_1 & & \rho_{n,n-1}\,\sigma_n\sigma_{n-1} & \sigma_n^2 \end{pmatrix}.$$

Cette matrice s'écrit également comme une espérance (de matrice). En effet, si l'on note $\mathbf{X}' := \mathbf{X} - \boldsymbol{m}$, alors :

$$\begin{aligned} C^{\mathbf{X}} &= \mathbf{E}\left(\mathbf{X}' \cdot {}^t\mathbf{X}'\right) \\ &= \mathbf{E}(\mathbf{X} \cdot {}^t\mathbf{X}) - \boldsymbol{m} \cdot {}^t\boldsymbol{m}. \end{aligned}$$

■ **Théorème 15.3** *La matrice de covariance* $\mathrm{C}^{\mathbf{X}}$ *est symétrique positive :* ${}^{\mathrm{t}}\mathbf{\Lambda}\mathrm{C}\mathbf{\Lambda} \geqslant 0$ *pour tout vecteur* $\mathbf{\Lambda} \in \mathbb{R}^n$.

DÉMONSTRATION : Si A est une matrice carrée, alors $\mathrm{A}\,{}^{\mathrm{t}}\mathrm{A}$ est toujours une matrice symétrique positive (A61), ce qui démontre le théorème. On peut également procéder par un calcul élémentaire : si $\mathbf{\Lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n)$ un vecteur de $\mathbb{R}^n$, alors en notant $\mathrm{Y} := \sum_{i=1}^n \lambda_i \mathrm{X}_i = \langle \mathbf{\Lambda} | \mathbf{X} \rangle$,

$$ {}^{\mathrm{t}}\mathbf{\Lambda}\mathrm{C}\mathbf{\Lambda} = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \mathrm{C}_{ij} = \mathbf{Cov}\left(\sum_{i=1}^{n} \lambda_i \mathrm{X}_i, \sum_{j=1}^{n} \lambda_j \mathrm{X}_j \right) = \mathbf{V}(\mathrm{Y}) \geqslant 0. $$ ∎

Remarquons que connaître les coefficients de $\mathrm{C}^{\mathbf{X}}$ est équivalent à connaître les quantités ${}^{\mathrm{t}}\boldsymbol{y}\mathrm{C}^{\mathbf{X}}\boldsymbol{x}$ pour $\boldsymbol{x}, \boldsymbol{y}$ parcourant $\mathbb{R}^n$: dans un sens, on a $\mathrm{C}_{ij} = {}^{\mathrm{t}}\boldsymbol{e}_i \mathrm{C}^{\mathbf{X}} \boldsymbol{e}_j = \langle \boldsymbol{e}_i | \mathrm{C}^{\mathbf{X}} \boldsymbol{e}_j \rangle$ et dans l'autre,

$$ {}^{\mathrm{t}}\boldsymbol{x}\mathrm{C}^{\mathbf{X}}\boldsymbol{y} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i\, y_j\, \mathrm{C}_{ij}. \tag{15.3} $$

■ **Proposition 15.4** *Soit* $\mathbf{X}$ *une variable aléatoire à valeurs dans* $\mathbb{R}^n$. *Alors sa matrice de covariance est caractérisée par la relation*

$$ \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \qquad {}^{\mathrm{t}}\boldsymbol{x}\mathrm{C}^{\mathbf{X}}\boldsymbol{y} = \mathbf{Cov}\left(\langle \boldsymbol{x} | \mathbf{X} \rangle , \langle \boldsymbol{y} | \mathbf{X} \rangle \right). \tag{15.4} $$

DÉMONSTRATION : Il suffit de développer le terme de droite pour obtenir l'expression (15.3). ∎

■ **Théorème 15.5 (Changement de base)** *Soit* $\mathrm{A} \in \mathfrak{M}_{m,n}(\mathbb{R})$ *une matrice de taille* $m \times n$ *et soit* $\boldsymbol{b}$ *un vecteur de* $\mathbb{R}^m$ *; la variable aléatoire* $\mathbf{Z} = \mathrm{A}\mathbf{X} + \boldsymbol{b}$, *à valeurs dans* $\mathbb{R}^m$, *a pour matrice de covariance* $\mathrm{C}^{\mathbf{Z}} = \mathrm{A}\,\mathrm{C}^{\mathbf{X}}\,{}^{\mathrm{t}}\mathrm{A}$.

DÉMONSTRATION : L'effet d'une translation étant nul sur un calcul de covariance, on a déjà $\mathrm{C}^{\mathbf{Z}} = \mathrm{C}^{\mathrm{A}\mathbf{X}}$. Ensuite, pour tous vecteurs $\boldsymbol{x}, \boldsymbol{y}$ de $\mathbb{R}^m$, on a

$$ \begin{aligned} {}^{\mathrm{t}}\boldsymbol{x}\mathrm{C}^{\mathrm{A}\mathbf{X}}\boldsymbol{y} &= {}^{\mathrm{t}}\boldsymbol{x}\,\mathbf{E}(\mathrm{A}\mathbf{X}'\,{}^{\mathrm{t}}\mathbf{X}'{}^{\mathrm{t}}\mathrm{A})\,\boldsymbol{y} = {}^{\mathrm{t}}\boldsymbol{x}\mathrm{A}\,\mathbf{E}(\mathbf{X}'\,{}^{\mathrm{t}}\mathbf{X}')\,{}^{\mathrm{t}}\mathrm{A}\boldsymbol{y} \qquad \textit{par linéarité} \\ &= {}^{\mathrm{t}}\boldsymbol{x}\,(\mathrm{A}\,\mathrm{C}^{\mathbf{X}}\,{}^{\mathrm{t}}\mathrm{A})\,\boldsymbol{y}, \end{aligned} $$

ce qui prouve que $\mathrm{C}^{\mathrm{A}\mathbf{X}} = \mathrm{A}\,\mathrm{C}^{\mathbf{X}}\,{}^{\mathrm{t}}\mathrm{A}$. ∎

■ **Théorème 15.6 (Variance d'une combinaison linéaire)** *Soit* $\mathbf{X}$ *un vecteur aléatoire de matrice de covariance* $\mathrm{C}^{\mathbf{X}} = (\mathrm{C}_{ij})_{ij}$ *; soit* $\boldsymbol{a}$ *un vecteur de* $\mathbb{R}^n$. *Alors*

$$ \mathbf{V}\left(\langle \boldsymbol{a} | \mathbf{X} \rangle \right) = \mathbf{V}\left(\sum_i a_i \mathrm{X}_i \right) = \sum_{i,j} a_i a_j \mathrm{C}_{ij} = \langle \boldsymbol{a} | \mathrm{C}^{\mathbf{X}} \boldsymbol{a} \rangle. $$

DÉMONSTRATION : Directement avec la formule (15.4) (et $\boldsymbol{x} = \boldsymbol{y} = \boldsymbol{a}$) ou, par bilinéarité,

$$ \mathbf{V}\left(\sum_i a_i \mathrm{X}_i \right) = \sum_{i,j} a_i\, a_i\, \mathbf{Cov}(\mathrm{X}_i, \mathrm{X}_j) = \sum_{i,j} a_i\, a_j\, \mathrm{C}_{ij} = \langle \boldsymbol{a} | \mathrm{C}^{\mathbf{X}}\, \boldsymbol{a} \rangle. $$ ∎

§ 234 **Inégalité de Cauchy-Schwarz**

Soient $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_n$ des variables aléatoires admettant des moments d'ordre 2. Notons R la matrice de coefficients $\mathrm{R}_{ij} = \mathbf{E}(\mathrm{X}_i\,\mathrm{X}_j)$. Si on l'interprète en terme de produit scalaire dans l'espace L^2, il s'agit tout simplement de la matrice de Gram des n vecteurs $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_n$, et donc :

■ **Proposition 15.7** *La matrice* R *est symétrique positive. Elle est définie positive si et seulement si les variables* $\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_n$ *sont linéairement indépendantes.*

Ce résultat généralise l'inégalité de Cauchy-Schwarz ; en effet, si l'on ne garde que les deux lignes i et j de la matrice R, et les deux colonnes i et j, on obtient la matrice 2×2

$$\begin{pmatrix} \mathbf{E}(X_i^2) & \mathbf{E}(X_i X_j) \\ \mathbf{E}(X_i X_j) & \mathbf{E}(X_j^2) \end{pmatrix}$$

qui, comme matrice extraite d'une matrice symétrique positive, est elle-même symétrique positive. Son déterminant est donc positif, ce qui s'écrit

$$\mathbf{E}(X_i X_j)^2 \leqslant \mathbf{E}(X_i^2)\,\mathbf{E}(X_j^2),$$

qui est bien l'inégalité de Cauchy-Schwarz classique.

15.3 FUBINI !!!

Le théorème de Fubini en pratique § 235

Soient $\mathbf{P}$ et $\mathbf{Q}$ des probabilités sur $\mathbb{R}$. Si $f : \mathbb{R}^2 \to \mathbb{R}$ est mesurable et si elle est *positive*, ou *intégrable* sur $\mathbb{R}^2$ muni de la probabilité produit $\mathbf{P} \otimes \mathbf{Q}$, alors les fonctions

$$x \longmapsto \int_{-\infty}^{+\infty} f(x,y)\,\mathbf{Q}(\mathrm{d}y) \qquad \text{et} \qquad y \longmapsto \int_{-\infty}^{+\infty} f(x,y)\,\mathbf{P}(\mathrm{d}x)$$

sont définies $\mathbf{P}$-presque partout et $\mathbf{Q}$-presque partout respectivement, mesurables et

$$\begin{aligned} \iint_{\mathbb{R}^2} f(x,y)\,(\mathrm{P}\otimes\mathrm{Q})(\mathrm{d}x,\mathrm{d}y) &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x,y)\,\mathbf{Q}(\mathrm{d}y) \right) \mathbf{P}(\mathrm{d}x) \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x,y)\,\mathbf{P}(\mathrm{d}x) \right) \mathbf{Q}(\mathrm{d}y). \qquad (15.5) \end{aligned}$$

En pratique, on retiendra que :

- si f est positive et que l'une des deux dernières intégrales est finie, alors les deux autres intégrales sont finies et lui sont égales ;
- si f n'est pas positive, on cherche si l'une des deux intégrales

$$\int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} |f(x,y)|\,\mathbf{Q}(\mathrm{d}y) \right) \mathbf{P}(\mathrm{d}x) \quad \text{ou} \quad \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} |f(x,y)|\,\mathbf{P}(\mathrm{d}x) \right) \mathbf{Q}(\mathrm{d}y)$$

est finie ; dans ce cas, les trois intégrales de l'équation (15.5) existent et sont égales.

Remarque 15.8 Le théorème de Fubini se généralise à des fonctions définies sur le produit $\mathrm{E} \times \mathrm{F}$ de deux espaces mesurés, par exemple $\mathbb{R} \times \Omega$ où $\mathbb{R}$ est muni de la tribu des boréliens et de la mesure de Lebesgue, et Ω un espace muni d'une tribu $\mathfrak{T}$ et d'une probabilité $\mathbf{P}$.

Application : densités marginales § 236

Soit $\mathbf{X} = (\mathrm{X}, \mathrm{Y})$ un vecteur aléatoire (nous traitons le cas $n = 2$ seulement, le cas général étant tout à fait similaire). On suppose qu'il admet une densité f, c'est-à-dire que sa loi $\mathbf{P}_{\mathbf{X}}$, qui est une probabilité sur $\mathbb{R}^2$, admet f pour densité :

$$\mathbf{P}\{\mathbf{X} \in \mathrm{A}\} = \mathbf{P}_{\mathbf{X}}(\mathrm{A}) = \iint_{\mathrm{A}} f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

pour tout borélien A de $\mathbb{R}^2$.

Pour déterminer la loi de la marginale X, fixons un borélien $A \in \mathfrak{B}$. Par positivité de f et en appliquant le théorème de Fubini :

$$\mathbf{P}\{X \in A\} = \mathbf{P}\{\mathbf{X} \in A \times \mathbb{R}\} = \iint_{A\times\mathbb{R}} f(x,y)\,dx\,dy = \int_A \left(\int_{-\infty}^{+\infty} f(x,y)\,dy\right) dx.$$

Ceci étant vrai pour tout borélien A, la marginale X admet pour densité

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy.$$

(Cette fonction peut éventuellement prendre la valeur $+\infty$ sur un ensemble de mesure nulle.)

§ 237 **Application : expression alternative de l'espérance**

Nous pouvons reprendre la discussion du paragraphe § 201, page 300 et montrer l'équivalence des deux points de vue. Dans le cas d'une variable positive, il s'agit donc de montrer le théorème suivant :

■ **Théorème 15.9** *Si* X *est une variable aléatoire positive, alors*

$$\mathbf{E}(X) = \int_0^{+\infty} \mathbf{P}\{X > x\}\,dx = \int_0^{+\infty} \big[1 - F(x)\big]\,dx.$$

DÉMONSTRATION : On applique une version du théorème de Fubini où l'espace de départ est $\Omega \times \mathbb{R}$, muni de la tribu $\mathfrak{T} \otimes \mathfrak{B}$ et de la mesure $\mathbf{P} \otimes \mu$, où μ désigne la mesure de Lebesgue sur $\mathfrak{B}$.

Notons alors

$$A := \{(\omega, x)\ ;\ 0 \leqslant x \leqslant X(\omega)\}$$

qui est l'ensemble des « points sous le graphe de X », et choisissons comme fonction $f = \mathbf{1}_A$, fonction constante égale à 1 sur A et 0 ailleurs.

Les deux fonctions du théorème de Fubini sont respectivement

$$\omega \longmapsto \int_{-\infty}^{+\infty} \mathbf{1}_A(\omega, x)\,dx = \int_0^{X(\omega)} dx = X(\omega)$$

et

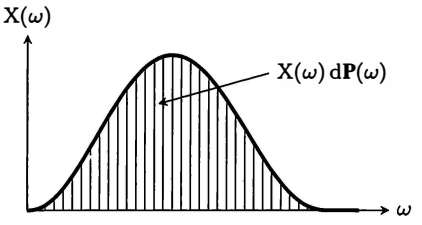
$$x \longmapsto \int_\Omega \mathbf{1}_A(\omega, x)\,d\mathbf{P}(\omega) = \begin{cases} \mathbf{P}\{X \geqslant x\} & \text{si } x \geqslant 0 \\ 0 & \text{sinon.} \end{cases}$$

Le théorème de Fubini donne alors

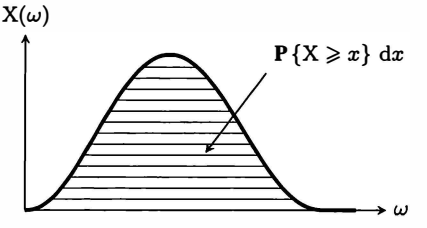
$$\mathbf{E}(X) = \int_\Omega X(\omega)\,d\mathbf{P}(\omega) = \int_0^{+\infty} \mathbf{P}\{X \geqslant x\}\,dx.$$

Une démonstratin identique avec $A' := \{(\omega, x)\ ;\ 0 \leqslant x < X(\omega)\}$ mène à la formule annoncée. ∎

Les deux intégrales précédentes peuvent être représentées symboliquement ainsi :



« $\mathbf{E}(X) = \int_\Omega X(\omega)\,d\mathbf{P}(\omega)$ »



« $\mathbf{E}(X) = \int_0^{+\infty} \mathbf{P}\{X \geqslant x\}\,dx$ »

15.4 CHANGEMENTS DE VARIABLES ALÉATOIRES

Changement de vecteur aléatoire

§ 238

Considérons un domaine ouvert D de $\mathbb{R}^n$, et $\varphi : \mathrm{D} \to \Delta$ un $\mathscr{C}^1$-difféomorphisme de D sur Δ. On rappelle qu'une telle application est caractérisée par le fait qu'elle est de classe $\mathscr{C}^1$, bijective et que son jacobien ne s'annule pas sur D. Le théorème de changement de variables A.33, page 686 permet d'obtenir le résultat suivant :

■ **Théorème 15.10** *Soit* $\mathbf{X} = (\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n)$ *un vecteur aléatoire admettant une densité* f. *Soit* $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ *un* $\mathscr{C}^1$*-difféomorphisme de* $\mathbb{R}^n$ *sur* $\Delta = \varphi(\mathbb{R}^n)$. *Si l'on pose* $\mathbf{Y} = \varphi(\mathbf{X})$, *alors* $\mathbf{Y}$ *est absolument continue et admet pour densité*

$$g(\boldsymbol{y}) = f\big(\varphi^{-1}(\boldsymbol{y})\big) \cdot \left|\frac{\mathrm{D}\boldsymbol{x}}{\mathrm{D}\boldsymbol{y}}\right| \cdot \mathbf{1}_\Delta(\boldsymbol{y}),$$

où l'on a noté $\mathrm{D}\boldsymbol{x}/\mathrm{D}\boldsymbol{y}$ *le jacobien de* φ^{-1}.

DÉMONSTRATION : Pour tout borélien $\mathrm{B} \subset \Delta$, notons $\mathrm{A} = \varphi^{-1}(\mathrm{B})$, de telle sorte que l'on ait $\{\mathrm{X} \in \mathrm{A}\} = \{\mathrm{Y} \in \mathrm{B}\}$; le théorème de changement de variables montre alors que

$$\mathbf{P}\{\mathrm{Y} \in \mathrm{B}\} = \mathbf{P}\{\mathrm{X} \in \mathrm{A}\} = \int_{\mathrm{A}} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{\mathrm{B}} f(\varphi^{-1}(\boldsymbol{y})) \left|\frac{\mathrm{D}\boldsymbol{x}}{\mathrm{D}\boldsymbol{y}}\right| \mathrm{d}\boldsymbol{y}.$$

Par ailleurs, si $\mathrm{B} \subset \Delta^{\mathrm{c}}$, on immédiatement $\mathbf{P}\{\mathrm{Y} \in \mathrm{B}\} = 0$. Au total, si B est un borélien, on le sépare en $\mathrm{B} = (\mathrm{B} \cap \Delta) + (\mathrm{B} \cap \Delta^{\mathrm{c}})$, union disjointe de deux boréliens[1], et on obtient

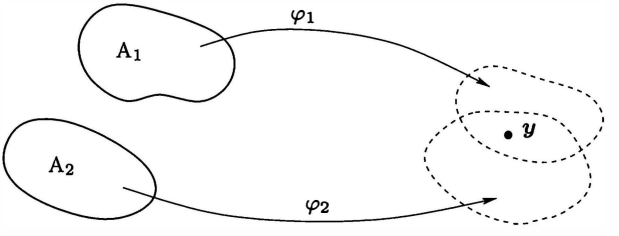
$$\begin{aligned}\mathbf{P}\{\mathrm{Y} \in \mathrm{B}\} = \mathbf{P}\{\mathrm{Y} \in (\mathrm{B} \cap \Delta)\} &= \int_{\mathrm{B}\cap\Delta} f(\varphi^{-1}(\boldsymbol{y})) \left|\frac{\mathrm{D}\boldsymbol{x}}{\mathrm{D}\boldsymbol{y}}\right| \mathrm{d}\boldsymbol{y} \\ &= \int_{\mathrm{B}} f(\varphi^{-1}(\boldsymbol{y})) \left|\frac{\mathrm{D}\boldsymbol{x}}{\mathrm{D}\boldsymbol{y}}\right| \mathbf{1}_\Delta(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}.\end{aligned}$$

Ceci étant vrai pour tout borélien B, le résultat s'ensuit. ∎

Il arrive que le changement de variable ne soit pas injectif, et que plusieurs portions du plan puissent être envoyées sur le même domaine. Le théorème précédent se généralise ainsi.

■ **Théorème 15.11** *Soit* $\mathbf{X} = (\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n)$ *un vecteur aléatoire admettant une densité* f. *Soient* $\mathrm{A}_1, \ldots, \mathrm{A}_p$ *des ouverts disjoints du plan. On suppose que* f *est identiquement nulle en dehors de ces ouverts. Soit* $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ *une application définie sur* $\mathbb{R}^n$, *telle que la restriction* φ_i *de* φ *à chacun des ouverts* A_i *soit un* $\mathscr{C}^1$*-difféomorphisme de* A_i *vers* $\varphi(\mathrm{A}_i)$. *Alors la variable* $\mathbf{Y} = \varphi(\mathbf{X})$ *admet pour densité*

$$g(\boldsymbol{y}) = \sum_{i=1}^{p} f\big(\varphi_i^{-1}(\boldsymbol{y})\big) \cdot \big|\mathrm{J}_{\varphi_i^{-1}}(\boldsymbol{y})\big| \cdot \mathbf{1}_{\varphi_i(\mathrm{A}_i)}(\boldsymbol{y}).$$



⚠15.3
⚠15.4

(1). Rappelons que Δ est ouvert, donc borélien.

§ 239 **Somme, produit et quotient de deux variables aléatoires**

Si le couple (X, Y) admet une densité conjointe f, on rappelle que l'on a montré (§ 178) que la variable $S = X + Y$ admet une densité f_S donnée, pour presque tout z, par

$$f_S(z) = \int_{-\infty}^{+\infty} f(z - s, s)\, ds.$$

On aimerait déduire du théorème 15.10 une densité de la variable $Z := XY$. Comme l'application $(x, y) \mapsto xy$ n'est certainement pas un $\mathscr{C}^1$-difféomorphisme[2], on utilise une petite astuce et on calcule plutôt la densité g du couple $(V, Z) := (X, XY)$. Une fois celle-ci déterminée, la densité de $Z = XY$ est la densité marginale du couple (V, Z) et se calcule par simple intégration. Notons $\varphi : (x, y) \mapsto (x, xy)$, application de $\mathbb{R}^2$ dans $\mathbb{R}^2$ dont le jacobien vaut

$$J_\varphi(x, y) = \begin{pmatrix} 1 & 0 \\ y & x \end{pmatrix} = x.$$

Notons D^+ le demi-plan $x > 0$ et D^- le demi-plan $x < 0$. L'application φ établit une bijection de $D^+ \cup D^-$ sur lui même, de bijection réciproque

$$(v, z) \longmapsto \left(v, \frac{z}{v}\right),$$

de jacobien $J_{\varphi^{-1}}(v, z) = 1/J_\varphi(x, y) = 1/v$. On est alors dans un cas courant : φ est « presque » un $\mathscr{C}^1$-difféomorphisme, son jacobien ne s'annulant que sur l'axe des abscisses, de mesure nulle. La fonction φ n'est donc un difféomorphisme que par morceaux. Dans ces cas-là, on applique le théorème 15.11 ; chaque point $\boldsymbol{y} = (v, z)$ de $D^+ \cup D^-$ admet un unique antécédent et on peut calculer en ces points la densité du couple image :

$$g(v, z) = f\left(v, \frac{z}{v}\right) \frac{1}{|v|}.$$

Enfin, par intégration, on en déduit la densité marginale de $Z = XY$:

$$f_Z(z) = \int_{-\infty}^{+\infty} f\left(v, \frac{z}{v}\right) \frac{dv}{|v|}.$$

Remarque 15.12 Si les variables X et Y sont indépendantes, alors $f = f_X \otimes f_Y$ et la formule précédente se réduit à

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(v)\, f_Y\left(\frac{z}{v}\right) \frac{dv}{|v|}.$$

De la même manière, et à titre d'exercice, on pourra établir qu'une densité de $Q := X/Y$ est

$$f_Q(q) = \int_{-\infty}^{+\infty} f(w, wq)\, |w|\, dw.$$

(2). Un $\mathscr{C}^1$-difféomorphisme ne peut exister qu'entre des ouverts de deux espaces de même dimension.

EXERCICES

♦ ***Exercice 15.1*** Soient X et Y deux variables aléatoires centrées et réduites ; notons ρ leur coefficient de corrélation, et posons $Z = X - \rho Y$. Montrer que Y et Z ne sont pas corrélées.

♦ ***Exercice 15.2 (Paradoxe téléphonique)*** Deux personnes entrent au même instant dans deux cabines téléphoniques [3] voisines ; on suppose que leurs temps de communication, notés X et Y, sont des variables aléatoires indépendantes, de même loi exponentielle de paramètre $\alpha > 0$.

i) On note T la variable aléatoire égale au temps d'attente avant qu'une des deux cabines se libère. Déterminer la loi de T et calculer son espérance.

ii) La première personne à terminer sa conversation attend la seconde pendant une durée notée U. Déterminer la loi de la variable U ; comparer alors l'attente moyenne U à la durée moyenne de communication $1/\alpha$.

♦ ***Exercice 15.3 (Statistique d'ordre)*** Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires réelles indépendantes. On note $X_{(1)} \leqslant X_{(2)} \leqslant \ldots, \leqslant X_{(n)}$ les valeurs *ordonnées* de ces variables, c'est-à-dire que $X_{(1)}$ est la plus petite des valeurs $X_1, X_2, \ldots, X_n$, $X_{(2)}$ la seconde plus petite, etc.

On suppose que les X_k admettent tous la même densité f. Montrer que $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ admet pour densité

$$g(y_1, y_2, \ldots, y_n) = n!\, f(y_1)\, f(y_2) \cdots f(y_n)\, \mathbf{1}_{(y_1 < y_2 < \cdots < y_n)}.$$

♦ ***Exercice 15.4 (Statistiques d'ordre)*** On reprend les hypothèses et les notations de l'exercice 15.3. On note F la fonction de répartition commune des X_k, et f leur densité.

i) Montrer que $X_{(n)}$ a pour densité $f_n = n\, F^{n-1}\, f$.

ii) Montrer que $X_{(1)}$ a pour densité $f_1 = n\,(1 - F)^{n-1}\, f$.

iii) Si $1 \leqslant k \leqslant r$, montrer que $X_{(k)}$ a pour densité

$$f_k = \frac{n!}{(k-1)!\,(n-k)!}\, F^{k-1}\,(1 - F)^{n-k}\, f.$$

♦ ***Exercice 15.5 (Inversibilité d'une matrice « aléatoire »)*** Soit M une matrice carrée d'ordre 2 :

$$M = \begin{pmatrix} A & C \\ B & D \end{pmatrix}$$

dont les coefficients A, B, C et D sont des variables aléatoires absolument continues et indépendantes. Montrer que M est presque sûrement inversible. (Cet exercice se généralise bien évidemment à des matrices de taille $n \times n$.)

Indications

◊ *Indication 15.3 :* On emploiera le théorème 15.11.

◊ *Indication 15.4 :* L'événement $\{X_{(k)} \leqslant t\}$ est caractérisé par la propriété : au moins k parmi les n variables $X_1, X_2, \ldots, X_n$ sont inférieures ou égales à t.

[3]. Pour nos lecteurs les plus jeunes : téléphone géant pouvant servir d'abri en cas de gros temps.

SOLUTIONS

♦ **Solution 15.1** $\mathbf{Cov}(Y,Z) = \mathbf{Cov}(Y, X-\rho Y) = \underbrace{\mathbf{Cov}(X,Y)}_{=\rho} - \rho \underbrace{\mathbf{Cov}(Y,Y)}_{=1} = 0.$

♦ **Solution 15.2**

i) T est défini comme $\min(X,Y)$. On a donc, pour tout $t > 0$:

$$\mathbf{P}\{T \geqslant t\} = \mathbf{P}\{X \geqslant t, Y \geqslant t\} = \mathbf{P}\{X \geqslant t\} \cdot \mathbf{P}\{Y \geqslant t\} = e^{-2\alpha t},$$

ce qui prouve que $F_T(t) = 1 - e^{2\alpha t}$ pour tout $t > 0$ et donc que $T \rightsquigarrow \mathscr{E}(2\alpha)$. D'après les tables des lois usuelles, on a donc $\mathbf{E}(T) = 1/2\alpha$: le temps d'attente moyen pour qu'une cabine se libère est la moitié du temps moyen de communication.

ii) On a $U = |X - Y|$, donc $U(\Omega) = \mathbb{R}^+$. Soit $u \geqslant 0$.

Le couple (X,Y) admet, par indépendance, une densité $f(x,y) = \alpha^2 e^{-\alpha(x+y)}$ sur $\mathbb{R}^{+2}$, nulle ailleurs. Ainsi, par symétrie,

$$F_U(u) = \mathbf{P}\{U \leqslant u\} = 2\alpha^2 \int_0^{+\infty} e^{-\alpha x} \left(\int_x^{x+u} e^{-\alpha y}\, dy \right) dx = 1 - e^{-\alpha u}.$$

En d'autres termes, $U \rightsquigarrow \mathscr{E}(\alpha)$. Notamment, et cela peut paraître paradoxal, le temps moyen d'attente du premier sorti est égal à la durée moyenne de communication !

♦ **Solution 15.3** Notons d'abord qu'une densité conjointe de $(X_1, X_2, \dots, X_n)$ est

$$f(x_1)\, f(x_2) \cdots f(x_n).$$

L'application φ qui à un n-uplet $(x_1, x_2, \dots, x_n)$ de réels distincts associe le n-uplet ordonné $(y_1, y_2, \dots, y_n)$ n'est évidemment pas une bijection (si l'on échange x_1 et x_2, l'image reste évidemment la même). Si en revanche l'on se fixe une permutation σ de $\{1, 2, \dots, n\}$ et que l'on considère l'ouvert

$$U_\sigma = \{(x_1, x_2, \dots, x_n)\ ;\ x_{\sigma(1)} < x_{\sigma(2)} < \cdots < x_{\sigma(n)}\},$$

alors la restriction de φ à U_σ est bijective et agit ainsi :

$$(x_1, x_2, \dots, x_n) \longmapsto (y_1, y_2, \dots, y_n) = (x_{\sigma(1)}, x_{\sigma(2)}, \cdots x_{\sigma(n)}).$$

C'est évidemment un $\mathscr{C}^1$-difféomorphisme de Jacobien égal à la signature de σ, donc ± 1. De plus, il y a $n!$ ouverts de ce type, chacun d'eux étant envoyé sur l'ouvert U_{Id}. Une densité de $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ est alors

$$\begin{aligned} g(y_1, y_2, \dots, y_n) &= \sum_{\sigma \in \mathfrak{S}_n} f(x_1)\, f(x_2) \cdots f(x_n) \cdot \mathbf{1}_{U_{\mathrm{Id}}}(y) \\ &= \sum_{\sigma \in \mathfrak{S}_n} f(x_{\sigma(1)})\, f(x_{\sigma(2)}) \cdots f(x_{\sigma(n)}) \cdot \mathbf{1}_{U_{\mathrm{Id}}}(y) \qquad \textit{par réarrangement} \\ &= \sum_{\sigma \in \mathfrak{S}_n} f(y_1)\, f(y_2) \cdots f(y_n) \cdot \mathbf{1}_{U_{\mathrm{Id}}}(y) = n!\, f(y_1)\, f(y_2) \cdots f(y_n) \cdot \mathbf{1}_{U_{\mathrm{Id}}}(y), \end{aligned}$$

ce qui est la formule voulue.

♦ **Solution 15.5** L'événement « M est non inversible » s'écrit également $\{AD - BC = 0\}$. Le quadruplet (A, B, C, D) admet, par indépendance des variables, une densité conjointe

$$f(a,b,c,d) = f_A(a) \cdot f_B(b) \cdot f_C(c) \cdot f_D(d).$$

Or, l'ensemble $Z := \{(a,b,c,d) \in \mathbb{R}^4\ ;\ ad - bc = 0\}$, une fois privé de l'origine, est une hypersurface de $\mathbb{R}^4$ (une variété différentiable de dimension 3), de mesure nulle, donc

$$\mathbf{P}\{AD - BC = 0\} = \int_{\mathbb{R}^4} \mathbf{1}_Z(a,b,c,d)\, f(a,b,c,d)\, da\, db\, dc\, dd = 0.$$

Chapitre 16

Indépendance (2/2)

La notion d'indépendance d'une famille d'événements permet de définir, de manière très générale, celle d'indépendance d'une famille de variables aléatoires quelconques. Des outils théoriques et des méthodes pratiques sont exposés ici concernant cette notion, avec des applications à la somme et au produit de variables aléatoires indépendantes.

Enfin, nous terminons ce chapitre par une introduction à la notion de temps d'arrêt, *permettant de généraliser l'identité de Wald, déjà rencontrée au chapitre 8.*

16.1 INDÉPENDANCE DE VARIABLES ALÉATOIRES

Indépendance de deux variables aléatoires § 240

On dit que deux variables aléatoires X et Y sont **indépendantes** si, pour tous événements A et B,

$$\mathbf{P}\{X \in A, Y \in B\} = \mathbf{P}\{X \in A\} \cdot \mathbf{P}\{Y \in B\}.$$

Cette définition mettant en jeu des événements du type $\{X \in A\}$, où A est un borélien, on introduit la notion de tribu engendrée par une variable aléatoire.

Définition 16.1 Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ et à valeurs réelles. La **tribu engendrée par X**, notée $\sigma(X)$, est l'ensemble des événements de la forme $\{X \in A\}$, où A décrit l'ensemble des boréliens de $\mathbb{R}$. C'est bien sûr une tribu.

Remarque 16.2 La tribu $\sigma(X)$ est donc, par construction, la plus petite tribu par rapport à laquelle la variable aléatoire X est mesurable. En effet, une telle tribu contient nécessairement tous les $\{X \in A\}$ pour $A \in \mathfrak{B}$, donc contient $\sigma(X)$.

Définition 16.3 Deux tribus $\mathfrak{T}_1$ et $\mathfrak{T}_2$ sont dites **indépendantes** si, quels que soient les événements $A_1 \in \mathfrak{T}_1$ et $A_2 \in \mathfrak{T}_2$, ces événements sont indépendants :

$$\forall A_1 \in \mathfrak{T}_1 \quad \forall A_2 \in \mathfrak{T}_2 \qquad \mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1)\,\mathbf{P}(A_2).$$

La notion de tribu engendrée par une variable permet de reformuler la définition de l'indépendance.

Définition 16.4 Deux variables aléatoires X et Y sont dites **indépendantes** si et seulement si leurs tribus engendrées $\sigma(X)$ et $\sigma(Y)$ le sont, ce qui revient à dire :

$$\forall A, B \in \mathfrak{B} \qquad \mathbf{P}\{X \in A, Y \in B\} = \mathbf{P}\{X \in A\} \cdot \mathbf{P}\{Y \in B\}.$$

Remarque 16.5 On peut étendre cette définition au cas où X et Y sont à valeurs dans $\mathbb{R}^d$ et $\mathbb{R}^q$, ou plus généralement à valeurs dans des espaces mesurables $(E, \mathfrak{E})$ et $(F, \mathfrak{F})$, l'indépendance ayant lieu si, pour tout $A \in \mathfrak{E}$ et tout $B \in \mathfrak{F}$, on a $\mathbf{P}\{X \in A, Y \in B\} = \mathbf{P}\{X \in A\} \cdot \mathbf{P}\{Y \in B\}$.

Exemple 16.6 Notons $\Omega := [0\,;1]^2$, muni de sa tribu des boréliens. Pour tout $\omega = (x, y)$ de Ω, on pose $X(\omega) = x$ et $Y(\omega) = y$. Alors les variables aléatoires X et Y sont intuitivement indépendantes — connaître Y n'apprend rien sur les valeurs possibles de X ! — ce que le théorème 16.11 confirmera bientôt.

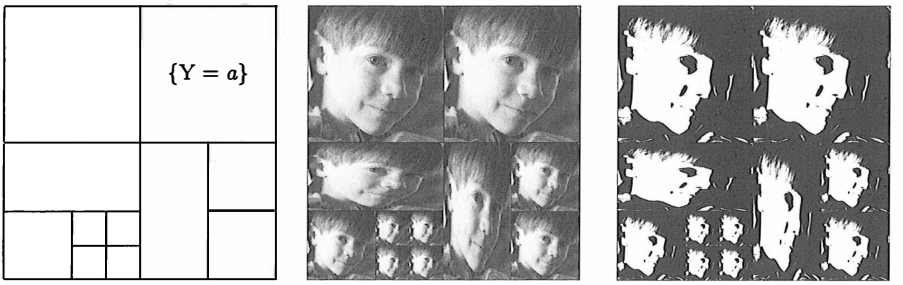
Notons $\mathbf{P}_{X,Y}$ la loi du couple (X, Y) :

$$\mathbf{P}_{X,Y}(A) = \mathbf{P}\{(X, Y) \in A\}$$

pour tout borélien A de $\mathbb{R}^2$ (ou de $\mathbb{R}^d \times \mathbb{R}^q$ si X et Y sont à valeurs vectorielles). L'indépendance de X et Y s'écrit alors, en utilisant la notion de probabilité produit (§ 230) :

$$\mathbf{P}_{X,Y} = \mathbf{P}_X \otimes \mathbf{P}_Y.$$

Remarque 16.7 (Une « image » pour représenter l'indépendance) La notion de variables indépendantes n'a pas d'équivalent dans l'analyse classique. Pour illustrer cette notion, choisissons d'abord comme univers Ω un carré du plan, de côté unité, muni de la probabilité uniforme. Considérons une variable aléatoire Y ne pouvant prendre qu'un nombre fini de valeurs distinctes, ce qui permet de partitionner Ω en régions (des événements de la forme $\{Y = y_k\}$) comme sur le dessin de gauche. Choisissons ensuite une variable aléatoire X prenant ses valeurs dans $[0\,;1]$ et convenons qu'une valeur de X est codée par un niveau de gris, avec $X = 0$ pour le noir et $X = 1$ pour le blanc. Les valeurs de X sur Ω sont représentées au milieu[1].



La variable X prend une valeur inférieure à 1/2 si celle-ci est codée par un pixel compris entre le noir et le gris moyen. Dans le dessin de droite, on a utilisé un seuil pour que les valeurs inférieures à 1/2 deviennent noires et les autres blanches. La probabilité que X prenne une valeur inférieure à 1/2 est dès lors égale à l'aire de la surface couverte par des pixels noirs dans le grand carré de l'image de droite ; la probabilité que X prenne une valeur inférieure à 1/2 *sachant que* $\{Y = a\}$ est égale à la *proportion* de la surface couverte par des pixels noirs dans le seul cadre en haut à droite.

Le lecteur pourra se convaincre que ces deux probabilités sont bien égales:

$$\mathbf{P}\{X \leqslant \tfrac{1}{2} \mid Y = a\} = \mathbf{P}\{X \leqslant \tfrac{1}{2}\}$$

(1). On se souvient que Ω ne possède pas, *a priori*, de structure particulière ; l'emplacement des différentes valeurs prises par X n'a donc aucune signification particulière, et l'image pourrait tout aussi bien représenter un magma informe.

ou encore

$$\mathbf{P}\{X \leqslant \tfrac{1}{2}, Y = a\} = \mathbf{P}\{X \leqslant \tfrac{1}{2}\} \cdot \mathbf{P}\{Y = a\}.$$

Les valeurs 1/2 et a pouvant être remplacées par toutes valeurs utiles, on a illustré l'indépendance de X et Y. ▵16.6

On utilisera couramment le théorème suivant :

■ **Théorème 16.8 (Indépendance des images)** *Soient* X *et* Y *deux variables aléatoires indépendantes. Si f et g sont des fonctions mesurables, alors $f(X)$ et $g(Y)$ sont indépendantes.*

DÉMONSTRATION : Pour tous boréliens A et B, les ensembles $f^{-1}(A)$ et $g^{-1}(B)$ sont des boréliens, donc

$$\begin{aligned}\mathbf{P}\{f(X) \in A, g(Y) \in B\} &= \mathbf{P}\{X \in f^{-1}(A), Y \in g^{-1}(B)\} \\ &= \mathbf{P}\{X \in f^{-1}(A)\} \cdot \mathbf{P}\{Y \in g^{-1}(B)\} \quad \textit{par indépendance} \\ &= \mathbf{P}\{f(X) \in A\} \cdot \mathbf{P}\{g(Y) \in B\}\end{aligned}$$

ce qui montre le résultat. ∎

Indépendance de classes et de tribus d'événements § 241

Généralisons encore la notion d'indépendance. Soient $\mathfrak{C}_1, \mathfrak{C}_2, \ldots, \mathfrak{C}_n$ des classes[2] d'événements, c'est-à-dire des parties de $\mathfrak{T}$. On dit que ces classes sont **indépendantes** si, pour tout choix d'éléments C_k de $\mathfrak{C}_k$ $(k = 1, \ldots, n)$, les événements $C_1, C_2, \ldots, C_n$ sont indépendants.

Bien sûr, puisqu'une tribu est généralement très grosse et défie la description, on peut chercher à caractériser l'indépendance de tribus par un test plus réduit ; en pratique, on peut se limiter aux éléments d'un « π-système », c'est-à-dire d'un ensemble de parties de Ω, stable par intersections finies[3] :

■ **Lemme 16.9 (Indépendance et π-systèmes)** *Soient $n \geqslant 1$ un entier et soient $\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_n$ des sous-tribus de $\mathfrak{T}$. On suppose que chacune des tribus $\mathfrak{A}_k$ est engendrée par un π-système $\mathfrak{C}_k$: $\mathfrak{A}_k = \sigma(\mathfrak{C}_k)$. Si $\mathfrak{C}_1, \mathfrak{C}_2, \ldots, \mathfrak{C}_n$ sont indépendantes, alors les tribus $\mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_n$ sont indépendantes.*

Ce lemme a une grande importance pratique, comme on le verra très prochainement. Il permet également de démontrer le lemme technique suivant, qui permettra de démontrer le lemme des coalitions pour les variables aléatoires (lemme 16.12).

■ **Lemme 16.10 (des coalitions)** *Soient n et k des entiers, et soient*

$$\mathfrak{A}_1, \ldots, \mathfrak{A}_n, \mathfrak{A}_{n+1}, \ldots, \mathfrak{A}_{n+k}$$

des tribus indépendantes. Alors les tribus

$$\mathfrak{B} = \sigma(\mathfrak{A}_1, \ldots, \mathfrak{A}_n) \qquad \textit{et} \qquad \mathfrak{C} = \sigma(\mathfrak{A}_{n+1}, \ldots, \mathfrak{A}_{n+k})$$

sont indépendantes.

Pour démontrer ces deux lemmes techniques, on fait usage du *lemme des classes monotones* ; afin de ne pas alourdir la lecture d'un chapitre déjà aride, nous en reportons la démonstration en annexe et nous invitons le lecteur à se reporter, quand le cœur lui en dira, au paragraphe A73, page 705.

(2). Ici, « classes » est simplement synonyme d'« ensembles ».

(3). Voir l'annexe B pour plus de détails sur les π-systèmes et leurs avantages.

§ 242 **Indépendance de n variables aléatoires**

On généralise les définitions précédentes ainsi. Soient $X_1, X_2, \dots, X_n$ des variables aléatoires réelles, définies sur le même espace probabilisé. On dit que $X_1, X_2, \dots, X_n$ sont **indépendantes** si leurs tribus engendrées le sont ; cela revient à dire que

$$\forall A_1, A_2, \dots, A_n \in \mathfrak{B} \qquad \mathbf{P}\{X_1 \in A_1,\ X_2 \in A_2, \dots, X_n \in A_n\} = \mathbf{P}\{X_1 \in A_1\}\ \mathbf{P}\{X_2 \in A_2\} \cdots \mathbf{P}\{X_n \in A_n\}. \quad (16.1)$$

On remarquera que la définition de l'indépendance d'événements (équation (5.1) page 113) implique en principe que certains événements $\{X_k \in A_k\}$ devraient pouvoir être omis dans la formule précédente ; c'est en fait le cas si l'on choisit $A_k = \mathbb{R}$.

Comme dans le cas de deux variables, on peut reformuler la définition d'indépendance : les n variables aléatoires $X_1, X_2, \dots, X_n$ sont indépendantes si et seulement si la loi du n-uplet $\mathbf{X} = (X_1, X_2, \dots, X_n)$ est le produit des lois individuelles

$$\mathbf{P}_{\mathbf{X}} = \mathbf{P}_{X_1} \otimes \mathbf{P}_{X_2} \otimes \cdots \otimes \mathbf{P}_{X_n}.$$

§ 243 **Caractérisations de l'indépendance de variables aléatoires**

La relation (16.1) implique évidemment que, pour tous réels $x_1, x_2, \dots, x_n$, on ait

$$\mathbf{P}\{X_1 \leqslant x_1, \dots, X_n \leqslant x_n\} = \mathbf{P}\{X_1 \leqslant x_1\} \cdots \mathbf{P}\{X_n \leqslant x_n\},$$

ce qui revient à dire que la fonction de répartition F de $\mathbf{X} = (X_1, X_2, \dots, X_n)$ est le produit $F = F_1 \otimes F_2 \otimes \cdots \otimes F_n$ des fonctions de répartition de chaque variable aléatoire, ce qui signifie simplement

$$\forall (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \qquad F(x_1, x_2, \dots, x_n) = F_1(x_1)\, F_2(x_2) \cdots F_n(x_n).$$

Réciproquement, supposons que cette relation soit vérifiée. Pour tout entier k, l'ensemble $\mathfrak{C}_k$ des événements de la forme $\{X_k \leqslant x_k\}$ est un π-système, et les classes $\mathfrak{C}_1, \mathfrak{C}_2, \dots, \mathfrak{C}_n$ sont indépendantes ; le lemme 16.9 permet de conclure que les tribus $\sigma(X_1), \sigma(X_2), \dots, \sigma(X_n)$ sont indépendantes, c'est-à-dire que $X_1, X_2, \dots, X_n$ le sont. On a donc démontré le théorème suivant :

■ **Théorème 16.11** X *et* Y *sont indépendantes si et seulement si l'une des conditions (équivalentes) suivantes est réalisée :*

i) $\mathbf{P}_{X,Y} = \mathbf{P}_X \otimes \mathbf{P}_Y$;

ii) $F_{X,Y}(x, y) = F_X(x)\, F_Y(y)$ *pour tous* $x, y \in \mathbb{R}$;

De même, $X_1, X_2, \dots, X_n$ *sont indépendantes si et seulement si l'une des conditions (équivalentes) suivantes est réalisée :*

i) $\mathbf{P}_{\mathbf{X}} = \mathbf{P}_{X_1} \otimes \mathbf{P}_{X_2} \otimes \cdots \otimes \mathbf{P}_{X_n}$;

ii) $F_{\mathbf{X}}(x_1, x_2, \dots, x_n) = F_1(x_1)\, F_2(x_2) \cdots F_n(x_n)$ *pour tous* $x_1, x_2, \dots, x_n \in \mathbb{R}$;

§ 244 **Coalitions**

Nous avons déjà rencontré le lemme des coalitions dans les cas discret et absolument continu. Sans surprise, le résultat énoncé est valable pour des variables aléatoires quelconques.

■ **Lemme 16.12 (des coalitions pour les variables aléatoires)** *Soient n un entier naturel, $X_1, X_2, \dots, X_n$ des variables aléatoires indépendantes et k un entier compris entre 1 et n. Soient $\varphi : \mathbb{R}^k \to \mathbb{R}$ et $\psi : \mathbb{R}^{n-k} \to \mathbb{R}$ des applications mesurables. Alors les variables $Y = \varphi(X_1, \dots, X_k)$ et $Z = \psi(X_{k+1}, \dots, X_n)$ sont indépendantes.*

Exemple 16.13 Si U, V, W, X, Y sont des variables aléatoires indépendantes, alors

$$Z := U \cos V \qquad S := W^2 + X^2 \qquad \text{et} \qquad T := \sqrt{Z^2 + 1}$$

sont indépendantes.

Indépendance de variables aléatoires discrètes § 245

Soient X et Y des variables aléatoires discrètes ; on note

$$X(\Omega) = \{x_k \ ; \ k \in K\} \qquad Y(\Omega) = \{y_\ell \ ; \ \ell \in L\}.$$

La classe $\mathfrak{C}_X$ des événements de la forme $\varnothing$ ou $\{X = x_k\}$ est un π-système engendrant la tribu $\sigma(X)$; de même, la classe $\mathfrak{C}_Y$ des événements de la forme $\varnothing$ ou $\{Y = y_\ell\}$ engendre $\sigma(Y)$. La définition de l'indépendance de X et Y peut alors se réduire à

$$\forall k \in K \quad \forall \ell \in L \qquad \mathbf{P}\{X = x_k, Y = y_\ell\} = \mathbf{P}\{X = x_k\}\ \mathbf{P}\{Y = y_\ell\}.$$

La généralisation à n variables est immédiate.

Indépendance de variables aléatoires à densité § 246

Supposons que $X_1, X_2, \ldots, X_n$ admettent des densités $f_1, f_2, \ldots, f_n$. D'après le théorème de Fubini, si l'on intègre la quantité positive $f_1(t_1)\, f_2(t_2) \cdots f_n(t_n)$ sur le domaine

$$\Delta(x_1, x_2, \ldots, x_n) =]-\infty ; x_1] \times]-\infty ; x_2] \times \cdots \times]-\infty ; x_n],$$

on obtient $F_1(x_1)\, F_2(x_2) \cdots F_n(x_n)$[4]. On en déduit que si les variables $X_1, \ldots, X_n$ sont indépendantes, alors la $f_1 \otimes f_2 \otimes \cdots \otimes f_n$ est une densité de la loi $\mathbf{P}_{\mathbf{X}}$. Réciproquement, si $f_1 \otimes f_2 \otimes \cdots \otimes f_n$ est une densité de $\mathbf{X}$, alors $X_1, X_2, \ldots, X_n$ sont indépendantes. On a donc prouvé :

■ **Théorème 16.14** *Des variables aléatoires $X_1, X_2, \ldots, X_n$ admettant des densités respectives $f_1, f_2, \ldots, f_n$ sont indépendantes si et seulement si $f_1 \otimes f_2 \otimes \cdots \otimes f_n$ est une densité de $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.*

Remarque 16.15 La densité d'une variable n'est définie qu'à une égalité presque partout près. Aussi, lorsque l'on connaît une densité f de $\mathbf{X}$ et les densités marginales $f_1, f_2, \ldots, f_n$ des variables $X_1, X_2, \ldots, X_n$, il suffit de vérifier que $f = f_1 \otimes f_2 \otimes \cdots \otimes f_n$ presque partout pour pouvoir conclure à l'indépendance.

On remarquera également que, si la densité conjointe s'écrit naturellement sous la forme d'un produit tensoriel de fonctions $f = f_1 \otimes f_2 \otimes \cdots \otimes f_n$, alors l'intégration par rapport à $n-1$ variables prouve, *via* le théorème de Fubini justifié par la positivité des fonctions, que les fonctions $f_1, f_2, \ldots, f_n$ sont bien des densités des variables $X_1, X_2, \ldots, X_n$.

Exemple 16.16 Supposons qu'un couple (X, Y) de variables aléatoires admette la densité

$$f(x,y) = \begin{cases} e^{-(x+y)} & \text{si } x \geqslant 0,\ y \geqslant 0, \\ 0 & \text{sinon.} \end{cases}$$

La fonction f peut se mettre sous la forme d'un produit tensoriel $f = f_X \otimes f_Y$, où $f_X(x) = f_Y(x) = e^{-x}$ si $x \geqslant 0$. On en déduit, par intégration sur une des variables, que f_X et f_Y sont les densités marginales de X et Y. En conclusion, X et Y sont indépendantes.

---

(4). En termes plus chics, intégrer la fonction $f_1 \otimes f_2 \otimes \cdots \otimes f_n$ sur le domaine $\Delta(x_1, x_2, \ldots, x_n)$ donne $F_1 \otimes F_2 \otimes \cdots \otimes F_n\,(x_1, x_2, \ldots, x_n)$.

Exemple 16.17 Supposons qu'un couple (X, Y) de variables aléatoires admette la densité

$$f(x,y) = \begin{cases} 2\,\mathrm{e}^{-(x+y)} & \text{si } 0 \leqslant y \leqslant x, \\ 0 & \text{sinon.} \end{cases}$$

On peut calculer par intégration les densités conjointes :

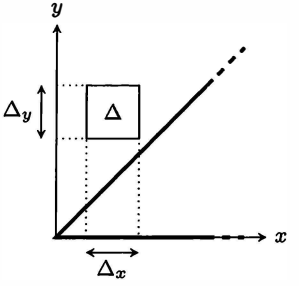
$$\forall x \geqslant 0 \qquad f_{\mathrm{X}}(x) = \int_0^{+\infty} f(x,y)\,\mathrm{d}y = 2\,\mathrm{e}^{-x} \int_0^{x} \mathrm{e}^{-y}\,\mathrm{d}y = 2\,\mathrm{e}^{-x}(1-\mathrm{e}^{-x})$$

et

$$\forall y \geqslant 0 \qquad f_{\mathrm{Y}}(y) = \int_0^{+\infty} f(x,y)\,\mathrm{d}x = 2\,\mathrm{e}^{-y} \int_y^{\infty} \mathrm{e}^{-x}\,\mathrm{d}x = 2\,\mathrm{e}^{-2y}.$$

Puisque f et $f_{\mathrm{X}} \otimes f_{\mathrm{Y}}$ ne sont pas presque partout égales, on en déduit que $f_{\mathrm{X}} \otimes f_{\mathrm{Y}}$ n'est *pas* une densité de (X, Y), et donc que les variables ne sont pas indépendantes.

Bien sûr, dans le cas présent, un argument graphique[(5)] est sans doute plus efficace pour montrer la non-indépendance de X et Y :



Dans le domaine Δ, la densité du couple est identiquement nulle ; les densités marginales en revanche sont strictement positives sur les projections Δ_x et Δ_y. Or, $\Delta = \Delta_x \times \Delta_x$, donc

$$0 = \mathbf{P}\{(\mathrm{X},\mathrm{Y}) \in \Delta\} = \mathbf{P}\{\mathrm{X} \in \Delta_x, \mathrm{Y} \in \Delta_y\} \quad \neq \quad \mathbf{P}\{\mathrm{X} \in \Delta_x\} \cdot \mathbf{P}\{\mathrm{Y} \in \Delta_y\} > 0.$$

16.2 SOMME ET PRODUIT DE VARIABLES ALÉATOIRES INDÉPENDANTES

§ 247 Espérance d'un produit

Le théorème suivant donne une condition nécessaire et suffisante d'indépendance de deux variables aléatoires.

■ **Théorème 16.18** *Deux variables aléatoires* X *et* Y *sont indépendantes si et seulement si, pour toutes fonctions f et g, réelles, mesurables et positives, on a l'égalité (dans $\mathbb{R}^+ \cup \{+\infty\}$) :*

$$\mathbf{E}\big[f(\mathrm{X})\,g(\mathrm{Y})\big] = \mathbf{E}f(\mathrm{X}) \cdot \mathbf{E}g(\mathrm{Y}). \tag{16.2}$$

L'équivalence est encore vraie si l'on remplace la condition de positivité par : f et g sont bornées (dans ce cas, l'égalité a lieu dans $\mathbb{R}$).

DÉMONSTRATION : *Supposons* X *et* Y *indépendantes.* Mettons à contribution la Cavalerie Légère du paragraphe § 215.

- La propriété voulue est vraie lorsque f et g sont des fonctions indicatrices : si $f = \mathbf{1}_{\mathrm{A}}$ et $g = \mathbf{1}_{\mathrm{B}}$, alors l'égalité

$$\mathbf{P}\{\mathrm{X} \in \mathrm{A}, \mathrm{Y} \in \mathrm{B}\} = \mathbf{P}\{\mathrm{X} \in \mathrm{A}\}\,\mathbf{P}\{\mathrm{Y} \in \mathrm{B}\}$$

s'écrit effectivement

$$\mathbf{E}\big[f(\mathrm{X})\,g(\mathrm{Y})\big] = \mathbf{E}f(\mathrm{X}) \cdot \mathbf{E}g(\mathrm{Y}).$$

(5). Cf. remarque 12.12 page 264.

- Par linéarité, la propriété est vraie pour les fonctions simples.
- Supposons maintenant f et g mesurables et positives. On peut les approcher par des suites croissantes $(f_n)_{n\geqslant 1}$ et $(g_n)_{n\geqslant 1}$ de fonctions simples positives. On obtient successivement

$$\begin{aligned}\mathbf{E}[f(\mathrm{X})\,g(\mathrm{Y})] &= \mathbf{E}\left(\lim_n \uparrow f_n(\mathrm{X})\,g_n(\mathrm{Y})\right)\\ &= \lim_{n\to\infty} \mathbf{E}[f_n(\mathrm{X})\,g_n(\mathrm{Y})] && \textit{(convergence monotone)}\\ &= \lim_{n\to\infty} \mathbf{E}[f_n(\mathrm{X})]\,\mathbf{E}[g_n(\mathrm{Y})] && \textit{(indépendance)}\\ &= \mathbf{E}f(\mathrm{X})\,\mathbf{E}g(\mathrm{Y}) && \textit{(convergence monotone).}\end{aligned}$$

- Si f et g sont bornées, on décompose $f = f^+ - f^-$ et $g = g^+ - g^-$ et on applique le résultat précédent (on notera que toutes les espérances sont alors finies).

Réciproquement, supposons la condition (16.2) *vérifiée.* Soient A et B deux événements. En notant $f := \mathbf{1}_{\mathrm{A}}$ et $g := \mathbf{1}_{\mathrm{B}}$, l'égalité de l'énoncé se traduit par

$$\mathbf{P}\{\mathrm{X}\in\mathrm{A},\mathrm{Y}\in\mathrm{B}\} = \mathbf{P}\{\mathrm{X}\in\mathrm{A}\}\,\mathbf{P}\{\mathrm{Y}\in\mathrm{B}\}.$$

On en déduit que X et Y sont indépendantes. ∎

Ce théorème a pour corollaire le résultat suivant, déjà connu dans le cas de variables discrètes comme de variables absolument continues :

■ **Théorème 16.19 (Espérance d'un produit)** *Soient* X *et* Y *des variables aléatoires indépendantes.* XY *est intégrable si et seulement si* X *et* Y *le sont et, si c'est le cas,*

$$\mathbf{E}(\mathrm{XY}) = \mathbf{E}(\mathrm{X})\cdot\mathbf{E}(\mathrm{Y}).$$

DÉMONSTRATION : On applique le théorème précédent aux fonctions positives $f, g : x \mapsto |x|$; on a alors, dans $[0\,;+\infty]$, l'égalité

$$\mathbf{E}\,|\mathrm{XY}| = \mathbf{E}\,|\mathrm{X}|\cdot\mathbf{E}\,|\mathrm{Y}|$$

qui prouve l'équivalence entre l'intégrabilité de XY et celles de X et Y.

Décomposons chaque variable en la différence de sa partie positive et de sa partie négative : $\mathrm{X} = \mathrm{X}^+ - \mathrm{X}^-$ et $\mathrm{Y} = \mathrm{Y}^+ - \mathrm{Y}^-$. Alors $\mathrm{X}^{\pm}$ est indépendante de $\mathrm{Y}^{\pm}$. Le théorème précédent, appliqué aux fonctions positives $f : x \mapsto x^+ = \max(x,0)$ et $g : x \mapsto x^- = \max(-x,0)$ montre que $\mathbf{E}(\mathrm{X}^+\mathrm{Y}^+) = \mathbf{E}(\mathrm{X}^+)\,\mathbf{E}(\mathrm{Y}^+)$, et de même pour les trois autres cas. La linéarité de l'espérance permet alors de conclure :

$$\begin{aligned}\mathbf{E}(\mathrm{XY}) &= \mathbf{E}[(\mathrm{X}^+ - \mathrm{X}^-)(\mathrm{Y}^+ - \mathrm{Y}^-)] = \mathbf{E}(\mathrm{X}^+\mathrm{Y}^+ - \mathrm{X}^+\mathrm{Y}^- - \mathrm{X}^-\mathrm{Y}^+ + \mathrm{X}^-\mathrm{Y}^-)\\ &= \mathbf{E}\mathrm{X}^+\,\mathbf{E}\mathrm{Y}^+ - \mathbf{E}\mathrm{X}^+\,\mathbf{E}\mathrm{Y}^- - \mathbf{E}\mathrm{X}^-\,\mathbf{E}\mathrm{Y}^+ + \mathbf{E}\mathrm{X}^-\,\mathbf{E}\mathrm{Y}^-\\ &= (\mathbf{E}\mathrm{X}^+ - \mathbf{E}\mathrm{X}^-)(\mathbf{E}\mathrm{Y}^+ - \mathbf{E}\mathrm{Y}^-) = \mathbf{E}(\mathrm{X})\,\mathbf{E}(\mathrm{Y}).\end{aligned}$$

∎

■ **Corollaire 16.20 (Indépendance et décorrélation)** *Si* X *et* Y *sont intégrables et indépendantes, alors le couple* (X, Y) *admet une covariance et* $\mathbf{Cov}(\mathrm{X},\mathrm{Y}) = 0$.

DÉMONSTRATION : Par définition, $\mathbf{Cov}(\mathrm{X},\mathrm{Y}) = \mathbf{E}(\mathrm{XY}) - \mathbf{E}(\mathrm{X})\,\mathbf{E}(\mathrm{Y})$, donc est nulle d'après le théorème précédent. ∎

Nous avons déjà vu des exemples montrant qu'il n'y a pas de réciproque (exemple 8.17 page 185 pour le cas discret, et remarque 12.24 page 269 pour le cas absolument continu).

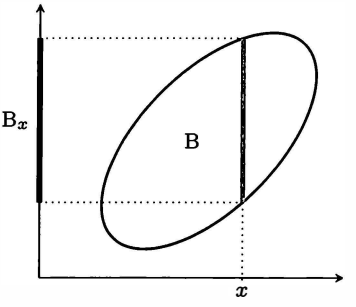
§248 **Formule de convolution**

Avant d'établir la formule de convolution dans le cas général, nous avons besoin d'un résultat intermédiaire de calcul intégral.

Soient X et Y deux variables aléatoires de lois respectives $\mathbf{P}_X$ et $\mathbf{P}_Y$. Soit B un borélien de $\mathbb{R}^2$. On note

$$B_x = \{y \in \mathbb{R} \,;\, (x,y) \in B\}$$

la **section** de B d'abscisse x. Le théorème de Fubini permet d'écrire



$$\begin{aligned}\mathbf{P}\{(X,Y) \in B\} &= \int_{-\infty}^{+\infty} \mathbf{P}_Y(B_x)\, \mathbf{P}_X(dx) \\ &= \int_{-\infty}^{+\infty} \mathbf{P}\{(x,Y) \in B\}\, \mathbf{P}_X(dx).\end{aligned}$$

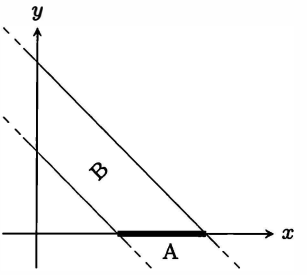
Définissons la fonction

$$s : (x,y) \longmapsto x+y,$$

application continue (et donc mesurable) de $\mathbb{R}^2$ dans $\mathbb{R}$. Si A est un borélien de $\mathbb{R}^2$, alors

$$B := s^{-1}(A) = \{(x,y) \in \mathbb{R}^2 \,;\, x+y \in A\}$$

est un borélien de $\mathbb{R}^2$.



Soient X et Y deux variables indépendantes, de fonctions de répartition F et G respectivement. On définit la **convolution $\mathbf{P}_X * \mathbf{P}_Y$ des mesures de probabilité $\mathbf{P}_X$** et $\mathbf{P}_Y$ comme la mesure de probabilité image de $\mathbf{P}_X \otimes \mathbf{P}_Y$ par s, c'est-à-dire la mesure définie par

$$\forall A \in \mathfrak{B} \qquad \mathbf{P}_X * \mathbf{P}_Y(A) := \mathbf{P}_X \otimes \mathbf{P}_Y\big(s^{-1}(A)\big).$$

Cette mesure est caractérisée par sa valeur sur les intervalles de la forme

$$A_a :=]-\infty\,;\, a\,].$$

Si l'on note

$$B_a := s^{-1}(A_a) = \{(x,y) \in \mathbb{R}^2 \,;\, x+y \in A_a\}$$

alors, d'après le calcul qui précède,

$$\begin{aligned}\mathbf{P}_X * \mathbf{P}_Y(A_a) &= \mathbf{P}_X \otimes \mathbf{P}_Y(B_a) = \mathbf{P}\{X+Y \leqslant a\} = \mathbf{P}\{X+Y \in A_a\} \\ &= \int_{-\infty}^{+\infty} \mathbf{P}\{Y \in A_a - x\}\ \mathbf{P}_X(dx) = \int_{-\infty}^{+\infty} \mathbf{P}\{Y \leqslant a-x\}\ \mathbf{P}_X(dx) \\ &= \int_{-\infty}^{+\infty} G(a-x)\, \mathbf{P}_X(dx). \qquad (16.3)\end{aligned}$$

Cela montre que la fonction de répartition de $\mathbf{P}_X * \mathbf{P}_Y$ est

$$H(a) = \int_{-\infty}^{+\infty} G(a-x)\, \mathbf{P}_X(dx) = \int_{-\infty}^{+\infty} G(a-x)\, dF(x). \qquad (16.4)$$

On la note également $H = F * G$. L'équation (16.3) prouve que la fonction H est également la fonction de répartition de la variable aléatoire $S = X + Y$.

Remarque 16.21 Puisque l'addition de variables aléatoires est commutative et associative, il en est de même de la loi de convolution des mesures et des fonctions de répartition :

$$\mathbf{P}_X * \mathbf{P}_Y = \mathbf{P}_Y * \mathbf{P}_X \qquad F * G = G * F$$
$$(\mathbf{P}_X * \mathbf{P}_Y) * \mathbf{P}_Z = \mathbf{P}_X * (\mathbf{P}_Y * \mathbf{P}_Z) \qquad (F * G) * H = F * (G * H)$$

et notamment

$$H(a) = \int_{-\infty}^{+\infty} F(a-x)\,dG(x).$$

■ **Théorème 16.22 (de convolution)** *Si* X *et* Y *sont deux variables aléatoires indépendantes, de fonctions de répartition respectives* F *et* G, *alors la loi de leur somme* $S = X+Y$ *est* $\mathbf{P}_S = \mathbf{P}_X * \mathbf{P}_Y$ *et sa fonction de répartition est* $F_S = F_X * F_Y$.

Formule de convolution pour les variables à densité § 249

Dans la formule (16.4), supposons que la fonction G soit absolument continue et admette une densité g. Écrivons que

$$G(a-x) = \int_{-\infty}^{a-x} g(s)\,ds = \int_{-\infty}^{a} g(u-x)\,du \qquad (u = s+x).$$

La positivité de g justifie l'application du théorème de Fubini, qui permet de transformer l'équation (16.4) en :

$$H(a) = (F * G)(a) = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{a} g(u-x)\,du\right] dF(x)$$
$$= \int_{-\infty}^{a} \left[\int_{-\infty}^{+\infty} g(u-x)\,dF(x)\right] du,$$

ce qui prouve que H admet pour densité la fonction, notée $F * g$, définie par

$$(F * g)(u) = \int_{-\infty}^{+\infty} g(u-x)\,dF(x).$$

Si de plus la variable X admet une densité f, alors, en vertu du théorème 14.27 page 320, le second membre de l'équation précédente s'écrit $\int_{-\infty}^{+\infty} g(u-x)\,f(x)\,dx$; la densité $F * g$ se note alors plutôt $f * g$, on l'appelle **convolution des densités f et g** :

$$(f * g)(u) = \int_{-\infty}^{+\infty} g(u-x)\,f(x)\,dx.$$

Elle est définie presque partout sur $\mathbb{R}$.

Remarque 16.23 (Convolution de distributions de Dirac) Si les fonctions de répartition de X et de Y se décomposent en la somme d'une fonction absolument continue et d'une fonction constante par morceaux, elles admettent des densités au sens des distributions (§ 200, page 299). On peut alors employer la formule de convolution au sens des distributions, en appliquant les règles suivantes : ◬16.3

$$(\delta_a * f)(x) = f(x-a) \qquad \text{et} \qquad \delta_a * \delta_b = \delta_{a+b}.$$

§ 250 **Somme de n variables aléatoires indépendantes**

Une des propriétés les plus remarquables des variables aléatoires indépendantes est l'additivité de leurs variances. Le théorème qui suit a déjà été vu dans le cas de variables discrètes, puis de variables absolument continues. Sa démonstration s'appuie uniquement sur la relation $\mathbf{E}(XY) = \mathbf{E}(X)\,\mathbf{E}(Y)$, vraie pour des variables indépendantes, et la relation

$$\mathbf{V}(X+Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\,\mathbf{Cov}(X,Y).$$

valable dans tous les cas (théorème 14.32).

■ **Théorème 16.24 (Égalité de Bienaymé)** *Soient* X *et* Y *deux variables aléatoires indépendantes et admettant un moment d'ordre* 2. *Alors* $X+Y$ *admet un moment d'ordre* 2 *et*

$$\mathbf{V}(X+Y) = \mathbf{V}(X) + \mathbf{V}(Y) \qquad \textit{et} \qquad \sigma_{X+Y} = \sqrt{\sigma_X^2 + \sigma_Y^2}.$$

De même si $X_1, X_2, \dots, X_n$ *sont des variables aléatoires indépendantes,*

$$\mathbf{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbf{V}(X_i).$$

16.3 TEMPS D'ARRÊT ET IDENTITÉ DE WALD (*)

§ 251 **Notion de temps d'arrêt**

La notion de temps d'arrêt est bien illustrée si on l'envisage comme une stratégie.

Exemple 16.25 (Deux stratégies pour gagner des fortunes à pile *ou* face*)*
Il existe un moyen bien simple de gagner (ou en tout cas de ne pas perdre) en jouant à *pile* ou *face*, à condition de pouvoir s'arrêter quand on le désire ; le voici. Notons $(X_n)_{n\geqslant 1}$ la suite de variables aléatoires définies par $X_n = 1$ si le n-ième tirage donne *pile*, $X_n = 0$ sinon, de sorte que les X_n soient des variables aléatoires indépendantes de même loi de Bernoulli $\mathscr{B}(1/2)$. Notons $S_n = X_1 + X_2 + \cdots + X_n$ le nombre de victoires et $G_n = S_n - (n - S_n) = 2S_n - n$ le gain à l'issue de n parties (nombre de victoires moins nombre de défaites).

Le théorème 20.11 page 422 montre (et on peut l'admettre entre-temps) que la suite $(G_n)_{n\geqslant 0}$ prendra des valeurs arbitrairement grandes, presque sûrement. Notamment, pour presque tout ω, il existe un entier n_ω tel que $G_{n_\omega}(\omega) = 1$ (on finit presque toujours par être, temporairement, gagnant) ; posons alors $N(\omega) := n_\omega$. La fonction N est définie sur Ω privé d'un ensemble négligeable (on peut poser $N(\omega) := +\infty$ sur cet ensemble). Voici alors la stratégie :

- si $N(\omega) = +\infty$, on ne joue pas du tout (on déclare forfait dès le début de la partie) ;
- sinon, on commence les paris et on arrête à la $N(\omega)$-ième partie (on a alors gagné 1 euro).

Bien sûr, cette « stratégie » n'est pas réalisable en pratique, et voici le premier écueil : peut-être qu'avant d'avoir pu gagner 1 euro, on est passé par une phase terrifiante où l'on a dû payer des milliards d'euros à l'adversaire. Comme aucun joueur ne dispose d'une réserve infinie d'argent, il faut amender la méthode.

Notons R la réserve d'argent à notre disposition en début de partie. Pour tout ω, définissons ensuite N comme ceci :

- Si la suite de terme général $G_n(\omega)$ passe en deçà de $-R$ *avant* de prendre la valeur $+1$, on pose $N(\omega) := +\infty$ et on ne joue pas.
- Sinon, on joue jusqu'à la partie dont le numéro $N(\omega)$ est le plus petit indice n tel que $G_n(\omega) = 1$.

Ainsi, quoi que la Destinée nous réserve, soit on ne joue pas, soit on finit par gagner un euro sans s'être jamais retrouvé dans l'impossibilité de poursuivre la partie.

La « stratégie » ainsi amendée n'est cependant toujours pas réalisable car elle souffre d'un second écueil, plus fondamental : elle suppose, dès le début de la partie, de connaître le résultat des tirages (jusqu'à un rang inconnu mais pouvant être arbitrairement grand). En pratique, une stratégie N *réalisable* est telle que l'instant d'arrêt n ne dépend *que* de la connaissance du jeu jusqu'à cet instant, c'est-à-dire des variables $X_1, X_2, \dots, X_n$ (on ne peut alors décider d'arrêter le jeu à un instant n qu'en fonction des tirages $X_1, X_2, \dots, X_n$, et pas des suivants).

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires. Un *temps d'arrêt* (adapté à cette suite) est une variable aléatoire N à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$ telle que, pour tout entier n, l'événement $\{N = n\}$ ne dépende que de $X_1, X_2, \ldots, X_n$: la connaissance des valeurs de X_1 à X_n suffit à décider si $N = n$ ou non, la connaissance de X_{n+1} et des suivantes étant inutile. Plus formellement :

Définition 16.26 Une variable aléatoire N à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$ est un **temps d'arrêt** adapté à la suite $(X_n)_{n\geqslant 1}$ si, pour tout entier $n \geqslant 1$, l'événement $\{N = n\}$ est élément de la tribu

$$\mathfrak{X}_n := \sigma(X_1, X_2, \ldots, X_n)$$

engendrée par $\sigma(X_1) \cup \cdots \cup \sigma(X_n)$.

■ **Proposition 16.27** *Puisque la suite de tribus $(\mathfrak{X}_n)_{n\geqslant 1}$ est croissante,* N *est un temps d'arrêt si et seulement si $\{N \leqslant n\} \in \mathfrak{X}_n$ pour tout entier n.* ⏏16.5

Exemple 16.28 Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires de Bernoulli, indépendantes et de même loi de Bernoulli $\mathscr{B}(p)$. On note N le plus petit entier n tel que $X_n = 1$:

$$\forall \omega \in \Omega \qquad N(\omega) = \inf \{n \in \mathbb{N}^* \,;\, X_n(\omega) = 1\}.$$

(On convient habituellement que $\inf \varnothing = +\infty$.) En d'autres termes, pour tout entier n, on a

$$\{N = n\} = \{X_1 = \cdots = X_{n-1} = 0, X_n = 1\}.$$

Cela montre que N est un temps d'arrêt.

Temps d'arrêt et identités de Wald

§ 252

Nous généralisons ici le résultat du paragraphe § 128. Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes, de même loi et admettant une espérance ; on note

$$S_n := X_1 + X_2 + \cdots + X_n.$$

Soit N un temps d'arrêt. On cherche l'espérance de la variable S_N définie par

$$\forall \omega \in \Omega \qquad S_N(\omega) := \sum_{k=1}^{N(\omega)} X_k(\omega)$$

On rappelle que, dans le cas où N était une variable aléatoire indépendante des X_n et admettant une espérance, on avait montré la relation $\mathbf{E}(S_N) = \mathbf{E}(N)\,\mathbf{E}(X)$. Bien sûr, ici, N n'est pas indépendante[6] des X_n ; malgré cela, le résultat reste valable :

■ **Théorème 16.29** *Si les X_n sont indépendants et admettent une espérance, si* N *est un temps d'arrêt admettant une espérance, alors S_N admet une espérance et*

$$\mathbf{E}(S_N) = \mathbf{E}(N)\,\mathbf{E}(X).$$

DÉMONSTRATION : Pour commencer, éliminons la variable N de la sommation définissant S_N, en remarquant que

$$\sum_{k=1}^{N} X_k = \sum_{k=1}^{\infty} X_k\, \mathbf{1}_{\{N\geqslant k\}}.$$

Pour tout entier k, l'événement $\{N \geqslant k\}$ est le complémentaire de $\{N < k\} = \{N \leqslant k-1\}$, il est donc dans la tribu engendrée par $X_1, \ldots, X_{k-1}$. Par le lemme des coalitions, $\mathbf{1}_{\{N\geqslant k\}}$ est indépendante de $X_k, X_{k+1}, \ldots$ et, notamment, de X_k. On en déduit que

$$\mathbf{E}(X_k\, \mathbf{1}_{\{N\geqslant k\}}) = \mathbf{E}(X_k)\,\mathbf{E}(\mathbf{1}_{\{N\geqslant k\}}) = \mathbf{E}(X_k)\,\mathbf{P}\{N \geqslant k\}$$

(6). Ou du moins pas *a priori*. À titre d'exercice, le lecteur pourra vérifier que si N est constante, et donc indépendante de tous les X_n, alors c'est bien un temps d'arrêt.

puis, sous réserve d'intégration terme à terme,

$$\mathbf{E}(S_N) = \mathbf{E}\left(\sum_{k=1}^{\infty} X_k \mathbf{1}_{\{N \geqslant k\}}\right) = \sum_{k=1}^{\infty} \mathbf{E}(X_k)\,\mathbf{P}\{N \geqslant k\}$$
$$= \mathbf{E}(X_1) \sum_{k=1}^{\infty} \mathbf{P}\{N \geqslant k\} = \mathbf{E}(X_1)\,\mathbf{E}(N),$$

la dernière égalité étant justifiée par le résultat établi dans l'exercice 7.5. Quant à l'intégration terme à terme, c'est une conséquence du théorème de Fubini (une des intégrations étant ici une somme discrète[7]) : il suffit de vérifier que la série de terme général $\mathbf{E}\big(|X_k|\,\mathbf{1}_{\{N \geqslant k\}}\big)$ converge. Or, quitte à travailler dans $\mathbb{R} \cup \{+\infty\}$, on peut effectuer le même raisonnement que précédemment et écrire que

$$\sum_{k=1}^{\infty} \mathbf{E}\big(|X_k|\,\mathbf{1}_{\{N \geqslant k\}}\big) = \sum_{k=1}^{\infty} \mathbf{E}\big(|X_k|\big)\,\mathbf{E}\big(\mathbf{1}_{\{N \geqslant k\}}\big) = \mathbf{E}\big(|X_1|\big)\,\mathbf{E}(N) < +\infty$$

ce qui valide l'utilisation du théorème de Fubini et l'intégration terme à terme. ∎

Exemple 16.30 (Faire tourner un programme dans un environnement instable) On désire faire tourner un programme P sur un ordinateur dont le système d'exploitation a des problèmes récurrents de stabilité[8]. On note τ le temps nécessaire pour que le programme finisse son travail (τ est une variable **déterministe**, c'est-à-dire ne dépendant pas de ω). Le système d'exploitation, lorsqu'il est lancé, fonctionne pendant un temps aléatoire avant de figer. On doit alors le redémarrer, et on relance aussitôt le programme P, qui reprend au début. Notons X_k le temps pendant lequel le système d'exploitation fonctionne après le k-ième redémarrage. On suppose que les variables aléatoires $X_1, X_2, \ldots$ sont indépendantes et identiquement distribuées.

On demande le temps moyen nécessaire pour obtenir le résultat final du programme P. On suppose connue la loi de X_1, et notamment

- la probabilité $\mathbf{P}\{X_1 < \tau\} = p$ que le système d'exploitation plante avant la fin du programme, et
- le temps moyen de fonctionnement du système, $\mathbf{E}(X_1)$.

Notons N le premier des indices n tels que $X_n \geqslant \tau$:

$$\{N = n\} = \{X_1 < \tau,\ X_2 < \tau, \ldots, X_{n-1} < \tau,\ X_n \geqslant \tau\}.$$

Le temps d'attente avant l'exécution complète du programme P est

$$T = X_1 + X_2 + \cdots + X_{N-1} + \tau.$$

La variable aléatoire $N - 1$ n'est hélas pas un temps d'arrêt, mais N l'est et l'identité de Wald s'écrit :

$$\begin{aligned}\mathbf{E}(T) &= \mathbf{E}(X_1 + X_2 + \cdots + X_{N-1} - X_N + X_N) + \tau \\ &= \mathbf{E}(X_1 + X_2 + \cdots + X_N) + \tau - \mathbf{E}(X_N) \\ &= \mathbf{E}(N)\,\mathbf{E}(X_1) - \mathbf{E}(X_N) + \tau.\end{aligned}$$

Pour calculer l'espérance de N, le plus simple est de remarquer que

$$\mathbf{P}\{N > k\} = \mathbf{P}\{X_1 < \tau, \ldots, X_k < \tau\} = \mathbf{P}\{X_1 < \tau\}^k = p^k$$

par indépendance et équidistribution. Ensuite, on obtient l'espérance de N par la formule

$$\mathbf{E}(N) = \sum_{k=0}^{\infty} \mathbf{P}\{N > k\} = \sum_{n=0}^{\infty} p^k = \frac{1}{1-p} = \frac{1}{\mathbf{P}\{X_1 \geqslant \tau\}}.$$

---

(7). Le théorème « analytique » correspondant est peut-être connu du lecteur sous la forme suivante : si $\sum f_n$ est une série de fonctions de I dans $\mathbb{R}$ de somme F, et si la série $\sum \int_I |f_n|$ converge, alors F est intégrable sur I et

$$\int_I F = \int_I \sum_{n=0}^{\infty} f_n = \sum_{n=0}^{\infty} \int_I f_n.$$

(8). On ne citera pas de nom.

Reste à déterminer la loi de X_N ; attention, celle-ci n'est évidemment pas égale à celle de X_1, puisque X_1 peut *a priori* prendre toutes les valeurs positives tandis que, par construction, X_N ne peut prendre que des valeurs au moins égales à τ. Un simple calcul montre que : *la loi de* X_N *est égale à la loi de* X_1 *conditionnée par l'événement* $\{X_1 \geqslant \tau\}$. En effet, pour tout réel x,

$$\begin{aligned}\mathbf{P}\{X_N \leqslant x\} &= \sum_{k=1}^{\infty} \mathbf{P}\{X_k \leqslant x, N = k\} = \sum_{k=1}^{\infty} \mathbf{P}\{X_1 < \tau, \dots, X_{k-1} < \tau, X_k \geqslant \tau, X_k \leqslant x\} \\ &= \sum_{k=1}^{\infty} p^{k-1}\, \mathbf{P}\{X_k \leqslant x, X_k \geqslant \tau\} = \mathbf{P}\{X_1 \leqslant x, X_1 \geqslant \tau\} \frac{1}{1-p} \\ &= \frac{\mathbf{P}\big(\{X_1 \leqslant x\} \cap \{X_1 \geqslant \tau\}\big)}{\mathbf{P}\{X_1 \geqslant \tau\}} = \mathbf{P}_{\{X_1 \geqslant \tau\}}\{X_1 \leqslant x\}.\end{aligned}$$

On en déduit l'espérance de X_N :

$$\mathbf{E}(X_N) = \frac{\mathbf{E}(X_1 \cdot \mathbf{1}_{\{X_1 \geqslant \tau\}})}{\mathbf{P}\{X_1 \geqslant \tau\}},$$

d'où l'on tire

$$\mathbf{E}(T) = \frac{\mathbf{E}(X_1)}{\mathbf{P}\{X \geqslant \tau\}} - \frac{\mathbf{E}(X_1 \cdot \mathbf{1}_{\{X_1 \geqslant \tau\}})}{\mathbf{P}\{X \geqslant \tau\}} + \tau = \frac{\mathbf{E}(X_1 \cdot \mathbf{1}_{\{X_1 < \tau\}})}{\mathbf{P}\{X \geqslant \tau\}} + \tau.$$

Application numérique : si le programme met $\tau = 10$ heures à s'exécuter et si la durée de stabilité de l'environnement suit une loi exponentielle de paramètre $\lambda = 1/2$ (la durée moyenne de stabilité de l'environnement est donc $\mathbf{E}(X_1) = 2$ heures), alors

$$\mathbf{E}(X_1 \cdot \mathbf{1}_{\{X_1 < \tau\}}) = \int_0^{\tau} x\lambda\, e^{-\lambda x}\, dx = \frac{1}{\lambda}\left[1 - (\lambda\tau + 1)\, e^{-\lambda\tau}\right] \approx 1,919,$$

tandis que $\mathbf{P}\{X > \tau\} = e^{-\lambda\tau} = e^{-5} \approx 0,00674$; on trouve donc

$$\mathbf{E}(T) \approx 295 \text{ heures.}$$

EXERCICES

♦ ***Exercice 16.1*** On reprend le modèle de natalité de l'exemple 8.1 page 176 : une famille a un nombre N d'enfants et, pour chaque naissance, la probabilité que ce soit un garçon est p et la probabilité que ce soit une fille est $q := 1 - p$. Autrement dit, les lois conditionnelles de X et Y sachant $\{N = n\}$ sont des lois binomiales $\mathscr{B}(n, p)$ et $\mathscr{B}(n, q)$.

On a déjà vu (exemple 8.4 page 178) que, si Y suit une loi de Poisson $\mathscr{P}(\lambda)$, alors X et Y sont indépendants. On veut établir la réciproque ; on suppose donc l'indépendance de X et Y. On suppose de plus que N est presque sûrement finie mais n'est pas presque sûrement nulle.

i) Montrer que $N(\Omega) = X(\Omega) = Y(\Omega) = \mathbb{N}$.

ii) Montrer que la propriété

$$\forall n \in \mathbb{N} \qquad \frac{\mathbf{P}\{N = n+1\}}{\mathbf{P}\{N = n\}} = \frac{\lambda}{n+1}$$

caractérise la loi de Poisson $\mathscr{P}(\lambda)$.

iii) Montrer que, pour tout entier n et tout couple (i, j) d'entiers tel que $i + j = n$, on peut écrire $n!\,\mathbf{P}\{N = n\}$ sous la forme d'un produit $\alpha_i\, \beta_j$.

iv) Conclure qu'il existe $\lambda > 0$ tel que $N \rightsquigarrow \mathscr{P}(\lambda)$.

♦ ***Exercice 16.2*** Soient A et B deux événements. Montrer que A et B sont indépendants si et seulement si leurs tribus engendrées $\sigma(A)$ et $\sigma(B)$ le sont.

♦ ***Exercice 16.3*** On reprend l'exemple du temps d'attente à un feu tricolore (exemple 13.13 page 296) ; on rappelle que la variable aléatoire X admet pour densité, au sens des distributions,

$$f(x) = (1 - r)\,\delta(x) + \mathbf{1}_{[0\,;\,r]}(x).$$

On suppose que, un peu plus loin sur le chemin de l'automobiliste, se trouve un second feu rouge, réglé de la même façon que le premier feu, mais dont l'instant de passage au vert est indépendant du premier. Le temps d'attente au second feu est une variable aléatoire Y, de même loi que X, et indépendante de X.

Calculer, par convolution de distributions, la densité de la somme $X + Y$. En déduire l'espérance de $X + Y$; retrouver ce résultat par linéarité.

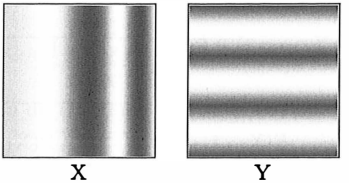
♦ ***Exercice 16.4*** Sous quelle condition nécessaire et suffisante une variable aléatoire X est-elle indépendante d'elle-même ?

♦ ***Exercice 16.5*** Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires ; soient N et N′ deux variables aléatoires à valeurs dans $\mathbb{N}^* \cup \{+\infty\}$. Montrer que :

i) si N est un temps d'arrêt, alors N + 1 est également un temps d'arrêt, mais que N − 1 n'est pas forcément un temps d'arrêt ;

ii) si N et N′ sont des temps d'arrêt, alors N + N′ est un temps d'arrêt.

♦ ***Exercice 16.6*** On choisit comme univers l'ensemble $\Omega := [0\,;1]^2$, muni de la probabilité uniforme. Expliquer pourquoi les variables X et Y représentées ci-dessous en niveaux de gris sont indépendantes.



X Y

SOLUTIONS

♦ **Solution 16.1**

i) Si $\mathbf{P}\{N=0\} = 0$, alors $\mathbf{P}\{X=0\}\cdot\mathbf{P}\{Y=0\} \neq \mathbf{P}\{X=0, Y=0\}$ donc $\mathbf{P}\{N=0\} > 0$. Supposons qu'il existe un entier n tel que $\mathbf{P}\{N=n\} = 0$; choisissons un tel n minimal. On a alors $\mathbf{P}\{X=n-1\} > 0$ et $\mathbf{P}\{Y=1\} > 0$ donc, par indépendance, $\mathbf{P}\{X=n-1, Y=1\} > 0$, ce qui est contradictoire. Ainsi, N peut prendre toutes les valeurs entières avec une probabilité non nulle.

Au vu des lois de X et Y, on a également $X(\Omega) = Y(\Omega) = \mathbb{N}$.

ii) Par récurrence, on a

$$\mathbf{P}\{N=n\} = \mathbf{P}\{N=0\}\cdot\frac{\lambda^n}{n!}$$

et la normalisation $\sum_{n=0}^{\infty} \mathbf{P}\{N=n\} = 1$ implique $\mathbf{P}\{N=0\} = e^{-\lambda}$.

iii) Soit n un entier. Pour tous entiers i, j tels que $i+j=n$, l'indépendance permet d'écrire

$$\mathbf{P}\{X=i, N=n\} = \mathbf{P}\{X=i, Y=j\} = \mathbf{P}\{X=i\}\cdot\mathbf{P}\{Y=j\}.$$

En écrivant $\mathbf{P}\{X=i, N=n\} = \mathbf{P}_{\{N=n\}}\{X=i\}\cdot\mathbf{P}\{N=n\}$, on obtient

$$\binom{n}{i} p^i q^j\, \mathbf{P}\{N=n\} = \mathbf{P}\{X=i\}\cdot\mathbf{P}\{Y=j\}$$

c'est-à-dire

$$n!\,\mathbf{P}\{N=n\} = \frac{i!\,\mathbf{P}\{X=i\}\cdot j!\,\mathbf{P}\{Y=j\}}{p^i q^j} = \alpha_i\,\beta_j.$$

Libérons la valeur de n : pour tous entiers i, j, le produit $\alpha_i\,\beta_j$ est égal à un nombre c_{i+j} ne dépendant que de $i+j$. Notamment, pour tous i, j, on a

$$\alpha_{i+1}\,\beta_{j-1} = \alpha_i\,\beta_j$$

ou encore

$$\forall (i,j) \in \mathbb{N}\times\mathbb{N}^* \qquad \frac{\alpha_{i+1}}{\alpha_i} = \frac{\beta_j}{\beta_{j-1}}.$$

Le membre de droite étant indépendant de i, on en déduit que la suite $(\alpha_n)_{n\geqslant 0}$ est géométrique ; et de même pour la suite $(\beta_n)_{n\geqslant 0}$, la raison étant la même que celle de $(\alpha_n)_{n\geqslant 0}$. On peut donc écrire $\alpha_i = \alpha_0\,\lambda^i$ et $\beta_j = \beta_0\,\lambda^j$, ce qui nous donne finalement

$$n!\,\mathbf{P}\{N=n\} = \alpha_0\,\beta_0\,\lambda^{i+j} = \alpha_0\,\beta_0\,\lambda^n$$

et, en passant au quotient,

$$(n+1)\cdot\frac{\mathbf{P}\{N=n+1\}}{\mathbf{P}\{N=n\}} = \lambda,$$

ce qui montre enfin que N suit la loi de Poisson de paramètre λ.

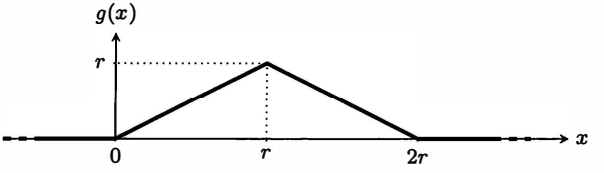
♦ **Solution 16.3** Une densité, au sens des distributions, de $X + Y$, est donnée par la convolution de f par elle-même :

$$\begin{aligned} f * f(x) &= \big((1-r)\,\delta + \mathbf{1}_{[0\,;\,r]}\big) * \big((1-r)\,\delta + \mathbf{1}_{[0\,;\,r]}\big)(x) \\ &= (1-r)^2\,\delta(x) + 2(1-r)\,\mathbf{1}_{[0\,;\,r]}(x) + g(x), \end{aligned}$$

où

$$g := \mathbf{1}_{[0\,;\,r]} * \mathbf{1}_{[0\,;\,r]}$$

a l'allure suivante :



L'espérance de $X + Y$ est donc

$$\int_{-\infty}^{+\infty} x\,[f * f]\,(x)\,\mathrm{d}x = 2(1-r)\,\frac{r^2}{2} + r^3 = r^2.$$

Par ailleurs, on peut calculer l'espérance de X, par exemple par la formule $\mathbf{E}(X) = \int_0^r (1 - F)$ (voir la figure de la page 296) ; on trouve alors $\mathbf{E}(X) = \mathbf{E}(Y) = r^2/2$, ce qui permet de retrouver le résultat précédent.

♦ **Solution 16.4** Notons F la fonction de répartition de X. Pour tout réel x, on a

$$F(x) = \mathbf{P}\{X \leqslant x\} = \mathbf{P}\{X \leqslant x, X \leqslant x\} = \mathbf{P}\{X \leqslant x\}^2 = F(x)^2$$

par indépendance de X avec elle-même. Par conséquent, F prend ses valeurs dans $\{0, 1\}$. Par ailleurs, elle prend obligatoirement ces deux valeurs, puisqu'elle tend vers 0 en $-\infty$ et vers 1 en $+\infty$. On pose alors $a = \inf\{x \in \mathbb{R}\ ;\ F(x) = 1\}$. Par continuité à droite de F, c'est d'ailleurs un minimum, et on a donc par croissance : $F(x) = 0$ pour tout $x < a$ et $F(x) = 1$ pour tout $x \geqslant a$. Ainsi, $\mathbf{P}\{X = a\} = 1$, c'est-à-dire que X est presque sûrement constante.

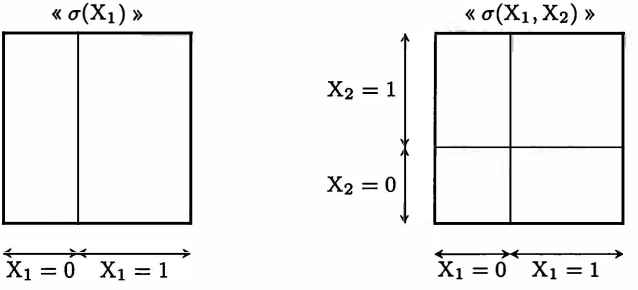
Réciproquement, supposons que X est presque sûrement constante égale à a. Soient A et B deux boréliens. Alors $\mathbf{P}\{X \in A, X \in B\}$ est égale à 1 si $a \in A$ et $a \in B$, et vaut 0 sinon. De même, $\mathbf{P}\{X \in A\}\,\mathbf{P}\{X \in B\}$ est égale à 1 si $a \in A$ et $a \in B$, et vaut 0 sinon. Ainsi, on a bien $\mathbf{P}\{X \in A, X \in B\} = \mathbf{P}\{X \in A\}\,\mathbf{P}\{X \in B\}$. Ceci étant vrai pour tout choix de A et B, on a prouvé que X est indépendante d'elle-même.

♦ **Solution 16.5** Soit N un temps d'arrêt. Pour tout $n \geqslant 2$, l'événement $\{N + 1 = n\}$ est égal à $\{N = n - 1\}$, donc est élément de $\mathfrak{X}_{n-1}$, tribu qui a le bon goût d'être incluse dans $\mathfrak{X}_n$. Cela suffit pour prouver que $N + 1$ est un temps d'arrêt.

Pour montrer que $N - 1$ n'est pas forcément un temps d'arrêt, considérons le cas où $(X_n)_{n \geqslant 1}$ est une suite de variables aléatoires indépendantes de même loi de Bernoulli $\mathscr{B}(p)$ pour $p \in\,]0\,;\,1[$, et définissons le temps d'arrêt N comme

$$N = \inf\{n \geqslant 1\ ;\ X_n = 1\}.$$

L'événement $\{N - 1 = 1\}$ est égal à l'événement $\{N = 2\}$, c'est-à-dire $\{X_1 = 0, X_2 = 1\}$, qui n'est pas dans la tribu $\mathfrak{X}_1 = \sigma(X_1)$. Dans cet exemple, $N - 1$ n'est donc pas un temps d'arrêt.



Enfin, soient N et N′ des temps d'arrêt. Notons $S = N + N'$. Pour tout entier n, on a

$$\{S = n\} = \bigcup_{k=1}^{n-1} \{N = k, N' = n - k\}.$$

Or, chaque événement $\{N = k, N' = n - k\} = \{N = k\} \cap \{N' = n - k\}$ est dans $\mathfrak{X}_{\max(k,n-k)}$, donc dans $\mathfrak{X}_n$; leur réunion est encore dans $\mathfrak{X}_n$.

♦ **Solution 16.6** L'univers $\Omega = [0\,;\,1]^2$ est muni de la probabilité produit $\mathbf{P} = \mathbf{P}_1 \otimes \mathbf{P}_2$, où $\mathbf{P}_1$ et $\mathbf{P}_2$ sont toutes deux égales à la probabilité uniforme (la mesure de Lebesgue) sur $[0\,;\,1]$.

Si l'on note génériquement $\omega = (u, v)$ les éléments de Ω, alors les variables X et Y sont de la forme $X(u, v) = f(u)$ et $Y(u, v) = g(v)$, où f et g sont des fonctions quelconques (boréliennes). Choisissons deux réels x et y et notons B_1 et B_2 les boréliens de $[0\,;\,1]$ définis par

$$B_1 := \{f \leqslant x\} = \{u \in [0\,;\,1]\ ;\ f(u) \leqslant x\}$$
$$B_2 := \{g \leqslant y\} = \{v \in [0\,;\,1]\ ;\ g(v) \leqslant y\}.$$

L'événement $\{X \leqslant x\}$ est alors égal à

$$\{X \leqslant x\} = \{(u, v) \in \Omega\ ;\ f(u) \leqslant x\} = B_1 \times [0\,;\,1].$$

De même,

$$\{Y \leqslant y\} = \{(u, v) \in \Omega\ ;\ g(v) \leqslant y\} = [0\,;\,1] \times B_2.$$

Ainsi,

$$\{X \leqslant x, Y \leqslant y\} = \Big(B_1 \times [0\,;\,1]\Big) \cap \Big([0\,;\,1] \times B_2\Big) = B_1 \times B_2.$$

Par définition de la probabilité produit, $\mathbf{P}(B_1 \times B_2) = \mathbf{P}_1(B_1) \cdot \mathbf{P}_2(B_2)$; comme de plus

$$\mathbf{P}\big(B_1 \times [0\,;\,1]\big) = \mathbf{P}_1(B_1) \qquad \text{et} \qquad \mathbf{P}\big([0\,;\,1] \times B_2\big) = \mathbf{P}_2(B_2),$$

on en déduit

$$\mathbf{P}\{X \leqslant x, Y \leqslant y\} = \mathbf{P}\{X \leqslant x\} \cdot \mathbf{P}\{Y \leqslant y\}.$$

Enfin, ce résultat étant vrai quelles que soient les valeurs de x et y, on conclut que les variables X et Y sont indépendantes.

Résumé du chapitre

- La notion d'indépendance de variables aléatoires est exprimée en termes d'indépendance de tribus engendrées par ces variables, puis en termes de lois produits.
- L'indépendance se caractérise par des relations de multiplication entre fonctions de répartition, ou de densité (voir tableau ci-dessous).
- On peut également caractériser l'indépendance par l'espérance.
- On généralise la formule de convolution pour la somme de variables indépendantes.
- Enfin, on introduit la notion de temps d'arrêt (section 16.3), permettant de généraliser les identités de Wald.

situation	conditions
variables discrètes	$\mathbf{P}\{X = x_k, Y = y_\ell\} = \mathbf{P}\{X = x_k\}\ \mathbf{P}\{Y = y_\ell\}$
variables à densité	$f(x, y) = f_X(x)\, f_Y(y)$ μ-presque partout
cas général	$F(x, y) = F_X(x)\, F_Y(y)$ pour tous x, y
	$\mathbf{E}\big[f(X)\, g(Y)\big] = \mathbf{E}f(X)\ \mathbf{E}g(Y)$ pour toutes $f, g \in \mathcal{L}^0_+(\mathbb{R}, \mathfrak{B})$

TABLE 16.1 — *Conditions à vérifier en pratique pour montrer l'indépendance de deux variables.*

Chapitre 17

Fonctions caractéristiques

Dans ce chapitre, nous présentons la définition et les principales propriétés de la fonction caractéristique d'une variable aléatoire. Il y a finalement peu de « vraies » probabilités présentes, seulement de la belle analyse assez standard, ce pourquoi nous ne nous attardons guère sur le sujet.

Cependant, les fonctions caractéristiques sont un outil d'une grande puissance dont on se passe difficilement, notamment lors de l'étude de la convergence de variables aléatoires ; elles permettent également une démonstration assez simple du théorème limite central. Le lecteur pressé peut se contenter, dans un premier temps, de parcourir les propriétés les plus importantes : théorème d'unicité, fonction caractéristique d'une somme de variables aléatoires indépendantes.

Variables aléatoires complexes

§ 253

Nous pouvons généraliser dans le corps $\mathbb{C}$ des complexes les notions de variable aléatoire (c'est-à-dire de fonction mesurable) et d'espérance.

Munissons tout d'abord $\mathbb{C}$ d'une tribu, en identifiant dans un premier temps le plan complexe au plan réel $\mathbb{R}^2$. L'ensemble des parties de $\mathbb{R}^2$ de la forme $A \times B$, où A et B sont des boréliens, n'est pas une tribu. On note $\mathfrak{B}^2$ ou $\mathfrak{B} \otimes \mathfrak{B}$ la tribu engendrée par l'ensemble de ces parties ; c'est une tribu sur $\mathbb{R}^2$, appelée **tribu des boréliens sur $\mathbb{R}^2$**. On peut vérifier[1] que cette tribu est également engendrée par les rectangles de $\mathbb{R}^2$. On la note également $\mathfrak{B}(\mathbb{C})$, l'espace $\mathbb{R}^2$ étant vu explicitement comme le plan complexe $\mathbb{C}$.

Soit X une **variable aléatoire complexe**, c'est-à-dire une application $X : \Omega \to \mathbb{C}$ mesurable par rapport aux tribus $\mathcal{T}$ et $\mathfrak{B}(\mathbb{C})$. Séparons la partie réelle de la partie imaginaire : il suffit, pour tout $\omega \in \Omega$, d'écrire $X(\omega)$ sous la forme $Y(\omega) + iZ(\omega)$ où $Y(\omega)$ et $Z(\omega)$ sont tous deux réels. On définit ainsi deux applications Y et Z vérifiant $X = Y + iZ$. Ce sont les images de l'application mesurable X par les fonctions continues « partie réelle » et « partie imaginaire », elles sont donc également mesurables[2] : ce sont des variables aléatoires.

Si Y et Z admettent toutes deux une espérance, on définit l'**espérance** de X comme

$$\mathbf{E}(X) \coloneqq \mathbf{E}(Y) + i\mathbf{E}(Z).$$

(1). Voir l'annexe B pour plus de détails.

(2). Corollaire 13.6 page 293.

Les inégalités

$$|Y| \leqslant |X| \qquad |Z| \leqslant |X| \qquad \text{et} \qquad |X| \leqslant |Y| + |Z|$$

montrent que X est intégrable si et seulement si Y et Z le sont.

§ 254 **Fonction caractéristique**

Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle ; pour tout réel t, la variable e^{itX} est de module 1, donc bornée, donc intégrable. On définit la **fonction caractéristique de X**, notée φ_X, par

$$\forall t \in \mathbb{R} \qquad \varphi_X(t) = \mathbf{E}(e^{itX}) = \int_\Omega e^{itX}\, d\mathbf{P} = \int_{-\infty}^{+\infty} e^{itx}\, dF(x).$$

(Les deux dernières formes sont équivalentes par le théorème de transfert ; on rappelle que l'on note $dF(x) = \mathbf{P}_X(dx)$ la mesure image $\mathbf{P}$ par X.)

Dans le cas où X admet une densité f, le théorème 14.28 permet d'écrire que

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f(x)\, dx,$$

c'est-à-dire que *la fonction caractéristique d'une variable aléatoire absolument continue est la transformée de Fourier de sa densité.* Dans le cas général, on dit plutôt que la fonction caractéristique est la transformée de Fourier de la mesure image $d\mathbf{P}_X$ [3].

Une réciproque partielle existe : si φ_X est une fonction intégrable sur $\mathbb{R}$, alors X admet une densité donnée par la formule d'inversion de Fourier :

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx}\, \varphi_X(t)\, dt.$$

Remarque 17.1 Il importe de voir que φ_X n'est pas une « transformée de Fourier de X », mais ne dépend en fait que de la *loi* de X.

Cette définition se généralise au cas des vecteurs aléatoires : si $\mathbf{X}$ est un tel vecteur sur $\mathbb{R}^d$, sa fonction caractéristique est définie, pour tout $\boldsymbol{u} \in \mathbb{R}^d$, par

$$\varphi_{\mathbf{X}}(\boldsymbol{u}) = \mathbf{E}\big(e^{i\langle \boldsymbol{u}|\mathbf{X}\rangle}\big) = \int_\Omega e^{i\langle \boldsymbol{u}|\mathbf{X}\rangle}\, d\mathbf{P}.$$

Si $\mathbf{X}$ admet une densité f, alors

$$\varphi_{\mathbf{X}}(\boldsymbol{u}) = \int_{\mathbb{R}^d} e^{i\langle \boldsymbol{u}|\boldsymbol{x}\rangle}\, f(\boldsymbol{x})\, d^d\boldsymbol{x}.$$

Exemple 17.2 (Fonction caractéristique d'une loi normale) Dans un premier temps, supposons que X suive une loi normale centrée réduite $\mathcal{N}(0\,;\,1)$. Sa fonction caractéristique est donc

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \underbrace{e^{itx}\, e^{-x^2/2}}_{h(x,t)}\, dx.$$

On peut dériver sous le signe somme en dominant la dérivée partielle de l'intégrande, uniformément par rapport à t :

$$\forall x \in \mathbb{R} \quad \forall t \in \mathbb{R} \qquad \left|\frac{\partial h}{\partial t}(x,t)\right| \leqslant x\, e^{-x^2/2},$$

---

[3]. Si $\mathbf{Q}$ est une probabilité sur $\mathbb{R}$, sa transformée de Fourier est la fonction définie par $\widehat{\mathbf{Q}}(t) = \int_{-\infty}^{+\infty} e^{itx}\, \mathbf{Q}(dx)$. Avec cette définition, on a donc $\varphi_X = \widehat{\mathbf{P}_X}$.

ce qui donne

$$\varphi'(t) = \frac{\mathrm{i}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x\, \mathrm{e}^{\mathrm{i}xt}\, \mathrm{e}^{-x^2/2}\, \mathrm{d}x = -\frac{t}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}xt}\, \mathrm{e}^{-x^2/2}\, \mathrm{d}x,$$

après intégration par parties ; ainsi, φ vérifie sur $\mathbb{R}$ l'équation différentielle $\varphi' + t\varphi = 0$, ce qui s'intègre en $\varphi(t) = \alpha\, \mathrm{e}^{-t^2/2}$. De la valeur $\varphi(0) = \mathbf{E}(1) = 1$, on tire finalement

$$\varphi(t) = \mathrm{e}^{-t^2/2}.$$

On aurait pu également calculer l'intégrale par la méthode des résidus, ou encore en développant $\mathrm{e}^{\mathrm{i}tx}$ en série et en intégrant terme à terme, sachant que $\mathbf{E}(\mathrm{X}^{2n}) = (2n!)/2^n\, n!$.

Le cas d'une variable de loi normale $\mathscr{N}(m\,;\,\sigma^2)$ est traité dans l'exemple 17.5.

Cas des variables aléatoires discrètes

§ 255

La formule de transfert permet également de traiter le cas où X est une variable aléatoire réelle discrète ; si l'on note $\Omega = \{x_k\ ;\ k \in \mathrm{K}\}$ et $p_k = \mathbf{P}\{\mathrm{X} = x_k\}$ pour tout $k \in \mathrm{K}$, alors

$$\varphi_{\mathrm{X}}(t) = \sum_{k \in \mathrm{K}} \mathrm{e}^{\mathrm{i}tx_k}\, p_k. \tag{17.1}$$

Exemple 17.3 (Variables de Bernoulli et de Poisson) Supposons que X suive une loi de Bernoulli $\mathscr{B}(p)$, c'est-à-dire que $p_0 = \mathbf{P}\{\mathrm{X} = 0\} = 1 - p$ et $p_1 = \mathbf{P}\{\mathrm{X} = 1\} = p$. Alors de la formule (17.1), on tire

$$\varphi_{\mathrm{X}}(t) = p\, \mathrm{e}^{\mathrm{i}t} + 1 - p.$$

De même, si X suit une loi de Poisson $\mathscr{P}(\lambda)$, alors

$$\varphi_{\mathrm{X}}(t) = \mathrm{e}^{-\lambda} \sum_{n=0}^{\infty} \mathrm{e}^{\mathrm{i}kt}\, \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda} \cdot \mathrm{e}^{\lambda\, \mathrm{e}^{\mathrm{i}t}} = \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}t} - 1)}.$$

On remarquera que la fonction caractéristique d'une variable aléatoire à valeurs dans $\mathbb{N}$ et sa fonction génératrice ont même forme : il suffit de remplacer t par $\mathrm{e}^{\mathrm{i}t}$ dans la fonction génératrice $\mathrm{G}(t) = \mathbf{E}(t^{\mathrm{X}})$ pour obtenir la fonction caractéristique $\varphi(t) = \mathbf{E}(\mathrm{e}^{\mathrm{i}t\mathrm{X}})$.

Propriétés de la fonction caractéristique

§ 256

Comme il est fréquent de centrer et réduire des variables aléatoires, voici le théorème permettant de traiter l'effet des translations et celui des dilatations.

■ **Théorème 17.4 (Translations et dilatations)** *Soient a et b des réels et* X *une variable aléatoire réelle. Alors* $\mathrm{Y} = a\mathrm{X} + b$ *a pour fonction caractéristique*

$$\varphi_{a\mathrm{X}+b}(t) = \mathrm{e}^{\mathrm{i}tb}\, \varphi_{\mathrm{X}}(at).$$

Si $\mathbf{X} = (\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_d)$ *est un vecteur aléatoire,* A *une matrice carrée de* $\mathfrak{M}_d(\mathbb{R})$ *et* $\boldsymbol{b}$ *un vecteur de* $\mathbb{R}^d$, *alors* $\mathbf{Y} = \mathrm{A}\mathbf{X} + \boldsymbol{b}$ *a pour fonction caractéristique*

$$\varphi_{\mathrm{A}\mathbf{X}+\boldsymbol{b}}(\boldsymbol{u}) = \mathrm{e}^{\mathrm{i}\langle \boldsymbol{u} | \boldsymbol{b} \rangle}\, \varphi_{\mathbf{X}}({}^{\mathrm{t}}\mathrm{A}\, \boldsymbol{u}).$$

DÉMONSTRATION : Traitons le cas vectoriel :

$$\varphi_{\mathbf{Y}}(\boldsymbol{u}) = \mathbf{E}\big(\mathrm{e}^{\mathrm{i}\langle \boldsymbol{u} | \mathrm{A}\mathbf{X}+\boldsymbol{b} \rangle}\big) = \mathrm{e}^{\mathrm{i}\langle \boldsymbol{u} | \boldsymbol{b} \rangle}\, \mathbf{E}\big(\mathrm{e}^{\mathrm{i}\langle \boldsymbol{u} | \mathrm{A}\mathbf{X} \rangle}\big) = \mathrm{e}^{\mathrm{i}\langle \boldsymbol{u} | \boldsymbol{b} \rangle}\, \mathbf{E}\big(\mathrm{e}^{\mathrm{i}\langle {}^{\mathrm{t}}\mathrm{A}\boldsymbol{u} | \mathbf{X} \rangle}\big) = \mathrm{e}^{\mathrm{i}\langle \boldsymbol{u} | \boldsymbol{b} \rangle}\, \varphi_{\mathbf{X}}({}^{\mathrm{t}}\mathrm{A}\, \boldsymbol{u}).$$

Le cas réel s'en déduit. ∎

Exemple 17.5 (Fonction caractéristique d'une loi normale — suite)
Si $\mathrm{X} \rightsquigarrow \mathscr{N}(m\,;\,\sigma^2)$, on peut l'écrire $\mathrm{X} = \sigma\mathrm{X}^* + m$, où X^* est la variable normée réduite correspondante. Puisque $\mathrm{X}^* \rightsquigarrow \mathscr{N}(0\,;\,1)$ et a pour fonction caractéristique $\varphi^*(t) = \mathrm{e}^{-t^2/2}$, on en déduit que X a pour fonction caractéristique

$$\varphi(t) = \mathrm{e}^{\mathrm{i}mt}\, \mathrm{e}^{-\sigma^2 t^2/2}.$$

■ **Proposition 17.6** *Soit* X *une variable aléatoire réelle et notons* φ *sa fonction caractéristique. Alors* $\varphi(0) = 1$, φ *est uniformément continue et* $\|\varphi\|_\infty = 1$.

DÉMONSTRATION : Par définition, $\varphi(0) = \mathbf{E}(1) = 1$. De plus, pour tout $t \in \mathbb{R}$,

$$|\varphi(t)| = |\mathbf{E}e^{itX}| \leqslant \mathbf{E}|e^{itX}| = 1.$$

Enfin, fixons momentanément un réel t ; pour tout $h \in \mathbb{R}$, on a

$$|\varphi(t+h) - \varphi(t)| = |\mathbf{E}(e^{i(t+h)X} - e^{itX})| = |\mathbf{E}(e^{itX}(1 - e^{ihX}))| \leqslant \mathbf{E}|1 - e^{ihX}| =: \psi(h),$$

où $\psi(h)$ est indépendante de t. Il suffit donc de montrer que ψ a pour limite 0 en 0, ce qui est immédiat par le théorème de convergence dominée. ∎

♦ ***Exercice 17.1*** En supposant que X est intégrable, et en utilisant l'inégalité $|e^{iz} - 1| \leqslant |z|$, montrer que φ est $\mathbf{E}|X|$-lipschitzienne.

§257 **Fonction caractéristique d'une somme de variables indépendantes**

Une des propriétés les plus utiles des fonctions caractéristiques est la suivante :

■ **Théorème 17.7** *Soient* $X_1, X_2, \ldots, X_n$ *des variables aléatoires réelles indépendantes, de fonctions caractéristiques respectives* $\varphi_1, \varphi_2, \ldots, \varphi_n$ *; alors la fonction caractéristique de leur somme* $S := X_1 + X_2 + \cdots + X_n$ *est*

$$\varphi_S(t) = \varphi_1(t)\, \varphi_2(t) \cdots \varphi_n(t).$$

DÉMONSTRATION : On rappelle que, si $Y_1, Y_2, \ldots, Y_n$ sont des variables aléatoires intégrables et indépendantes, alors $\mathbf{E}(Y_1 Y_2 \cdots Y_n) = \mathbf{E}(Y_1) \cdot \mathbf{E}(Y_2) \cdots \mathbf{E}(Y_n)$. Ici, l'indépendance de $(X_1, X_2, \ldots, X_n)$ implique celle de $e^{itX_1}, \ldots, e^{itX_n}$, ce qui permet d'écrire que

$$\mathbf{E}(e^{itS}) = \mathbf{E}(e^{itX_1}\, e^{itX_2} \cdots e^{itX_n}) = \mathbf{E}(e^{itX_1})\, \mathbf{E}(e^{itX_2}) \cdots \mathbf{E}(e^{itX_n})$$

ce qui est le résultat annoncé. ∎

Remarque 17.8 Dans le cas de variables indépendantes admettant une densité, on retrouve la relation bien connue

$$\mathscr{F}(f_1 * \cdots * f_n) = \widehat{f_1} \cdots \widehat{f_n}$$

(la transformée de Fourier d'un produit de convolution est le produit des transformées de Fourier). Le théorème précédent généralise donc cette propriété aux transformées de Fourier des mesures de probabilité.

§258 **Théorème d'unicité**

Le résultat suivant prouve que la fonction caractéristique d'une variable aléatoire caractérise entièrement sa loi. Nous l'admettons ici[4], l'idée étant essentiellement d'inverser la transformée de Fourier

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx}\, dF_X(t)$$

d'une fonction de répartition.

■ **Théorème 17.9** *Soient* X *et* Y *deux variables aléatoires de même fonction caractéristique ; alors elles ont même loi* ($\mathbf{P}_X = \mathbf{P}_Y$).

Exemple 17.10 Une variable aléatoire admet pour fonction caractéristique $\varphi(t) = e^{-t^2/2}$ si et seulement si elle suit la loi $\mathscr{N}(0\,;\,1)$.

(4). Le lecteur pourra trouver des preuves différentes de ce résultat dans de nombreux ouvrages, par exemple dans Jacod et Protter [53] ou dans Shiryaev [91].

Exemple 17.11 (Somme de variables aléatoires normales indépendantes) Soient X_1 et X_2 deux variables aléatoires indépendantes de lois respectives $\mathcal{N}(m_1 ; \sigma_1^2)$ et $\mathcal{N}(m_2 ; \sigma_2^2)$. Leurs fonctions caractéristiques sont respectivement

$$\varphi_1(t) = e^{im_1 t} e^{-\sigma_1^2 t^2/2} \qquad \text{et} \qquad \varphi_2(t) = e^{im_2 t} e^{-\sigma_2^2 t^2/2}.$$

Par indépendance, la fonction caractéristique de la variable $S = X_1 + X_2$ est (théorème 17.7) :

$$\varphi_X(t) = \varphi_1(t)\,\varphi_2(t) = e^{i(m_1+m_2)t} e^{-(\sigma_1^2+\sigma_2^2)t^2/2}.$$

Le théorème d'unicité permet alors de conclure que $S \rightsquigarrow \mathcal{N}(m_1 + m_2 ; \sigma_2^1 + \sigma_2^2)$. On retrouve donc le résultat du théorème 12.18 page 266, avec une agréable économie de calculs.

Fonction caractéristique et indépendance

§ 259

Le théorème précédent permet de montrer que des variables $X_1, X_2, \ldots, X_d$ sont indépendantes si et seulement si la fonction caractéristique du vecteur $\mathbf{X}$ est le produit direct (tensoriel) des fonctions caractéristiques de chaque variable :

■ **Théorème 17.12** *Soit $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ un vecteur aléatoire. Alors les variables aléatoires $X_1, X_2, \ldots, X_d$, de fonctions caractéristiques $\varphi_1, \varphi_2, \ldots, \varphi_d$, sont indépendantes si et seulement si, pour tout $\boldsymbol{u} = (u_1, u_2, \ldots, u_d)$,*

$$\varphi_{\mathbf{X}}(\boldsymbol{u}) = \varphi_1(u_1)\,\varphi_2(u_2)\cdots\varphi_d(u_d),$$

c'est-à-dire $\varphi_{\mathbf{X}} = \varphi_1 \otimes \varphi_2 \otimes \cdots \otimes \varphi_d$.

DÉMONSTRATION : Supposons les X_k indépendants, alors

$$\varphi_{\mathbf{X}}(u_1, u_2, \ldots, u_d) = \mathbf{E}\big(e^{i\langle \boldsymbol{u}|\mathbf{X}\rangle}\big) = \mathbf{E}\big(e^{i\sum u_k X_k}\big) = \mathbf{E}\big(e^{iu_1 X_1}\big)\cdot\mathbf{E}\big(e^{iu_2 X_2}\big)\cdots\mathbf{E}\big(e^{iu_k X_k}\big)$$

par indépendance.

Réciproquement, supposons $\varphi_{\mathbf{X}} = \varphi_1 \otimes \varphi_2 \otimes \cdots \otimes \varphi_d$. Dans le second membre, on reconnaît, d'après ce qui précède, la fonction caractéristique d'un vecteur $(Y_1, Y_2, \ldots, Y_d)$ dont les composantes sont indépendantes et tel que la loi de Y_k soit celle de X_k. En d'autres termes, $\varphi_{\mathbf{X}}$ est la fonction caractéristique de la loi $\mathbf{P}_{X_1} \otimes \mathbf{P}_{X_2} \otimes \cdots \otimes \mathbf{P}_{X_d}$. Le théorème d'unicité prouve que cette loi est la loi de $\mathbf{X}$; ainsi (§ 242), les marginales $X_1, X_2, \ldots, X_d$ sont indépendantes. ∎

Régularité de la fonction caractéristique

§ 260

Le phénomène est bien connu avec la transformation de Fourier : plus une fonction admet des moments d'ordre élevé, plus sa transformée de Fourier est régulière[(5)]. La version probabiliste en est la suivante.

■ **Théorème 17.13** *Soit n un entier naturel. Soit X une variable aléatoire réelle admettant un moment d'ordre n. Alors sa fonction caractéristique est de classe $\mathscr{C}^n$ et ses dérivées sont données par*

$$\varphi^{(\ell)}(t) = i^\ell\, \mathbf{E}(X^\ell e^{itX})$$

pour $\ell = 1, \ldots, n$.

(Version vectorielle) : *Si $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ est une variable à valeurs dans $\mathbb{R}^d$ admettant des moments jusqu'à l'ordre n, alors φ est de classe $\mathscr{C}^n$ est ses dérivées partielles sont*

$$\frac{\partial^n \varphi}{\partial u_{i_1}\,\partial u_{i_2}\cdots\partial u_{i_k}}(\boldsymbol{u}) = i^n\, \mathbf{E}(X_{i_1} X_{i_2}\cdots X_{i_k}\, e^{i\langle \boldsymbol{u}|\mathbf{X}\rangle}).$$

(5). Voir par exemple Gasquet et Witomski [36] ou tout autre ouvrage standard traitant de transformation de Fourier.

DÉMONSTRATION : Rappelons que si X admet un moment d'ordre n, alors elle admet des moments d'ordres 1 à n. Calculons d'abord la dérivée de φ ; pour cela, on considère le taux d'accroissement

$$\frac{\varphi(t+h)-\varphi(t)}{h} = \mathbf{E}\left(e^{itX}\,\frac{e^{ihX}-1}{h}\right).$$

En écrivant $|e^{ihx}-1| = |2i\,e^{ihx/2}\sin(hx/2)| = 2|\sin hx/2| \leqslant |hx|$, on tire

$$\frac{|e^{ihx}-1|}{|h|} \leqslant |x| \qquad \text{et donc} \qquad \left|e^{itX}\,\frac{e^{ihX}-1}{h}\right| \leqslant |X|.$$

Puisque $|X|$ est intégrable, cette majoration permet d'utiliser le théorème de convergence dominée et d'en déduire que φ est dérivable en t et que

$$\varphi'(t) = \lim_{h\to 0} \mathbf{E}\left(e^{itX}\,\frac{e^{ihX}-1}{h}\right) = i\,\mathbf{E}(X\,e^{itX}).$$

Le même raisonnement, répété autant de fois que nécessaire, permet de montrer la formule générale.

Le cas vectoriel se traite de façon similaire. ∎

■ **Corollaire 17.14** *Si* X *admet un moment d'ordre* n, *alors pour tout* $k = 1, \ldots, n$,

$$\mathbf{E}(X^k) = (-i)^k\,\varphi^{(k)}(0).$$

DÉMONSTRATION : Il suffit d'appliquer la formule précédente en $t = 0$. ∎

§ 261 **Fonctions caractéristiques usuelles**

Les fonctions caractéristiques des lois usuelles sont listées ici, et rappelées dans le tableau récapitulatif de la page 719.

Nom	$X(\Omega)$	$\varphi(t)$
Bernoulli $\mathscr{B}(p)$	$\{0,1\}$	$1-p+p\,e^{it}$
Binomiale $\mathscr{B}(n;p)$	$[0\,;\,n]$	$(1-p+p\,e^{it})^n$
Géométrique $\mathscr{G}(p)$	$\mathbb{N}^*$	$\dfrac{p\,e^{it}}{1+(1-p)\,e^{it}}$
Poisson $\mathscr{P}(\lambda)$	$\mathbb{N}$	$e^{\lambda(e^{it}-1)}$

Nom	$X(\Omega)$	$\varphi(t)$
Uniforme (a,b)	$[a\,;\,b]$	$\dfrac{1}{b-a}\,\dfrac{e^{itb}-e^{ita}}{it}$
Normale $\mathscr{N}(m;\sigma^2)$	$\mathbb{R}$	$e^{imt}\,e^{-\sigma^2t^2/2}$
Exponentielle $\mathscr{E}(\alpha)$	$\mathbb{R}^+$	$\dfrac{1}{1-it/\alpha}$
Cauchy $\mathscr{C}(m\,;\,a)$	$\mathbb{R}$	$e^{imt-\lvert at\rvert}$

§ 262 **Le théorème de continuité de Paul Lévy**

En avance sur les chapitres suivants, mais pour des raisons de complétude de chapitre, citons un des résultats les plus importants concernant les fonctions caractéristiques (une version un peu plus générale est donnée dans le chapitre 19, théorème 19.44 page 401).

■ **Théorème 17.15 (P. Lévy)** *Soit* $(X_n)_{n\geqslant 1}$ *une suite de variables aléatoires à valeurs dans* $\mathbb{R}$ *(ou* $\mathbb{R}^n$*), de fonctions caractéristiques respectives* $(\varphi_n)_{n\geqslant 1}$. *Soit* X *une variable aléatoire à valeurs dans le même espace, de fonction caractéristique* φ. *Alors il y a équivalence entre les énoncés*

i) $(X_n)_{n\geqslant 1}$ *converge en loi vers* X ;

ii) $(\varphi_n)_{n\geqslant 1}$ *converge simplement vers* φ.

Ce théorème fameux est utilisé pour démontrer des théorèmes limites, dont le plus célèbre est le théorème limite central.

EXERCICES

♦ ***Exercice 17.2 (Lois discrètes)*** Calculer les fonctions caractéristiques des lois suivantes :

i) loi de Bernoulli $\mathscr{B}(p)$;
ii) loi de Poisson $\mathscr{P}(\lambda)$;
iii) loi binomiale $\mathscr{B}(n\,;\,p)$;
iv) loi géométrique $\mathscr{G}(p)$.

♦ ***Exercice 17.3*** Montrer les convergences suivantes par la méthode des fonctions caractéristiques, c'est-à-dire en utilisant le théorème de continuité de Paul Lévy. On pourra commencer par calculer les fonctions caractéristiques des diverses lois incriminées, et vérifier le résultat dans la table A86.

i) Si $X_n \rightsquigarrow \mathscr{P}(\lambda_n)$ et $\lambda_n \to \lambda$, alors X_n converge en loi vers la loi de Poisson $\mathscr{P}(\lambda)$.
ii) Si $X_n \rightsquigarrow \mathscr{E}(\alpha_n)$ et $\alpha_n \to \alpha > 0$, alors X_n converge en loi vers la loi exponentielle $\mathscr{E}(\alpha)$.
iii) Si $X_n \rightsquigarrow \mathscr{B}(n\,;\,p_n)$ et si $n\,p_n \to \lambda$, alors X_n converge en loi vers la loi de Poisson $\mathscr{P}(\lambda)$.

♦ ***Exercice 17.4 (Somme de lois de Cauchy)***

i) Déterminer la fonction caractéristique d'une loi de Cauchy $\mathscr{C}(m\,;\,a)$.
ii) Déterminer la loi de la somme de deux variables aléatoires X et X′ indépendantes, de lois de Cauchy respectives $\mathscr{C}(m\,;\,a)$ et $\mathscr{C}(m'\,;\,a')$.
iii) En déduire que, si $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes, de même loi de Cauchy $\mathscr{C}(m\,;\,1)$, et si l'on pose $S_n = X_1 + X_2 + \cdots + X_n$, alors $Y_n = S_n/n$ a même loi $\mathscr{C}(m\,;\,1)$.

♦ ***Exercice 17.5*** Soit X une variable aléatoire ; soit $\alpha > 0$. Montrer que φ_X est α-périodique si et seulement si $\mathbf{P}\{X \in (2\pi/\alpha)\mathbb{Z}\} = 1$.

Indications

◊ *Indication 17.4 :* Calculer la transformée de Fourier de la fonction de densité $f(x) = \dfrac{2}{\pi}\,\dfrac{1}{1+x^2}$ par la méthode des résidus. Alternativement, on peut calculer $\varphi(t) = \dfrac{1}{\pi}\displaystyle\int_{-\infty}^{+\infty} \frac{\cos ut}{1+u^2}\,\mathrm{d}u$ en effectuant, pour $t > 0$, le changement de variable $x = ut$. En utilisant le fait que la fonction $(x,t) \mapsto t/(x^2+t^2)$ est de laplacien nul, on en déduit par dérivation sous le signe somme puis intégrations par parties que $\varphi'' - \varphi = 0$ sur $]\,0\,;+\infty[$, d'où $\varphi(t) = \mathrm{e}^{-t}$ sur $]\,0\,;+\infty[$. On trouve finalement $\varphi(t) = \mathrm{e}^{-|t|}$ pour tout réel t.

◊ *Indication 17.5 :* Si φ_X est α-périodique, montrer que $\mathbf{E}(1-\cos\alpha X) = 0$.

SOLUTIONS

♦ **Solution 17.2**

i) Si $X \rightsquigarrow \mathscr{B}(p)$, alors par la formule de transfert discrète, $\varphi_X(t) = p\,\mathrm{e}^{\mathrm{i}t} + 1 - p$.

ii) Si $X \rightsquigarrow \mathscr{B}(n\,;\,p)$, alors $\varphi_X(t) = \displaystyle\sum_{k=0}^{n} \binom{n}{k} p^k\,(1-p)^{n-k}\,\mathrm{e}^{\mathrm{i}kt} = \left[p\,\mathrm{e}^{\mathrm{i}t} + 1 - p\right]^n$. On peut également retrouver ce résultat en remarquant que X_n a même loi que la somme de n variables de loi $\mathscr{B}(p)$ indépendantes et en appliquant le théorème 17.7.

iii) Si $X \rightsquigarrow \mathscr{P}(\lambda)$, alors $\varphi_X(t) = \mathrm{e}^{-\lambda} \displaystyle\sum_{n=0}^{\infty} \frac{\lambda^n}{n!}\,\mathrm{e}^{\mathrm{i}nt} = \mathrm{e}^{-\lambda}\,\mathrm{e}^{\lambda\,\mathrm{e}^{\mathrm{i}t}} = \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}t}-1)}$.

iv) $\varphi_X(t) = \displaystyle\sum_{n=1}^{\infty} p(1-p)^{n-1}\,\mathrm{e}^{\mathrm{i}nt} = p\,\mathrm{e}^{\mathrm{i}t} \sum_{n=1}^{\infty} \left[(1-p)\,\mathrm{e}^{\mathrm{i}t}\right]^{n-1} = \dfrac{p\,\mathrm{e}^{\mathrm{i}t}}{1-(1-p)\,\mathrm{e}^{\mathrm{i}t}}$.

♦ **Solution 17.3**

i) La fonction caractéristique de X_n est $\varphi_n : t \mapsto \mathrm{e}^{\lambda_n(\mathrm{e}^{\mathrm{i}t}-1)}$, de limite simple la fonction $\varphi : t \mapsto \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}t}-1)}$ caractéristique de $\mathscr{P}(\lambda)$. Le théorème de Lévy permet de conclure.

ii) La fonction caractéristique de la loi $\mathscr{E}(\alpha_n)$ est $\varphi_n : t \mapsto (1 - \mathrm{i}t/\alpha_n)^{-1}$, de limite simple $\varphi : t \mapsto (1-\mathrm{i}t/\alpha)^{-1}$, fonction caractéristique de $\mathscr{E}(\alpha)$.

iii) La fonction caractéristique de X_n vaut $\varphi_n(t) = \left[p_n\,e^{it} + 1 - p_n\right]^n = \left[1 + p_n(e^{it} - 1)\right]^n$, et pour tout $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \left[1 + p_n(e^{it} - 1)\right]^n = e^{\lambda(e^{it}-1)}$$

(cf. A11), où l'on reconnaît la fonction caractéristique d'une loi de Poisson.

♦ **Solution 17.4**

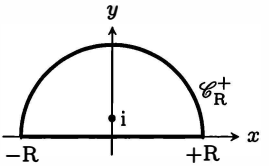
i) Le calcul par résidus n'offre pas de difficultés[6], à condition de bien séparer les cas $t > 0$ et $t < 0$. Supposons $t > 0$; on calcule

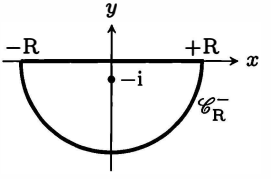
$$\varphi(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{e^{itx}}{1+x^2}\,dx$$

comme limite de l'intégrale de $z \mapsto \frac{e^{itz}}{1+z^2}$ sur le contour $\mathscr{C}_R^+$ (voir ci-dessous) dans la limite $R \to +\infty$; on vérifie que la partie de l'intégrale sur le demi-cercle tend vers 0 (lemme de Jordan) et $\varphi(t)$ est obtenue comme résidu au pôle i :

$$\varphi(t) = \frac{2i\pi}{\pi}\,\mathrm{Res}\left(\frac{e^{itz}}{1+z^2}\,;\,i\right) = 2i \lim_{z\to i}(z-i)\frac{e^{itz}}{1+z^2} = e^{-t}.$$

De même, si $t < 0$, on intègre sur le contour $\mathscr{C}_R^-$ et l'on trouve $\varphi(t) = e^t$.





En conclusion, et puisque $\varphi(0) = 1$, on a montré que pour tout t réel, $\varphi(t) = e^{-|t|}$.
Enfin, par translation/dilatation, la fonction caractéristique d'une loi $\mathscr{C}(m\,;\,a)$ est celle de la variable $aX + m$, c'est-à-dire

$$\varphi_{m,a}(t) = e^{imt}\,e^{-a|t|}.$$

(Le caractère non dérivable de φ en 0 est représentatif de la faible vitesse de décroissance de la densité, et notamment du fait qu'elle n'admet pas de moment d'ordre 1.)

ii) La fonction caractéristique de la somme des variables aléatoires indépendantes est le produit de chaque fonction caractéristique, c'est-à-dire

$$\varphi_{X+X'}(t) = e^{imt}\,e^{-a|t|}\cdot e^{im't}\,e^{-a'|t|} = e^{i(m+m')t}\,e^{-(a+a')|t|},$$

qui est la fonction caractéristique de la loi de Cauchy $\mathscr{C}(m+m'\,;\,a+a')$. Le théorème d'unicité permet de conclure que $X + X' \rightsquigarrow \mathscr{C}(m+m'\,;\,a+a')$.

iii) La fonction caractéristique de chaque X_k est $\varphi(t) = e^{imt}\,e^{-|t|}$, celle de X_k/n est $\varphi(t/n)$ et celle de S_k est

$$\psi_n(t) = \left[\varphi\left(\frac{t}{n}\right)\right]^n = \left(e^{imt/n}\,e^{-|t|/n}\right)^n = \varphi(t).$$

On en déduit, toujours grâce au théorème d'unicité, que $Y_n \rightsquigarrow \mathscr{C}(m\,;\,1)$.

♦ **Solution 17.5** Supposons que φ_X soit une fonction α-périodique. Alors $\varphi_X(\alpha) = \varphi_X(0) = 1$, ce qui s'écrit $\mathbf{E}(\cos\alpha X + i\sin\alpha X) = 1$. On en déduit que $\mathbf{E}(\cos\alpha X) = 1$ et donc $\mathbf{E}(1-\cos\alpha X) = 0$. Or la variable aléatoire $1-\cos\alpha X$ est positive ; étant d'espérance nulle, elle est donc nulle presque sûrement. Ainsi, $\mathbf{P}\{\cos\alpha X = 1\} = \mathbf{P}\{X \in (2\pi/\alpha)\mathbb{Z}\} = 1$.

Réciproquement, supposons que X soit presque sûrement à valeurs dans $(2\pi/\alpha)\mathbb{Z}$. Alors pour tout réel t on a

$$\varphi_X(t+\alpha) = \mathbf{E}\left(e^{i(t+\alpha)X}\right) = \sum_{n\in\mathbb{Z}} e^{i(t+\alpha)(2\pi n/\alpha)}\,\mathbf{P}\left\{X = \frac{2\pi n}{\alpha}\right\}$$

$$= \sum_{n\in\mathbb{Z}} e^{it(2\pi n/\alpha)}\,\mathbf{P}\left\{X = \frac{2\pi n}{\alpha}\right\} = \varphi_X(t).$$

---

[6]. La méthode des résidus et les lemmes de Jordan sont expliqués dans de nombreux ouvrages d'analyse complexe, par exemple le très pédagogique Greene & Krantz [43], le classique Rudin [83] ou, à l'usage des physiciens, la référence [2].

Chapitre 18

Vecteurs gaussiens (*)

Les vecteurs gaussiens, généralisation naturelle à $\mathbb{R}^d$ des lois normales sur $\mathbb{R}$, sont présentés en deux temps. D'abord sur $\mathbb{R}^2$, dans une approche assez pédestre ; quelques aspects numériques (droites de corrélation, ellipses d'égale densité,...), utiles en physique ou en traitement du signal, sont abordés. Sur $\mathbb{R}^d$ ensuite, de manière plus théorique, où l'on caractérise un vecteur gaussien $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ par le fait que $\langle \boldsymbol{u} | \mathbf{X} \rangle$ suit une loi normale pour tout vecteur $\boldsymbol{u}$ de $\mathbb{R}^d$.

18.1 LOIS NORMALES SUR $\mathbb{R}^2$

Lois normales et extension § 263

La loi normale $\mathcal{N}(m\,;\,\sigma^2)$ sur $\mathbb{R}$ (on dit aussi loi gaussienne **univariée**), d'espérance m et de variance σ^2 admet pour densité

$$\mathfrak{n}_{m,\sigma}(x) = \frac{1}{\sigma}\,\mathfrak{n}\left(\frac{x-m}{\sigma}\right) := \frac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-(x-m)^2/2\sigma^2}. \qquad (18.1)$$

(Voir page 231 pour des rappels élémentaires.) Cette loi est d'une importance fondamentale en théorie des probabilités et dans ses applications, notamment en statistiques. En effet, comme le montre le théorème limite central, la somme d'un grand nombre de variables aléatoires indépendantes et de même loi, correctement renormalisée, suit approximativement une loi normale, et ce, *quels que soient les détails de la loi commune à ces variables* ; quelques hypothèses simples sont naturellement requises, comme l'existence d'un moment d'ordre 2. Afin d'inclure tous les cas de figure, on peut également définir la **loi normale dégénérée** $\mathcal{N}(m\,;\,0)$, de variance nulle, comme la loi certaine de valeur m (ou **mesure de Dirac** en m), dont la fonction de répartition vaut $\mathrm{F}(x) = 0$ si $x < m$ et $\mathrm{F}(x) = 1$ si $x \geqslant m$[1].

Cette fâcheuse dichotomie entre lois normales de variances strictement positives (continues, et admettant une densité) et dégénérées (discontinues, et n'en

(1). Par définition de la convergence en loi (§ 287), on a convergence en loi de $\mathcal{N}(m\,;\,\sigma^2)$ vers $\mathcal{N}(m\,;\,0)$ lorsque $\sigma \to 0$. On remarquera que, au sens des distributions, l'expression (18.1) converge vers la distribution de Dirac en m.

admettant pas), disparaît lorsqu'on les caractérise, non par leurs fonctions de répartition, mais par leurs fonctions caractéristiques. Si X est une variable aléatoire de loi $\mathcal{N}(m\,;\,\sigma^2)$, sa fonction caractéristique est

$$\varphi(u) = \mathrm{e}^{imu}\,\mathrm{e}^{-\sigma^2u^2/2},$$

et ce résultat est encore valable pour $\sigma = 0$. Réciproquement, une telle fonction caractéristique caractérise une loi normale $\mathcal{N}(m\,;\,\sigma^2)$. Nous avons donc là un moyen de caractériser toutes les lois normales dans une même formule, ce qui justifie la définition suivante :

Définition 18.1 Une variable aléatoire réelle X est dite **gaussienne** (ou **normale**) s'il existe deux réels m et $\sigma \geqslant 0$ tels que la fonction caractéristique de X soit

$$\varphi(u) = \mathrm{e}^{imu}\,\mathrm{e}^{-\sigma^2u^2/2}.$$

On peut cependant définir une loi normale générique en restant dans le monde réel (c'est-à-dire sans passer par le monde de Fourier). En effet, si X est non dégénérée, alors la variable centrée réduite $\mathrm{N} = \mathrm{X}^* = (\mathrm{X}-m)/\sigma$ suit la loi normale $\mathcal{N}(0\,;\,1)$ et on peut écrire $\mathrm{X} = \sigma\mathrm{N} + m$. Si X est dégénérée, on peut également l'écrire $\mathrm{X} = 0\cdot\mathrm{N} + m$ presque sûrement.

Définition 18.2 (alternative) Une variable aléatoire réelle X est dite **gaussienne** (ou **normale**) s'il existe une variable aléatoire N normale centrée réduite ainsi que deux réels α et β tels que $\mathrm{X} = \alpha\mathrm{N} + \beta$ presque sûrement. Si $\alpha = 0$, la variable aléatoire est une gaussienne **dégénérée**, n'admettant pas de densité, et presque sûrement constante égale à β. Si $\alpha \neq 0$, X suit la loi $\mathcal{N}(\beta\,;\,\alpha^2)$ et admet pour densité

$$f(x) = \frac{1}{|\alpha|\,\sqrt{2\pi}}\,\mathrm{e}^{-(y-\beta)^2/2\alpha^2}.$$

Le résultat élémentaire suivant nous sera utile :

■ **Proposition 18.3 (Stabilité)** *Si* X *est une variable gaussienne et si* λ *et* μ *sont des réels quelconques,* $\lambda\mathrm{X}+\mu$ *est encore gaussienne. Si* X *et* Y *sont des lois gaussiennes indépendantes, alors* $\mathrm{X}+\mathrm{Y}$ *est encore gaussienne.*

Démonstration : Écrivons X sous la forme $\mathrm{X} = \alpha\mathrm{N} + \beta$, alors $\lambda\mathrm{X} + \mu = \lambda\alpha\mathrm{N} + \lambda\beta + \mu$ est gaussienne. Si $\mathrm{Y} = \alpha'\mathrm{N}' + \beta'$, alors $\mathrm{X} + \mathrm{Y} = \alpha\mathrm{N} + \alpha'\mathrm{N}' + \beta + \beta'$. Les cas $\alpha = 0$ ou $\alpha' = 0$ sont immédiats. Le cas où ils sont non nuls a déjà été traité par convolution (théorème 12.18 page 266) : $\mathrm{X} \rightsquigarrow \mathcal{N}(\beta\,;\,\alpha^2)$ et $\mathrm{Y} \rightsquigarrow \mathcal{N}(\beta'\,;\,\alpha'^2)$ sont indépendants, leur somme $\mathrm{X}+\mathrm{Y}$ suit la loi normale $\mathcal{N}(\beta+\beta'\,;\,\alpha^2+\alpha'^2)$. ∎

Remarque 18.4 Sans l'hypothèse d'indépendance, rien ne garantit que la somme de deux variables gaussiennes soit gaussienne. Soit X une variable normale centrée réduite, et ε une variable indépendante de X vérifiant $\mathbf{P}\{\varepsilon = +1\} = \mathbf{P}\{\varepsilon = -1\} = 1/2$. Posons $\mathrm{Y} = \varepsilon\,\mathrm{X}$. Alors Y suit encore la loi $\mathcal{N}(0\,;\,1)$, mais la somme $\mathrm{X}+\mathrm{Y}$ prend la valeur 0 avec une probabilité $1/2$, ce qui implique qu'elle n'admet pas de densité. Comme elle n'est pas non plus presque sûrement constante, elle n'est donc pas gaussienne. (On peut vérifier que la loi de $\mathrm{X}+\mathrm{Y}$ est $\frac{1}{2}\delta_0 + \frac{1}{2}\mathcal{N}(0\,;\,2)$.)

§ 264 **Lois normales sur $\mathbb{R}^2$: cas non dégénéré**

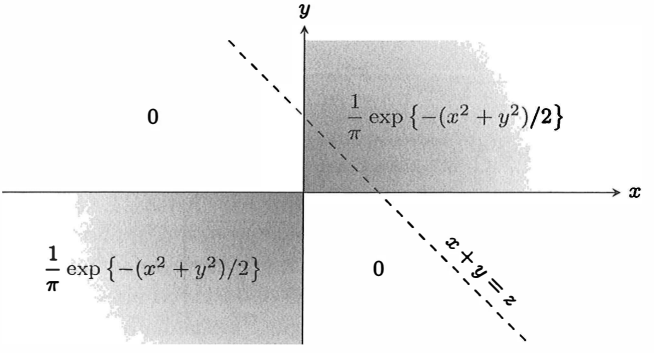
Cherchons à généraliser la notion de loi normale pour une variable aléatoire à valeurs dans $\mathbb{R}^2$. Tout d'abord, on peut commencer par choisir deux variables aléatoires X et Y de lois gaussiennes non dégénérées, respectivement égales à $\mathcal{N}(m\,;\,\sigma^2)$ et $\mathcal{N}(m'\,;\,\sigma'^2)$, et s'intéresser à la loi du couple (X,Y).

Cette loi n'est pas déterminée simplement par les lois marginales de X et de Y ; par exemple, si X et Y sont indépendantes (corrélation minimale) alors le couple (X, Y) admet une densité

$$(x,y) \longmapsto f_X(x)\, f_Y(y) = \left(\frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-m)^2/2\sigma^2}\right)\left(\frac{1}{\sigma'\sqrt{2\pi}}\, e^{-(x-m')^2/2\sigma'^2}\right)$$

tandis que, si X = Y presque sûrement (corrélation maximale !), alors le couple (X, Y) n'admet pas de densité. À vrai dire, n'imposer que les lois marginales permet d'observer les situations les plus diverses.

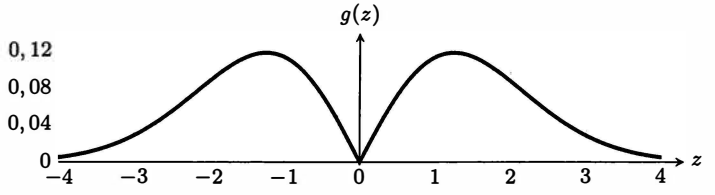
Exemple 18.5 (Une peu sympathique densité de couple dont les marginales sont gaussiennes) Considérons un couple (X, Y) dont la densité vaut $f(x,y) = \frac{1}{\pi}\exp\{-(x^2+y^2)/2\}$ si $xy \geqslant 0$, et 0 sinon :



Il est aisé de vérifier que X et Y suivent, séparément, la même loi normale $\mathcal{N}(0\,;\,1)$. En revanche, la variable aléatoire Z = X + Y a une loi très différente : d'après la formule donnant la densité d'une somme[(2)], Z admet pour densité

$$g(z) = \int_{-\infty}^{+\infty} f(t, z-t)\,dt = \int_0^z f(t, z-t)\,dt = \frac{e^{-z^2/4}}{\pi}\int_{-|z|/2}^{|z|/2} e^{-u^2}\,du,$$

qui n'est évidemment pas gaussienne, comme on le voit sur le graphe ci-dessous :



Ce phénomène est dû au fait qu'on a choisi une densité qui donnait un rôle essentiel aux axes principaux de $\mathbb{R}^2$. Une rotation du plan, ou une transformation linéaire générique

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}$$

fait disparaître tout caractère gaussien.

Parmi tous les couples de variables $\mathbf{X} = (X, Y)$ dont les lois marginales sont gaussiennes, nous choisirons, dans ce paragraphe, ceux admettant une densité dont la forme généralise la densité à une dimension, à savoir

$$f(x,y) = \frac{\sqrt{\det A}}{2\pi}\, e^{-\langle (x-m)|A(x-m)\rangle/2} \qquad (18.2)$$

(2). Voir paragraphe § 178, page 261.

où $\boldsymbol{x} = (x, y)$, A est une matrice symétrique définie positive, $\langle \boldsymbol{u}|\boldsymbol{v}\rangle = {}^t\boldsymbol{u}\,\boldsymbol{v}$ est le produit scalaire canonique de $\mathbb{R}^2$, identifié à l'espace $\mathfrak{M}_{2,1}(\mathbb{R})$ des vecteurs colonnes, et $\boldsymbol{m}$ est un vecteur de $\mathbb{R}^2$. Le préfacteur est celui qui normalise correctement la fonction (voir paragraphe suivant). Le fait que A soit symétrique définie positive nous assure que $\langle \boldsymbol{x}|\mathrm{A}\boldsymbol{x}\rangle > 0$ pour tout vecteur $\boldsymbol{x}$ non nul, et que la fonction f est intégrable. On l'appelle **loi gaussienne non dégénérée** de paramètres $\boldsymbol{m}$ et A^{-1} et on la note $\mathscr{N}(\boldsymbol{m}\,;\,\mathrm{A}^{-1})$.

Une telle loi dépend de cinq paramètres (trois paramètres pour définir la matrice symétrique A, deux paramètres pour le vecteur $\boldsymbol{m}$). Il est aisé de calculer son espérance et sa matrice de covariance en fonction de A et $\boldsymbol{m}$. Sans trop de surprise, on a

$$\mathbf{E}(\mathbf{X}) = \boldsymbol{m} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}.$$

Quant à la matrice de covariance, dont on notera les coefficients

$$\mathrm{C} = \begin{pmatrix} \sigma_x^2 & \rho\,\sigma_x\,\sigma_y \\ \rho\,\sigma_x\,\sigma_y & \sigma_y^2 \end{pmatrix},$$

elle est reliée à la matrice A par

$$\mathrm{A} = \frac{1}{1-\rho^2}\begin{pmatrix} \dfrac{1}{\sigma_x^2} & -\dfrac{\rho}{\sigma_x\,\sigma_y} \\ -\dfrac{\rho}{\sigma_x\,\sigma_y} & \dfrac{1}{\sigma_y^2} \end{pmatrix} = \mathrm{C}^{-1}.$$

Ainsi, notre loi normale se note, de manière cohérente avec le cas unidimensionnel,

$$\mathscr{N}(\boldsymbol{m}\,;\,\mathrm{C}).$$

L'équation (18.2) s'écrit, en fonction des variances de X et Y et du coefficient de corrélation ρ :

$$f(x,y) = \frac{1}{2\pi\,\sigma_x\sigma_y\sqrt{1-\rho^2}}\,\exp\left\{-\frac{1}{2(1-\rho^2)}\right.$$
$$\left.\times\left[\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2\rho(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right]\right\}$$

C'est la forme générale d'une loi gaussienne non dégénérée dans $\mathbb{R}^2$. Les caractéristiques numériques d'une telle loi normale seront exposées plus loin (section 18.2), mais une fois que l'on aura pris un peu de hauteur par rapport à cette dernière formule.

§ 265 **Réduction d'une loi normale non dégénérée sur $\mathbb{R}^2$**

La théorie générale des formes quadratiques (ou des endomorphismes symétriques), rappelée dans l'annexe A, permet d'écrire la partie quadratique

$$q(\boldsymbol{x}) = \big\langle (\boldsymbol{x}-\boldsymbol{m})\big|\mathrm{A}(\boldsymbol{x}-\boldsymbol{m})\big\rangle$$

sous forme diagonale : il existe une matrice Ω du groupe spécial orthogonal $\mathrm{SO}(2,\mathbb{R})$ telle que la matrice $\mathrm{D} = {}^t\Omega\mathrm{A}\Omega$ soit une matrice diagonale, à coefficients diagonaux strictement positifs.

Notons $1/\lambda^2$ et $1/\mu^2$ ces coefficients diagonaux, qui sont les valeurs propres de A et sont *strictement positives.* Elles vérifient donc

$$\frac{1}{\lambda^2\,\mu^2} = \text{dét}(\mathrm{A}) = \frac{1}{\sigma_x^2\,\sigma_y^2\,(1-\rho^2)}.$$

On effectue le changement de variable

$$\boldsymbol{u} = \begin{pmatrix} u \\ v \end{pmatrix} = {}^{\mathrm{t}}\Omega(\boldsymbol{x} - \boldsymbol{m})$$

ce qui permet d'exprimer la forme quadratique dans ces nouvelles coordonnées :

$$\tilde{q}(u,v) = q(x,y) = q(\Omega\,\boldsymbol{u} + \boldsymbol{m}) = \langle \Omega\boldsymbol{u} | \mathrm{A}\Omega\boldsymbol{u}\rangle = {}^{\mathrm{t}}\boldsymbol{u}\,{}^{\mathrm{t}}\Omega\mathrm{A}\Omega\boldsymbol{u} = \langle \boldsymbol{u} | \mathrm{D}\boldsymbol{u}\rangle = \frac{u^2}{\lambda^2} + \frac{v^2}{\mu^2}.$$

La fonction $\tilde{f}$ définie par

$$\tilde{f}(\boldsymbol{u}) := f(\Omega\,\boldsymbol{u} + \boldsymbol{m}) = \frac{1}{2\pi\,\lambda\mu}\,\mathrm{e}^{-\langle \boldsymbol{u}|\mathrm{D}\boldsymbol{u}\rangle/2} = \frac{1}{2\pi\,\lambda\mu}\,\mathrm{e}^{-(u^2/\lambda^2 + v^2/\mu^2)/2}$$

est la densité du couple $\mathbf{X}$ « vue dans les nouvelles coordonnées », autrement dit la densité du couple ${}^{\mathrm{t}}\Omega(\mathbf{X} - \boldsymbol{m})$. Cette densité est bien normalisée (son intégrale vaut 1) car $\text{dét}\,\Omega = 1$. Ainsi, *par une simple rotation, on observe une densité gaussienne correspondant à des variables indépendantes.*

On peut même faire une transformation plus complète pour obtenir une densité gaussienne de matrice de covariance égale à l'identité. Au passage, changeons de point de vue : au lieu d'effectuer une transformation passive (on change les coordonnées), faisons une transformation active (c'est-à-dire que l'on va considérer un nouveau vecteur $\mathbf{X}'$ sans changer de système de coordonnées). La matrice C se diagonalise au moyen de la (même) matrice orthogonale Ω :

$$\mathrm{C} = \Omega\Delta\,{}^{\mathrm{t}}\Omega$$

où $\Delta = \mathrm{diag}(\lambda^2, \mu^2) = \mathrm{D}^{-1}$ est une matrice diagonale, et notons symboliquement $\sqrt{\Delta} := \mathrm{diag}(\lambda, \mu)$, de telle sorte que $\mathrm{D}^{-1} = \sqrt{\Delta}^2 = \Delta = {}^{\mathrm{t}}\Omega\,\mathrm{V}\,\Omega$. Si l'on définit la variable aléatoire $\mathbf{X}'$ par

$$\mathbf{X} = (\Omega\sqrt{\Delta})\mathbf{X}' + \boldsymbol{m} \qquad \textit{i.e.} \qquad \mathbf{X}' := \sqrt{\Delta}^{-1}\,{}^{\mathrm{t}}\Omega(\mathbf{X} - \boldsymbol{m}),$$

elle admet une densité

$$g(\boldsymbol{x}') = f(\Omega\sqrt{\Delta}\boldsymbol{x}' + \boldsymbol{m}) \cdot \left|\frac{\mathrm{D}\boldsymbol{x}}{\mathrm{D}\boldsymbol{x}'}\right|,$$

où le jacobien $\mathrm{D}\boldsymbol{x}/\mathrm{D}\boldsymbol{x}'$ vaut $\text{dét}(\Omega\sqrt{\Delta}) = \text{dét}\,\sqrt{\Delta} = \sqrt{\text{dét}\,\mathrm{V}} = \lambda\mu$, ce qui donne

$$g(\boldsymbol{x}') = \frac{1}{2\pi}\,\mathrm{e}^{-\langle \Omega\sqrt{\Delta}\boldsymbol{x}' | \mathrm{A}\Omega\sqrt{\Delta}\boldsymbol{x}'\rangle/2} = \frac{1}{2\pi}\,\mathrm{e}^{-\langle \boldsymbol{x}' | \boldsymbol{x}'\rangle/2},$$

c'est-à-dire la loi normale $\mathscr{N}(\mathbf{0}\,;\,\mathrm{I}_2)$.

■ **Théorème 18.6** *La variable aléatoire* $\mathbf{X}'$ *suit la loi normale centrée réduite, de matrice de covariance égale à l'identité.*

Remarque 18.7 On retrouve ce résultat facilement si l'on se souvient qu'une transformation affine envoie une loi gaussienne sur une loi gaussienne. La moyenne de $\mathbf{X}'$ est, par linéarité, $\mathbf{E}(\sqrt{\Delta}^{-1}\,{}^{\mathrm{t}}\Omega(\mathbf{X} - \boldsymbol{m})) = \sqrt{\Delta}^{-1}\,{}^{\mathrm{t}}\Omega\,\mathbf{E}(\mathbf{X} - \boldsymbol{m}) = \mathbf{0}$. Quant à la matrice de covariance de $\mathbf{X}'$, elle vaut $\sqrt{\Delta}^{-1}\,{}^{\mathrm{t}}\Omega\mathrm{V}\Omega\sqrt{\Delta}^{-1} = \mathrm{I}_n$ (voir théorème 15.5 page 334).

Remarque 18.8 Plutôt que de factoriser la matrice V sous la forme $V = \Omega\sqrt{\Delta}\cdot\sqrt{\Delta}{}^t\Omega$, on aurait pu prendre n'importe quelle matrice Λ telle que $V = \Lambda\cdot{}^t\Lambda$, et le vecteur $\mathbf{X}''$ défini par $\mathbf{X} - \boldsymbol{m} = \Lambda\mathbf{X}''$ aurait également suivi la loi $\mathcal{N}(\mathbf{0}\,;\,I_2)$. Un cas intéressant est la décomposition de Cholesky de V :

$$V = L\,{}^tL$$

où L est triangulaire inférieure, à coefficients diagonaux positifs. On définit alors $\mathbf{X}''$ par $\mathbf{X} - \boldsymbol{m} = L\mathbf{X}''$, l'inversion de la matrice L étant aisée puisqu'elle est triangulaire. L'avantage est que la décomposition de Cholesky est assez rapide à mettre en œuvre et bien optimisée dans les logiciels de calcul formel.

18.2 ASPECTS NUMÉRIQUES DE LOIS NORMALES SUR $\mathbb{R}^2$

Dans toute cette section, (X, Y) est un couple suivant une loi normale non dégénérée, de densité

$$f(x,y) = \frac{\sqrt{\det A}}{2\pi}\,e^{-\langle(x-m)|A(x-m)\rangle/2} \tag{18.3}$$

déjà donnée dans les équations (18.2) et suivantes.

§ 266 **Lois marginales**

La loi marginale de X s'obtient en intégrant f par rapport à sa deuxième variable :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy.$$

On commence par un changement de variable $v = (y - m_y)/\sigma_y$ qui donne, en notant $u = (x - m_x)/\sigma_x$:

$$\begin{aligned} f_X(x) &= \frac{1}{2\pi\,\sigma_x\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[u^2 - 2\rho uv + v^2\right]\right\}\,dv \\ &= \frac{1}{2\pi\,\sigma_x\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[(1-\rho^2)\,u^2 + (v-\rho u)^2\right]\right\}\,dv \\ &= \frac{e^{-u^2/2}}{2\pi\,\sigma_x\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\exp\left\{-\frac{(v-\rho u)^2}{2(1-\rho^2)}\right\}\,dv \end{aligned}$$

Un dernier changement de variable $t = (v - \rho u)/\sqrt{1-\rho^2}$ donne

$$f_X(x) = \frac{e^{-u^2/2}}{2\pi\,\sigma_x}\int_{-\infty}^{+\infty} e^{-t^2/2}\,dt = \frac{e^{-u^2/2}}{\sqrt{2\pi}\,\sigma_x} = \frac{e^{-(x-m_x)^2/2\sigma_x^2}}{\sqrt{2\pi}\,\sigma_x}.$$

Ainsi, X suit la loi normale de moyenne m_x et d'écart-type σ_x, et on a montré le résultat suivant :

■ **Proposition 18.9** *Si* $\mathbf{X} = (X, Y)$ *suit une loi normale non dégénérée, alors ses marginales suivent des lois normales non dégénérées. Si* $\mathbf{X}$ *admet pour moyenne et pour matrice de covariances*

$$\boldsymbol{m} = \mathbf{E}(\mathbf{X}) = \begin{pmatrix} m_x \\ m_y \end{pmatrix} \qquad \text{et} \qquad V = C^{\mathbf{X}} = \begin{pmatrix} \sigma_x^2 & \rho\,\sigma_x\,\sigma_y \\ \rho\,\sigma_x\,\sigma_y & \sigma_y^2 \end{pmatrix},$$

alors $X \rightsquigarrow \mathcal{N}\left(m_x\,;\,\sigma_x^2\right)$ *et* $Y \rightsquigarrow \mathcal{N}\left(m_y\,;\,\sigma_y^2\right)$.

Combinaisons linéaires de X et Y § 267

Nous pouvons généraliser le calcul précédent en projetant, non pas selon les axes principaux de $\mathbb{R}^2$, mais selon tout autre système d'axes, ce qui revient en pratique à étudier une combinaison linéaire arbitraire de X et de Y.

Une telle combinaison linéaire $\alpha X + \beta Y$ s'écrit également $\langle \boldsymbol{u}|\mathbf{X}\rangle$, avec $\boldsymbol{u} = (\alpha, \beta)$. La densité de la variable aléatoire $Z = \langle \boldsymbol{u}|\mathbf{X}\rangle$ peut être obtenue par un calcul direct de la fonction de répartition F_Z, calcul quelque peu fastidieux que nous présentons rapidement (on verra plus loin une méthode bien plus intéressante pour déterminer cette loi, en utilisant les fonctions caractéristiques). Pour tout $z \in \mathbb{R}$, on a

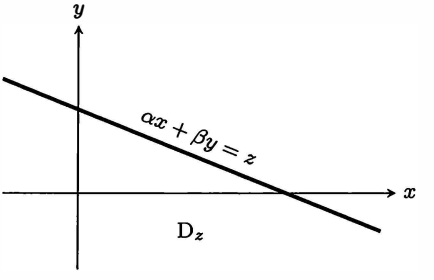
$$F_Z(z) = \mathbf{P}\{Z \leqslant z\} = \iint_{D_z} f(\boldsymbol{x})\, d^2\boldsymbol{x},$$

où D_z est le domaine du plan défini par

$$D_z = \{\boldsymbol{x} \in \mathbb{R}^2 \,;\, \langle \boldsymbol{u}|\boldsymbol{x}\rangle \leqslant z\} = \{(x, y) \in \mathbb{R}^2 \,;\, \alpha x + \beta y \leqslant z\}$$

Supposons par exemple que $\alpha > 0$, alors

$$D_z = \left\{(x, y) \in \mathbb{R}^2 \,;\, x \leqslant \frac{z - \beta y}{\alpha}\right\}.$$



On obtient

$$F_Z(z) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{(z-\beta y)/\alpha} f(x, y)\, dx \right) dy$$

et, par dérivation, une densité de Z :

$$f_Z(z) = \frac{1}{\alpha} \int_{-\infty}^{+\infty} f\left(\frac{z - \beta y}{\alpha}, y\right) dy.$$

Les techniques générales de calcul d'une intégrale gaussienne (mise sous forme canonique du terme quadratique dans l'exponentielle) prouvent que f_Z est une fonction gaussienne, et plus précisément

$$f_Z(z) = \frac{1}{2\pi\sqrt{\Sigma}} \exp\left\{-\frac{(z - \langle \boldsymbol{u}|\boldsymbol{m}\rangle)^2}{2\Sigma^2}\right\}, \qquad \text{où} \qquad \Sigma^2 = \alpha^2\sigma_x^2 + 2\alpha\beta\rho\,\sigma_x\sigma_y + \beta^2\sigma_y^2.$$

Le calcul est identique dans le cas $\alpha < 0$; le cas $\alpha = 0$ ($Z = \beta Y$) est immédiat et conduit à la même formule finale. Ainsi :

■ **Proposition 18.10** *Si* $\mathbf{X} = (X, Y)$ *suit une loi normale non dégénérée sur* $\mathbb{R}^2$, *toute combinaison linéaire* $Z = \langle \boldsymbol{u}|\mathbf{X}\rangle$, $\boldsymbol{u} \neq \mathbf{0}$, *suit une loi normale non dégénérée.*

Cette propriété permet de généraliser à $\mathbb{R}^n$ la notion de loi normale, dégénérée ou non.

§ 268 **Densités conditionnelles**

La densité marginale de X ne s'annulant pas sur $\mathbb{R}$, la densité conditionnelle de Y sachant $\{X = x\}$ est

$$f^{Y}_{\{X=x\}}(y) =: f(y \mid x) = \frac{f(x,y)}{f_X(x)}.$$

Un calcul (peu engageant) d'intégrale gaussienne prouve que cette densité est encore gaussienne ; tous calculs effectués, on obtient

$$f(y \mid x) = \frac{1}{\sqrt{2\pi}\,\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_y^2(1-\rho^2)}\big[y - m(y \mid x)\big]^2\right\},$$

où l'on a noté

$$m(y \mid x) = m_y + \frac{\rho\sigma_y}{\sigma_x}(x - m_x).$$

En d'autres termes, sous la probabilité conditionnée $\mathbf{P}_{\{X=x\}}$, la variable Y suit le loi normale de moyenne $m(y \mid x)$ et de variance $(1-\rho^2)\,\sigma_y^2$.

La droite d'équation

$$y = m(y \mid x) = m_y + \frac{\rho\sigma_y}{\sigma_x}(x - m_x) \tag{18.4}$$

est la **droite de régression de Y par rapport à X**. Son interprétation géométrique sera plus claire sur l'ellipse de covariance.

Bien évidemment, nous pouvons également calculer la densité de X sachant $\{Y = y\}$, les résultats étant de la même forme. La droite

$$x = m(x \mid y) = m_x + \frac{\rho\sigma_x}{\sigma_y}(y - m_y) \tag{18.5}$$

est la **droite de régression de Y par rapport à X**.

§ 269 **Ellipses d'égale densité**

Nous avons vu qu'une simple translation, suivie d'une rotation, permet de réduire la partie quadratique $q(x,y)$ en

$$q(x,y) = \frac{u^2}{\lambda^2} + \frac{v^2}{\mu^2},$$

où (u,v) est relié à (x,y) par

$$\begin{pmatrix} u \\ v \end{pmatrix} = {}^{t}\Omega \begin{pmatrix} x - m_x \\ y - m_y \end{pmatrix}.$$

Les courbes $\mathscr{E}_k$ d'équation $q(x,y) = k^2$ sont appelées **ellipses d'égale densité** ; en un point d'une telle courbe, la densité est en effet constante et vaut

$$f(x,y) = \frac{1}{2\pi\,\sigma_x\sigma_y\sqrt{1-\rho^2}}\,\mathrm{e}^{-k^2/2}.$$

L'ellipse $\mathscr{E}_1$ est appelée **ellipse de corrélation**. Elle est tangente au rectangle dont la largeur est le segment $[m_x - \sigma_x\,;\, m_x + \sigma_x]$ et la hauteur $[m_y - \sigma_y\,;\, m_y + \sigma_y]$. On remarquera que les abscisses et les ordonnées extrémales de cette ellipse sont indépendantes des coefficients de corrélation.

La démonstration n'est qu'une affaire de calcul : un point (x,y) de $\mathscr{E}_1$ est d'abscisse extrémale s'il vérifie la double condition

$$q(x,y) = 1 \qquad \text{et} \qquad \frac{\partial q}{\partial y}(x,y) = 0,$$

ce qui nous donne un système de deux équations dont les solutions sont

$$x = m_x \pm \sigma_x \qquad y = m_y + \frac{\rho\sigma_y}{\sigma_x}(x - m_x) = m_y \pm \rho\sigma_y.$$

On remarquera que le point extrémal (x,y) est sur la droite D de régression d'équation (18.4).

Une coupe verticale de l'ellipse de corrélation donne deux points de même abscisse ; le lieu des milieux de ces points est la droite D de régression de y par rapport à x (voit figure 18.1).

Il est aisé d'intégrer la densité sur l'intérieur d'une ellipse $\mathscr{E}_k$, le changement de variable $x' = \sqrt{\Delta}^{-1}\,{}^{t}\Omega(x - m)$ transformant précisément l'ellipse en le cercle de rayon k, et le jacobien $\mathrm{D}(x,y)/\mathrm{D}(x',y') = \text{dét}\,\sqrt{\Delta} = \lambda\mu = \sigma_x\sigma_y\sqrt{1-\rho^2}$ annulant le préfacteur de f :

$$\iint_{\mathscr{E}_k} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{2\pi}\iint_{\mathrm{D}_k} \mathrm{e}^{-(x'\mid x')/2}\,\mathrm{d}x' = \frac{1}{2\pi}\iint_{\mathrm{D}_k} \mathrm{e}^{-(u^2+v^2)/2}\,\mathrm{d}u\,\mathrm{d}v,$$

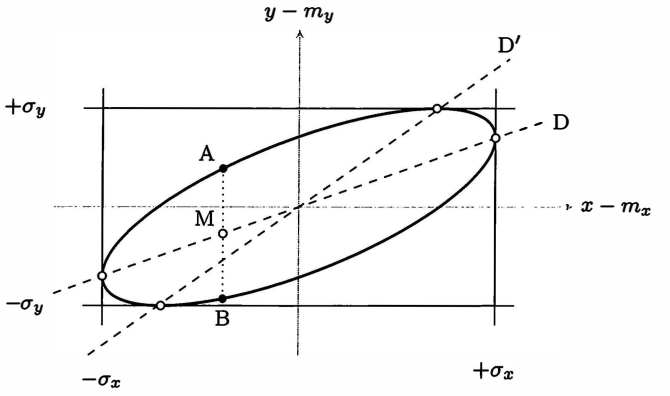


FIGURE 18.1 — *Ellipse de corrélation et droites de régression. La droite* D *est la droite de régression de* y *par rapport à* x *; la droite* D′ *est celle de régression de* x *par rapport à* y*. Le point* M *est le milieu du segment* [A ; B].

et un ultime changement de variable cartésien / polaire donne

$$\iint_{\mathscr{E}_k} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^{2\pi} \mathrm{d}\theta \int_0^k r\,\mathrm{e}^{-r^2/2}\,\mathrm{d}r = 1 - \mathrm{e}^{-k^2/2}.$$

■ **Théorème 18.11 (Contenu des ellipses d'égale densité)** *Le contenu en probabilité d'une ellipse* $\mathscr{E}_k$ *est*

$$\mathbf{P}\{\mathbf{X} \in \mathscr{E}_k\} = 1 - \mathrm{e}^{-k^2/2}.$$

Notamment, l'ellipse de corrélation contient $1 - \mathrm{e}^{-1/2} \approx 39,4\,\%$ *de la probabilité totale. L'ellipse* $\mathscr{E}_2$ *contient* $\approx 86,5\,\%$ *de la probabilité totale (voir figure 18.2).*

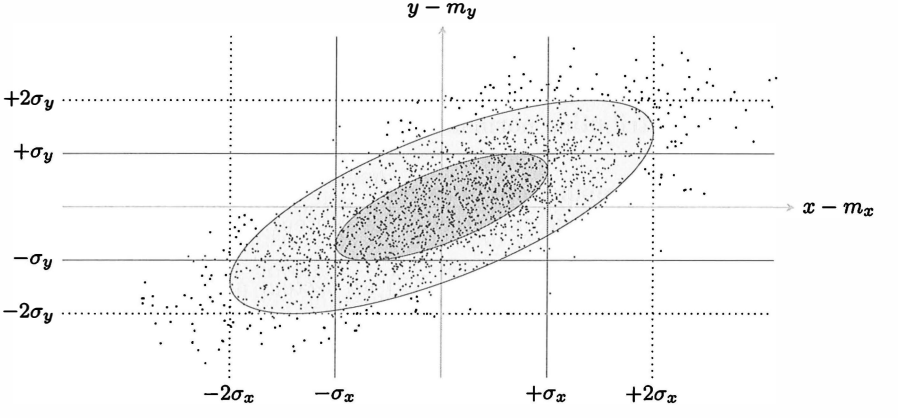


FIGURE 18.2 — *Ellipses d'équation* $q(x,y) = 1$ *et* $q(x,y) = 4$*. La première est contenue dans le* rectangle de corrélation*, de largeur* $2\sigma_x$ *et de hauteur* $2\sigma_y$*. On a également représenté 2000 points tirés aléatoirement selon la loi gaussienne correspondante*(4)*. Les valeurs numériques choisies dans cette illustration sont* $\sigma_x = 2$, $\sigma_y = 1$ *et* $\rho = 0,7$.

(4). Le chapitre 25 explique en détail comment obtenir, informatiquement, une telle série de points.

18.3 VECTEURS GAUSSIENS

§ 270 **Définition des vecteurs gaussiens**

Les variables gaussiennes (éventuellement dégénérées) sur $\mathbb{R}$ sont des fonctions affines de variables normales centrées réduites. Soit $\mathrm{E} \simeq \mathbb{R}^d$ un espace vectoriel de dimension finie d, muni d'un produit scalaire $\langle \cdot | \cdot \rangle$. La généralisation naturelle de la notion de densité gaussienne non dégénérée est

$$f(\boldsymbol{x}) = \frac{\sqrt{\det \mathrm{A}}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2}\langle \boldsymbol{x}-\boldsymbol{m} | \mathrm{A}(\boldsymbol{x}-\boldsymbol{m})\rangle\right]$$

où A est une matrice symétrique définie positive, ce qui donne une fonction caractéristique égale à

$$\varphi_{\mathbf{X}}(\boldsymbol{u}) = \exp\left\{\mathrm{i}\langle \boldsymbol{m}|\boldsymbol{u}\rangle - \frac{1}{2}\langle \boldsymbol{u}|\mathrm{C}\boldsymbol{u}\rangle\right\},$$

où C est la matrice inverse de A. Un calcul direct permet de montrer que C est la matrice de covariance de $\mathbf{X}$, tandis que $\boldsymbol{m}$ est son espérance. La matrice C est définie positive.

Le cas dégénéré est traité de manière identique, on autorise simplement la matrice C à n'être que symétrique positive, sans être forcément définie positive ; on verra que dans ce cas, $\mathbf{X}$ n'admet pas de densité.

Définition 18.12 Soit $\mathbf{X}$ un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. On dit que $\mathbf{X}$ est un **vecteur gaussien** si sa fonction caractéristique est de la forme

$$\varphi_{\mathbf{X}}(\boldsymbol{u}) = \exp\left\{\mathrm{i}\langle \boldsymbol{m}|\boldsymbol{u}\rangle - \frac{1}{2}\langle \boldsymbol{u}|\mathrm{C}\boldsymbol{u}\rangle\right\}.$$

Dans ce cas, les paramètres $\boldsymbol{m}$ et C sont l'espérance et la matrice de covariance de $\mathbf{X}$:

$$\mathbf{E}(\mathbf{X}) = \boldsymbol{m} \qquad \text{et} \qquad \mathrm{C}^{\mathbf{X}} = \mathrm{C}.$$

Ces deux paramètres caractérisent entièrement la loi de $\mathbf{X}$, notée $\mathscr{N}(\boldsymbol{m}\,;\,\mathrm{C})$, que l'on appelle parfois **distribution normale multivariée**.

§ 271 **Caractérisation par les combinaisons linéaires**

Le théorème qui suit permet de caractériser les vecteurs gaussiens.

■ **Théorème 18.13** *Soit $\mathbf{X} = (\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_d)$ un vecteur aléatoire. Alors $\mathbf{X}$ est un vecteur gaussien si et seulement si toute combinaison linéaire $\sum_{i=1}^{d} a_i \mathrm{X}_i$ est une variable gaussienne.*

DÉMONSTRATION : Supposons que toute combinaison linéaire $\langle \boldsymbol{a}|\mathbf{X}\rangle = \sum a_i \mathrm{X}_i$ soit gaussienne. Notons $\mathrm{V} = \mathrm{C}^{\mathbf{X}}$ la matrice de covariance de $\mathbf{X}$ et $\boldsymbol{m} = \mathbf{E}(\mathbf{X})$ son espérance. Par définition d'une fonction caractéristique, pour tout vecteur $\boldsymbol{u} \in \mathbb{R}^d$,

$$\varphi_{\mathbf{X}}(\boldsymbol{u}) = \mathbf{E}\left(\mathrm{e}^{\mathrm{i}\langle \mathbf{X}|\boldsymbol{u}\rangle}\right) = \varphi_{\langle \mathbf{X}|\boldsymbol{u}\rangle}(1).$$

Or la variable aléatoire $\mathrm{Y} = \langle \mathbf{X}|\boldsymbol{u}\rangle$ étant gaussienne par hypothèse, sa fonction caractéristique est

$$\varphi_{\langle \mathbf{X}|\boldsymbol{u}\rangle}(t) = \mathrm{e}^{\mathrm{i}\mu t - \sigma^2 t^2/2}$$

où μ est son espérance et σ^2 sa variance. Par linéarité,

$$\mu = \mathbf{E}\big(\langle \mathbf{X}|\boldsymbol{u}\rangle\big) = \langle \boldsymbol{m}|\boldsymbol{u}\rangle$$

et, par le théorème 15.6 page 334, sa variance est

$$\mathbf{V}\left(\langle \mathbf{X}|\boldsymbol{u}\rangle\right) = \langle \boldsymbol{u}|\mathrm{C}\boldsymbol{u}\rangle .$$

Ainsi,

$$\varphi_{\mathbf{X}}(\boldsymbol{u}) = \varphi_{\langle \mathbf{X}|\boldsymbol{u}\rangle}(1) = \exp\left\{ \mathrm{i}\langle \boldsymbol{m}|\boldsymbol{u}\rangle - \frac{1}{2}\langle \boldsymbol{u}|\mathrm{C}\boldsymbol{u}\rangle \right\}.$$

Réciproquement, supposons que $\mathbf{X}$ soit un vecteur gaussien. Soit $\boldsymbol{a} = (a_1, a_2, \ldots, a_d)$ un vecteur de $\mathbb{R}^d$, et construisons la combinaison linéaire $\mathrm{Y} = \langle \boldsymbol{a}|\mathbf{X}\rangle = \sum a_i \mathrm{X}_i$. La fonction caractéristique de Y est

$$\varphi_{\mathrm{Y}}(t) = \mathbf{E}(\mathrm{e}^{\mathrm{i}t\mathrm{Y}}) = \mathbf{E}\big(\mathrm{e}^{\mathrm{i}\langle t\boldsymbol{a}|\mathbf{X}\rangle}\big) = \varphi_{\mathbf{X}}(t\boldsymbol{a}) = \exp\left\{ \mathrm{i}t\langle \boldsymbol{m}|\boldsymbol{a}\rangle - \frac{t^2}{2}\langle \boldsymbol{a}|\mathrm{C}\boldsymbol{a}\rangle \right\}.$$

Comme de plus $\langle \boldsymbol{a}|\mathrm{C}\boldsymbol{a}\rangle$ est positif, la fonction φ_{Y} est bien la fonction caractéristique d'une variable gaussienne, de moyenne $\mu = \langle \boldsymbol{m}|\boldsymbol{a}\rangle$ et de variance $\sigma^2 = \langle \boldsymbol{a}|\mathrm{C}\boldsymbol{a}\rangle$. ∎

■ **Corollaire 18.14** *Soit $\mathbf{X} = (\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n)$ un vecteur gaussien ; alors, pour tout $k \in \{1, 2, \ldots, n\}$ et tout choix d'indices $i_1, i_2, \ldots, i_k$ (distincts ou non), le vecteur $\widetilde{\mathbf{X}} = (\mathrm{X}_{i_1}, \mathrm{X}_{i_2}, \ldots, \mathrm{X}_{i_k})$ est un vecteur gaussien de $\mathbb{R}^k$.*

Indépendance et décorrélation

§ 272

Dans le monde merveilleux des variables gaussiennes, les composantes d'un vecteur sont indépendantes si et seulement si leur corrélation est nulle !

■ **Théorème 18.15** *Soit $\mathbf{X} = (\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_d)$ une variable gaussienne. Alors les vecteurs $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_d$ sont indépendants si et seulement si $\mathbf{Cov}(\mathrm{X}_i, \mathrm{X}_j) = 0$ pour tout $i \neq j$, ce qui revient à dire que leur matrice de covariance est diagonale*

Démonstration : Le sens direct est immédiat : des variables indépendantes sont décorrélées. Réciproquement, si la matrice de covariance est diagonale, cela implique que la fonction caractéristique se factorise sous la forme

$$\varphi_{\mathbf{X}}(u_1, u_2, \ldots, u_d) = \varphi_1(u_1) \cdot \varphi_2(u_2) \cdots \varphi_d(u_d),$$

où l'on a noté

$$\varphi_k(u_k) = \exp\left\{ \mathrm{i}m_k u_k - \frac{1}{2}u_k^2 \sigma_k^2 \right\}$$

avec $m_k = \mathbf{E}(\mathrm{X}_k)$ et $\sigma_k = \sigma_{\mathrm{X}_k}$. D'après le théorème 17.12 page 361, on conclut que $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_d$ sont indépendantes. ∎

Décomposition d'un vecteur gaussien

§ 273

Le théorème qui précède montre qu'un vecteur gaussien peut, au moyen d'une simple rotation, se transformer en un nouveau vecteur dont les composantes sont indépendantes. En effet, notons $\mathrm{C} = \mathrm{C}^{\mathbf{X}}$ la matrice de covariance de $\mathbf{X}$. C'est une matrice symétrique positive, on peut donc la diagonaliser au moyen d'une matrice orthogonale Q :

$$\mathrm{C} = \mathrm{Q}\Delta{}^{\mathrm{t}}\mathrm{Q}$$

où Δ est une matrice diagonale dont les termes diagonaux sont tous positifs ; notons-la

$$\Delta = \begin{pmatrix} \lambda_1^2 & & & \\ & \lambda_2^2 & & \\ & & \ddots & \\ & & & \lambda_d^2 \end{pmatrix}.$$

On peut toujours choisir la matrice Q de déterminant $+1$ (quitte, par exemple, à changer le signe de sa dernière colonne) ; c'est donc une matrice de rotation de l'espace $\mathbb{R}^d$. Posons alors

$$\mathbf{Y} = {}^{\mathrm{t}}\mathrm{Q} \cdot (\mathbf{X} - \boldsymbol{m}). \tag{18.6}$$

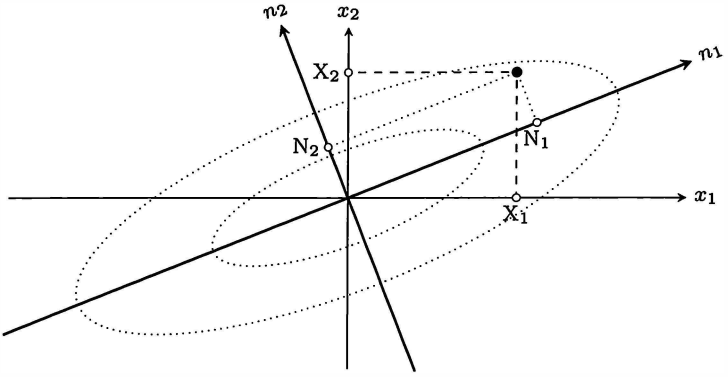


FIGURE 18.3 — *Directions propres de la matrice de covariance d'un couple gaussien centré. Dans le repère $(0, n_1, n_2)$, les coordonnées d'un vecteur sont les valeurs de N_1 et N_2. On a représenté également deux ellipses d'égale densité.*

Puisque $\mathbf{X}$ est gaussien, $\mathbf{X}' = (\mathbf{X} - \boldsymbol{m})$ est encore gaussien et centré. De plus, $\mathbf{Y}$ est lui-même gaussien car pour tout vecteur $\boldsymbol{a} \in \mathbb{R}^d$,

$$\langle \boldsymbol{a} | \mathbf{Y} \rangle = \langle \boldsymbol{a} | {}^{\mathrm{t}}\mathrm{Q}\,\mathbf{X}' \rangle = \langle \mathrm{Q}\,\boldsymbol{a} | \mathbf{X}' \rangle$$

est combinaison linéaire des composantes $X'_1, X'_2, \ldots, X'_d$ de $\mathbf{X}'$, donc gaussien.

Enfin, la matrice de covariance de $\mathbf{Y}$,

$$\mathrm{C}^{\mathbf{Y}} = {}^{\mathrm{t}}\mathrm{Q}\,\mathrm{C}\,\mathrm{Q} = \Delta$$

est diagonale, ce qui prouve que les composantes $Y_1, Y_2, \ldots, Y_d$ de $\mathbf{Y}$ sont indépendantes. Chaque composante suit une loi normale $\mathcal{N}(0\,;\,\lambda_i^2)$. On a donc prouvé que :

■ **Proposition 18.16** *Si $\mathbf{X}$ est un vecteur gaussien d'espérance $\boldsymbol{m}$, on peut l'écrire sous la forme $\mathbf{X} = \mathrm{Q}\mathbf{Y} + \boldsymbol{m}$, où Q est une matrice orthogonale (matrice de rotation de $\mathbb{R}^d$) et $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_d)$ est un vecteur formé de variables aléatoires réelles, indépendantes et de lois respectives $\mathcal{N}(0\,;\,\lambda_k)$, où les λ_k^2 sont les éléments du spectre de la matrice de covariance $\mathrm{C}^{\mathbf{X}}$.*

Supposons de plus que la matrice de covariance de $\mathbf{X}$ est inversible, c'est-à-dire que $\lambda_k > 0$ pour $k = 1, \ldots, d$. Posons $N_k = \lambda_k^{-1} Y_k$, de telle sorte que les N_k suivent toutes la loi $\mathcal{N}(0\,;\,1)$ et soient indépendantes. On a alors

$$\mathbf{X} = \mathrm{Q}\mathbf{Y} + \boldsymbol{m} = (\mathrm{Q}\sqrt{\Delta})\,\mathbf{N} + \boldsymbol{m},$$

où l'on a noté $\sqrt{\Delta} = \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_d)$.

Sur la figure 18.3, en plus des ellipses d'égale densité d'un vecteur $\mathbf{X}$ de $\mathbb{R}^2$ (déjà illustrées à la figure 18.2), on représente les *directions* associées aux vecteurs N_1 et N_2 ; ce sont les axes principaux des ellipses.

Remarque 18.17 Réciproquement, si l'on se donne une collection $N_1, N_2, \ldots, N_d$ de variables normales de loi $\mathcal{N}(0\,;\,1)$ et mutuellement indépendantes, alors les variables $Y_k = \lambda_k N_k$ suivent les lois $\mathcal{N}(0\,;\,\lambda_k^2)$ et sont indépendantes, donc $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_d)$ est un vecteur gaussien de matrice de covariance Δ et

$$\widetilde{\mathbf{X}} = \mathrm{Q}\mathbf{Y} + \boldsymbol{m} = (\mathrm{Q}\sqrt{\Delta})\,\mathbf{N} + \boldsymbol{m}$$

est un vecteur gaussien de même espérance $\boldsymbol{m}$ et de même matrice de covariance C que $\mathbf{X}$; ainsi, $\widetilde{\mathbf{X}}$ est de même loi que $\mathbf{X}$.

Bien sûr, puisque nous sommes partis de variables aléatoires N_k dont *seule la loi était précisée*, il n'y a aucune raison pour que $\widetilde{\mathbf{X}}$ soit égale à $\mathbf{X}$, alors que l'équation (18.6) était une vraie égalité, pas une égalité en loi. Cette propriété est cependant utile pour simuler, informatiquement, un vecteur gaussien de loi $\mathscr{N}(\boldsymbol{m}\,;\,\mathrm{C})$ donnée ; elle sera mise en œuvre au chapitre 25.

Densité d'un vecteur gaussien

§ 274

Dans le cas où la matrice de covariance C du vecteur $\mathbf{X}$ est inversible, c'est-à-dire de rang r, il est aisé de constater que $\mathbf{X}$ admet une densité. En effet, on peut l'écrire sous la forme

$$\mathbf{X} = \mathrm{A}\,\mathbf{N} + \boldsymbol{m},$$

où $\mathrm{A} = \mathrm{Q}\sqrt{\Delta}$ est une matrice inversible. Or, le vecteur $\mathbf{N}$ admet la densité

$$f(y_1, y_2, \ldots, y_d) = \frac{1}{(2\pi)^{d/2}}\,\mathrm{e}^{-(y_1^2+\cdots+y_d^2)/2}.$$

Puisque la transformation $\boldsymbol{x} \mapsto \mathrm{A}\,\boldsymbol{x} + \boldsymbol{m}$ est un $\mathscr{C}^1$-difféomorphisme de $\mathbb{R}^2$ (dont le jacobien vaut $\det \mathrm{A} = \lambda_1 \cdots \lambda_n$), en vertu du théorème 15.10 page 337 le vecteur $\mathbf{X}$ admet pour densité

$$\begin{aligned} f_{\mathbf{X}}(x_1, x_2, \ldots, x_d) &= \frac{1}{(2\pi)^{d/2}\,\lambda_1 \cdots \lambda_d}\mathrm{e}^{-\langle \mathrm{A}^{-1}(\boldsymbol{x}-\boldsymbol{m})\,|\,\mathrm{A}^{-1}(\boldsymbol{x}-\boldsymbol{m})\rangle/2} \\ &= \frac{1}{(2\pi)^{d/2}\,\lambda_1 \cdots \lambda_d}\mathrm{e}^{-\langle (\boldsymbol{x}-\boldsymbol{m})\,|\,\mathrm{C}^{-1}(\boldsymbol{x}-\boldsymbol{m})\rangle/2} \end{aligned}$$

puisque $\mathrm{A}^{\mathrm{t}}\mathrm{A} = \mathrm{C}$. (Voir également § 377.)

Dans le cas dégénéré, c'est-à-dire si au contraire la matrice $\mathrm{C}^{\mathbf{X}}$ n'est pas inversible, alors le vecteur $\mathbf{X}$ vit dans un sous-espace affine de dimension strictement inférieure à d. En effet, il existe un vecteur $\boldsymbol{a}$ non nul tel que $\mathrm{C}^{\mathbf{X}}\boldsymbol{a} = 0$. La variance de $\langle \boldsymbol{a}|\mathbf{X}\rangle$ est (théorème 15.6) $\langle \boldsymbol{a}|\mathrm{C}^{\mathbf{X}}\boldsymbol{a}\rangle = 0$, ce qui veut dire que la variable $\langle \boldsymbol{a}|\mathbf{X}\rangle$ est presque sûrement constante et égale à son espérance. En d'autres termes, il existe une relation linéaire $\sum a_i\,\mathrm{X}_i$ presque sûrement constante — ou, en termes géométriques : la variable $\mathbf{X}$ vit (presque sûrement) dans l'hyperplan affine passant par $\boldsymbol{m}$ et orthogonal au vecteur $\boldsymbol{a}$. Si $\mathbf{X}$ admettait une densité, et puisque la mesure d'un hyperplan est nulle[(5)], on en déduirait que $\mathbf{P}\{\mathrm{X} \in \mathrm{H}\} = 0$, ce qui est absurde. Donc $\mathbf{X}$ n'admet pas de densité.

■ **Théorème 18.18 (Densité d'un vecteur gaussien)** *Un vecteur gaussien* $\mathbf{X}$ *admet une densité si et seulement si sa matrice de covariance* $\mathrm{C} = \mathrm{C}^{\mathbf{X}}$ *est inversible. Si c'est le cas, sa densité est*

$$f_{\mathbf{X}}(x_1, x_2, \ldots, x_d) = \frac{1}{(2\pi)^{d/2}\,\sqrt{\det(\mathrm{C})}}\,\mathrm{e}^{-q(\boldsymbol{x})/2}$$

où $q(\boldsymbol{x}) = \langle \boldsymbol{x} - \boldsymbol{m}\,|\,\mathrm{C}^{-1}(\boldsymbol{x} - \boldsymbol{m})\rangle$.

(5). Par rapport à la mesure de Lebesgue de $\mathbb{R}^d$.

§ 275 **Ellipsoïdes d'égale densité et loi du χ^2**

Supposons la forme quadratique q non dégénérée : le vecteur gaussien $\mathbf{X}$ admet alors une densité. Les ellipsoïdes $\mathscr{E}_k$ d'équation $q(x) = k^2$ sont les **ellipsoïdes d'égale densité**, la densité de probabilité y prenant effectivement la valeur constante $(2\pi)^{-d/2}(\det \mathrm{C})^{-1/2}\,\mathrm{e}^{-k^2/2}$. Tout comme dans le cas bidimensionnel, le contenu en probabilité de ces ellipses est une fonction de k seulement ; on le calcule par le changement de variable $x' = \mathrm{A}^{-1}(x - m)$, ce qui mène à

$$\begin{aligned}\mathbf{P}\{\mathbf{X} \in \mathscr{E}_k\} &= \mathbf{P}\left\{\langle \mathbf{N}|\mathbf{N}\rangle \leqslant k^2\right\} \\ &= \frac{1}{(2\pi)^{d/2}} \int \cdots \int_{\mathrm{D}_k} \mathrm{e}^{-(y_1^2+y_2^2+\cdots+y_d^2)/2}\,\mathrm{d}y_1 \cdots \mathrm{d}y_d,\end{aligned}$$

où D_k est la sphère centrée à l'origine et de rayon r. Une intégration en coordonnées « sphériques » transforme cette intégrale en

$$\mathbf{P}\{q(\mathrm{X}) \leqslant k^2\} = \mathbf{P}\{\mathbf{X} \in \mathscr{E}_k\} = \frac{1}{2^{d/2}\,\Gamma(d/2)} \int_0^{k^2} \mathrm{e}^{-t/2}\, t^{d/2-1}\,\mathrm{d}t.$$

On dit que la quantité $q(\mathrm{X}) = \langle \mathbf{N}|\mathbf{N}\rangle = \mathrm{N}_1^2 + \cdots + \mathrm{N}_d^2$ suit la **loi du χ^2 à d degrés de liberté**.

Remarque 18.19 Dans le cas où la loi gaussienne est dégénérée, c'est-à-dire si le rang de la matrice de covariance est $r < d$, alors le contenu en probabilité de l'ellipsoïde $\mathscr{E}_k$ suit la loi du χ^2 à r degrés de liberté.

§ 276 **Lemme des moments (théorème de Wick)**

Les moments d'ordre 2 (c'est-à-dire la matrice de covariance) d'une distribution gaussienne centrée déterminent entièrement sa loi, et donc les valeurs de ses moments de tous ordres. Le lemme des moments donne une expression explicite des moments d'ordre k.

■ **Théorème 18.20 (Lemme des moments)** *Soit $(\mathrm{X}_1, \ldots, \mathrm{X}_n)$ un vecteur gaussien centré de $\mathbb{R}^n$. Soient p un entier et $k_1, \ldots, k_p$ des indices de $\{1, \ldots, n\}$.*

- *Si p est impair, alors $\mathbf{E}(\mathrm{X}_{k_1}\mathrm{X}_{k_2}\ldots\mathrm{X}_{k_p}) = 0$.*
- *Si p est pair, alors*

$$\mathbf{E}(\mathrm{X}_{k_1}\mathrm{X}_{k_2}\ldots\mathrm{X}_{k_p}) = \sum_{\text{paires}} \mathbf{E}(\mathrm{X}_{i_1}\mathrm{X}_{j_1}) \cdot \mathbf{E}(\mathrm{X}_{i_2}\mathrm{X}_{j_2}) \cdots \mathbf{E}(\mathrm{X}_{i_k}\mathrm{X}_{j_k}),$$

où la sommation a lieu sur toutes les manières distinctes de décomposer le p-uplet d'entiers $(k_1, k_2, \ldots, k_p)$ en $p/2$ paires[6] disjointes $\{i_1, j_1\}, \ldots, \{i_k, j_k\}$.

Ce théorème a des applications cruciales en mécanique quantique et en théorie des champs ; il est connu des physiciens sous le nom de *théorème de Wick*[7]. L'usage de ce théorème est simplifié par une notation pratique, le *symbole de contraction* « $\overbrace{\ \ }$ » qui indique le calcul d'une espérance. L'espérance d'une quantité étant notée entre crochets $\langle\cdots\rangle$, le théorème de Wick s'écrit, pour $p = 4$ et des choix d'indices $\mathrm{X}_1\mathrm{X}_2\mathrm{X}_3\mathrm{X}_4$:

$$\langle \mathrm{X}_1\mathrm{X}_2\mathrm{X}_3\mathrm{X}_4\rangle = \langle \mathrm{X}_1\mathrm{X}_2\rangle\,\langle \mathrm{X}_3\mathrm{X}_4\rangle + \langle \mathrm{X}_1\mathrm{X}_3\rangle\,\langle \mathrm{X}_2\mathrm{X}_4\rangle + \langle \mathrm{X}_1\mathrm{X}_4\rangle\,\langle \mathrm{X}_2\mathrm{X}_3\rangle$$

ce qui s'écrit sous la forme de la somme sur toutes les contractions possibles :

$$\langle \mathrm{X}_1\mathrm{X}_2\mathrm{X}_3\mathrm{X}_4\rangle = \langle \overbrace{\mathrm{X}_1\,\mathrm{X}_2}\,\overbrace{\mathrm{X}_3\,\mathrm{X}_4}\rangle + \langle \overbrace{\mathrm{X}_1\,\overbrace{\mathrm{X}_2\,\mathrm{X}_3}\,\mathrm{X}_4}\rangle + \langle \overbrace{\mathrm{X}_1\,\overbrace{\mathrm{X}_2\,\mathrm{X}_3}\,\mathrm{X}_4}\rangle$$

Remarque 18.21 En prenant $n = 1$, on retrouve facilement la valeur des moments d'ordre pair d'une loi normale centrée réduite. Par exemple,

$$\mathbf{E}(\mathrm{X}^4) = \langle \mathrm{X}_1\mathrm{X}_1\mathrm{X}_1\mathrm{X}_1\rangle = \langle \overbrace{\mathrm{X}_1\,\mathrm{X}_1}\,\overbrace{\mathrm{X}_1\,\mathrm{X}_1}\rangle + \langle \overbrace{\mathrm{X}_1\,\overbrace{\mathrm{X}_1\,\mathrm{X}_1}\,\mathrm{X}_1}\rangle + \langle \overbrace{\mathrm{X}_1\,\overbrace{\mathrm{X}_1\,\mathrm{X}_1}\,\mathrm{X}_1}\rangle = 3\,\langle \mathrm{X}^2\rangle^2 = 3.$$

De même, $\mathbf{E}(\mathrm{X}^6) = 15\,\langle \mathrm{X}^2\rangle^3 = 15$ (il y a 15 pairages possibles dans un ensemble à 6 éléments). Une analyse combinatoire simple permet de retrouver les résultats de l'exercice 11.6 page 254 : $\mathbf{E}(\mathrm{X}^{2n}) = (2n-1)\cdot(2n-3)\cdots 3\cdot 1 = (2n)!/2^n\,n!$.

---

(6). Les indices $k_1, \ldots, k_p$ doivent être rendus artificiellement distincts, même s'ils ont des valeurs identiques ; la paire $\{3, 3\}$ est donc considérée comme valide. En revanche, les paires $\{3, 4\}$ et $\{4, 3\}$ sont identiques et ne sont comptées qu'une seule fois.

(7). Le théorème des physiciens est plus compliqué et fait intervenir des champs quantiques [52], mais les calculs de fonctions de corrélation à plusieurs points peuvent être mis en parallèle avec les calculs d'espérances du lemme des moments.

EXERCICES

♦ ***Exercice 18.1 (Théorème de Cochran)*** Soit $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vecteur gaussien de loi centrée réduite $\mathcal{N}(\mathbf{0}\ ;\ I_n)$ sur l'espace $\mathbb{R}^n$, muni de sa base canonique $(e_1, e_2, \dots, e_n)$.

On choisit $(d_1, d_2, \dots, d_k)$ des entiers positifs tels que $d_1 + d_2 + \cdots + d_k = n$ et on note F_1 l'espace engendré par les d_1 premiers vecteurs, F_2 celui engendré par les d_2 suivants,... et F_k celui engendré par les d_k derniers, de sorte que

$$E = F_1 \overset{\perp}{\oplus} F_2 \overset{\perp}{\oplus} \dots \overset{\perp}{\oplus} F_k.$$

Enfin, on note $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k$ les projections orthogonales de $\mathbf{X}$ sur les sous-espaces respectifs $F_1, F_2, \dots, F_k$.

i) Déterminer la loi de $\|\mathbf{X}_i\|^2$ pour $i = 1, \dots, k$.

ii) En déduire le résultat suivant (on prononce [kokran]) :

■ **Théorème 18.22 (Cochran)** *Soient $F_1, F_2, \dots, F_k$ des sous-espaces vectoriels de $E = \mathbb{R}^n$, en somme directe orthogonale. Si $\mathbf{X} \rightsquigarrow \mathcal{N}(\mathbf{0}\ ;\ I_n)$ et si l'on note $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k$ les projections orthogonales de $\mathbf{X}$ sur chacun de ces sous-espaces vectoriels, alors*

$$\|\mathbf{X}_i\|^2 \rightsquigarrow \chi^2_{d_i}$$

où $d_i = \dim F_i$ pour $i = 1, \dots, k$.

♦ ***Exercice 18.2*** En n'utilisant que les formules classiques de l'espérance et de la variance d'une somme, retrouver rapidement les paramètres de la loi (normale) de $Z = aX + bY$ lorsque (X, Y) est un vecteur gaussien de paramètres $m_1, m_2, \sigma_1, \sigma_2$ et ρ.

♦ ***Exercice 18.3*** Soit $\mathbf{X} = (X, Y)$ un vecteur gaussien de matrice de covariance

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

et de moyenne $\mathbf{E}(\mathbf{X}) = (m_1, m_2)$. Montrer que

$$\mathbf{P}\{X \geqslant m_1, Y \geqslant m_2\} = \frac{1}{4} + \frac{1}{2\pi} \operatorname{Arc\,sin} \rho.$$

Interpréter graphiquement ce résultat pour les valeurs extrêmes $\rho = -1$, $\rho = 0$ et $\rho = +1$.

♦ ***Exercice 18.4*** Soit (X, Y) un couple gaussien, caractérisé par des écarts-types σ_x et σ_y, ainsi qu'un coefficient de corrélation ρ. Montrer qu'il existe des coefficients a et b tels que $\mathbf{E}(Y \,\|\, X) = aX + b$. (On peut utiliser la définition de l'espérance conditionnelle donnée au paragraphe § 192.)

SOLUTIONS

♦ **Solution 18.1**

i) Puisque $\mathbf{X}_1 = (X_1, \dots, X_{d_1}, 0, \dots, 0)$, la quantité $\|\mathbf{X}_1\|^2$ est la somme de d_1 carrés de variables normales centrées réduites indépendantes ; ainsi, $\|\mathbf{X}_1\|^2$ suit la loi du khi-deux à d_1 degrés de liberté (voir remarque 10.18 page 235).

De même, $\|\mathbf{X}_i\|^2$ suit la loi $\chi^2_{d_i}$.

ii) On se ramène au cas précédent en choisissant une base orthonormée directe adaptée à la décomposition $E = F_1 \overset{\perp}{\oplus} \cdots \overset{\perp}{\oplus} F_k$ et en notant que la loi $\mathcal{N}(\mathbf{0}\ ;\ I_n)$ est invariante par rotation : si P est une matrice orthogonale, le vecteur $P\mathbf{X}$ est toujours gaussien, de matrice de covariance $C^{P\mathbf{X}} = PC^{\mathbf{X}}{}^tP = PI_n{}^tP = I_n$.

♦ **Solution 18.2** Par linéarité, l'espérance de $Z = aX + bY$ est $\mathbf{E}(Z) = am_1 + bm_2$. Sa variance est

$$\mathbf{V}(aX + bY) = a^2\,\mathbf{V}(X) + b^2\,\mathbf{V}(Y) + 2ab\,\mathbf{Cov}(X, Y) = a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2.$$

Ainsi, $Z \rightsquigarrow \mathcal{N}\left(am_1 + bm_2\ ;\ a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2\right)$.

♦ **Solution 18.3** Notons $f(\rho)$ la probabilité cherchée (nous verrons qu'elle ne dépend effectivement d'aucun autre paramètre que ρ). Après le changement de variable

$$u = \frac{x - m_1}{\sigma_1} \qquad \text{et} \qquad v = \frac{y - m_2}{\sigma_2},$$

on obtient

$$f(\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_0^{+\infty} \int_0^{+\infty} \mathrm{e}^{-(u^2 - 2\rho uv + v^2)/2(1-\rho^2)}\, \mathrm{d}u\, \mathrm{d}v$$

On met l'argument de l'exponentielle sous forme canonique, ce qui mène à

$$\begin{aligned} f(\rho) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_0^{+\infty} \left(\int_0^{+\infty} \mathrm{e}^{-(u-\rho v)^2/2(1-\rho^2)}\, \mathrm{d}u \right) \mathrm{e}^{-v^2/2}\, \mathrm{d}v \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_0^{+\infty} \left(\int_{-\rho v}^{+\infty} \mathrm{e}^{-u^2/2(1-\rho^2)}\, \mathrm{d}u \right) \mathrm{e}^{-v^2/2}\, \mathrm{d}v. \end{aligned}$$

Un dernier changement de variable $t = u/\sqrt{1-\rho^2}$ donne

$$f(\rho) = \int_0^{+\infty} \left[1 - \mathfrak{N}\left(-\frac{\rho v}{\sqrt{1-\rho^2}} \right) \right] \mathfrak{n}(v)\, \mathrm{d}v.$$

Le plus simple maintenant est sans doute de dériver sous le signe somme (en vérifiant soigneusement la domination uniforme de la dérivée partielle de l'intégrande) :

$$f'(\rho) = \frac{\mathrm{d}}{\mathrm{d}\rho} \left[\frac{\rho}{\sqrt{1-\rho^2}} \right] \cdot \int_0^{+\infty} v \cdot \mathfrak{n}\left(\frac{\rho v}{\sqrt{1-\rho^2}} \right) \mathfrak{n}(v)\, \mathrm{d}v$$

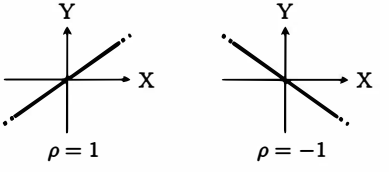
et cette intégrale gaussienne se calcule directement (remplacer la fonction $\mathfrak{n}$ par son expression) :

$$f'(\rho) = \frac{1}{2\pi} \frac{1}{\sqrt{1-\rho^2}}.$$

On vérifie ensuite que $f(0) = 1/4$: lorsque $\rho = 0$, les vecteurs X et Y sont indépendants. Il s'ensuit que

$$\mathbf{P}\{\mathrm{X} \geqslant m_1, \mathrm{Y} \geqslant m_2\} = \frac{1}{4} + \frac{1}{2\pi} \operatorname{Arc\,sin} \rho.$$

Quant à l'interprétation graphique : si $\rho = 1$, les vecteurs X et Y sont (presque sûrement) liés positivement et ne prennent leurs valeurs que sur une diagonale du plan (voir figure ci-dessous), donc $f(1) = \mathbf{P}\{\mathrm{X} \geqslant m_1, \mathrm{Y} \geqslant m_2\} = \mathbf{P}\{\mathrm{X} \geqslant m_1\} = 1/2$. Si $\rho = -1$, ils sont négativement liés, ce qui fait que les événements $\{\mathrm{X} \geqslant m_1\}$ et $\{\mathrm{Y} \geqslant m_2\}$ sont presque incompatibles et $f(-1) = 0$.



$\rho = 1$ $\qquad$ $\rho = -1$

♦ **Solution 18.4** La densité conditionnelle $f(y \mid x)$ de Y, calculée au paragraphe § 268, est celle d'une loi normale centrée en

$$m(y \mid x) = m_y + \frac{\rho\sigma_y}{\sigma_x}(x - m_x).$$

Par conséquent,

$$\mathbf{E}(\mathrm{Y} \mid \mathrm{X} = x) = m(y \mid x) = m_y + \frac{\rho\sigma_y}{\sigma_x}(x - m_x)$$

et donc

$$\mathbf{E}(\mathrm{Y} \,\|\, \mathrm{X}) = m_y + \frac{\rho\sigma_y}{\sigma_x}(\mathrm{X} - m_x).$$

Quatrième partie

Théorèmes limites

Chapitre 19

Convergence

On introduit dans ce chapitre les principales notions de convergence de suites de variables aléatoires : convergence presque sûre, convergence en probabilité, convergence en loi. Les relations, parfois subtiles, entre ces notions, sont développées, et des critères simples les caractérisant sont exposés.

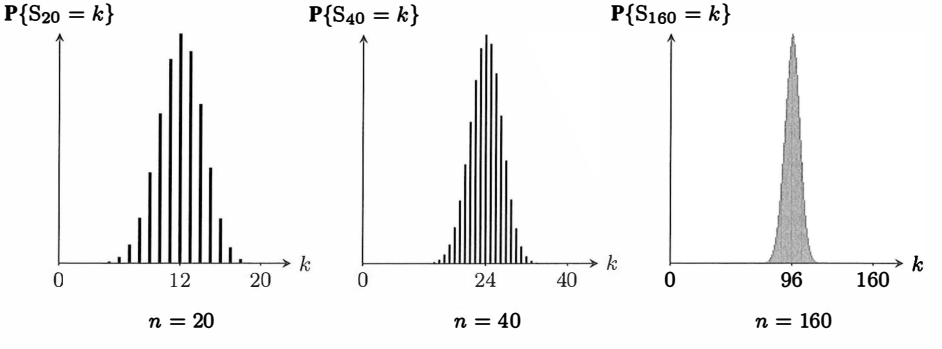
Ce chapitre est, par essence, très technique. Il est cependant important de s'y plonger un temps suffisant pour bien se pénétrer des techniques exposées. Plus que tout autre chapitre, celui-ci doit être lu un crayon à la main. Le minimum syndical *à connaître pour pouvoir aborder les lois des grands nombres et le théorème limite central sont indiqués par un ou deux symboles « ◀ » en marge.*

Afin de faire oublier un peu l'aridité du sujet, on termine par deux applications simples. Dans la première, on montre que, dans les espaces euclidiens de grande dimension, presque tous les vecteurs sont presque orthogonaux. *Dans la seconde, on étudie une propriété caractérisant les suites de réels dont la partie décimale est équidistribuée sur* $[0\,;1[$.

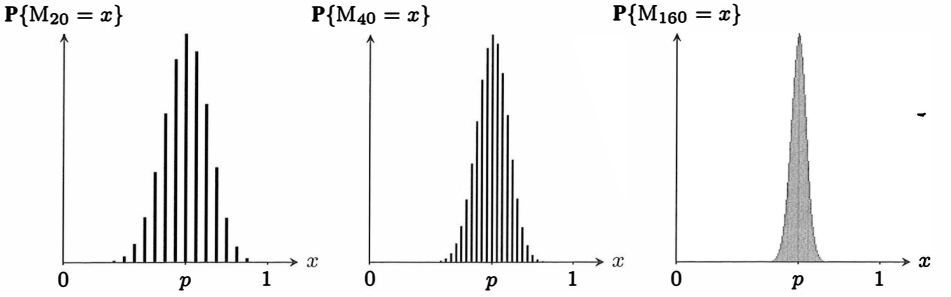
Introduction aux théorèmes limites

§ 277

Nous avons déjà rencontré plusieurs résultats typiques de ce que l'on appelle des *théorèmes limites*, notamment dans le cas d'une succession infinie de tirages à *pile* ou *face*. Ci-dessous, nous représentons l'histogramme de la loi du nombre de *pile* obtenu lors de n tirages : $S_n = X_1 + X_2 + \cdots + X_n$, où X_k est la variable valant 1 si le k-ième tirage amène *pile* et 0 sinon (la probabilité d'obtenir *pile* étant ici $p = 0,6$).



Dans cette succession d'histogrammes, nous avons naturellement divisé l'échelle des ordonnées par n, parce que S_n prend ses valeurs dans $[0\,;\,n]$, et pour une raison de mise en page (on voulait que les schémas aient la même taille). Ce choix revient à considérer non pas la variable S_n, mais $M_n := S_n/n$, c'est-à-dire la moyenne de n tirages indépendants. L'expérience — celle qui est à la base de la notion même de probabilité — montre que M_n « devrait tendre » vers p, ce qu'illustre le resserrement du pic.



Ce « devrait tendre » peut recevoir plusieurs interprétations.

- La première est l'*interprétation déterministe* : pour toute réalisation ω du hasard, on a $M_n(\omega) \to p$.
  Cette interprétation est évidemment totalement inadéquate avec notre modélisation : dans l'espace Ω, on trouve notamment la suite

$$\omega_1 = \text{« FFFFFFFFFF}\cdots\text{ »},$$

  pour laquelle $M_n(\omega) = 0$ pour tout n, ainsi que la suite

$$\omega_2 = \text{« PPPPPPPPPP}\cdots\text{ »},$$

  pour laquelle $M_n(\omega) = 1$ pour tout n. Ainsi, quelle que soit la valeur de p, on peut trouver un élément de Ω qui ne vérifie pas la condition « $M_n(\omega) \to p$ ».
- La seconde est l'*interprétation forte* : avec une probabilité 1, la suite de terme général $M_n(\omega)$ va tendre vers p, ou encore

$$\mathbf{P}\big\{\omega \in \Omega\,;\ M_n(\omega) \to p\big\} = 1.$$

  À cette interprétation forte est associée la notion de *convergence presque sûre* : à moins d'un manque de chance incroyable, on peut être positivement certain que la suite observée des moyennes successives converge vers p. Ce résultat est connu sous le nom de loi (forte) des grands nombres[1].
- La troisième interprétation, qui est l'*interprétation faible*, s'écrit

$$\forall \varepsilon > 0 \qquad \mathbf{P}\big\{\omega \in \Omega\,;\ \big|M_n(\omega) - p\big| > \varepsilon\big\} \xrightarrow[n\to\infty]{} 0$$

  et signifie que l'ensemble des moyennes qui restent écartées de p de plus que ε se réduit comme peau de chagrin quand n tend vers l'infini. L'interprétation intuitive de limite est plus délicate, puisqu'elle ne dit rien à ω fixé, mais n'impose qu'une série de conditions globales, valables chacune pour un ε donné. À cette interprétation faible est associée la notion de *convergence en probabilité*.

(1). Et on notera dès à présent que l'essence d'un théorème limite est, à partir d'un modèle basé sur le hasard, d'établir des *certitudes*, au sens probabiliste du terme, c'est-à-dire une propriété vraie avec une probabilité 1.

- Enfin, il est intéressant d'étudier la convergence dans l'espace L^q des variables aléatoires admettant un moment d'ordre q : a-t-on $\mathbf{E}\,|\mathrm{M}-p|^q \to 0$? Si l'interprétation de cette propriété est moins immédiatement claire, le cas $q=2$ est particulièrement intéressant puisque l'on bénéficie des puissants outils de la géométrie des espaces de Hilbert.

On peut aller plus loin que la propriété : « (presque sûrement), la suite de terme général $\mathrm{M}_n = \mathrm{S}_n/n$ va converger vers p ». On peut s'intéresser aux *fluctuations* autour de la valeur limite p, ce qui revient à dilater l'échelle d'observation d'un facteur convenable. Les fluctuations de M_n autour de sa valeur moyenne se font sur une échelle dont l'ordre de grandeur est l'écart-type $\boldsymbol{\sigma}_{\mathrm{M}_n} = \boldsymbol{\sigma}_{\mathrm{X}}/\sqrt{n}$, on considère donc la variable réduite $\mathrm{M}_n/\boldsymbol{\sigma}_{\mathrm{M}_n}$ — ou, plus précisément, et pour neutraliser l'effet de la divergence vers $+\infty$ de son espérance, la variable centrée réduite

$$\mathrm{M}_n^* = \mathrm{S}_n^* := \frac{\mathrm{X}_1 + \mathrm{X}_2 + \cdots + \mathrm{X}_n - n\,\mathbf{E}(\mathrm{X})}{\sqrt{n}\,\boldsymbol{\sigma}_{\mathrm{X}}}.$$

On ne peut pas espérer de résultat du même type que le précédent pour le comportement limite de M_n^*, puisque par définition ses fluctuations sont de l'ordre de l'unité. Tout ce que l'on peut espérer, c'est une estimation de la répartition de ces fluctuations, c'est-à-dire une loi limite. La notion de convergence associée est celle de *convergence en loi*, et le théorème régissant cette convergence est le *théorème limite central.*

Dans ce chapitre, nous passons en revue les différents modes de convergence (en probabilité, presque sûre et en loi), tandis que les trois prochains chapitres sont consacrés à l'étude des principaux théorèmes limites (loi du 0–1, loi des grands nombres, théorème limite central) et à leurs applications dans les techniques d'approximations.

19.1 CONVERGENCE PRESQUE SÛRE, CONVERGENCE EN PROBABILITÉ

Dans toute cette section, on considère des variables $\mathrm{X}_1, \mathrm{X}_2, \ldots$ et X, toutes définies sur le même espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$.

Convergence presque sûre § 278

On rappelle que la suite $(\mathrm{X}_n)_{n\geqslant 1}$ **converge simplement vers X** lorsque

$$\forall \omega \in \Omega \qquad \lim_{n\to\infty} \mathrm{X}_n(\omega) = \mathrm{X}(\omega).$$

Cette notion est trop forte dans la plupart des situations.

Exemple 19.1 Reprenons en effet la suite $(\mathrm{M}_n)_{n\geqslant 1}$ de l'introduction (§ 277) : il n'y a pas convergence simple de $(\mathrm{M}_n)_{n\geqslant 1}$ vers la variable constante $\mathrm{X} : \omega \mapsto p$. Il y a en revanche convergence simple vers X sur une partie stricte $\Omega' = \Omega \setminus \mathrm{N}$ de Ω, où l'on s'est contenté de définir

$$\Omega' = \{\omega \in \Omega\,;\ \mathrm{X}_n(\omega) \to p\}.$$

Une question naturelle est alors : quelle est la taille (pour la mesure $\mathbf{P}$) de l'ensemble Ω' ?

On dit que $(\mathrm{X}_n)_{n\geqslant 1}$ **converge presque sûrement** vers X si l'ensemble des ω tels ◀◀ que $\mathrm{X}_n(\omega) \to \mathrm{X}(\omega)$ est de probabilité 1 :

$$\mathbf{P}\Big\{\omega \in \Omega\,;\ \lim_{n\to\infty} \mathrm{X}_n(\omega) = \mathrm{X}(\omega)\Big\} = 1.$$

On écrit alors « $\mathrm{X}_n \xrightarrow{\text{ps}} \mathrm{X}$ », ou « $\mathrm{X}_n \to \mathrm{X}$ p.s. », ou encore $\mathbf{P}\{\mathrm{X}_n \to \mathrm{X}\} = 1$.

Remarque 19.2 Dit en d'autres termes : si l'on effectue l'expérience, ce qui revient à confier à la Fortune le soin de nous attribuer un ω, on peut être *positivement certain* d'observer que $X_n(\omega) \to X(\omega)$; non que l'inverse soit impossible, mais on a plus de chances de gagner à l'EuroMillion, de recevoir par erreur le prix Nobel de littérature et d'obtenir cent *face* d'affilée avec une pièce honnête, le tout dans la même heure, que de ne pas observer la convergence susdite.

Remarque 19.3 Le fait que $\{X_n \to X\}$ soit un événement sera prouvé rigoureusement au paragraphe § 279.

Bien entendu, dire que $X_n \xrightarrow{ps} X$ est la même chose que de dire $(X_n - X) \xrightarrow{ps} 0$. Enfin, on a la propriété immédiate de linéarité :

■ **Proposition 19.4** *Si* $X_n \xrightarrow{ps} X$ *et* $Y_n \xrightarrow{ps} Y$, *alors* $\alpha X_n + \beta Y_n \xrightarrow{ps} \alpha X + \beta Y$.

Cette propriété reste vraie si les variables sont à valeurs dans $[-\infty\,;+\infty]$, à condition que les combinaisons linéaires aient toutes un sens.

§ 279 **Critères de convergence presque sûre (∗)**

Dans ce paragraphe, nous donnons quelques critères pratiques de convergence presque sûre. *Nous commençons par relier la convergence presque sûre à des probabilités, en montrant que* $X_n \to X$ *p.s. si et seulement si, pour tout* $\varepsilon > 0$, *l'événement* $\{|X_n - X| > \varepsilon\ ;\ i.s.\}$ *est négligeable.*

Il est plus simple en fait de décrire le complémentaire de $\{X_n \to X\}$ en revenant à la définition de la limite d'une suite : dire que $X_n(\omega)$ ne tend pas vers $X(\omega)$, c'est dire qu'il existe un $\varepsilon > 0$ tel que, *pour une infinité d'indices* n, on ait $|X_n(\omega) - X(\omega)| > \varepsilon$. Autrement dit,

$$\{X_n \to X\}^c = \bigcup_{\varepsilon>0} \{\, |X_n - X| > \varepsilon\ ;\ \text{i.s.}\}.$$

Une telle union n'est pas directement utilisable (elle a lieu sur un ensemble non dénombrable), mais on peut évidemment se restreindre à des ε rationnels, ou même à une suite de limite nulle comme $\varepsilon_k = 1/k$:

$$\{X_n \to X\}^c = \bigcup_{\varepsilon\in\mathbb{Q},\ \varepsilon>0} \{\, |X_n - X| > \varepsilon\ ;\ \text{i.s.}\} = \bigcup_{k=1}^{\infty} \{|X_n - X| > \tfrac{1}{k}\ ;\ \text{i.s.}\}$$

Au passage, ces égalités prouvent que : $\{X_n \to X\}$ *est bien un événement de la tribu* $\mathfrak{T}$.

Puisqu'une union dénombrable d'événements est de probabilité nulle si et seulement si tous ces événements sont de probabilité nulle, on en déduit qu'il y a équivalence entre les énoncés suivants :

i) $\mathbf{P}\{X_n \to X\} = 1$;

ii) $\mathbf{P}\big(\{X_n \to X\}^c\big) = 0$;

iii) $\mathbf{P}\{|X_n - X| > \varepsilon\ ;\ \text{i.s.}\} = 0$ pour tout $\varepsilon \in \mathbb{Q}, \varepsilon > 0$;

iv) $\mathbf{P}\{|X_n - X| > \varepsilon\ ;\ \text{i.s.}\} = 0$ pour tout $\varepsilon > 0$.

La propriété annoncée est donc démontrée.

Remarquons que l'événement $\{|X_n - X| > \varepsilon\ ;\ \text{i.s.}\}$ est une intersection décroissante :

$$A(\varepsilon) := \{|X_n - X| > \varepsilon\ ;\ \text{i.s.}\} = \lim \downarrow B_N(\varepsilon),$$

où

$$B_N(\varepsilon) := \bigcup_{n\geqslant N} \{\, |X_n - X| > \varepsilon\} = \{\exists n \geqslant N \quad |X_n - X| > \varepsilon\},$$

ce qui fait que

$$\mathbf{P}\big(\mathrm{A}(\varepsilon)\big) = 0 \iff \lim_{\mathrm{N}\to\infty} \mathbf{P}\big(\mathrm{B}_{\mathrm{N}}(\varepsilon)\big) = 0.$$

En résumé, on a démontré le résultat suivant :

■ **Théorème 19.5 (CNS de convergence presque sûre)** *Il y a équivalence entre les énoncés*

i) $\mathrm{X}_n \to \mathrm{X}$ *p.s. ;*

ii) $\mathbf{P}\{|\mathrm{X}_n - \mathrm{X}| > \varepsilon \text{ ; } i.s.\} = 0$ *pour tout* $\varepsilon > 0$ *;*

iii) $\mathbf{P}\{\exists n \geqslant \mathrm{N} \quad |\mathrm{X}_n - \mathrm{X}| > \varepsilon\} \xrightarrow[\mathrm{N}\to\infty]{} 0$ *pour tout* $\varepsilon > 0$.

iv) $\mathbf{P}\left\{\sup_{n\geqslant \mathrm{N}} |\mathrm{X}_n - \mathrm{X}| > \varepsilon\right\} \xrightarrow[\mathrm{N}\to\infty]{} 0$ *pour tout* $\varepsilon > 0$.

En corollaire, deux critères simples permettant de conclure qu'une suite $(\mathrm{X}_n)_{n\geqslant 1}$ converge presque sûrement vers X :

■ **Corollaire 19.6 (Critère pratique de convergence p.s.)**
Si la série $\sum \mathbf{P}\{|\mathrm{X}_n - \mathrm{X}| > \varepsilon\}$ *converge pour tout* $\varepsilon > 0$, *alors* $\mathrm{X}_n \to \mathrm{X}$ *presque sûrement.*

DÉMONSTRATION : Soit $\varepsilon > 0$; puisque $\sum \mathbf{P}\{|\mathrm{X}_n - \mathrm{X}| > \varepsilon\}$ converge, le premier lemme de Borel-Cantelli entraîne que l'événement $\{|\mathrm{X}_n - \mathrm{X}| > \varepsilon \text{ ; i.s.}\}$ est de probabilité nulle. On conclut avec le théorème 19.5. ▮

■ **Corollaire 19.7 (Critère pratique de convergence p.s.)**
Si $\sum \mathbf{E}\,|\mathrm{X}_n - \mathrm{X}|$ *converge, alors* $\mathrm{X}_n \to \mathrm{X}$ *presque sûrement.*

DÉMONSTRATION : D'après l'inégalité de Markov, pour tout $\varepsilon > 0$, on a

$$\mathbf{P}\{\,|\mathrm{X}_n - \mathrm{X}| > \varepsilon\} \leqslant \frac{\mathbf{E}\,|\mathrm{X}_n - \mathrm{X}|}{\varepsilon}$$

donc la série de terme général $\mathbf{P}\{\,|\mathrm{X}_n - \mathrm{X}| > \varepsilon\}$ converge et on peut appliquer la proposition précédente. ▮

Remarque 19.8 *Une autre démarche* est la suivante : puisque $\sum \mathbf{E}\,|\mathrm{X}_n - \mathrm{X}|$ converge, cela implique que la variable aléatoire $\mathrm{S} = \sum_{n=0}^{\infty} |\mathrm{X}_n - \mathrm{X}|$, définie et à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$, est presque sûrement finie : il existe une partie N négligeable de Ω telle que

$$\forall \omega \in \Omega \setminus \mathrm{N} \qquad \sum \big[\mathrm{X}_n(\omega) - \mathrm{X}(\omega)\big] \quad \text{converge absolument}$$

(voir corollaire 14.24 page 316). Une démonstration élémentaire de ce résultat est proposé dans l'exercice 19.6 page 412. Puisque le terme général d'une série convergente tend vers 0, on en déduit que

$$\forall \omega \in \Omega \setminus \mathrm{N} \qquad \mathrm{X}_n(\omega) \to \mathrm{X}(\omega),$$

ce qui est le résultat demandé.

Sous la condition de convergence de la série $\sum \mathbf{E}\,|\mathrm{X}_n - \mathrm{X}|$, on a donc montré une condition plus forte : $\sum(\mathrm{X}_n - \mathrm{X})$ converge (absolument) presque sûrement. On peut également montrer (corollaire 14.24 page 316) qu'alors la série s'intègre terme à terme :

$$\mathbf{E} \sum_{n=0}^{\infty} (\mathrm{X}_n - \mathrm{X}) = \sum_{n=0}^{\infty} \mathbf{E}(\mathrm{X}_n - \mathrm{X}).$$

Remarque 19.9 (Moyennes des tirages à pile *ou* face*)* C'est ce résultat qui a été utilisé, au chapitre 9, pour montrer la loi forte des grands nombres dans le cas particulier de suites de variables de Bernoulli indépendantes (théorème 9.13 page 212).

§280 **Convergence en probabilité**

▶▶ On dit qu'une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ **converge en probabilité** vers une variable aléatoire X, également définie sur $(\Omega, \mathfrak{T}, \mathbf{P})$, si, pour tout $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbf{P}\Big\{\omega \in \Omega\,;\ \big|X_n(\omega) - X(\omega)\big| > \varepsilon\Big\} = 0,$$

ce que l'on écrira plus succinctement

$$\mathbf{P}\big\{\,|X_n - X| > \varepsilon\big\} \xrightarrow[n\to\infty]{} 0.$$

On notera « $X_n \xrightarrow{\mathbf{P}} X$ » cette propriété.

Remarque 19.10 D'après la définition de la limite, cela revient à dire que, pour tout $\varepsilon > 0$ et pour tout $\eta > 0$, il existe un entier $N_{\varepsilon,\eta}$ tel que

$$\forall n \geqslant N_{\varepsilon,\eta} \qquad \mathbf{P}\big\{\,|X_n - X| > \varepsilon\big\} < \eta.$$

Bien entendu, dire que $X_n \xrightarrow{\mathbf{P}} X$ est la même chose que de dire $(X_n - X) \xrightarrow{\mathbf{P}} 0$. Enfin, on a la propriété de linéarité :

■ **Proposition 19.11** *Si* $X_n \xrightarrow{\mathbf{P}} X$ *et* $Y_n \xrightarrow{\mathbf{P}} Y$, *alors* $\alpha X_n + \beta Y_n \xrightarrow{\mathbf{P}} \alpha X + \beta Y$.

Exemple 19.12 Les inégalités de Markov et de Bienaymé-Tchebychev sont des outils naturels pour prouver la convergence en probabilité. Considérons par exemple une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires centrées, dont les écarts-types respectifs σ_n tendent vers 0 quand n tend vers l'infini. L'inégalité de Bienaymé-Tchebychev montre que

$$\forall \varepsilon > 0 \qquad \mathbf{P}\big\{\,|X_n| > \varepsilon\big\} \leqslant \frac{\sigma_n^2}{\varepsilon^2} \xrightarrow[n\to\infty]{} 0$$

pour tout $\varepsilon > 0$, et donc que $X_n \xrightarrow{\mathbf{P}} 0$.

§281 **Une caractérisation par l'espérance**

D'un point de vue théorique, il est souvent plus aisé de manipuler et de majorer des espérances que des probabilités. La caractérisation suivante est donc fondamentale.

■ **Proposition 19.13** *Soit* $\psi : \mathbb{R} \to \mathbb{R}^+$ *une fonction positive, continue, bornée, telle que* $\psi(0) = 0$, *croissante et strictement croissante sur un voisinage de* 0. *Alors*

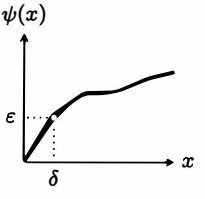
$$(X_n \xrightarrow{\mathbf{P}} X) \iff \lim_{n\to\infty} \mathbf{E}\psi\big(\,|X_n - X|\big) = 0.$$

DÉMONSTRATION : Notons M un majorant de ψ sur $\mathbb{R}^+$. Notons $Y_n = |X_n - X|$; il s'agit donc de prouver que $Y_n \xrightarrow{\mathbf{P}} 0$ si et seulement si $\mathbf{E}\psi(Y_n) \to 0$.

• *Supposons dans un premier temps que* $Y_n \xrightarrow{\mathbf{P}} 0$.

Soit $\varepsilon > 0$. Il existe $\delta > 0$ tel que $\psi(\delta) \leqslant \varepsilon$; on pose alors $\eta = \min(\delta, \varepsilon)$, de sorte que $\eta \leqslant \varepsilon$ et $\psi(x) \leqslant \varepsilon$ sur $[0\,;\,\eta]$. Pour tout entier n, on a

$$\begin{aligned} 0 \leqslant \psi(Y_n) &= \psi(Y_n)\left(\mathbf{1}_{\{Y_n > \eta\}} + \mathbf{1}_{\{Y_n \leqslant \eta\}}\right) \\ &\leqslant \psi(Y_n)\,\mathbf{1}_{\{Y_n > \eta\}} + \varepsilon\,\mathbf{1}_{\{Y_n \leqslant \eta\}} \leqslant M\,\mathbf{1}_{\{Y_n > \eta\}} + \varepsilon. \end{aligned}$$



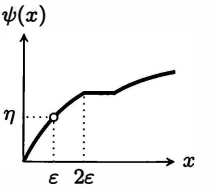
Évaluons l'espérance de part et d'autre :

$$0 \leqslant \mathbf{E}\varphi(Y_n) \leqslant M\,\mathbf{P}\{Y_n > \eta\} + \varepsilon,$$

donc, pour n suffisamment grand, $0 \leqslant \mathbf{E}\psi(Y_n) \leqslant 2\varepsilon$.

Ceci étant vrai pour tout $\varepsilon > 0$, on a prouvé que $\lim_{n\to\infty} \mathbf{E}\psi(Y_n) = 0$.

• *Réciproquement, supposons que* $\mathbf{E}\psi(\mathrm{Y}_n) \to 0$.
Soit $\varepsilon > 0$; sans restreindre la généralité de la démonstration, on peut le choisir suffisamment petit pour que ψ soit strictement croissante sur $[0\,;\,2\varepsilon]$. Posons $\eta = \psi(\varepsilon)$, de sorte que $\{\psi(\mathrm{Y}_n) > \eta\} = \{\mathrm{Y}_n > \varepsilon\}$. La variable $\psi(\mathrm{Y}_n)$ étant positive et bornée, elle est intégrable et on peut appliquer l'inégalité de Markov :



$$\eta\,\mathbf{P}\{\psi(\mathrm{Y}_n) > \eta\} \leqslant \mathbf{E}\psi(\mathrm{Y}_n)$$

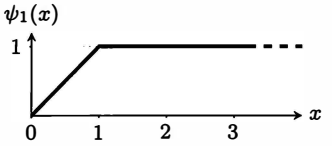
et donc

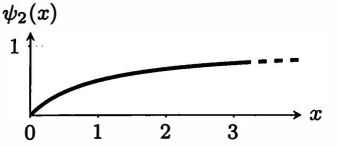
$$\mathbf{P}\{\mathrm{Y}_n > \varepsilon\} \leqslant \frac{\mathbf{E}\psi(\mathrm{Y}_n)}{\eta} \xrightarrow[n\to\infty]{} 0,$$

ce qui achève la démonstration. ∎

En pratique, on peut choisir une des fonctions suivantes :

$$\psi_1(x) = x \wedge 1 \qquad \text{ou} \qquad \psi_2(x) = \frac{x}{1+x}.$$





La première de ces fonctions convenant tout particulièrement bien lorsque l'on a affaire à des fonctions indicatrices, c'est celle que l'on utilisera par la suite.

■ **Corollaire 19.14 (Un critère pratique de convergence en probabilité)**

$$\left(\mathrm{X}_n \xrightarrow{\mathrm{P}} \mathrm{X}\right) \iff \left(\mathbf{E}\big[|\mathrm{X}_n - \mathrm{X}| \wedge 1\big] \to 0\right).$$

Convergence et images de variables aléatoires § 282
Si $g : \mathbb{R} \to \mathbb{R}$ est une application, que peut-on dire de la suite de terme général $g(\mathrm{X}_n)$ lorsque $\mathrm{X}_n \xrightarrow{\mathrm{P}} \mathrm{X}$? Lorsque $\mathrm{X}_n \xrightarrow{\mathrm{ps}} \mathrm{X}$? De manière générale, rien ; mais si g est continue, le théorème suivant montre que les convergences sont toujours valables.

■ **Théorème 19.15** *Soit* $g : \mathbb{R} \to \mathbb{R}$ *une fonction continue.*

i) Si $\mathrm{X}_n \xrightarrow{\mathrm{ps}} \mathrm{X}$, *alors* $g(\mathrm{X}_n) \xrightarrow{\mathrm{ps}} g(\mathrm{X})$.

ii) Si $\mathrm{X}_n \xrightarrow{\mathrm{P}} \mathrm{X}$, *alors* $g(\mathrm{X}_n) \xrightarrow{\mathrm{P}} g(\mathrm{X})$.

La première propriété découle immédiatement de la continuité de g. La seconde demande un peu plus de travail et nous l'admettons ici.

Comparaison entre convergence en probabilité et convergence presque sûre § 283
Le théorème suivant montre que la convergence presque sûre est plus forte que la convergence en probabilité, mais qu'une réciproque partielle existe.

■ **Théorème 19.16 (CVPS $\Longrightarrow$ CVP et réciproque partielle)** ◀

i) Si $\mathrm{X}_n \xrightarrow{\mathrm{ps}} \mathrm{X}$, *alors* $\mathrm{X}_n \xrightarrow{\mathrm{P}} \mathrm{X}$.

ii) Si $\mathrm{X}_n \xrightarrow{\mathrm{P}} \mathrm{X}$, *alors on peut extraire une sous-suite vérifiant* $\mathrm{X}_{\varphi(n)} \xrightarrow{\mathrm{ps}} \mathrm{X}$.

DÉMONSTRATION : Notons, pour tout n entier, $Y_n := |X - X_n|$.

i) Soit $\varepsilon > 0$ et notons $B_N(\varepsilon) := \{\exists n \geqslant N \quad Y_n > \varepsilon\}$ pour tout entier N. La convergence presque sûre de Y_n vers 0 montre que $\mathbf{P}(B_n(\varepsilon)) \to 0$ (théorème 19.5) ; or $\{Y_n > \varepsilon\} \subset B_n(\varepsilon)$ pour tout entier n, ce qui prouve que $\mathbf{P}\{Y_n > \varepsilon\} \to 0$. ⏢19.10

ii) Cette propriété a déjà été démontrée, en utilisant le lemme de Borel-Cantelli (exemple 5.20 page 119). Nous en donnons une seconde preuve, plus rapide.

On suppose donc que $X_n \xrightarrow{\mathbf{P}} X$; d'après le corollaire 19.14, la suite de terme général $\mathbf{E}(Y_n \wedge 1)$ tend vers 0, on peut donc trouver une extractrice $\varphi : \mathbb{N} \to \mathbb{N}$ telle que

$$\forall n \in \mathbb{N} \qquad \mathbf{E}(Y_{\varphi(n)} \wedge 1) \leqslant \frac{1}{2^n}.$$

Le série $\sum \mathbf{E}(Y_{\varphi(n)} \wedge 1)$ converge donc (corollaire 19.7) $Y_{\varphi(n)} \wedge 1 \to 0$ presque sûrement, donc $Y_{\varphi(n)} \to 0$ presque sûrement et finalement $X_{\varphi(n)} \to X$ presque sûrement.

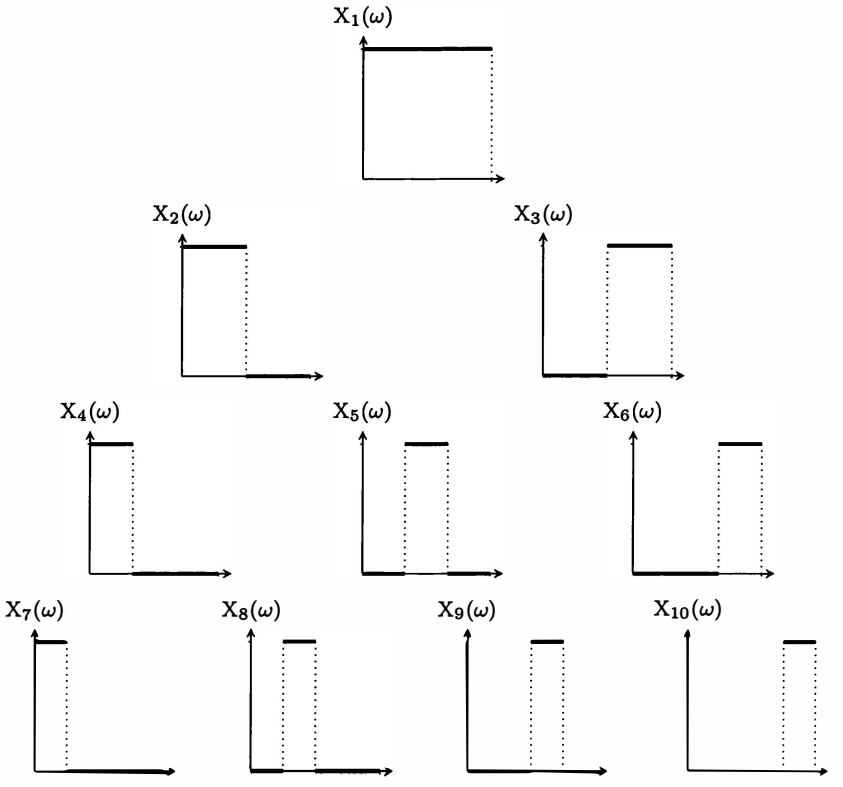
∎

Il n'y a cependant pas de réciproque complète : il se peut que l'on ait $X_n \xrightarrow{\mathbf{P}} X$ sans que l'on ait $X_n \xrightarrow{\text{ps}} X$. En revanche, lorsque les deux convergences ont lieu, elles mènent à la « même limite » au sens où : si $X_n \xrightarrow{\mathbf{P}} X$ et si $X_n \xrightarrow{\text{ps}} Y$, alors $X = Y$ presque sûrement[2].

Exemple 19.17 (Convergence en probabilité mais pas presque sûre) On considère la suite des intervalles

$$[0;1] \quad [0;\tfrac{1}{2}] \quad [\tfrac{1}{2};1] \quad [0;\tfrac{1}{3}] \quad [\tfrac{1}{3};\tfrac{2}{3}] \quad [\tfrac{2}{3};1] \quad [0;\tfrac{1}{4}]$$

On définit la suite $(X_n)_{n\geqslant 1}$ de fonctions sur $[0;1[$ comme fonctions indicatrices des intervalles précédents.



(2). Le lecteur se convaincra que si X et X′ diffèrent sur un ensemble négligeable, et si $X_n \xrightarrow{\mathbf{P}} X$, alors $X_n \xrightarrow{\mathbf{P}} X'$ également ; de même pour la convergence presque sûre.

etc. Chaque fonction X_n est donc nulle sauf sur un intervalle de la forme $\left[\frac{k}{p}\,;\,\frac{k+1}{p}\right]$. Alors la suite $(X_n)_{n\geqslant 1}$ converge en probabilité vers la variable nulle X (définie par $X(\omega)=0$ pour tout $\omega\in[0\,;\,1]$) : pour tout $\varepsilon>0$, $\mathbf{P}\{|X_n|>\varepsilon\}=1/k$, où k est l'unique entier tel que $k(k-1)/2\leqslant n<k(k+1)/2$. En revanche, tout réel $\omega\in[0\,;\,1]$ est tel que $X_n(\omega)=1$ une infinité de fois, ce qui montre que $\{X_n\to 0\}=\varnothing$!

Une distance sur l'espace des variables aléatoires (∗)

§ 284

Notons $L^0=L^0(\Omega,\mathfrak{T},\mathbf{P})$ l'espace des variables aléatoires sur $(\Omega,\mathfrak{T},\mathbf{P})$, définies à une égalité presque sûre près[3]. Pour toutes variables $X,Y\in L^0$, on pose

$$d(X,Y):=\mathbf{E}\big(\,|X-Y|\wedge 1\big)$$

(ou bien $\mathbf{E}\varphi(X-Y)$, où φ est une fonction quelconque vérifiant les propriétés énoncées dans le théorème 19.13).

■ **Théorème 19.18** *L'application d est une distance sur* $L^0(\Omega,\mathfrak{T},\mathbf{P})$, *qui caractérise la convergence en probabilité, au sens où*

$$(X_n\xrightarrow{\mathbf{P}}X)\iff\big(d(X_n,X)\to 0\big).$$

De plus, $L^0(\Omega,\mathfrak{T},\mathbf{P})$ *est complet pour cette distance.*

La démonstration qui suit est essentiellement intéressante parce que, tout comme dans le théorème 19.16, elle met en œuvre une méthode classique de construction de la limite (presque sûre) d'une suite de variables aléatoires.

Démonstration : Vérifions d'abord les axiomes qu'une distance doit vérifier. La positivité et la symétrie sont immédiates. Si X, Y et Z sont des variables aléatoires, l'inégalité triangulaire permet d'écrire que

$$|X-Z|\wedge 1\leqslant|X-Y|\wedge 1+|Y-Z|\wedge 1$$

puis, par croissance de l'espérance, $d(X,Z)\leqslant d(X,Y)+d(Y,Z)$. Enfin, $d(X,Y)=0$ implique que $|X-Y|\wedge 1=0$ presque sûrement, et donc $X=Y$ presque sûrement.

Le corollaire 19.14 montre que $X_n\xrightarrow{\mathbf{P}}X$ si et seulement si $d(X_n,X)\to 0$.

Montrons enfin que (L^0,d) est complet. Soit $(X_n)_{n\geqslant 1}$ une suite de Cauchy ; on peut construire par récurrence une extractrice $\varphi:\mathbb{N}^*\to\mathbb{N}^*$ telle que

$$\forall n\geqslant 1\qquad d(X_{\varphi(n+1)},X_{\varphi(n)})\leqslant\frac{1}{2^n},$$

ce qui nous assure de la convergence de la série de terme général

$$d(X_{\varphi(n+1)},X_{\varphi(n)})=\mathbf{E}\big(\,|X_{\varphi(n+1)}-X_{\varphi(n)}|\wedge 1\big).$$

Le théorème de convergence monotone, et plus précisément le corollaire 14.24 page 316, montrent que la série $\sum|X_{\varphi(n+1)}-X_{\varphi(n)}|\wedge 1$ converge presque sûrement. Notamment, $|X_{\varphi(n+1)}-X_{\varphi(n)}|<1$ presque sûrement à partir d'un certain rang et donc $\sum|X_{\varphi(n+1)}-X_{\varphi(n)}|$ converge presque sûrement. On en déduit que $\sum(X_{\varphi(n+1)}-X_{\varphi(n)})$ converge presque sûrement et on peut poser

$$X:=X_{\varphi(1)}+\sum_{n=1}^{\infty}(X_{\varphi(n+1)}-X_{\varphi(n)}),$$

qui est bien une variable aléatoire. Par construction, la sous-suite $(X_{\varphi(n)})$ converge presque sûrement vers X, donc converge également en probabilité, c'est-à-dire dans l'espace (L^0,d). On conclut avec une propriété générale des espaces métriques : toute suite de Cauchy admettant une sous-suite convergente de limite X est elle-même convergente de limite X. ▮

---

(3). Plus précisément, notons $\mathcal{L}^0=\mathcal{L}^0(\Omega,\mathfrak{T},\mathbf{P})$ l'espace de toutes les variables aléatoires sur $(\Omega,\mathfrak{T},\mathbf{P})$; on définit la relation d'équivalence « $\equiv$ » par : $X\equiv Y$ si et seulement si $X=Y$ presque sûrement. On note $L^0(\Omega,\mathfrak{T},\mathbf{P})$ l'espace quotient, c'est-à-dire l'espace des classes d'équivalences de $\mathcal{L}^0$.

19.2 CONVERGENCE EN MOYENNE ET DANS L^p (*)

On introduit de nouveaux types de convergence pour des variables aléatoires définies sur un même espace $(\Omega, \mathfrak{T}, \mathbf{P})$ et admettant des moments d'ordre p donné.

§ 285 **Convergence en moyenne et dans L^p**

On dit qu'une suite $(X_n)_{n\geqslant 1}$ **converge en moyenne** vers X si les variables aléatoires $X - X_n$ $(n \in \mathbb{N})$ sont intégrables et si

$$\lim_{n\to\infty} \mathbf{E}\,|X_n - X| = 0.$$

On écrit alors $X_n \xrightarrow{L^1} X$. On dit également que $(X_n)_{n\geqslant 1}$ **converge dans L^1** vers X.

Remarque 19.19 Si $X = X'$ presque sûrement, si $X_n = X'_n$ presque sûrement pour tout entier n et si $X_n \xrightarrow{L^1} X$, alors $X'_n \xrightarrow{L^1} X'$. La convergence en moyenne est réellement une propriété de l'espace L^1, et non de $\mathcal{L}^1$.

Remarque 19.20 Si $\mathbf{E}\,|X_n - X| \to 0$, alors $\mathbf{E}X_n \to \mathbf{E}X$ par simple inégalité triangulaire :

$$|\mathbf{E}X_n - \mathbf{E}X| = |\mathbf{E}(X_n - X)| \leqslant \mathbf{E}\,|X_n - X|\,.$$

De la même façon, on dit que $(X_n)_{n\geqslant 1}$ **converge dans L^2 vers X**, et on écrit $X_n \xrightarrow{L^2} X$, si $\mathbf{E}\,|X_n - X|^2 \to 0$.

Remarque 19.21 Puisque l'application $X \mapsto (\mathbf{E}X^2)^{1/2}$ est une norme, si $X_n \xrightarrow{L^2} X$, alors $\mathbf{E}X_n^2 \to \mathbf{E}X^2$ par inégalité triangulaire $\big|\,\|X_n\| - \|X\|\,\big| \leqslant \|X_n - X\|$.

Plus généralement, si p est un entier strictement positif, on dit que $(X_n)_{n\geqslant 1}$ **converge vers X dans L^p** si $\mathbf{E}\,|X_n - X|^p \to 0$.

§ 286 **Convergence en probabilité et convergence dans L^p (*)**

Un premier résultat montre que la convergence dans un espace L^p implique la convergence en probabilité.

■ **Théorème 19.22** *Soit p un entier, $p \geqslant 1$.*

i) Si $X_n \xrightarrow{L^1} X$, alors $X_n \xrightarrow{P} X$.

ii) Si $X_n \xrightarrow{L^p} X$ et si $1 \leqslant q \leqslant p$, alors $X_n \xrightarrow{L^q} X$.

iii) Notamment, si $X_n \xrightarrow{L^p} X$, alors $X_n \xrightarrow{P} X$.

DÉMONSTRATION : D'après l'inégalité de Markov, $\varepsilon\,\mathbf{P}\{|X_n - X| > \varepsilon\} \leqslant \mathbf{E}\,|X_n - X|$ pour tout $\varepsilon > 0$. La première assertion en découle. Le seconde provient des inégalités de Lyapounov (cf. exercice 14.1 page 329), et la troisième découle naturellement des deux premières. ▮

Remarque 19.23 Une réciproque partielle existe, que nous donnons sans démonstration :

■ **Proposition 19.24** *Si $X_n \xrightarrow{P} X$ et si $(X_n)_{n\geqslant 1}$ est bornée dans L^q $(q \geqslant 2)$, alors, pour tout entier p tel que $1 \leqslant p < q$, X_n converge vers X dans L^p.*

19.3 CONVERGENCE EN LOI

Convergence en loi

§ 287

Nous introduisons ici une quatrième notion de convergence : celle de convergence en loi (que l'on appelle également **convergence étroite**[4]).

Définition 19.25 (Convergence en loi, première version) ◀◀

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles définies chacune sur son propre espace probabilisé $(\Omega_n, \mathcal{T}_n, \mathbf{P}_n)$, et soit X une variable aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathcal{T}, \mathbf{P})$. Si l'on note F_n la fonction de répartition de X_n et F celle de X, on dit que la suite $(X_n)_{n\geqslant 1}$ **converge en loi** vers X, et on écrit

$$X_n \overset{L}{\Longrightarrow} X \qquad \text{ou} \qquad F_n \overset{L}{\Longrightarrow} F.$$

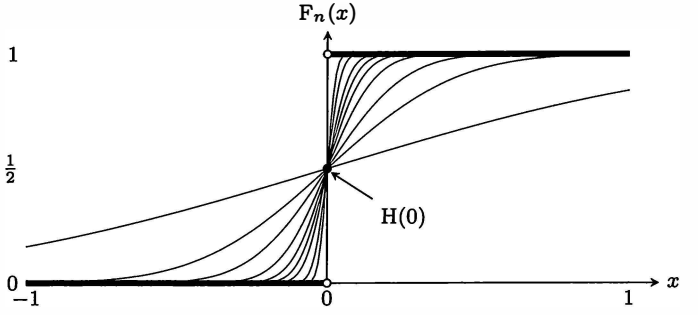
si la suite $(F_n)_{n\geqslant 1}$ converge simplement vers F en tout point de continuité de F ; autrement dit, si, pour tout $x \in \mathbb{R}$ tel que $F(x^-) = F(x)$, on a $F_n(x) \to F(x)$. ⏢19.7

Remarque 19.26 (Continuité et discontinuité d'une fonction de répartition)
Toute fonction de répartition étant continue à droite, la continuité de F en x est donc équivalente à la continuité à gauche en x ; enfin, par croissance de F, la limite à gauche existe, et la continuité en x est donc équivalente à la relation : $F(x^-) = F(x)$.

Par ailleurs, l'ensemble des points de discontinuité de F est nécessairement fini ou dénombrable (voir A.4, page 676), ce qui veut dire que la condition $F_n(x) \xrightarrow[n\to\infty]{} F(x)$ doit avoir lieu sur $\mathbb{R}$ privé, au plus, d'un ensemble D dénombrable. Celui-ci peut avoir l'apparence d'un « gros » ensemble (il peut notamment être dense, comme $\mathbb{Q}$), mais il reste « petit » au sens de la mesure (il est mesurable et de mesure nulle).

Exemple 19.27 (Le problème avec les atomes) Considérons une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires réelle, telle que X_n suit une loi normale $\mathcal{N}(0\ ;\ 1/n)$. Si l'on note $\mathfrak{N}$ la fonction de répartition de la loi normale centrée réduite, alors la fonction de répartition de X_n est $F_n(x) = \mathfrak{N}(nx)$. La suite $(F_n)_{n\geqslant 0}$ converge simplement vers la fonction

$$H : x \longmapsto \begin{cases} 1 & \text{si } x > 0, \\ 1/2 & \text{si } x = 0, \\ 0 & \text{si } x < 0. \end{cases}$$



Mais attention ! cette fonction n'est *pas* la fonction de répartition d'une variable aléatoire, car elle n'est pas continue à droite en 0. En revanche, une variable aléatoire X de loi certaine : $\mathbf{P}\{X = 0\} = 1$ a pour fonction de répartition

$$F : x \longmapsto \begin{cases} 1 & \text{si } x \geqslant 0, \\ 0 & \text{si } x < 0. \end{cases}$$

(4). *Convergence in distribution* et *weak convergence* en anglais, respectivement. Les deux notions sont légèrement différentes, comme il sera expliqué plus loin.

Notamment, pour tout $x \neq 0$, on a bien $F_n(x) \to F(x)$, et 0 est l'unique point de discontinuité (ou : **atome**) de F. Ainsi, la suite $(X_n)_{n\geqslant 1}$ converge en loi vers X.

Mais au fait... y a-t-il une situation que l'on modéliserait par cette suite de variables aléatoires ?

Considérons par exemple une suite $(Y_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, suivant toutes la même loi normale centrée réduite : $Y_n \rightsquigarrow \mathcal{N}(0\,;\,1)$. Ce peuvent être des variables égales au résultat de mesures indépendantes d'une certaine grandeur dans une expérience de physique (par exemple un bruit blanc gaussien à différents instants), dans un laboratoire situé, disons, à Lyon. Si l'on note $X_n = (Y_1 + \cdots + Y_{n^2})/n^2$ la moyenne des n^2 premières mesures, alors X_n suit la loi $\mathcal{N}(0\,;\,1/n)$. Quant à la variable aléatoire X, définissons la comme la fonction indicatrice d'une période de canicule en terre Adélie : $X = 0$ s'il fait froid (probabilité 1), $X = 1$ si, contre toute attente, il y a une canicule (probabilité 0). Alors, comme on l'a dit, la suite $(X_n)_{n\geqslant 1}$ converge en loi vers X — bien que, et cet exemple tiré par les cheveux n'a pas d'autre but que de l'illustrer, les variables aléatoires X_n et la variable X n'aient aucun rapport entre elles, et qu'elles ne soient même pas, *a priori*, définies sur le même univers[(5)].

L'exemple précédent illustre le fait que la convergence en loi n'implique aucune relation de « proximité » entre les variables aléatoires X_n et la variable X.

Pour être rigoureux, on ne devrait d'ailleurs pas dire que X_n converge vers X, mais que *la loi de* X_n *converge vers la loi de* X, et écrire

$$\mathbf{P}_{X_n} \overset{L}{\Longrightarrow} \mathbf{P}_X, \qquad \text{ou} \qquad F_{X_n} \overset{L}{\Longrightarrow} F_X.$$

La convergence en loi n'est pas une histoire de convergence de variables aléatoires, contrairement à la convergence en probabilité ou à la convergence presque sûre.

Remarque 19.28 (Attention aux manipulations hâtives) Les énoncés

$$X_n \overset{L}{\Longrightarrow} X \qquad \text{et} \qquad X_n - X \overset{L}{\Longrightarrow} 0$$

ne sont pas équivalents !

De manière plus générale, et contrairement à ce qui se passe avec la convergence presque sûre ou la convergence en probabilité (théorèmes 19.4 et 19.11), si $X_n \overset{L}{\Longrightarrow} X$ et $Y_n \overset{L}{\Longrightarrow} Y$, rien n'oblige à avoir $\alpha X_n + \beta Y_n \overset{L}{\Longrightarrow} \alpha X + \beta Y$! Prenons par exemple deux variables X et Y, indépendantes, de même loi normale centrée réduite $\mathcal{N}(0\,;\,1)$. Posons $X_n = X$ et $Y_n = -X$ pour tout $n \in \mathbb{N}$. On a bien $X_n \overset{L}{\Longrightarrow} X$ et $Y_n \overset{L}{\Longrightarrow} X$ (puisque $-X$ a même loi que X), mais cependant $X_n + Y_n = 0$ et ne converge pas en loi vers 2X, qui suit la loi $\mathcal{N}(0\,;\,2)$.

Rabâchons : s'il y a une notion de convergence en loi, il n'y a pas de notion de (variable) limite : seules les lois ont une limite.

§ 288 **Convergence en loi *versus* convergence étroite**

On peut caractériser la convergence en loi d'une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires réelles sans passer par la fonction de répartition :

■ **Théorème 19.29** *Une suite* $(X_n)_{n\geqslant 1}$ *de variables aléatoires réelles converge en loi vers une variable aléatoire* X *si et seulement si, pour toute fonction* $f : \mathbb{R} \to \mathbb{R}$ *continue et bornée, on a*

$$\lim_{n\to\infty} \mathbf{E}\big(f(X_n)\big) = \mathbf{E}\big(f(X)\big). \tag{19.1}$$

La démonstration de ce théorème est reportée en annexe, à la fin de ce chapitre.

La propriété (19.1), appelée **convergence étroite**, est équivalente à la convergence en loi lorsque l'on traite de variables aléatoires à valeurs réelles ; elle permet de généraliser la notion de convergence en loi à des variables aléatoires à valeurs dans $\mathbb{R}^d$ (ou dans $\mathbb{C}^d$), comme suit.

(5). Bien entendu, il est toujours possible de choisir Ω suffisamment gros pour englober à la fois le climat de la terre Adélie et les mesures physiques effectuées dans un laboratoire à Lyon, mais ce n'est en rien nécessaire.

Définition 19.30 (Convergence en loi, deuxième version) Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires et soit X une variable aléatoire. On suppose que X et les X_n sont toutes à valeurs dans le même espace $E = \mathbb{R}^d$ (mais elles peuvent être définies sur des espaces probabilisés différents). On dit que $(X_n)_{n\geqslant 1}$ **converge en loi** (ou : **étroitement**) vers X si, pour toute fonction $f : E \to \mathbb{R}$ continue et bornée, on a

$$\lim_{n\to\infty} \mathbf{E}\big(f(X_n)\big) = \mathbf{E}\big(f(X)\big).$$

Remarque 19.31 On peut limiter la classe des fonctions f dans la définition précédente :

■ **Théorème 19.32** $X_n \overset{L}{\Longrightarrow} X$ *si et seulement si* $\mathbf{E}f(X_n) \to \mathbf{E}f(X)$ *pour toute fonction* $f : \mathbb{R} \to \mathbb{R}$ *bornée et lipschitzienne.*

Le résultat est encore valable si $\mathbf{E}f(X_n) \to \mathbf{E}f(X)$ *pour toute fonction de classe* $\mathscr{C}^\infty$ *à support borné.*

On démontre la première propriété approchant toute fonction f bornée par une suite croissante de fonctions lipschitziennes de limite f et en appliquant le théorème de convergence monotone. Le lecteur trouvera une preuve claire et complète dans Jacod et Protter [53].

Remarque 19.33 (Cas de fonctions discontinues) On peut admettre que la fonction $f : \mathbb{R}^d \to \mathbb{R}$ admette des discontinuités, du moment qu'elle est continue sur une partie A de $\mathbb{R}^d$ telle que $\mathbf{P}\{X \in A\} = 1$.

Remarque 19.34 (Convergence en loi et espérances) Supposons que chaque variable aléatoire X_n admette une espérance $\mathbf{E}(X_n)$, et que la suite $(X_n)_{n\geqslant 1}$ converge en loi vers la loi d'une variable aléatoire X. Peut-on en déduire que X admet une espérance ? La réponse est *non.* Et dans le cas où X admet une espérance, peut-on en déduire $\mathbf{E}(X_n) \to \mathbf{E}(X)$? Là encore, la réponse est hélas *non.*

Voici des contre-exemples pour chaque questions.

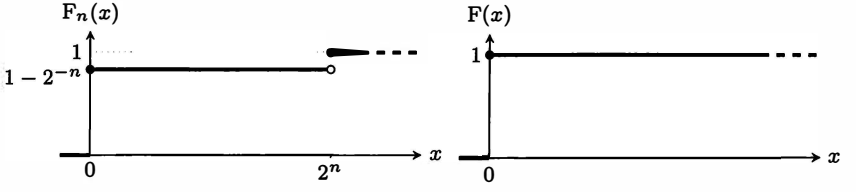
- Si Y_n est une variable aléatoire admettant pour densité la fonction

$$f_n(x) = \begin{cases} \dfrac{1}{\operatorname{Arc\,tan} n \cdot (1+x^2)} & \text{si } |x| \leqslant n \\ 0 & \text{sinon,} \end{cases}$$

  alors $\mathbf{E}(Y_n) = 0$ pour tout $n \in \mathbb{N}$, mais le lecteur est invité à vérifier que $(Y_n)_{n\geqslant 1}$ converge en loi vers la loi de Cauchy $\mathscr{C}(0,1)$ qui n'admet pas d'espérance.
- On considère une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires discrètes, dont la loi est donnée, pour tout entier n, par

$$\mathbf{P}\{X_n = 0\} = 1 - \frac{1}{2^n} \qquad \text{et} \qquad \mathbf{P}\{X_n = 2^n\} = \frac{1}{2^n}.$$

  Notons enfin X la variable constante égale à 0. La suite $(X_n)_{n\geqslant 1}$ converge en loi vers X, comme on le voit sur les graphes des fonctions de répartition de X_n et de X :



  L'espérance de X_n vaut $\mathbf{E}(X_n) = 1$ pour tout $n \geqslant 1$, tandis que celle de X est évidemment nulle. *Il n'y a pas de contradiction avec le théorème 19.29*, puisque la fonction f utilisée est l'identité, qui n'est pas bornée sur $\mathbb{R}^+$.

Remarque 19.35 (Concernant l'algèbre dans $\mathbb{R} \cup \{\pm\infty\}$*)* Nous avons évoqué (§ 26) la règle axiomatique de calcul : $0 \times (\pm\infty) = 0$, utilisée lors du calcul d'une espérance lorsque 0 est la probabilité $\mathbf{P}\{X = \pm\infty\}$. Cette règle stipule qu'un gain infini, mais une probabilité nulle, doit être négligé lors du calcul de l'espérance. L'exemple précédent permet d'illustrer le bien fondé de cette règle : les variables X_n peuvent prendre une valeur qui tend vers $+\infty$, avec une probabilité qui tend vers 0, le produit étant toujours égal à 1 — mais l' « espérance limite » est bien nulle, ce que l'on pourrait écrire symboliquement comme « $0 \cdot (+\infty) = 0$ ».

■ **Corollaire 19.36** *Si* $X_n \xrightarrow{L} X$ *et si* $g : \mathbb{R} \to \mathbb{R}$ *est une fonction continue, alors* $g(X_n) \xrightarrow{L} g(X)$.

Plus généralement, si $(X_n)_{n\geqslant 1}$ *est une suite à valeurs dans* $\mathbb{R}^n$, *si* $g : \mathbb{R}^n \to \mathbb{R}$ *est une application continue sur un ensemble* $A \subset \mathbb{R}^d$ *tel que* $X \in A$ *presque sûrement et si* $X_n \xrightarrow{L} X$, *alors* $g(X_n) \xrightarrow{L} g(X)$.

DÉMONSTRATION : Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue et bornée. Alors $f \circ g$ est également continue et bornée donc $\mathbf{E}f(g(X_n)) \to \mathbf{E}f(g(X))$. ▮

↫19.12

§289 **Convergence en loi et convergence en probabilité**

Montrons maintenant un des résultats essentiels du chapitre, qui montre que la convergence en loi est la plus faible des notions déjà rencontrées (convergence presque sûre, en probabilité, en moyenne).

►► ■ **Théorème 19.37** *Si* $X_n \xrightarrow{P} X$, *alors* $X_n \xrightarrow{L} X$.

Notamment, si $X_n \xrightarrow{L^p} X$ *ou si* $X_n \xrightarrow{ps} X$, *alors* $X_n \xrightarrow{L} X$.

DÉMONSTRATION : Nous démontrons ce résultat dans le cadre des variables aléatoires à valeurs réelles (il reste vrai pour les variables à valeurs dans un espace vectoriel normé, mais nous l'admettons ici).

Notons F_n la fonction de répartition de X_n et F celle de X. Soit x un point de continuité de F, c'est-à-dire tel que $F(x^+) = F(x^-) = F(x)$.

Soit $\varepsilon > 0$. Il existe un réel η tel que $|F(y) - F(x)| \leqslant \varepsilon$ dès lors que $x - \eta \leqslant y \leqslant x + \eta$.

On peut écrire, en partitionnant Ω selon la valeur de X puis par croissance de la mesure de probabilité :

$$\begin{aligned} F_n(x) = \mathbf{P}\{X_n \leqslant x\} &= \mathbf{P}\{X_n \leqslant x, X \leqslant x+\eta\} + \mathbf{P}\{X_n \leqslant x, X > x+\eta\} \\ &\leqslant F(x+\eta) + \mathbf{P}\{|X_n - X| > \eta\}. \end{aligned}$$

De la même façon,

$$\begin{aligned} F(x-\eta) = \mathbf{P}\{X \leqslant x-\eta\} &= \mathbf{P}\{X \leqslant x-\eta, X_n \leqslant x\} + \mathbf{P}\{X \leqslant x-\eta, X_n > x\} \\ &\leqslant F_n(x) + \mathbf{P}\{|X_n - X| > \eta\}, \end{aligned}$$

ce qui donne

$$F(x-\eta) - \mathbf{P}\{|X_n - X| > \eta\} \leqslant F_n(x) \leqslant F(x+\eta) + \mathbf{P}\{|X_n - X| > \eta\}.$$

Puisque $\mathbf{P}\{|X_n - X| > \eta\} \to 0$, il existe un rang N à partir duquel $\mathbf{P}\{|X_n - X| > \eta\} \leqslant \varepsilon$. Ainsi, pour tout $n \geqslant N$,

$$F(x) - 2\varepsilon \leqslant F(x-\eta) - \varepsilon \leqslant F_n(x) \leqslant F(x+\eta) + \varepsilon \leqslant F(x) + 2\varepsilon,$$

c'est-à-dire $|F_n(x) - F(x)| \leqslant 2\varepsilon$.

Ceci étant vrai pour tout $\varepsilon > 0$, on a bien montré que $F_n(x) \to F(x)$. ▮

Il n'y a évidemment pas de réciproque, ne serait-ce que parce que les X_n ne sont pas forcément définies sur le même univers.

§290 **Convergence en loi de variables discrètes**

Dans ce paragraphe, nous démontrons deux résultats permettant de caractériser la convergence en loi d'une suite de variables prenant leurs valeurs dans $\mathbb{Z}$ (ou dans un ensemble dénombrable A n'admettant pas de point d'accumulation[6]). Le premier caractérise la convergence en loi en terme de convergence des probabilités atomiques ; le second, réservé aux variables à valeurs dans $\mathbb{N}$, montre que la convergence en loi est équivalente à la convergence *simple* des fonctions génératrices. Les démonstrations sont techniques, mais élémentaires.

(6). On dit également que les points de A sont isolés : pour tout $a \in A$, il existe un voisinage $]a-\varepsilon\,;\,a+\varepsilon[$ ne contenant aucun autre point de A.

■ **Proposition 19.38 (Convergence en loi de variables discrètes)**
Si X *et* X_n *(*$n \in \mathbb{N}$*) sont réelles et à valeurs dans un même ensemble dénombrable* ◀◀ $A = \{x_k \ ; \ k \in K\}$ discret, *c'est-à-dire dont tous les points sont isolés, alors* $X_n \xrightarrow{L} X$ *si et seulement si*

$$\mathbf{P}\{X_n = x_k\} \xrightarrow[n\to\infty]{} \mathbf{P}\{X = x_k\}$$

pour tout $k \in K$.

DÉMONSTRATION : L'ensemble A est supposé discret. On définit

$$p_k^{[n]} := \mathbf{P}\{X_n = x_k\} \qquad \text{et} \qquad p_k := \mathbf{P}\{X = x_k\},$$

pour tous entiers n et k.

• Supposons $X_n \xrightarrow{L} X$. On aimerait écrire que, si k est un entier fixé, alors $F_n(x_k) \to F(x_k)$ quand $n \to \infty$ grâce à la convergence en loi. Malheureusement, x_k est justement un point de discontinuité de F. En revanche, puisque nous avons supposé x_k isolé, il existe un réel $\varepsilon > 0$ tel que l'intervalle $]\,x_k\,;\,x_k+\varepsilon\,]$ ne rencontre par A (par exemple, si $A = \mathbb{Z}$, on peut choisir $\varepsilon = 1/2$). Ainsi, les fonctions F_n ainsi que la fonction F sont constantes sur $[\,x_k\,;\,x_k+\varepsilon\,]$, et de plus $x_k+\varepsilon$ est un point de continuité de F. On peut alors écrire

$$F_n(x_k) = F_n(x_k+\varepsilon) \xrightarrow[n\to\infty]{} F(x_k+\varepsilon) = F(x_k).$$

Ceci étant établi, on en déduit que pour tout entier k :

$$p_k^{[n]} = F_n(x_k) - F_n(x_{k-1}) \xrightarrow[n\to\infty]{} F(x_k) - F(x_{k-1}) = p_k.$$

• Réciproquement, supposons $p_k^{[n]} \to p_k$ pour tout entier k. Nous allons montrer que $(F_n)_{n\geqslant 1}$ converge uniformément vers F sur $\mathbb{R}$ tout entier[7].

En écrivant que, pour tous entiers k et n,

$$|p_k^{[n]} - p_k| = p_k^{[n]} + p_k - 2\min(p_k^{[n]}, p_k), \qquad (*)$$

et en remarquant que $0 \leqslant \min(p_k^{[n]}, p_k) \leqslant p_k$, on peut sommer la relation $(*)$ pour k parcourant K et obtenir

$$\sum_{k\in K} |p_k^{[n]} - p_k| = 2 - 2\sum_{k\in K} \min(p_k^{[n]}, p_k).$$

À k fixé, la suite de terme général $\min(p_k^{[n]}, p_k)$ converge vers p_k, tout en étant majorée par p_k ; la version discrète du théorème de convergence dominée (paragraphe A17, page 674) permet de conclure que

$$\lim_{n\to\infty} \sum_{k\in K} \min(p_k^{[n]}, p_k) = \sum_{k\in K} p_k = 1$$

et donc que

$$\lim_{n\to\infty} \sum_{k\in K} |p_k^{[n]} - p_k| = 0.$$

Pour conclure, pour tout $x \in \mathbb{R}$, on peut écrire

$$F_n(x) = \sum_{\substack{k\in K \\ x_k \leqslant x}} p_k^{[n]} \qquad \text{et} \qquad F(x) = \sum_{\substack{k\in K \\ x_k \leqslant x}} p_k,$$

et donc majorer

$$|F_n(x) - F(x)| = \sum_{\substack{k\in K \\ x_k \leqslant x}} |p_k^{[n]} - p_k| \leqslant \sum_{k\in K} |p_k^{[n]} - p_k| \xrightarrow[n\to\infty]{} 0,$$

ce qui achève la preuve. ∎

---

[7]. Les points de discontinuité de F sont exactement ceux des F_n, ce qui explique que cette convergence soit un peu plus forte que ce qui est demandé dans la définition de la convergence en loi.

Remarque 19.39 (Cas général de variables discrètes (*)) Dans la démonstration du sens direct, on a utilisé le fait que l'ensemble dénombrable A ait tous ses points isolés. Si l'ensemble A possède des points d'accumulation, ce sens direct n'est plus valable (le sens réciproque, quant à lui, est toujours valide).

Prenons par exemple une suite $(X_n)_{n\geqslant 0}$ de variables constantes : $X_n = 1/n$. (La loi de X_n est alors notée $\delta_{1/n}$.) Alors il est aisé de vérifier que $\delta_{1/n} \xrightarrow{L} \delta_0$ (cf. exercice 19.8). Si l'on note $x_0 = 0$ et $x_k = 1/k$ pour tout entier $k \geqslant 1$, on a donc

$$p_k^{[n]} = \begin{cases} 1 & \text{si } k = n, \\ 0 & \text{sinon.} \end{cases} \qquad \text{et} \qquad p_k = \begin{cases} 1 & \text{si } k = 0, \\ 0 & \text{sinon.} \end{cases}$$

Notamment, $p_k^{[n]} \to 0 = p_k$ pour tout $k \geqslant 1$, mais $p_0^{[n]} = 0 \not\to 1 = p_0$.

On peut rétablir les choses en imposant que la limite simple des probabilités ponctuelles soit encore une loi de probabilité (c'est-à-dire de somme égale à 1). Sous la condition

$$\sum_{k\in \mathrm{K}} \lim_{n\to\infty} p_k^{[n]} = 1,$$

l'hypothèse « $X_n \xrightarrow{L} X$ » implique que $p_k^{[n]} \to p_k$ pour tout entier n (la répartition des masses de probabilité ne peut plus « s'échapper à l'infini »).

En revanche, le sens réciproque (si $p_k^{[n]} \to p_k$ pour tout k alors $X_n \xrightarrow{L} X$) est toujours vrai, que K soit discret ou non.

Remarque 19.40 Si toutes les variables X_n sont certaines, les trois convergences (en loi, en pro-
↑19.8 babilité, presque sûre) sont équivalentes.

Le **théorème de continuité** suivant permet de caractériser la convergence en loi de fonctions à valeurs dans $\mathbb{N}$.

▶ ■ **Théorème 19.41 (Convergence en loi et fonctions génératrices)**
Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires à valeurs dans $\mathbb{N}$, de fonctions génératrices G_n et de lois $\mathbf{P}\{X_n = k\} = p_k^{[n]}$. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$, de fonction génératrice G et de loi $\mathbf{P}\{X = k\} = p_k$. Alors il y a équivalence entre les énoncés

i) $X_n \xrightarrow{L} X$;

ii) $\lim_{n\to\infty} p_k^{[n]} = p_k$ *pour tout* $k \in \mathbb{N}$;

iii) $(G_n)_{n\geqslant 1}$ *converge simplement vers* G *sur* $[0\,;1[$.

DÉMONSTRATION : L'équivalence entre *i)* et *ii)* a été établie au théorème précédent.

• *ii)* $\Longrightarrow$ *iii)*. Soit $u \in [0\,;1[$. Fixons un réel $\varepsilon > 0$. La majoration $|p_k^{[n]} - p_k| \leqslant 1$ permet de d'écrire l'inégalité suivante, où le second membre est bien défini :

$$|G_n(u) - G(u)| \leqslant \sum_{k=0}^{\infty} |p_k^{[n]} - p_k|\, u^k.$$

On peut bien sûr rendre $|p_k^{[n]} - p_k|$ epsilonesque pour un nombre arbitraire, mais fini, d'entre eux ; pour les autres, on peut majorer

$$\sum_{k=\mathrm{K}+1}^{\infty} |p_k^{[n]} - p_k|\, u^k \leqslant \sum_{k=\mathrm{K}+1}^{\infty} u^k = \frac{u^{\mathrm{K}+1}}{1-u}.$$

On choisit donc un entier K qui rende le second membre inférieur à ε. Ce nombre étant fixé, il existe un entier N tel que

$$\forall n \geqslant \mathrm{N} \qquad \sum_{k=0}^{\mathrm{K}} |p_k^{[n]} - p_k| \leqslant \varepsilon.$$

Alors, pour tout $n \geqslant N$, on aura

$$|G_n(u) - G(u)| \leqslant \sum_{k=0}^{K} |p_k^{[n]} - p_k| + \sum_{k=K+1}^{\infty} |p_k^{[n]} - p_k| u^k \leqslant 2\varepsilon.$$

• *iii)* $\Longrightarrow$ *ii)* : tout d'abord, rappelons un résultat bien connu, conséquence du théorème de Bolzano-Weierstrass :

■ **Lemme 19.42** *Soit $(x_n)_{n\geqslant 1}$ une suite réelle bornée, alors elle admet au moins une valeur d'adhérence ℓ. Si elle n'en admet pas d'autre, alors elle converge vers ℓ.*

Nous allons l'appliquer une infinité de fois avant de conclure par un « argument diagonal ». La suite $(p_1^{[n]})_{n\geqslant 0}$ étant bornée, on peut en extraire une suite $(p_1^{[\varphi_1(n)]})_{n\geqslant 0}$ convergente, dont on note ℓ_1 la limite.

La suite $(p_2^{[\varphi_1(n)]})_{n\geqslant 0}$ étant bornée, on peut en extraire une suite $(p_2^{[\varphi_2(n)]})_{n\geqslant 0}$ convergente, dont on note ℓ_2 la limite (l'extractrice φ_2 est donc de la forme $\varphi_1 \circ \xi$, où ξ est une autre extractrice).

On construit ainsi de proche en proche des extractrices φ_k vérifiant : $(p_k^{[\varphi_k(n)]})_{n\geqslant 0}$ converge vers une limite ℓ_k.

Enfin, on définit $\varphi : \mathbb{N} \to \mathbb{N}$ par $\psi(n) = \varphi_n(n)$. Alors $(\psi(n))_{n\geqslant m}$ est une sous-suite de $(\varphi_m(n))_{n\geqslant 0}$ pour tout m, ce qui prouve que $(p_k^{[\psi(n)]})_{n\geqslant 0}$ converge, et ce pour tout $k \in \mathbb{N}$. D'après la première partie de la démonstration, la suite de terme général $G_{\psi(n)}$ converge simplement vers la fonction définie par

$$H(u) = \sum_{k=0}^{\infty} \ell_k u^k,$$

et donc $H = G$; les coefficients d'une série entière se déduisant de son comportement au voisinage de 0, on en déduit que $\ell_k = p_k$ pour tout $k \in \mathbb{N}$.

Nous n'avons pour l'instant prouvé que l'existence d'une sous-suite convergente. Si, pour un certain entier q la suite $(p_q^{[n]})_{n\geqslant 0}$ ne convergeait pas vers p_q, elle admettrait, d'après le lemme ci-dessus, une sous-suite convergeant vers une autre valeur p'_q. On pourrait alors refaire la même construction que précédemment, en commençant par l'indice q au lieu de l'indice 1, et construire par procédé diagonal une extraction ψ' telle que $p_k^{[\psi'(n)]} \to p'_k$ pour tout k, on trouverait que la fonction $H' : u \mapsto \sum p'_k u^k$ serait égale à G, donc que $p'_k = p_k$ pour tout k, et finalement $p'_q = p_q$, en contradiction avec l'hypothèse. ∎

Convergence en loi de variables à densité

§ 291

Soient X_n $(n \in \mathbb{N})$ et X des variables aléatoires, à valeurs dans le même espace $\mathbb{R}$ ou $\mathbb{R}^d$. Supposons de plus que ces variables admettent des densités respectives f_n et f. Quel est le lien entre les propriétés :

$$X_n \xRightarrow{L} X \qquad \text{et} \qquad f_n \to f \ ?$$

L'exemple de la figure 19.1 page suivante montre que la convergence en loi n'implique pas la convergence des densités — ni uniformément, ni simplement, ni même presque partout. En revanche, la réciproque est vraie.

■ **Théorème 19.43** *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires (à valeurs dans $\mathbb{R}$ ou $\mathbb{R}^d$) ; on suppose que chaque X_n admet une densité f_n. Soit X une variable aléatoire (à valeurs dans le même espace), de densité f.*

i) Si $(f_n)_{n\geqslant 1}$ converge simplement vers f, alors $X_n \xRightarrow{L} X$.

ii) Plus généralement, si $(f_n)_n$ converge presque partout vers f, alors $X_n \xRightarrow{L} X$.

DÉMONSTRATION : Nous pouvons démontrer aisément le résultat dans le cas de variables à valeurs réelles.

Supposons que $(f_n)_{n\geqslant 1}$ converge simplement (ou presque partout) vers f. Alors, pour tout x réel,

$$\left|\mathrm{F}_n(x) - \mathrm{F}(x)\right| \leqslant \int_{-\infty}^{x} \left|f_n(t) - f(t)\right| \mathrm{d}t \leqslant \int_{-\infty}^{+\infty} \left|f_n(t) - f(t)\right| \mathrm{d}t. \qquad (*)$$

Pour tout $t \in \mathbb{R}$, on peut écrire

$$\left|f_n(t) - f(t)\right| = f_n(t) + f(t) - 2\min\left(f_n(t), f(t)\right)$$

et on a l'encadrement

$$0 \leqslant g_n(t) := \min\left(f_n(t), f(t)\right) \leqslant f(t)$$

qui montre que la suite $(g_n)_{n\geqslant 1}$ est dominée par la fonction intégrable f. De plus, elle converge simplement (ou presque partout) vers f ; le théorème de convergence dominée permet de conclure que

$$\int_{-\infty}^{+\infty} g_n(t)\,\mathrm{d}t \xrightarrow[n\to\infty]{} \int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = 1$$

ce qui montre que

$$\int_{-\infty}^{+\infty} \left|f_n(t) - f(t)\right| \mathrm{d}t = \int_{-\infty}^{+\infty} \left[f_n(t) + f(t) - 2g_n(t)\right] \mathrm{d}t \xrightarrow[n\to\infty]{} 0.$$

La relation $(*)$ prouve alors que $(\mathrm{F}_n)_{n\geqslant 1}$ converge uniformément vers F.

Une démonstration plus générale, utilisant la convergence étroite, peut être trouvée dans de nombreux ouvrages (par exemple [53]). ∎

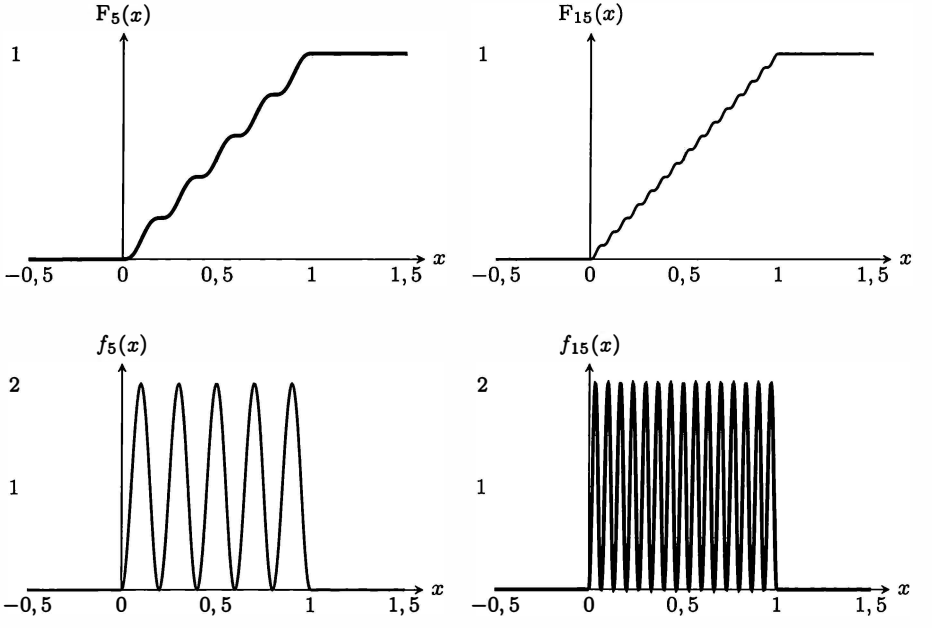


FIGURE 19.1 — En haut : *les fonctions de répartition* $\mathrm{F}_n(x) = x - \sin(2\pi nx)/2\pi n$ *pour les valeurs* $n = 5$ *et* $n = 15$. En bas : *les densités de probabilité* $f_n(x) = 1 - \cos(2\pi nx)$ *correspondantes.*

Convergence en loi et fonctions caractéristiques § 292

La convergence en loi est équivalente à la condition « $\mathbf{E}\big(f(X_n)\big) \to \mathbf{E}\big(f(X)\big)$ pour toute fonction f continue et bornée ». Le théorème de Paul Lévy stipule que l'on peut extraordinairement réduire la classe des fonctions f à tester : il suffit que la condition précédente soit vraie pour toutes les exponentielles complexes $x \mapsto e^{itx}$.

Nous reprenons l'énoncé du théorème que nous avions cité dans le chapitre 17, dans une version un peu plus élaborée.

■ **Théorème 19.44 (th. de continuité de Paul Lévy)** *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires, toutes à valeurs dans le même espace $\mathbb{R}$ ou $\mathbb{R}^n$ (mais pas forcément définies sur le même espace probabilisé). Notons $(\varphi_n)_{n\geqslant 1}$ leurs fonctions caractéristiques respectives.* ◀◀

Soit enfin X une variable aléatoire à valeurs dans le même espace $\mathbb{R}$ ou $\mathbb{R}^n$.

i) Si $X_n \xRightarrow{L} X$, alors $(\varphi_n)_{n\geqslant 1}$ converge simplement vers la fonction caractéristique φ de X ; de plus, cette convergence est uniforme sur tout segment de $\mathbb{R}$.

ii) Réciproquement, si la suite $(\varphi_n)_{n\geqslant 1}$ converge simplement vers la fonction φ caractéristique de X, alors $(X_n)_{n\geqslant 1}$ converge en loi vers X.

iii) Plus généralement, si la suite $(\varphi_n)_{n\geqslant 1}$ converge simplement vers une fonction ψ qui est de plus continue en 0, *alors ψ est la fonction caractéristique d'une (unique) loi $\mathscr{L}$, et la suite $(X_n)_{n\geqslant 1}$ converge en loi vers $\mathscr{L}$.*

C'est un théorème dont la partie réciproque est difficile ; nous l'admettons ici. Le lecteur désireux de lire une démonstration de ce théorème peut se reporter, par exemple, à Billingsley [9, section 26] ou à Zuily & Queffelec [102].

Remarque 19.45 La convergence simple de $(\varphi_n)_{n\geqslant 1}$ est aisée à montrer sous l'hypothèse que $X_n \xRightarrow{L} X$. Fixons momentanément $t \in \mathbb{R}$; alors $\mathbf{E}\big(f(X_n)\big) \to \mathbf{E}\big(f(X)\big)$ pour la fonction continue et bornée $f : x \mapsto e^{itx}$, c'est-à-dire que la suite de terme général $\varphi_n(t)$ converge vers $\varphi(t)$. La convergence uniforme est plus délicate, mais la véritable difficulté du théorème est le sens réciproque.

Enfin, puisque la fonction caractéristique d'une variable aléatoire est toujours continue, le théorème de Lévy revient à dire que, si $(\varphi_n)_{n\geqslant 1}$ converge simplement vers une fonction ψ, alors il y a équivalence entre les énoncés

i) ψ est continue en 0 ;

ii) ψ est la fonction caractéristique d'une variable aléatoire X ;

iii) ψ est la fonction caractéristique d'une variable aléatoire X et $X_n \xRightarrow{L} X$.

Convergence en loi vers une constante § 293

Il existe un cas particulier notable où la convergence en loi implique la convergence en probabilité.

■ **Théorème 19.46** *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires définies sur le même espace $(\Omega, \mathfrak{T}, \mathbf{P})$, et convergeant en loi vers une variable aléatoire X presque sûrement égale à la constante a. Alors la convergence a lieu en probabilité :*

$$\big(X_n \xRightarrow{L} a\big) \implies \big(X_n \xrightarrow{P} a\big).$$

DÉMONSTRATION : Notons $\psi(x) := x \wedge 1$; la fonction $f : t \mapsto \psi\big(|t-a|\big)$ est continue et bornée, donc la convergence en loi de X_n vers X permet d'écrire que $\lim_{n\to\infty} \mathbf{E}f(X_n) = \mathbf{E}f(X) = 0$ puisque $f(X) = 0$ p.s. Le corollaire 19.14 permet de conclure que $X_n \xrightarrow{P} X$. ∎

§ 294 **Convergence en loi d'un couple, théorème de Slutsky**

Si $X_n \xrightarrow{L} X$ et si $Y_n \xrightarrow{L} Y$, on pourrait s'attendre à des résultats du style : $X_n Y_n \xrightarrow{L} XY$, $X_n + Y_n \xrightarrow{L} X + Y$, voire $(X_n, Y_n) \xrightarrow{L} (X, Y)$. Hélas, ces résultats ne sont pas vrais en toute généralité[8].

Cependant, il existe un cas particulier où la convergence souhaitée a lieu, c'est lorsque Y_n converge vers une constante. Cette situation est très fréquente lorsque l'on cherche à construire des intervalles de confiance, et c'est principalement dans ce cas-là que l'on utilisera le théorème de Slutsky (voir chapitre 22).

■ **Théorème 19.47 (Slutsky)** *Soient* $(X_n)_{n\geqslant 1}$ *et* $(Y_n)_{n\geqslant 1}$ *deux suites de variables aléatoires, définies sur le* même *espace probabilisé* $(\Omega, \mathfrak{T}, \mathbf{P})$, *à valeurs respectivement dans* $\mathbb{R}^d$ *et* $\mathbb{R}^m$. *On suppose qu'il existe une variable aléatoire* X *sur* $(\Omega, \mathfrak{T}, \mathbf{P})$ *et une constante* $a \in \mathbb{R}^m$ *telles que :*

— $(X_n)_{n\geqslant 1}$ *converge en loi vers* X ;

— $(Y_n)_{n\geqslant 1}$ *converge en loi vers* a.

Alors la suite $(X_n, Y_n)_{n\geqslant 1}$ *converge en loi vers* (X, a).

Le corollaire qui nous intéresse au premier plan est celui-ci :

■ **Corollaire 19.48 (Slutsky)** *Si* $X_n \xrightarrow{L} X$ *et si* $Y_n - X_n \xrightarrow{P} 0$, *alors* $Y_n \xrightarrow{L} X$.

Si le théorème général est difficile à démontrer[9], ce corollaire pratique découle assez directement du théorème 19.32 page 395.

DÉMONSTRATION : Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction bornée et lipschitzienne. Il existe deux constantes positives M et k telles que $|f(x)| \leqslant M$ et $|f(y) - f(x)| \leqslant k\,|y - x|$ pour tous réels x et y. Soit $\varepsilon > 0$. Pour tout $n \in \mathbb{N}$, on a

$$|\mathbf{E}(f(Y_n) - f(X))| \leqslant \underbrace{\mathbf{E}|f(Y_n) - f(X_n)|}_{a} + \underbrace{\mathbf{E}|f(X_n) - f(X)|}_{b}.$$

On peut majorer le terme « a » :

$$\begin{aligned} a &= \mathbf{E}\big[|f(Y_n) - f(X_n)|\,\mathbf{1}_{\{|X_n - Y_n| > \varepsilon\}}\big] + \mathbf{E}\big[|f(Y_n) - f(X_n)|\,\mathbf{1}_{\{|X_n - Y_n| \leqslant \varepsilon\}}\big] \\ &\leqslant 2M\,\mathbf{P}\big\{|Y_n - X_n| > \varepsilon\big\} + k\varepsilon. \end{aligned}$$

Le premier terme du membre de droite vers 0 quand n tend vers l'infini par définition de la convergence en probabilité ; de même, le terme « b » tend vers 0 par convergence en loi de X_n vers X. Ainsi, pour n suffisamment grand, on aura $|\mathbf{E}(f(Y_n) - f(X))| \leqslant (k+2)\varepsilon$.

On a donc prouvé que $|\mathbf{E}(f(Y_n) - f(X))| \to 0$. Ceci étant vrai pour tout choix de fonction f bornée et lipschitzienne, le théorème 19.32 montre que $Y_n \xrightarrow{L} X$. ∎

Remarque 19.49 On peut énoncer, grâce au théorème général, de nombreuses autres propriétés — comme la suivante, que le lecteur pourra démontrer à titre d'exercice : *si* $(X_n)_{n\geqslant 1}$ *et* $(Y_n)_{n\geqslant 1}$ *sont deux suites de variables aléatoires réelles, telles que que* $X_n \xrightarrow{L} X$ *et que* $Y_n \xrightarrow{L} a$, *où* a *est une (variable aléatoire) constante, alors* $X_n Y_n \xrightarrow{L} aX$.

---

(8). Des contre-exemples sont donnés dans Ouvrard, tome II [77]. Ils sont vrais en revanche si l'on suppose que X et Y sont indépendantes et que, pour tout entier n, les variables aléatoires X_n et Y_n sont également indépendantes. Il suffit, pour le montrer, de remarquer que la fonction caractéristique du couple (X_n, Y_n) est

$$\varphi_{(X_n, Y_n)}(u, v) = \varphi_{X_n}(u) \cdot \varphi_{Y_n}(v)$$

et que le membre de droite converge vers $\varphi_X(u)\,\varphi_Y(v)$ d'après le théorème de Lévy (sens direct), ce qui est encore égal, par indépendance de X et Y, à $\varphi_{(X,Y)}(u, v)$. Le théorème de Lévy, dans le sens réciproque cette fois-ci, permet de conclure que (X_n, Y_n) converge en loi vers (X, Y). Enfin, les fonctions $(x, y) \mapsto x + y$ et $(x, y) \mapsto xy$ étant continues, on en déduit que $X_n + Y_n \xrightarrow{L} X + Y$ et $X_n Y_n \xrightarrow{L} XY$.

(9). Cf. par exemple Ouvrard [77].

19.4 UNE APPLICATION : PRODUIT SCALAIRE DE VECTEURS UNITAIRES ALÉATOIRES

On pose la question suivante : *si $\boldsymbol{u}$ et $\boldsymbol{v}$ sont deux vecteurs choisis aléatoirement sur la sphère unité S_{n-1} de l'espace euclidien $\mathbb{R}^n$, quelle est la loi de leur produit scalaire $\mathrm{X} = \langle \boldsymbol{u} | \boldsymbol{v} \rangle$?*

Vecteur aléatoire sur la sphère S_n § 295

La première chose à faire est évidemment de définir proprement ce qu'on appelle « choisir aléatoirement des vecteurs sur la sphère unité ». Il s'agit donc de munir

$$\mathrm{S}_{n-1} = \left\{ \boldsymbol{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \ ; \ x_1^2 + x_2^2 + \cdots + x_n^2 = 1 \right\}$$

d'une probabilité uniforme ; pour cela, nous allons paramétrer S_{n-1} en coordonnées polaires.

Rappelons qu'un vecteur $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ de $\mathbb{R}^n$ peut s'écrire sous la forme

$$\begin{cases} x_1 = r \ \cos\theta_1 \\ x_2 = r \ \sin\theta_1 \ \cos\theta_2 \\ x_3 = r \ \sin\theta_1 \ \sin\theta_2 \ \cos\theta_3 \\ x_4 = r \ \sin\theta_1 \ \sin\theta_2 \ \sin\theta_3 \ \cos\theta_4 \\ \quad \vdots \\ x_n = r \ \sin\theta_1 \ \sin\theta_2 \ \sin\theta_3 \ \sin\theta_4 \cdots \sin\theta_{n-2} \ \cos\theta_{n-1}, \end{cases}$$

où $\theta_1, \dots, \theta_{n-2} \in [\,0\,;\,\pi\,]$ et $\theta_{n-1} \in [\,0\,;\,2\pi\,]$. Le jacobien de l'application

$$\Phi : (r, \theta_1, \theta_2, \dots, \theta_{n-1}) \longmapsto (x_1, x_2, \dots, x_n)$$

est

$$\mathrm{J}_n = r^{n-1} \ (\sin\theta_1)^{n-2} \ (\sin\theta_2)^{n-3} \ \cdots \ (\sin\theta_{n-3})^2 \ (\sin\theta_{n-2}) \qquad (19.2)$$

(exercice 19.11 page 413). Il induit sur la sphère S_{n-1} une mesure

$$\mathrm{dS} = (\sin\theta_1)^{n-2} \ (\sin\theta_2)^{n-3} \ \cdots \ (\sin\theta_{n-3})^2 \ (\sin\theta_{n-2}) \cdot \mathrm{d}\theta_1 \, \mathrm{d}\theta_2 \ \dots \ \mathrm{d}\theta_{n-1}.$$

Loi du produit scalaire § 296

Le problème étant invariant par rotation, il est équivalent au problème suivant : le vecteur $\boldsymbol{v} = (1, 0, \dots, 0)$ étant fixé dans S_{n-1}, on cherche la loi de $\mathrm{X}_n = \langle \boldsymbol{u} | \boldsymbol{v} \rangle$ lorsque $\boldsymbol{u}$ parcourt S_{n-1} avec la mesure de probabilité uniforme sur S_{n-1}, donnée par

$$\mathbf{P}(\mathrm{d}\theta_1, \dots, \mathrm{d}\theta_{n-1}) = \frac{1}{\sigma_n} \ (\sin\theta_1)^{n-2} \ (\sin\theta_2)^{n-3} \ \cdots \\ \cdots \ (\sin\theta_{n-3})^2 \ (\sin\theta_{n-2}) \cdot \mathrm{d}\theta_1 \, \mathrm{d}\theta_2 \ \dots \ \mathrm{d}\theta_{n-1}$$

où $\sigma_n = \int_{\mathrm{S}_{n-1}} \mathrm{dS}$ est la constante de normalisation. Or, si $\boldsymbol{u} = (x_1, x_2, \dots, x_n)$ est repéré par ses coordonnées polaires $(\theta_1, \theta_2, \dots, \theta_{n-1})$, on a $\langle \boldsymbol{u} | \boldsymbol{v} \rangle = \cos\theta_1$, de sorte que, si c est un réel tel que $-1 \leqslant c \leqslant 1$, et que l'on pose $\alpha = \operatorname{Arc\,cos} c$, on obtient

$$\mathbf{P}\{ \langle \boldsymbol{u} | \boldsymbol{v} \rangle \geqslant c \} = \mathbf{P}\{\cos\theta_1 \geqslant c\} = \mathbf{P}\{0 \leqslant \theta_1 \leqslant \alpha\} = \mathrm{K}_n \int_0^\alpha (\sin\theta_1)^{n-2} \, \mathrm{d}\theta_1,$$

où K_n est le résultat de l'intégration de dS sur les variables θ_2 à θ_{n-1}. Notons F_n la fonction de répartition de X_n, on a donc montré

$$\forall c \in [-1\,;1] \qquad 1 - F_n(c) = K_n \int_0^{\operatorname{Arc}\cos c} (\sin\theta_1)^{n-2}\,\mathrm{d}\theta_1,$$

ce qui montre que X_n est absolument continue et admet une densité f_n, nulle en dehors de $[-1\,;+1]$, et valant

$$f_n(c) = -K_n\,(\sin \operatorname{Arc}\cos c)^{n-2} \cdot \operatorname{Arc}\cos'(c) = K_n\big(\sqrt{1-c^2}\big)^{n-3}$$

sur $]-1\,;+1[$. On en déduit également la valeur de la constante :

$$K_n = \frac{\Gamma(n/2)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)}.$$

Remarque 19.50 Lorsque $n = 3$, la loi trouvée est la loi uniforme sur $]-1\,;+1[$. Lorsque $n = 2$, on trouve la loi de l'Arc sinus. Lorsque $n = 4$, c'est la **loi du demi-cercle**.

§ 297 **Loi limite**

Les densités trouvées au paragraphe précédent se concentrent au voisinage de 0 quand la dimension n de l'espace augmente, comme on peut le voir sur la figure 19.2.

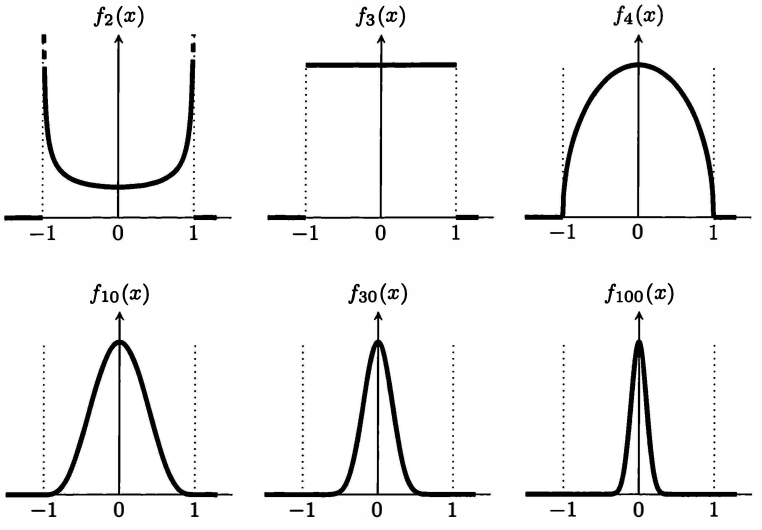


FIGURE 19.2 — *La fonction de densité de X_n, pour différentes valeurs de la dimension n. (L'échelle en ordonnée n'est pas respectée.)*

Il est alors aisé de constater que la suite $(X_n)_{n\geqslant 1}$ converge en loi[10] vers la (loi de la) variable aléatoire constante égale à 0, comme c'est le cas chaque fois que la suite des densités est une suite d'approximations de l'unité (suite de Dirac, pour les physiciens). Heuristiquement, on peut ainsi dire que :

en grande dimension, (presque) tous les vecteurs unitaires sont (presque) orthogonaux.

(10). On rappelle qu'il n'est nul besoin que les variables aléatoires X_n soient définies sur le même espace.

19.5 SECONDE APPLICATION : SUITES ÉQUIDISTRIBUÉES

Suites équidistribuées § 298

On dit qu'une suite $(u_k)_{k\geqslant 1}$, à valeurs dans $[\,0\,;\,1\,[$, est **uniformément distribuée**, ou **équidistribuée** si la densité moyenne d'éléments de la suite est uniforme sur $[\,0\,;\,1\,]$. Pour donner un sens rigoureux à cette idée, on demande que, pour tout intervalle $\mathrm{I} \subset [\,0\,;\,1\,]$,

$$\lim_{n\to\infty} \frac{\operatorname{Card}\{k \in [\![1\,;\,n]\!]\ ;\ u_k \in \mathrm{I}\}}{n} = \ell(\mathrm{I}),$$

où l'on a noté $\ell(\mathrm{I})$ la longueur de I. On rappelle que, si $\mathrm{I} = [\,a\,;\,b\,]$ ou $\mathrm{I} =]\,a\,;\,b\,[$, cette longueur vaut $b-a$ et qu'on peut également l'écrire $\ell(\mathrm{I}) = \int_0^1 \mathbf{1}_{\mathrm{I}}(t)\,\mathrm{d}t$. Cette écriture intégrale permet de reformuler la définition précédente de manière plus agréable :

Définition 19.51 Une suite $(u_k)_{k\geqslant 1}$ à valeurs dans $[\,0\,;\,1\,[$ est dite **équidistribuée** si

$$\frac{1}{n}\sum_{k=1}^{n} f(u_k) \xrightarrow[n\to\infty]{} \int_0^1 f \tag{19.3}$$

pour toute fonction f indicatrice d'un intervalle de $[\,0\,;\,1\,[$.

Si l'on note δ_{u_k} la mesure de Dirac en u_k, la condition précédente revient à dire qu'on a la convergence en mesure (« en loi » par abus de langage) suivante :

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{u_k} \overset{\text{L}}{\Longrightarrow} \mathscr{U}\big([\,0\,;\,1\,[\big).$$

En effet, si X_n est une variable aléatoire de loi $\mathbf{P}_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{u_k}$, sa fonction de répartition est alors

$$\mathrm{F}_n(x) = \int_0^x \mathrm{d}\mathbf{P}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{[\,0\,;\,x\,[}(u_k)$$

et la suite $(u_k)_{k\geqslant 0}$ est équidistribuée si et seulement si, pour tout x, la quantité $\mathrm{F}_n(x)$ tend vers $\int \mathbf{1}_{[\,0\,;\,x\,[} = x$.

Critère de Weyl § 299

Le critère suivant d'équidistribution d'une suite a été introduit par Hermann WEYL [99] en 1916[11] :

■ **Théorème 19.52 (Weyl)** *Une suite $(u_k)_{k\geqslant 1}$ à valeurs dans $[\,0\,;\,1\,[$ est équidistribuée si et seulement si, pour tout $m \in \mathbb{Z}$ non nul,*

$$\frac{1}{n}\sum_{k=1}^{n} \mathrm{e}^{2\mathrm{i}\pi m u_k} \xrightarrow[n\to\infty]{} 0. \tag{19.4}$$

(Si $m = 0$, le membre de gauche est évidemment toujours égal à 1.)

(11). Le critère d'équidistribution modulo 1 de la suite $(na)_{n\in\mathbb{N}}$ lorsque a est irrationnel a été prouvé par H. WEYL en 1910 et, indépendamment, par Piers BOHL (1909) et Wacław SIERPIŃSKI (1910).

DÉMONSTRATION : Supposons la suite équidistribuée. La convergence en loi permet d'écrire que

$$\frac{1}{n}\sum_{k=1}^{n} f(u_k) \xrightarrow[n\to\infty]{} \int_0^1 f$$

pour toute fonction f continue et bornée. C'est vrai notamment pour les fonctions du type $x \mapsto \mathrm{e}^{2\mathrm{i}\pi m x}$ pour m entier non nul, ce qui prouve le sens direct(12).

Réciproquement, supposons que la relation (19.4) a lieu pour tout $m \in \mathbb{Z}^*$. Alors la relation (19.3) est valable pour tout polynôme trigonométrique : si $\mathrm{P}(x) = \sum_{j=-d}^{d} c_j \mathrm{e}^{2\mathrm{i}\pi j x}$, alors

$$\mathrm{S}_n(\mathrm{P}) := \frac{1}{n}\sum_{k=1}^{n} \mathrm{P}(u_k) \xrightarrow[n\to\infty]{} c_0 = \int_0^1 \mathrm{P}(t)\,\mathrm{d}t. \tag{19.5}$$

Par densité des polynômes trigonométriques dans l'espace des fonctions continues et 1-périodiques muni de la norme infinie (A36), nous allons montrer que :

■ **Lemme 19.53** *La relation* (19.3) *est encore valide pour toute fonction continue sur le segment* $[0\,;1]$ *et vérifiant* $f(0) = f(1)$.

DÉMONSTRATION DU LEMME : Soit $f : [0\,;1] \to \mathbb{C}$ une fonction continue telle que $f(0) = f(1)$. Soit $\varepsilon > 0$. Il existe un polynôme trigonométrique P tel que $\|\mathrm{P} - f\|_\infty \leqslant \varepsilon$. Par la relation (19.5), il existe un entier n_0 tel que, pour tout $n \geqslant n_0$, on ait $\left|\mathrm{S}_n(\mathrm{P}) - \int_0^1 \mathrm{P}\right| \leqslant \varepsilon$. Alors,

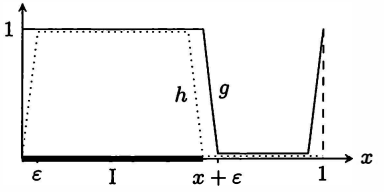
$$\left|\mathrm{S}_n(f) - \int_0^1 f\right| \leqslant \left|\mathrm{S}_n(f) - \mathrm{S}_n(\mathrm{P})\right| + \left|\mathrm{S}_n(\mathrm{P}) - \int_0^1 \mathrm{P}\right| + \int_0^1 |\mathrm{P} - f| \leqslant 3\varepsilon. \quad ▮$$

Nous ne pouvons évidemment pas appliquer ce résultat à une fonction f indicatrice, qui ne vérifie *a priori* pas $f(0) = f(1)$, mais on peut contourner cette difficulté.

Soit donc I un intervalle de $[0\,;1]$; il est clair que l'on peut se contenter de choisir un intervalle de la forme $[0\,;x]$(13) ; notons $f = \mathbf{1}_\mathrm{I}$ sa fonction indicatrice. On construit alors deux fonctions g et h vérifiant

$$0 \leqslant h \leqslant f \leqslant g \qquad \text{et} \qquad 0 \leqslant \int_0^1 (g - h) \leqslant \varepsilon,$$

selon le schéma ci-dessous :



Par le lemme précédent, on peut trouver un rang N tel que, pour tout $n \geqslant \mathrm{N}$, on ait

$$\left|\mathrm{S}_n(g) - \int_0^1 g\right| \leqslant \varepsilon \qquad \text{et} \qquad \left|\mathrm{S}_n(h) - \int_0^1 h\right| \leqslant \varepsilon.$$

---

(12). On peut donner également l'argument suivant : la convergence en loi implique la convergence simple des fonctions caractéristiques. Or, la fonction caractéristique de la loi $\mathbf{P}_n$ est

$$\varphi_n(2\pi\nu) = \int_0^1 \mathrm{e}^{2\mathrm{i}\pi\nu x}\,\mathrm{d}\mathbf{P}_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{e}^{2\mathrm{i}\pi\nu u_k},$$

tandis que la fonction caractéristique de la loi uniforme sur $[0\,;1[$ est

$$\varphi(2\pi\nu) = \int_0^1 \mathrm{e}^{2\mathrm{i}\pi\nu x}\,\mathrm{d}x = \frac{\mathrm{e}^{2\mathrm{i}\pi\nu} - 1}{2\mathrm{i}\pi\nu}$$

pour tout $\nu \neq 0$, et notamment $\varphi(2\pi m) = 0$ pour tout $m \in \mathbb{Z}^*$.

(13). Par différence, on pourra montrer la relation (19.3) pour toute fonction indicatrice, ou même remarquer que les intervalles $[0\,;x]$ mènent directement à la fonction de répartition, et permettent de conclure sur la convergence en loi.

On a alors la majoration

$$\left|\int_0^1 f - \mathrm{S}_n(f)\right| \leqslant \underbrace{\int_0^1 |f-g|}_{\leqslant \varepsilon} + \underbrace{\left|\int_0^1 g - \mathrm{S}_n(g)\right|}_{\leqslant \varepsilon} + |\mathrm{S}_n(g) - \mathrm{S}_n(f)| \leqslant 2\varepsilon + |\mathrm{S}_n(g) - \mathrm{S}_n(f)|.$$

De plus, pour tout $n \geqslant \mathrm{N}$, on a

$$0 \leqslant \mathrm{S}_n(g) - \mathrm{S}_n(f) \leqslant \mathrm{S}_n(g) - \mathrm{S}_n(h) \leqslant \left|\mathrm{S}_n(g) - \int_0^1 g\right| + \int_0^1 |g-h| + \left|\int_0^1 h - \mathrm{S}_n(h)\right| \leqslant 3\varepsilon,$$

ce qui donne

$$\left|\int_0^1 f - \mathrm{S}_n(f)\right| \leqslant 5\varepsilon$$

et conclut la preuve. ∎

Ce théorème s'interprète heuristiquement de la manière suivante. On considère une marche aléatoire de pas de longueur 1, dans la direction « aléatoire » $2m\pi u_k$. Si la suite $(u_n)_{n\geqslant 1}$ remplit « uniformément » l'intervalle $[\,0\,;\,1\,[$, alors les différents pas devraient, statistiquement, se compenser, et le marcheur ne s'éloignera jamais si vite que, après une homothétie de rapport $1/n$, sa distance ne tende vers 0.

Application aux suites ($n\alpha \mod 1$) § 300

Notons $\langle x\rangle = x - \lfloor x\rfloor$ la partie fractionnaire d'un réel x. Alors on déduit facilement du critère de Weyl le résultat suivant :

■ **Théorème 19.54 (Weyl-Jacobi)** *Soit α un nombre irrationnel. Alors la suite $\langle n\alpha\rangle$ est équidistribuée.*

DÉMONSTRATION : Il suffit de montrer que cette suite vérifie le critère de Weyl. Soit $m \in \mathbb{Z}^*$. Le nombre α étant irrationnel, $2\pi m\alpha$ ne saurait être congru à 0 modulo 2π, et donc $z := \mathrm{e}^{2\mathrm{i}\pi m\alpha} \neq 1$. On a alors $\mathrm{e}^{2\mathrm{i}\pi m\langle n\alpha\rangle} = \mathrm{e}^{2\mathrm{i}\pi mn\alpha} = z^n$, donc

$$\frac{1}{n}\sum_{k=1}^{n} \mathrm{e}^{2\mathrm{i}\pi m\langle k\alpha\rangle} = \frac{1}{n}\sum_{k=1}^{n} z^k = \frac{1}{n}\cdot\frac{z^{n+1}-z}{1-z} \xrightarrow[n\to\infty]{} 0$$

ce qui achève la preuve. ∎

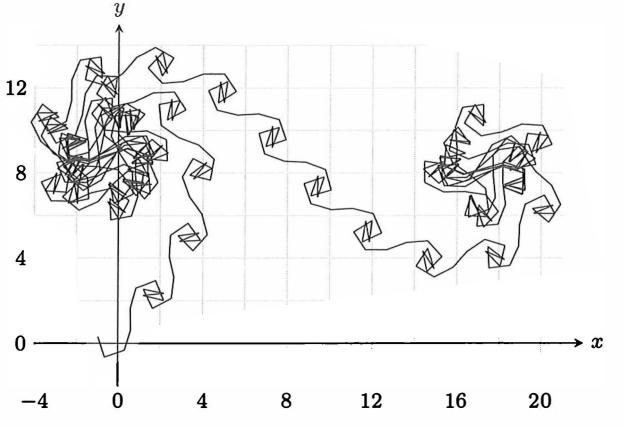


FIGURE 19.3 — *Suite des 500 premières valeurs de $\sum_{k=1}^{n} \mathrm{e}^{2\mathrm{i}\pi u_k}$ pour la suite (équidistribuée) $u_k = k^2\sqrt{2}/\pi$. On remarquera la beauté du motif.*

Remarque 19.55 (Autres suites équidistribuées) Si α est un irrationnel, il est possible de fabriquer d'autres suites équidistribuées, par exemple :

— la suite de terme général $n^2\alpha$ (Weyl, 1916, voir figure 19.3) ;

— la suite de terme général $p_n\,\alpha$, où p_n est le n-ième entier premier (I. M. Vinogradov, 1935) ; voir la figure 19.4 et *Les nombres premiers* de Gérald Tenenbaum et Michel Mendès France [97] ;

— la suite de terme général $2^n\,\alpha$, mais seulement pour presque tous les irrationnels α.

Il existe par ailleurs un théorème remarquable : pour presque tout réel $\alpha > 1$, la suite de terme général α^n est équidistribuée modulo 1. (Par contre, on ne connaît pas encore avec certitude un seul de ces réels... Voir par exemple l'article de Dress et Mendès France paru dans *La Recherche* [26].)

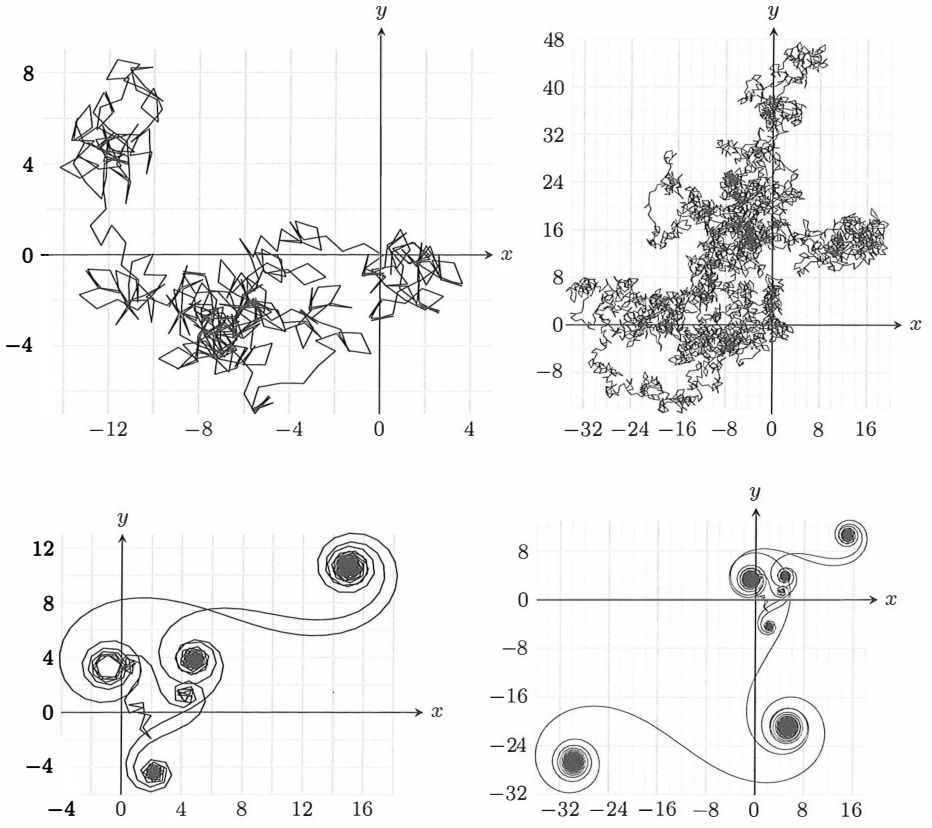


FIGURE 19.4 — (En haut, à gauche :) *les 500 premières valeurs de* $\sum_{k=1}^{n} \mathrm{e}^{2\mathrm{i}\pi u_k}$ *pour la suite équidistribuée* $u_k = p_k\sqrt{2}$, p_k *étant le* k*-ième entier premier. La courbe obtenue a l'allure d'une marche aléatoire.* (En haut, à droite) : *les 6000 premières valeurs de la même suite, confirmant bien l'allure « aléatoire » de la marche.* (En bas, à gauche) : *les 500 premières valeurs de la suite associée à* $u_k = \sqrt{2}\,k\ln k$. *Cette fois-ci, la marche est tout sauf aléatoire ! La succession des spirales continue si l'on trace les 2500 premières valeurs* (en bas, à droite). *La distance à l'origine croît lentement, laissant subodorer que la suite* $(u_k)_k$ *est équidistribuée (ce qui est le cas).*

Exemple 19.56 (Une suite non équidistribuée) Bien que la suite de terme général $u_n = \langle \ln n \rangle$ soit dense dans $[\,0\,;\,1\,[$, elle n'est pas équidistribuée (elle est « plus dense » vers 1 que vers 0). Le graphe de la suite des sommes partielles $\mathrm{S}_n = \sum_{k=1}^{n} \mathrm{e}^{2\mathrm{i}\pi \ln k}$ présente une forme de spirale, et on peut en effet montrer (en prenant $m = 1$) que $|\mathrm{S}_n| \sim n/\sqrt{1+4\pi^2}$, donc que $\mathrm{S}_n/n \not\to 0$.

19.6 RÉSUMÉ SYNOPTIQUE

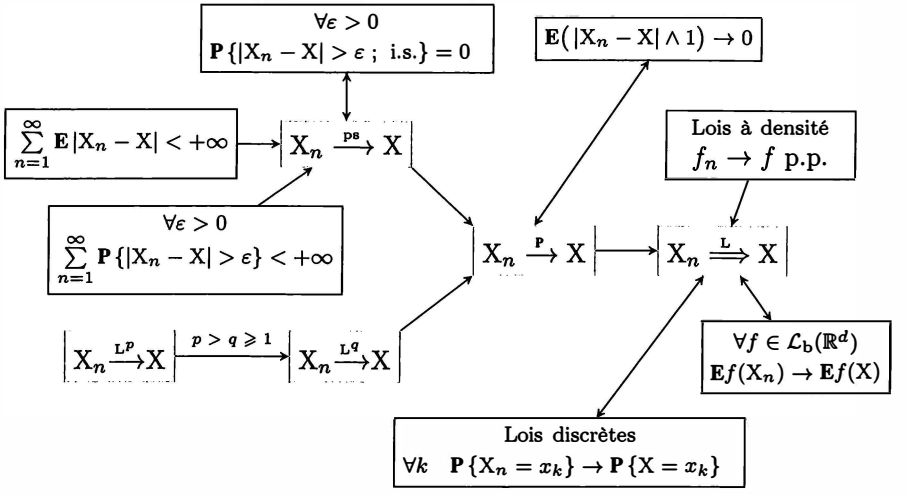


FIGURE 19.5 — *Relations entre les principaux types de convergence.*

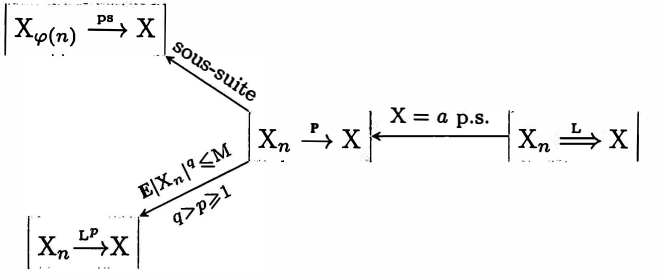


FIGURE 19.6 — *Réciproques conditionnelles entre les principaux types de convergence.*

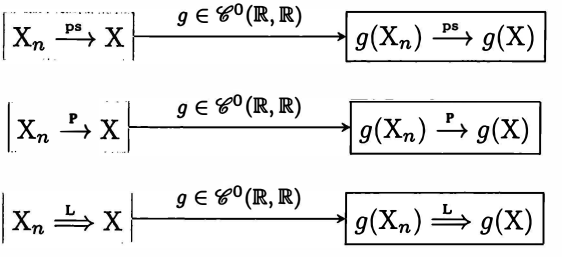


FIGURE 19.7 — *Convergence des images de variables aléatoires.*

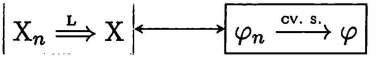


FIGURE 19.8 — *Convergence en loi et fonctions caractéristiques*

19.7 ANNEXE : CONVERGENCE ÉTROITE

§ 301 **Théorème de représentation de Skorokhod**

Bien que la convergence en loi n'implique nullement la convergence en probabilité ni presque sûre, le théorème suivant montre que l'on peut *représenter* une convergence en loi par une suite de variables aléatoires qui converge presque sûrement.

■ **Théorème 19.57 (Skorokhod)** *Soit* $(X_n)_{n\geqslant 1}$ *une suite de variables aléatoires qui converge en loi vers* X. *Alors il existe un espace probabilisé* $(\Omega^*, \mathfrak{T}^*, \mathbf{P}^*)$ *et des variables aléatoires* $(Y_n)_{n\geqslant 1}$ *et* Y *sur cet espace telles que* Y_n *a même loi que* X_n, Y *a même loi que* X *et* $Y_n \xrightarrow{ps} Y$.

DÉMONSTRATION : On va choisir pour espace probabilisé l'espace $\Omega^* = [0\,;\,1]$ muni de la tribu des boréliens et de la mesure habituelle de Lebesgue. On note F_n et F les fonctions de répartition de X_n et X, et on note Y_n et Y leurs pseudo-inverses :

$$Y_n(\omega) = \inf\{x \in \mathbb{R}\,;\, F_n(x) \geqslant \omega\} \qquad Y(\omega) = \inf\{x \in \mathbb{R}\,;\, F(x) \geqslant \omega\}.$$

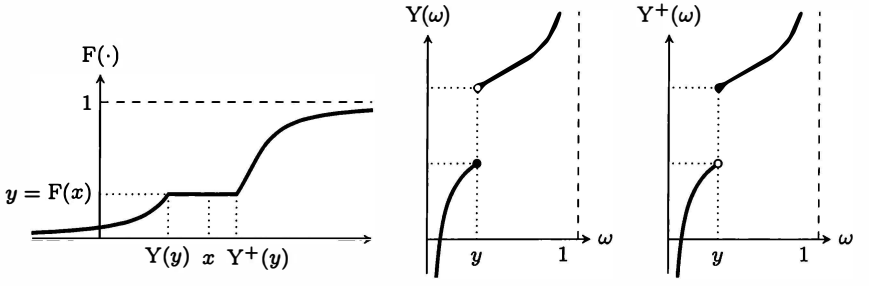
Alors (théorème A.24 page 680) Y_n a pour fonction de répartition F_n, et de même Y admet F pour fonction de répartition.

Il reste à montrer que $Y_n \to Y$ presque sûrement ; nous allons montrer que la convergence a lieu en tout point de continuité de Y, ce qui répondra à la question car l'ensemble des discontinuités de Y est dénombrable (théorème A.20 page 677).

Pour cela, définissons

$$Y_n^+(\omega) = \inf\{x \in \mathbb{R}\,;\, F_n(x) > \omega\} \qquad Y^+(\omega) = \inf\{x \in \mathbb{R}\,;\, F(x) > \omega\}.$$

La différence entre Y et Y^+ est visible sur la figure ci-dessous :



Soit maintenant ω un réel fixé dans $[0\,;\,1]$. Si z est un point de continuité de F tel que $z > Y^+(\omega)$, alors $F(z) > \omega$ et, pour tout entier n à partir d'un certain rang, $F_n(z) > \omega$, donc $z \geqslant Y_n^+(\omega)$. Notamment, $\limsup Y_n^+(\omega) \leqslant z$. Comme l'ensemble des points de continuité de F est le complémentaire d'un ensemble dénombrable, on peut appliquer ce résultat à une suite $(z_k)_{k\geqslant 1}$ qui décroît vers $Y^+(\omega)$, et on obtient

$$\limsup Y_n^+(\omega) \leqslant Y^+(\omega).$$

On effectue le même travail de l'autre côté : si z est un point de continuité de F tel que $z < Y(\omega)$, alors $F(z) < \omega$ et de même, à partir d'un certain rang, $F_n(z) < \omega$ donc $z < Y_n(\omega)$. Notamment, $\liminf Y_n(\omega) \geqslant z$. On applique ceci à une suite $(z_k)_{k\geqslant 1}$ qui croît vers $Y(\omega)$ et on obtient

$$Y(\omega) \leqslant \liminf Y_n(\omega).$$

Lorsque ω est un point de continuité de Y, on a $Y(\omega) = Y^+(\omega)$, et les inégalités

$$Y(\omega) \leqslant \liminf Y_n(\omega) \leqslant \limsup Y_n(\omega) \leqslant \limsup Y_n^+(\omega) \leqslant Y^+(\omega) = Y(\omega)$$

montrent que $\liminf Y_n(\omega) = \limsup Y_n(\omega)$, c'est-à-dire que $Y_n(\omega) \to Y(\omega)$. ∎

§ 302 **Équivalence de la convergence en loi et de la convergence étroite**

Le théorème de représentation de Skorokhod permet de prouver que, si l'on se limite aux variables aléatoires réelles, la convergence en loi (définition 19.25) est équivalente à la convergence étroite (définition 19.30).

■ **Théorème 19.58** *Soient* $(X_n)_{n\geqslant 1}$ *une suite de variables aléatoires et* X *une variable aléatoire, toutes à valeurs dans* $\mathbb{R}$. *Alors il y a équivalence entre les énoncés*

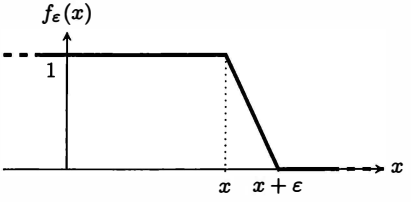
i) $X_n \xrightarrow{L} X$;

ii) pour toute fonction f *continue et bornée,* $\mathbf{E}f(X_n) \to \mathbf{E}f(X)$.

DÉMONSTRATION : Supposons que $X_n \xrightarrow{L} X$. Soit f une fonction continue et bornée sur $\mathbb{R}$. On utilise le théorème de représentation de Skorokhod et on fabrique $(Y_n)_{n\geqslant 1}$ et Y comme précisé. Par continuité de f, $f(Y_n) \xrightarrow{ps} f(Y)$; or $|f(Y_n)| \leqslant \|f\|_\infty$ pour tout $n \geqslant 1$ donc, d'après le théorème de convergence dominée, $\mathbf{E}f(Y_n) \to \mathbf{E}f(Y)$.

Réciproquement, supposons que $\mathbf{E}f(X_n) \to \mathbf{E}f(X)$ pour toute fonction f continue et bornée. On veut montrer que $F_n(x) \to F(x)$ en tout point x de continuité de F. Soit donc x un tel point de continuité.

Fixons momentanément un réel $\varepsilon > 0$. On définit la fonction f_ε de la manière suivante :



Alors f_ε est continue et bornée, donc $\mathbf{E}f_\varepsilon(X_n) \to \mathbf{E}f_\varepsilon(X)$. De plus, on a les inégalités élémentaires

$$F_n(x) = \mathbf{P}\{X_n \leqslant x\} = \mathbf{E}(\mathbf{1}_{\{X_n \leqslant x\}}) \leqslant \mathbf{E}f_\varepsilon(X_n)$$

et

$$\mathbf{E}f_\varepsilon(X) \leqslant \mathbf{E}(\mathbf{1}_{\{X_n \leqslant x+\varepsilon\}}) = F(x+\varepsilon).$$

On en déduit que

$$\limsup F_n(x) \leqslant \limsup \mathbf{E}f_\varepsilon(X_n) = \mathbf{E}f_\varepsilon(X) \leqslant F(x+\varepsilon).$$

Ceci étant valable pour tout $\varepsilon > 0$, on en déduit finalement que

$$\limsup F_n(x) \leqslant F(x).$$

On effectue le même travail avec la fonction $g_\varepsilon : x \mapsto f_\varepsilon(x+\varepsilon)$, et on obtient

$$F(x) = F(x^-) \leqslant \liminf F_n(x).$$

De ces deux inégalités, on tire que $\lim F_n(x) = F(x)$. ∎

EXERCICES

♦ ***Exercice 19.1 (De l'uniforme à l'exponentielle)*** Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes, de même loi uniforme $\mathscr{U}([0\,;1])$. On $M_n := \max(X_1, X_2, \dots, X_n)$ et $Y_n := n(1-M_n)$ pour tout entier n non nul. Calculer la fonction de répartition de Y_n et montrer que la suite $(Y_n)_{n\geqslant 1}$ converge en loi vers une loi à préciser.

♦ ***Exercice 19.2*** Trouver des variables aléatoires X et X_n $(n \geqslant 1)$ telles que $X_n \to X$ presque sûrement, ainsi qu'une fonction $\varphi : \mathbb{R} \to \mathbb{R}$ telle que $\varphi(X_n)$ ne converge pas vers $\varphi(X)$ presque sûrement.

♦ ***Exercice 19.3*** On considère une suite de variables aléatoires $(X_n)_{n\geqslant 1}$ telle que X_n suit la loi uniforme sur $[0\,;\frac{1}{n}]$. Cette suite converge-t-elle en loi ?

On considère maintenant une suite de variables aléatoires $(Y_n)_{n\geqslant 1}$ telle que Y_n suit la loi uniforme sur $[0\,;n]$. Converge-t-elle en loi ?

♦ ***Exercice 19.4*** On suppose que $(X_n)_{n\geqslant 1}$ converge vers X dans L^2. Montrer que $\mathbf{V}(X_n)$ tend vers $\mathbf{V}(X)$.

♦ ***Exercice 19.5*** On définit la suite $(X_n)_{n\geqslant 1}$ de variables aléatoires discrètes par $X_n(\Omega) = \{1/n, 2/n, \dots, n/n\}$ et $\mathbf{P}\{X_n = k/n\} = \alpha_n k$ pour tout $n \geqslant 1$ et tout $k \in [\![1\,;n]\!]$.

Après avoir déterminé la valeur de α_n pour tout entier n non nul, montrer que $(X_n)_{n\geqslant 1}$ converge en loi vers $\sqrt{U}$, où U est une variable de loi uniforme sur $[0\,;1]$.

♦ ***Exercice 19.6 (Un critère de convergence p.s. d'une série)*** Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires admettant chacune une espérance finie. On suppose que la série $\sum \mathbf{E}|X_n|$ converge.

i) Pour tout $n \in \mathbb{N}^*$, on pose $S_n = |X_0| + |X_1| + \cdots + |X_n|$. Majorer la quantité $k\mathbf{P}\{S_n \geqslant k\}$ par une quantité indépendante de n et de k.

ii) Montrer que l'événement

$$A := \bigcap_{k=0}^{\infty} \bigcup_{n=1}^{\infty} \{S_n \geqslant k\}$$

est négligeable.

iii) En déduire que la série $\sum |X_n|$ converge presque sûrement, puis que $\sum X_n$ converge presque sûrement.

♦ ***Exercice 19.7 (Convergence en loi d'une suite de Dirac)*** On rappelle que, si a est un réel, la mesure de Dirac δ_a est définie sur $\mathfrak{B}$ par $\delta_a(A) = 1$ si $a \in A$ et $\delta_a(A) = 0$ sinon. Soit $(a_n)_{n\geqslant 1}$ une suite de réels. Donner une condition nécessaire et suffisante sur cette suite pour que $\delta_{a_n} \xrightarrow{L} \delta_0$.

♦ ***Exercice 19.8 (Convergence de variables presque sûres)*** Soient X et X_n (pour $n \geqslant 1$) des variables aléatoires réelles presque sûrement constantes : $X_n = a_n$ presque sûrement et $X = a$ presque sûrement. Montrer que les énoncés suivants sont équivalents :

i) $a_n \to a$;

ii) $X_n \xrightarrow{ps} X$;

iii) $X_n \xrightarrow{P} X$;

iv) $X_n \xrightarrow{L} X$.

♦ ***Exercice 19.9 (Minimum d'un nombre aléatoire de lois exponentielles)*** Soit $(N_n)_{n\geqslant 1}$ une suite de variables aléatoires telles que $N_n - 1$ suive la loi de Poisson $\mathscr{P}(n\lambda)$ pour tout entier $n \geqslant 1$. Soit $(M_{n,k})_{n,k\geqslant 1}$ une famille de variables aléatoires indépendantes suivant des lois exponentielles : $M_{n,k} \rightsquigarrow \mathscr{E}(\alpha/n)$. On note alors $S_n = \min(M_{n,1}, M_{n,2}, \dots, M_{n,N_n})$. Montrer que $(S_n)_{n\geqslant 1}$ converge en loi et déterminer la loi limite.

♦ ***Exercice 19.10 (CVPS $\Longrightarrow$ CVP)*** Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires et X une variable aléatoire, toutes définies sur le même espace $(\Omega, \mathfrak{T}, \mathbf{P})$. On suppose que $X_n \xrightarrow{ps} X$ et on veut montrer que $X_n \xrightarrow{P} X$ (selon une méthode différente de celle du théorème 19.16). Montrer que $\mathbf{E}(|X_n - X| \wedge 1) \to 0$ et conclure.

♦ ***Exercice 19.11 (Jacobien de la transformation polaire)*** Montrer la formule (19.2).

♦ ***Exercice 19.12 (Convergence en loi des images)*** Montrer le corollaire 19.36 sans utiliser la notion de convergence étroite (uniquement la définition de la convergence en loi avec les fonctions de répartition), mais en utilisant le théorème de représentation de Skorokhod.

♦ ***Exercice 19.13*** Soit $(Y_n)_{n\geqslant 0}$ une suite de variables aléatoires et identiquement distribuées. On suppose que Y_1 admet une espérance. Montrer que la suite de terme général Y_n/n converge presque sûrement vers 0.

♦ ***Exercice 19.14*** Soit $(E_n)_{n\geqslant 0}$ une suite de variables aléatoires de même loi exponentielle de paramètre $\lambda > 0$. On note S la somme de la série positive $\sum E_n/2^n$, variable aléatoire à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$.

Montrer que S est presque sûrement finie, c'est-à-dire $\mathbf{P}\{S = +\infty\} = 0$.

♦ ***Exercice 19.15 (Convergence en loi et en probabilité, mais pas presque sûre)*** On veut montrer que la convergence en probabilité n'implique pas la convergence presque sûre. Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes, telles que $X_n \rightsquigarrow \mathscr{B}(1/n)$. Montrer que $(X_n)_{n\geqslant 1}$ converge en loi et en probabilité vers la variable nulle, mais ne converge pas presque sûrement vers 0.

Indications

◊ *Indication 19.2 :* Bien entendu, la fonction φ ne doit pas être continue !

◊ *Indication 19.3 :* Montrer que la suite $(F_n)_{n\geqslant 1}$ des fonctions de répartition de $(Y_n)_{n\geqslant 1}$ converge simplement vers la fonction nulle.

◊ *Indication 19.4 :* Montrer d'abord que $\mathbf{E}X_n \to \mathbf{E}X$, puisque $\mathbf{E}X_n^2 \to \mathbf{E}X^2$.

◊ *Indication 19.7 :* Cf. exercice suivant.

◊ *Indication 19.11 :* Montrer que $\dfrac{\partial J_n}{\partial \theta_{n-1}} = 0$, et en déduire que

$$J_n(r, \theta_1, \dots, \theta_{n-2}, \theta_{n-1}) = J_n(r, \theta_1, \dots, \theta_{n-2}, 0) = r \sin\theta_1 \cdots \sin\theta_{n-2}\, J_{n-1},$$

puis conclure par récurrence. (On se souviendra que la dérivée d'un déterminant s'obtient comme somme des déterminants dans lesquelles on a dérivé successivement chacune des lignes.)

◊ *Indication 19.13 :* On montrera, par une comparaison série-intégrale, que, pour tout $\epsilon > 0$, la série $\sum \mathbf{P}\{|Y_1| > n\epsilon\}$ converge.

◊ *Indication 19.14 :* On vérifiera que $\{S = +\infty\} \subset \{E_k > 2^{k/2} \text{ ; i.s.}\}$ et on emploiera le lemme de Borel-Cantelli.

◊ *Indication 19.15 :* Prouver, avec Borel-Cantelli, que $\mathbf{P}\{X_n = 1 \text{ ; i.s.}\} = 1$.

SOLUTIONS

♦ **Solution 19.1** Les variables M_n sont à valeurs dans $[0\,;1]$ et, pour tout $x \in [0\,;1]$,

$$\mathbf{P}\{M_n \leqslant x\} = \mathbf{P}\{X_1 \leqslant x, \dots X_n \leqslant x\} = x^n$$

par indépendance. Les variables Y_n sont à valeurs dans $[0\,;n]$ et, pour tout $t \in [0\,;n]$,

$$F_{Y_n}(t) = \mathbf{P}\{1 - M_n \leqslant t/n\} = \mathbf{P}\{M_n \geqslant 1 - t/n\} = 1 - \left(1 - \frac{t}{n}\right)^n.$$

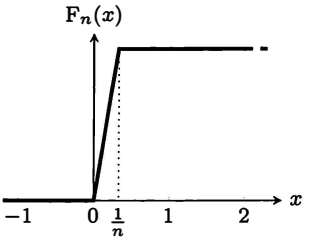
Ainsi, si l'on fixe un réel t positif, alors à partir d'un certain rang (égal à $\lfloor 1/t \rfloor + 1$), on aura $0 \leqslant t \leqslant n$ et la valeur $F_{Y_n}(t)$ sera égale à la quantité précédente, donc

$$\lim_{n\to\infty} F_{Y_n}(t) = 1 - e^{-t}.$$

La suite $(Y_n)_{n\geqslant 1}$ converge donc en loi vers la loi exponentielle $\mathscr{E}(1)$.

♦ **Solution 19.2** Pour tout $\omega \in \Omega$, posons $X_n(\omega) = -1/n$ pour tout $n \in \mathbb{N}$ et $X(\omega) = 0$; alors X_n converge simplement vers X, donc $X_n \to X$ p.s. *a fortiori*. Si l'on pose $\varphi : x \mapsto \lfloor x \rfloor$, alors $\varphi(X_n)$ est la variable aléatoire constante égale à -1 pour tout $n \geqslant 1$ tandis que $\varphi(X)$ est la variable aléatoire égale à 0.

♦ **Solution 19.3** Pour tout $n \in \mathbb{N}^*$, la fonction de répartition F_n de X_n vaut 0 sur $\mathbb{R}^-$, 1 sur $[1/n\,; +\infty[$, tandis que $F_n(x) = nx$ pour tout $x \in [0\,;n]$:

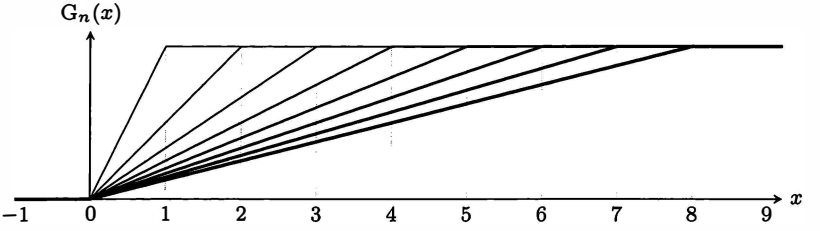


La suite $(F_n)_{n\geqslant 1}$ converge simplement vers la fonction de répartition de la variable nulle,

$$F : x \longmapsto \begin{cases} 1 & \text{si } x \geqslant 0 \\ 0 & \text{si } x < 0, \end{cases}$$

la convergence ayant lieu en tout point x différent de 0, seul point de discontinuité de F. Ainsi, $(X_n)_{n\geqslant 0}$ converge en loi vers la (loi de la) variable nulle.

Pour tout $n \geqslant 1$, la fonction de répartition G_n de Y_n vaut 0 sur $\mathbb{R}^-$ et 1 sur $[n\,;+\infty[$, et $G_n(x) = x/n$ pour tout $x \in [0\,;n]$:



Ainsi, la suite $(G_n)_{n\geqslant 1}$ converge simplement vers la fonction nulle, qui n'est *pas* une fonction de répartition. Conclusion : la suite $(Y_n)_{n\geqslant 1}$ ne converge pas en loi (ni donc en probabilité ni presque sûrement)[14].

♦ **Solution 19.4** La convergence dans L^2 implique la convergence dans L^1 (théorème 19.22), ce qui montre que $\mathbf{E}\,|X_n - X| \to 0$ et notamment que $\mathbf{E}X_n \to \mathbf{E}X$.

Par ailleurs, $X \mapsto \sqrt{\mathbf{E}\,|X|^2}$ est une norme, donc

$$\left|\sqrt{\mathbf{E}\,|X_n|^2} - \sqrt{\mathbf{E}\,|X|^2}\right| = \big|\,\|X_n\| - \|X\|\,\big| \leqslant \|X_n - X\| = \sqrt{\mathbf{E}\,|X_n - X|^2} \to 0,$$

ce qui montre que $\mathbf{E}\,|X_n|^2 \to \mathbf{E}\,|X|^2$. Par suite,

$$\mathbf{V}(X_n) = \mathbf{E}\,|X_n|^2 - (\mathbf{E}X_n)^2 \xrightarrow[n\to\infty]{} \mathbf{E}\,|X_n|^2 - (\mathbf{E}X)^2 = \mathbf{V}(X).$$

♦ **Solution 19.6**

i) Pour tous $n, k \geqslant 1$, l'inégalité de Markov permet d'écrire

$$k\,\mathbf{P}\{S_n \geqslant k\} \leqslant \mathbf{E}(S_n) = \sum_{k=1}^{n} \mathbf{E}\,|X_k| \leqslant \sum_{k=1}^{\infty} \mathbf{E}\,|X_k| =: \alpha.$$

ii) Notons $B_k = \bigcup_{n=1}^{\infty} \{S_n \geqslant k\}$: c'est une union croissante, ce qui permet d'écrire que

$$\mathbf{P}(B_k) = \lim_{n\to\infty} \mathbf{P}\{S_n \geqslant k\}$$

et mène à la majoration

$$\mathbf{P}(B_k) \leqslant \frac{\alpha}{k}.$$

Enfin, A étant égale à l'intersection des B_k, on a $\mathbf{P}(A) \leqslant \mathbf{P}(B_k)$ pour tout $k \in \mathbb{N}$, et donc $\mathbf{P}(A) = 0$.

---

(14). Bien sûr, on ne peut s'empêcher de ressentir une sorte de convergence ; de fait, on dit que la suite des lois de Y_n converge faiblement vers 0 (la mesure nulle).

iii) Notons C l'événement décrit par « la série $\sum |X_n|$ diverge ». En utilisant la croissance de la suite $(S_n)_{n\geqslant 1}$, on peut l'écrire sous la forme

$$C = \{\forall k \in \mathbb{N} \quad \exists n \in \mathbb{N} \quad S_n \geqslant k\} = \bigcap_{k=0}^{\infty} \bigcup_{n=1}^{\infty} \{S_n \geqslant k\} = A.$$

Ainsi, $\sum |X_n|$ converge presque sûrement, donc converge.

♦ **Solution 19.8** Les théorèmes généraux montrent que *i)* $\iff$ *ii)* $\implies$ *iii)* $\implies$ *iv)*. Il reste à montrer *iv)* $\implies$ *i)*.

Par contraposée, supposons que $(a_n)_{n\geqslant 1}$ ne converge pas vers a ; alors il existe $\varepsilon > 0$ tel que $|a_n - a| > \varepsilon$ une infinité de fois. L'une des deux propriétés suivantes est donc vraie : $a_n > a + \varepsilon$ une infinité de fois, ou $a_n < a - \varepsilon$ une infinité de fois. Dans le premier cas, on a $F(a+\varepsilon) = 1$ mais $F_n(a+\varepsilon) = 0$ une infinité de fois, ce qui montre que l'on n'a pas la convergence en loi puisque $a + \varepsilon$ est un point de continuité de F. Dans le second cas, on a $F(a - \varepsilon) = 0$ mais $F_n(a - \varepsilon) = 1$ une infinité de fois et on conclut de même.

♦ **Solution 19.9** Chaque variable aléatoire S_n est à valeurs dans $\mathbb{R}^+$. Fixons un réel $x \geqslant 0$ et un entier n. En décomposant l'événement $\{S_n \geqslant x\}$ selon le système complet des $\{N_n = \ell\}$, $\ell \geqslant 0$, on obtient

$$\begin{aligned}
\{S_n \geqslant x\} &= \sum_{\ell=1}^{\infty} \mathbf{P}\{S_n \geqslant x, N_n = \ell\} \\
&= \sum_{\ell=1}^{\infty} \mathbf{P}\{M_{n,1} \geqslant x, M_{n,2} \geqslant x, \dots, M_{n,\ell} \geqslant x, N_n = \ell\} \\
&= \sum_{\ell=1}^{\infty} \mathbf{P}\{M_{n,1} \geqslant x\} \cdot \mathbf{P}\{M_{n,2} \geqslant x\} \cdots \mathbf{P}\{M_{n,\ell} \geqslant x\} \cdot \mathbf{P}\{N_n - 1 = \ell - 1\} \\
&= \sum_{\ell=1}^{\infty} e^{-\ell\alpha x/n} e^{-n\lambda} \frac{(n\lambda)^{\ell-1}}{(\ell-1)!} = e^{-\alpha x/n} e^{-n\lambda} \cdot \exp(n\lambda e^{-\alpha x/n}).
\end{aligned}$$

On développe ensuite $n\lambda e^{-\alpha x/n} = n\lambda\big(1 - \alpha x/n + o(1/n)\big) = n\lambda - \alpha\lambda x + o(1)$, ce qui donne

$$e^{-\alpha x/n} e^{-n\lambda} \cdot \exp(n\lambda e^{-\alpha x/n}) = \exp\left\{-n\lambda - \frac{\alpha x}{n} + n\lambda - \alpha\lambda x + o(1)\right\} \xrightarrow[n\to\infty]{} e^{-\alpha\lambda x}.$$

On en déduit que $(S_n)_{n\geqslant 1}$ converge en loi vers la loi exponentielle $\mathscr{E}(\alpha\lambda)$.

♦ **Solution 19.10** On utilise le corollaire 19.14 pour montrer que $Y_n := |X_n - X| \xrightarrow{P} 0$. Puisque $0 \leqslant Y_n \wedge 1 \leqslant 1$ et que la constante 1 est intégrable, on peut appliquer le théorème de convergence dominée :

$$\lim_{n\to\infty} \mathbf{E}(Y_n \wedge 1) = \mathbf{E}\Big(\lim_{n\to\infty} Y_n \wedge 1\Big) = \mathbf{E}(0) = 0,$$

ce qui prouve que $Y_n \xrightarrow{P} 0$ et donc $X_n \xrightarrow{P} X$.

♦ **Solution 19.12** On suppose $X_n \xRightarrow{L} X$ et on note F_n $(n \geqslant 1)$ et F les fonctions de répartition de X_n et X. Le théorème de représentation de Skorokhod permet de construire $(Y_n)_{n\geqslant 1}$ et Y telles que $Y_n \xrightarrow{ps} Y$ et $F_{Y_n} = F_n$ pour tout entier n. Par continuité de φ, si $Y_n(\omega) \to Y(\omega)$, alors $\varphi\big(Y_n(\omega)\big) \to \varphi\big(Y(\omega)\big)$:

$$\{Y_n \to Y\} \subset \{\varphi(Y_n) \to \varphi(Y)\}.$$

Puisque l'ensemble de gauche est de probabilité 1, il en est de même de celui de droite, et donc $\varphi(Y_n) \xrightarrow{ps} \varphi(Y)$. On en déduit que $\varphi(Y_n) \xRightarrow{L} \varphi(Y)$. Puisque $\varphi(Y_n)$ a même fonction de répartition que $\varphi(X_n)$, et de même pour $\varphi(Y)$ et $\varphi(X)$, le théorème est prouvé.

♦ **Solution 19.13** Notons F la fonction de répartition de $|Y_1|$. Puisque $|Y_1|$ est intégrable, la fonction $1 - F$ est intégrable sur $\mathbb{R}^+$ (voir § 237, page 336). La monotonie de $1 - F$ autorise une classique comparaison série-intégrale, permettant de conclure que la série $\sum \mathbf{P}\{|Y_1| > n\varepsilon\}$ converge.

Il reste maintenant à exprimer la propriété $Y_n/n \to 0$. Pour tout $\omega \in \Omega$, on a l'équivalence

$$\frac{Y_n(\omega)}{n} \xrightarrow[n\to\infty]{} 0 \iff \forall \varepsilon > 0 \quad \exists N \in \mathbb{N} \quad \forall n \geqslant N \qquad |Y_n(\omega)| \leqslant \varepsilon.$$

Remarquer que l'on peut également se limiter aux ε *rationnels* et, par exemple, dans $]0\,;1]$. Dans le langage probabiliste, on a donc

$$\left\{\frac{Y_n}{n}\to 0\right\}=\bigcap_{\varepsilon\in\mathbb{Q}\cap]0\,;1]}\bigcup_{N=0}^{\infty}\bigcap_{n=N}^{\infty}\{|Y_n|\leqslant n\varepsilon\}$$

et, par passage au complémentaire,

$$\begin{aligned}\left\{\frac{Y_n}{n}\not\to 0\right\}&=\bigcup_{\varepsilon\in\mathbb{Q}\cap]0\,;1]}\bigcap_{N=0}^{\infty}\bigcup_{n=N}^{\infty}\{|Y_n|>n\varepsilon\}\\&=\bigcup_{\varepsilon\in\mathbb{Q}\cap]0\,;1]}\{|Y_n|>n\varepsilon\ ;\ \text{i.s.}\}=\bigcup_{\varepsilon\in\mathbb{Q}\cap]0\,;1]}\{|Y_1|>n\varepsilon\ ;\ \text{i.s.}\}.\end{aligned}$$

Or, d'après le lemme de Borel-Cantelli, $\mathbf{P}\{|Y_n|>n\varepsilon\ ;\ \text{i.s.}\}$ donc, par somme dénombrable, $\mathbf{P}\{Y_n/n\not\to 0\}=0$. En d'autres termes, la suite $(Y_n/n)_{n\geqslant 1}$ converge presque sûrement vers 0.

♦ **Solution 19.14** Les variables aléatoires E_k sont toutes positives. Ainsi, la suite des sommes partielles de la série $\sum 2^{-n}E_n$ est croissante, donc converge dans $\mathbb{R}\cup\{+\infty\}$. On note S sa somme, c'est une variable aléatoire à valeurs dans $\mathbb{R}\cup\{+\infty\}$.

S'il existe un rang à partir duquel $E_k\leqslant 2^{k/2}$, alors la série converge par comparaison avec la série géométrique de raison $1/\sqrt{2}$. Ainsi,

$$\bigcup_{N=0}^{\infty}\bigcap_{k=N}^{\infty}\left\{E_k\leqslant 2^{k/2}\right\}\subset\{S<+\infty\}.$$

Par passage au complémentaire,

$$\{S=+\infty\}\subset\bigcap_{N=0}^{\infty}\bigcup_{k=N}^{\infty}\left\{E_k>2^{k/2}\right\}=\left\{E_k>2^{k/2}\ ;\ \text{i.s.}\right\}.$$

Or les E_k suivent la loi exponentielle de paramètre λ, donc $\mathbf{P}\{E_k>2^{k/2}\}=e^{-2^{k/2}\lambda}$, ce qui prouve que $\sum\mathbf{P}\{E_k>2^{k/2}\}$ converge. Le lemme de Borel-Cantelli (sens trivial) prouve que $\mathbf{P}\{E_k>2^{k/2}\ ;\ \text{i.s.}\}=0$, et donc $\mathbf{P}\{S=+\infty\}=0$.

♦ **Solution 19.15** Si l'on note F la fonction de répartition de la variable nulle, $F(x)=1$ si $x\geqslant 0$ et 0 sinon. Notons F_n la fonction de répartition de X_n. Alors pour tout $x>0$, $F_n(x)=1$ dès que $1/n\leqslant x$; pour tout $x<0$, $F_n(x)=0=F(x)$. Ainsi, $(F_n)_{n\geqslant 0}$ converge vers F en tout point de continuité de F, donc $X_n\xrightarrow{\text{L}}0$.

Pour tout $\varepsilon>0$, $\mathbf{P}\{|X_n-0|>\varepsilon\}=1/n\to 0$. Ainsi, $X_n\xrightarrow{\text{P}}0$.

Enfin $\sum\mathbf{P}\{X_n=1\}=\sum 1/n$ diverge donc, par Borel-Cantelli, $\mathbf{P}\{X_n=1\ ;\ \text{i.s.}\}=1$, ce qui montre que presque sûrement $(X_n)_{n\geqslant 0}$ ne converge pas vers 0.

Résumé du chapitre

On définit trois grands types de convergence pour une suite de variables aléatoires :

- la convergence presque sûre, qui est la convergence simple sur un ensemble de probabilité 1 ;
- la convergence en probabilité — $\mathbf{P}\{|X_n-X|>\varepsilon\}\to 0$ pour tout $\varepsilon>0$ — légèrement plus faible que la précédente, et qui est une conséquence naturelle de l'inégalité de Bienaymé-Tchebychev dans de nombreuses situations ;
- la convergence en moyenne — $\mathbf{E}\,|X_n-X|\to 0$.

À ces trois types de convergence on ajoute la convergence en loi, qui ne concerne que la loi des variables : $F_n(x)\to F(x)$ en tout point x de continuité de F.

Chapitre 20

Loi des grands nombres

GUILDENSTERN (Flips a coin) : *The law of averages, if I have got this right, means that if six monkeys were thrown up in the air for long enough they would land on their tails about as often as they would land on their —*

ROSENCRANTZ : *Heads.* (He picks up the coin.)

Tom STOPPARD, *Rosencrantz & Guildenstern are dead*[1].

Ce chapitre présente deux grands résultats sur les suites de variables aléatoires.

- *Le premier est la loi du zéro-un (ou loi du tout-ou-rien) de Kolmogorov qui stipule qu'un événement, ne dépendant que de manière asymptotique d'une suite $(X_n)_{n\geqslant 0}$ quelconque de variables aléatoires indépendantes, est nécessairement de probabilité 0 ou 1.*
- *Le second est la très fameuse loi des grands nombres, déclinée dans plusieurs versions, et dont l'importance n'est plus à souligner. C'est le théorème qui permet notamment de relier la probabilité abstraite à la « fréquence limite du nombre de cas favorables sur le nombre total de cas » lorsque l'on répète une expérience un grand nombre de fois. Une démonstration relativement élémentaire de la version « forte » de ce théorème est donnée en annexe.*

20.1 LA LOI DU TOUT OU RIEN (OU DU 0–1) (∗)

Événements de queue, tribu asymptotique § 303

Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires ; on appelle **événement de queue** associé à cette suite tout événement qui peut être déterminé par l'ensemble des valeurs prises par cette suite, mais qui n'est pas modifié lorsque l'on change arbitrairement un nombre fini de ces valeurs. En d'autres termes, un événement de queue est un événement qui est entièrement déterminé par le comportement asymptotique de la suite des valeurs de $(X_n)_{n\geqslant 0}$.

(1). GUILDENSTERN *(Lance une pièce)* : D'après la loi des grands nombres, si j'ai bien compris, si on lance six singes un grand nombre de fois, ils retomberont sur leurs fesses à peu près aussi souvent que sur leur —

ROSENCRANTZ : Face. *(Il ramasse la pièce.)*

Exemple 20.1 Notons $A = \{(X_n)_{n\geqslant 0} \text{ converge}\}$, c'est-à-dire

$$A = \Big\{\omega \in \Omega\ ;\ (X_n(\omega))_{n\in\mathbb{N}} \text{ converge}\Big\}.$$

Alors A est un événement asymptotique : pour une réalisation du hasard ω donnée, savoir si A est réalisé ou non, c'est-à-dire savoir si $\omega \in A$ ou non, ne dépend que du comportement asymptotique de la suite $(X_n(\omega))_{n\in\mathbb{N}}$.

Plus formellement, l'ensemble des événements de queue forme une tribu, que l'on construit ainsi. Pour tout entier n, on note $\sigma(X_n)$ la tribu engendrée par X_n, et $\mathfrak{X}_n$ celle engendrée par $X_0, X_1, \dots, X_n$:

$$\mathfrak{X}_n = \sigma(X_0, X_1, \dots, X_n) = \sigma\Big(\sigma(X_0) \cup \sigma(X_1) \cup \cdots \cup \sigma(X_n)\Big).$$

Enfin, on définit

$$\mathfrak{Q}_n = \sigma(X_n, X_{n+1}, X_{n+2}, \dots) := \sigma\Big(\bigcup_{k\geqslant n} \sigma(X_k)\Big),$$

tribu engendrée par les variables aléatoires X_k $(k \geqslant n)$. La tribu $\mathfrak{Q}_n$ est la tribu[(2)] des événements ne dépendant que des valeurs de la suite $(X_k)_{k\geqslant n}$. La **tribu asymptotique**, ou **tribu des événements de queue**, est l'intersection de toutes ces tribus :

$$\mathfrak{Q}_\infty = \bigcap_{n=0}^{\infty} \sigma(X_n, X_{n+1}, X_{n+2}, \dots) = \bigcap_{n=0}^{\infty} \mathfrak{Q}_n.$$

Ses éléments sont les **événements de queue**, encore appelés **événements asymptotiques**.

Exemple 20.2 Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires ; l'ensemble $A := \{\sum X_n \text{ converge}\}$ est un événement asymptotique. En effet, on peut écrire, pour tout $n \in \mathbb{N}$:

$$A = \Big\{ \sum_{k\geqslant n} X_k \text{ converge}\Big\} \in \mathfrak{Q}_n$$

ce qui montre que $A \in \bigcap \mathfrak{Q}_n = \mathfrak{Q}_\infty$. De même, les événements suivants sont des événements de queue :

$$\Big\{\limsup_n X_n = +\infty\Big\} \qquad \Big\{\limsup_n \frac{X_1 + \cdots + X_n}{n} < +\infty\Big\}$$

Enfin, si

$$Q := \Big\{\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{} 0\Big\}$$

est bien un événement de queue, en revanche l'événement $\{X_1 + \cdots + X_n \to 0\}$ n'est pas un événement de queue : changer la valeur de X_1, par exemple, modifie la limite de $X_1 + X_2 + \cdots + X_n$. Dans l'événement Q, au contraire, modifier la valeur de X_1, ou d'un nombre arbitraire mais fini de X_k, ne modifie pas la convergence de la suite des moyennes.

---

(2). On remarquera que $\mathfrak{G}_n := \bigcup_{k\geqslant n} \sigma(X_k)$ est un π-système, et que c'est même une algèbre. En effet, si A et B sont des éléments de $\mathfrak{G}_n$, alors il existe des entiers p et $q \geqslant n$ tels que $A \in \sigma(X_p)$ et $B \in \sigma(X_q)$, donc

$$A \cup B \in \mathfrak{G}_{\min(p,q)} \subset \mathfrak{G}_n.$$

En revanche, ce n'est pas *a priori* une tribu, et on définit donc $\mathfrak{Q}_n$ comme la tribu engendrée par $\mathfrak{G}_n$; elle est beaucoup plus grosse que $\mathfrak{G}_n$. Cette tribu $\mathfrak{Q}_n$ est la plus petite tribu rendant mesurables toutes les applications $X_n, X_{n+1}, \dots$. Ses éléments sont les événements A dont la réalisation ne dépend que des valeurs des X_k, $k \geqslant n$: pour chaque réalisation ω, on peut savoir (de manière théorique !) si $\omega \in A$ par la simple inspection des valeurs $X_n(\omega), X_{n+1}(\omega), \dots$; la structure d'une tribu engendrée est cependant si complexe qu'on ne peut généralement pas donner de moyen *effectif* de décider si $\omega \in A$ ou non.

Certaines variables aléatoires d'intérêt sont $\mathfrak{Q}_\infty$-mesurables, par exemple la variable Y, à valeurs dans $\mathbb{R} \cup \{\pm\infty\}$ et définie par

$$Y = \limsup_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Loi du 0–1 de Kolmogorov § 304

La **loi du tout ou rien** de Kolmogorov, encore appelée **loi du zéro-un**, est un théorème remarquable, qui nous permet de construire des *certitudes*, et même de montrer que ces certitudes sont, d'une certaine manière, des cas généraux, et non des cas isolés de la théorie des probabilités. Il concerne des suites $(X_n)_{n\geqslant 0}$ de variables indépendantes, dont on ne précise, par ailleurs, rien sur la loi (elles peuvent être de même loi ou non, admettre des espérances ou non, etc.)

■ **Théorème 20.3 (Loi du 0–1 de Kolmogorov)** *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes. Notons $\mathfrak{Q}_\infty$ sa tribu de queue. Alors tout événement de queue $A \in \mathfrak{Q}_\infty$ vérifie $\mathbf{P}(A) = 0$ ou $\mathbf{P}(A) = 1$.*[3]

DÉMONSTRATION : Nous allons montrer le résultat suivant (quelque peu étonnant à première vue) : la tribu $\mathfrak{Q}_\infty$ est indépendante d'elle-même, ce qui implique notamment que *tout élément de queue est indépendant de lui-même.*

Il sera fait grand usage du lemme 16.9, que l'on rappelle : *si $\mathfrak{G}_1$ et $\mathfrak{G}_2$ sont des π-systèmes indépendants, alors les tribus engendrées $\mathfrak{T}_1 = \sigma(\mathfrak{G}_1)$ et $\mathfrak{T}_2 = \sigma(\mathfrak{G}_2)$ sont indépendantes.*

— *1re étape :* les tribus $\mathfrak{X}_{n-1}$ et $\mathfrak{Q}_n$ sont indépendantes. C'est une généralisation du lemme des coalitions que l'on peut démontrer aisément : $\mathfrak{Q}_n$ est engendré par les π-systèmes $\bigcup_k \mathfrak{X}_n^k = \bigcup_k \sigma(X_n, X_{n+1}, \ldots, X_{n+k})$ qui sont tous indépendants de $\mathfrak{X}_{n-1}$, et le lemme 16.9 permet de conclure.

— *2e étape :* puisque $\mathfrak{Q}_\infty \subset \mathfrak{Q}_n$, on en déduit que $\mathfrak{X}_{n-1}$ et $\mathfrak{Q}_\infty$ sont indépendantes.

— *3e étape :* notons $\mathfrak{A}$ la tribu engendrée par la classe $\mathfrak{G} = \bigcup_{n\geqslant 1} \mathfrak{X}_{n-1}$. La classe $\mathfrak{G}$ est un π-système (vérification immédiate car c'est une union croissante), indépendant de $\mathfrak{Q}_\infty$, le lemme 16.9 permet encore de conclure que $\mathfrak{A}$ est indépendant de $\mathfrak{Q}_\infty$.

— *Pour conclure,* $\mathfrak{A}$ est, par définition, $\mathfrak{Q}_0$, et la tribu de queue $\mathfrak{Q}_\infty$ est incluse dans $\mathfrak{A}$, donc $\mathfrak{Q}_\infty$ est indépendante d'elle-même !

En particulier, si $A \in \mathfrak{Q}_\infty$, alors A est indépendant de lui-même, c'est-à-dire que $\mathbf{P}(A) = \mathbf{P}(A \cap A) = \mathbf{P}(A) \cdot \mathbf{P}(A) = \mathbf{P}(A)^2$, ou encore $\mathbf{P}(A) \in \{0, 1\}$. ∎

Remarque 20.4 On peut voir ce théorème comme une généralisation des lemmes de Borel-Cantelli. Mais, alors que Borel-Cantelli nous disait franchement si l'événement $\{A_n \text{ ; i.s.}\}$ était de probabilité 0 ou 1, la loi du tout ou rien de Kolmogorov nous fait une réponse de Normand : c'est peut-être 1, c'est peut-être 0 !

Cela peut laisser une désagréable impression d'inachevé ; il est alors nécessaire de lever l'indétermination, par exemple en prouvant une inégalité stricte $\mathbf{P}(A) > 0$. Nous verrons une application spectaculaire au paragraphe § 362.

■ **Théorème 20.5** *Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires indépendantes, de tribu de queue $\mathfrak{Q}_\infty$. Si Y est une variable aléatoire mesurable par rapport à $\mathfrak{Q}_\infty$, alors Y est presque sûrement constante : il existe un élément c de $\mathbb{R} \cup \{\pm\infty\}$ tel que $\mathbf{P}\{Y = c\} = 1$.*

DÉMONSTRATION : Pour tout $x \in \mathbb{R}$, on a $F(x) = \mathbf{P}\{Y \leqslant x\} \in \{0, 1\}$. Notons

$$c := \sup\{x \in \mathbb{R} \text{ ; } \mathbf{P}\{Y \leqslant x\} = 0\}.$$

Si c est égal à $-\infty$, cela signifie que Y est presque sûrement égale à $-\infty$. De même, si $c = +\infty$, alors Y est presque sûrement égale à $+\infty$.

(3). Autrement dit, A est $\varnothing$ ou Ω, à un ensemble négligeable près.

Supposons désormais que $-\infty < c < +\infty$. Pour tout $n \geqslant 1$, on a $\mathbf{P}\{Y \leqslant c - 1/n\} = 0$ et, par l'axiome 3′ :

$$\mathbf{P}\{Y < c\} = \mathbf{P}\left(\bigcup_{n=1}^{\infty} \{Y \leqslant c - 1/n\}\right) = \lim_{n\to\infty} \mathbf{P}\{Y \leqslant c - 1/n\} = 0.$$

De même, $\mathbf{P}\{Y \leqslant c + 1/n\} = 1$ et, par l'axiome 3″,

$$\mathbf{P}\{Y \leqslant c\} = \mathbf{P}\left(\bigcap_{n=1}^{\infty} \mathbf{P}\{Y \leqslant c + 1/n\}\right) = \lim_{n\to\infty} \mathbf{P}\{Y \leqslant c + 1/n\} = 1.$$

On en déduit que $\mathbf{P}\{Y = c\} = 1$. ∎

Exemple 20.6 (Rayon de convergence d'une série entière aléatoire (1/2))
On rappelle que le rayon d'une série entière $\sum a_n z^n$ est donné par le théorème de Cauchy : dans $[0\,;+\infty]$ on a la relation

$$\frac{1}{R} = \limsup |a_n|^{1/n}.$$

Soit $(A_n)_{n\geqslant 0}$ une suite de variables aléatoires indépendantes et de même loi, à valeurs réelles ou complexes. Le rayon de la série entière $\sum A_n z^n$ est maintenant une variable aléatoire :

$$\frac{1}{R(\omega)} = \limsup |A_n(\omega)|^{1/n}.$$

Cette variable aléatoire étant $\mathfrak{Q}_\infty$-mesurable, elle est presque sûrement constante.

Suite au paragraphe § 462 page 652.

§ 305 Loi du tout ou rien pour les événements

On définit de la même façon une tribu de queue pour des événements $A_0, A_1, A_2, \ldots$ de $\mathfrak{T}$:

$$\mathfrak{Q}_n = \sigma(A_n, A_{n+1}, A_{n+2}, \ldots) \qquad \mathfrak{Q}_\infty = \bigcap_{n=0} \mathfrak{Q}_n.$$

■ **Théorème 20.7** *Si $(A_n)_{n\geqslant 0}$ est une suite d'événements indépendants, alors tout événement A de la tribu de queue est de probabilité 0 ou 1.*

§ 306 Loi du tout ou rien de Hewitt-Savage (∗)

Il existe une intéressante extension des résultats précédents à une classe d'événements plus large que les tribus asymptotiques, due à Edwin Hewitt et Leonard Savage [47]. Soit $X = (X_n)_{n\geqslant 0}$ une suite de variables aléatoires ; on considère une **permutation finie** $\rho : \mathbb{N} \to \mathbb{N}$, c'est-à-dire une bijection telle que $\rho(n) = n$ pour tous les entiers sauf, peut-être, un nombre fini d'entre eux. On note $\rho(X) = (X_{\rho(n)})_{n\in\mathbb{N}} = (X_{\rho(1)}, X_{\rho(2)}, \ldots, X_{\rho(n)}, \ldots)$. Si B est un événement de $\mathbb{R}^{\mathbb{N}}$, on dit que l'événement $A = \{X \in B\}$ est **symétrique** si $A = \{\rho(X) \in B\}$ pour toute permutation finie ρ.

Exemple 20.8 (Un événement symétrique) Dans une succession infinie de tirages à *pile* ou *face*, effectués avec une pièce équilibrée ($p = 1/2$), notons S_n la somme du nombre de *pile* au cours des n premiers tirages. L'événement $A = \{S_n = 0$; i.s.$\}$ n'est pas un événement asymptotique (il dépend évidemment de tous les tirages : changez le résultat du premier tirage, tous les S_n sont modifiés de 2, et A_n est effectivement modifié). Cependant, c'est un événement symétrique. En effet, pour toute permutation finie ρ, il existe un entier N à partir duquel tous les entiers vérifient $\rho(n) = n$; si l'on note S'_n les sommes partielles définies par la suite $\rho(X)$, on a donc $S_n = S'_n$ pour tout $n \geqslant N$, et donc $A = \{S'_n = 0$; i.s.$\}$.

■ **Proposition 20.9** *Soit $X = (X_n)_{n\geqslant 0}$ une suite de variables aléatoires. Tout événement de la tribu de queue est symétrique.*

Ainsi, la notion d'événement symétrique généralise celle d'événement de queue. Le lecteur trouvera chez Billingsley [9] ou Shiryaev [91] une démonstration du théorème qui suit.

■ **Théorème 20.10 (Loi du tout ou rien de Hewitt-Savage)** *Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires indépendantes et identiquement distribuées. Soit B un événement de $\mathbb{R}^{\mathbb{N}}$; on suppose que l'événement $A = \{X \in B\}$ est symétrique. Alors $\mathbf{P}(A) = 0$ ou $\mathbf{P}(A) = 1$.*

Une application aux séries aléatoires § 307

Considérons l'espace $\Omega = \{-1,+1\}^{\mathbb{N}^*}$ des suites $\varepsilon = (\varepsilon_n)_{n\geqslant 1}$ à valeurs dans $\{-1,+1\}$. Pour chaque suite ε, la série $\sum \varepsilon_n/n$ est soit convergente, soit divergente. Par exemple, à la suite constante $(1)_{n\in\mathbb{N}^*}$ est associée la série harmonique $\sum 1/n$, divergente ; à la suite $\big((-1)^n\big)_{n\in\mathbb{N}^*}$ est associée la série alternée $\sum(-1)^n/n$, convergente.

Que peut-on dire de l'ensemble C des suites $(\varepsilon_n)_{n\geqslant 1}$ telles que la série $\sum \varepsilon_n/n$ converge ? S'il est aisé de donner des critères suffisants d'appartenance à C [4], une étude exhaustive est plus difficile. Du point de vue probabiliste, on peut se poser une question différente : quelle est la « taille » de l'ensemble C, par rapport à la taille de Ω ?

Munissons donc Ω d'une probabilité quelconque pour laquelle les variables aléatoires $X_n : \varepsilon \mapsto \varepsilon_n$ sont indépendantes[5]. Alors, la loi du tout ou rien de Kolmogorov nous affirme que

$$\mathbf{P}(C) = 0 \quad \text{ou} \quad \mathbf{P}(C) = 1.$$

En effet, l'événement C fait partie de la tribu asymptotique de la suite $(X_n)_{n\geqslant 0}$ de variables aléatoires à valeurs dans $\{-1,+1\}$.

Bien sûr, nous n'avons pas vraiment répondu à la question « l'ensemble C est-il "gros" ou "petit" ? » ; nous savons seulement que C ne peut pas être de taille intermédiaire. Mais ce résultat, déjà remarquable en soi, a été obtenu très facilement comme conséquence de la loi du tout ou rien ; de plus, il est valable quelle que soit la probabilité dont on munit Ω.

Le lecteur est invité à faire fonctionner son intuition et à deviner, dans le cas où Ω est muni de la probabilité correspondant à une succession de tirages à *pile* ou *face* avec une pièce équilibrée, si l'on a $\mathbf{P}(C) = 0$, ou si au contraire $\mathbf{P}(C) = 1$; nous reviendrons ultérieurement sur ce problème (section 33.2).

Une application aux marches aléatoires § 308

Voici une application relativement simple de la loi du tout ou rien de Kolmogorov, qui montre que, dans un jeu de *pile* ou *face* équilibré, le gain S_n d'un joueur, cumulé après n parties, va (presque sûrement) prendre des valeurs alternativement positives et négatives, et de plus en plus grandes.

---

(4). Par exemple, si ε est telle que tous les paquets $A_{p,q} = \varepsilon_p + \varepsilon_{p+1} + \cdots + \varepsilon_q$ sont de taille uniformément bornée par un réel M, alors la série $\sum \varepsilon_n/n$ converge.

(5). On a vu qu'une telle probabilité existe ; le moyen le plus naturel d'en exhiber une est de considérer tout d'abord une algèbre, celle des parties de Ω de la forme

$$A^{i_1,\ldots,i_n}_{a_1,\ldots,a_n} = \{\varepsilon \in \Omega \,;\, \varepsilon_{i_1} = a_1, \ldots, \varepsilon_{i_n} = a_n\}$$

(*cylindres* de Ω). On définit alors une probabilité sur cette algèbre par la formule :

$$\mathbf{P}\big(A^{i_1,\ldots,i_n}_{a_1,\ldots,a_n}\big) = p^k\,(1-p)^{n-k},$$

où k est le nombre d'éléments de $a_1, \ldots, a_n$ égaux à $+1$. L'interprétation de cette probabilité est la suivante : on identifie le choix d'une suite ε à une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, vérifiant $\mathbf{P}\{X_n = +1\} = p$ et $\mathbf{P}\{X_n = -1\} = 1-p$.

Dans le cas $p = 1/2$, si l'on envoie Ω sur l'intervalle $[0\,;\,1]$ en posant $x(\varepsilon) = \sum_{n=1}^{\infty} \dfrac{(1+\varepsilon_n)}{2^{n+1}}$, la probabilité précédente a pour image la mesure de Lebesgue. (L'application $\varepsilon \mapsto x(\varepsilon)$ n'est pas tout à fait une bijection, mais la restriction de cette application à un ensemble $\Omega' = \Omega \setminus N$, où N est négligeable, établit une bijection de Ω' sur $[0\,;\,1[$.)

■ **Théorème 20.11** *Soit $(X_n)_{n\geqslant 0}$ une suite de Rademacher, c'est-à-dire une suite de variables aléatoires indépendantes, de même loi*

$$\mathbf{P}\{X_n = 1\} = \mathbf{P}\{X_n = -1\} = \frac{1}{2}.$$

Si l'on note $S_n = X_1 + X_2 + \cdots + X_n$, alors

$$\mathbf{P}\Big\{ \sup_n S_n = +\infty \quad \textit{et} \quad \inf_n S_n = -\infty \Big\} = 1.$$

Le point essentiel de la démonstration de ce théorème est de remarquer que les événements $\{\sup S_n = +\infty\}$ et $\{\inf S_n = -\infty\}$ sont dans la tribu de queue. En effet, pour tout k entier, on a

$$\begin{aligned}\{\sup S_n = +\infty\} &= \Big\{ \sup_{n\geqslant k}(S_n - S_k) = +\infty \Big\} \\ &= \Big\{ \sup_{n\geqslant k}(X_{k+1} + X_{k+2} + \cdots + X_n) = +\infty \Big\} \in \mathfrak{Q}_k,\end{aligned}$$

ce qui montre que $\{\sup S_n = +\infty\}$ est dans l'intersection $\mathfrak{Q}_\infty$ de tous les $\mathfrak{Q}_k$: c'est donc un événement de queue. D'après la loi du tout ou rien, cet événement ne peut donc être que de probabilité 0 ou 1. De même pour l'événement $\{\inf S_n = -\infty\}$. D'ailleurs, un simple argument de symétrie montre que

$$\mathbf{P}\{\sup S_n = +\infty\} = \mathbf{P}\{\inf S_n = -\infty\}.$$

Ce que nous allons faire, pour prouver le théorème précédent, est de montrer que

$$\mathbf{P}\{\sup S_n = +\infty\} + \mathbf{P}\{\inf S_n = -\infty\} > 0, \qquad (*)$$

ce qui suffit pour conclure.

On commence par montrer quelques résultats intermédiaires. Le premier affirme que, presque sûrement, une suite de 1 (de taille $2p+1$ fixée à l'avance) apparaîtra, une infinité de fois, dans la succession des tirages[6].

■ **Lemme 20.12** *Soit p un entier, $p \geqslant 1$. Alors, avec une probabilité 1,*

$$\mathrm{Card}\{k \in \mathbb{N} \,;\, X_{k+1} = X_{k+2} = \cdots = X_{k+2p+1} = 1\} = +\infty.$$

DÉMONSTRATION : Ce résultat s'appuie sur le lemme de Borel-Cantelli : si $\sum \mathbf{P}(A_n)$ diverge et si les événements A_n sont indépendants, alors $\{A_n \,;\, \text{i.s.}\}$ avec la probabilité 1.

Notons A_k l'événement $A_k = \{X_{k+1} = \cdots = X_{k+2p+1} = 1\}$, qui est de probabilité constante égale à $1/2^{2p+1}$. Malheureusement, les A_k ne sont pas indépendants : A_k et A_{k+1} font tous deux intervenir X_{k+2} par exemple. En revanche, si l'on note

$$A'_k := A_{(2p+1)k},$$

alors les A'_k sont deux à deux indépendants (ils ne font intervenir que des variables aléatoires X_n avec des jeux d'indices disjoints, et le lemme des coalitions s'applique).

Enfin, chaque événement A'_k étant de probabilité $1/2^{2p+1}$, la série $\sum \mathbf{P}(A'_k)$ diverge grossièrement ; on applique le lemme de Borel-Cantelli pour obtenir $\mathbf{P}\{A'_k \,;\, \text{i.s.}\} = 1$, ce qui prouve que

$$\mathbf{P}\{A_k \,;\, \text{i.s.}\} = 1 \qquad (**)$$

également, et donc le résultat annoncé. ∎

(6). Ce résultat et sa démonstration ont déjà été vus sous la forme anecdotique des singes tapant au hasard à la machine (exemple 5.16 page 117).

Dans un deuxième temps, on montre que, presque sûrement, la suite $(S_n)_{n\geqslant 1}$ n'est pas confinée à un domaine borné donné :

■ **Lemme 20.13** *Soit p un entier, $p \geqslant 1$. Alors*

$$\mathbf{P}\Big\{ -p \leqslant \inf_n S_n \leqslant \sup_n S_n \leqslant p \Big\} = 0.$$

DÉMONSTRATION : Notons L_p l'événement $L_p = \{-p \leqslant \inf S_n \leqslant \sup S_n \leqslant p\}$. Alors, chacun des événements A_k définis dans le lemme précédent est dans le complémentaire L_p^c de L_p. En effet, si $\omega \in A_k$, alors l'écart $S_{k+2p+1} - S_k$ est égal à $2p+1$, ce qui montre que l'une des deux valeurs S_k et S_{k+2p+1} est en dehors de $[-p\,;+p]$.

On a l'inclusion

$$\{A_k \text{ ; i.s.}\} = \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k \subset \bigcup_{k=0}^{\infty} A_k,$$

(qui se traduit par : « si une infinité de A_k est réalisée, alors il y en a au moins un qui est réalisé ! »). Comme le membre de gauche est de probabilité 1 (équation (**) dans la démonstration précédente), le membre de droite est également de probabilité 1. Enfin, puisque chaque A_k est inclus dans L_p^c, leur union l'est également, donc L_p^c est de probabilité 1, et $\mathbf{P}(L_p) = 0$. ∎

DÉMONSTRATION DU THÉORÈME 20.11 : On remarque d'abord que

$$L_\infty := \Big\{ -\infty < \inf_n S_n \leqslant \sup_n S_n < +\infty \Big\} = \bigcup_{p=0}^{\infty} \Big\{ -p \leqslant \inf_n S_n \leqslant \sup_n S_n \leqslant p \Big\} = \bigcup_{p=0}^{\infty} L_p.$$

Puisque la suite $(L_p)_{p\geqslant 0}$ est croissante, on a $L_\infty = \lim_{p\to\infty} \uparrow L_p$ et donc (axiome 3′)

$$\mathbf{P}(L_\infty) = \mathbf{P}\big(\lim_p \uparrow L_p \big) = \lim_{p\to\infty} \mathbf{P}(L_p) = 0.$$

Par passage au complémentaire, on obtient

$$\mathbf{P}\{ \inf S_n = -\infty \text{ ou } \sup S_n = +\infty \} = 1,$$

ce qui mène directement à l'inégalité (*) et achève la preuve. ∎

■ **Corollaire 20.14** *Une marche aléatoire équilibrée sur $\mathbb{Z}$ repasse, presque sûrement, une infinité de fois par 0.*

20.2 LOIS DES GRANDS NOMBRES

Les divers résultats regroupés sous la dénomination de *loi des grands nombres* forment une succession de théorèmes, de plus en plus fins, qui permettent notamment de justifier que l'espérance d'une variable aléatoire est le résultat moyen de la mesure expérimentale résultant d'expériences identiques répétées un grand nombre de fois. Plus généralement, on étudie la convergence de la suite de terme général

$$Y_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

vers une certaine limite ; cette convergence peut être en probabilité (auquel cas on a une *loi faible des grands nombres*), presque sûrement (c'est alors une *loi forte*) ou encore dans l'espace L^2 ou L^p. Les hypothèses sur les X_k peuvent être assez restrictives (les variables ont toutes la même loi et admettent une variance), ou un peu moins (les variables ont toutes la même espérance), mais elles doivent toujours être *indépendantes*[7].

(7). Encore que, là aussi, on puisse affaiblir les hypothèses, comme dans la *loi forte échangeable*, voir l'annexe 20.4 et, plus spécifiquement, la remarque 20.32.

§ 309 La loi faible des grands nombres

Dans le premier de ces théorèmes, on exige que les variables aléatoires aient une espérance ainsi qu'un moment d'ordre 2, ce qui les oblige à ne pas trop se disperser autour de leur valeur moyenne. Cette version de la loi faible des grands nombres se démontre aisément grâce à l'inégalité de Bienaymé-Tchebychev, ce qui explique sa popularité.

■ **Théorème 20.15 (Loi faible)** *Soit* $(\Omega, \mathfrak{T}, \mathbf{P})$ *un espace probabilisé. Soit* $(X_n)_{n\geqslant 1}$ *une suite de variables aléatoires indépendantes, de même espérance* m *et de même variance* σ^2. *Alors la suite de terme général*

$$Y_n = \frac{X_1 + \cdots + X_n}{n}$$

converge en probabilité vers (la variable aléatoire constante égale à) m :

$$\forall \varepsilon > 0 \qquad \lim_{n\to\infty} \mathbf{P}\{\, |Y_n - m| \geqslant \varepsilon \,\} = 0.$$

DÉMONSTRATION : Soit $\varepsilon > 0$. Pour tout $n \in \mathbb{N}$, l'inégalité de Bienaymé-Tchebychev s'écrit

$$\mathbf{P}\{\, |Y_n - m| \geqslant \varepsilon \,\} \leqslant \frac{\sigma^2_{Y_n}}{\varepsilon^2}.$$

Or, puisque les variables $X_1, \ldots, X_n$ sont indépendantes, l'égalité de Bienaymé (théorème 16.24 page 350) permet d'écrire que $\sigma^2_{nY_n} = \sigma^2_{X_1+\cdots+X_n} = n\sigma^2$, et donc

$$\sigma^2_{Y_n} = \frac{\sigma^2}{n},$$

ce qui donne

$$\mathbf{P}\{\, |Y_n - m| \geqslant \varepsilon \,\} \leqslant \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0. \qquad \blacksquare$$

Remarque 20.16 Il existe d'autres théorèmes de type « loi faible », où l'on diminue les hypothèses faites sur la suite $(X_n)_{n\geqslant 0}$ (par exemple, les variances peuvent être différentes, du moment qu'elles ne fluctuent « pas trop »).

Notons une autre version de la loi faible, due à Alexandre KHINTCHINE [59], n'imposant pas que les variables aléatoires admettent une variance, mais requérant en revanche qu'elles suivent toutes la même loi. La conclusion est la même. Sa démonstration n'est pas aussi élémentaire que la précédente, mais offre un bel exemple des techniques basées sur l'usage des fonctions caractéristiques.

■ **Théorème 20.17 (Khintchine)** *Soit* $(X_n)_{n\geqslant 1}$ *une suite de variables aléatoires, indépendantes et de même loi, admettant une espérance* m. *Alors la suite de terme général*

$$Y_n = \frac{X_1 + \cdots + X_n}{n}$$

converge en probabilité vers m.

DÉMONSTRATION : Notons φ la fonction caractéristique des variables X_i : $\varphi(t) = \mathbf{E}(e^{itX})$. Par indépendance des X_k, la fonction caractéristique de $X_1 + \cdots + X_n$ est φ^n, et celle de Y_n est

$$\varphi_n(t) = \mathbf{E}(e^{itY_n}) = \left[\varphi\left(\frac{t}{n}\right)\right]^n.$$

Puisque $\mathbf{E}|X| < +\infty$, on a

$$\varphi(t) = 1 + it\,\mathbf{E}(X) + o(t) = 1 + imt + o(t).$$

Notamment, à t fixé, on a

$$\varphi\left(\frac{t}{n}\right) = 1 + \frac{imt}{n} + \underset{n\to\infty}{o}\left(\frac{1}{n}\right)$$

ce qui mène à

$$\varphi_n(t) = \left[1 + \frac{imt}{n} + o\left(\frac{1}{n}\right)\right]^n \xrightarrow[n\to\infty]{} e^{itm}$$

(voir A11, page 670). Puisque la fonction $t \mapsto e^{itm}$ est continue en 0, le théorème de continuité de Paul Lévy[8] montre que $(Y_n)_{n\geqslant 1}$ converge en loi vers m :

$$Y_n \xrightarrow{L} m.$$

Le théorème 19.46 page 401 permet alors de conclure que la convergence a lieu en probabilité. ∎

La loi forte des grands nombres

§ 310

À la plupart des énoncés de type « loi faible » correspondent des énoncés de type « loi forte », où la convergence en probabilité est remplacée par la conclusion plus forte de convergence presque sûre.

■ **Théorème 20.18 (Loi forte des grands nombres de Kolmogorov)** *Soit* X *une variable aléatoire. Soit* $(X_n)_{n\geqslant 1}$ *une suite de variables aléatoires, indépendantes et de même loi que* X. *Enfin, soit* m *un réel. Alors, la suite de terme général*

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

converge presque sûrement vers m *si et seulement si* X *est intégrable et* $\mathbf{E}(X) = m$.

La démonstration que Kolmogorov proposa de ce théorème est assez délicate, et n'est pas abordée ici. Cependant,

— on peut démontrer la loi forte en ajoutant des hypothèses sur la loi des X_k, par exemple en supposant qu'elles admettent une variance, ou un moment d'ordre 4, ce qui simplifie l'argumentation[9], et est suffisant dans la plupart des cas ;

— le mathématicien Luca PRATELLI a proposé, en 1989, une démonstration élémentaire de la loi forte [80], dans le cas où les variables aléatoires admettent une espérance (mais sans que l'on impose l'existence de moments d'ordre plus élevé) ; nous en donnons une version légèrement modifiée en annexe, à la fin de ce chapitre.

Remarque 20.19 (Illustration numérique) On peut illustrer la loi des grands nombres par une simulation numérique. Considérons deux variables aléatoires X et X′, pouvant prendre chacune les valeurs entières 1, 2, 3,... On suppose qu'elles suivent les lois définies respectivement par :

$$\forall k \geqslant 1 \qquad \mathbf{P}\{X = k\} = \frac{1}{k^2} - \frac{1}{(k+1)^2} \qquad \text{et} \qquad \mathbf{P}\{X' = k\} = \frac{1}{k^4} - \frac{1}{(k+1)^4}.$$

De la propriété

$$\mathbf{P}\{X = k\} = \mathbf{P}\{X \geqslant k\} - \mathbf{P}\{X \geqslant k+1\} = \frac{1}{k^2} - \frac{1}{(k+1)^2},$$

on déduit, par télescopage,

$$\mathbf{P}\{X \geqslant k\} = \sum_{i=k}^{\infty} \mathbf{P}\{X = i\} = \frac{1}{k^2}$$

et donc, en utilisant la formule alternative de l'espérance (exercice 7.5 page 166), on en déduit que X admet une espérance, valant

$$\mathbf{E}(X) = \sum_{k=0}^{\infty} \mathbf{P}\{X > k\} = \sum_{k=1}^{\infty} \mathbf{P}\{X \geqslant k\} = \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6} \approx 1,6449,$$

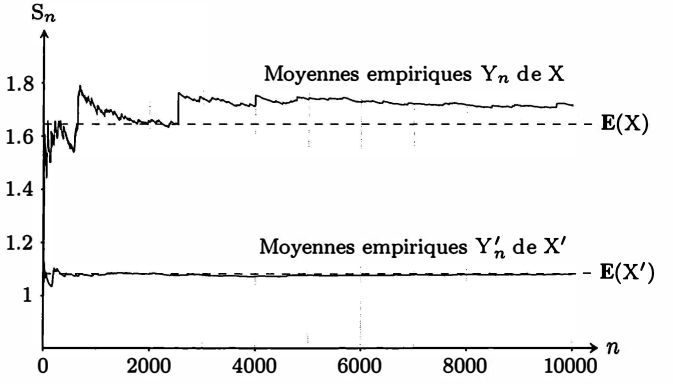
(8). Théorème 19.44 page 401.

(9). Voir exercice 20.2.

mais n'admet pas de variance. La variable X′ admet à la fois une espérance et une variance ; son espérance peut se calculer par la même méthode :

$$\mathbf{E}(X') = \sum_{k=1}^{\infty} \mathbf{P}\{X' \geqslant k\} = \sum_{k=1}^{\infty} \frac{1}{k^4} = \frac{\pi^4}{90} \approx 1,0823.$$

Une simulation informatique permet de représenter une réalisation de la suite des moyennes empiriques pour chacune des deux variables. Dans les deux cas, on observe une convergence assez nette vers la valeur numérique de l'espérance, comme on pouvait s'y attendre. Cependant, la convergence de la suite des moyennes empiriques de X′ est plus « rapide » que celle de X, grâce à l'existence du moment d'ordre 2 de X′. Pour comparaison, la moyenne observée après $n = 10\,000$ tirages pour chacune des variables est $Y_n(\omega) = 1,7144$ et $Y'_n(\omega) = 1,0799$.



Des simulations avec d'autres lois permettraient de mettre en évidence des vitesses de convergence différentes. Il apparaît que la loi des grands nombres a une utilité essentiellement « théorique » : elle ne permet pas de quantifier la convergence, ni de donner de valeur de n pour laquelle la moyenne Y_n observée donnerait une valeur acceptable de $\mathbf{E}(X)$. Le théorème limite central permet d'aller plus loin en donnant une approximation de la loi des fluctuations de Y_n autour de $\mathbf{E}(X)$.

Remarque 20.20 (Illustration lorsque les X_k ne sont pas intégrables) La loi de Cauchy ne possède pas de moment d'ordre 1 ; d'après la loi forte des grands nombres, si l'on se donne une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, de même loi de Cauchy $\mathscr{C}(0\,;\,1)$, alors la suite de terme général $Y_n = (X_1 + X_2 + \cdots + X_n)/n$ ne converge pas presque sûrement vers une constante. L'exercice 17.4 page 363 montre que Y_n suit également la loi de Cauchy $\mathscr{C}(0\,;\,1)$; notamment, la suite $(Y_n)_{n\geqslant 1}$ ne converge pas non plus en probabilité vers une constante, c'est-à-dire qu'elle ne suit pas la loi faible des grands nombres. Ce comportement est illustré sur la figure de la page 250.

Remarque 20.21 (Autres formes de la loi forte) On peut élargir le champ d'application du théorème de Kolmogorov en ne supposant plus la suite $(X_n)_{n\geqslant 1}$ identiquement distribuée, mais en la remplaçant par une condition plus faible qui empêche les variances $\mathbf{V}(X_n)$ de croître trop rapidement. Nous en donnons un énoncé sans démonstration ; le lecteur pourra en trouver, par exemple, chez Shiryaev [91, chapitre IV, § 3]. On note comme d'habitude $S_n = X_1 + X_2 + \cdots + X_n$.

■ **Théorème 20.22 (Kolmogorov)** *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes, admettant toutes un moment d'ordre 2. Soit $(b_n)_{n\geqslant 1}$ une suite croissante de réels telle que $b_n \to +\infty$ et telle que la série $\sum \mathbf{V}(X_n)/b_n^2$ converge. Alors la suite de terme général*

$$Y_n = \frac{(X_1 - \mathbf{E}X_1) + (X_2 - \mathbf{E}X_2) + \cdots + (X_n - \mathbf{E}X_n)}{b_n} = \frac{S_n - \mathbf{E}S_n}{b_n}$$

converge presque sûrement vers 0.

■ **Corollaire 20.23** *Si $\sum \mathbf{V}(X_n)/n^2$ converge, alors $(S_n - \mathbf{E}S_n)/n$ converge presque sûrement vers 0. C'est notamment le cas si les variances sont bornées par une constante commune.*

Remarque 20.24 (« Nombre de cas favorables... ») Soit B un événement de Ω. On effectue un grand nombre de fois une expérience, dont le résultat est déclaré un succès si B est réalisé, et un échec sinon ; on compte, après n réalisations de l'expérience, le nombre de cas favorables sur le nombre de cas total. Une telle situation est modélisée par des variables aléatoires $X_1, X_2, \dots, X_n$ indépendantes, de même loi que $\mathbf{1}_B$. Le nombre de cas favorables est alors $S_n = X_1 + X_2 + \cdots + X_n$, et la loi des grands nombres nous apprend que, presque sûrement,

$$\lim_{n\to\infty} \frac{S_n}{n} = \mathbf{E}(\mathbf{1}_B) = \mathbf{P}(B).$$

Ainsi, la modélisation d'une situation aléatoire typique est bien conforme au modèle statistique intuitif : avec une probabilité 1, le nombre de cas favorables sur le nombre de cas total va tendre vers une limite, et celle-ci a pour valeur le nombre $\mathbf{P}(B)$. Ce résultat, à lui seul, justifie la cohérence de l'approche mathématique sous-jacente à toute la théorie.

De même, si A est un événement tel que $\mathbf{P}(A) > 0$, si l'on considère une suite $(Y_n)_{n\geqslant 1}$ de variables aléatoires indépendantes de même loi que $\mathbf{1}_{A\cap B}$ et que l'on note $T_n = Y_1 + Y_2 + \cdots + Y_n$, alors presque sûrement

$$\lim_{n\to\infty} \frac{T_n}{S_n} = \frac{\mathbf{E}(\mathbf{1}_{A\cap B})}{\mathbf{E}(\mathbf{1}_B)} = \frac{\mathbf{P}(A\cap B)}{\mathbf{P}(A)} = \mathbf{P}(B \mid A).$$

Cette remarque est à la base de la **méthode de rejet** : pour calculer la probabilité de B conditionnellement à A, on effectue un grand nombre de tirages, parmi lesquels *on ne retient que ceux pour lesquels* A *est réalisé*, puis on calcule le nombre de cas favorables (B est également réalisé) sur le « nombre total » de cas.

Cas de la divergence (*) § 311

Que dire d'une suite de variables aléatoires dont l'espérance est infinie ? Tout d'abord, rappelons qu'une variable aléatoire réelle $X = X^+ - X^-$ peut être non intégrable pour deux raisons très distinctes :

— si X^+ et X^- sont simultanément non intégrables, c'est-à-dire si $\mathbf{E}(X^+) = \mathbf{E}(X^-) = +\infty$, alors X n'a pas d'espérance ;

— si X^- est intégrable et $\mathbf{E}(X^+) = +\infty$, alors X n'est pas intégrable mais on définit malgré tout son espérance comme

$$\mathbf{E}(X) := \mathbf{E}(X^+) - \mathbf{E}(X^-) = +\infty.$$

De même, si $\mathbf{E}(X^-) = +\infty$ et $\mathbf{E}(X^+) < +\infty$, alors on pose $\mathbf{E}(X) := -\infty$.

Supposons que X admette une espérance égale à $+\infty$. Intuitivement, si l'on fait une suite de tirages, on devrait observer une moyenne qui tend vers $+\infty$. C'est bien ce que dit la version étendue de la loi des grands nombres.

■ **Théorème 20.25** *Soit* X *une variable aléatoire ; soit* $(X_n)_{n\geqslant 1}$ *une suite de variables aléatoires indépendantes de même loi que* X. *On suppose que* $\mathbf{E}(X^-) < +\infty$ *et* $\mathbf{E}(X^+) = +\infty$, *de sorte que* $\mathbf{E}(X) = +\infty$. *Alors la suite de terme général*

$$Y_n = \tfrac{1}{n}\,(X_1 + X_2 + \cdots + X_n)$$

diverge vers $+\infty$ *presque sûrement.*

DÉMONSTRATION : Posons $T_n := \frac{1}{n}\,(X_1^- + X_2^- + \cdots + X_n^-)$ pour tout entier n ; alors $T_n \to \mathbf{E}(X^-)$ p.s. par la loi des grands nombres.

De même, posons $Z_n := \frac{1}{n}\,(X_1^+ + X_2^+ + \cdots + X_n^+)$; il suffit de prouver que $Z_n \to +\infty$ presque sûrement pour pouvoir conclure. Pour cela, on se ramène d'abord à des variables intégrables en tronquant les X_k^+ : pour tout entier $\alpha \geqslant 1$, on pose

$$X_k^{[\alpha]} = X_k^+ \wedge \alpha = \begin{cases} X_k^+ & \text{si } 0 \leqslant X_k^+ \leqslant \alpha \\ \alpha & \text{si } X_k^+ > \alpha. \end{cases}$$

Alors $Z_n \geqslant (X_1^{[\alpha]} + X_2^{[\alpha]} + \cdots + X_n^{[\alpha]})/n$, le membre de droite convergeant presque sûrement vers $\mathbf{E}(X^{[\alpha]})$ puisque les $X_k^{[\alpha]}$ sont indépendantes et de même loi que $X^{[\alpha]} = X^+ \wedge \alpha$. Cela prouve qu'il existe des ensembles Ω_α (pour $\alpha = 1, 2, \dots$) de probabilité 1, tels que

$$\forall \omega \in \Omega_\alpha \qquad \liminf_n Z_n(\omega) \geqslant \mathbf{E}(X^{[\alpha]}).$$

Enfin, lorsque $\alpha \to \infty$, on a $X^{[\alpha]} \uparrow X^+$ donc, par théorème de convergence monotone, $\mathbf{E}(X^{[\alpha]}) \uparrow \mathbf{E}(X^+) = +\infty$. Sur l'ensemble $\Omega' = \bigcap_{\alpha\geqslant 1} \Omega_\alpha$, de probabilité égale à 1, on a donc $\liminf Z_n(\omega) = +\infty$, ou encore : $Z_n \to +\infty$ presque sûrement.

On en déduit que $Y_n = T_n + Z_n$ converge presque sûrement vers $+\infty$. ∎

Remarque 20.26 (Que se passe-t-il si $\mathbf{E}(X)$ *n'est pas définie ?)* Si $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes et de même loi que X, et si $\mathbf{E}(X^+) = \mathbf{E}(X^-) = +\infty$, trois scénarios mutuellement exclusifs sont possibles — et trois seulement — selon que la divergence de $\mathbf{E}(X^+)$ est « plus forte » ou « moins forte » que celle de $\mathbf{E}(X^-)$.

- Dans le premier scénario, $\mathbf{P}\{Y_n \to +\infty\} = 1$.
- Dans le deuxième scénario, $\mathbf{P}\{Y_n \to -\infty\} = 1$.
- Dans le troisième (qui est notamment celui associé à la loi de Cauchy), la suite effectue une infinité d'allers et retours arbitrairement grands, et

$$\mathbf{P}\{\limsup Y_n = +\infty \quad \text{et} \quad \liminf Y_n = -\infty\} = 1.$$

(Voir par exemple la référence [32].)

Il convient de remarquer que, même si $\lim Y_n = +\infty$ presque sûrement, il se peut que cette divergence vers l'infini soit d'une extrême lenteur (par exemple en $\ln n$) ; c'est ce qui arrive dans le fameux « paradoxe de Saint-Petersbourg » (§ 467).

20.3 THÉORÈME DE GLIVENKO-CANTELLI

§312 **Fonction de répartition empirique**

Dans de nombreux domaines d'application, il est nécessaire d'estimer la fonction de répartition F d'une variable aléatoire X ; le théorème de Glivenko-Cantelli permet de justifier l'approximation de F par une fonction dite *empirique*, construite d'après des données expérimentales. Pour modéliser une telle situation, on considère une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, de même loi que X, qui représentent le résultat d'une « même » expérience répétée un grand nombre de fois. Pour une réalisation ω du hasard (et une seule), on cherche à estimer la loi de X à partir de l'information donnée par $X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$.

On considère une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires. Pour chaque $\omega \in \Omega$, on calcule pour tout entier n la **fonction de répartition empirique**, qui est la fonction $F_n(x,\omega)$ qui compte la proportion de tirages X_k $(1 \leqslant k \leqslant n)$ inférieurs à x :

$$F_n(x,\omega) = \tfrac{1}{n} \operatorname{Card}\{k \in [\![1\,;\,n]\!]\ ;\ X_k(\omega) \leqslant x\}.$$

Elle est donc constante par morceaux, et fait un « saut » de hauteur $1/n$ à chacune des valeurs $X_1(\omega), \ldots, X_n(\omega)$. En termes plus synthétiques,

$$F_n(x,\cdot) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{]-\infty\,;\,x]}(X_k).$$

En pratique, lorsque l'on mène une expérience[10], on note $a_1, a_2, \ldots, a_n$ les valeurs prises par $X_1(\omega), \ldots, X_n(\omega)$, valeurs que l'on aura pris soin d'ordonner : $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$. La fonction de répartition empirique $F_n(\cdot,\omega)$ est alors une fonction en escalier, prenant la valeur 0 entre $-\infty$ et a_1, puis la valeur $1/n$ entre a_1 et a_2,... , la valeur $(n-1)/n$ entre a_{n-1} et a_n, et enfin la valeur 1 entre a_n et $+\infty$. Insistons sur le fait que, par nature, ω est *fixé* ; la figure 20.1 page ci-contre montre deux réalisations de $x \mapsto F_n(x,\omega)$, pour deux valeurs ω_1 et ω_2 données[11], pour une même valeur $n = 4$[12].

(10). Expérience « physique » aussi bien que « numérique ».

(11). C'est-à-dire que nous avons réalisé deux expériences numériques distinctes.

(12). Sur ces figures, nous avons relié les traits horizontaux qui sont bien entendu les seuls à faire partie du graphe de $F_n(\cdot,\omega)$; les traits verticaux ne sont là que pour améliorer la lisibilité du graphe.

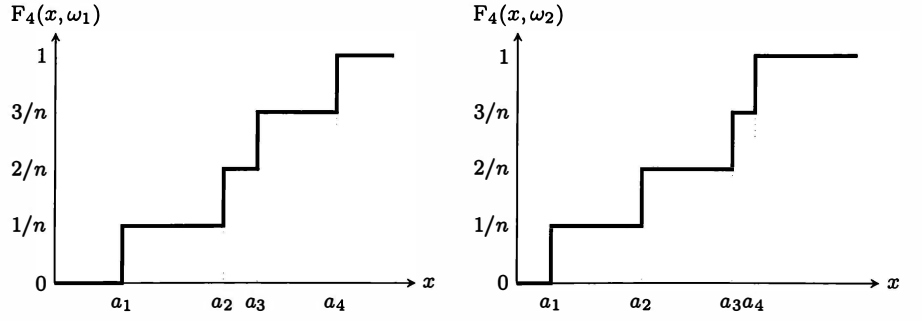


FIGURE 20.1 — *Les fonctions de répartition empiriques* $F_4(\cdot,\omega)$*, pour deux valeurs distinctes de* ω.

Le théorème de Glivenko-Cantelli affirme que si $(X_n)_{n\geqslant 0}$ est une suite d'échantillons de X, alors la suite des fonctions de répartition empiriques converge uniformément vers la fonction de répartition de X.

■ **Théorème 20.27 (Glivenko-Cantelli)** *Soit* X *une variable aléatoire de fonction de répartition* F. *Soit* $(X_n)_{n\geqslant 1}$ *une suite de variables aléatoires indépendantes et de même loi que* X. *Alors, presque sûrement, la suite de terme général* $F_n(x,\cdot)$ *converge uniformément vers* F :

$$\mathbf{P}\left\{\omega\in\Omega\,;\ \sup_{x\in\mathbb{R}}\left|F_n(x,\omega)-F(x)\right|\xrightarrow[n\to\infty]{}0\right\}=1.$$

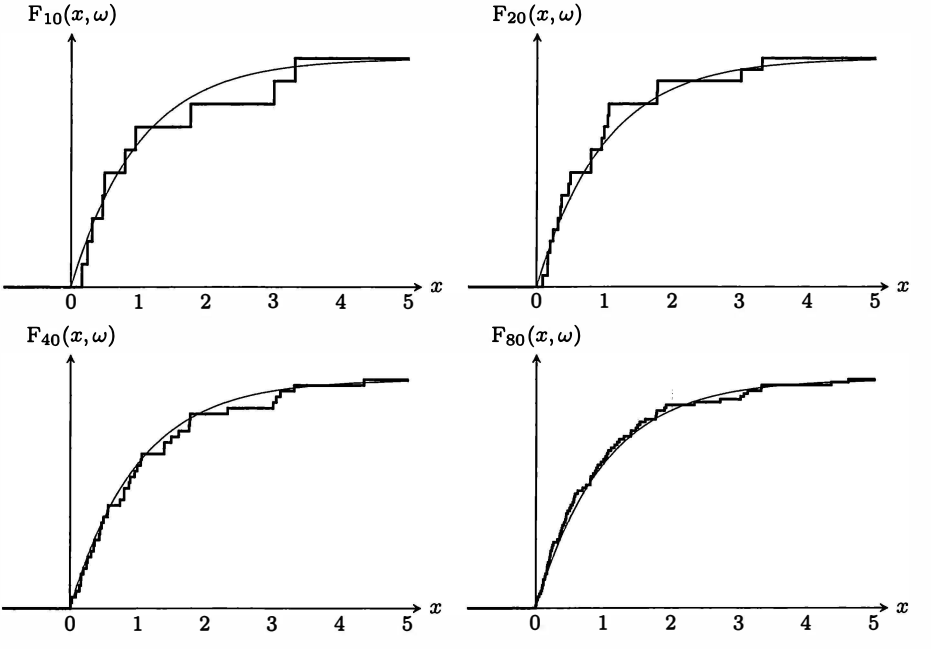


FIGURE 20.2 — *Les fonctions* $F_n(\cdot,\omega)$ *pour* $n=10$, $n=20$, $n=40$ *et* $n=80$, *pour* une *valeur de* ω *donnée. En trait mince, la fonction de répartition de la loi* $\mathscr{E}(1)$.

Ce théorème est, pour une bonne part, une conséquence de la loi des grands nombres ; nous ne le démontrons pas ici[13].

Remarque 20.28 On remarquera simplement un résultat plus faible : toujours d'après la loi des grands nombres, pour chaque x, on sait qu'il existe une partie Ω_x de Ω, de probabilité 1, telle que

$$\forall \omega \in \Omega_x \qquad \lim_{n\to\infty} \mathrm{F}_n(x,\omega) = \mathrm{F}(x)$$

Le théorème de Glivenko-Cantelli montre que cette propriété a lieu pour un ensemble Ω' indépendant de x[14], et de façon uniforme.

Si l'on simule informatiquement une variable aléatoire X suivant la loi exponentielle $\mathscr{E}(1)$[15] et que l'on construit les graphes de la fonction de répartition empirique, pour une *même* valeur de ω, et pour différentes valeurs de n, on obtient des résultats semblables à ceux présentés à la figure 20.2.

§313 **Application : le test de Kolmogorov-Smirnov**

Un des traits remarquables du théorème de Glivenko-Cantelli est que la vitesse de convergence de F_n vers F peut être contrôlée de manière absolue, sous la seule réserve que F soit *continue*. Andreï Kolmogorov a montré que, avec les notations précédentes, on a pour tout réel $r > 0$:

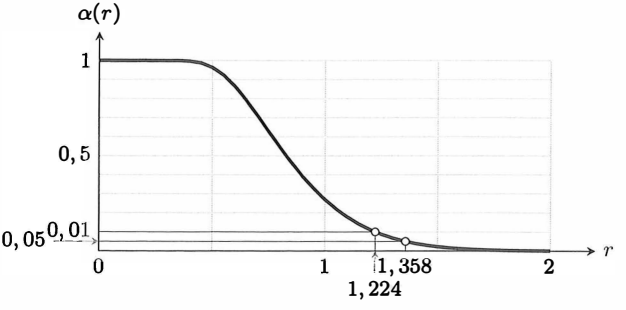
$$\mathbf{P}\left\{ \|\mathrm{F}_n - \mathrm{F}\|_\infty \geqslant \frac{r}{\sqrt{n}} \right\} \xrightarrow[n\to\infty]{} \alpha(r), \tag{20.1}$$

où α est une fonction bien tabulée, pouvant s'exprimer sous la forme

$$\alpha(r) = 2\sum_{k=0}^{\infty} (-1)^{k-1} \mathrm{e}^{-2k^2r^2} = 1 - \frac{\sqrt{2\pi}}{r} \sum_{k=1}^{\infty} \mathrm{e}^{-(2k-1)^2\pi^2/8r^2}$$

(la seconde se déduisant de la première par la formule sommatoire de Poisson). Quelques valeurs utiles de la variable r sont listées ci-dessous, la valeur $\alpha(r)$ étant le risque accepté :

r	1,2238	1,3581	1,4802	1,9495
$\alpha(r)$	0,1	0,05	0,025	0,01



[13]. Le lecteur pourra trouver des démonstrations agréables de ce théorème dans plusieurs ouvrages, notamment [9, 77, 63].

[14]. Étant donné qu'il y a une infinité *non dénombrable* de tels x, l'intersection de tous les Ω_x n'est pas *a priori* de probabilité 1. En réalité, les fonctions étant toutes continues à droite, on peut se limiter aux x rationnels : $\sup_{x\in\mathbb{R}} |\mathrm{F}_n(x,\omega) - \mathrm{F}(x)| = \sup_{q\in\mathbb{Q}} |\mathrm{F}_n(q,\omega) - \mathrm{F}(q)|$. L'intersection de tous les Ω_q est alors de probabilité 1.

[15]. Les méthodes permettant d'effectuer une telle simulation sont présentées dans le chapitre 25, et plus précisément dans l'exemple 25.8 page 515.

Si l'on connaît un échantillon $(x_1, x_2, \ldots, x_n)$ de valeurs provenant d'une loi de fonction de répartition F *continue* donnée, on construit la fonction de répartition empirique

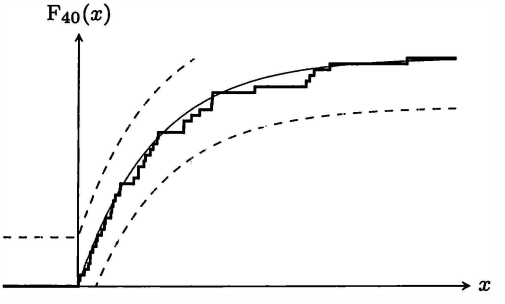
$$\mathrm{F}_n(x) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{]-\infty\,;\,x]}(x_k)$$

puis l'on détermine $\|\mathrm{F}_n - \mathrm{F}\|_\infty$ (il suffit pour cela de calculer les valeurs de $\mathrm{F}(x_k) - \mathrm{F}_n(x_k^-)$ et de $\mathrm{F}(x_k) - \mathrm{F}_n(x_k^+)$, et de prendre le plus grand de ces $2n$ nombres en valeur absolue). On choisit ensuite un risque α (une valeur « classique » est $\alpha = 0,05$) et on fixe la valeur de r correspondante ($r \approx 1,3581$ pour $\alpha = 0,05$). Le **test de Kolmogorov-Smirnov** consiste à dire que l'hypothèse

$$\mathscr{H}_0 \ : \quad \text{« l'échantillon suit la loi F »}$$

est rejetée, au risque α, si $\|\mathrm{F}_n - \mathrm{F}\|_\infty > r/\sqrt{n}$. Le choix de la **rejeter** traduit le fait que, si l'hypothèse $\mathscr{H}_0$ était acceptée, alors l'écart $\|\mathrm{F}_n - \mathrm{F}\|_\infty$ ne pourrait prendre une valeur aussi grande qu'avec une probabilité inférieure ou égale à α — événement dont on pourra juger la réalisation trop improbable pour être crédible.

Exemple 20.29 On considère des données $x_1 = 3,0566$, $x_2 = 1,4957$,... jusqu'à $x_{40} = 1,7762$, et on construit la fonction de répartition empirique F_{40} associée. L'hypothèse « ces données proviennent d'une loi exponentielle de paramètre 1 » peut subir le test de Kolmogorov-Smirnov, puisque la fonction de répartition F de la loi $\mathscr{E}(1)$ est continue. Si l'on choisit le risque $\alpha = 0,05$, on prend $r = 1,3581$, et donc $r/\sqrt{n} \approx 0,215$. On trace ci-dessous le graphe de F_n, ainsi, en grisé, que la zone entourant le graphe de F à $\pm 0,215$ près. On constate sans peine que le graphe de F_n est entièrement dans la zone grisée, ce qui fait que l'on ne peut pas, au risque de $\alpha = 5$ %, rejeter l'hypothèse faite.



Remarque 20.30 Le test de Kolmogorov-Smirnov est un **test non-paramétrique** : l'équation (20.1) ne fait référence à aucun paramètre (comme le paramètre de la loi exponentielle dont on veut mesurer l'adéquation pour décrire la liste des données expérimentales). Pour plus de renseignements sur la notion de test statistique, le lecteur est renvoyé, par exemple, à un très clair ouvrage de Jourdain [54].

20.4 ANNEXE : DÉMONSTRATION DE LA LOI DES GRANDS NOMBRES (*)

Nous proposons ici une démonstration élémentaire de la loi forte des grands nombres, suivant dans les grandes lignes le raisonnement de Luca PRATELLI[16] [80].

On se donne une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, de même loi, intégrables et d'espérance m ; on note

$$Y_n := \frac{X_1 + X_2 + \cdots + X_n}{n} - m = \frac{\tilde{X}_1 + \tilde{X}_2 + \cdots + \tilde{X}_n}{n},$$

où $\tilde{X}_k := X_k - m$ est la variable centrée issue de X_k.

■ **Théorème 20.31 (Pratelli)** *Sous les hypothèses précédentes,* $Y_n \xrightarrow{ps} 0$.

Remarque 20.32 La démonstration de Pratelli est un peu plus générale, puisqu'elle s'applique à des suites de variables aléatoires *échangeables*, c'est-à-dire telles que, pour tout entier $n \geqslant 1$ et toute permutation σ de $\{1, 2, \ldots, n\}$, les vecteurs $(X_1, X_2, \ldots, X_n)$ et $(X_{\sigma(1)}, X_{\sigma(2)}, \ldots, X_{\sigma(n)})$ ont même loi. Les suites de variables aléatoires indépendantes et identiquement distribuées sont échangeables ; les variables échangeables sont identiquement distribuées, mais pas forcément indépendantes.

La structure de la démonstration est la suivante :

- par une technique classique de troncature, on montre que la suite $(Y_n)_{n\geqslant 1}$ converge vers 0 dans L^1, c'est-à-dire $\mathbf{E}\,|Y_n| \to 0$. Au passage, on fait appel à une version de la loi des grands nombres pour les variables de l'espace L^4 (exercice 20.2).
- On en déduit, par une inégalité astucieuse, que $\mathbf{P}\{\sup_{k\geqslant n} |Y_k| > \varepsilon\} \to 0$ pour tout $\varepsilon > 0$, ce qui d'après le théorème 19.5 montre la convergence presque sûre de $(Y_n)_{n\geqslant 1}$ vers 0.

Pour simplifier les notations, on supposera désormais que $m = 0$ (et on utilisera donc les X_k au lieu des $\tilde{X}_k$).

Commençons par récrire le résultat de Cantelli :

■ **Proposition 20.33 (Cantelli)** *Soit* $(Z_n)_{n\geqslant 1}$ *une suite de variables aléatoires indépendantes, identiquement distribuées, admettant des moments d'ordre* 4 *; si l'on note*

$$M_n = \frac{Z_1 + Z_2 + \cdots + Z_n - n\mathbf{E}Z_1}{n},$$

alors $\mathbf{E}\,|M_n| \to 0$.

Cette proposition est démontrée à la première question de l'exercice 20.2. Voici maintenant le premier résultat intermédiaire :

■ **Lemme 20.34 (Loi des grands nombres dans L^1)** *La suite* $(Y_n)_{n\geqslant 1}$ *tend vers* 0 *dans* L^1, *c'est-à-dire que* $\mathbf{E}\,|Y_n| \to 0$.

DÉMONSTRATION : Soit $\varepsilon > 0$. On va montrer qu'il existe une fonction f continue et *bornée* telle que les variables $Z_k = f(X_k)$ soient bornées, indépendantes, identiquement distribuées, centrées, et telles que $\mathbf{E}\,|Z_k - X_k| \leqslant \varepsilon$.

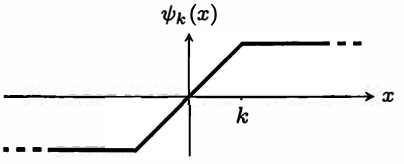
On commence par construire une variable Z_1 qui soit bornée, ε-proche de X_1 dans L^1 (c'est-à-dire $\mathbf{E}\,|Z_1 - X_1| \leqslant \varepsilon$) et centrée, selon des techniques vues dans les chapitres 13

(16). La démonstration originelle de Pratelli utilise la complétude de l'espace L^2. La version plus élémentaire que nous donnons ici est inspirée des notes de cours de Julien MICHEL.

et 14. Notons $\varphi_k : x \mapsto \min(x, k)$, et décomposons $X_1 = X_1^+ - X_1^-$. La suite de terme général $\varphi_k(X_1^\pm)$ croît vers $X_1^\pm$ donc $\mathbf{E}\varphi_k(X_1^\pm) \to \mathbf{E}(X_1^\pm)$. Notamment, on peut trouver un entier k tel que, en notant $X_1' = \varphi_k(X_1^+) - \varphi_k(X_1^-)$, on ait

$$\mathbf{E}\,|X_1 - X_1'| \leqslant \frac{\varepsilon}{2}.$$

On remarquera que l'on a en fait écrit $X_1' = \psi_k(X_1)$, où ψ_k est la fonction continue suivante :



Enfin, on centre cette variable : $Z_1 := X_1' - \mathbf{E}(X_1')$. Au total, on a écrit Z_1 sous la forme $Z_1 = f(X_1)$, où $f = \psi_k - m_1$ et $m_1 = \mathbf{E}(X_1')$. Comme $\mathbf{E}(X_1')$ est $\varepsilon/2$-proche de $\mathbf{E}(X_1) = 0$, on en déduit que $\mathbf{E}\,|Z_1 - X_1| \leqslant \varepsilon$.

On définit de même la suite de terme général $Z_n = f(X_n)$: ce sont encore des variables aléatoires indépendantes, identiquement distribuées, centrées, et de plus bornées, donc admettant des moments de tous ordres.

On peut donc appliquer le résultat de Cantelli : $\mathbf{E}\,|M_n| \to 0$. Notamment, pour tout n suffisamment grand, on a $\mathbf{E}\,|M_n| \leqslant \varepsilon$. Or, par linéarité, $\mathbf{E}\,|Y_n - M_n| \leqslant \varepsilon$, donc $\mathbf{E}\,|Y_n| \leqslant 2\varepsilon$ pour n suffisamment grand. ∎

Voici maintenant le résultat qui est au cœur de la démonstration de Pratelli :

■ **Lemme 20.35** *Pour tout $n \geqslant 1$ et tout $\varepsilon > 0$,*

$$\varepsilon\,\mathbf{P}\left\{\sup_{k \geqslant n} |Y_k| > \varepsilon\right\} \leqslant \mathbf{E}\,|Y_n|\,.$$

DÉMONSTRATION : Soient $n \geqslant 1$ et $\varepsilon > 0$. Il suffit en pratique de démontrer que, pour tout $m \geqslant n$, on a

$$\varepsilon\,\mathbf{P}\left\{\sup_{n \leqslant k \leqslant m} |Y_k| > \varepsilon\right\} \leqslant \mathbf{E}\,|Y_n|\,.$$

Fixons donc l'entier $m > n$. Notons maintenant T le plus grand entier k de $[1\,;m]$ vérifiant $|Y_k| > \varepsilon$, c'est-à-dire tel que

$$\{T = k\} = \big\{\,|Y_k| > \varepsilon, \quad |Y_{k+1}| \leqslant \varepsilon, \ldots, |Y_m| \leqslant \varepsilon\big\}.$$

(On convient que $T = 0$ si tous les $|Y_k|$ sont inférieurs ou égaux à ε.) L'ensemble dont on cherche à majorer la probabilité s'écrit

$$\left\{\sup_{n \leqslant k \leqslant m} |Y_k| > \varepsilon\right\} = \{T \geqslant n\} = \bigcup_{k=n}^{m} \{T = k\}\,.$$

Comme l'union précédente est disjointe, on va simplement majorer individuellement la probabilité de chaque événement. Pour cela, écrivons d'abord que

$$\varepsilon\,\mathbf{P}\{T = k\} = \varepsilon \int_{\{T=k\}} \mathrm{d}\mathbf{P} \leqslant \int_{\{T=k\}} |Y_k|\,\mathrm{d}\mathbf{P}$$

puisque $|Y_k|$ prend justement des valeurs $> \varepsilon$ sur l'événement $\{T = k\}$. On sépare enfin cette intégrale en deux :

$$\varepsilon\,\mathbf{P}\{T = k\} \leqslant \int_{\{T=k,\,Y_k>0\}} Y_k\,\mathrm{d}\mathbf{P} - \int_{\{T=k,\,Y_k<0\}} Y_k\,\mathrm{d}\mathbf{P}$$

C'est là qu'interviennent les hypothèses faites sur les variables $\tilde{X}_i$, qui sont centrées, indépendantes et identiquement distribuées. Puisque, pour chaque intégrale, on intègre sur un

domaine qui est *symétrique* en $(\tilde{X}_1, \dots, \tilde{X}_k)$, c'est-à-dire invariant par toute permutation $(X_1, \dots, X_k) \mapsto (X_{\sigma(1)}, \dots, X_{\sigma(k)})$ des variables, on en déduit que[17]

$$\int_{\{T=k, Y_k>0\}} Y_k \, d\mathbf{P} = \frac{1}{k} \sum_{i=1}^{k} \int_{\{T=k, Y_k>0\}} \tilde{X}_i \, d\mathbf{P} = \int_{\{T=k, Y_k>0\}} \tilde{X}_1 \, d\mathbf{P}$$
$$= \frac{1}{n} \sum_{i=1}^{n} \int_{\{T=k, Y_k>0\}} \tilde{X}_i \, d\mathbf{P} = \int_{\{T=k, Y_k>0\}} Y_n \, d\mathbf{P}.$$

De même,

$$\int_{\{T=k, Y_k<0\}} Y_k \, d\mathbf{P} = \int_{\{T=k, Y_k<0\}} Y_n \, d\mathbf{P}.$$

Par conséquent,

$$\varepsilon \, \mathbf{P}\{T = k\} \leqslant \int_{\{T=k, Y_k>0\}} Y_n \, d\mathbf{P} - \int_{\{T=k, Y_k<0\}} Y_n \, d\mathbf{P} \leqslant \int_{\{T=k\}} |Y_n| \, d\mathbf{P}$$

et il ne reste plus qu'à sommer ces inégalités pour toutes les valeurs de k pour trouver, les ensembles étant disjoints :

$$\mathbf{P}\left\{ \sup_{n \leqslant k \leqslant m} |Y_k| > \varepsilon \right\} = \sum_{k=n}^{m} \{T = k\} \leqslant \sum_{k=n}^{m} \int_{\{T=k\}} |Y_n| \, d\mathbf{P} \leqslant \int_{\Omega} |Y_n| \, d\mathbf{P} = \mathbf{E}|Y_n|,$$

ce qu'il fallait démontrer. ∎

Une fois ces deux lemmes en poche, on peut démontrer le théorème général. D'après le théorème 20.34, on a $\mathbf{E}|Y_n| \to 0$, et le lemme 20.35 montre que $\mathbf{P}\{\sup_{k \geqslant n} |Y_k| > \varepsilon\} \to 0$ quand n tend vers l'infini, et donc, par le théorème 19.5, $Y_n \to 0$ presque sûrement.

EXERCICES

♦ ***Exercice 20.1*** Soit $(X_n)_{n \geqslant 0}$ une suite de variables indépendantes, de même loi, admettant un moment d'ordre 2. On note m leur espérance et σ leur écart-type. Montrer que la suite de terme général

$$Y_n := \frac{1}{n} \sum_{k=1}^{n} (X_k - m)^2$$

converge presque sûrement vers σ^2.

♦ ***Exercice 20.2 (Un cas particulier de loi forte des grands nombres par Cantelli)*** On suppose que les variables $(X_n)_{n \geqslant 1}$ sont indépendantes, identiquement distribuées, de même loi que X, et qu'elles admettent un moment d'ordre 4. On va montrer le **théorème de Cantelli** : *la suite de terme général* $Y_n = S_n/n = (X_1 + X_2 + \cdots + X_n)/n$ *converge presque sûrement vers* $m = \mathbf{E}(X)$.

i) Quitte à poser $X'_k = X_k - m$ et $S'_n = S_n - nm$, on peut supposer que $m = 0$. Montrer que

$$\mathbf{E}(S_n^4) = n\,\mathbf{E}(X^2) + 3n(n-1)\,\mathbf{E}(X^4).$$

ii) En déduire que $Y_n^4 \to 0$ presque sûrement (on utilisera le corollaire 19.7 page 387).

iii) Démontrer que la même conclusion a lieu si l'on suppose seulement que les variables X_k sont indépendantes et admettent un moment d'ordre 4 uniformément majoré : $\mathbf{E}(X_k^4) \leqslant C$ pour tout entier k. *(On utilisera l'inégalité de Cauchy-Schwarz pour majorer uniformément* $\mathbf{E}(X_k^2)$*.)*

♦ ***Exercice 20.3*** Soit $f : [0\,;1] \to \mathbb{R}$ une fonction continue. Pour tout entier $n \geqslant 1$, on pose

$$I_n := \int_{[0\,;1]^n} f\left(\frac{x_1 + x_2 + \cdots + x_n}{n} \right) dx_1 \, dx_2 \cdots dx_n.$$

Montrer que $(I_n)_{n \geqslant 1}$ admet une limite que l'on déterminera.

---

[17]. Il apparaît ici que l'indépendance des variables n'est pas une nécessité, seul le caractère *échangeable* étant utilisé.

♦ ***Exercice 20.4*** Soient $f, g : [0\,;1] \to \mathbb{R}$ deux fonctions continues. On suppose de plus que g est à valeurs strictement positives. Montrer que

$$\lim_{n\to\infty} \int_{[0\,;1]^n} \frac{f(x_1)+f(x_2)+\cdots+f(x_n)}{g(x_1)+g(x_2)+\cdots+g(x_n)}\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_n = \frac{\int_0^1 f}{\int_0^1 g}.$$

♦ ***Exercice 20.5 (Autour de la loi forte...)*** Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées. Soit $f : \mathbb{R}^3 \to \mathbb{R}$ une fonction mesurable (borélienne). On pose, pour tout entier n :

$$Y_n := f(X_n, X_{n+1}, X_{n+2})$$

On suppose de plus que Y_1 admet une espérance.

i) Les variables aléatoires Y_n sont-elles indépendantes ?

ii) On note $U_n = Y_{3n}$, $V_n = Y_{3n+1}$ et $W_n = Y_{3n+2}$. Montrer que la suite de terme général $(U_1+U_2+\cdots+U_n)/n$ converge presque sûrement vers $\mathbf{E}Y_1$; et de même pour les moyennes successives des V_i et celles des W_i.

iii) Montrer que la suite de terme général Y_n/n converge presque sûrement vers 0.

iv) En déduire que la suite de terme général $M_n = (Y_1+Y_2+\cdots+Y_n)/n$ converge presque sûrement vers $\mathbf{E}Y_1$.

Indications

◊ *Indication 20.3 :* I_n est l'espérance de $f(M_n)$, où M_n est une variable aléatoire à déterminer.

◊ *Indication 20.4 :* Interpréter l'intégrale comme l'espérance de M_n/M'_n, où M_n et M'_n sont deux variables aléatoires à déterminer.

◊ *Indication 20.5 : iii)* : voir l'exercice 19.13.

SOLUTIONS

♦ **Solution 20.2**

i) Écrivons $S_n^4 = \sum_i \sum_j \sum_k \sum_\ell X_i X_j X_k X_\ell$, de sorte que

$$\mathbf{E}(S_n^4) = \sum_i \sum_j \sum_k \sum_\ell \mathbf{E}(X_i X_j X_k X_\ell).$$

Dans cette somme, les quatre indices varient, indépendamment, entre 1 et n. Les termes sont alors de plusieurs types, selon que certains indices sont égaux ou non. Par exemple, si $i = j = k \neq \ell$, on a

$$\mathbf{E}(X_i X_j X_k X_\ell) = \mathbf{E}(X_i^3)\,\mathbf{E}(X_\ell) = 0$$

puisque $\mathbf{E}(X_\ell) = 0$. Si par exemple $i = j \neq k = \ell$, on a, par indépendance,

$$\mathbf{E}(X_i X_j X_k X_\ell) = \mathbf{E}(X_i^2)\,\mathbf{E}(X_k^2).$$

Chaque fois qu'un indice se retrouve seul avec cette valeur, on obtient un terme nul, ce qui élimine une grande partie des termes. Il ne reste donc que les termes où tous les indices sont égaux, qui sont au nombre de n ; et les termes où les indices sont égaux par paires. Il y a 3 manières de choisir deux paires d'indices parmi 4, et pour chaque choix, il y a $n(n-1)$ valeurs possibles (n valeurs pour le premier couple d'indices, et $n-1$ pour le second couple). On trouve alors bien

$$\mathbf{E}(S_n^4) = n\,\mathbf{E}(X^2) + 3n(n-1)\,\mathbf{E}(X^4).$$

ii) Notamment, puisque $Y_n = S_n/n$,

$$\mathbf{E}(Y_n^4) \leqslant \frac{\mathbf{E}(X^2)}{n^3} + \frac{3\,\mathbf{E}(X^4)}{n^2}$$

est le terme général d'une série convergente. D'après le corollaire 19.7, la suite de terme général Y_n^4 converge presque sûrement vers 0[18], et il en est de même de Y_n.

---

(18). On rappelle brièvement le raisonnement : par l'inégalité de Markov, pour $\varepsilon > 0$, on a

$$\mathbf{P}\{\,|Y_n| > \varepsilon\} = \mathbf{P}\{\,|Y_n|^4 > \varepsilon^4\} \leqslant \frac{\mathbf{E}(Y_n^4)}{\varepsilon^4},$$

ce qui montre que la série de terme général $\mathbf{P}\{\,|Y_n| > \varepsilon\}$ converge, donc par le premier lemme de Borel-Cantelli, $\mathbf{P}\{\,|Y_n| > \varepsilon \text{ ; i.s.}\} = 0$, ce qui est une simple traduction de $\mathbf{P}\{Y_n \to 0\} = 1$ d'après les raisonnements vus au paragraphe § 279 page 386.

iii) Si l'on ne suppose plus la suite identiquement distribuée, la majoration de $\mathbf{E}(S_n^4)$ reste valable en remarquant que, par Cauchy-Schwarz, on a une majoration uniforme des produits de variances :

$$\mathbf{E}(X_i^2)\,\mathbf{E}(X_k^2) = \mathbf{E}(X_i^2 X_k^2) \leqslant \sqrt{\mathbf{E}(X_i^4)\,\mathbf{E}(X_k^4)} \leqslant C.$$

♦ **Solution 20.3** Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes de même loi uniforme sur $[0\,;1]$. On pose alors

$$M_n := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

pour tout entier $n \geqslant 1$. Le n-uplet $(X_1, X_2, \ldots, X_n)$ admet une densité uniforme sur $[0\,;1]^n$ (produit des densités de chaque X_i, par indépendance) ; la formule de transfert permet donc d'interpréter l'intégrale I_n comme $I_n = \mathbf{E}f(M_n)$. Or, d'après la loi forte des grands nombres, la suite $(M_n)_{n\geqslant 1}$ converge presque sûrement vers $\mathbf{E}(X_1) = 1/2$; par continuité, $f(M_n)$ converge presque sûrement vers $f(1/2)$. Enfin, f étant continue sur un compact, elle est bornée par un réel $A > 0$, donc la suite de terme général $f(M_n)$ est dominée par A. Le théorème de convergence dominée permet alors de conclure que

$$\lim_{n\to\infty} I_n = \lim_{n\to\infty} \mathbf{E}f(M_n) = f\left(\frac{1}{2}\right).$$

♦ **Solution 20.4** Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes de même loi uniforme sur $[0\,;1]$. On pose $Y_i = f(X_i)$ et $Z_i = g(X_i)$, et

$$M_n := \frac{Y_1 + \cdots + Y_n}{n} \qquad \text{et} \qquad M'_n := \frac{Z_1 + \cdots + Z_n}{n}.$$

La formule de transfert permet d'interpréter l'intégrale étudiée comme

$$\int_{[0\,;1]^n} \frac{f(x_1) + f(x_2) + \cdots + f(x_n)}{g(x_1) + g(x_2) + \cdots + g(x_n)}\,dx_1\,dx_2\cdots dx_n = \mathbf{E}\left(\frac{M_n}{M'_n}\right).$$

Or, d'après la loi des grands nombres, $(M_n)_{n\geqslant 1}$ converge presque sûrement vers $\mathbf{E}(Y_1) = \mathbf{E}f(X_1) = \int_0^1 f$, tandis que $(M'_n)_{n\geqslant 1}$ converge presque sûrement vers $\mathbf{E}(Z_1) = \mathbf{E}g(X_1) = \int_0^1 g$. Notamment,

$$\frac{M_n}{M'_n} \xrightarrow{\text{ps}} \frac{\int_0^1 f}{\int_0^1 g}.$$

Notons enfin $A = \sup_{[0\,;1]} |f|$ et $B = \inf_{[0\,;1]} |g|$. Ce dernier nombre est strictement positif (g est continue sur un compact, donc atteint son minimum). La suite de terme général $|M_n/M'_n|$ est dominée par la constante A/B, le théorème de convergence dominée permet de conclure que

$$\mathbf{E}\left(\frac{M_n}{M'_n}\right) \xrightarrow[n\to\infty]{} \frac{\int_0^1 f}{\int_0^1 g}.$$

♦ **Solution 20.5**

i) Les variables Y_n n'ont aucune raison d'être indépendantes ; plus précisément, si $|p-n| \geqslant 3$, Y_p et Y_n sont indépendantes (lemme des coalitions), mais Y_n et Y_{n+1} ne le sont pas *a priori*.

ii) Les U_n sont indépendantes, de même loi égale à la loi de Y_1. La loi forte des grands nombres montre que $(U_1 + \cdots + U_n)/n$ converge presque sûrement vers $\mathbf{E}Y_1$.

iii) Voir l'exercice 19.13.

iv) En sommant les relations trouvées,

$$\frac{1}{3n}(U_1 + \cdots + U_n + V_1 + \cdots + V_n + W_1 + \cdots + W_n) \xrightarrow{\text{ps}} \mathbf{E}Y_1,$$

ou encore $M_{3n} \xrightarrow{\text{ps}} \mathbf{E}Y_1$.

Il ne reste plus qu'à remarquer que

$$M_{3n+1} = \frac{3n}{3n+1}\,Y_{3n} + \frac{1}{3n+1}\,Y_{3n+1} \xrightarrow{\text{ps}} \mathbf{E}Y_1$$

et

$$M_{3n+2} = \frac{3n}{3n+2}\,Y_{3n} + \frac{1}{3n+1}(Y_{3n+1} + Y_{3n+2}) \xrightarrow{\text{ps}} \mathbf{E}Y_1$$

pour conclure $M_n \xrightarrow{\text{ps}} \mathbf{E}Y_1$.

Chapitre 21

Le théorème limite central

Le célèbre théorème limite central permet d'aller au-delà du résultat de la loi des grands nombres : si $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes et de même loi, admettant une espérance m et une variance σ^2, alors les fluctuations de la moyenne $M_n = (X_1 + \cdots + X_n)/n$ autour de m se font sur une échelle typique en $\sigma/\sqrt{n}$ et suivent, asymptotiquement, une loi précise : la loi normale $\mathcal{N}(0\,;\,\sigma^2/n)$. En d'autres termes, on peut approcher la loi de $X_1 + X_2 + \cdots + X_n$ par la loi normale $\mathcal{N}(nm\,;\,n\sigma^2)$.

D'autres formes du théorème permettent aux variables X_n de suivre des lois différentes, à condition que leurs variances ne deviennent pas trop petites.

Les conséquences de ce théorème fondamental sont innombrables, et tout d'abord en statistiques : si l'on ajoute des quantités inconnues mais indépendantes, le résultat est (presque) une loi normale !

En conclusion du chapitre, nous présentons un résultat connu sous le nom de loi du logarithme itéré, *qui permet d'encadrer les fluctuations de $S_n/\sqrt{n}$.*

Au-delà de la loi des grands nombres

§ 314

Dans le paragraphe § 277, nous avons vu l'évolution de la loi du nombre de *pile* obtenus au cours d'une succession de tirages à *pile* ou *face* indépendants. Notamment, la moyenne

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

converge (en probabilité et presque sûrement) vers une constante m, qui est ici la probabilité d'obtenir *pile* à chaque tirage. Cette propriété reste vraie sous l'hypothèse plus générale que les X_i sont des variables aléatoires indépendantes, de même loi, admettant une espérance m.

Dans la vie réelle, on ne peut pas réitérer une infinité de fois une expérience, pas plus qu'on ne saurait lancer une infinité de fois la pièce. Une question naturelle est alors : peut-on quantifier la vitesse de cette convergence ? En d'autres termes, quelle est l'importance des fluctuations autour de la valeur limite dans une figure telle que celles de la page 384 ?

La largeur typique de ces fluctuations est donnée simplement par l'écart-type de M_n, s'il est défini ; supposons désormais que les X_i admettent un moment d'ordre 2 et notons $\sigma = \sigma_{X_1}$, l'hypothèse d'indépendance des variables $X_1, X_2, \ldots, X_n$ permet alors de conclure que leurs variances s'ajoutent (égalité de Bienaymé), et donc

que l'écart-type

$$\sigma_{M_n} = \frac{\sqrt{n}\,\sigma}{n} = \frac{\sigma}{\sqrt{n}}$$

de M_n tend vers 0.

Changeons alors d'échelle pour rendre visible les fluctuations voulues, en ramenant cet écart-type à 1. Profitons de l'occasion pour centrer la variable. On pose donc

$$Y_n = \frac{M_n - \mathbf{E}(M_n)}{\sigma_{M_n}} = \frac{M_n - m}{\sigma/\sqrt{n}}.$$

Cette variable, par construction, est d'espérance nulle et d'écart-type unité : c'est la variable centrée réduite issue de M_n. C'est d'ailleurs également la variable centrée réduite issue de $S_n = X_1 + X_2 + \cdots + X_n$ (1) :

$$Y_n = M_n^* = S_n^* = \frac{(X_1 + X_2 + \cdots + X_n) - nm}{\sigma_{X_1}\sqrt{n}}.$$

Exemple 21.1 (Lois de Poisson) Soit $(X_n)_{n\geqslant 1}$ des variables aléatoires indépendantes, suivant la même loi de Poisson $\mathscr{P}(\lambda)$. Alors $\mathbf{E}(X_1) = \lambda$ et $\sigma_{X_1} = \sqrt{\lambda}$; la variable Y_n est définie par

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n - n\lambda}{\sqrt{n\lambda}}.$$

Il est bien entendu inutile d'espérer un théorème de style « loi forte des grands nombres », c'est-à-dire une convergence presque sûre vers une variable aléatoire donnée ; même une convergence en probabilité est, en fait, exclue (2). Tout ce que nous pouvons espérer, c'est trouver la loi de distribution de Y_n. Or, et c'est là qu'arrive un petit miracle, sous les hypothèses très générales que nous avons faites, la suite $(Y_n)_{n\geqslant 1}$ converge, en loi, vers une loi normale et ce, indépendamment de tout détail de la loi des X_n.

§ 315 **Le théorème limite central**

Il existe de nombreux théorèmes qui quantifient le comportement gaussien des fluctuations de la moyenne de variables aléatoires (plus ou moins) indépendantes et (plus ou moins) identiquement distribuées. Toutes ces variantes portent le nom de « théorème limite central », en suivant la dénomination de Polyà (3). Voici la forme la plus simple et la plus fréquente de ces théorèmes :

■ **Théorème 21.2 (Théorème limite central)** *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes, de même loi, et admettant une variance $\sigma^2 > 0$. Notons $m = \mathbf{E}(X_1)$ leur espérance. Posons enfin $S_n = X_1 + X_2 + \cdots + X_n$ et*

$$Y_n = S_n^* = \frac{S_n - nm}{\sigma\sqrt{n}}.$$

Alors la suite $(Y_n)_{n\geqslant 1}$ converge en loi vers la loi normale centrée réduite :

$$Y_n \xrightarrow{L} \mathscr{N}(0\,;\,1).$$

(1). Voir l'exercice 7.7 page 167, évidemment valable pour tout type de variables aléatoires, et pas seulement les variables discrètes.

(2). Cf. remarque 21.6 page suivante.

(3). Qui a appelé ce théorème *zentraler Grenzwertsatz*, ou « théorème central (fondamental) sur les limites ». Les anglo-saxons l'appellent *central limit theorem*, ce qui fait qu'il a longtemps été connu en France sous le nom de « théorème central limite ». Jugeant sans doute cette appellation trop franglaise, certains l'ont renommé « théorème de la limite centrée », ce qui paraît plutôt difficile à justifier.

Remarque 21.3 (Convergence uniforme) La conclusion du théorème précédent (Y_n converge en loi vers $\mathcal{N}(0\,;\,1)$) peut s'écrire sous les formes (équivalentes) suivantes :

— pour tout $x \in \mathbb{R}$, $\mathbf{P}\{Y_n \leqslant x\} \xrightarrow[n\to\infty]{} \mathfrak{N}(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-t^2/2}\,dt$;

— pour tout $a < b$ réels, $\mathbf{P}\{a \leqslant Y_n \leqslant b\} \xrightarrow[n\to\infty]{} \mathfrak{N}(b) - \mathfrak{N}(a) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{a}^{b} e^{-t^2/2}\,dt$.

La première formulation revient à dire que : la suite des fonctions de répartition F_n de S_n^* converge simplement vers $\mathfrak{N}$. La continuité de $\mathfrak{N}$ entraîne que cette convergence est en réalité *uniforme*[4] :

■ **Théorème 21.4** *Sous les hypothèses du théorème 21.2, la suite de terme général* F_n *converge uniformément vers* $\mathfrak{N}$.

Remarque 21.5 (Forme alternative) On peut formuler le résultat du théorème limite central sous la forme suivante : posons

$$Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_n - m).$$

Alors $Z_n \xRightarrow{L} \mathcal{N}(0\,;\,\sigma^2)$. Cette dernière formulation a pour intérêt d'être encore valable si $\sigma = 0$ (cas dégénéré), et de se généraliser plus facilement aux suites de vecteurs aléatoires (§ 319).

La démonstration proposée ici, très classique, repose sur la technique des fonctions caractéristiques.

DÉMONSTRATION DU THÉORÈME LIMITE CENTRAL : Notons φ la fonction caractéristique commune à toutes les variables $X_n - m$. Puisque

$$Y_n = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} (X_n - m),$$

et par indépendance des variables, sa fonction caractéristique ψ_n est

$$\psi_n(t) = \mathbf{E}\left[\exp\left(it\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_n - m)\right)\right] = \prod_{k=1}^{n} \mathbf{E}\exp\left(it\frac{X_n - m}{\sigma\sqrt{n}}\right) = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n.$$

On cherche maintenant la limite simple de ψ_n. Le comportement de φ à l'origine est, d'après le théorème 17.13 page 361

$$\varphi(t) = \varphi(0) + \varphi'(0)\,t + \frac{\varphi''(0)\,t^2}{2} + o(t^2)$$

avec

$$\varphi(0) = 1 \qquad \varphi'(0) = i\,\mathbf{E}(X_n - m) = 0 \qquad \text{et} \qquad \varphi''(0) = i^2\,\mathbf{E}\big((X_n - m)^2\big) = -\sigma^2,$$

c'est-à-dire

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2).$$

On en déduit (A11, page 670) que

$$\lim_{n\to\infty} \psi_n(t) = \lim_{n\to\infty} \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n = e^{-t^2/2}.$$

On reconnaît dans le membre de droite la fonction caractéristique de la loi normale centrée réduite. Le théorème de continuité de Paul Lévy (page 362) permet de conclure que S_n^* converge en loi vers la loi $\mathcal{N}(0\,;\,1)$. ▮

≙21.1

Remarque 21.6 (Du mode de convergence dans le théorème limite central) Si $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes, identiquement distribuées, d'espérance m et d'écart-type σ, on sait que S_n^* converge en loi vers $\mathcal{N}(0\,;\,1)$. Est-il possible que la convergence se fasse selon un mode plus contraignant, c'est-à-dire qu'il existe une variable aléatoire L telle que S_n^* converge presque sûrement, ou en probabilité, vers L ?

(4). Voir le théorème A.27 paragraphe A35, page 682.

La réponse est négative pour les deux types de convergence. Pour simplifier, on suppose que $\mathbf{E}(X_1) = 0$ et $\sigma_{X_1} = 1$ et on note

$$Y_n = S_n^* = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}.$$

Supposons maintenant que la suite $(Y_n)_{n\geqslant 1}$ converge en probabilité vers une certaine variable aléatoire L. Puisque la convergence en probabilité implique la convergence en loi, la variable L suit la loi normale centrée réduite. La sous-suite (Y_{2n}) converge également en probabilité vers L. Si maintenant l'on forme la quantité

$$Z_n = \frac{X_{n+1} + X_{n+2} + \cdots + X_{2n}}{\sqrt{n}},$$

on a convergence en loi de Z_n vers la loi $\mathcal{N}(0\,;\,1)$: cela se montre comme pour le théorème limite central, en vérifiant que, par indépendance des X_k, la fonction caractéristique de Z_n est

$$\varphi_{Z_n}(t) = \left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \xrightarrow[n\to\infty]{} e^{-t^2/2}.$$

Par ailleurs, on a

$$Z_n = \frac{\sqrt{2n}\,Y_{2n} - \sqrt{n}\,Y_n}{\sqrt{n}} = \sqrt{2}\,Y_{2n} - Y_n$$

et cette quantité converge en probabilité vers αL, où $\alpha = \sqrt{2} - 1$. Or la loi de αL est une loi normale $\mathcal{N}(0\,;\,\alpha^2)$, d'où la contradiction. Ainsi, il ne saurait y avoir de convergence en probabilité de la suite $(Y_n)_{n\geqslant 1}$. *A fortiori*, la convergence presque sûre ne peut pas avoir lieu non plus.

En conclusion, à l'échelle « $1/n$ », la loi des grands nombres assure que les fluctuations de S_n s'annulent et ne laissent que la certitude :

$$\frac{(S_n - \mathbf{E}S_n)}{n} \xrightarrow{\text{ps}} 0.$$

C'est à l'échelle « $1/\sqrt{n}$ » précisément que résident les fluctuations de S_n autour de sa valeur moyenne ; celles-ci sont de nature intrinsèquement probabiliste, et non déterministe. Aux échelles intermédiaires $1/n^\alpha$ pour $\alpha > 1/2$, on a convergence en loi vers 0 ; aux échelles $1/n^\alpha$ pour $\alpha < 1/2$, il ne saurait y avoir de convergence, même en loi.

≏21.2

Remarque 21.7 (Un contre-exemple) Outre l'indépendance des variables, une des hypothèses capitales du théorème limite central est celle de l'existence de moments d'ordre 1 et 2 pour les X_k ; l'existence du second empêche notamment un étalement trop important de la distribution de S_n. Si les X_i ne possèdent pas de tels moments, l'opération de centrage et de réduction ne sera pas possible. Choisissons par exemple la loi de Cauchy $\mathcal{C}(0,1)$ comme loi initiale ; alors $M_n := (X_1 + X_2 + \cdots + X_n)/n$ suit également la loi de Cauchy $\mathcal{C}(0,1)$ (cf. exercice 17.4) et aucun changement d'échelle $Y_n = \alpha_n M_n + \beta_n$ ne permettra de convergence vers une loi normale.

§ 316 Autres formes du théorème limite central (*)

Il est possible d'affaiblir les hypothèses permettant de conclure à la convergence en loi vers la loi $\mathcal{N}(0\,;\,1)$. Dans le théorème suivant, l'indépendance des variables aléatoires X_n est conservée, mais l'hypothèse d'une loi commune est remplacée par une condition sur l'étalement des X_n, qui ne doivent pas se resserrer autour de leur valeurs moyennes : la somme de leurs variances doit diverger. En même temps, on demandera que leurs moments d'ordre $2+\varepsilon$ soient bornée pour au moins une valeur $\varepsilon > 0$.

■ **Théorème 21.8 (Théorème limite central)** *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles indépendantes. On note $\sigma_n^2 = \mathbf{V}(X_n)$ et on suppose que la série $\sum \sigma_n^2$ diverge. On suppose de plus qu'il existe deux réels $\varepsilon > 0$ et M tels que $\mathbf{E}\,|X_j|^{2+\varepsilon} \leqslant M$ pour tout $j \in \mathbb{N}$. Alors la suite de terme général*

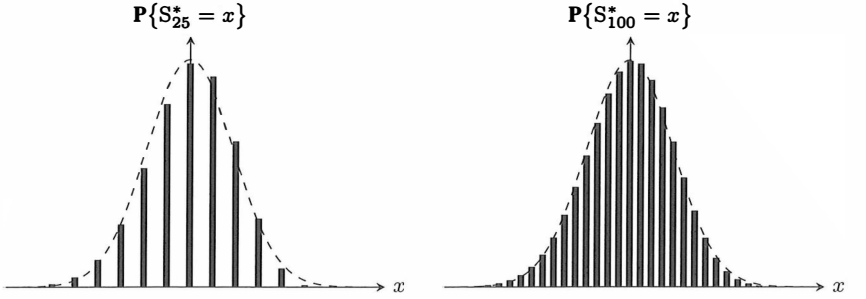
$$Y_n = S_n^* = \frac{(X_1 - \mathbf{E}X_1) + (X_2 - \mathbf{E}X_2) + \cdots + (X_n - \mathbf{E}X_n)}{\sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}}$$

converge en loi vers la loi $\mathcal{N}(0\,;\,1)$.

Remarque 21.9 D'autres conditions, plus générales, existent. Le lecteur est renvoyé, par exemple, à Shiryaev [91] pour la condition de Lindeberg et d'autres conditions moins standards.

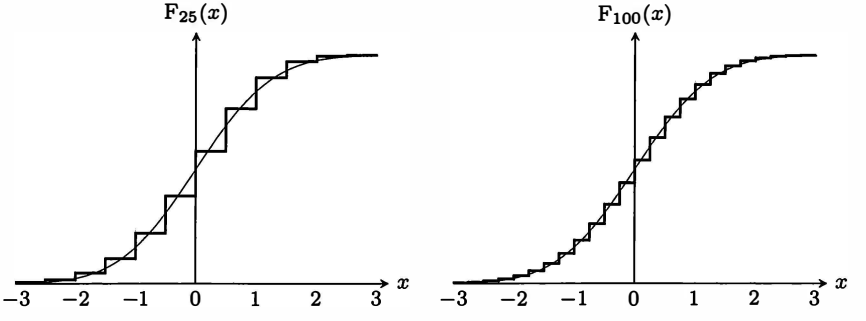
Forme locale du théorème limite central §317

Il est tentant de fournir une illustration du théorème limite central en utilisant des graphes de ce style :



(Ici, S_n est la somme de 25, puis de 100 variables indépendantes de Bernoulli de paramètre $p = 0,8$.) Ce que l'on fait apparaître sur un tel graphe, c'est un histogramme de la loi, c'est-à-dire l'équivalent discret d'une densité de probabilité. La « convergence » est manifeste et l'histogramme se rapproche formidablement bien de la densité de la loi normale, la fameuse courbe en cloche.

Cependant, le théorème limite central offre pour conclusion une convergence *en loi*, c'est-à-dire la convergence d'une suite de fonctions répartition ; ainsi, les graphes précédents n'illustrent *pas* le théorème limite central. Si l'on veut illustrer notre théorème en gardant le même exemple, il nous faut plutôt faire apparaître la fonction de répartition de S_n^* conjointement avec la fonction $\mathfrak{N}$:

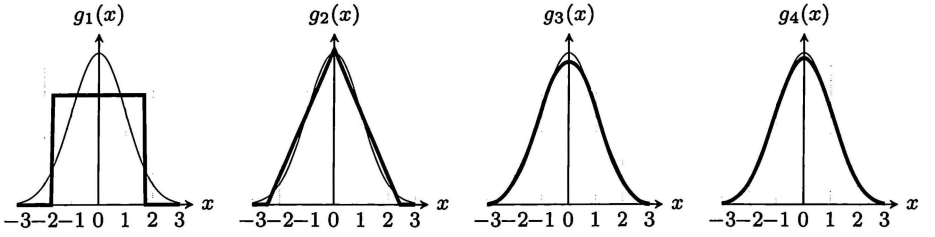


D'ailleurs, on sait que la convergence (même uniforme) de F_n vers $\mathfrak{N}$ n'implique pas, *a priori*, la convergence de $F_n' = f_n$ vers $\mathfrak{N}' = \mathfrak{n}$. Cependant, les graphes « en densité » ont l'air bien convaincant, et c'est encore plus flagrant si, au lieu de sommer des variables aléatoires discrètes, on somme des variables aléatoires absolument continues, comme le montre l'exemple ci-après.

Exemple 21.10 (Somme de lois uniformes indépendantes) Considérons une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, identiquement distribuées, de loi uniforme sur $[-1\,;+1]$. Leur espérance est $\mathbf{E}(X) = 0$, et leur écart-type est $\sigma = \sigma_X = 1/\sqrt{3}$. Formons la somme $S_n = X_1 + X_2 + \cdots + X_n$, de densité $f_n = f * f * \cdots * f$. Une récurrence simple montre que la densité de $S_n = X_1 + X_2 + \cdots + X_n$ est polynomiale par morceaux et vaut [33] :

$$f_n(x) = \frac{1}{2^n\,(n-1)!} \sum_{k\leqslant \left\lfloor \frac{n+x}{2} \right\rfloor} (-1)^k \binom{n}{k} (n+x-2k)^{n-1}$$

pour $-n \leqslant x \leqslant n$, et $f_n(x) = 0$ sinon. La densité g_n de la somme réduite $S_n^* = S/\sigma\sqrt{n}$ est représentée ci-dessous, pour les valeurs $n = 1$ à 4. En trait mince, la densité de la loi normale centrée réduite. À partir de $n = 10$, la différence entre g_n et $\mathfrak{n}$ n'est plus guère visible à cette échelle.



Aurait-on un résultat de type « théorème limite central » sur les fonctions de densité, c'est-à-dire un théorème affirmant que : $(f_n)_{n\geqslant 0}$ converge vers $\mathfrak{n}$? Eh bien oui ! c'est notre chance, sous des conditions très simples, cette dernière convergence a bien lieu.

■ **Théorème 21.11 (Théorème limite central local)** *Soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires indépendantes, identiquement distribuées, admettant une densité f bornée, une espérance m et un écart-type $\sigma > 0$. Si l'on note g_n la densité de*

$$Y_n = S_n^* = \frac{X_1 + X_2 + \cdots + X_n - nm}{\sigma\sqrt{n}},$$

alors la suite $(g_n)_{n\geqslant 1}$ converge uniformément vers la densité $\mathfrak{n}$ de la loi normale $\mathcal{N}(0\,;\,1)$:

$$\lim_{n\to\infty} g_n(x) = \mathfrak{n}(x) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}.$$

Une démonstration de ce théorème se trouve dans l'ouvrage de Rényi [87].

Remarque 21.12 Comme la densité de $S_n = X_1 + X_2 + \cdots + X_n$ est $f^{*n} = f * f * \cdots * f$, celle de S_n^* est donc

$$g_n = \sigma\sqrt{n}\, f^{*n}\big((x + mn)\,\sigma\sqrt{n}\big).$$

Ainsi, à la normalisation précédente près, convoluer une fonction avec elle-même un grand nombre de fois la rapproche d'une fonction gaussienne.

§318 **Évaluation de la vitesse de convergence**

Sous les hypothèses du théorème 21.2, les fonctions de répartition F_n de $Y_n = S_n^*$ convergent uniformément vers $\mathfrak{N}$ — c'est donc un comportement asymptotique. Les **inégalités de Berry-Esseen** permettent de majorer *non asymptotiquement* (c'est-à-dire à n fixé) la distance uniforme $\|F_n - \mathfrak{N}\|_\infty$, sous l'hypothèse supplémentaire de l'existence d'un moment d'ordre trois de $|X - \mathbf{E}X|$: si l'on pose $M_3 = \mathbf{E}\big|X - \mathbf{E}X\big|^3$, alors

$$\|F_n - \mathfrak{N}\|_\infty \leqslant c\,\frac{M_3}{\sigma^3\sqrt{n}}$$

La constante c optimale n'est pas connue, mais on peut prendre $c = 0,8$[5]. Ces inégalités permettent, en théorie, de justifier les techniques d'approximation normale.

(5). Jacod et Protter signalent [53] que la meilleure estimation connue est $c = 0,7975$.

Remarque 21.13 Ces inégalités sont, en pratique, peu utilisées, car elles sont loin d'être optimales. Pour être sûr d'obtenir une différence de 10^{-2} avec M_3 et σ de l'ordre de l'unité, il faut effectuer un nombre d'expériences de l'ordre de 10 000. En réalité, la convergence est, très souvent, bien plus rapide.

Extension aux variables vectorielles (∗) §319

Le théorème limite central admet une généralisation naturelle aux vecteurs aléatoires à valeurs dans $\mathbb{R}^d$. Comme dit à la remarque 21.5, c'est l'énoncé « $(S_n - \mathbf{E}S_n)/\sqrt{n}$ converge en loi vers $\mathcal{N}(0\,;\,\sigma^2)$ » qui se généralise convenablement, et permet de traiter aussi bien le cas non dégénéré (dét $C \neq 0$) que le cas dégénéré (pour lequel la loi normale correspondante n'admet pas de densité). La démonstration est essentiellement la même que celle du théorème 21.2.

■ **Théorème 21.14 (Théorème limite central)** *Soit $(\mathbf{X}_n)_{n\geqslant 1}$ une suite de vecteurs aléatoires à valeurs dans $E = \mathbb{R}^d$, indépendants et de même loi. On note $m = \mathbf{E}(X_1)$ leur espérance, et C leur matrice de covariance (de coefficients $C_{ij} = \mathbf{Cov}(X_i, X_j)$, donc). Si l'on note $\mathbf{Z}_n = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n$, alors la suite de vecteurs de terme général*

$$\mathbf{Y}_n = \frac{1}{\sqrt{n}}(\mathbf{Z}_n - nm)$$

converge en loi vers la loi normale $\mathcal{N}(0\,;\,C)$.

Utilité du théorème limite central §320

La principale utilisation du théorème limite central est dans la justification de l'approximation normale de certaines lois.

Notamment, lorsqu'une mesure est entachée d'erreurs X_k indépendantes les unes des autres et identiquement distribuées, la moyenne sur un nombre n grand de mesures, de ces erreurs, obéit *grosso modo* à une loi normale de paramètre $\sigma/\sqrt{n}$, où σ est l'écart-type de la distribution des X_k. Toute la force du théorème limite central est de ramener cette somme de variables aléatoires de loi *a priori* inconnue à une variable gaussienne.

Le chapitre 22 est consacré aux techniques d'approximations de lois ; plusieurs exercices illustrent, notamment, les techniques d'approximations gaussiennes.

La physique et la loi du $1/\sqrt{n}$ §321

Les fluctuations d'une somme $S_n = X_1 + X_2 + \cdots + X_n$ de variables aléatoires, indépendantes et de même loi de carré intégrable, sont de l'ordre de $\sigma_{S_n} = \sqrt{n}\,\sigma_X$, tandis que sa valeur moyenne est $\mathbf{E}(S_n) = n\,\mathbf{E}(X)$. Ainsi, le rapport des fluctuations par rapport à la quantité S_n est proportionnel à $1/\sqrt{n}$.

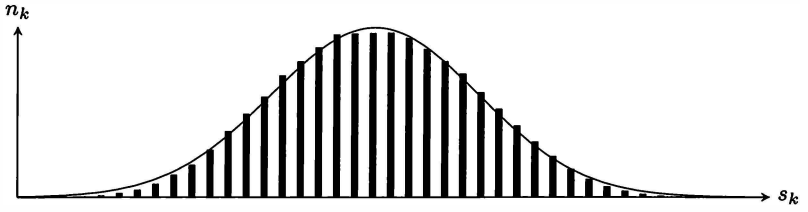
Le physicien Erwin SCHRÖDINGER[6] note [89] que de très nombreuses lois de la physique, et notamment celles qui concernent les systèmes macroscopiques (exemple : la thermodynamique d'un gaz), ne sont valables qu'en tant que lois sur les moyennes. Par exemple, la pression au sein d'un gaz est une quantité qui fluctue très légèrement au cours du temps, car la force créée par un gaz sur un paroi est le résultat d'un très grand nombre de chocs moléculaires. Le nombre de molécules d'air compris dans un certain volume est lui aussi une variable aléatoire, dont la loi doit être proche d'une loi de Poisson $\mathcal{P}(\lambda)$, donc d'une loi normale $\mathcal{N}(\lambda\,;\,\lambda)$. Si l'on s'attend à une centaine de molécules, les fluctuations observées seront de l'ordre d'une dizaine de molécules. Dans un volume plus grand, où l'on s'attend à un million de molécules, les fluctuations seront de l'ordre d'un millier de molécules. Aux échelles macroscopiques, le nombre de molécules attendu prendra

(6). Erwin SCHRÖDINGER (1887–1961), physicien autrichien émigré en Irlande, prix Nobel en 1933, célèbre pour sa fameuse *équation de Schrödinger* décrivant l'évolution de la fonction d'onde d'une particule quantique non relativiste.

des valeurs immenses, par exemple 10^{20} : les fluctuations ne seront plus alors que de 10^{10}, soit *dix milliard de fois moins.*

C'est pourquoi, poursuit Schrödinger, un organisme évolué — suffisamment évolué pour un jour développer des théories physiques — doit nécessairement être suffisamment « gros » pour qu'à son échelle, les fluctuations des lois physico-chimiques soient négligeables. Un organisme de taille microscopique observera un monde dans lequel les fluctuations sont aussi importantes que les valeurs moyennes, et il ne pourra pas espérer déduire des lois de ses observations.

Avant de conclure ce chapitre, voici un histogramme, issu d'une expérience réelle[7]. Paul a mesuré le « bruit d'avalanche » dans la jonction base-émetteur d'un transistor bipolaire. Ce type de bruit, appelé « bruit blanc », possède la propriété suivante. Si l'on échantillonne le signal $S(t)$ au cours du temps, la distribution des valeurs de S suit une loi normale. Voici l'histogramme de ces valeurs — en abscisses, les valeurs s_k prises par S et, en ordonnées, le nombre n_k de mesures dont le résultat est dans l'intervalle $[\,s_k\,;\,s_k+\delta s\,[$ —, conjointement avec une densité normale de même moyenne et de même variance. L'adéquation est excellente.



§ 322 **Au-delà du théorème limite central : la loi du logarithme itéré**

Le théorème limite central et la loi des grands nombres sont deux théorèmes permettant de quantifier la croissance de la suite de terme général

$$S_n = X_1 + X_2 + \cdots + X_n,$$

lorsque $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes et de même loi centrée et réduite, en l'observant à deux *échelles* distinctes :

- avec la loi des grands nombres, on a $S_n/n \to 0$ presque sûrement ;
- avec le théorème limite central, alors $S_n/\sqrt{n} \xrightarrow{\mathrm{L}} \mathscr{N}(0\,;\,1)$.

La suite de terme général $Y_n := S_n/\sqrt{n}$, dont la loi est asymptotiquement la loi $\mathscr{N}(0\,;\,1)$, peut prendre, accidentellement, des valeurs arbitrairement grandes. Par exemple, il est possible de montrer que

$$\mathbf{P}\left\{\limsup_{n\to\infty} Y_n = +\infty\right\} = 1,$$

c'est-à-dire que Y_n va prendre des valeurs arbitrairement grandes une infinité de fois. Si l'on diminue un tout petit peu l'échelle, d'un facteur $1/\ln n$, on ne voit plus rien[8] :

$$\mathbf{P}\left\{\limsup_{n\to\infty} \frac{Y_n}{\ln n} = +\infty\right\} = 0.$$

(7). Réalisée par un de mes anciens élèves, Paul Cottalorda, pour son TIPE, et que je remercie vivement à cette occasion.

(8). C'est une conséquence directe du théorème de Kolmogorov 20.22 page 426 avec $b_n = \sqrt{n}\ln n$, puisque la série de Bertrand $\sum 1/n\ln^2 n$ converge (A7) : on a donc $Y_n/\ln n \to 0$ presque sûrement.

Les fluctuations typiques de $(Y_n)_{n\geqslant 1}$ sont donc entre ces deux échelles 1 et $1/\ln n$. La **loi du logarithme itéré** nous apprend que la taille typique des fluctuations de $(Y_n)_{n\geqslant 1}$ est de l'ordre de $\sqrt{2\ln\ln n}$:

■ **Théorème 21.15 (Loi du logarithme itéré)** *Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes, de même loi, d'espérance nulle et de variance égale à 1. Alors*

$$\mathbf{P}\left\{\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} \cdot \frac{1}{\sqrt{2\ln\ln n}} = 1\right\} = 1.$$

En d'autres termes, avec une probabilité égale à 1, la suite de terme général $Y_n/\sqrt{2\ln\ln n}$ prendra une infinité de fois des valeurs arbitrairement proches de 1, mais un nombre fini de fois seulement des valeurs strictement supérieures à $1+\varepsilon$:

$$\forall \varepsilon > 0 \qquad \begin{cases} \mathbf{P}\left\{S_n \geqslant (1+\varepsilon)\sqrt{2n\ln\ln n}\ ;\ \text{i.s.}\right\} = 0 \\ \mathbf{P}\left\{S_n \geqslant (1-\varepsilon)\sqrt{2n\ln\ln n}\ ;\ \text{i.s.}\right\} = 1. \end{cases}$$

Par symétrie (prendre $X'_n = -X_n$), ces résultats sont encore vrais dans le domaine négatif

$$\forall \varepsilon > 0 \qquad \begin{cases} \mathbf{P}\left\{S_n \leqslant -(1+\varepsilon)\sqrt{2n\ln\ln n}\ ;\ \text{i.s.}\right\} = 0 \\ \mathbf{P}\left\{S_n \leqslant -(1-\varepsilon)\sqrt{2n\ln\ln n}\ ;\ \text{i.s.}\right\} = 1. \end{cases}$$

Géométriquement, on peut se le représenter ainsi : si l'on trace les « courbes » d'équation $y = \pm(1-\varepsilon)\sqrt{2n\ln\ln n}$, alors la suite $(S_n)_{n\geqslant 1}$ les traversera une infinité de fois. Les courbes voisines $y = \pm(1+\varepsilon)\sqrt{2n\ln\ln n}$ ne seront, elles, traversées qu'un nombre fini de fois.

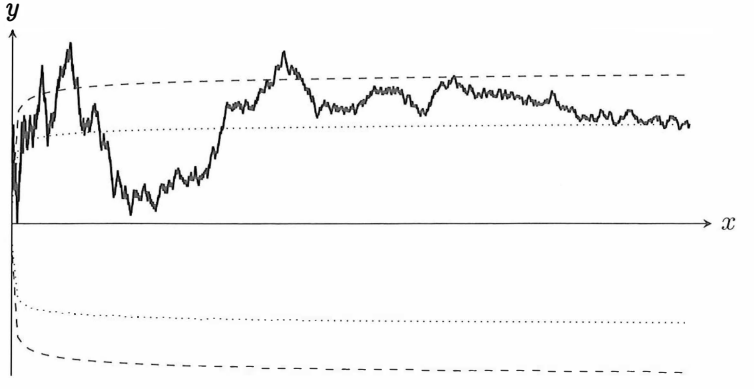


FIGURE 21.1 — *Illustration de la loi du logarithme itéré. Les deux courbes extérieures, d'équation $y = \pm(1+\varepsilon)\sqrt{\ln\ln n}$, ne seront (presque sûrement) traversées qu'un nombre fini de fois par la courbe des $Y_n = S_n/\sqrt{n}$; au contraire, les deux courbes intérieures, d'équation $y = \pm(1-\varepsilon)\sqrt{\ln\ln n}$, le seront (presque sûrement) une infinité. On notera à quel point ces courbes sont aplaties.*

La preuve de ce résultat est assez délicate, et nous l'admettons ici. Le lecteur pourra la trouver, par exemple, dans Shiryaev [91]. Dans le cas particulier d'une

succession infinie de tirages à *pile* ou *face*, une démonstration d'une grande clarté est donnée dans l'ouvrage de Lesigne [71].

EXERCICES

♦ ***Exercice 21.1 (Convergence vers la loi de Poisson)*** Redémontrer qu'une suite de lois de Bernoulli $\mathscr{B}(n\,;\,p_n)$, telle que $np_n \to \lambda$ converge vers la loi de Poisson $\mathscr{P}(\lambda)$, en utilisant la méthode des fonctions caractéristiques mise en œuvre dans la démonstration du théorème limite central.

♦ ***Exercice 21.2*** Soit $\sigma > 0$. Vérifier que, si X et Y sont deux variables aléatoires gaussiennes centrées indépendantes, d'écart type σ, alors $(X+Y)/\sqrt{2}$ suit également la loi $\mathscr{N}(0\,;\,\sigma^2)$.

On suppose que $\mathscr{L}$ est une loi de variable aléatoire continue telle que :

— si X suit la loi $\mathscr{L}$, alors X possède une espérance et une variance, et sa variance est σ^2 ;

— si X et Y sont indépendantes et de loi $\mathscr{L}$, alors $(X+Y)/\sqrt{2}$ suit la loi $\mathscr{L}$.

i) Montrer que $\mathscr{L}$ est une loi centrée.

ii) Montrer que $\mathscr{L} = \mathscr{N}(0\,;\,\sigma^2)$.

✦

Les exercices concernant les approximations normales de diverses lois sont regroupés dans le chapitre 22, page 464.

Indications

◊ *Indication 21.2 :* Montrer, par récurrence sur n, que si l'on définit des variables aléatoires X_i $(i = 1, \ldots, 2^n)$ mutuellement indépendantes et toutes de loi $\mathscr{L}$, alors $(X_1 + X_2 + \cdots + X_{2^n})/\sqrt{2^n}$ suit la loi $\mathscr{L}$.

SOLUTIONS

♦ **Solution 21.1** La fonction caractéristique de la loi $\mathscr{B}(n\,;\,p_n)$ est $\varphi_n(t) = \left[1 + p_n(e^{it} - 1)\right]^n$; si l'on écrit $p_n = \lambda/n + o(1/n)$, on peut appliquer le résultat du paragraphe A11, page 670, ce qui donne

$$\lim_{n\to\infty} \varphi_n(t) = e^{\lambda(e^{it}-1)},$$

le membre de droite étant la fonction caractéristique de la loi de Poisson de paramètre λ. Le théorème de continuité de Paul Lévy permet de conclure.

♦ **Solution 21.2**

i) Notons m l'espérance de la loi $\mathscr{L}$. Si X et Y suivent la loi $\mathscr{L}$, alors l'espérance de $(X+Y)/\sqrt{2}$ est $\sqrt{2}m$; si de plus elles sont sont indépendantes, elle vaut m. Cela n'est possible que si $m = 0$: $\mathscr{L}$ est une loi centrée.

ii) La propriété proposée en indication se montre par une récurrence immédiate. Par ailleurs, le théorème limite central montre que la suite de terme général

$$Y_n := \frac{X_1 + X_2 + \cdots + X_{2^n}}{\sigma\sqrt{2^n}}$$

(tous les termes sont des variables aléatoires centrées et réduites) converge en loi vers la loi normale $\mathscr{N}(0\,;\,1)$. Par conséquent, $\frac{X_1+X_2+\cdots+X_{2^n}}{\sqrt{2^n}}$ converge en loi vers $\mathscr{N}(0\,;\,\sigma^2)$; comme toutes les variables aléatoires de cette dernière suite suivent la loi $\mathscr{L}$, la fonction de répartition de $\mathscr{L}$ et celle de $\mathscr{N}(0\,;\,\sigma^2)$ sont identiques sur $\mathbb{R}$, et $\mathscr{L} = \mathscr{N}(0\,;\,\sigma^2)$.

Résumé du chapitre

La somme (ou la moyenne) de n variables aléatoires indépendantes et de même loi, une fois centrée et réduite, converge en loi vers la loi normale centrée réduite et ce, indépendamment du détail de la loi de ces variables (la seule contrainte étant qu'elles doivent admettre un moment d'ordre 2). C'est le théorème limite central (théorème 21.2). Il n'y a en revanche aucune convergence en probabilité ni presque sûre.

Chapitre 22

Techniques d'approximation

Le chapitre précédent montre la convergence de certaines lois vers une loi normale. Or, il peut être intéressant d'approcher une distribution par une autre — notamment parce qu'une loi normale admet une densité. Nous exposons ici les techniques d'approximations : poissonienne d'une loi binomiale, normale d'une loi binomiale ou d'une loi de Poisson.

22.1 APPROXIMATIONS POISSONIENNES

Convergence vers une loi de Poisson § 323

Rappelons l'énoncé du théorème 6.21, démontré au chapitre 6.

■ **Théorème 22.1 (Limite d'épreuves de Bernoulli)** *Soit $(p_n)_{n\geqslant 1}$ une suite de réels de $]\,0\,;\,1\,[$ telle que $\lim n p_n = \lambda$. Pour chaque entier n, on note X_n le nombre de succès obtenus dans la succession de n épreuves de Bernoulli indépendantes, et de même loi $\mathscr{B}(p_n)$; ainsi, $X_n \rightsquigarrow \mathscr{B}(n\,;\,p_n)$. Alors, pour tout entier k, $\mathbf{P}\{X_n = k\} \to \mathrm{e}^{-\lambda}\,\lambda^k/k!$, c'est-à-dire que la suite $(X_n)_{n\geqslant 1}$ converge en loi vers $\mathscr{P}(\lambda)$.*

Cet énoncé peut se généraliser en n'imposant plus aux n épreuves de Bernoulli d'avoir le même paramètre p_n. Pour chaque valeur de n, on considère des variables $X_{n,1}, X_{n,2}, \ldots, X_{n,n}$ indépendantes et on forme leur somme S_n. Les variables $X_{n,k}$ suivent une loi de Bernoulli de paramètre $p_{n,k}$.

					somme
$p_{1,1}$					q_1
$p_{2,1}$	$p_{2,2}$				q_2
$p_{3,1}$	$p_{3,2}$	$p_{3,3}$			q_3
$\cdots$	$\cdots$	$\cdots$	$\cdots$		$\cdots$
$p_{n,1}$	$p_{n,2}$	$p_{n,3}$	$\cdots$	$p_{n,n}$	q_n
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

Ce théorème, comme celui qui suit, illustre le fait qu'une distribution de Poisson peut être obtenue en collationnant un grand nombre d'événements rares (et indépendants).

■ **Théorème 22.2 (Poisson)** *Soient $p_{n,k}$ $(1 \leqslant k \leqslant n)$ des réels positifs, et $X_{n,k}$ des variables aléatoires indépendantes de lois de Bernoulli respectives $\mathscr{B}(p_{n,k})$. On note $S_n = \sum_{k=1}^{n} X_{n,k}$. On suppose d'une part que*

$$q_n = p_{n,1} + p_{n,2} + \cdots + p_{n,n} \xrightarrow[n\to\infty]{} \lambda$$

et que, de plus, la famille des $p_{n,k}$ tend uniformément vers 0 quand $n \to \infty$:

$$\lim_{n\to\infty} \max_{1\leqslant k\leqslant n} p_{n,k} = 0.$$

Alors $(S_n)_{n\geqslant 1}$ converge en loi vers la loi de Poisson de paramètre λ : $S_n \overset{L}{\Longrightarrow} \mathscr{P}(\lambda)$.

DÉMONSTRATION : Nous savons que la convergence en loi peut se montrer par la convergence simple des fonctions génératrices (théorème 19.41 page 398). Notons G_n la fonction génératrice de S_n et fixons un réel $t \in [0\,;1[$; par indépendance des variables aléatoires $X_{n,1}, X_{n,2}, \ldots, X_{n,k}$, on peut l'écrire

$$G_n(t) = \mathbf{E}(t^{S_n}) = \prod_{k=1}^{n} \mathbf{E}(t^{X_{n,k}}) = \prod_{k=1}^{n} \big(1-p_{n,k}(1-t)\big) = \exp\left\{\sum_{k=1}^{n} \ln\big(1-p_{n,k}(1-t)\big)\right\}.$$

L'idée est maintenant d'approcher chacun des logarithmes par $-p_{n,k}(1-t)$, par exemple en remarquant que $\ln(1-x)+x \sim -x^2/2$ et donc que $\ln(1-x)+x \geqslant -x^2$ sur tout un voisinage de 0 ; en fait, il est aisé de vérifier que

$$\forall x \in \left[0\,;\tfrac{1}{2}\right] \qquad -x^2 \leqslant \ln(1-x)+x \leqslant 0.$$

Il existe un entier N tel que, pour tout $n \geqslant N$, tous les $p_{n,k}$ sont compris entre 0 et 1/2, et donc

$$\big|\ln\big(1-p_{n,k}(1-t)\big) + p_{n,k}(1-t)\big| \leqslant p_{n,k}^2.$$

Par inégalité triangulaire, pour tout $n \geqslant N$,

$$\left|\sum_{k=1}^{n} \ln\big(1-p_{n,k}(1-t)\big) + p_{n,k}(1-t)\right| \leqslant \sum_{k=1}^{n} p_{n,k}^2 \leqslant \Big(\max_{1\leqslant k\leqslant n} p_{n,k}\Big) \cdot \sum_{j=1}^{n} p_{n,j}.$$

Le membre de droite est le produit d'une quantité qui tend vers 0 et d'une autre qui tend vers λ, donc tend vers 0. Ainsi,

$$\lim_{n\to\infty} \sum_{k=1}^{n} \ln\big(1-p_{n,k}(1-t)\big) = -\lim_{n\to\infty} \sum_{k=1}^{n} p_{n,k}(1-t) = -\lambda(1-t)$$

et $G_n(t) \to H(t) = e^{-\lambda(1-t)}$. On reconnaît en la fonction H la fonction génératrice de la loi de Poisson. Le théorème 19.41 permet de conclure. ∎

⇗22.1

§ 324 **Approximation poissonienne d'une loi binomiale**

Les théorèmes précédents donnent un comportement limite, lorsque n tend vers l'infini, d'une suite de lois paramétrées par n. En pratique, il arrive fréquemment que la valeur de n soit fixée, et « grande ». On peut alors faire comme si la limite était atteinte, et approcher une loi binomiale par une loi de Poisson.

En pratique, on approche la loi binomiale $\mathscr{B}(n\,;\,p)$ par la loi de Poisson $\mathscr{P}(np)$, les résultats étant corrects lorsque le paramètre p est suffisamment petit (disons $p \leqslant 0,1$), l'entier n suffisamment grand (disons $n \geqslant 30$).

⇗22.2

Exemple 22.3 (Paris répétés) Un joueur achète tous les jours un ticket de jeu à gratter ; il a une chance sur 30 de remporter un lot. On demande la probabilité qu'il gagne une fois en avril (qui comporte 30 jours) ? Qu'il gagne deux fois ? Qu'il gagne au moins une fois ? Au moins deux fois ?

Si l'on note X le nombre de succès durant le mois d'avril, X suit exactement la loi binomiale $\mathscr{B}(n\,;\,p)$, où $n = 30$ et $p = 1/30$. (On a supposé que les tirages de tickets étaient indépendants[(1)].)

(1). Récemment, des joueurs ont porté plainte contre une société de jeux car les tickets gagnants étaient répartis par rouleaux chez les débitants ; lorsqu'un ticket gagnant avait été pris du rouleau, les braves marchands ne les proposaient plus à leurs meilleurs clients, créant ainsi une inégalité entre les joueurs. Les plaignants ont eu gain de cause.

Puisque p est suffisamment petit et n suffisamment grand on peut approcher la loi de X par la loi de Poisson $\mathscr{P}(\lambda)$, où $\lambda = np = 1$. On peut alors calculer

$$\mathbf{P}\{X = 1\} \approx e^{-\lambda}\lambda \approx 0,368$$

$$\mathbf{P}\{X \geqslant 1\} = 1 - \mathbf{P}\{X = 0\} \approx 1 - e^{-\lambda} \approx 0,632$$

$$\mathbf{P}\{X = 2\} \approx e^{-\lambda}\frac{\lambda^2}{2} \approx 0,184$$

$$\mathbf{P}\{X \geqslant 2\} = 1 - \mathbf{P}\{X = 0\} - \mathbf{P}\{X = 1\} \approx 1 - e^{-\lambda} - e^{-\lambda}\lambda \approx 0,264.$$

Un calcul exact mène à

$$\mathbf{P}\{X = 1\} = np(1-p)^{n-1} \approx 0,374 \qquad \mathbf{P}\{X \geqslant 1\} = 1 - \mathbf{P}\{X = 0\} \approx 0,638$$
$$\mathbf{P}\{X = 2\} = 0,187 \qquad \mathbf{P}\{X \geqslant 2\} \approx 0,264.$$

Comme on le voit, les approximations ne sont bonnes qu'à 10^{-2} près.

Les approximations poissoniennes sont plus utiles lorsque les calculs des coefficients binomiaux sont délicats car faisant intervenir des quotients de nombres très grands (voir exercice 22.3). Cependant, les machines actuelles permettant d'effectuer des calculs de plus en plus fins, l'intérêt d'une approximation poissonienne pour un calcul numérique est assez limité. ⇗22.3

22.2 APPROXIMATION NORMALE

Approximation normale d'une loi binomiale ; correction de continuité

§ 325

Le théorème limite central justifie l'approximation normale de certaines variables aléatoires. Le premier exemple dans l'histoire des probabilités est le théorème de de Moivre-Laplace[2] :

■ **Théorème 22.4 (De Moivre-Laplace)** *Soit $p \in]0\,;\,1[$ et soit $(S_n)_{n\geqslant 1}$ une suite de variables aléatoires telles que $S_n \rightsquigarrow \mathscr{B}(n\,;\,p)$; alors, pour tout réel x,*

$$\mathbf{P}\left\{\frac{S_n - np}{\sqrt{np(1-p)}} \leqslant x\right\} \xrightarrow[n\to\infty]{} \mathfrak{N}(x).$$

Aujourd'hui, ce théorème peut être vu comme une conséquence (presque triviale) du théorème limite central: pour cela, il suffit d'interpréter S_n comme la somme de n variables indépendantes de Bernoulli de paramètre p, la variable centrée réduite correspondante étant justement

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}.$$

Ce serait évidemment oublier qu'à l'époque, la démonstration de ce théorème était un véritable tour de force, obligeant à effectuer des approximations uniformes assez fines des quantités $\binom{n}{k} p^k (1-p)^{n-k}$, conduisant tout d'abord à des estimations locales, puis à la limite intégrale[3].

Tout comme pour la loi de Poisson, on utilise souvent ce théorème pour justifier l'utilisation d'une loi normale comme loi approchée d'une autre loi. Par exemple, une loi binomiale $\mathscr{B}(n\,;\,p)$ peut être approchée par la loi normale de même espérance et de même variance, c'est-à-dire $\mathscr{N}(np\,;\,np(1-p))$. Une telle approximation n'est satisfaisante que si les deux paramètres np et $n(1-p)$ sont assez grands (disons plus grands que 15 chacun).

(2). Issu des travaux d'Abraham de Moivre (1667–1754), publiés dans *The Doctrine of Chance* (1718) ; et de Pierre Simon, marquis de Laplace (1749–1827), auteur d'une *Théorie analytique des probabilités* (1812), ouvrage d'une ampleur et d'une influence considérables.

(3). Le lecteur intéressé pourra trouver une démonstration rigoureuse dans l'ouvrage de Lesigne [71], ouvrage en tous points remarquable et dont nous conseillons très vivement la lecture !

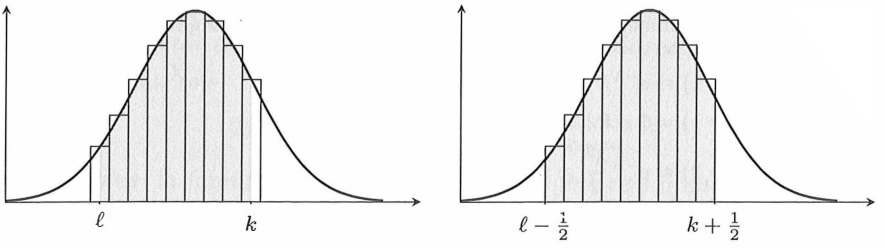


FIGURE 22.1 — *Correction de continuité. L'aire totale des rectangles représente la probabilité qu'une variable* X, *de loi binomiale* $\mathscr{B}(n\,;\,p)$ *avec* $n = 40$ *et* $p = 0,5$, *prenne ses valeurs entre* $\ell = 15$ *et* $k = 23$. À gauche : *l'aire grisée est la probabilité qu'une variable* N, *de loi normale* $\mathscr{N}(np\,;\,np(1-p))$, *prenne ses valeurs dans* $[\ell\,;\,k]$. À droite : *la probabilité que* N *soit comprise entre* $\ell - 1/2$ *et* $k + 1/2$.

Remarque 22.5 La différence entre la fonction de répartition de la loi normale (continue) et celle de la loi binomiale (présentant des sauts) ne saurait être inférieure à la moitié du plus grand saut de discontinuité, qui a lieu autour de la valeur $k = \lfloor np \rfloor$, et de valeur approximative

$$\Delta(k,n,p) \approx \frac{1}{\sqrt{2\pi\, np(1-p)}}.$$

Dans cette approximation, on peut donc écrire, si $X \rightsquigarrow \mathscr{B}(n\,;\,p)$ et pour tout k :

$$\mathbf{P}\{X \leqslant k\} = \mathbf{P}\left\{\frac{X - np}{\sqrt{np(1-p)}} \leqslant \frac{k - np}{\sqrt{np(1-p)}}\right\} \approx \mathfrak{N}\left(\frac{k - np}{\sqrt{np(1-p)}}\right). \qquad (*)$$

En réalité, *cette approximation n'est pas satisfaisante, et peut être améliorée.* On remarquera d'abord que, dans le membre de droite, la quantité calculée varie continûment avec le paramètre k, tandis que la membre de gauche effectue des sauts chaque fois que k prend une valeur entière. La différence entre le membre de gauche est le membre de droite peut être réduite en effectuant ce que l'on appelle une **correction de continuité** : au lieu d'écrire par exemple

$$\mathbf{P}\{X = k\} = \mathbf{P}\{X \leqslant k\} - \mathbf{P}\{X \leqslant k-1\} \approx \mathfrak{N}\left(\frac{k - np}{\sqrt{np(1-p)}}\right) - \mathfrak{N}\left(\frac{k - 1 - np}{\sqrt{np(1-p)}}\right),$$

on symétrise la situation en

$$\mathbf{P}\{X = k\} = \mathbf{P}\left\{X \leqslant k + \tfrac{1}{2}\right\} - \mathbf{P}\left\{X \leqslant k - \tfrac{1}{2}\right\} \approx \mathfrak{N}\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \mathfrak{N}\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

On obtient alors plus généralement les formules suivantes, lorsque k et ℓ sont des entiers :

$$\mathbf{P}\{X \leqslant k\} \approx \mathfrak{N}\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

$$\mathbf{P}\{\ell \leqslant X \leqslant k\} \approx \mathfrak{N}\left(\frac{k + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \mathfrak{N}\left(\frac{\ell - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

Remarque 22.6 Dans la dernière formule, l'événement $\{\ell \leqslant X \leqslant k\}$ correspond à $k-\ell+1$ valeurs distinctes ; dans le membre de droite, la différence des primitives correspond à l'intégration d'une densité sur intervalle de même largeur $k-\ell+1$.

On retiendra que, lorsque ℓ et k sont des entiers « proches », et notamment petits devant $\sqrt{n}$, la correction de continuité augmente singulièrement la précision de l'approximation. ⇗22.4

Exemple 22.7 On lance 50 fois une pièce équilibrée. On demande une estimation de la probabilité que le nombre de *pile* obtenus soit compris entre 20 et 30.

On note X la variable aléatoire égale au nombre de *pile*. On sait qu'alors $X \rightsquigarrow \mathscr{B}(50\ ;\ 1/2)$, mais les calculs peuvent être difficiles à mener à la calculatrice, notamment à cause des coefficients binomiaux $\binom{50}{k}$. En revanche, on peut essayer une approximation par une loi normale. La variable aléatoire X a pour espérance 25 et pour écart type $\sqrt{50/4} = \sqrt{12,5}$. Puisque $np = nq = 25 \geqslant 15$, cette approximation est raisonnable. La probabilité p cherchée, sans correction de continuité, est alors environ

$$p = \mathbf{P}\{20 \leqslant X \leqslant 30\} \approx \mathfrak{N}\left(\frac{30-25}{\sqrt{25/2}}\right) - \mathfrak{N}\left(\frac{20-25}{\sqrt{25/2}}\right)$$
$$\approx \mathfrak{N}(1{,}414) - \mathfrak{N}(-1{,}414) \approx 0{,}8426.$$

On obtient cependant un meilleur résultat si l'on effectue une *correction de continuité* :

$$p = \mathbf{P}\{19{,}5 \leqslant X \leqslant 30{,}5\} \approx \mathfrak{N}\left(\frac{30{,}5-25}{\sqrt{25/2}}\right) - \mathfrak{N}\left(\frac{19{,}5-25}{\sqrt{25/2}}\right)$$
$$\approx \mathfrak{N}(1{,}56) - \mathfrak{N}(-1{,}56) = 2\,\mathfrak{N}(1{,}56) - 1 \approx 0{,}881.$$

La valeur exacte (obtenue sans approximation, directement avec la loi binomiale), est en effet ⇗22.5

$$p \approx 0{,}881\ 079\ 547\ 4.$$

Remarque 22.8 L'approximation *locale* par la loi normale consiste à écrire

$$\mathbf{P}\{X = k\} \approx \frac{1}{\sqrt{np(1-p)}}\,\mathfrak{n}\left(\frac{k-np}{\sqrt{np(1-p)}}\right),$$

qui est cohérente avec les formules précédentes, utilisées conjointement avec l'égalité des accroissements finis. Historiquement, c'est plutôt de cette approximation que l'on a tiré la forme intégrale du théorème limite central.

Approximation normale d'une loi de Poisson

§ 326

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires *indépendantes*, suivant chacune la même loi de Poisson $\mathscr{P}(\lambda)$. Alors

$$S_n := X_1 + \ldots X_n \rightsquigarrow \mathscr{P}(n\lambda),$$

et notamment, $\mathbf{E}(S_n) = \mathbf{V}(S_n) = n\lambda$. Une application directe du théorème limite central prouve que la suite de variables aléatoires centrées réduites

$$S_n^* := \frac{S_n - n\lambda}{\sqrt{n\lambda}}$$

converge en loi vers la loi normale centrée réduite. De manière plus générale, cela démontre le théorème suivant :

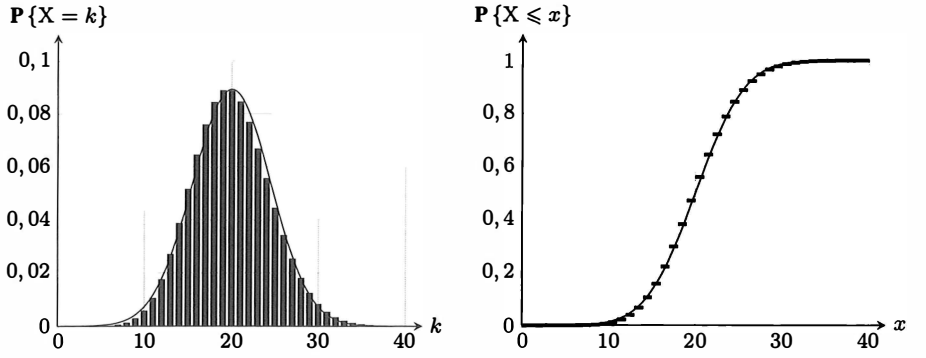


FIGURE 22.2 — À gauche : *Comparaison entre l'histogramme de la loi de Poisson* $\mathscr{P}(20)$ *et la densité normale* $\mathscr{N}(20\,;\sqrt{20})$. À droite : *Comparaison entre les deux fonctions de répartition.*

■ **Théorème 22.9 (Approximation d'une loi de Poisson)** *Soit* $\lambda > 0$. *Soit* $(\mathrm{Y}_n)_{n\geqslant 1}$ *une suite de variables aléatoires telles que* $\mathrm{Y}_n \rightsquigarrow \mathscr{P}(n\lambda)$. *Si l'on pose*

$$\mathrm{T}_n := \frac{\mathrm{Y}_n - n\lambda}{\sqrt{n\,\lambda}},$$

alors la suite $(\mathrm{T}_n)_{n\geqslant 1}$ *converge* en loi *vers la loi* $\mathscr{N}(0\,;\,1)$:

$$\forall x \in \mathbb{R} \qquad \lim_{n\to\infty} \mathbf{P}\left[\frac{\mathrm{Y}_n - n\lambda}{\sqrt{n\lambda}} \leqslant x\right] = \mathfrak{N}(x).$$

Ce théorème justifie l'approximation d'une loi de Poisson $\mathscr{P}(\mu)$ par une loi normale $\mathscr{N}(\mu\,;\,\mu)$ lorsque μ est « assez grand », c'est-à-dire que l'on écrit

$$\mathbf{P}\{\mathrm{X} \leqslant x\} \approx \mathfrak{N}\left(\frac{x-\mu}{\sqrt{\mu}}\right).$$

Comme précédemment, la correction de continuité pousse plutôt à écrire, si ℓ et k sont des entiers,

$$\mathbf{P}\{\mathrm{X} \leqslant k\} \approx \mathfrak{N}\left(\frac{k+\frac{1}{2}-\mu}{\sqrt{\mu}}\right)$$
$$\mathbf{P}\{\ell \leqslant \mathrm{X} \leqslant k\} \approx \mathfrak{N}\left(\frac{k+\frac{1}{2}-\mu}{\sqrt{\mu}}\right) - \mathfrak{N}\left(\frac{\ell-\frac{1}{2}-\mu}{\sqrt{\mu}}\right).$$

L'erreur maximale entre la fonction de répartition F_μ de la loi $\mathscr{P}(\mu)$ et la fonction de répartition de la loi normale $\mathscr{N}(\mu\,;\,\mu)$ ne saurait être inférieure à la moitié du plus grand des sauts de discontinuités de F_μ (puisque $x \mapsto \mathfrak{N}\big((x-\mu)/\sqrt{\mu}\big)$ est continue). Elle est de l'ordre de 0,05 pour $\mu = 15$, et décroît avec μ. Aussi, si l'on ne cherche à estimer des probabilités qu'à 5% près, on retiendra que l'approximation normale de la loi de Poisson exige que $\mu \geqslant 15$; si l'on veut une précision de 0,025, il faut monter à $\mu \geqslant 60$.

§ 327 **Approximation normale et inégalité de Bienaymé-Tchebychev**

Soit N^* une variable aléatoire de loi normale $\mathscr{N}(0\,;\,1)$. Pour tout $\varepsilon > 0$, on sait calculer précisément la probabilité de l'événement $\{\,|\mathrm{N}^*| > \varepsilon\,\}$, qui vaut

$$\mathbf{P}\{\,|\mathrm{N}^*| > \varepsilon\,\} = \mathfrak{N}(-\varepsilon) + \big(1 - \mathfrak{N}(\varepsilon)\big) = 2 - 2\,\mathfrak{N}(\varepsilon) \qquad (22.1)$$

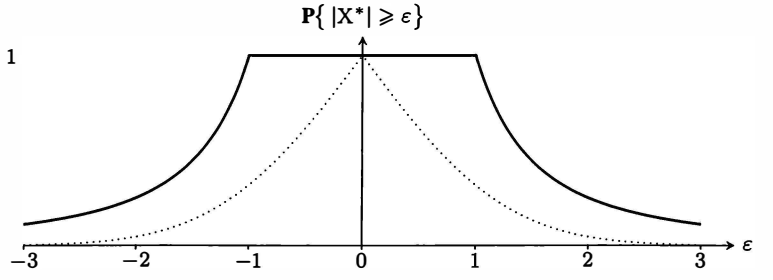
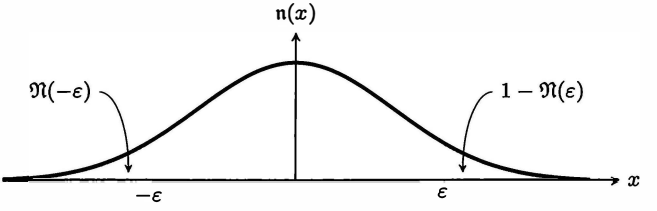


FIGURE 22.3 — *Comparaison des fonctions* $1/\varepsilon^2$ *et* $2 - 2\mathfrak{N}(\varepsilon)$ *données respectivement par Bienaymé-Tchebychev (en trait plein) et par l'approximation normale (en pointillés).*



Si X* est une variable aléatoire centrée et réduite, mais de loi par ailleurs inconnue, on ne sait pas toujours calculer la probabilité de $\{\,|X^*| > \varepsilon\}$: on peut du moins la majorer, par l'inégalité de Bienaymé-Tchebychev :

$$\mathbf{P}\{\,|X^*| > \varepsilon\} \leqslant \frac{1}{\varepsilon^2}. \tag{22.2}$$

Cependant, si X* a une loi « proche » de la loi normale, la relation (22.1) remplacera avantageusement (22.2). Sur la figure 22.3, nous représentons sur un même graphe les deux estimations précédentes, la valeur $2 - 2\mathfrak{N}(\varepsilon)$, exacte pour une variable normale, ainsi que la majoration $1/\varepsilon^2$ valable pour toute variable admettant une variance unité (plus exactement, on représente la quantité utile $\min(1, 1/\varepsilon^2)$). Pour des valeurs « petites » (disons $\varepsilon \leqslant 2$), l'inégalité de Bienaymé-Tchebychev fournit une borne médiocre mais acceptable. Ce n'est plus le cas si ε dépasse quelques unités, le rapport entre les deux fonctions étant de plusieurs ordres de grandeur (pour $\varepsilon = 3$, il est déjà de l'ordre de 500). ⇗22.6

Estimer un nombre de lancers nécessaires § 328

On peut estimer la probabilité d'un événement A en multipliant les expériences (indépendantes) et en calculant la proportion de succès : c'est le résultat de la loi des grands nombres et la base des méthodes dites de Monte-Carlo. *Combien* de tirages faut-il effectuer pour obtenir un résultat *raisonnablement correct* ? La réponse dépend de deux critères : la précision voulue, et le risque acceptable. Par exemple : on voudrait calculer la probabilité $p := \mathbf{P}(A)$ avec une précision de 10^{-2}, et un risque de 5 %, c'est-à-dire une certitude de 95 % (on accepte que, dans 5 % des cas, la mesure effectuée diffère de la valeur p de plus que 10^{-2}).

On définit la variable $X := \mathbf{1}_A$, valant 1 si l'événement A est réalisé et 0 sinon : X suit la loi de Bernoulli de paramètre p, et donc $p = \mathbf{E}(X)$. Notons $\sigma = \sigma_X = \sqrt{p(1-p)}$ son écart-type. On considère alors une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, identiquement distribuées, de même loi que X,

on note $S_n = X_1 + X_2 + \cdots + X_n$ et on définit la variable aléatoire centrée réduite

$$S_n^* = \frac{S_n - np}{\sqrt{n}\,\sigma}.$$

Si l'on admet que la variable aléatoire S_n^* suit approximativement la loi normale centrée réduite $\mathcal{N}(0\,;\,1)$, alors pour tout $\varepsilon > 0$, on a

$$\mathbf{P}\{\,|S_n^*| > \varepsilon\,\} \approx 2 - 2\mathfrak{N}(\varepsilon).$$

Si l'on accepte un risque α, on cherche ε pour que $2 - 2\mathfrak{N}(\varepsilon) = \alpha$, c'est-à-dire

$$\varepsilon = \mathfrak{N}^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Voici les valeurs de ε pour les valeurs les plus courantes de α

α	0,1 %	1 %	5 %	10 %	20 %
ε	3,29	2,576	1,96	1,645	1,28

Par ailleurs,

$$|S_n^*| > \varepsilon \iff \left|\frac{S_n}{n} - p\right| > \frac{\varepsilon\sigma}{\sqrt{n}}$$

Effectuer une « mesure » empirique de p, c'est calculer (pour la valeur de ω qui nous a été attribuée par dame Fortune) la « valeur moyenne » $S_n(\omega)/n$. L'écart à la valeur réelle celui-ci sera plus proche de p qu'une précision τ fixée à l'avance et sous le risque α si

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| \geqslant \tau\right\} \leqslant \alpha \qquad \text{pour} \qquad \tau = \frac{\varepsilon\sigma}{\sqrt{n}}.$$

On en déduit que le nombre de tirages à effectuer est $n = \varepsilon^2\sigma^2/\tau^2$.

Si, pour une raison ou pour une autre, on pense ne pas pouvoir utiliser l'approximation normale, un raisonnement semblable peut-être mené à l'aide de l'inégalité de Bienaymé-Tchebychev, le lien entre ε et le risque α étant alors simplement $\alpha = 1/\varepsilon^2$, et la probabilité $\mathbf{P}\{\,|S_n^*| > \varepsilon\,\}$ étant désormais *majorée* par α (et non plus égale à α).

En résumé :

- on fixe un risque α et une précision τ ;
- on déduit ε de α par la formule

$$\varepsilon = \begin{cases} \mathfrak{N}^{-1}(1 - \alpha/2) & \text{dans l'approximation normale,} \\ 1/\sqrt{\alpha} & \text{par Bienaymé-Tchebychev.} \end{cases}$$

- on en déduit le nombre de tirages nécessaires $n = \varepsilon^2\sigma^2/\tau^2$ (arrondi à l'entier supérieur).

Exemple 22.10 (Lancers d'un dé équilibré) On jette n fois un dé équilibré à 6 faces. Combien de lancers faut-il effectuer pour que la proportion de « 1 » soit égale à 1/6 à 10^{-2} près, avec une probabilité d'au moins 90 % ?

Le risque accepté étant de 10 %, on prend la valeur $\varepsilon = \mathfrak{N}^{-1}(0{,}95) \approx 1{,}645$. L'écart-type de la variable X, de loi de Bernoulli $\mathscr{B}(1/6)$, est $\sigma = \sqrt{p(1-p)} = \sqrt{5}/6$, et le nombre de tirages cherché est

$$n \approx \frac{5\varepsilon^2}{36\,10^{-4}} \approx 3758{,}36.$$

Ainsi, 3759 tirages suffisent pour répondre au problème.

Bien entendu, puisque nous avons effectué une approximation *non contrôlée* (de la loi binomiale par une loi normale), on prendra quelque distance avec les chiffres significatifs obtenus...

La méthode a l'air belle et solide, mais souffre d'un défaut majeur : on ne connaît pas toujours l'écart-type σ de X ! On peut alors essayer simplement de la majorer (par exemple, dans le cas où X suit une loi de Bernoulli $\mathscr{B}(p)$, son écart-type $\sigma = \sqrt{p(1-p)}$ est, quelle que soit la valeur de p, compris entre 0 et 1/2). On effectue le dernier calcul avec un majorant $\Sigma \geqslant \sigma$, et on calcule $n = \varepsilon^2\Sigma^2/\tau^2$.[4]

Exemple 22.11 (Surbooking) Un avion comporte 300 places. La probabilité qu'un passager ayant réservé une place ne se présente pas à l'embarquement est de 5%. On demande combien la compagnie aérienne peut vendre de places pour que la probabilité de voir se présenter plus de passagers que de places disponibles soit inférieure à 1% ?

On note n le nombre de places vendues et X le nombre de passagers se présentant effectivement à l'embarquement. En notant $N = 300$ et $p = 1 - q = 0{,}95$, la variable X suit la loi $\mathscr{B}(n\,;\,p) \approx \mathscr{N}(np\,;\,npq)$. On cherche donc un entier n tel que $\mathbf{P}\{X \leqslant N\} \geqslant 0{,}99$, ou encore

$$\mathbf{P}\left\{\frac{X - np}{\sqrt{npq}} \leqslant \frac{N - np}{\sqrt{npq}}\right\} \geqslant 0{,}99$$

ce qui, avec une approximation normale, mène à

$$\Pi\left(\frac{N - np}{\sqrt{npq}}\right) \geqslant 0{,}99.$$

Posons $v := \Pi^{-1}(0{,}99) \approx 2{,}33$, la condition devient alors

$$\frac{N - np}{\sqrt{npq}} \geqslant v \qquad (*)$$

Déjà, cela veut dire que le numérateur doit être positif, et donc $n \leqslant N/p \approx 315$ (c'est la valeur pour laquelle, dans l'approximation normale, on a une chance sur deux de se voir présenter plus de passagers qu'il n'y a de places : en effet le nombre de passagers suit approximativement une loi gaussienne de moyenne np, donc $N \geqslant np$ avec une probabilité 1/2).

Multiplions maintenant par le dénominateur et élevons au carré :

$$(N - np)^2 \geqslant v^2 npq$$

ou encore $an^2 + bn + c \geqslant 0$ avec $a = p^2$, $b = -p(2N + v^2 q)$ et $c = N^2$. On trouve numériquement

$$a = 0{,}902\,5 \qquad b \approx -570{,}26 \qquad c = 90\,000$$

et donc un discriminant $\Delta \approx 294{,}04 > 0$ et des racines

$$n' \approx 306{,}43 \qquad \text{et} \qquad n'' \approx 325{,}43.$$

L'inégalité cherchée ne peut être vérifiée qu'en dehors du segment $[n'\,;\,n'']$, et la condition cherchée impose également $n \leqslant N/p$, ce qui donne comme condition finale :

$$n \leqslant n' = 306{,}43.$$

Ainsi, *la compagnie aérienne peut proposer jusqu'à 306 places.*

On peut simplifier un peu le raisonnement précédent en remarquant que, dans l'inégalité (*), on connaissait déjà un ordre de grandeur de n dans le dénominateur : $n \leqslant N/p \approx 315$ (qui correspond à une probabilité 1/2 d'avoir des ennuis), et $n \geqslant N$ (qui correspond à une probabilité nulle d'avoir des ennuis). Ainsi, il ne coûte pas grand chose de remplacer, dans le dénominateur, n par sa valeur la plus restrictive $n \approx N/p$; l'inégalité (*) devient

$$N - np \geqslant v\sqrt{Nq}$$

ou encore

$$n \leqslant \frac{N - 2{,}33\sqrt{Nq}}{p} \approx 306{,}29$$

ce qui conduit toujours à proposer, au plus, 306 places.

▵22.7

(4). On essaiera toutefois d'obtenir la majoration la plus fine possible, étant donné que le carré de Σ entre en jeu.

Dernière remarque, le nombre de tirages à effectuer croît comme le *carré* de la précision cherchée. En revanche, il dépend de manière plus subtile du risque accepté. Par Bienaymé-Tchebychev, il est simplement proportionnel à $1/\alpha$ (diminuer le risque d'un facteur $1/2$ revient à doubler le nombre de tirages). Dans l'approximation normale, ε croît très lentement quand $\alpha \to 0$[5]. On peut donc exiger un risque très faible sans accroître pour cela le nombre de tirages nécessaires de manière démesurée.

§ 329 **Résumé des conditions d'approximation**

Nous résumons ici les divers résultats utiles, en ajoutant l'approximation déjà vue d'une loi hypergéométrique par une loi binomiale (exercice 6.3). Toutes les conditions pratiques sont usuelles, mais à prendre avec du recul.

Loi et paramètre		Loi approchée	Conditions pratiques	
Hypergéométrique	$\mathscr{H}(\mathrm{N}, n, p)$	$\mathscr{B}(n\,;\,p)$	$\mathrm{N} \geqslant 10\,n$	
Binomiale	$\mathscr{B}(n\,;\,p)$	$\mathscr{P}(np)$	$p \leqslant 0{,}1$	$n \geqslant 30$
Poisson	$\mathscr{P}(\mu)$	$\mathscr{N}(\mu\,;\,\mu)$	$\lambda \geqslant 15$	
Binomiale	$\mathscr{B}(n\,;\,p)$	$\mathscr{N}(np\,;\,np(1-p))$	$np \geqslant 15$	$n(1-p) \geqslant 15$

TABLE 22.1 — *Table des lois approchées.*

22.3 RÉGRESSION LINÉAIRE

§ 330 **Droite de régression**

Supposons connues deux jeux de données $x_1, x_2, \ldots, x_n$ et $y_1, y_2, \ldots, y_n$; chaque couple (x_i, y_i) de réels est supposé être le résultat d'une expérience commune (par exemple, la taille et le poids de l'individu numéro i, ou deux quantités physiques obtenues lors de la réalisation de la i-ième expérience d'une série). Si ces données sont suffisamment corrélées, on peut en tirer profit pour donner une approximation de y lorsqu'une donnée incomplète x est connue.

Une des méthodes possible est celle de la recherche de la **droite de régression de y par rapport à x**, qui est la droite

$$\mathrm{D} : \quad y = ax + b$$

minimisant les écarts quadratiques en y, c'est-à-dire celle qui rende minimale la quantité

$$\Delta(a,b) := \sum_{i=1}^{n} (y_i - ax_i - b)^2.$$

(voir figure 22.4). Les valeurs de a et b minimisant Δ sont obtenues par dérivation :

$$0 = \frac{\partial \Delta}{\partial a}(a,b) = \sum_{i=1}^{n} 2x_i(ax_i + b - y_i)$$

$$0 = \frac{\partial \Delta}{\partial b}(a,b) = \sum_{i=1}^{n} 2(ax_i + b - y_i).$$

---

[5]. À vitesse logarithmique, puisque $\alpha \underset{\varepsilon \to 0+}{\sim} \sqrt{\frac{2}{\pi}} \frac{\mathrm{e}^{-\varepsilon^2/2}}{\varepsilon}$, selon l'équivalent (A.2), annexe A.22.

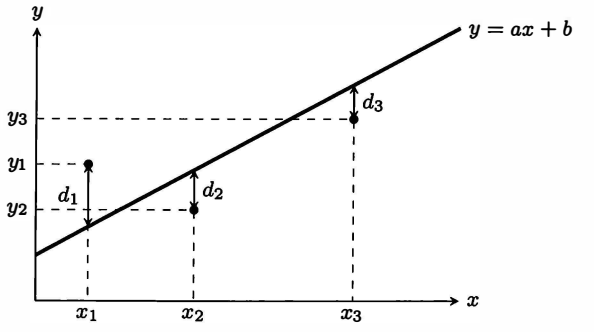


FIGURE 22.4 — *Somme des écarts quadratiques en y :* $\Delta(a,b) = \sum d_i^2$, *où l'on a posé* $d_i = y_i - (ax_i + b)$.

Introduisons les notations suivantes :

$$\langle x \rangle := \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \langle x^2 \rangle := \frac{1}{n}\sum_{i=1}^{n} x_i^2 \qquad \langle xy \rangle := \frac{1}{n}\sum_{i=1}^{n} x_i\, y_i$$

Les équations précédentes s'écrivent

$$\begin{cases} a\,\langle x^2 \rangle + b\,\langle x \rangle = \langle xy \rangle \\ a\,\langle x \rangle + b = \langle y \rangle\,, \end{cases}$$

et ce système se résout en

$$\begin{cases} a = \dfrac{\langle xy \rangle - \langle x \rangle\,\langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2} \\ b = \langle y \rangle - a\,\langle x \rangle\,. \end{cases}$$

La droite D a donc pour équation

$$\mathrm{D} \ : \quad y = \frac{\langle xy \rangle - \langle x \rangle\,\langle y \rangle}{\langle x^2 \rangle - \langle x \rangle^2}\big(x - \langle x \rangle\big) + \langle y \rangle\,.$$

Remarque 22.12 Si l'on interprète tous les points comme les réalisations possibles et équiprobables d'un couple (X, Y), les relations précédentes s'écrivent

$$\begin{cases} a = \dfrac{\mathbf{E}(\mathrm{XY}) - \mathbf{E}(\mathrm{X})\mathbf{E}(\mathrm{Y})}{\mathbf{E}(\mathrm{X}^2) - \mathbf{E}(\mathrm{X})^2} = \dfrac{\mathbf{Cov}(\mathrm{X},\mathrm{Y})}{\mathbf{V}(\mathrm{X})} \\ b = \mathbf{E}(\mathrm{Y}) - a\mathbf{E}(\mathrm{X}). \end{cases}$$

La droite D a alors pour équation

$$y - \mathbf{E}(\mathrm{Y}) = \frac{\mathbf{Cov}(\mathrm{X},\mathrm{Y})}{\mathbf{V}(\mathrm{X})}\,(x - \mathbf{E}(\mathrm{X})).$$

Elle peut servir à effectuer des prédictions de la valeur y, lorsque seule la valeur x est connue, et est souvent utilisée dans les sciences expérimentales, par exemple lorsque l'on subodore une loi affine mais que les mesures de y sont entachées d'erreurs.

Droite de régression par rapport à y §331

On définit symétriquement la **droite de régression de x par rapport à y** comme la droite d'équation

$$\mathrm{D}' \ : \quad x = a'y + b'$$

minimisant les écarts quadratiques en x, c'est-à-dire rendant minimale la quantité

$$\Delta'(a', b') := \sum_{i=1}^{n} (x_i - a' y_i - b')^2.$$

Elle a pour coefficients

$$\begin{cases} a' = \dfrac{\langle xy \rangle - \langle x \rangle \langle y \rangle}{\langle y^2 \rangle - \langle y \rangle^2} \\ b' = \langle x \rangle - a' \langle y \rangle . \end{cases}$$

Remarque 22.13 On notera que le produit aa' vaut, dans l'interprétation probabiliste des données,

$$aa' = \frac{\mathbf{Cov}(\mathrm{X}, \mathrm{Y})^2}{\mathbf{V}(\mathrm{X})\ \mathbf{V}(\mathrm{Y})} = \rho^2(\mathrm{X}, \mathrm{Y}),$$

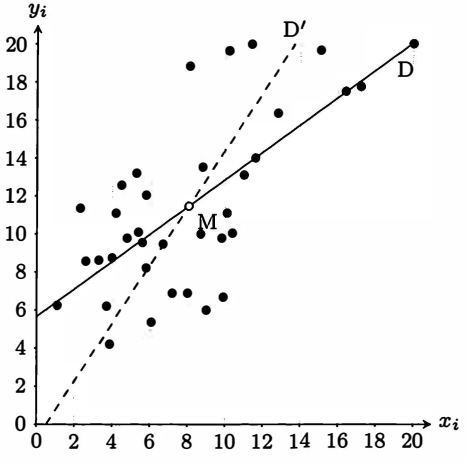
c'est-à-dire le carré du coefficient de corrélation. Géométriquement, plus les droites de régression font un angle ouvert entre elles, plus le coefficient de corrélation est faible, et donc plus les variables sont décorrélées.

Le point d'intersection des deux droites de corrélation vérifie le système

$$\begin{cases} y - \langle y \rangle = a \left[x - \langle x \rangle \right] \\ y - \langle y \rangle = \dfrac{1}{a'} \left[x - \langle x \rangle \right]. \end{cases}$$

Si $aa' = \rho^2 \neq 1$, c'est-à-dire si les points ne sont pas tous alignés, alors l'unique point d'intersection est le point M de coordonnées $\left(\langle x \rangle , \langle y \rangle \right)$.

Exemple 22.14 (Notes à deux examens) On porte sur le graphe ci-dessous les couples (x_i, y_i) où x_i est la note à l'examen d'analyse, et y_i la note à l'examen d'algèbre, pour l'élève numéro i d'une classe de 36. En trait plein, la droite D, en pointillés, la droite D′.



Les valeurs numériques des différents coefficients sont

$$a = 0,719 \qquad b = 5,663 \qquad a' = 0,658 \qquad b' = 0,520 \qquad \rho = \sqrt{aa'} = 0,688.$$

L'écartement des deux droites est caractéristique de la valeur relativement faible du coefficient de corrélation. Le point M, intersection des deux droites de corrélation, a pour coordonnées $\left(\langle x \rangle , \langle y \rangle \right)$: c'est le barycentre de tous les points noirs.

Corrélation et causalité § 332

Il est crucial d'établir une distinction nette entre les notions d'indépendance, de corrélation et de causalité. On a déjà vu à plusieurs reprises que deux variables aléatoires peuvent être *dépendantes* tout en étant *décorrélés*.

La définition suivante (qui frise le truisme) se situe en dehors du domaine mathématique. Deux événements sont *liés par une relation causale*, ou bien *en causalité*, si l'un est la cause de l'autre.

Exemple 22.15 (Accidents de la route) Les statistiques nous apprennent que les accidents de la route sont plus fréquents les samedis que les autres jours de la semaine ; il y a donc une corrélation numérique, que l'on peut expliciter ainsi. On note $\Omega = \{\text{Lu, Ma, Me, Je, Ve, Sa, Di}\}$ l'ensemble des jours de la semaine, X la variable aléatoire égale à la fonction caractéristique du samedi, Y le nombre moyen d'accident pour chaque jour. Alors $\rho(X, Y) > 0$.

Cette corrélation découle d'un lien de causalité : l'arrivée du samedi s'accompagne statistiquement d'une plus grande consommation d'alcool, qui elle-même est cause d'accidents.

Deux événements peuvent être corrélés *sans* être en lien de causalité. Un exemple est donné par Henri Broch [14] : dans les villages alsaciens, les statistiques montrent que le nombre de cigognes est fortement corrélé au nombre de naissances par an (le coefficient de corrélation est proche de 1, voir figure 22.5). Doit-on y voir la preuve d'une causalité directe (*les cigognes apportent les enfants*) ?

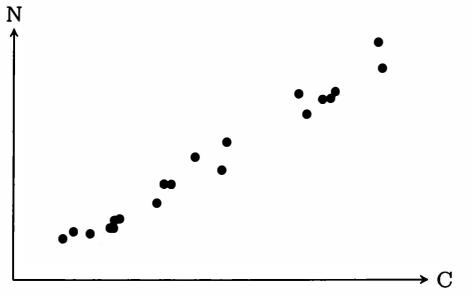


FIGURE 22.5 — *Nombre de naissances (N) et nombre de cigognes (C) dans un échantillon de villages alsaciens.*

Au risque de déflorer le mystère : la réponse est négative, il n'y a pas de tel lien de causalité. En revanche, on pourra trouver un autre type de causalité, indirecte : d'une part, plus il y a de familles, plus il y a de naissances ; de l'autre côté, plus il y a de familles, plus il y a de foyers, donc de cheminées, donc de place pour les cigognes... La causalité n'est donc pas à chercher entre les deux événements qui nous intéressent, mais depuis un troisième.

Au risque de lasser le lecteur, répétons que les notions de dépendance et de corrélation sont purement numériques : leur interprétation a beau être parfois délicate, leur établissement est purement mathématique[(6)]. En revanche, la notion de causalité sort du champ mathématique.

David RUELLE, dans son excellent ouvrage *Hasard et chaos* [84, 85], cite le cas plus étonnant d'événements *liés par une relation causale*, mais décorrélés. Considérons d'une part la position de Vénus sur le fond du ciel et, d'autre part, le

(6). En ce qui concerne l'indépendance, elle est souvent introduite comme *postulat*, comme un choix de modélisation ; rien, sinon la confrontation à l'expérience, qui doit sans cesse être renouvelée, ne permet d'établir par ailleurs cette indépendance. C'est l'origine des questions que se pose Rosencrantz (à moins que ce ne soit Guildenstern) dans la pièce *Rosencrantz et Guildenstern sont morts* [96], quand il obtient *face* une centaine de fois de suite.

temps (météorologique) un mois après. L'atmosphère terrestre est un système chaotique [94], c'est-à-dire extrêmement sensible aux conditions initiales. La position de Vénus un jour donné a donc toutes les chances de modifier profondément les conditions météo en un point donné, pour peu que l'on daigne attendre disons un mois. Il y a donc une causalité (et une dépendance) forte entre la position de Vénus et le temps qu'il fait. Pourtant, *aucune corrélation ne pourra être observée.* En effet, les autres facteurs — jour de l'année (est-ce la Saint-Médard ?), ensoleillement des derniers mois, activité magnétique solaire, position des autres planètes, pollutions en tous genres et autres cataclysmes qui nous détraquent le temps — jouent un rôle crucial dans la dynamique atmosphérique et jouent un rôle (au moins) aussi important que celui de Vénus. Si l'on se cantonne à la position de cette dernière et que l'on « moyenne » sur les autres variables, on ne verra plus aucune trace de l'influence vénusienne : causalité, donc — mais décorrélation.

22.4 ESTIMATION

§ 333 **Un exemple : moyenne empirique**

Soit X une variable aléatoire réelle de loi donnée, admettant une espérance $m = \mathbf{E}(\mathrm{X})$. On peut estimer l'espérance de X en calculant des moyennes empiriques : on répète l'expérience dont le résultat est X et on calcule la moyenne des résultats $x_1, x_2, \ldots, x_n$ obtenus.

Pour modéliser cette situation, on considère un ***n*-échantillon** de X, c'est-à-dire une suite finie $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$ de variables aléatoires mutuellement indépendantes et de même loi que X[7]. Pour une réalisation ω du hasard donné, on considère les valeurs observées $x_k = \mathrm{X}_k(\omega)$, et on en calcule la moyenne. On note alors

$$\mathrm{M}_n(\omega) = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\mathrm{X}_1(\omega) + \mathrm{X}_2(\omega) + \cdots + \mathrm{X}_n(\omega)}{n}.$$

Par linéarité de l'espérance, la relation $\mathrm{M}_n = \frac{1}{n}(\mathrm{X}_1 + \cdots + \mathrm{X}_n)$ conduit immédiatement au résultat suivant :

■ **Proposition 22.16** *La variable aléatoire* M_n *admet une espérance* $\mathbf{E}(\mathrm{M}_n) = \mathbf{E}(\mathrm{X})$.

On dit que la variable aléatoire M_n est un **estimateur sans biais** de $m = \mathbf{E}(\mathrm{X})$. Notamment, si l'on fait croître indéfiniment n, on va obtenir des estimations de précision arbitrairement grande de m, en vertu de la loi forte des grands nombres : pour *presque tout* ω, la suite de terme général $\mathrm{M}_n(\omega)$ converge vers m.

§ 334 **Un exemple : variance empirique**

Supposons toujours observées des valeurs $x_1, x_2, \ldots, x_n$; on cherche, à partir de ces valeurs, une estimation de la variance de X. Pour cela, on suppose que X admet une espérance $m = \mathbf{E}(\mathrm{X})$ et une variance $\sigma^2 = \mathbf{V}(\mathrm{X})$. Une estimation « naturelle » de la variance est de calculer la moyenne des écarts quadratiques à la moyenne *observée* ; à partir des données $x_1, x_2, \ldots, x_n$, on calcule $m_n = \frac{1}{n}(x_1 + \cdots + x_n)$ puis

$$v_n = \frac{1}{n}\sum_{k=1}^{n}(x_k - m_n)^2.$$

(7). Pour mémoire, une telle suite existe ; il suffit de remplacer l'espace Ω par Ω^n pour pouvoir la construire explicitement. Le chapitre C traite de tous ces problèmes d'existence de suites de variables aléatoires.

Du point de vue probabiliste, on considère une suite $X_1, X_2, \ldots, X_n$ de variables aléatoires indépendantes, de même loi que X, et on pose

$$M_n = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad \text{et} \qquad V_n = \frac{1}{n}\sum_{k=1}^{n}(X_k - M_n)^2.$$

Calculons l'espérance de V_n. Par linéarité et équidistribution, on a

$$\mathbf{E}(V_n) = \mathbf{E}\big[(X_1 - M_n)^2\big] = \mathbf{E}(X_1^2 - 2X_1M_n + M_n^2) = \mathbf{E}(X_1^2) - 2\mathbf{E}(X_1M_n) + \mathbf{E}(M_n^2).$$

— Le premier terme vaut $\mathbf{E}(X_1^2) = \mathbf{E}(X^2) = \mathbf{E}(X)^2 + \mathbf{V}(X) = m^2 + \sigma^2$.

— Le second se décompose en

$$\mathbf{E}(X_1M_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}(X_1X_i).$$

Dans cette somme, il faut séparer la contribution $i = 1$ des autres. En effet, pour $i = 1$, on a $\mathbf{E}(X_1^2) = m^2 + \sigma^2$ comme déjà vu ; pour $i \neq 1$, on a par indépendance $\mathbf{E}(X_1X_i) = \mathbf{E}(X_1)\,\mathbf{E}(X_i) = \mathbf{E}(X)^2 = m^2$. Ainsi,

$$\mathbf{E}(X_1M_n) = m^2 + \frac{\sigma^2}{n}.$$

— Enfin, on développe le produit

$$\mathbf{E}(M_n^2) = \frac{1}{n^2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{E}(X_iX_j)\right]$$

et on sépare les n contributions pour lesquelles $i = j$ des $n(n-1)$ contributions pour lesquelles $i \neq j$. On obtient alors

$$\mathbf{E}(M_n^2) = \frac{1}{n^2}\big[n(n-1)\,m^2 + n(m^2 + \sigma^2)\big] = m^2 + \frac{\sigma^2}{n}.$$

Au total, on obtient donc

$$\mathbf{E}(V) = \frac{n-1}{n}\,\sigma^2.$$

Surprise ! On n'obtient pas la valeur σ^2 à laquelle on pouvait s'attendre, mais une valeur différente. Si l'on pose $V^* = nV/(n-1)$, alors $\mathbf{E}(V^*) = \sigma^2$; en d'autres termes, le calcul de

$$v_n^* = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - m_n)^2$$

nous donne une « bonne » estimation de la variance de X.

■ **Théorème 22.17** *La variable*

$$V_n = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - M_n)^2$$

est un estimateur sans biais de $\mathbf{V}(X)$.

§335 **Notion d'estimateur**

On cherche, de manière générale, à estimer un paramètre θ qui gouverne la loi d'une variable aléatoire X — ou plus généralement une fonction $g(\theta)$ de ce paramètre. Un n-échantillon de X est un n-uplet $(X_1, X_2, \dots, X_n)$ de variables aléatoires indépendantes, de même loi que X. Un **estimateur** T_n de $g(\theta)$ est une fonction de $X_1, X_2, \dots, X_n$. Le **biais** de T_n est

$$b(T_n) := \mathbf{E}(T_n) - g(\theta)$$

et T_n est dit **sans biais** (ou : **non biaisé**) si son biais est nul. On dit que T_n est **asymptotiquement sans biais** si $\mathbf{E}(T_n) \to g(\theta)$ quand n tend vers l'infini.

Exemple 22.18 La variable $\frac{1}{n}\sum_{k=1}^{n}(X_k - M_n)^2$, où $M_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, est un estimateur biaisé, mais asymptotiquement sans biais, de $\mathbf{V}(X)$.

22.8

Qu'un estimateur soit sans biais, c'est sans doute agréable, mais on désire évidemment davantage : lorsque n tend vers l'infini, on voudrait que T_n converge vers $g(\theta)$. On dit que T_n est un **estimateur convergent** de $g(\theta)$ si $T_n \xrightarrow{P} g(\theta)$, ou encore

$$\mathbf{P}\{\, |T_n - g(\theta)| \geqslant \varepsilon \,\} \xrightarrow[n\to\infty]{} 0$$

pour tout $\varepsilon > 0$. Bien entendu, pour prouver cette propriété, l'inégalité de Bienaymé-Tchebychev est tout indiquée !

Exemple 22.19 Toujours avec les notations du paragraphe §333, la variance de M_n est $\mathbf{V}(X)/n$; ainsi, pour tout $\varepsilon > 0$,

$$\mathbf{P}\{\, |M_n - m| \geqslant \varepsilon \,\} \leqslant \frac{\sigma_{M_n}}{\varepsilon^2} = \frac{\sigma_X}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$

(On reconnaît la loi faible des grands nombres.) Ainsi, M_n est un estimateur sans biais et convergent de $m = \mathbf{E}(X)$.

§336 **Risque quadratique**

Pour terminer cette très courte section sur les estimateurs, introduisons la notion de risque quadratique, qui quantifie les fluctuations que peut présenter un estimateur autour de sa valeur moyenne. Si T_n est un estimateur de $g(\theta)$ et s'il admet un moment d'ordre 2, on appelle **risque quadratique** de T_n la quantité

$$r(T_n\ ;\ \theta) := \mathbf{E}\big[(T_n - g(\theta))^2\big].$$

■ **Proposition 22.20 (Risque quadratique)** *Avec les notations précédentes,*

$$r(T_n\ ;\ \theta) = \mathbf{V}(T_n) + b(T_n)^2.$$

Notamment, si T_n est un estimateur non biaisé, alors $r(T_n\ ;\ \theta) = \mathbf{V}(T_n)$.

DÉMONSTRATION : C'est un simple développement :

$$\begin{aligned} r(T_n\ ;\ \theta) &= \mathbf{E}(T_n^2) - 2\,\mathbf{E}(T_n)\,g(\theta) + g(\theta)^2 = \mathbf{V}(T_n) + \mathbf{E}(T_n)^2 - 2\,\mathbf{E}(T_n)\,g(\theta) + g(\theta)^2 \\ &= \mathbf{V}(T_n) + \big[\mathbf{E}(T_n) - g(\theta)\big]^2 = \mathbf{V}(T_n) + b(T_n)^2. \end{aligned}$$

∎

Plus le risque quadratique moyen est petit, plus l'estimateur est efficace. Entre deux estimateurs donnés, on préférera celui qui, à n fixé, présente le risque quadratique le plus petit.

Remarque 22.21 Si $r(T_n\ ;\ \theta)$ tend vers 0, alors T_n est un estimateur convergent de $g(\theta)$ (conséquence de l'inégalité de Markov).

Intervalles de confiance § 337

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles, indépendantes et de même loi, et admettant un moment d'ordre 2. Notons $m := \mathbf{E}(X_1)$ et $\sigma^2 := \mathbf{V}(X_1)$ leur moyenne et leur variance.

Lorsque l'on mesure la moyenne empirique

$$M_n := \frac{X_1 + X_2 + \cdots + X_n}{n},$$

on obtient une estimation de la quantité *inconnue* m. Fixons un **risque** α (typiquement, $\alpha = 0,05$ ou $\alpha = 0,01$). L'**intervalle de confiance** au risque α est défini comme

$$I_n := \left[M_n - \frac{\varepsilon\,\sigma}{\sqrt{n}} ; M_n + \frac{\varepsilon\,\sigma}{\sqrt{n}} \right],$$

où ε est le réel défini par

$$\mathbf{P}\{|N| > \varepsilon\} = \alpha,$$

où N est une variable de loi normale $\mathscr{N}(0\,;\,1)$, c'est-à-dire

$$\varepsilon = \mathfrak{N}^{-1}\left(1 - \frac{\alpha}{2}\right).$$

C'est un intervalle aléatoire, centré sur la variable aléatoire M_n. La probabilité que m appartienne à I_n peut également s'écrire

$$\mathbf{P}\{m \in I_n\} = \mathbf{P}\left\{M_n \in \left[m - \frac{\varepsilon\,\sigma}{\sqrt{n}} ; m + \frac{\varepsilon\,\sigma}{\sqrt{n}}\right]\right\} \xrightarrow[n\to\infty]{} \int_{-\varepsilon}^{+\varepsilon} \mathfrak{n}(t)\,dt = 1 - \alpha$$

en vertu du théorème limite central.

Remarque 22.22 On prendra garde à ne pas confondre l'intervalle de confiance avec l'**intervalle de fluctuation**

$$J_n := \left[m - \frac{\varepsilon\sigma}{\sqrt{n}} ; m + \frac{\varepsilon\sigma}{\sqrt{n}} \right],$$

qui est l'intervalle *fixe* (c'est-à-dire ne dépendant pas de ω) dans lequel se trouve la moyenne empirique M_n avec une probabilité asymptotiquement égale à $1 - \alpha$. Cet intervalle a un intérêt plus théorique, puisqu'en général c'est bien m que l'on cherche à estimer. Cependant, les énoncés $M_n \in J_n$ et $m \in I_n$ sont équivalents.

Le plus souvent, lorsque l'on cherche à estimer le paramètre m, la variance σ^2 est également inconnue[8]. On remplace donc σ par une estimation, par exemple

$$\Sigma_n^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - M_n^2.$$

La loi forte des grands nombres montre que $\frac{1}{n}\sum_{k=1}^{n} X_k^2$ converge presque sûrement vers $\mathbf{E}(X_k^2)$, tandis que M_n converge presque sûrement vers m, ce qui fait que Σ_n converge presque sûrement (et donc en probabilité) vers σ^2 d'après la loi des grands nombres.

On fait alors appel au théorème de Slutsky 19.47 page 402 : puisque la suite de terme général $\sqrt{n}(M_n - m)$ converge en loi vers la loi $N_\sigma := \mathscr{N}(0\,;\,\sigma^2)$ et que

(8). Par exemple, si les X_k suivent une même loi de Bernoulli $\mathscr{B}(p)$, le paramètre inconnu m est égal à p, et $\sigma^2 = p(1-p)$.

Σ_n^2 tend vers la constante σ^2, le couple $\left(\sqrt{n}(\mathrm{M}_n - m), \Sigma_n^2\right)$ converge en loi vers $(\mathrm{N}_\sigma, \sigma^2)$. D'après le corollaire 19.36, la suite de terme général

$$\mathrm{Z}_n := \frac{\sqrt{n}(\mathrm{M}_n - m)}{\Sigma_n}$$

converge donc en loi vers $\mathscr{N}(0\ ;\ 1)$, ce qui signifie que

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{\sqrt{n}(\mathrm{M}_n - m)}{\Sigma_n}\right| \leqslant \varepsilon\right\} = \int_{-\varepsilon}^{+\varepsilon} \mathrm{n}(t)\,\mathrm{d}t = 1 - \alpha$$

ou encore

$$\lim_{n\to\infty} \mathbf{P}\left\{m \in \left[\mathrm{M}_n - \frac{\varepsilon\,\Sigma_n}{\sqrt{n}}\ ;\ \mathrm{M}_n + \frac{\varepsilon\,\Sigma_n}{\sqrt{n}}\right]\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon}^{+\varepsilon} \mathrm{e}^{-t^2/2}\,\mathrm{d}t = 1 - \alpha.$$

Ainsi, pour n suffisamment grand, l'« intervalle de confiance » calculé avec l'estimation Σ_n de l'écart-type joue bien le rôle d'un intervalle de confiance au risque α.

EXERCICES

♦ ***Exercice 22.1*** Montrer que, sous la condition $\sum_{k=1}^{n} p_k = \lambda$ la variable $\mathrm{Y} = \sum_{k=1}^{n} \mathrm{X}_k$ où les X_k sont des variables indépendantes de lois respectives $\mathrm{X}_k \rightsquigarrow \mathscr{B}(p_k)$, est de variance maximale si les p_k sont tous égaux.

♦ ***Exercice 22.2 (Nuits agitées à la maternité)*** Dans la maternité d'une petite localité de Syldavie, le nombre de naissances par nuit suit une loi de Poisson ; on dénombre en moyenne une naissance par nuit. On suppose de plus que le nombre de naissances une nuit donnée est indépendant du nombre de naissances les autres nuits.

On note N le nombre des nuits d'une année durant lesquelles il y a au moins 4 naissances — nuits mémorables, car le personnel était complètement débordé.

i) Déterminer la loi de N, ainsi qu'une approximation de cette loi.

ii) Quel est le nombre N′ de nuits agitées ayant lieu pendant la pleine lune (le plupart des gens voient la lune « pleine » lorsqu'elle est simplement bien gibbeuse, ce qui correspond peu ou prou à 5 nuits sur 28) ?

iii) Quelle est la probabilité que le nombre N′ de nuits de pleine lune agitées soit supérieur ou égal à 3 ?

♦ ***Exercice 22.3 (Approximation poissonienne)*** Lors d'une élection municipale dans une ville comptant $n = 100\,000$ électeurs, quatre candidats se présentent au poste de maire. Parmi les 100 000 bulletins de vote, 10 000 sont pour le candidat Z. Quelle est la probabilité que, en tirant au hasard un échantillon de 1000 bulletins, exactement 100 soient pour le candidat Z ? On donnera une valeur numérique en utilisant une calculatrice par une méthode exacte, puis par une approximation poissonienne.

♦ ***Exercice 22.4 (Argenter un miroir de télescope)*** On désire vaporiser une mince couche d'argent sur une plaque de verre de surface S, modélisée par N emplacements possibles. Chaque atome d'argent, une fois vaporisé, va occuper, de manière indépendante et selon une distribution uniforme, un des N emplacements. (Cette modélisation est évidemment quelque peu simpliste, et quiconque a fabriqué lui-même un télescope le sait.)

i) On suppose que l'épaisseur moyenne de la couche d'argent est de 5 atomes. Quelle est la loi de probabilité de l'épaisseur d'argent pour un site donné ? Comment peut-on l'approcher ? Quelle est la fraction de la surface sur laquelle l'épaisseur est effectivement de 5 atomes ? Et comprise entre 4 et 6 atomes ?

ii) On suppose désormais l'épaisseur moyenne égale à 50 atomes. Quelle est la fraction de la surface sur laquelle l'épaisseur est effectivement de 50 atomes ? Et comprise entre 40 et 60 atomes ?

♦ ***Exercice 22.5 (Les étables aux cent vaches)*** Un exploitant agricole possède 100 vaches qui se répartissent au hasard entre deux étables, qui contiennent chacune n places. Quelle est la valeur minimale de n permettant à chaque vache de trouver une place, avec une probabilité supérieure à 95 % ?

♦ ***Exercice 22.6 (Une aventure de M. Maladroit)*** M. Maladroit fait tomber une boîte de 9000 dés par terre.

i) Estimer la probabilité d'obtenir un nombre de « 6 » compris entre 1450 et 1550, grâce à une approximation normale.

ii) Comparer avec la minoration donnée par l'inégalité de Bienaymé-Tchebychev.

♦ ***Exercice 22.7*** Une semaine avant les élections, un institut de sondage décide de sonder un échantillon de taille n d'électeurs, afin de déterminer la proportion p d'électeurs qui votera pour le candidat K. Il aimerait que, avec un risque de 5 %, il puisse estimer p à 0,03 près (à « trois points près » en plus ou en moins). En d'autres termes, si X_n est le nombre de personnes votant pour K parmi les n choisis, il voudrait que

$$\mathbf{P}\left\{\left|\frac{X_n}{n} - p\right| \leqslant 0,03\right\} \geqslant 0,95.$$

i) Quelle valeur minimale de n doit-il prendre ?

ii) Si, voyant le coût d'un sondage, l'institut décide de n'interroger finalement que 100 personnes, quelle précision a-t-il sur la proportion p, toujours avec un risque de 5 % ? Et avec un risque de 10 % ? Et de 50 % ?

♦ ***Exercice 22.8 (Estimation d'un paramètre)*** Soit ℓ un réel, $\ell > 0$. Soit X une variable aléatoire de loi uniforme sur $[0\,;\,\ell]$. On considère un n-échantillon $(X_1, X_2, \dots, X_n)$ de même loi que X. On voudrait, à partir de cet échantillon, estimer le paramètre ℓ. On propose deux méthodes.

i) On pose $M_n = \max(X_1, X_2, \dots, X_n)$. Montrer que $T_n = \left(\frac{n+1}{n}\right) M_n$ est un estimateur sans biais de ℓ.

ii) On pose $Y_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$. Montrer que $U_n := 2Y_n$ est un estimateur sans biais de ℓ.

iii) Quel est celui de ces deux estimateurs pour qui le risque quadratique est le plus faible ?

Indications

◊ *Indication 22.1 :* Utiliser la méthode des multiplicateurs de Lagrange (A47, page 687).

◊ *Indication 22.4 :* Selon les cas, une approximation poissonienne ou une approximation normale seront appropriées.

◊ *Indication 22.7 :* Utiliser une approximation normale de la loi binomiale, et se souvenir que $\sqrt{p(1-p)} \leqslant 1/2$ dans tous les cas.

SOLUTIONS

♦ **Solution 22.1** Notons $v(p_1, p_2, \dots, p_n) = \mathbf{V}(Y) = \sum_{k=1}^{n} p_k(1-p_k)$ et $C(p_1, p_2, \dots, p_n) = \sum_{k=1}^{n} p_k$. On cherche à maximiser la fonction v sous la contrainte linéaire $C(p_1, p_2, \dots, p_n) = \lambda$. D'après la méthode des multiplicateurs de Lagrange, il existe un scalaire μ tel que $\mathbf{grad}\, v = \mu\, \mathbf{grad}\, C$, c'est-à-dire

$$0 = \frac{\partial v}{\partial p_k} - \mu \frac{\partial C}{\partial p_k} = 1 - 2p_k - \mu$$

pour $k = 1, \dots, n$. Notamment, tous les p_k sont nécessairement égaux : $p_k = \lambda/n$. L'unique point critique est $\boldsymbol{p} = (\lambda/n, \dots, \lambda/n)$.

Enfin, la matrice hessienne de v est diagonale et manifestement définie négative, donc v admet bien un maximum local en $\boldsymbol{p}$.

♦ **Solution 22.2** Tout d'abord, le nombre de naissances par nuit suit la loi de Poisson de paramètre $\lambda = 1$ (puisque le paramètre de la loi de Poisson est égal à l'espérance de la loi). Ainsi, la probabilité qu'une nuit soit agitée est

$$p = \mathbf{P}\{\text{nb. de naissances } \geqslant 4\} = \mathrm{e}^{-1} \sum_{n=4}^{\infty} \frac{1}{k!} = \mathrm{e}^{-1}\left(\mathrm{e} - 1 - 1 - \frac{1}{2!} - \frac{1}{3!}\right) = 1 - \frac{8}{3\mathrm{e}} \approx 0,019.$$

i) Chaque nuit étant agitée avec une probabilité p, le nombre de nuits agitée dans une année non bissextile suit la loi binomiale $\mathscr{B}(N_0\,;\,p)$, où $N_0 = 365$.
Une telle loi binomiale peut être approchée par une loi de Poisson $\mathscr{P}(N_0 p)$, acceptable car « p est petit et N_0 grand ».

ii) Le nombre moyen de jours de « pleine lune apparente » est $n = (5/28)N_0 \approx 65{,}18$. Le nombre moyen de nuits de pleine lune agitées suit la loi binomiale $\mathscr{B}(n\,;\,p)$, ou encore la loi approchée de Poisson $\mathscr{P}(np)$; son espérance est donc $np \approx 1{,}24$, c'est-à-dire en moyenne un peu plus d'une nuit par an, ce qui n'est guère.

iii) Pour calculer la probabilité que $N' \geqslant 3$, utilisons l'approximation poissonienne :

$$\mathbf{P}\{N' \geqslant 3\} = 1 - \sum_{k=0}^{2} \mathbf{P}\{N' = k\} \approx 1 - e^{-np}\left(\frac{(np)^0}{0!} + \frac{(np)}{1!} + \frac{(np)^2}{2!}\right) \approx 0{,}129.$$

Ainsi, environ une année sur 7, la maternité connaît au moins trois nuits de pleine lune agitées.

De très bonne foi, des obstétriciens ou des sages-femmes pourront ainsi confirmer que, « très clairement, il y a davantage de naissance les nuits de pleine lune » : une nuit fort agitée où brillait la pleine lune se gravera durablement dans leur mémoire, alors que les nuits de pleine lune, mais calmes, disparaîtront vite dans l'oubli. En réalité, *un simple comptage des naissances, associé à un calcul des phases lunaires, montre qu'il n'y a pas de corrélation entre le nombre de naissances et la phase de la lune*[9].

♦ **Solution 22.3** Notons X le nombre de personnes votant pour le candidat Z dans l'échantillon de mille personnes ; alors $X \rightsquigarrow \mathscr{B}(1000, \frac{1}{10})$. Ainsi, un calcul exact donne

$$\mathbf{P}\{X = 100\} = \binom{1000}{100} \cdot \left(\frac{1}{10}\right)^{100} \cdot \left(1 - \frac{1}{10}\right)^{990} \approx 0{,}042.$$

Cependant, au moment de l'impression de ce livre, de nombreuses calculatrices de poche n'arrivent pas à effectuer ce calcul (produit d'un nombre très grand et d'un nombre très petit, sortant du domaine représentable par la machine). On peut dans ce cas approcher la loi réelle par une loi de Poisson de paramètre $\lambda = 1000/10 = 100$ et

$$\mathbf{P}\{X = 100\} \approx e^{-\lambda}\,\frac{\lambda^{100}}{100!} \approx 0{,}0399.$$

Si la calculette a du mal à évaluer ce quotient ($\lambda^{100} = 10^{200}$, nombre qui sort peut être du domaine accepté par la machine), on peut la programmer pour qu'elle effectue le calcul sous la forme

$$e^{-100}\,\frac{100^{100}}{100!} = \prod_{k=1}^{100} \frac{100}{k\,e},$$

qui lui permet de rester dans des domaines raisonnables.

♦ **Solution 22.4**

i) Le nombre d'atomes envoyés est 5N et chaque site est occupé, pour un atome donné, avec la probabilité 1/N ; l'épaisseur sur le site i est donc une variable aléatoire X_i suivant la loi binomiale $\mathscr{B}(5N\,;\,1/N)$. Le nombre N étant évidemment très grand, on peut approcher cette loi binomiale par une loi de Poisson de paramètre 5 : $X_i \rightsquigarrow \mathscr{P}(5)$.
Pour chaque site i, la probabilité que le nombre d'atomes soit égal à 5 est

$$\mathbf{P}\{X_i = 5\} \approx e^{-5}\,\frac{5^5}{5!} \approx 0{,}175.$$

---

(9). Plus précisément, il existe une corrélation, significative, mais *très faible*, bien qu'elle ait pris un peu d'ampleur dans les dernières décennies. La différence du nombre de naissances entre les nuits de pleine lune et les autres est totalement inobservable à l'échelle d'une maternité, même importante, mais se décèle sur les relevés du nombre de naissances, jour après jour, à l'échelle du pays. Son explication est plus subtile. Le nombre de naissances n'est pas décorrélé des jours fériés et des dimanches, car il est aisé de retarder ou d'avancer un peu une naissance pour raison médicales ; dans ce cas, il est évidemment tentant pour le médecin, si l'on doit déclencher un accouchement avant terme, de ne pas le faire un dimanche, ni le jour de Noël, ni le jour de Pâques, etc. Or, certains jours fériés du calendrier français coïncident avec des fêtes chrétiennes dont la date est calculée selon un algorithme luni-solaire (Pâques est fixé au premier dimanche suivant la première pleine lune après l'équinoxe de printemps). Ce minuscule effet explique la — très faible, et expérimentalement inobservable à l'échelle d'une maternité — corrélation entre naissances et phase lunaire ; si l'on retire les jours fériés de l'étude, la corrélation devient statistiquement compatible avec la valeur zéro.

et la fraction de surface cherchée est donc de 17, 5%. (En raisonnant comme si les épaisseurs sur chaque site étaient des variables indépendantes ; ce n'est pas tout à fait le cas.)
La probabilité que ce nombre soit compris entre 4 et 6 est

$$\mathbf{P}\{4 \leqslant X_i \leqslant 6\} = e^{-5}\left\{\frac{5^4}{4!} + \frac{5^5}{5!} + \frac{5^6}{6!}\right\} \approx 0,492.$$

Ainsi, presque la moitié de la surface a une épaisseur comprise entre 4 et 6 atomes.

ii) Chaque variable X_i suit une loi $\mathscr{P}(50)$, qui peut être approchée (théorème 22.9) par la loi normale $\mathscr{N}(50\ ;\ 50)$, c'est-à-dire que $Z_i := (X_i - 50)/\sqrt{50}$ suit approximativement la loi normale centrée réduite $\mathscr{N}(0\ ;\ 1)$. On peut calculer directement

$$\mathbf{P}\{X_i = 50\} = e^{-50}\frac{50^{50}}{50!} \approx 0,056$$

ou bien, par l'approximation normale et avec correction de continuité :

$$\begin{aligned}\mathbf{P}\{X_i = 50\} &= \mathbf{P}\{49,5 \leqslant X_i \leqslant 50,5\}\\ &= \mathbf{P}\left\{\frac{-0,5}{\sqrt{50}} \leqslant Z_i \leqslant \frac{0,5}{\sqrt{50}}\right\} \approx 2\mathfrak{N}\left(\frac{0,5}{\sqrt{50}}\right) - 1 \approx 0,056.\end{aligned}$$

Pour le second calcul, il est nettement plus simple d'utiliser l'approximation normale (avec correction de continuité)

$$\begin{aligned}\mathbf{P}\{40 \leqslant X_i \leqslant 60\} &= \mathbf{P}\{39,5 \leqslant X_i \leqslant 60,5\}\\ &= \mathbf{P}\left\{\frac{-10,5}{\sqrt{50}} \leqslant Z_i \leqslant \frac{10,5}{\sqrt{50}}\right\} \approx 2\mathfrak{N}\left(\frac{10,5}{\sqrt{50}}\right) - 1 \approx 0,862.\end{aligned}$$

(Un calcul exact de la somme des 21 termes donne $\approx 0,859$, soit une différence de $3 \cdot 10^{-3}$ seulement.)
Par rapport à la première question, on remarque qu'il y a nettement moins de chance d'avoir l'épaisseur précise (50 atomes), mais nettement plus d'être dans la fourchette de $\pm 10\,\%$.

♦ **Solution 22.5** Notons X le nombre de vaches choisissant l'étable de gauche ; alors X suit la loi $\mathscr{B}(100\ ;\ \frac{1}{2})$, donc, $\mathbf{E}(X) = 50$ et $\mathbf{V}(X) = 25$ d'où $\sigma_X = 5$. On peut approcher $\mathscr{B}(100\ ;\ \frac{1}{2})$ par $\mathscr{N}(50\ ;\ 25)$. Pour que toutes les vaches soient bien logées, il est nécessaire que X vérifie la double condition

$$100 - n \leqslant X \leqslant n.$$

La probabilité de cet événement, modulo l'approximation normale et en effectuant une correction de continuité, est

$$\begin{aligned}\mathbf{P}\{100 - n \leqslant X \leqslant n\} &= \mathbf{P}\{99,5 - n \leqslant X \leqslant n + 0,5\}\\ &\approx \mathfrak{N}\left(\frac{n - 49,5}{5}\right) - \mathfrak{N}\left(\frac{99,5 - n - 50}{5}\right) \approx 2\mathfrak{N}\left(\frac{n - 49,5}{5}\right) - 1.\end{aligned}$$

Pour que cette quantité soit supérieure à 0,95, il suffit que $\mathfrak{N}\big((n - 49,5)/5\big)$ soit supérieure à 0,975 ; un bref coup d'œil à la table de la loi normale page 720 mène à la condition

$$\frac{n - 49,5}{5} \geqslant 1,96$$

ou encore $n \geqslant 59,3$. Il *suffit* donc de prendre $n = 60$.

♦ **Solution 22.6** Le nombre de « 6 » visibles est une variable aléatoire X de loi binomiale $\mathscr{B}(N\ ;\ p)$, où $N = 9000$ et $p = 1/6$. On l'approche par la loi normale $\mathscr{N}(Np\ ;\ Np(1-p)) = \mathscr{N}(1500\ ;\ 1250)$ (une correction de continuité ne changerait presque rien ici). Notons $m = Np = 1500$ et $\sigma = \sqrt{Np(1-p)}$. On peut alors écrire

$$\mathbf{P}\{1450 \leqslant X \leqslant 1550\} = \mathbf{P}\{|X - 1500| \leqslant 50\} = \mathbf{P}\left\{\frac{|X - m|}{\sigma} \leqslant \frac{50}{\sigma}\right\} \approx 2\mathfrak{N}\left(\frac{50}{\sigma}\right) - 1 \approx 0,843.$$

Une estimation à la Bienaymé-Tchebychev donnerait

$$\mathbf{P}\left\{\frac{|X - m|}{\sigma} \geqslant \frac{50}{\sigma}\right\} \leqslant \frac{\sigma^2}{50^2} = \frac{1}{2} \qquad \text{et donc} \qquad \mathbf{P}\left\{\frac{|X - m|}{\sigma} < \frac{50}{\sigma}\right\} \geqslant \frac{1}{2},$$

ce qui comme d'habitude est une minoration peu performante.

♦ **Solution 22.7** Notons tout d'abord que $X_n \rightsquigarrow \mathscr{B}(n\,;\,p)$. On approche la loi binomiale par une loi normale : $\mathscr{B}(n\,;\,p) \approx \mathscr{N}(np\,;\,npq)$, où $q := 1-p$. Ainsi,

$$\mathbf{P}\left\{\left|\frac{X_n}{n} - p\right| \leqslant \varepsilon\right\} = \mathbf{P}\{n\varepsilon - np \leqslant X_n \leqslant n\varepsilon + np\}$$
$$\approx \mathfrak{N}\left(\frac{n\varepsilon}{\sqrt{npq}}\right) - \mathfrak{N}\left(-\frac{n\varepsilon}{\sqrt{npq}}\right) = 2\mathfrak{N}\left(\sqrt{n}\frac{\varepsilon}{\sqrt{pq}}\right) - 1.$$

(Nous avons négligé la correction de continuité ici.) Cette quantité vaut au moins 0,95 lorsque

$$2\mathfrak{N}\left(\sqrt{n}\frac{\varepsilon}{\sqrt{pq}}\right) \geqslant 1,95,$$

ou encore, en lisant la table de la loi normale, $\sqrt{n}\varepsilon/\sqrt{pq} \geqslant 1,96$, ce qui donne

$$\sqrt{n} \geqslant \sqrt{pq}\,\frac{1,96}{0,03} = \sqrt{pq}\cdot 65,3.$$

Comme $\sqrt{pq} = \sqrt{p(1-p)}$ vaut, au maximum, $1/2$, il suffit donc de prendre $\sqrt{n} \geqslant 32,7$ et donc $n \geqslant 1069$.

Si, nonobstant le résultat de ce calcul, l'institut de sondage décide de ne sonder que $n = 100$ personnes, l'incertitude sur p, *dans le pire des cas*[10], est de $3\cdot\sqrt{10} \approx 0,09$ (« 9 points ») avec un risque de 5 %. Autrement dit, on peut s'attendre à ce que, une fois sur vingt, le résultat du sondage soit grossièrement faux à plus de 9 points près[11] !

Avec un risque de 10 %, on doit résoudre

$$\mathfrak{N}\left(\sqrt{n}\,\frac{\varepsilon}{\sqrt{pq}}\right) = 0,95,$$

soit $\sqrt{n}\varepsilon/\sqrt{pq} = 1,65$, et donc $\varepsilon = 0,825$, ou encore environ 8 points d'incertitude : *si p est proche de $1/2$, alors une fois sur 10 le sondage se trompera de 8 points ou plus.*

De même, un risque de 50 % mène à $\sqrt{n}\varepsilon/\sqrt{pq} = 0,68$ et donc $\varepsilon \approx 0,42$: un sondage sur deux mène à une erreur de 4 points ou plus[12] !

Conclusion : heureusement que les instituts de sondage ne se contentent pas d'une centaine de personnes interrogées (ni même d'un millier) avant de déclarer que Monsieur X bat Monsieur Y d'un point et demi !

Remarque 22.23 Ajoutons tout de même que les instituts de sondage peuvent améliorer quelque peu leurs résultats par la *méthode des quotas*, qui a pour effet de réduire les écarts-types ; mais cette méthode présuppose la connaissance d'une partie du résultat, ce qui est, scientifiquement, pour le moins contestable — et conduit à des résultats plus ou moins cocasses à presque chaque élection.

♦ **Solution 22.8**

i) La fonction de répartition F_n de M_n est reliée à celle de X par $F_n(x) = F(x)^n$; on en déduit sa densité

$$f_n(x) = \frac{n}{\ell^n}\,x^{n-1} \qquad \text{si } 0 \leqslant x \leqslant \ell$$

et son espérance $\mathbf{E}M_n = \ell n/(n+1)$. On en déduit que $(n+1)\,M_n/n$ est un estimateur sans biais de ℓ.

ii) Par linéarité, $\mathbf{E}(2Y_n) = 2\mathbf{E}(X) = \ell$.

iii) Quelques lignes de calcul montrent que la variance de M_n est

$$\mathbf{V}(M_n) = \frac{n\ell^2}{(n+1)^2(n+2)}$$

et donc que celle de T_n est $\frac{\ell^2}{n(n+2)}$.

La variance de U_n se calcule par l'égalité de Bienaymé, et vaut $\ell^2/3n$, ce qui est nettement moins bon. L'estimateur T_n est « meilleur » que l'estimateur U_n. (On notera que les estimateurs sont ici convergents, car le risque quadratique tend vers 0, celui de T_n plus rapidement que celui de U_n.)

---

(10). C'est-à-dire lorsque pq est maximal, donc lorsque $p = 1/2$.

(11). Si p est au contraire proche de 0 ou de 1, l'écart-type est plus faible et l'incertitude est réduite. Lors d'un deuxième tour d'élections, la valeur de p est généralement proche de $1/2$, sauf accident ou république bananière.

(12). Toujours dans le cas de l'écart-type le plus défavorable.

Chapitre 23

Espérance conditionnelle

Ce qui est simple est faux, ce qui est compliqué est inutilisable.

Paul VALÉRY

L'espérance d'une variable aléatoire peut être vue comme sa meilleure approximation sous forme d'un scalaire et, plus précisément, comme sa meilleure approximation quadratique, puisque $\mathbf{E}\,|Y - m|^2$ *est minimale lorsque* $m = \mathbf{E}(Y)$[1]. *Lorsque l'on connaît un renseignement numérique limitant le domaine de* Ω *dans lequel se trouve* ω *(c'est-à-dire un renseignement de la forme «* $\omega \in A$ *»), on peut calculer,* via *la probabilité* $\mathbf{P}_A$ *conditionnée par ce renseignement, l'« espérance conditionnée »* $\mathbf{E}_A(Y)$ *de* Y. *Si ce renseignement est le résultat d'une seconde variable aléatoire* X, *l'espérance conditionnelle de* Y *est donc une quantité dépendant de* ω *au travers de la valeur de* X ; *c'est donc une variable aléatoire mesurable par rapport à* $\sigma(X)$.

Ce chapitre expose tout d'abord la théorie de l'espérance conditionnelle lorsque X *est une variable aléatoire discrète — et que tout va bien. Dans un second temps, les résultats sont généralisés au cas d'une variable quelconque — et les difficultés jaillissent, qui sont résolues de deux manières différentes : par un théorème issu de la théorie de la mesure (Radon-Nikodým), ou par une interprétation géométrique de meilleure approximation quadratique dans le cas de variables de* L^2.

23.1 LE CAS FINI OU DÉNOMBRABLE

Dans toute cette section, les variables aléatoires sont définies sur un même espace probabilisé $(\Omega, \mathcal{T}, \mathbf{P})$ et discrètes, ce qui permettra de définir de manière élémentaire la notion d'espérance conditionnelle puis de démontrer par simple calcul quelques propriétés essentielles ; on vérifiera que celles-ci sont encore vraies pour des variables quelconques, bien que les démonstrations fassent alors appel à des outils plus sophistiqués.

---

[1]. Théorème 14.38 page 325. On en rappelle l'interprétation géométrique : l'espérance de X est égale à sa projection orthogonale sur la droite des variables aléatoires déterministes (= constantes) $D = \mathrm{Vect}(\mathbf{1})$.

§338 **Espérance conditionnelle sachant une partition**

On se donne une **décomposition**, ou partition finie, $\mathfrak{D} = \{D_1, D_2, \dots, D_n\}$ de Ω, c'est-à-dire un ensemble d'événements de Ω, deux à deux disjoints et vérifiant

$$\Omega = D_1 + D_2 + \cdots + D_n.$$

On suppose de plus qu'aucun de ces événements n'est négligeable : $\mathbf{P}(D_k) > 0$ pour tout k. Les éléments D_k sont appelés les **atomes** de $\mathfrak{D}$. (De manière intuitive, $\mathfrak{D}$ représente une *connaissance partielle* de l'état du monde : pour une réalisation ω du hasard donnée, un observateur connaît lequel des événements $D_1, D_2, \dots, D_n$ est réalisé, c'est-à-dire quel est l'unique entier k tel que $\omega \in D_k$.)

Soit Y une variable discrète admettant une espérance ; l'ensemble des valeurs prises par Y est noté $Y(\Omega) = \{y_i \,;\, i \in I\}$ (l'ensemble I étant fini ou dénombrable), de telle sorte que l'on puisse écrire

$$Y = \sum_{i \in I} y_i \, \mathbf{1}_{\{Y = y_i\}}.$$

Pour une réalisation ω du hasard donnée, on sait qu'il existe un unique k^* tel que $\omega \in D_{k^*}$. La valeur moyenne de Y connaissant le « renseignement » D_{k^*}, autrement dit l'espérance de Y sachant D_{k^*}, est l'espérance de Y selon la mesure de probabilité $\mathbf{P}_{D_{k^*}}$:

$$\mathbf{E}(Y \mid D_{k^*}) = \sum_{i \in I} y_i \, \mathbf{P}\{Y = y_i \mid D_{k^*}\}. \tag{23.1}$$

Remarque 23.1 On peut, moyennant un jeu d'écriture, interpréter cette quantité comme une valeur moyenne de Y sur la partie D_{k^*} :

$$\mathbf{E}(Y \mid D_{k^*}) = \sum_{i \in I} y_i \, \frac{\mathbf{P}(\{Y = y_i\} \cap D_{k^*})}{\mathbf{P}(D_{k^*})} = \sum_{i \in I} y_i \, \frac{\mathbf{E}(\mathbf{1}_{\{Y = y_i\}} \cdot \mathbf{1}_{D_{k^*}})}{\mathbf{P}(D_{k^*})} = \frac{\mathbf{E}(Y \cdot \mathbf{1}_{D_{k^*}})}{\mathbf{P}(D_{k^*})}. \tag{23.2}$$

L'**espérance conditionnelle de Y sachant $\mathfrak{D}$** est la variable aléatoire définie par la formule (23.1), dans laquelle l'indice k^* dépend de ω ; de manière concise,

$$\mathbf{E}(Y \,\|\, \mathfrak{D}) := \sum_{k=1}^{n} \mathbf{E}(Y \mid D_k) \, \mathbf{1}_{D_k}. \tag{23.3}$$

C'est donc bien une variable aléatoire ; pour tout $\omega \in D_k$, on a

$$\mathbf{E}(Y \,\|\, \mathfrak{D})(\omega) = \sum_{i \in I} y_i \, \mathbf{P}\{Y = y_i \mid D_k\} \tag{23.4}$$

et l'inégalité $\mathbf{P}\{Y = y_i \mid D_k\} \leqslant \mathbf{P}\{Y = y_i\} / \mathbf{P}(D_k)$ montre que la série du membre de droite de l'équation (23.4) converge absolument.

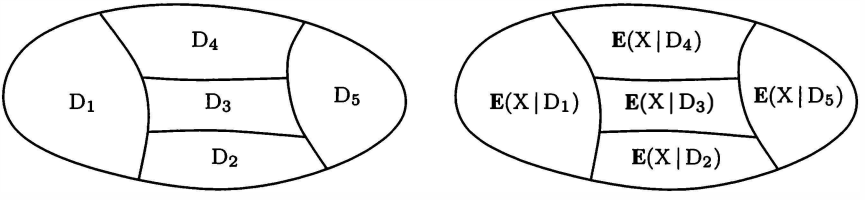


FIGURE 23.1 — *Une partition finie* $\mathfrak{D} = \{D_1, \dots, D_5\}$ *et les valeurs de* $\mathbf{E}(Y \,\|\, \mathfrak{D})$ *sur* Ω.

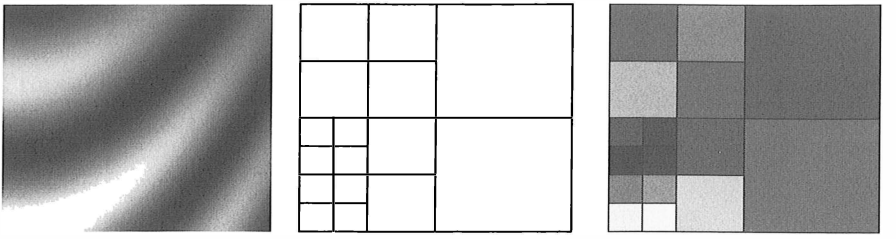


FIGURE 23.2 — (À gauche) : *les valeurs de* Y *sont représentées par des niveaux de gris.* (Au milieu) : *une partition* $\mathfrak{D}$ *de* Ω. (À droite) : *l'espérance conditionnelle de* Y *selon* $\mathfrak{D}$.

Un autre point de vue § 339

Injectons la relation (23.1) dans l'équation (23.3) et intervertissons les sommes :

$$\mathbf{E}(\mathrm{Y} \,\|\, \mathfrak{D}) := \sum_{i \in \mathrm{I}} y_i \left(\sum_{k=1}^{n} \mathbf{P}\{\mathrm{Y} = y_i \,|\, \mathrm{D}_k\}\, \mathbf{1}_{\mathrm{D}_k} \right).$$

Dans la parenthèse, nous voyons apparaître une quantité appelée **probabilité conditionnelle de $\{\mathrm{Y} = y_i\}$ sachant $\mathfrak{D}$** :

$$\mathbf{P}\{\mathrm{Y} = y_i \,\|\, \mathfrak{D}\} := \sum_{k=1}^{n} \mathbf{P}\{\mathrm{Y} = y_i \,|\, \mathrm{D}_k\}\, \mathbf{1}_{\mathrm{D}_k}.$$

L'espérance conditionnelle de Y sachant $\mathfrak{D}$ est donc

$$\mathbf{E}(\mathrm{Y} \,\|\, \mathfrak{D}) := \sum_{i \in \mathrm{I}} y_i\, \mathbf{P}\{\mathrm{Y} = y_i \,\|\, \mathfrak{D}\}.$$

On a donc deux manières de voir l'espérance conditionnelle $\mathbf{E}(\mathrm{Y} \,\|\, \mathfrak{D})$, le passage de l'une à l'autre se faisant simplement par une interversion de sommes :

selon la partition	selon les valeurs de Y
pour chaque D_k : • on considère la probabilité conditionnée $\mathbf{P}_{\mathrm{D}_k}$; • on calcule l'espérance $\mathbf{E}(\mathrm{Y} \,|\, \mathrm{D}_k)$ selon cette probabilité : $\mathbf{E}(\mathrm{Y} \,|\, \mathrm{D}_k) = \sum_{i \in \mathrm{I}} y_i\, \mathbf{P}\{\mathrm{Y} = y_i \,|\, \mathrm{D}_k\}$, puis on définit $\mathbf{E}(\mathrm{Y} \,\|\, \mathfrak{D}) = \sum_{k=1}^{n} \mathbf{E}(\mathrm{Y} \,|\, \mathrm{D}_k)\, \mathbf{1}_{\mathrm{D}_k}$	pour chaque $\mathrm{A}_i = \{\mathrm{Y} = y_i\}$: • on calcule la probabilité conditionnelle $\mathbf{P}\{\mathrm{Y} = y_i \,\|\, \mathfrak{D}\} = \sum_{k=1}^{n} \mathbf{P}\{\mathrm{Y} = y_i \,|\, \mathrm{D}_k\}\, \mathbf{1}_{\mathrm{D}_k}$, • on multiplie par la valeur y_i, puis l'on somme ces valeurs : $\mathbf{E}(\mathrm{Y} \,\|\, \mathfrak{D}) = \sum_{i \in \mathrm{I}} y_i\, \mathbf{P}\{\mathrm{Y} = y_i \,\|\, \mathfrak{D}\}$,

Ces définitions et propriétés se généralisent au cas où $\mathfrak{D} = \{\mathrm{D}_k \,;\, k \in \mathrm{K}\}$ est dénombrable, la convergence absolue des séries autorisant l'interversion des sommations :

$$\mathbf{E}(\mathrm{Y} \,\|\, \mathfrak{D}) = \sum_{k \in \mathrm{K}} \mathbf{E}(\mathrm{Y} \,|\, \mathrm{D}_k)\, \mathbf{1}_{\mathrm{D}_k}. \tag{23.5}$$

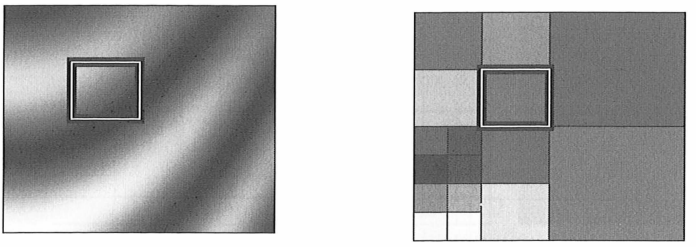


FIGURE 23.3 — *Sur un atome* D_k, *les variables* Y (à gauche) *et* $\mathrm{Z} = \mathbf{E}(\mathrm{Y} \,\|\, \mathfrak{D})$ (à droite) *ont même valeur moyenne.*

§340 **La propriété fondamentale de l'espérance conditionnelle**

Par concision, notons $\mathrm{Z} := \mathbf{E}(\mathrm{Y} \,\|\, \mathfrak{D})$; cette variable prend, sur tout atome D_k, la valeur constante

$$z_k = \mathbf{E}(\mathrm{Y} \,|\, \mathrm{D}_k) = \frac{\mathbf{E}(\mathrm{Y}\,\mathbf{1}_{\mathrm{D}_k})}{\mathbf{P}(\mathrm{D}_k)}, \tag{23.6}$$

comme vu à l'équation (23.2), ce que l'on interprète ainsi : Z *et* Y *prennent la même valeur moyenne sur chaque* D_k (voir figure 23.3). Par ailleurs, l'équation (23.3) implique que, pour tout k, on a $\mathrm{Z}\,\mathbf{1}_{\mathrm{D}_k} = \mathbf{E}(\mathrm{Y} \,|\, \mathrm{D}_k)\,\mathbf{1}_{\mathrm{D}_k}$ donc, en prenant l'espérance dans chaque membre :

$$\mathbf{E}(\mathrm{Z}\,\mathbf{1}_{\mathrm{D}_k}) = \mathbf{E}(\mathrm{Y} \,|\, \mathrm{D}_k)\,\mathbf{P}(\mathrm{D}_k) = \mathbf{E}(\mathrm{Y}\,\mathbf{1}_{\mathrm{D}_k}).$$

Maintenant, tout élément G de la tribu $\mathfrak{G} := \sigma(\mathfrak{D})$ est union dénombrable disjointe d'éléments D_k, donc par σ-additivité

$$\forall \mathrm{G} \in \mathfrak{G} \qquad \mathbf{E}(\mathrm{Z}\,\mathbf{1}_{\mathrm{G}}) = \mathbf{E}(\mathrm{Y}\,\mathbf{1}_{\mathrm{G}}),$$

ce que l'on peut écrire :

$$\forall \mathrm{G} \in \mathfrak{G} \qquad \int_{\mathrm{G}} \mathrm{Z}\,\mathrm{d}\mathbf{P} = \int_{\mathrm{G}} \mathrm{Y}\,\mathrm{d}\mathbf{P}. \tag{23.7}$$

On a l'impression d'avoir compliqué les affaires, mais c'est bien cette égalité fondamentale sur les éléments de $\mathfrak{G}$ qui permettra de définir l'espérance conditionnelle dans le cas général, quand les atomes de $\mathfrak{G}$ seront (éventuellement) *tous* de probabilité nulle.

§341 **Propriétés élémentaires**

Si certains éléments D_k de la partition sont de probabilité nulle, la formule (23.5) n'a plus guère de sens et on se contente d'une définition plus faible :

$$\forall \omega \in \Omega \qquad \mathbf{E}(\mathrm{Y} \,\|\, \mathfrak{D})(\omega) = \begin{cases} \mathbf{E}(\mathrm{Y} \,|\, \mathrm{D}_k) & \text{si } \omega \in \mathrm{D}_k \quad \text{et} \quad \mathbf{P}(\mathrm{D}_k) > 0, \\ \textit{une valeur arbitraire} & \text{si } \omega \in \mathrm{D}_k \quad \text{et} \quad \mathbf{P}(\mathrm{D}_k) = 0. \end{cases}$$

Les valeurs arbitraires ne sont définies que sur un ensemble N négligeable, et n'interviennent donc pas, par exemple, si l'on calcule l'espérance de la variable $\mathbf{E}(\mathrm{Y} \,\|\, \mathfrak{D})$. Choisir des valeurs particulières sur N, c'est se fixer une **version** de l'espérance conditionnelle. L'espérance conditionnelle elle-même est plutôt la classe de toutes les versions possibles mais, par abus de langage, on ne distinguera pas toujours l'une de l'autre ; tout au plus ajoutera-t-on un « presque sûrement » lorsque l'on voudra rappeler cette indétermination.

On déduit des propriétés de l'espérance que :

■ **Proposition 23.2** *Si* X *et* Y *sont des variables aléatoires discrètes admettant une espérance et si* $\mathfrak{D} = \{D_k \ ; \ k \in \mathbb{N}\}$ *est une partition finie ou dénombrable de* Ω, *alors*

i) ***linéarité :*** $\mathbf{E}(\alpha X + \beta Y \,\|\, \mathfrak{D}) = \alpha\, \mathbf{E}(X \,\|\, \mathfrak{D}) + \beta\, \mathbf{E}(Y \,\|\, \mathfrak{D})$;

ii) $\mathbf{E}(X \,\|\, \Omega) = \mathbf{E}(X)$ *(variable aléatoire constante) ;*

iii) Si X *est (presque sûrement) constante égale à* C, *alors* $\mathbf{E}(X \,\|\, \mathfrak{D}) = C$ *p.s. ;*

Par « moyenne des moyennes », on déduit le théorème suivant :

■ **Théorème 23.3** *Si* Y *est une variable aléatoire discrète admettant une espérance et si* $\mathfrak{D} = \{D_k \ ; \ k \in \mathbb{N}\}$ *est une partition dénombrable de* Ω, *alors la variable aléatoire* $\mathbf{E}(Y \,\|\, \mathfrak{D})$ *admet une espérance et*

$$\mathbf{E}\big(\mathbf{E}(Y \,\|\, \mathfrak{D})\big) = \mathbf{E}(Y).$$

DÉMONSTRATION : Pour fixer les notations, on suppose que Y prend des valeurs $\{y_i \ ; \ i \in \mathbb{N}\}$. On rappelle que la variable aléatoire $Z = \mathbf{E}(Y \,\|\, \mathfrak{D})$ prend, sur chaque D_k, la valeur

$$\mathbf{E}(Y \,|\, D_k) = \sum_{i=0}^{\infty} y_i\, \mathbf{P}\{Y = y_i \,|\, D_k\}.$$

Supposons momentanément qu'elle admet une espérance ; celle-ci est alors somme de la série de terme général $\alpha_k = \mathbf{P}(D_k) \cdot \sum_i y_i\, \mathbf{P}\{Y = y_i \,|\, D_k\}$, c'est-à-dire que

$$\begin{aligned}
\mathbf{E}(Z) = \sum_{k=0}^{\infty} \alpha_k &= \sum_{k=0}^{\infty} \sum_{i=0}^{\infty} y_i\ \mathbf{P}\big(\{Y = y_i\} \cap D_k\big) \\
&= \sum_{i=0}^{\infty} y_i \sum_{k=0}^{\infty} \mathbf{P}\big(\{Y = y_i\} \cap D_k\big) \qquad \textit{(interversion de sommes)} \\
&= \sum_{i=0}^{\infty} y_i\, \mathbf{P}\left(\sum_{k=0}^{\infty} \{Y = y_i\} \cap D_k\right) \qquad \textit{(union disjointe)} \\
&= \sum_{i=0}^{\infty} y_i\, \mathbf{P}\{Y = y_i\} = \mathbf{E}(Y).
\end{aligned}$$

Pour justifier la convergence, on fait appel au théorème de Fubini. Étant donné que la série de la dernière ligne converge absolument (par définition de l'espérance de Y), la famille

$$\Big\{y_i\, \mathbf{P}\big(\{Y = y_i\} \cap D_k\big) \ ; \ i, k \in \mathbb{N}\Big\}$$

est sommable, l'interversion est licite et la série définissant $\mathbf{E}(Z)$ converge absolument. ∎

Remarque 23.4 Dans ce qui précède, nous avons toujours supposé que Y admettait une espérance. Cependant, il peut tout à fait arriver que Y n'admette pas d'espérance, mais qu'elle admette une espérance conditionnelle $\mathbf{E}(Y \,\|\, \mathfrak{D})$ selon une partition (cette variable aléatoire, en revanche, n'admettra pas d'espérance).

On peut construire un exemple simple en considérant une variable aléatoire Y prenant ses valeurs dans $\mathbb{N}^*$ et telle que $\mathbf{P}\{Y = n\} = \alpha/n^2$ pour tout entier $n \geqslant 1$; la constante α valant $6/\pi^2$. On choisit pour partition le système complet d'événement $\mathfrak{D} = \mathfrak{D}_Y$ induit par Y. Bien que Y n'admette pas d'espérance, elle admet une espérance conditionnelle selon $\mathfrak{D}_Y$ et cette espérance conditionnelle n'est rien d'autre que Y elle-même : $\mathbf{E}(Y \,\|\, \mathfrak{D}_Y) = Y$.

Dans la suite, on continuera de supposer que Y admet une espérance ; ainsi, dans tous les cas, les espérances conditionnelles étudiées seront bien définies.

§342 **Espérance conditionnelle sachant une variable aléatoire discrète**
Soit X une variable aléatoire discrète ; on note $\mathfrak{D}_X = \{\{X = x_k\} \ ; \ k \in K\}$ le système complet d'événements engendré par X. Si Y est une variable aléatoire discrète, on définit l'**espérance conditionnelle de Y sachant X** par

$$\mathbf{E}(Y \,\|\, X) := \mathbf{E}(Y \,\|\, \mathfrak{D}_X) = \sum_{k \in K} \mathbf{E}(Y \,|\, X = x_k)\, \mathbf{1}_{\{X = x_k\}},$$

où, chaque fois que $\mathbf{P}\{X = x_k\} > 0$, on a

$$\mathbf{E}(Y \,|\, X = x_k) := \frac{\mathbf{E}(Y \cdot \mathbf{1}_{\{X = x_k\}})}{\mathbf{P}\{X = x_k\}}.$$

■ **Théorème 23.5** *Si* X *et* Y *sont indépendantes, alors*

$$\mathbf{E}(Y \,\|\, X) = \mathbf{E}(Y) \quad p.s.$$

DÉMONSTRATION : Par définition, Y est indépendante de $\sigma(X)$ et donc de $\mathbf{1}_{D_k}$ pour tout $k \in K$. Notamment, si D_k est non négligeable,

$$\mathbf{E}(Y \,|\, D_k) = \frac{\mathbf{E}(Y \cdot \mathbf{1}_{D_k})}{\mathbf{P}(D_k)} = \frac{\mathbf{E}(Y) \cdot \mathbf{E}(\mathbf{1}_{D_k})}{\mathbf{P}(D_k)} = \mathbf{E}(Y).$$

Ainsi, $\sum \mathbf{E}(Y \,|\, D_k)\, \mathbf{1}_{D_k} = \mathbf{E}(Y)$ sur l'union des D_k non négligeables, qui est de probabilité 1. ∎

Ce résultat est intuitif si l'on se souvient de l'interprétation de l'indépendance de Y par rapport à X : sur chaque événement lié à X, et notamment sur chaque $D_k = \{X = x_k\}$, la variable Y doit exhiber, à un facteur $\mathbf{P}(D_k)$ près, la même répartition que sur Ω tout entier. Graphiquement, sur une partition finie, on peut imaginer Y comme à la figure 23.4.

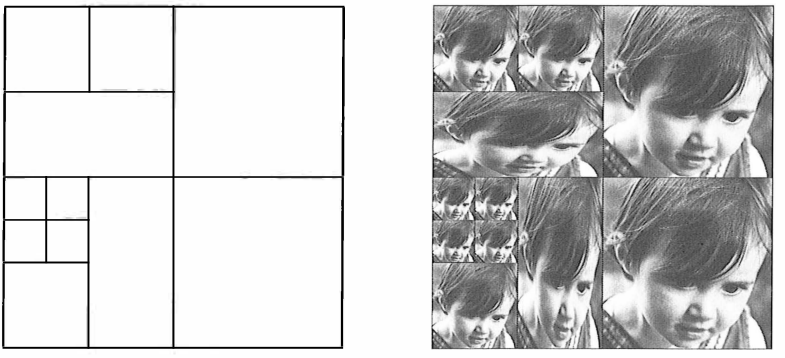


FIGURE 23.4 — (À gauche) *: une partition finie* $\mathfrak{D}$. (À droite) *: allure d'une variable* Y *indépendante de* $\mathfrak{D}$.

Exemple 23.6 (Loi de Bernoulli conditionnée par une loi de Poisson) On se donne une variable aléatoire réelle X suivant une loi de Poisson de paramètre λ. On note, pour tout k entier, $D_k := \{X = k\}$, événement de probabilité $\mathbf{P}(D_k) = e^{-\lambda} \lambda^k / k!$. On définit une variable aléatoire Y à valeurs entières par sa loi conditionnée par X : pour tout entier k,

$$\mathbf{P}\{Y = k \,|\, X = k\} = p \qquad \text{et} \qquad \mathbf{P}\{Y = 0 \,|\, X = k\} = 1 - p.$$

Alors

$$\mathbf{E}(Y \,|\, D_k) = \sum_{i=0}^{\infty} i\, \mathbf{P}\{Y = i \,|\, D_k\} = k\, \mathbf{P}\{Y = k \,|\, D_k\} = kp,$$

tous les autres termes étant nuls. Ainsi, $\mathbf{E}(Y \,\|\, X) = \sum kp\, \mathbf{1}_{\{X = k\}} = pX$.

Fonctions mesurables selon une partition dénombrable § 343

Soit $\mathfrak{D} = \{D_k \,;\, k \in K\}$ une partition finie ou dénombrable. Une variable aléatoire réelle Y est mesurable selon $\mathfrak{D}$ (on dit encore : **$\mathfrak{D}$-mesurable**) si et seulement si $Y^{-1}(B)$ est élément de la tribu engendrée $\sigma(\mathfrak{D})$, pour tout borélien $B \in \mathfrak{B}(\mathbb{R})$. En particulier, les événements $\{Y = y\}$ sont tous dans $\sigma(\mathfrak{D})$: ce sont donc des unions de D_k, ce qui permet d'écrire

$$Y = \sum_{k \in K} y_k \mathbf{1}_{D_k}$$

(où cette fois-ci les y_k ne sont pas nécessairement distincts). En d'autres termes :

■ **Proposition 23.7** *Soit $\mathfrak{D}$ une partition finie ou dénombrable. Une variable aléatoire Y est $\mathfrak{D}$-mesurable si et seulement si elle est constante sur les atomes de $\mathfrak{D}$.*

On vérifie alors immédiatement que :

■ **Théorème 23.8** *Si X, Y sont des variables aléatoires, si Y est $\mathfrak{D}$-mesurable et si X admet une espérance conditionnelle selon $\mathfrak{D}$, alors XY admet une espérance conditionnelle et*

$$\mathbf{E}(XY \,\|\, \mathfrak{D}) = Y\, \mathbf{E}(X \,\|\, \mathfrak{D}).$$

Notamment,

$$\mathbf{E}(Y \,\|\, \mathfrak{D}) = Y.$$

Enfin, considérons une partition $\mathfrak{D}'$ plus fine que $\mathfrak{D}$, c'est-à-dire que chaque élément de $\mathfrak{D}$ peut s'écrire comme union d'éléments de $\mathfrak{D}'$; cela revient à dire que $\sigma(\mathfrak{D}) \subset \sigma(\mathfrak{D}')$, et l'on écrit $\mathfrak{D}' \preccurlyeq \mathfrak{D}$ (attention au sens !).

■ **Proposition 23.9 (« Tower property »)** *Si $\mathfrak{D}$ et $\mathfrak{D}'$ sont des partitions de Ω et si $\mathfrak{D}' \preccurlyeq \mathfrak{D}$, alors*

$$\mathbf{E}\big(\mathbf{E}(Y \,\|\, \mathfrak{D}') \,\|\, \mathfrak{D}\big) = \mathbf{E}(Y \,\|\, \mathfrak{D}).$$

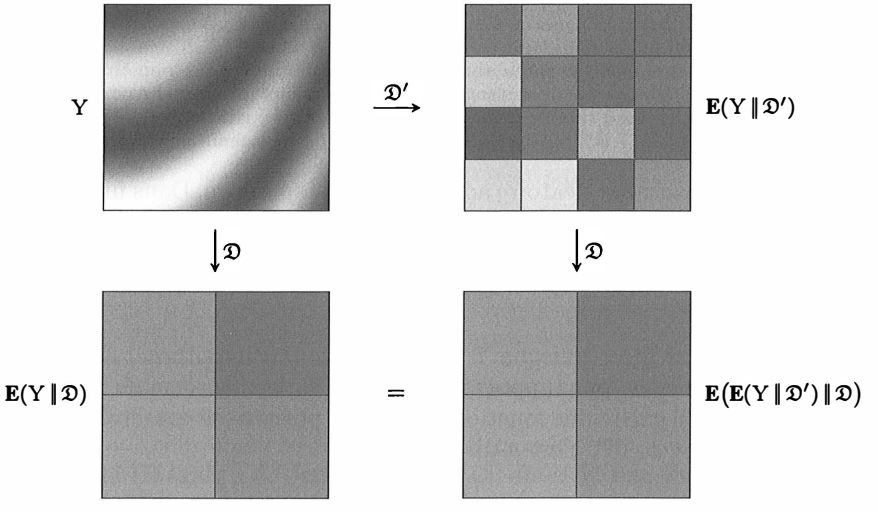


FIGURE 23.5 — *Propriété de la tour. Si $\mathfrak{D}'$ est une partition plus fine que $\mathfrak{D}$, alors la variable $\mathbf{E}(Y \,\|\, \mathfrak{D}')$, vue à travers le filtre $\mathfrak{D}$, n'est rien d'autre que $\mathbf{E}(Y \,\|\, \mathfrak{D})$.*

23.2 LE CAS GÉNÉRAL

Lorsque X est une variable à densité, la partition de Ω en ensembles $X^{-1}(\{x\})$ ne permet plus de faire les calculs précédents : tous les événements $D_x = \{X = x\}$ sont de probabilité nulle et on ne peut plus définir aucune des quantités $\mathbf{P}(A \mid D_x)$ comme $\mathbf{P}(A \cap D_x)/\mathbf{P}(D_x)$, quotient de deux quantités nulles.

On ne considérera plus les événements D_x eux-mêmes, mais la tribu $\sigma(X)$ dans son ensemble. Dans le cas fini ou dénombrable, savoir si ω est, ou non, élément de chaque D_k est équivalent à savoir si ω est élément, ou non, de chaque élément G de $\mathfrak{G} = \sigma(\mathfrak{D})$. Le gain est évidemment que, même si les atomes de $\mathfrak{G}$ sont de probabilité nulle, en revanche des unions (non dénombrables) de tels atomes, si elles font partie de $\mathfrak{G}$, seront de probabilité strictement positive.

§ 344 **Espérance d'une variable aléatoire sachant un événement**

Avant de plonger au cœur du problème, donnons un sens à $\mathbf{E}(X \mid A)$ lorsque A est un événement de probabilité non nulle et X une variable aléatoire quelconque.

Lorsque X est une variable aléatoire discrète, la construction de la section précédente montrait que

$$\mathbf{E}(X \mid A) = \frac{\mathbf{E}(X\mathbf{1}_A)}{\mathbf{P}(A)}.$$

C'est la définition que, encouragés par la Cavalerie Légère (§ 215), on choisit pour une variable aléatoire quelconque ; elle est cohérente avec l'interprétation de cette quantité comme espérance de la variable X dans l'espace A muni de la tribu des éléments de la forme $B \cap A$ pour B parcourant $\mathfrak{T}$ et de la probabilité $\mathbf{P}_A$ définie
⏢23.2 par $\mathbf{P}_A(B \cap A) = \mathbf{P}(B \cap A)/\mathbf{P}(A)$.

§ 345 **Espérance conditionnelle sachant une tribu**

Soit $\mathfrak{G}$ une sous-tribu de $\mathfrak{T}$. Comme précédemment, $\mathfrak{G}$ peut, de manière intuitive, représenter une connaissance partielle de Ω : un observateur, s'il ne peut pas déterminer ω de façon précise, sait néanmoins quels sont les événements $G \in \mathfrak{G}$ auxquels ω appartient.

Nous ne pouvons plus utiliser la méthode du paragraphe § 338 pour définir une espérance conditionnelle selon une variable aléatoire générale ; pourtant, l'intuition nous dit que cela devrait pouvoir exister.

Exemple 23.10 Reprenons l'exemple du paragraphe § 190, page 274 : on tire un nombre X au hasard dans $[0\,;\,1]$ puis, ce résultat étant connu, on tire un nombre Y au hasard dans $[0\,;\,X]$ (avec une probabilité uniforme à chaque fois). Intuitivement, la loi conditionnelle de Y sachant X est la loi uniforme $\mathscr{U}([0\,;\,X])$ et l'espérance conditionnelle de Y sachant X est donc $\mathbf{E}(Y \,\|\, X) = X/2$.

Soit Y une variable aléatoire admettant une espérance. Dans un premier temps, supposons Y positive. On définit sur $(\Omega, \mathfrak{G})$ une mesure positive $\mathbf{Q}_Y$ par :

$$\forall G \in \mathfrak{G} \qquad \mathbf{Q}_Y(G) = \int_G Y \, d\mathbf{P} = \mathbf{E}(Y \cdot \mathbf{1}_G).$$

Cette mesure est finie (puisque Y est intégrable par hypothèse) et elle est de plus absolument continue par rapport à la mesure $\mathbf{P}$; le théorème de Radon-Nikodým nous assure qu'il existe une fonction $Z : \Omega \to \mathbb{R}$ positive, $\mathfrak{G}$-mesurable et intégrable, telle que $d\mathbf{Q}_Y = Z \cdot d\mathbf{P}$, c'est-à-dire

$$\forall G \in \mathfrak{G} \qquad \int_G Z \, d\mathbf{P} = \mathbf{Q}_Y(G) = \int_G Y \, d\mathbf{P}. \tag{23.8}$$

On reconnaît la propriété fondamentale (23.7) de l'espérance conditionnelle ! On appelle **espérance conditionnelle de Y sachant $\mathfrak{G}$** et on note $Z = \mathbf{E}(Y \,\|\, \mathfrak{G})$ la variable aléatoire Z ainsi définie. Elle est unique à une égalité presque sûre près.

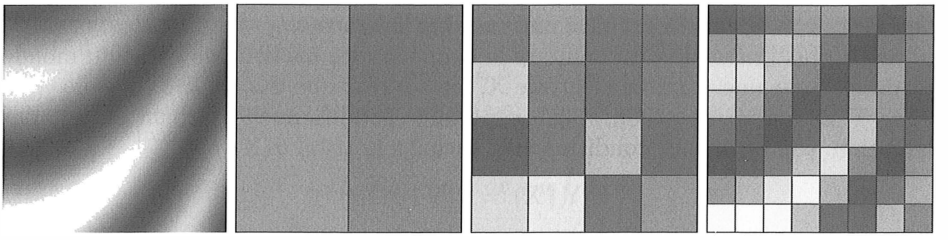


FIGURE 23.6 — *Les valeurs de la variable aléatoire* Y *sont représentés par des niveaux de gris. On représente également* $\mathbf{E}(Y \,\|\, \mathfrak{D}_i)$, *où* $\mathfrak{D}_1$, $\mathfrak{D}_2$ *et* $\mathfrak{D}_3$ *sont des partitions de plus en plus fines de* Ω.

Si Y n'est pas positive, on la décompose en $Y = Y^+ - Y^-$ et l'on pose

$$\mathbf{E}(Y \,\|\, \mathfrak{G}) := \mathbf{E}(Y^+ \,\|\, \mathfrak{G}) - \mathbf{E}(Y^- \,\|\, \mathfrak{G}).$$

En résumé, voici la définition de l'espérance conditionnelle[2] :

Définition 23.11 Si Y est une variable aléatoire intégrable et $\mathfrak{G}$ une sous-tribu de la tribu $\mathfrak{T}$, l'**espérance conditionnelle** de Y sachant $\mathfrak{G}$ est l'unique (classe de) variable aléatoire $\mathfrak{G}$-mesurable telle que

$$\forall G \in \mathfrak{G} \qquad \int_G Y \, d\mathbf{P} = \int_G \mathbf{E}(Y \,\|\, \mathfrak{G}) \, d\mathbf{P}, \tag{23.9}$$

ce qui s'écrit aussi :

$$\forall G \in \mathfrak{G} \qquad \mathbf{E}(Y \cdot 1_G) = \mathbf{E}\big(\mathbf{E}(Y \,\|\, \mathfrak{G}) \cdot \mathbf{1}_G\big)$$

Dans cette définition, la notion importante est la *mesurabilité par rapport à* $\mathfrak{G}$. D'une certaine manière, l'espérance conditionnelle est une manière de voir Y à travers un « verre dépoli » qui est une tribu $\mathfrak{G}$ plus grossière que $\mathfrak{T}$ (voir figure 23.6).

Remarque 23.12 (Unicité de l'espérance conditionnelle) L'existence et l'unicité de l'espérance conditionnelle proviennent du théorème de Radon-Nikodým, que nous avons admis ici. L'unicité seule est cependant démontrable par des moyens élémentaires. Supposons que Z et Z′ soient deux variables aléatoires $\mathfrak{G}$-mesurables vérifiant l'équation (23.8) et qui ne soient pas égales presque sûrement. Quitte à les renommer, supposons $\mathbf{P}\{Z' > Z\} > 0$. On a, par croissance de la suite d'événements $\{Z' > Z + 1/n\}$ vers $\{Z' > Z\}$:

$$\lim_{n\to\infty} \mathbf{P}\{Z' > Z + \tfrac{1}{n}\} = \mathbf{P}\{Z' > Z\} > 0,$$

donc il existe un entier n tel que $\mathbf{P}\{Z' > Z + 1/n\} > 0$. Or, l'ensemble $G = \{Z' > Z + 1/n\}$ est élément de $\mathfrak{G}$ puisque $Z' - Z$ est $\mathfrak{G}$-mesurable, donc

$$\mathbf{E}(Z' - Z \,|\, G) \geqslant \frac{1}{n} \mathbf{P}(G) > 0$$

ou encore $\mathbf{E}(Z' \, \mathbf{1}_G) > \mathbf{E}(Z \, \mathbf{1}_G)$, en contradiction avec les hypothèses.

Tout comme dans la construction de l'espérance, l'espérance conditionnelle vérifie une inégalité importante :

■ **Théorème 23.13 (Inégalité de Jensen)** *Soient* $\varphi : \mathbb{R} \to \mathbb{R}$ *une fonction convexe et* X *une variable aléatoire intégrable. Alors*

$$\varphi\big(\mathbf{E}(X \,\|\, \mathfrak{G})\big) \leqslant \mathbf{E}(\varphi(X) \,\|\, \mathfrak{G}) \qquad p.s.$$

(2). C'est encore Kolmogorov, en 1933, qui a proposé ce théorème/définition.

§346 **Espérance conditionnelle sachant une variable aléatoire**
Le cas le plus intéressant est celui où la connaissance partielle de l'état du monde est donnée par une variable aléatoire X, c'est-à-dire que $\mathfrak{G} = \sigma(X)$.

Si Y admet une espérance, son **espérance conditionnelle sachant X** est, par définition, son espérance conditionnelle sachant la tribu $\sigma(X)$:

$$\mathbf{E}(Y \,\|\, X) := \mathbf{E}\big(Y \,\|\, \sigma(X)\big).$$

§347 **Lemme de factorisation**
Avant de continuer, faisons une halte pour rappeler l'interprétation de la mesurabilité d'une variable aléatoire Y par rapport à la tribu $\sigma(X)$.

■ **Théorème 23.14 (Doob-Dynkin)** *Une variable aléatoire* Y *est mesurable selon la tribu* $\sigma(X)$ *si et seulement si il existe une fonction borélienne* $f : \mathbb{R} \to \mathbb{R}$ *telle que* $Y = f(X)$.

Remarque 23.15 (Idée de la preuve) L'idée qui sous-tend ce théorème est la suivante. Si x est un élément de $X(\Omega)$, alors les ensembles $X^{-1}(\{x\})$ sont les *atomes* de la tribu $\sigma(X)$ — ils sont insécables, au sens où il n'existe pas dans $\sigma(X)$ d'élément inclus dans $X^{-1}(\{x\})$, à part bien sûr l'ensemble vide. Par conséquent, la fonction Y ne peut prendre qu'une seule valeur sur chacun des $X^{-1}(\{x\})$; cette valeur est notée $f(x)$.

Dans la preuve complète que le lecteur trouvera en annexe, au paragraphe A70, page 701, on adopte le point de vue de la Cavalerie Légère : on suppose $Y \geqslant 0$, on l'approche par une suite croissante de variables simples Y_n et on construit une factorisation explicite $Y_n = f_n(X)$; le cas général se traite en séparant $Y = Y^+ - Y^-$. Cette construction permet de montrer le caractère borélien de f, ce que l'on ne peut pas faire avec une factorisation directe.

Dit d'une manière très imagée : *la tribu* $\sigma(X)$ *engendrée par* X *contient tous les renseignement délivrés par la fonction* X *— à part ses valeurs.*

§348 **Propriétés de l'espérance conditionnelle**
Nous donnons ici quelques propriétés parmi les plus utiles sur l'espérance conditionnelle. On remarquera que leur démonstration est parfois plus rapide que dans le cas discret déjà traité.

■ **Théorème 23.16** *Soit* X *une variable aléatoire. Soient* Y *et* Y′ *deux variables aléatoires admettant une espérance et* $\mathfrak{G}$ *une sous-tribu de* $\mathfrak{T}$. *Les égalités suivantes entre espérances conditionnelles sont toujours à prendre au sens presque sûr.*

i) ***Linéarité :*** $\mathbf{E}(\lambda Y + \mu Y \,\|\, \mathfrak{G}) = \lambda \mathbf{E}(Y \,\|\, \mathfrak{G}) + \mu \mathbf{E}(Y' \,\|\, \mathfrak{G})$.

ii) ***Moyenne des moyennes :*** $\mathbf{E}\big(\mathbf{E}(Y \,\|\, \mathfrak{G})\big) = \mathbf{E}(Y)$.

iii) Si $Y = a$ *p.s., alors* $\mathbf{E}(Y \,\|\, \mathfrak{G}) = a$.

iv) ***Positivité :*** *Si* $Y \geqslant 0$ *p.s., alors* $\mathbf{E}(Y \,\|\, \mathfrak{G}) \geqslant 0$.

v) ***Croissance :*** *Si* $Y \leqslant Y'$ *p.s., alors* $\mathbf{E}(Y \,\|\, \mathfrak{G}) \leqslant \mathbf{E}(Y' \,\|\, \mathfrak{G})$.

vi) ***Inégalité triangulaire :*** $\big|\mathbf{E}(Y \,\|\, \mathfrak{G})\big| \leqslant \mathbf{E}\big(|Y| \,\|\, \mathfrak{G}\big)$.

vii) ***Indépendance :*** *Si* X *et* Y *sont indépendantes, alors* $\mathbf{E}(Y \,\|\, X) = \mathbf{E}(Y)$.

viii) Si Y *est* $\mathfrak{G}$*-mesurable, alors* $\mathbf{E}(Y \,\|\, \mathfrak{G}) = Y$.

ix) $\mathbf{E}(X \,\|\, X) = X$.

x) ***« Tower property » :*** *Si* $\mathfrak{G}$ *et* $\mathfrak{G}'$ *sont des tribus et* $\mathfrak{G}' \preccurlyeq \mathfrak{G}$, *alors*

$$\mathbf{E}\big(\mathbf{E}(Y \,\|\, \mathfrak{G}') \,\|\, \mathfrak{G}\big) = \mathbf{E}(Y \,\|\, \mathfrak{G}).$$

xi) ***Inégalité de Jensen :*** *Si* φ *est convexe et* X *intégrable, alors*

$$\varphi\big(\mathbf{E}(X \,\|\, \mathfrak{G})\big) \leqslant \mathbf{E}(\varphi(X) \,\|\, \mathfrak{G}) \quad p.s.$$

DÉMONSTRATION : Pour montrer qu'une variable Z est égale (p.s.) à l'espérance conditionnelle d'une variable aléatoire Y sachant $\mathfrak{G}$, il suffit de montrer qu'elle est $\mathfrak{G}$-mesurable et qu'elle satisfait $\int_G Z\,d\mathbf{P} = \int_G Y\,d\mathbf{P}$ pour tout $G \in \mathfrak{G}$.

i) Notons $Z = \mathbf{E}(Y \,\|\, \mathfrak{G})$ et $Z' = \mathbf{E}(Y' \,\|\, \mathfrak{G})$. Alors $\lambda Z + \mu Z'$ est $\mathfrak{G}$-mesurable et, pour tout $G \in \mathfrak{G}$,

$$\begin{aligned}\int_G (\lambda Z + \mu Z')\,d\mathbf{P} &= \lambda \int_G Z\,d\mathbf{P} + \mu \int_G Z'\,d\mathbf{P} \\ &= \lambda \int_G Y\,d\mathbf{P} + \mu \int_G Y'\,d\mathbf{P} = \int_G (\lambda Y + \mu Y')\,d\mathbf{P},\end{aligned}$$

ce qui montre que $\mathbf{E}(\lambda Y + \mu Y' \,\|\, \mathfrak{G}) = \lambda Z + \mu Z'$ comme voulu.

ii) Puisque $\Omega \in \mathfrak{G}$,

$$\mathbf{E}(\mathbf{E}(Y \,\|\, \mathfrak{G})) = \int_\Omega \mathbf{E}(Y \,\|\, \mathfrak{G})\,d\mathbf{P} = \int_\Omega Y\,d\mathbf{P} = \mathbf{E}(Y).$$

iii) Immédiat, puisque $\int_G a\,d\mathbf{P} = \int_G Y\,d\mathbf{P}$ pour tout $G \in \mathfrak{G}$.

iv) Notons $Z = \mathbf{E}(Y \,\|\, \mathfrak{G})$ et supposons par l'absurde que $\mathbf{P}\{Z < 0\} > 0$. Puisque la suite d'événements $\{Z + 1/n < 0\}$ croît vers $\{Z < 0\}$, on a $\mathbf{P}\{Z + 1/n < 0\} \to \mathbf{P}\{Z < 0\} > 0$, donc il existe un entier n tel que $\mathbf{P}\{Z + 1/n < 0\} > 0$. Notons $G = \{Z < -1/n\}$, alors G est élément de $\mathfrak{G}$ puisque Z est $\mathfrak{G}$-mesurable. On a alors $\int_G Y\,d\mathbf{P} = \int_G Z\,d\mathbf{P} \leqslant \int_G (-1/n)\,d\mathbf{P} = -\mathbf{P}(G)/n < 0$, en contradiction avec la positivité presque sûre de Y qui implique que $\int_G Y\,d\mathbf{P} \geqslant 0$.

v) Découle du point précédent et de l'additivité.

vi) Découle du point précédent et que l'inégalité $-|Y| \leqslant Y \leqslant |Y|$.

vii) Pour tout $G \in \mathfrak{G}$, $\int_G \mathbf{E}(Y)\,d\mathbf{P} = \mathbf{E}(Y) \int_G d\mathbf{P} = \mathbf{E}(Y)\,\mathbf{E}(\mathbf{1}_G) = \mathbf{E}(Y\,\mathbf{1}_G)$ par indépendance, c'est-à-dire $\int_G \mathbf{E}(Y)\,d\mathbf{P} = \int_G Y\,d\mathbf{P}$.

viii) Immédiat par définition.

ix) C'est une tautologie : X est bien $\sigma(X)$-mesurable et $\int_G X\,d\mathbf{P} = \int_G X\,d\mathbf{P}$ pour tout $G \in \sigma(X)$!

∎

Théorèmes de convergence

§ 349

Les principaux théorèmes de convergence de l'espérance d'une suite ont leur équivalent pour l'espérance conditionnelle.

■ **Théorème 23.17 (Convergence monotone conditionnelle)** *Si* $0 \leqslant Y_n \uparrow Y$ *presque sûrement, alors* $\mathbf{E}(Y_n \,\|\, \mathfrak{G}) \to \mathbf{E}(Y \,\|\, \mathfrak{G})$ *presque sûrement.*

DÉMONSTRATION : Notons $Z_n = \mathbf{E}(Y_n \,\|\, \mathfrak{G})$ pour tout $n \in \mathbb{N}$. Par croissance de l'espérance conditionnelle, la suite $(Z_n)_{n\geqslant 1}$ est presque sûrement croissante (pour tout $n \in \mathbb{N}$, on a $Z_n \leqslant Z_{n+1}$ sur un événement A_n de probabilité 1, donc la suite $(Z_n)_{n\geqslant 1}$ est croissante sur l'intersection $A = \bigcap A_n$, également de probabilité 1). On pose $Z := \limsup Z_n$, alors $Z_n \uparrow Z$ presque sûrement et Z est $\mathfrak{G}$-mesurable. Enfin, par double application du théorème de convergence monotone,

$$\int_G Y\,d\mathbf{P} = \lim_{n\to\infty} \int_G Y_n\,d\mathbf{P} = \lim_{n\to\infty} \int_G Z_n\,dP = \int_G Z\,d\mathbf{P}$$

pour tout $G \in \mathfrak{G}$, ce qui montre que $Z = \mathbf{E}(Y \,\|\, \mathfrak{G})$. ∎

■ **Théorème 23.18 (Convergence dominée conditionnelle)** *Si* $Y_n \to Y$ *presque sûrement et s'il existe une variable aléatoire* T *intégrable telle que* $|Y_n| \leqslant T$ *pour tout* $n \in \mathbb{N}$, *alors* $\lim_{n\to\infty} \mathbf{E}(Y_n \,\|\, \mathfrak{G}) = \mathbf{E}(Y \,\|\, \mathfrak{G})$ *presque sûrement.*

DÉMONSTRATION : Posons $M_n = \sup_{k\geqslant n} |Y_k - Y|$ pour tout entier n. Alors, $|Y_n - Y| \leqslant M_n$ et $M_n \downarrow 0$ presque sûrement. De plus, $|M_n| \leqslant 2T$; le théorème de convergence monotone conditionnelle (appliqué à $2T - M_n$) montre que $\mathbf{E}(M_n \,\|\, \mathfrak{G}) \to 0$. Enfin, l'inégalité triangulaire $\big|\mathbf{E}(Y_n \,\|\, \mathfrak{G}) - \mathbf{E}(Y \,\|\, \mathfrak{G})\big| \leqslant \mathbf{E}(M_n \,\|\, \mathfrak{G})$ permet de conclure. ∎

Une fois munis du théorème de convergence monotone, nous pouvons faire appel à la Cavalerie Légère. Le lecteur est invité, sur ce principe, à démontrer le théorème suivant qui montre que, si l'on est muni de l'information véhiculée par la tribu $\mathfrak{G}$, alors toute fonction $\mathfrak{G}$-mesurable agit comme une constante.

■ **Théorème 23.19** *Si* Y *est* $\mathfrak{G}$*-mesurable, si* Z *et* YZ *sont intégrables, alors*

$$\mathbf{E}(\mathrm{YZ} \,\|\, \mathfrak{G}) = \mathrm{Y}\, \mathbf{E}(\mathrm{Z} \,\|\, \mathfrak{G}) \qquad p.s.$$

§ 350 **Densités conditionnelles (∗)**
Revenons au cas d'un couple (X, Y) admettant une densité conjointe $f(x,y)$. On rappelle qu'on a défini la *densité conditionnelle de* Y *par rapport à* X comme

$$f_{\mathrm{Y}}(y \,|\, x) = f^{\mathrm{Y}}_{\{\mathrm{X}=x\}}(y) := \begin{cases} \dfrac{f(x,y)}{f_{\mathrm{X}}(x)} & \text{si } f_{\mathrm{X}}(x) > 0 \\ \rho(x) & \text{si } f_{\mathrm{X}}(x) = 0. \end{cases}$$

où ρ est une densité arbitraire et $f_{\mathrm{X}}(x) = \int f(x,y)\,\mathrm{d}y$ la densité marginale de X. Cette fonction nous permet de calculer des espérances conditionnelles comme le montre le résultat suivant :

■ **Proposition 23.20** *Soit* $\varphi : \mathbb{R} \to \mathbb{R}$ *une fonction mesurable, telle que* $\varphi(\mathrm{Y})$ *soit intégrable. Définissons la fonction* ψ *par*

$$\psi(x) = \int_{-\infty}^{+\infty} \varphi(y)\, f_{\mathrm{Y}}(y \,|\, x)\, \mathrm{d}y.$$

Alors la variable $\mathrm{Z} = \psi(\mathrm{X})$ *est (une version de) l'espérance conditionnelle* $\mathbf{E}\big(\varphi(\mathrm{Y}) \,\|\, \mathrm{X}\big)$ *de* $\varphi(\mathrm{Y})$ *sachant* X.

Démonstration : Soit $\mathrm{G} \in \sigma(\mathrm{X})$; on doit montrer que $\int_{\mathrm{G}} \mathrm{Z}\,\mathrm{d}\mathbf{P} = \int_{\mathrm{G}} \varphi(\mathrm{Y})\,\mathrm{d}\mathbf{P}$. L'événement G est de la forme $\mathrm{G} = \{\mathrm{X} \in \mathrm{B}\}$, où B est un borélien de $\mathbb{R}$. Le théorème de transfert et celui de Fubini permettent d'écrire

$$\begin{aligned} \int_{\mathrm{G}} \mathrm{Z}\,\mathrm{d}\mathbf{P} &= \int_{\Omega} \psi(\mathrm{X})\, \mathbf{1}_{\mathrm{B}}(\mathrm{X})\,\mathrm{d}\mathbf{P} = \int_{-\infty}^{+\infty} \psi(x)\, \mathbf{1}_{\mathrm{B}}(x)\, f_{\mathrm{X}}(x)\,\mathrm{d}x \\ &= \iint_{\mathbb{R}^2} \varphi(y)\, f_{\mathrm{Y}}(y \,|\, x)\, \mathbf{1}_{\mathrm{B}}(x)\, f_{\mathrm{X}}(x)\,\mathrm{d}x\,\mathrm{d}y. \end{aligned}$$

De même, par théorème de transfert et Fubini,

$$\int_{\mathrm{G}} \varphi(\mathrm{Y})\,\mathrm{d}\mathbf{P} = \int_{\Omega} \varphi(\mathrm{Y})\, \mathbf{1}_{\mathrm{B}}(\mathrm{X})\,\mathrm{d}\mathbf{P} = \iint_{\mathbb{R}^2} \varphi(y)\, \mathbf{1}_{\mathrm{B}}(x)\, f(x,y)\,\mathrm{d}x\,\mathrm{d}y,$$

et ces deux quantités sont égales car $f(x,y) = f_{\mathrm{X}}(x)\, f_{\mathrm{Y}}(y \,|\, x)$ pour presque tous $x, y \in \mathbb{R}$. ▮

§ 351 **Le point de vue hilbertien : l'espérance conditionnelle comme projection**
Dans le cadre de l'espace L^2 des variables aléatoires de carré intégrable, l'espérance conditionnelle admet une interprétation particulièrement éclairante.

Si $\mathfrak{G}$ est une sous-tribu de $\mathfrak{T}$, notons $\mathrm{L}^2(\Omega, \mathfrak{G}, \mathbf{P})$ l'espace des fonctions $\mathfrak{G}$-mesurables et de carré intégrable[3]. C'est un sous-espace vectoriel de $\mathrm{L}^2(\Omega, \mathfrak{T}, \mathbf{P})$ et, qui plus est, c'est encore un espace complet. On peut donc lui appliquer le théorème de projection orthogonale :

■ **Proposition 23.21** *Si* Y *est une variable aléatoire de* $\mathrm{L}^2(\Omega, \mathfrak{T}, \mathbf{P})$*, il existe, à une égalité presque sûre près, une unique variable aléatoire* $\mathrm{Z} \in \mathrm{L}^2(\Omega, \mathfrak{G}, \mathbf{P})$ *qui minimise* $\mathbf{E}\,|\mathrm{Y} - \mathrm{Z}|^2$ *; cette variable aléatoire est le projeté orthogonal de* Y *sur* $\mathrm{L}^2(\Omega, \mathfrak{G}, \mathbf{P})$.

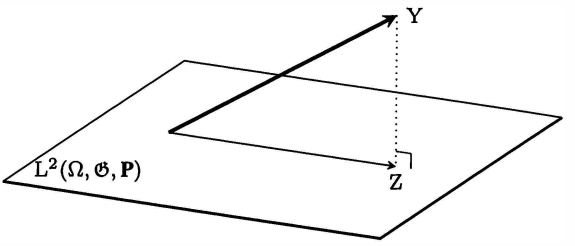


FIGURE 23.7 — *Projection orthogonale d'une variable* Y *sur l'espace* $L^2(\Omega, \mathfrak{G}, \mathbf{P})$.

(Voir figure 23.7). On a orthogonalité entre $Y-Z$ et tout vecteur U de $L^2(\Omega, \mathfrak{G}, \mathbf{P})$:

$$\forall U \in L^2(\Omega, \mathfrak{G}, \mathbf{P}) \qquad \langle Y - Z | U \rangle = 0$$

Notamment, pour tout $G \in \mathfrak{G}$, si l'on pose $U := \mathbf{1}_G$, on obtient

$$\langle Y - Z | \mathbf{1}_G \rangle = \mathbf{E}\big[(Y - Z)\mathbf{1}_G\big] = 0$$

ou encore

$$\mathbf{E}(Y\,\mathbf{1}_G) = \mathbf{E}(Z\,\mathbf{1}_G),$$

ce qui est justement (avec le fait que Z est $\mathfrak{G}$-mesurable), la caractérisation de l'espérance conditionnelle de Y sachant $\mathfrak{G}$.

■ **Théorème 23.22** *Si* $Y \in L^2(\Omega, \mathfrak{T}, \mathbf{P})$ *et si* $\mathfrak{G}$ *est une sous-tribu de* $\mathfrak{T}$, *alors l'espérance conditionnelle de* Y *sachant* $\mathfrak{G}$ *est le projeté orthogonal de* Y *sur* $L^2(\Omega, \mathfrak{G}, \mathbf{P})$.

Ainsi, dans l'espace L^2, l'existence et l'unicité de l'espérance conditionnelle découlent directement du théorème de projection orthogonale ; la quantité $\mathbf{E}(Y \,\|\, \mathfrak{G})$ est de plus caractérisée comme *la meilleure approximation quadratique* $\mathfrak{G}$*-mesurable de* Y.

Remarque 23.23 (Tower property) La *tower property* 23.9 est une conséquence géométrique immédiate de ce point de vue géométrique : projeter orthogonalement sur un sous-espace vectoriel V de E, puis sur un sous-espace W de V, revient à projeter directement sur W.

Remarque 23.24 (Existence de l'espérance conditionnelle dans L^1*)* On peut maintenant étendre le résultat démontré dans L^2 aux variables de L^1 par les techniques usuelles de Cavalerie Légère (approximations et monotonie).

Dans un premier temps, supposons $Y \geqslant 0$ et approchons-la par une suite $(Y_n)_{n\geqslant 1}$ croissante de variables bornées : $Y_n \uparrow Y$ et $Y_n \in L^2$. On note $Z_n = \mathbf{E}(Y_n \,\|\, \mathfrak{G})$, ou plus précisément, on choisit Z_n comme une version de $\mathbf{E}(Y_n \,\|\, \mathfrak{G})$. Remarquons d'abord que la suite $(Z_n)_{n\geqslant 1}$ est (presque sûrement) croissante par croissance de l'espérance conditionnelle (proposition 23.16). On pose alors

$$Z = \limsup_{n\to\infty} Z_n.$$

Sur un ensemble de probabilité 1, c'est une limite simple, et cette définition assure que Z est $\mathfrak{G}$-mesurable et $Z_n \uparrow Z$ presque sûrement. Le théorème de convergence monotone assure alors que

$$\int_G Z\,d\mathbf{P} = \lim_n \uparrow \int_G Z_n\,d\mathbf{P} = \lim_n \uparrow \int_G Y_n\,d\mathbf{P} = \int_G Y\,d\mathbf{P}$$

pour tout $G \in \mathfrak{G}$: ainsi, Z est (une version de) l'espérance conditionnelle de Y sachant $\mathfrak{G}$.

Dans le cas où Y n'est pas positive, il suffit de la séparer en $Y = Y^+ - Y^-$ pour conclure.

Cette méthode, conjointement à la remarque 23.12, permet ainsi de se passer entièrement du théorème de Radon-Nikodým pour prouver l'existence et l'unicité de l'espérance conditionnelle.

(3). Par abus, on a encore noté **P** la restriction de **P** à $\mathfrak{G}$.

23.3 MARTINGALES

Munis de la notion d'espérance conditionnelle, nous pouvons aborder, très brièvement, les *martingales*, qui sont devenues, au cours des dernières décennies, des objets fondamentaux et des outils de base extrêmement puissants de la théorie des probabilités.

§ 352 **Ma première martingale**

Considérons une suite $(X_n)_{n\geqslant 0}$ de variables aléatoires *indépendantes*, de même loi et admettant une espérance nulle, par exemple des variables de Rademacher :

$$\mathbf{P}\{X_n = -1\} = \tfrac{1}{2} \qquad \text{et} \qquad \mathbf{P}\{X_n = +1\} = \tfrac{1}{2},$$

et posons

$$S_n = X_1 + \cdots + X_n$$

pour tout entier $n \geqslant 1$. On s'intéresse à l'espérance conditionnelle de S_{n+1} sachant S_n, c'est-à-dire à « si l'on connaît la valeur de S_n, que peut-on en déduire pour la valeur moyenne de S_{n+1} ? ». Puisque S_n est une variable aléatoire discrète (et même finie), cette espérance conditionnelle se calcule selon la méthode du paragraphe § 342, page 474 : pour chacune des valeurs k que peut prendre S_n (les valeurs de k telles que $-n \leqslant k \leqslant n$ et k a la même parité que n), la loi conditionnelle de S_{n+1} sachant $S_n = k$ est

$$\mathbf{P}\{S_{n+1} = k+1 \,|\, S_n = k\} = \tfrac{1}{2} \qquad \text{et} \qquad \mathbf{P}\{S_{n+1} = k-1 \,|\, S_n = k\} = \tfrac{1}{2},$$

et donc

$$\mathbf{E}(S_{n+1} \,|\, S_n = k) = \tfrac{1}{2}(k+1) + \tfrac{1}{2}(k-1) = k.$$

En d'autres termes, on a montré que

$$\mathbf{E}(S_{n+1} \,\|\, S_n) = S_n. \tag{23.10}$$

Ce résultat élémentaire reste d'ailleurs valable sans hypothèse spécifique sur la loi commune des variables X_n (mise part la nullité de l'espérance) :

$$\mathbf{E}(S_{n+1} \,\|\, S_n) = \mathbf{E}(S_n + X_{n+1} \,|\, S_n) = \mathbf{E}(S_n \,\|\, S_n) + \mathbf{E}(X_{n+1} \,\|\, S_n)$$

par linéarité ; l'indépendance de X_{n+1} par rapport à $X_1, \ldots, X_n$, et donc à S_n, permet d'écrire que

$$\mathbf{E}(X_{n+1} \,\|\, S_n) = \mathbf{E}(X_{n+1}) = 0.$$

Enfin, l'égalité $\mathbf{E}(S_n \,\|\, S_n) = S_n$ montre que l'équation (23.10) est encore vérifiée. Cette équation se généralise encore sous la forme

$$\mathbf{E}(S_{n+1} \,\|\, S_1, S_2, \ldots, S_n) = S_n, \tag{$*$}$$

qui découle du fait que

$$\sigma(S_1, \ldots, S_n) = \sigma(X_1, X_2, \ldots, X_n) \supset \sigma(S_n).$$

L'équation $(*)$ fait de la suite $(S_n)_{n\geqslant 1}$ un cas particulier de *martingale*.

§ 353 **Filtrations, martingales**

Donnons d'abord une version restreinte de la définition d'une **martingale** $(M_n)_{n\geqslant 0}$: c'est une suite de variables aléatoires intégrables telle que

$$\mathbf{E}(M_{n+1} \,\|\, M_0, M_1, \ldots, M_n) = M_n.$$

Autrement dit, *lorsque l'on connaît les valeurs de* $M_0, M_1, \ldots, M_n$, *la meilleure estimation que l'on puisse faire de* M_{n+1} *est* M_n.

Cette définition admet une généralisation, un peu plus abstraite. On appelle **filtration** une suite croissante $(\mathfrak{F}_n)_{n\geqslant 0}$ de sous-tribus de $\mathfrak{T}$:

$$\mathfrak{F}_0 \subset \mathfrak{F}_1 \subset \cdots \subset \mathfrak{F}_n \subset \cdots$$

Rappelons que l'on dit également que $\mathfrak{F}_0$ est *plus grossière* que $\mathfrak{F}_1$, elle-même plus grossière que $\mathfrak{F}_2$ etc. : toute partie de Ω, mesurable dans $\mathfrak{F}_n$, l'est également dans $\mathfrak{F}_{n+1}$.

Exemple 23.25 (Filtration naturelle) Si $(X_n)_{n\geqslant 0}$ est une suite de variables aléatoires et si l'on pose

$$\mathfrak{F}_n := \sigma(X_0, X_1, \ldots, X_n),$$

alors $(\mathfrak{F}_n)_{n\geqslant 1}$ est une filtration. Elle décrit l'information que l'on possède à l'instant n et qui est la donnée des valeurs $X_0(\omega), X_1(\omega), \ldots, X_n(\omega)$.

Rappelons que, si Y_n est une variable aléatoire $\mathfrak{F}_n$-mesurable, alors il existe une fonction borélienne $f_n : \mathbb{R}^{n+1} \to \mathbb{R}$ telle que $Y_n = f_n(X_0, X_1, \ldots, X_n)$.

Une suite $(M_n)_{n\geqslant 0}$ est appelée une **martingale** relativement à $(\mathfrak{F}_n)_{n\geqslant 0}$ si, pour tout $n \geqslant 0$:

i) M_n est $\mathfrak{F}_n$-mesurable ;

ii) M_n est intégrable ;

iii) $\mathbf{E}(M_{n+1} \,\|\, \mathfrak{F}_n) = M_n$.

Une suite $(M_n)_{n\geqslant 0}$ vérifiant la première propriété (M_n est mesurable par rapport à $\mathfrak{F}_n$ pour tout $n \geqslant 0$) est dite **adaptée** à la filtration $(\mathfrak{F}_n)_{n\geqslant 0}$.

Exemple 23.26 Dans le paragraphe précédent, la suite $(S_n)_{n\geqslant 1}$ est une martingale relativement à la filtration naturelle $\mathfrak{F}_n := \sigma(X_1, X_2, \ldots, X_n)$.

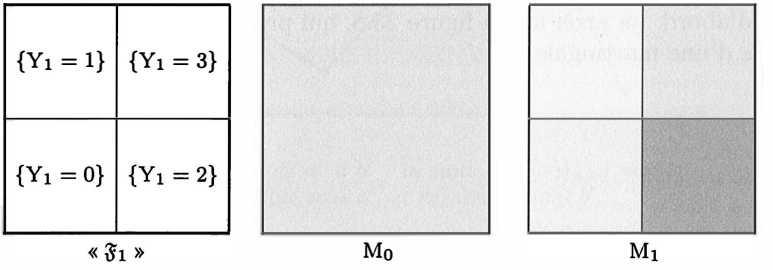
Exemple 23.27 Soient X une variable aléatoire intégrable et $(\mathfrak{F}_n)_{n\geqslant 0}$ une filtration. Posons $M_n := \mathbf{E}(X \,\|\, \mathfrak{F}_n)$ pour tout entier n ; alors M_n est intégrable et $\mathfrak{F}_n$ mesurable par construction et de plus le théorème 23.16 (*tower property*) montre que $\mathbf{E}(M_{n+1} \,\|\, \mathfrak{F}_n) = M_n$. La martingale ici présente est donc d'un type particulier, puisqu'elle dérive d'une variable aléatoire X (on dit que c'est une martingale **fermée**).

Il faut imaginer $(M_n)_{n\geqslant 0}$ comme la suite des projections d'une fonction intégrable X sur une suite d'espaces de plus en plus gros : ce sont donc des fonctions prenant des valeurs de plus en plus précises au fur et à mesure que l'on obtient des données au travers des tribus $\mathfrak{F}_n$.

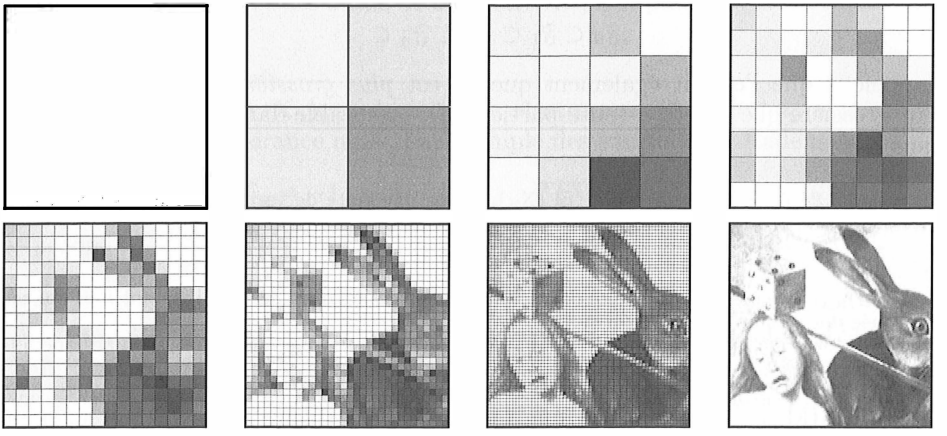
Donnons un exemple. Soit $(Y_n)_{n\geqslant 1}$ une suite de variables aléatoires de même loi $\mathscr{B}(3\ ;\ p)$ et posons $\mathfrak{F}_n = \sigma(Y_1, Y_2, \ldots, Y_n)$ pour tout entier n. Soit X une variable aléatoire intégrable (qui peut être fonction des Y_k), et posons $M_n = \mathbf{E}(X \,|\, \mathfrak{F}_n)$. Notons $M_0 = \mathbf{E}(X)$ la valeur moyenne de X. Une fois le premier tirage effectué, on a des informations sur X et, plus précisément, on connaît les valeurs

$$\mathbf{E}(X \,|\, Y_1 = 0) \qquad \mathbf{E}(X \,|\, Y_1 = 1) \qquad \mathbf{E}(X \,|\, Y_1 = 2) \qquad \text{et} \qquad \mathbf{E}(X \,|\, Y_1 = 3).$$

Ces quatre valeurs définissent entièrement $M_1 = \mathbf{E}(X \,\|\, Y_1) = \sum_k \mathbf{E}(X \,|\, Y_1 = k)\, \mathbf{1}_{\{Y_1 = k\}}$. Graphiquement, si l'on représente $\mathfrak{F}_1$ par ses atomes et les variables M_0 et M_1 par des niveaux de gris sur Ω, on obtient par exemple :



À mesure que n augmente et que l'on obtient davantage d'informations au travers des valeurs prises par les variables $Y_1, \ldots, Y_n$, on obtient une description de plus en plus fine de la distribution de X, comme dans la séquence ci-dessous. (On n'obtient pas toujours un détail du *Jardin des délices*, cependant.)



§ 354 **Sur- et sous-martingales**

Une suite $(M_n)_{n\geqslant 0}$ est appelée une **sous-martingale** relativement à $(\mathfrak{F}_n)_{n\geqslant 0}$ si, pour tout $n \geqslant 0$:

i) M_n est $\mathfrak{F}_n$-mesurable ;

ii) M_n est intégrable ;

iii) $\mathbf{E}(M_{n+1} \,\|\, \mathfrak{F}_n) \geqslant M_n$.

Ainsi, une sous-martingale est *croissante* en moyenne : $\mathbf{E}(M_{n+1}) \geqslant \mathbf{E}(M_n)$.

De même, une **sur-martingale** vérifie les deux premiers points de la définition précédente, ainsi que l'inégalité $\mathbf{E}(M_{n+1} \,\|\, \mathfrak{F}_n) \leqslant M_n$.

Remarque 23.28 $(M_n)_{n\geqslant 0}$ est une sur-martingale si et seulement si $(-M_n)_{n\geqslant 0}$ est une sous-martingale. ✦ Une suite $(M_n)_{n\geqslant 0}$ est une martingale (respectivement sur- ou sous-martingale) si et seulement si $(M_n - M_0)_{n\geqslant 0}$ en est une. Ainsi, on peut toujours étudier une martingale dont le premier terme est nul.

Si $(M_n)_{n\geqslant 0}$ est une sous-martingale, on a immédiatement

$$\mathbf{E}(M_{n+p} \,\|\, \mathfrak{F}_n) \geqslant M_n$$

pour tous entiers n et p.

§ 355 **Théorème de convergence des martingales**

La propriété la plus frappante des martingales est que, sous des hypothèses extrêmement réduites, on peut montrer qu'elles convergent presque sûrement (et dans $\mathcal{L}^1$ moyennant une hypothèse de plus).

Mais d'abord, un arrêt sur la figure 23.8, qui propose une autre représentation graphique d'une martingale.

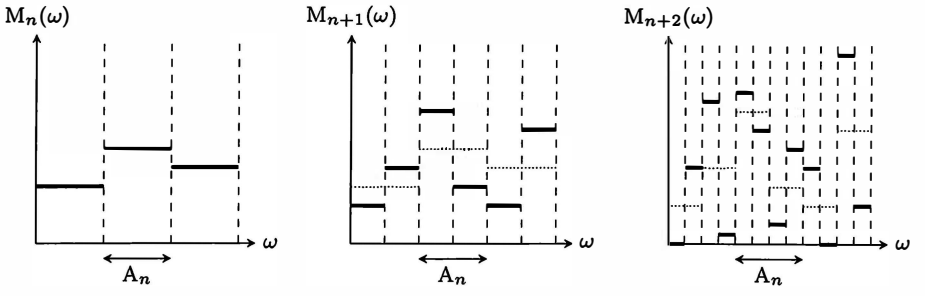


FIGURE 23.8 — *Trois étapes d'une martingale positive. Sur l'événement* A_n *de* $\mathfrak{F}_n$, *les variables* M_n, M_{n+1} *et* M_{n+2} *prennent la même valeur moyenne.*

Lorsque l'on passe d'une étape n à l'étape $n+1$, la variable peut prendre des valeurs très différentes, la seule condition étant que, sur un événement A_n de $\mathfrak{F}_n$, M_n et M_{n+1} aient même valeur moyenne. Toute valeur « plus grande » engendre automatiquement une valeur « plus petite » quelque part sur A_n. Notamment, une simple minoration de la suite $(M_n)_{n\geqslant 0}$ — par exemple une condition de positivité — empêche que les valeurs prises par les M_n ne deviennent arbitrairement grandes trop souvent.

Mieux encore, une telle condition va montrer la convergence presque sûre. Le principe est le suivant. Dire que la suite $(M_n)_{n\geqslant 0}$ converge presque sûrement revient à dire que, presque sûrement, $\liminf M_n = \limsup M_n$. Si l'on note

$$\Lambda := \{\liminf M_n < \limsup M_n\}$$

l'ensemble des ω en lesquels on n'a pas convergence simple, on peut décomposer

$$\Lambda = \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} \Lambda(a,b)$$

où

$$\Lambda(a,b) := \{\liminf M_n \leqslant a < b \leqslant \limsup M_n\}.$$

Que se passe-t-il en un point ω d'un des événements $\Lambda(a,b)$? La suite de terme général $M_n(\omega)$ doit s'approcher une infinité de fois de $\liminf M_n < a$ comme de $\limsup M_n > b$; notamment, si l'on choisit des points a' et b' tels que $a < a' < b' < b$, elle doit effectuer une infinité de « remontées » de a' à b'.

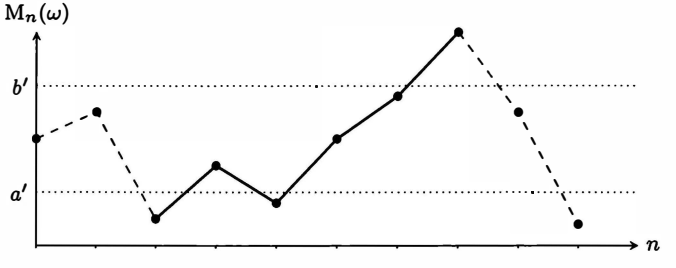


FIGURE 23.9 — *Une remontée de* a' *à* b' *: la suite* $(M_n(\omega))_{n\geqslant 0}$ *monte depuis un point situé sous* a', *et dépasse le point* b'.

Or, une inégalité permet de contrôler le nombre $U_n(a',b')$ de remontées de a' à b' entre les instants 0 et n, pour une martingale ou une sous-martingale :

■ **Théorème 23.29 (Inégalité des remontées de Doob)** *Si* $(M_n)_{n\geqslant 0}$ *est une sous-martingale, alors*

$$\mathbf{E}\big[U_n(a',b')\big] \leqslant \frac{1}{(b'-a')}\mathbf{E}\big[(M_n-a')^+\big]$$

où l'on a noté $(M_n-a')^+ := \max(M_n-a',0)$ *la partie positive de* M_n-a'.

Notamment, en prenant la limite $n\to\infty$ (grâce au théorème de convergence monotone, puisque $U_n(a',b')$ croît avec n), et par l'inégalité $(m-a')^+ \leqslant m^+ + |a'|$, on obtient que $\mathbf{E}\big[U_\infty(a',b')\big] < +\infty$ sous la simple condition que $\mathbf{E}(M_n^+)$ soit majoré par une constante. Cette propriété montre que $U_\infty(a',b')$ est presque sûrement finie, c'est-à-dire que la suite $\big(M_n(\omega)\big)$ ne remonte de a' à b', presque sûrement, qu'un nombre fini de fois ; on en déduit que $\mathbf{P}(\Lambda(a',b')) = 0$ et donc, par inclusion, que $\mathbf{P}(\Lambda(a,b)) = 0$. Enfin, par union dénombrable, on obtient $\mathbf{P}(\Lambda) = 0$. Cela montre le théorème suivant :

■ **Théorème 23.30** *Soit* $(M_n)_{n\geqslant 0}$ *une martingale ou une sous-martingale. On suppose qu'il existe une constante* K *telle que* $\mathbf{E}(M_n^+) \leqslant K$ *pour tout entier* n. *Alors la suite* $(M_n)_{n\geqslant 0}$ *converge presque sûrement vers une variable limite* M *à valeurs réelles, et intégrable.*

Notamment, une martingale bornée dans l'espace L^1 converge presque sûrement.

Remarque 23.31 Le fait que M soit intégrable vient du fait que, si l'on écrit $|M_n| = 2M_n^+ - M_n$, la définition d'une sous-martingale implique que

$$\mathbf{E}\,|M_n| \leqslant 2\,\mathbf{E}(M_n^+) - \mathbf{E}(M_0) \leqslant 2K - \mathbf{E}(M_0) < +\infty$$

puis, par le lemme de Fatou, $\mathbf{E}\,|M| = \mathbf{E}(\lim|M_n|) \leqslant \liminf \mathbf{E}\,|M_n| < +\infty$. Cela entraîne également le fait que M soit presque sûrement finie.

En pratique, on utilise souvent le corollaire suivant, dont les hypothèses sont très faciles à vérifier :

■ **Théorème 23.32 (Convergence de surmartingales positives)**
Toute surmartingale positive (ou toute sous-martingale négative) converge presque sûrement vers une variable limite M *à valeurs réelles, et intégrable.*

DÉMONSTRATION : Si $(M_n)_{n\geqslant 0}$ est une sous-martingale négative, alors $\mathbf{E}(M_n^+) = 0$ et le théorème précédent s'applique. Si $(M_n)_{n\geqslant 0}$ est une surmartingale positive, alors $(-M_n)_{n\geqslant 0}$ est une sous-martingale négative et on est ramenés au cas précédent. ∎

La condition « $\mathbf{E}(M_n^+) \leqslant K$ » est un cas particulier d'une propriété, appelée **intégrabilité uniforme**, que nous n'abordons pas ici[4], et qui est vérifiée notamment dans les deux cas suivants :

- s'il existe une variable aléatoire Y intégrable qui domine $(M_n)_{n\geqslant 0}$: $|M_n| \leqslant Y$ pour tout entier n ;
- s'il existe deux réels $p > 1$ et $K \geqslant 0$ tels que $\mathbf{E}\,|M_n|^p \leqslant K$ pour tout entier n.

(4). Cf. par exemple Williams [100] ou Jacod et Protter [53].

■ **Théorème 23.33 (Convergence des martingales dans L^1)**
Si $(M_n)_{n\geqslant 0}$ est une martingale uniformément intégrable, elle converge presque sûrement vers une variable aléatoire M intégrable ; de plus, la convergence a lieu dans L^1, c'est-à-dire que $\mathbf{E}\,|M_n - M| \to 0$, et $M_n = \mathbf{E}(M \,|\, \mathfrak{F}_n)$ pour tout entier n.

Remarque 23.34 Dans les théorèmes précédents (notamment le théorème 23.30 et ses divers corollaires), la convergence presque sûre de $(M_n)_{n\geqslant 0}$ vers une variable $M \in L^1$ n'impliquait pas la convergence *dans l'espace* L^1. Ici, la condition plus forte d'uniforme intégrabilité permet de conclure à cette convergence.

Applications § 356

Citons ici quelques applications de la théorie des martingales.

- La loi du zéro-un de Kolmogorov peut se prouver aisément à l'aide de martingales.
- La loi des grands nombres de Kolmogorov peut également être vue comme une conséquence d'un des théorèmes de convergence de martingales : si $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes de même loi et admettant une espérance $m = \mathbf{E}(X_1)$, alors la suite de terme général $M_n = (X_1 + X_2 + \cdots + X_n)/n$ converge, presque sûrement et dans L^1, vers m.
- Le théorème de Radon-Nikodým peut se démontrer, entre autres, à l'aide de martingales [100, chapitre 14].
- Le problème des processus de branchement, dont nous avions vu un exemple avec le modèle de Galton-Watson, se traite bien plus facilement en termes de martingales [6].
- Les martingales ont une interprétation naturelle en terme de stratégies ; des outils puissants ont été développés en analyse financière [6].

De nombreux ouvrages sont consacrés aux martingales et à leurs applications (voir les conseils de lecture en fin de chapitre).

23.4 ANNEXE : PROBABILITÉS CONDITIONNELLES (*)

On peut relier les probabilités conditionnelles aux espérances conditionnelles en définissant $\mathbf{P}(A \,\|\, \mathfrak{G}) := \mathbf{E}(\mathbf{1}_A \,\|\, \mathfrak{G})$. Cependant, une construction directe de l'espérance conditionnelle est possible ; la formule précédente en sera une conséquence.

Probabilité conditionnelle sachant une partition dénombrable § 357

On se donne une partition finie $\mathfrak{D} = \{D_1, D_2 \ldots, D_n\}$ telle que $\mathbf{P}(D_k) > 0$ pour $k = 1, \ldots, n$.

Soit $A \in \mathfrak{T}$ un événement. Nous voulons calculer la probabilité de A conditionnée, non par la connaissance particulière de la réalisation d'un des D_k, mais par la connaissance de $\mathfrak{D}$ tout entier. Si c'est D_1 qui est réalisé, la probabilité qui nous intéresse est $\mathbf{P}(A \,|\, D_1)$; si l'on sait que D_2 est réalisé, c'est $\mathbf{P}(A \,|\, D_2)$ etc. On notera donc

$$\forall \omega \in \Omega \qquad \mathbf{P}(A \,\|\, \mathfrak{D})(\omega) := \begin{cases} \mathbf{P}(A \,|\, D_1) & \text{si } \omega \in D_1, \\ \quad\vdots & \quad\vdots \\ \mathbf{P}(A \,|\, D_n) & \text{si } \omega \in D_n. \end{cases}$$

De manière plus concise :

$$\mathbf{P}(A \,\|\, \mathfrak{D}) = \sum_{k=1}^{n} \mathbf{P}(A \,|\, D_k)\, \mathbf{1}_{D_k}. \tag{23.11}$$

Notons que $\mathbf{P}(A \,\|\, \mathfrak{D})$ *est une variable aléatoire*, qu'on appelle **probabilité conditionnelle de A par rapport à $\mathfrak{D}$**, ou **sachant $\mathfrak{D}$**. Cette variable est constante sur chacun des atomes D_k de $\mathfrak{D}$; elle est donc *mesurable* par rapport à la tribu $\sigma(\mathfrak{D})$ engendrée par $\mathfrak{D}$.

D'après sa définition même, la probabilité conditionnelle sachant $\mathfrak{D}$ est additive : si A et B sont deux événements incompatibles, alors

$$\mathbf{P}(A + B \,\|\, \mathfrak{D}) = \mathbf{P}(A \,\|\, \mathfrak{D}) + \mathbf{P}(A \,\|\, \mathfrak{D}),$$

et on vérifie sans peine qu'elle est σ-additive.

Si $\mathfrak{D} = \{\Omega\}$ (partition triviale), alors $\mathbf{P}(A \,\|\, \Omega) = \mathbf{P}(A)$ (fonction constante). Ensuite, et c'est une propriété importante, on peut évaluer, par linéarité, l'espérance de $\mathbf{P}(A \,\|\, \mathfrak{D})$ écrite sous la forme (23.11) :

$$\mathbf{E}\big(\mathbf{P}(A \,\|\, \mathfrak{D})\big) = \sum_{k=1}^{n} \mathbf{P}(A \,|\, D_k)\, \mathbf{E}\left(\mathbf{1}_{D_k}\right) = \sum_{k=1}^{n} \mathbf{P}(A \,|\, D_k)\, \mathbf{P}(D_k) = \mathbf{P}(A) \qquad (*)$$

d'après la formule des probabilités totales.

Enfin, ces définitions se généralisent sans problème au cas où $\mathfrak{D} = \{D_k \,;\, k \in \mathbb{N}\}$ est une partition dénombrable vérifiant $\mathbf{P}(D_k) > 0$ pour tout entier k. Comme il n'est pas rare d'avoir à utiliser des partitions pour lesquelles $\mathbf{P}(D_k)$ est nulle pour certains indices k, on définit, pour tout $\omega \in \Omega$:

$$\mathbf{P}(A \,\|\, \mathfrak{D})(\omega) = \begin{cases} \mathbf{P}(A \,|\, D_k) & \text{si } \omega \in D_k \text{ et } \mathbf{P}(D_k) > 0 \\ \textit{une valeur arbitraire de } [0\,;1] & \text{si } \omega \in D_k \text{ et } \mathbf{P}(D_k) = 0. \end{cases}$$

La variable aléatoire $\mathbf{P}(A \,\|\, \mathfrak{D})$ est donc définie de manière explicite sur presque tout Ω, et de manière arbitraire sur un ensemble N négligeable, à savoir $N = \bigcup_{k \in K} D_k$, l'union portant sur l'ensemble K des indices tels que $\mathbf{P}(D_k) = 0$. La propriété $(*)$ reste encore valable, les valeurs arbitraires étant toutes écrasées par un facteur $\mathbf{P}(D_k) = 0$. Ainsi, la véritable nature de $\mathbf{P}(A \,\|\, \mathfrak{D})$ n'est pas d'être une variable aléatoire, mais une classe d'équivalence de variables aléatoires à une égalité presque sûre près. Choisir *un* représentant particulier s'appelle choisir une **version** de $\mathbf{P}(A \,\|\, \mathfrak{D})$.

■ **Proposition 23.35 (Espérance d'une probabilité conditionnelle)**
Soient $\mathfrak{D}$ une partition dénombrable de Ω et A un événement, alors $\mathbf{E}\big(\mathbf{P}(A \,\|\, \mathfrak{D})\big) = \mathbf{P}(A)$.

Enfin, on remarquera que, si $A \in \mathfrak{T}$ est un événement, alors $\mathbf{E}(\mathbf{1}_A \,\|\, \mathfrak{D}) = \mathbf{P}(A \,\|\, \mathfrak{D})$, ce qui affermit encore les liens entre probabilités conditionnelles et espérances conditionnelles.

§ 358 **Probabilité conditionnelle sachant une variable aléatoire**
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire discrète, prenant un ensemble de valeurs fini ou dénombrable que l'on notera $\{x_k \,;\, k \in K\}$. On note $\mathfrak{D}_X$ le système complet d'événements engendré par X :

$$\mathfrak{D}_X = \big\{\{X = x_k\} \,;\, k \in K\big\},$$

qui est donc une partition de Ω. Si A est un événement, on définit la **probabilité conditionnelle de A sachant X** comme la variable aléatoire

$$\mathbf{P}(A \,\|\, X) := \mathbf{P}(A \,\|\, \mathfrak{D}_X) = \sum_{k=0}^{\infty} \mathbf{P}(A \,|\, X = x_k)\, \mathbf{1}_{\{X = x_k\}}.$$

Encore une fois, on doit lire dans cette équation la définition d'une variable aléatoire *à une égalité presque partout près*, certains événements $\{X = x_k\}$ pouvant être négligeables.

Enfin, si $X_1, X_2, \ldots, X_d$ sont des variables aléatoires discrètes, le système complet d'événements engendrés par $X_1, X_2, \ldots, X_d$ est

$$\mathfrak{D}_{X_1,\ldots,X_d} = \big\{\{X_1 = x_{k_1}^1\} \cap \cdots \cap \{X_d = x_{k_d}^d\} \,;\, (k_1, \ldots, k_d) \in \mathbb{N}^d\big\},$$

où l'on a noté $\{x_k^i \,;\, k \in \mathbb{N}\}$ l'ensemble des valeurs prises par X_i. On pose alors

$$\mathbf{P}(A \,\|\, X_1, \ldots, X_d) := \mathbf{P}(A \,\|\, \mathfrak{D}_{X_1,\ldots,X_d})$$

à une égalité presque sûre près.

§ 359 **Probabilité conditionnelle sachant une tribu**
Soient $\mathfrak{G}$ une sous-tribu de $\mathfrak{T}$ et $A \in \mathfrak{T}$ un événement. On peut définir une mesure finie μ_A sur l'espace mesurable $(\Omega, \mathfrak{G})$ par la formule

$$\forall G \in \mathfrak{G} \qquad \mu_A(G) := \mathbf{P}(G \cap A).$$

La mesure finie μ_A est absolument continue par rapport à la mesure $\mathbf{P}$[5] ; le théorème de Radon-Nikodým assure qu'il existe une fonction $f : \Omega \to \mathbb{R}$ qui soit $\mathfrak{G}$-mesurable et intégrable (par rapport à $\mathbf{P}$) telle que

$$\forall G \in \mathfrak{G} \qquad \mu_A(G) = \int_G f \, d\mathbf{P}.$$

---

(5). Plus exactement, elle est absolument continue par rapport à la mesure $\mathbf{P}'$, restriction de $\mathbf{P}$ à l'espace $(\Omega, \mathfrak{G})$: pour tout $G \in \mathfrak{G}$ tel que $\mathbf{P}(G) = 0$, on a $\mu_A(G) = 0$.

Cette fonction est appelée **probabilité conditionnelle de A sachant la tribu $\mathcal{G}$**, on la note $\mathbf{P}(A \,\|\, \mathcal{G})$.

Par construction, $\mathbf{P}(A \,\|\, \mathcal{G})$ est donc une variable aléatoire $\mathcal{G}$-mesurable, vérifiant

$$\forall G \in \mathcal{G} \qquad \int_G \mathbf{P}(A \,\|\, \mathcal{G})\, d\mathbf{P} = \int_G \mathbf{1}_A \, d\mathbf{P} \tag{23.12}$$

le membre de droite n'étant qu'une réécriture de $\mathbf{P}(G \cap A)$. Il est crucial de noter que cette relation *caractérise* la probabilité conditionnelle sachant $\mathcal{G}$. Plus précisément, si l'on exhibe une variable aléatoire $\mathcal{G}$-mesurable Z vérifiant

$$\forall G \in \mathcal{G} \qquad \int_G Z \, d\mathbf{P} = \int_G \mathbf{1}_A \, d\mathbf{P},$$

alors cette variable est presque sûrement l'espérance conditionnelle : $Z = \mathbf{P}(A \,\|\, \mathcal{G})$ presque sûrement. C'est une limitation naturelle du théorème de Radon-Nikodým, montrant que, comme dans le cas fini, une probabilité conditionnelle n'est pas tant une variable aléatoire qu'une classe d'équivalence de variables aléatoires (à une égalité p.s. près). Le choix d'un représentant de cette classe d'équivalence s'appelle une **version** de la probabilité conditionnelle.

Remarque 23.36 Bien que f puisse prendre ses valeurs dans $\overline{\mathbb{R}^+} = \mathbb{R}^+ \cup \{+\infty\}$, l'ensemble $\{f = +\infty\}$ est de mesure nulle, puisque f est intégrable. Ainsi, on peut toujours décider de choisir une version de f qui ne prenne que des valeurs finies sur Ω.

Le lecteur est invité à vérifier que, si $\mathcal{G}$ est la tribu engendrée par une partition finie ou dénombrable $\mathcal{D} = \{D_k \,;\, k \in K\}$, alors la définition (23.11) de $\mathbf{P}(A \,\|\, \mathcal{D})$ vérifie la relation (23.12), ce qui montre que la probabilité sachant la partition $\mathcal{D}$ coïncide (presque sûrement) avec la probabilité sachant $\mathcal{G} = \sigma(\mathcal{D})$.

Enfin, la relation (23.12) montre que

$$\mathbf{P}(A \,\|\, \mathcal{G}) = \mathbf{E}(\mathbf{1}_A \,\|\, \mathcal{G}),$$

ce qui permet de définir la notion de probabilité conditionnelle directement à partir de celle d'espérance conditionnelle (et, notamment, si l'on aime le point de vue hilbertien, de s'affranchir du théorème de Radon-Nikodým).

EXERCICES

♦ ***Exercice 23.1 (Quelques cas élémentaires)***
En guise d'échauffement : sur l'univers $\Omega = [0\,;\,1]$ muni de sa tribu des boréliens et de la probabilité uniforme, calculer $Z := \mathbf{E}(Y \,\|\, X)$ dans les cas suivants :

i)	$X_1(\omega) = 1$,	$Y_1(\omega) = \sin \pi\omega$;
ii)	$X_2(\omega) = \mathbf{1}_{[0\,;\,1/2]}(\omega)$,	$Y_2(\omega) = \cos \pi\omega$;
iii)	$X_3(\omega) = \omega^3$,	$Y_3(\omega) = \sin \pi\omega$
iv)	$X_4(\omega) = 2\omega \wedge 1$,	$Y_4(\omega) = \sin \pi\omega$;
v)	$X_5(\omega) = (\omega - \frac{1}{2})^2$,	$Y_5(\omega) = \cos \pi\omega$.

♦ ***Exercice 23.2 (Espérance sachant un événement)*** Soit (X, Y) un couple de variables aléatoires de densité conjointe $f(x, y) = e^{-x-y}$ si $x, y \geqslant 0$ et $f(x, y) = 0$ sinon. Calculer l'espérance conditionnelle de $X + Y$ sachant l'événement $\{X < Y\}$.

♦ ***Exercice 23.3 (Espérance conditionnelle sachant la partie entière)***
Le calcul explicite de l'espérance conditionnelle d'une variable aléatoire par rapport à une autre est souvent difficile ! Voici un exemple où il se calcule aisément.

Soit X une variable aléatoire de loi exponentielle $\mathcal{E}(1)$. On note $Y := \lfloor X \rfloor$ et $Z := X - \lfloor X \rfloor$.

i) Calculer l'espérance conditionnelle $\mathbf{E}(X \,\|\, Y)$.

ii) Montrer que Y et Z sont indépendantes.

iii) En déduire $\mathbf{E}(X \,\|\, Z)$.

Indications

◊ *Indication 23.2 :* Si A est un événement de probabilité non nulle, $\mathbf{E}(Z \,|\, A) = \mathbf{E}(Z\,\mathbf{1}_A)/\mathbf{P}(A)$.

◊ *Indication 23.3 :* Pour la troisième question, utiliser la relation $\mathbf{E}(X \,\|\, Z) = \mathbf{E}(Y \,\|\, Z) + \mathbf{E}(Z \,\|\, Z)$. Le premier terme se calcule grâce à l'indépendance de Y et de Z ; le second n'est autre que Z.

SOLUTIONS

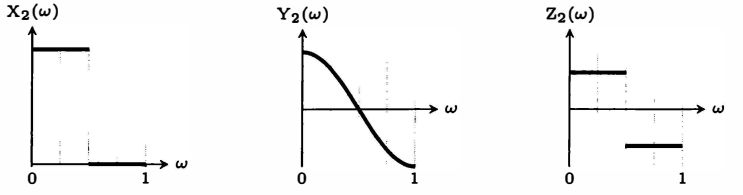
♦ **Solution 23.1** Pour chaque cas, nous donnons un argument heuristique, suivi d'une solution rigoureuse.

i) Il est clair que connaître la valeur de X_1 ne renseigne *en rien* sur la valeur de ω, et donc sur celle de $Y_1(\omega)$; on s'attend donc à ce que $\mathbf{E}(Y_1 \| X_1) = \mathbf{E}(Y_1) = 2/\pi$.

Pour le prouver, il suffit de noter que $\sigma(X_1) = \{\varnothing, \Omega\}$ est la tribu grossière, c'est-à-dire que $[0\,;\,1]$ tout entier est un atome de $\sigma(X_1)$; la variable $\mathbf{E}(Y_1 \| X_1)$, qui doit être $\sigma(X_1)$-mesurable, est donc constante.

ii) Si l'on sait que $X_2(\omega) = 1$, alors cela signifie que $0 \leqslant \omega \leqslant 1/2$; conditionnellement à l'événement $[0\,;\,1/2]$, la variable aléatoire Y_2 a pour valeur moyenne $2/\pi$. Si au contraire on sait que $X_2(\omega) = 0$, alors $\omega \in]1/2\,;\,1]$; conditionnellement à cet événement, Y_2 a pour valeur moyenne $-2/\pi$.

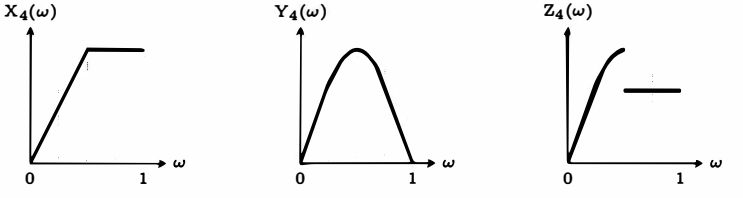
Plus formellement, la tribu engendrée par X_2 est $\{\varnothing, [0\,;\,\frac{1}{2}],]\frac{1}{2}\,;\,1], [0\,;\,1]\}$. Ainsi, la variable Z_2 est constante sur chacun des intervalles $[0\,;\,\frac{1}{2}]$ et $]\frac{1}{2}\,;\,1]$, et y prend les valeurs $\mathbf{E}(Y_2 \,|\, X_2 = 1)$ et $\mathbf{E}(Y_2 \,|\, X_2 = 0)$ respectivement. On peut résumer ces deux résultats par la formule $Z_2 = \mathbf{E}(Y_2 \| X_2) = \frac{4}{\pi}(X_2 - \frac{1}{2})$.



iii) Cette fois-ci, connaître $X(\omega)$ permet de retrouver précisément la valeur de ω, et donc celle de ω ; en d'autres termes, Y est X-mesurable. On en déduit que $\mathbf{E}(Y \| X) = Y$. On peut donner un autre argument : la tribu engendrée par X est précisément la tribu des boréliens de $[0\,;\,1]$, donc la variable Y vérifie bien la propriété fondamentale (23.8).

iv) Si la valeur de X est égale à 1, tout ce qu'on sait, c'est que ω est compris entre $1/2$ et 1. Conditionnellement à cet événement, la variable Y prend donc la valeur moyenne $2/\pi$. Si la valeur de X est dans $[0\,;\,1[$, alors on peut en déduire précisément la valeur de ω par la relation $\omega = X(\omega)/2$, et donc la valeur de Y. Ainsi, $\mathbf{E}(Y \| X)$ vaut Y sur $[0\,;\,1/2[$, et $2/\pi$ sur $[1/2\,;\,1]$:

$$\mathbf{E}(Y \,|\, X) = Y \cdot \mathbf{1}_{[0\,;\,1/2[} + \tfrac{2}{\pi} \cdot \mathbf{1}_{[1/2\,;\,1]}.$$



v) Cette fois-ci, connaître X_5 ne permet de connaître ω qu'à la symétrie $\omega \mapsto 1 - \omega$ près. Or, Y prend des valeurs opposées sur les points ω et $1 - \omega$. On en déduit que $\mathbf{E}(Y_5 \| X_5) = 0$. On peut le démontrer plus formellement : la tribu engendrée par X_5 est celle des boréliens G tels que $1 - G = G$. Pour un tel borélien G, on a bien $\mathbf{E}(Y_5 \cdot \mathbf{1}_G) = 0 = \mathbf{E}(0 \cdot \mathbf{1}_G)$, et la fonction nulle est bien X_5-mesurable.

♦ **Solution 23.2** Par symétrie ou calcul direct, l'événement $A = \{X < Y\}$ est de probabilité $1/2$. L'espérance de $Z := X + Y$ sachant A vaut (équation (23.6)) : $\mathbf{E}(Z \,|\, A) = \mathbf{E}(Z\mathbf{1}_A)/\mathbf{P}(A)$. Le calcul est ensuite immédiat :

$$\begin{aligned}\mathbf{E}(Z\mathbf{1}_A) &= \iint_{y>x} (x+y)\, f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^{+\infty} \left(\int_x^{+\infty} (x+y)\, \mathrm{e}^{-x-y}\,\mathrm{d}y \right) \mathrm{d}x \\ &= \int_0^{+\infty} (2x+1)\, \mathrm{e}^{-2x} = 1,\end{aligned}$$

et finalement $\mathbf{E}(X + Y \,|\, X < Y) = 2$.

♦ **Solution 23.3**

i) Puisque Y est une variable discrète, on peut écrire $\mathbf{E}(X \,\|\, Y) = \sum_{k=0}^{\infty} \mathbf{E}(X \mid Y = k)\, \mathbf{1}_{\{Y=k\}}$. Or,

$$\mathbf{E}(X \mid Y = k) = \frac{\mathbf{E}(X \cdot \mathbf{1}_{\{Y=k\}})}{\mathbf{P}\{Y = k\}} = \frac{1}{\mathbf{P}\{Y = k\}} \int_{\{Y=k\}} X \, d\mathbf{P} = \frac{1}{\mathbf{P}\{Y = k\}} \int_k^{k+1} x\, f(x)\, dx,$$

où l'on a noté $f(x) = e^{-x}$ la densité de X. Ainsi,

$$\mathbf{E}(X \mid Y = k) = \frac{\int_k^{k+1} x\, e^{-x}\, dx}{\int_k^{k+1} e^{-x}\, dx} = k + \alpha,$$

où $\alpha := (e - 2) / (e - 1)$ est une constante indépendante de k. Ainsi,

$$\mathbf{E}(X \,\|\, Y) = \sum_{k=0}^{\infty} (k + \alpha)\, \mathbf{1}_{\{Y=k\}} = Y + \alpha.$$

On remarquera que $\mathbf{E}(X \,\|\, Y)$ est effectivement Y-mesurable, puisqu'on a pu l'écrire comme une fonction (borélienne) de Y.

ii) Pour montrer l'indépendance de Y et Z, il suffit de montrer que, pour tout entier k et tout réel $t \in [0\,;\, 1[$, on a

$$\mathbf{P}\{Y = k, Z \leqslant t\} = \mathbf{P}\{Y = k\}\, \mathbf{P}\{Z \leqslant t\}.$$

Or,

$$\mathbf{P}\{Y = k, Z \leqslant t\} = \mathbf{P}\{k \leqslant X \leqslant k + t\} = \int_k^{k+t} e^{-x}\, dx = e^{-k}(1 - e^{-t}),$$

tandis que

$$\mathbf{P}\{Y = k\} = \int_k^{k+1} e^{-x}\, dx = e^{-k}(1 - e^{-1})$$

et

$$\mathbf{P}\{Z \leqslant t\} = \sum_{k=0}^{\infty} \mathbf{P}\{k \leqslant X \leqslant k + t\} = \sum_{k=0}^{\infty} e^{-k}(1 - e^{-t}) = \frac{1 - e^{-t}}{1 - e^{-1}},$$

ce qui est conforme à nos attentes.

iii) Tout d'abord, par linéarité, $\mathbf{E}(X \,\|\, Z) = \mathbf{E}(Y \,\|\, Z) + \mathbf{E}(Z \,\|\, Z)$. Par indépendance de Y et Z, $\mathbf{E}(Y \,\|\, Z) = \mathbf{E}(Y)$; de plus, $\mathbf{E}(Z \,\|\, Z) = Z$. On en conclut que

$$\mathbf{E}(X \,\|\, Z) = \mathbf{E}(Y) + Z = \frac{1}{e - 1} + Z.$$

Là encore, la fonction trouvée est évidemment Z-mesurable.

Conseils de lecture

Concernant les martingales, une référence très abordable et proposant de nombreuses applications est :

- Michel Benaïm et Nicole El Karoui, *Promenade aléatoire — Chaînes de Markov et simulations ; martingales et stratégies* [6].

Concernant la théorie des martingales et de l'espérance conditionnelle, on trouvera une présentation synthétique (réduite à l'essentiel) ainsi que les démonstrations des principaux théorèmes dans deux ouvrages :

- David Williams, *Probability with martingales* [100].
- Jean Jacod et Philip Protter, *L'essentiel en théorie des probabilités* [53].

Cinquième partie

Quelques applications

Chapitre 24

Quelques résultats élégamment établis par la théorie des probabilités

Dans ce chapitre, nous présentons quelques résultats d'analyse démontrés avec des méthodes, des théorèmes ou des idées provenant du monde probabiliste.

- *La formule de Bernstein est une formule découlant rapidement du théorème limite central, mais plus difficile à démontrer avec des outils élémentaires d'analyse.*
- *Le même théorème limite central permet une démonstration relativement élémentaire de la formule de Stirling.*
- *Deux théorèmes (§ 361 et § 362) sont représentatifs de la* méthode probabiliste, *historiquement introduite par Erdős. Il s'agit de prouver l'existence (certaine !) d'un objet possédant telle propriété, en montrant par exemple que la mesure de l'ensemble des objets possédant cette propriété est strictement positive. Le langage même de l'aléatoire pourrait d'ailleurs être totalement gommé au profit de celui de* mesure *— mais l'intuition initiale étant probabiliste, ne boudons pas notre plaisir.*
- *Le théorème de Weierstrass sur la densité de l'espace des fonctions polynomiales a reçu, au début du siècle, une démonstration particulièrement simple et élégante par Bernstein ; c'est l'occasion de voir l'apparition très naturelle des polynômes du même nom.*

La formule de Bernstein

§ 360

Un exemple des plus célèbres d'application « non probabiliste » des probabilités est la **formule de Bernstein** :

$$\lim_{n\to+\infty} \mathrm{e}^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \frac{1}{2}.$$

Ce n'est pas à vrai dire une application tellement impressionnante des probabilités : peu de mathématiciens se réveillent le matin en se disant « Tiens, j'aimerais bien démontrer cette formule » et finissent pas se dire que les probabilités peuvent les

y aider. C'est plutôt un exemple de formule qui est donné, presque gratuitement et avec un minimum de calculs, en appliquant le théorème limite central dans une situation élémentaire.

Pour prouver cette formule, la première chose à faire est d'interpréter la quantité

$$u_n = \mathrm{e}^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}$$

comme une probabilité ; la forme de u_n suggère l'utilisation d'une loi de Poisson de paramètre n. Si X_n est une variable aléatoire de loi $\mathscr{P}(n)$, on peut en effet écrire

$$u_n = \mathbf{P}\{\mathrm{X}_n = 0\} + \mathbf{P}\{\mathrm{X}_n = 1\} + \cdots + \mathbf{P}\{\mathrm{X}_n = n\} = \mathbf{P}\{\mathrm{X}_n \leqslant n\}.$$

Or, une loi de Poisson de paramètre n peut être vue comme la somme de n lois de Poisson indépendantes de paramètre 1 : si $\mathrm{Y}_1, \ldots, \mathrm{Y}_n$ sont des variables indépendantes suivant la même loi $\mathscr{P}(1)$, la variable

$$\mathrm{X}'_n = \mathrm{Y}_1 + \mathrm{Y}_2 + \cdots + \mathrm{Y}_n$$

suit la loi $\mathscr{P}(n)$. Invoquons maintenant le théorème limite central; pour cela, on construit comme d'habitude la suite de variables aléatoires centrées et réduite, en posant

$$\mathrm{Z}_n = \frac{\mathrm{X}'_n - \mathbf{E}(\mathrm{X}'_n)}{\sigma_{\mathrm{X}'_n}} = \frac{\mathrm{X}'_n - n}{\sqrt{n}},$$

et il ne reste plus qu'à remarquer que

$$u_n = \mathbf{P}\{\mathrm{X}'_n \leqslant n\} = \mathbf{P}\{\mathrm{Z}_n \leqslant 0\}$$

⏎24.1 et que cette quantité converge vers $\Pi(0) = \frac{1}{2}$ d'après le théorème limite central.

Remarque 24.1 (Démonstration sans probabilités) On peut tout à fait prouver ce résultat en n'utilisant que des outils standards d'analyse, ne requérant qu'un niveau de Mathématiques Spéciales ou de deuxième année de Licence. Tout d'abord, interprétons la quantité u_n comme le polynôme de Taylor d'ordre n associé à la fonction $f = \exp$: pour tous $x, t \in \mathbb{R}$, on a

$$\mathrm{e}^{-x+t} = f(-x+t) = \sum_{k=0}^{n} \frac{f^{(k)}(-x)}{k!}\, t^k + \int_{-x}^{-x+t} \frac{(-x+t-s)^n}{n!}\, \mathrm{e}^s \, \mathrm{d}s$$

c'est-à-dire, en prenant $x = t = n$:

$$1 = u_n + \int_{-n}^{0} \frac{(-s)^n}{n!}\, \mathrm{e}^s \, \mathrm{d}s = u_n + \mathrm{e}^{-n} \int_0^n \frac{(n-v)^n}{n!}\, \mathrm{e}^v \, \mathrm{d}v = u_n + \mathrm{R}_n.$$

On doit donc prouver que la suite $(\mathrm{R}_n)_{n \geqslant 1}$ converge et a pour limite $1/2$. Pour cela, modifions son expression en effectuant le changement de variable $y = v/n$:

$$\mathrm{R}_n = \mathrm{e}^{-n} \int_0^1 \frac{n^n (1-y)^n}{n!}\, \mathrm{e}^{ny}\, n \, \mathrm{d}y = \underbrace{n \frac{\mathrm{e}^{-n} n^n}{n!}}_{\alpha_n} \underbrace{\int_0^1 (1-y)^n \, \mathrm{e}^{ny} \, \mathrm{d}y}_{\mathrm{I}_n} = \alpha_n \, \mathrm{I}_n.$$

L'équivalent de Stirling $n! \sim n^n \, \mathrm{e}^{-n} \sqrt{2\pi n}$ nous fournit $\alpha_n \sim \sqrt{n/2\pi}$. Quant à l'intégrale I_n, écrivons-la sous la forme

$$\mathrm{I}_n = \int_0^1 \mathrm{e}^{n\varphi(u)} \, \mathrm{d}u \qquad \text{avec} \qquad \varphi(u) = \ln(1-u) + u.$$

Il ne reste plus qu'à déterminer un équivalent de I_n ; pour cela, la méthode de Laplace est tout indiquée[1] puisque φ admet un maximum en 0 : $\varphi(0) = 0$, $\varphi'(0) = 0$ et $\varphi''(0) = -1$. On

(1). En prenant garde qu'un facteur 1/2 se rajoute à la formule habituelle $\mathrm{I}_n \sim \mathrm{e}^{n\varphi(0)} \sqrt{-2\pi/n\,\varphi''(0)}$, car le maximum est atteint au bord de l'intervalle : on n'a donc que la moitié de la contribution gaussienne. Le lecteur pourra trouver des explications sur la méthode de Laplace dans de nombreux ouvrages d'analyse.

décompose

$$\varphi(u) = -\frac{u^2}{2} - u^3\, w(u),$$

où w est une fonction bornée au voisinage de 0 et positive sur $[0\,;1[$, et on compare I_n à la contribution principale, qui est l'intégrale $\int_0^1 e^{-nu^2/2}\,du$. Pour commencer, on isole le voisinage de 1 (où w diverge) : on pose par exemple $\alpha := 1 - 1/2e$, et l'on majore

$$I_n'' := \int_\alpha^1 (1-y)^n\, e^{ny}\, du \leqslant \frac{1-\alpha}{2^n}$$

donc $\alpha_n\, I_n'' \to 0$ quand $n \to \infty$. Si l'on note $I_n' := \int_0^\alpha e^{n\varphi(u)}\,du$, alors

$$\left| I_n' - \int_0^\alpha e^{-nu^2/2}\,du \right| = \left| \int_0^1 e^{-nu^2/2} \left[e^{-nu^3\, w(u)} - 1 \right] du \right|$$

On note M un majorant de w est bornée sur $[0\,;\alpha]$; alors $\left| e^{-nu^3\, w(u)} - 1 \right| \leqslant Mn\,u^3$, et donc

$$\begin{aligned} \left| I_n' - \int_0^\alpha e^{-nu^2/2}\,du \right| &\leqslant Mn \int_0^\alpha e^{-nu^2/2}\, u^3\,du \\ &\leqslant \frac{M}{n} \int_0^{\alpha\sqrt{n}} e^{-y^2/2}\, y^3\,dy \leqslant \frac{M}{n} \int_0^{+\infty} e^{-y^2/2}\, y^3\,dy = \frac{K}{n}. \end{aligned}$$

Enfin,

$$\int_0^\alpha e^{-nu^2/2}\,du = \frac{1}{\sqrt{n}} \int_0^{\sqrt{n}\alpha} e^{-y^2/2}\,dy \underset{n\to\infty}{\sim} \frac{1}{\sqrt{n}} \int_0^{+\infty} e^{-y^2/2}\,dy = \frac{\sqrt{\pi}}{\sqrt{2n}},$$

ce qui montre finalement que $\alpha_n\, I_n' \to 1/2$ et donc que $R_n \to 1/2$, ce qui achève la démonstration.

Remarque 24.2 (Sur la formule de Stirling) Nous avons fait usage de la formule de Stirling :

$$n! \underset{n\to\infty}{\sim} \left(\frac{n}{e}\right)^n \sqrt{2\pi n},$$

dont une démonstration élémentaire est la suivante. On pose pour tout n entier :

$$a_n := n!\,(e/n)^n/\sqrt{n} \qquad \text{et} \qquad u_n := \ln(a_{n+1}/a_n).$$

On vérifie que $u_n = O(1/n^2)$, ce qui entraîne que la série $\sum u_n$ converge et donc que la suite $(a_n)_{n\geqslant 1}$ converge vers une limite ℓ strictement positive. Pour montrer que cette limite vaut $\ell = \sqrt{2\pi}$, on peut utiliser astucieusement les intégrales de Wallis et leur équivalent connu.

Voici une approche probabiliste. Considérons une suite $(X_n)_{n\geqslant 1}$ de variable indépendantes, de même loi de Poisson $\mathscr{P}(1)$. Notons $S_n = X_1 + X_2 + \cdots + X_n$ et

$$S_n^* = \frac{X_1 + X_2 + \cdots + X_n - n}{\sqrt{n}}$$

la variable centrée réduite correspondante. D'après le théorème limite central, la suite $(S_n^*)_{n\geqslant 1}$ converge en loi vers la loi normale $\mathscr{N}(0\,;\,1)$. Notons $\varphi : x \mapsto \max(0, -x)$ la fonction « partie négative ». C'est une fonction continue, mais hélas non bornée, ce qui ne permet pas d'écrire directement que : $\varphi(S_n^*) \xrightarrow{L} \varphi(N)$, où N est la loi normale centrée réduite. Des majorations fines permettent cependant de montrer ce résultat somme toute intuitif[2] :

■ **Lemme 24.3** *L'espérance de $\varphi(S_n^*)$ converge vers l'espérance de $\varphi(N)$:*

$$\mathbf{E}\varphi(S_n^*) \xrightarrow[n\to\infty]{} \mathbf{E}\varphi(N) = \frac{1}{\sqrt{2\pi}}.$$

Or, on sait que la variable S_n suit une loi de Poisson $\mathscr{P}(n)$, c'est-à-dire qu'elle prend la valeur entière $k \geqslant 0$ avec la probabilité

$$P_n(k) := \mathbf{P}\{S_n = k\} = e^{-n}\,\frac{n^k}{k!}.$$

---

[2]. L'idée est d'approcher φ par la suite des fonctions définies par $\varphi_n(x) = -x$ si $-n \leqslant x \leqslant 0$ et $\varphi_n(x) = 0$ sinon.

Dans le tableau ci-dessous, on résume les valeurs possiblement prises par S_n^* et par $\varphi(S_n^*)$, et les probabilités associées :

valeur de S_n	0	1	$\cdots$	$n-1$	n	$n+1$	$\cdots$
valeur de S_n^*	$\frac{-n}{\sqrt{n}}$	$\frac{-n+1}{\sqrt{n}}$		$\frac{-1}{\sqrt{n}}$	0	$\frac{1}{\sqrt{n}}$	
Probabilité	$P_n(0)$	$P_n(1)$	$\cdots$	$P_n(n-1)$	$P_n(n)$	$P_n(n+1)$	$\cdots$
valeur de $\varphi(S_n^*)$	$\frac{n}{\sqrt{n}}$	$\frac{n-1}{\sqrt{n}}$		$\frac{1}{\sqrt{n}}$	0		
Probabilité	$P_n(0)$	$P_n(1)$	$\cdots$	$P_n(n-1)$	C^{te}		

On en déduit que l'espérance de $\varphi(S_n^*)$ vaut

$$\mathbf{E}\varphi(S_n^*) = \sum_{k=0}^{n} \frac{n-k}{\sqrt{n}}\, e^{-n}\, \frac{n^k}{k!} = \frac{e^{-n}}{\sqrt{n}} \left[\sum_{k=0}^{n} \frac{n^{k+1}}{k!} - \sum_{k=1}^{n} \frac{n^k}{(k-1)!} \right]$$

$$= \frac{e^{-n}}{\sqrt{n}} \cdot \frac{n^{n+1}}{n!} \qquad \textit{après télescopage}$$

et cette quantité convergeant vers $1/\sqrt{2\pi}$, on en déduit l'équivalent de Stirling.

§ 361 L'existence de nombres complètement normaux

Dans le chapitre 9, on a défini un nombre **normal en base d** comme un nombre x dont le développement propre[3] en base d :

$$x = \lfloor x \rfloor + \sum_{n=1}^{\infty} \frac{a_n}{d^n}$$

est statistiquement équilibré :

$$\mathscr{P}_i(x) := \lim_{n\to\infty} \frac{1}{n} \operatorname{Card}\left\{ k \in [\![1\,;n]\!] \;;\; a_k = i \right\} = \frac{1}{d}$$

pour $i = 0, \ldots, d-1$. Le réel $\mathscr{P}_i(x)$, s'il existe, est la *proportion* de coefficients égaux à i dans le développement de x. Il est très facile de fabriquer des nombres normaux, par exemple en posant, en base $d = 10$,

$$x := 0,123456789012345678901234\cdots.$$

Un réel x est **complètement normal en base d** si, quels que soient l'entier N et le motif $B = b_1 b_2 \cdots b_N$ formé en concaténant des entiers de $\{0, 1, \ldots, d-1\}$, le proportion de motifs B

$$\mathscr{P}_B(x) := \lim_{n\to\infty} \frac{1}{n} \operatorname{Card}\left\{ k \in [\![1\,;n]\!] \;;\; (a_k, a_{k+1}, \ldots, a_{k+N}) = (b_1, b_2, \ldots, b_N) \right\}$$

existe et vaut $1/d^N$. On connaît de tels nombres, comme le *nombre de Champernowne*

$$x = 0,1234567891011121314151617181920212223\cdots$$

---

(3). Les nombres de la forme k/d^n où k est entier sont les seuls à admettre deux développements distincts en base d. L'un d'eux est fini, on l'appelle **développement propre**. Par exemple, en base 10, les nombres

$$0,42 = 0 + \frac{4}{10} + \frac{2}{100} \qquad \text{et} \qquad 0,41999\cdots = 0 + \frac{4}{10} + \frac{1}{100} + \sum_{n=3}^{\infty} \frac{9}{10^n}$$

sont égaux.

qui est complètement normal en base 10.

Pour terminer, un nombre est **complètement normal** s'il est complètement normal en base d, pour toute valeur de l'entier $d \geqslant 2$. À l'heure actuelle,

- on ne connaît aucun nombre complètement normal ;
- on sait que de tels nombres existent (et même qu'ils sont extrêmement nombreux).

La preuve de leur existence, déjà présentée au chapitre 9, a ceci de fascinant qu'elle démontre que, dans l'intervalle $[0\,;\,1]$, l'ensemble des nombres normaux est de mesure 1 — et donc non vide ! Certes, de nombreux passages de la preuve ne reposent que sur de la théorie élémentaire de la mesure, mais l'idée centrale (montrer que presque tous les nombres sont complètement normaux dans une base d donnée) provient de la loi forte des grands nombres.

Erdős fut le grand initiateur des techniques de démonstration de ce style : pour prouver que certains ensembles sont non vides, on prouve qu'ils sont de probabilité strictement positive. Leur application à l'arithmétique — le domaine des mathématiques où l'idée de hasard semblait la moins naturelle — marqua de nombreux esprits.

Points singuliers du cercle de convergence d'une série entière §362

Voici un dernier exemple de preuve d'existence d'un certain type d'objets, de nature purement analytique cette fois-ci. Nous l'empruntons à l'excellent ouvrage de Zuily & Queffélec[4].

Soit $\sum a_n z^n$ une série entière de rayon 1 ; rappelons que le rayon d'une série entière est déterminé par les modules $|a_n|$ seulement. La théorie générale des fonctions analytiques nous apprend qu'il existe (au moins) un réel θ tel que la somme de cette série n'admet pas de prolongement analytique en $z = \mathrm{e}^{\mathrm{i}\theta}$. En d'autres points, au contraire, la série peut admettre un prolongement analytique ; de tels points sont dits **réguliers**. Les points qui ne sont pas réguliers sont dits **singuliers**.

Exemple 24.4 La série entière $\sum z^n$, de rayon 1, a pour somme $1/(1-z)$ sur le disque ouvert $\mathrm{B}(0\,;\,1)$; elle admet 1 comme point singulier. Tous les autres points du cercle unité sont réguliers, la fonction $z \mapsto 1/(1-z)$ se prolongeant analytiquement sur $\mathbb{C} \setminus \{1\}$.

La série $\sum z^n/n^2$, de rayon 1, converge en tout point du disque fermé. Cependant, la fonction f somme de cette série entière, si elle est analytique sur le disque ouvert, n'admet pas de prolongement analytique au voisinage de $z = 1$ (la fonction f' diverge au voisinage de ce point). En revanche, f admet un prolongement sur l'ouvert $\mathbb{C} \setminus [1\,;+\infty[$ (la primitive, nulle en 0, de $z \mapsto -\ln(1-z)/z$, bien définie sur ce domaine simplement connexe). Ainsi, 1 est le seul point singulier de la série entière.

Une série entière de rayon 1 admet au moins un point singulier sur le cercle unité. Le théorème suivant montre que, si l'on modifie correctement les arguments (les phases) des coefficients a_n, on peut s'arranger pour que *tous les points* du cercle unité soient des points singuliers — on dit alors que le cercle unité est une **coupure** pour la série entière.

■ **Théorème 24.5 (Steinhaus)** *Soit $\sum a_n z^n$ une série entière de rayon 1. Alors, il existe une suite réelle $(\theta_n)_{n\geqslant 0}$ telle que la série entière $\sum a_n \mathrm{e}^{\mathrm{i}\theta_n} z^n$ admette le cercle unité $\mathscr{C} = \mathscr{C}(0\,;\,1)$ comme coupure.*

(4). Claude Zuily et Hervé Queffélec, *Analyse pour l'Agrégation* [102] ; un ouvrage qui regorge d'exemples passionnants et de remarques subtiles.

Pour montrer l'existence d'une telle perturbation des phases, nous allons prouver que *presque toutes* les perturbations possibles mènent au résultat voulu. Autrement dit, si l'on modifie les phases aléatoirement, alors presque sûrement la série $\sum a_n \, e^{i\theta_n} \, z^n$ admet $\mathscr{C}$ comme coupure.

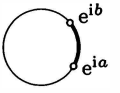
DÉMONSTRATION : On considère l'espace $\Omega = \mathscr{C}^{\mathbb{N}}$ des suites $\omega = (\omega_n)_{n\geqslant 0} = (e^{i\theta_n})_{n\geqslant 0}$ à valeurs dans $\mathscr{C}$, muni de la tribu produit $\mathfrak{T} := \bigotimes_{n\geqslant 0} \mathfrak{T}_n$, où $\mathfrak{T}_n$ est la tribu des boréliens du cercle $\mathscr{C}$, et de la probabilité produit $\mathbf{P} := \bigotimes_{n\geqslant 0} \mathbf{P}_n$, où $\mathbf{P}_n$ est la probabilité uniforme sur $(\mathscr{C}, \mathfrak{T}_n)$ (cf. A68). On définit la suite $(X_n)_{n\geqslant 0}$ de variables aléatoires sur Ω par la formule

$$X_n(\omega) := \omega_n.$$

Par construction de la probabilité produit, les X_n sont des variables aléatoires complexes indépendantes, de même loi uniforme sur $\mathscr{C}$.

Notons $\mathscr{A}$ l'ensemble des arcs fermés de $\mathscr{C}$, et non réduits à un point. Un arc A de $\mathscr{A}$ est donc un ensemble de la forme

$$A = \left\{ e^{i\theta} \; ; \; a \leqslant \theta \leqslant b, \quad 0 < b - a \leqslant 2\pi \right\}.$$



Pour tout $\omega = (e^{i\theta_n})_{n\geqslant 0}$, notons

$$f_\omega : z \longmapsto \sum_{n=0}^{\infty} a_n \, e^{i\theta_n} \, z^n$$

la somme de la série entière aléatoire $\sum a_n \, e^{i\theta_n} \, z^n = \sum a_n \, X_n(\omega) \, z^n$. Enfin, pour tout arc $A \in \mathscr{A}$, notons R_A l'ensemble des ω tels que tous les points de A sont réguliers pour f_ω.

Le point le plus technique est de montrer que les R_A sont des événements, c'est-à-dire des éléments de la tribu $\mathfrak{T}$, et nous l'admettons ici. En revanche, les R_A sont des événements asymptotiques par rapport aux variables $(X_n)_{n\geqslant 0}$; comme celles-ci sont indépendantes, la loi du zéro-un montre que

$$\mathbf{P}(R_A) = 0 \quad \text{ou} \quad \mathbf{P}(R_A) = 1$$

pour tout arc A.

Maintenant, une conséquence du fait que la loi uniforme sur $\mathscr{C}$ est invariante par rotation $z \mapsto e^{i\theta} z$, est que pour tout suite $\omega = (e^{i\theta_n})_{n\geqslant 0}$, la suite $(X_n \, e^{i\theta_n})_{n\geqslant 0}$ a même loi que la suite $(X_n)_{n\geqslant 0}$. Si A et A′ sont des arcs de même longueur, ils se déduisent l'un de l'autre par une rotation : $A' = e^{-i\theta} A$; si l'on note $X'_n = X_n \, e^{in\theta}$, alors pour tout $z \in \mathscr{C}$ on a

$$f_\omega(z \, e^{i\theta}) = \sum_{n=0}^{\infty} a_n \, X'_n(\omega) \, z^n$$

d'où l'on déduit l'équivalence entre les propositions

i) $\big(X'_n(\omega)\big)_{n\geqslant 0} \in R_{A'}$;

ii) pour tout $z \in A'$, $z e^{i\theta}$ est régulier pour f_ω ;

iii) pour tout $z \in A$, z est régulier pour f_ω ;

iv) $\omega = \big(X_n(\omega)\big)_{n\geqslant 0} \in R_A$;

Enfin, puisque les suites $(X_n)_{n\geqslant 0}$ et $(X'_n)_{n\geqslant 0}$ ont même loi, on en déduit que $\mathbf{P}(R_A) = \mathbf{P}(R_{A'})$.

Reste à savoir si l'on a $\mathbf{P}(R_A) = 0$ ou $\mathbf{P}(R_A) = 1$ pour un arc A donné. Nous allons prouver que, pour tout arc A, on a en fait $\mathbf{P}(R_A) = 0$.

Supposons en effet que ce ne soit pas le cas, et qu'il existe un arc A tel que $\mathbf{P}(R_A) = 1$. On peut recouvrir le cercle $\mathscr{C}$ tout entier avec un nombre fini de copies de A, après rotations adéquates :

$$\mathscr{C} = \bigcup_{n=1}^{N} A_n.$$

Ces arcs ayant même longueur, on en déduit que $\mathbf{P}(R_{A_n}) = 1$ pour $n = 1, \ldots, N$ et donc

$$\mathbf{P}\left(\bigcap_{n=1}^{N} R_{A_n} \right) = 1.$$

Notamment, l'ensemble $\bigcap_{n=1}^{N} R_{A_n}$ est non vide, il admet donc au moins un élément $\omega = (e^{i\theta_n})_n$ dans cet ensemble. Alors, tous les points du cercle sont réguliers pour f_ω, ce qui est faux (une série entière de rayon 1 possède au moins un point singulier).

On a donc montré que :

$$\forall A \in \mathscr{A} \qquad \mathbf{P}(R_A) = 0.$$

C'est presque terminé ! Limitons-nous maintenant à la classe $\mathscr{A}_0$ des arcs dont les extrémités sont des racines de l'unité ; ainsi, $\mathscr{A}_0$ est un ensemble dénombrable. Notons enfin

$$R := \bigcup_{A \in \mathscr{A}_0} R_A.$$

C'est une réunion dénombrable d'événements de probabilité nulle, donc un événement de probabilité nulle. Son complémentaire $S = R^c$ est donc un ensemble de probabilité 1. Il suffit maintenant de remarquer que, par construction, S est l'ensemble des ω qui n'appartiennent à aucun des R_A, c'est-à-dire tels que tout point z de $\mathscr{C}$ est singulier pour f_ω (si z est un point régulier de f_ω, il existe tout un arc centré sur z et formé de points réguliers). ∎

Le théorème de Weierstrass et les polynômes de Bernstein

§ 363

Nous rappelons l'énoncé du théorème de Weierstrass sur l'approximation des fonctions continues sur un segment :

■ **Théorème 24.6 (Weierstrass)** *Soient* $[a\,;b]$ *un segment de* $\mathbb{R}$ *et* $f : [a\,;b] \to \mathbb{C}$ *une fonction continue. Alors il existe une suite* $(P_n)_{n\geqslant 0}$ *de polynômes qui converge uniformément vers* f *sur* $[a\,;b]$.

Nous allons montrer ce résultat dans le cas où $[a\,;b] = [0\,;1]$, une translation et une dilatation permettant de revenir au cas général. L'idée de Bernstein[(5)] est d'approcher la valeur $f(x)$ de la fonction en un point $x \in [0\,;1]$ par la valeur $f(X_n)$, où $(X_n)_{n\geqslant 0}$ est une variable aléatoire à valeurs dans $[0\,;1]$, dont la distribution se concentre rapidement en x lorsque n tend vers l'infini. La valeur moyenne de $f(X_n)$ est alors une moyenne sur un échantillonnage des valeurs de f autour de x. Une telle suite est aisée à fabriquer. Fixons donc momentanément x dans $[0\,;1]$, et construisons une marche aléatoire

$$S_n^x = Y_1^x + Y_2^x + \cdots + Y_n^x,$$

où les Y_k^x sont des variables aléatoires de Bernoulli de paramètre x, c'est-à-dire vérifiant

$$\mathbf{P}\{Y_k^x = 0\} = 1 - x \qquad \text{et} \qquad \mathbf{P}\{Y_k^x = 1\} = x.$$

La variable aléatoire S_n^x suit une loi binomiale $\mathscr{B}(n\,;\,x)$, son espérance vaut nx et sa variance $nx(1-x)$. Enfin, notons

$$X_n^x = \frac{S_n^x}{n} = \frac{Y_1^x + Y_2^x + \cdots + Y_n^x}{n},$$

et remarquons que X_n^x prend bien ses valeurs dans $[0\,;1]$, que $\mathbf{E}(X_n^x) = x$ et que sa variance peut être majorée *uniformément* par rapport à x :

$$\mathbf{V}(X_n^x) = \frac{x(1-x)}{n} \leqslant \frac{1}{4n}.$$

(5). Sergueï Natanovitch BERNSTEIN (1880–1960), mathématicien ukrainien, ayant étudié à Paris à Supélec, et enseigné à Leningrad et Moscou. Il résolut le 19e problème de Hilbert dans sa thèse de doctorat. À ne pas confondre avec son homonyme Felix Bernstein (1878–1956), mathématicien allemand, élève de Cantor, qui démontra le théorème que son maître avait conjecturé, désormais connu sous le nom de théorème de Cantor-Schröder-Bernstein (*s'il existe une injection de* E *dans* F *et une injection de* F *dans* E, *alors il existe une bijection de* E *sur* F).

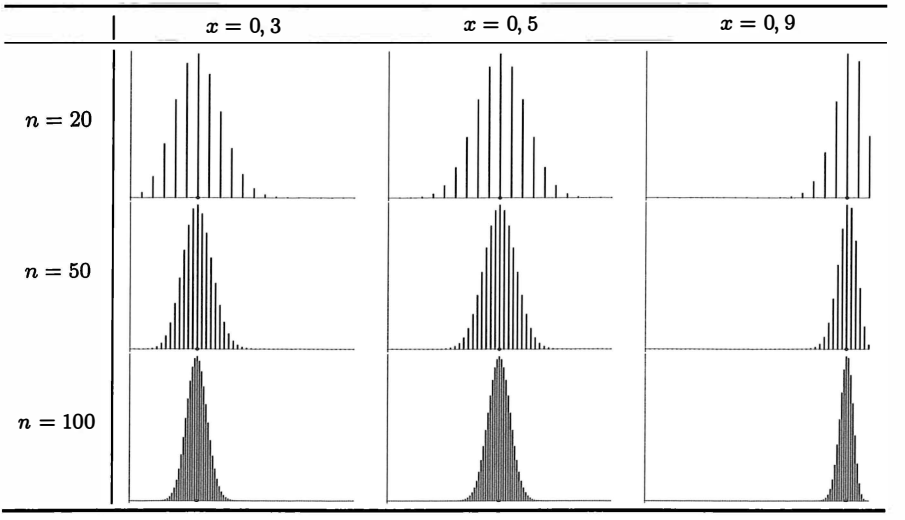


FIGURE 24.1 — *La distribution des variables* X_n^x *sur* $[0;1]$, *pour différentes valeurs de* x *et de* n. *On notera la convergence vers un « pic » centré en* x.

La loi de X_n^x, représentée sous forme d'histogramme sur la figure 24.1, est donnée par

$$\forall k \in [\![0;n]\!] \qquad \mathbf{P}\left\{X_n^x = \frac{k}{n}\right\} = \binom{n}{k} x^k (1-x)^{n-k},$$

et l'espérance de $f(X_n^x)$ est

$$\mathbf{E}\big[f(X_n^x)\big] = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right),$$

qui a le bon goût d'être une fonction polynomiale en x.

Nous pouvons maintenant écrire le résultat de Bernstein :

■ **Théorème 24.7 (Bernstein, 1913)** *Soit* $f : [0;1] \to \mathbb{C}$ *une fonction continue. Pour tout entier* n, *on définit la fonction polynomiale*

$$\begin{aligned} B_n : \ [0;1] &\longrightarrow \mathbb{C}, \\ x &\longmapsto \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}. \end{aligned}$$

Alors la suite $(B_n)_{n\geqslant 0}$ *converge uniformément vers* f *sur* $[0;1]$.

Pour tout $x \in [0;1]$, écrivons

$$\begin{aligned} \big|B_n(x) - f(x)\big| &= \mathbf{E}\Big[\big|f(X_n^x) - f(x)\big|\Big] \\ &= \sum_{k=0}^{n} \left|f\left(\frac{k}{n}\right) - f(x)\right| \mathbf{P}\left\{X_n^x = \frac{k}{n}\right\} \qquad (24.1) \\ &= \sum_{k=0}^{n} \left|f\left(\frac{k}{n}\right) - f(x)\right| \binom{n}{k} x^k (1-x)^{n-k}. \end{aligned}$$

Nous devons donc montrer que cette quantité converge (uniformément par rapport à x) vers 0 quand n tend vers l'infini. Avant toute chose, remarquons que la fonction f, continue sur un segment :

- est bornée — on notera $\mathrm{M} = \|f\|_\infty = \sup |f(x)|$;
- est uniformément continue (théorème de Heine, voir A18).

Soit $\varepsilon > 0$. Par uniforme continuité de f, il existe un réel $\delta > 0$ tel que, pour tous réels x, y de $[0\,;1]$ tels que $|x-y| < \delta$, on ait $|f(x) - f(y)| < \varepsilon$. Par ailleurs, l'inégalité de Bienaymé-Tchebychev permet d'écrire que, pour tout $n \in \mathbb{N}$ et tout $x \in [0\,;1]$,

$$\mathbf{P}\big\{\,|\mathrm{X}_n^x - x| > \delta\big\} \leqslant \frac{\mathbf{V}(\mathrm{X}_n^x)}{\delta^2} \leqslant \frac{1}{4n\,\delta^2}.$$

Notamment, il existe un entier N, indépendant de x, tel que

$$\forall n \geqslant \mathrm{N} \qquad \mathbf{P}\big\{\,|\mathrm{X}_n^x - x| > \delta\big\} \leqslant \frac{\varepsilon}{2\mathrm{M}}.$$

Or, la somme de l'équation (24.1) peut se décomposer en deux, en découpant l'ensemble Ω en

$$\Omega = \big\{|\mathrm{X}_n^x - x| > \delta\big\} \cup \big\{|\mathrm{X}_n^x - x| \leqslant \delta\big\},$$

ce que l'on écrira

$$\begin{aligned}\big|\mathrm{B}_n(x) - f(x)\big| = &\sum_{\substack{k \text{ tel que}\\ \left|\frac{k}{n}-x\right|>\delta}} \left|f\left(\frac{k}{n}\right) - f(x)\right| \binom{n}{k} x^k\,(1-x)^{n-k} \\ &+ \sum_{\substack{k \text{ tel que}\\ \left|\frac{k}{n}-x\right|\leqslant\delta}} \left|f\left(\frac{k}{n}\right) - f(x)\right| \binom{n}{k} x^k\,(1-x)^{n-k} \qquad (24.2)\end{aligned}$$

Dans la première somme, on majore $\big|f(k/n) - f(x)\big|$ par 2M, ce qui conduit à la majoration suivante, valable pour tout $n \geqslant \mathrm{N}$:

$$\begin{aligned}\sum_{\substack{k \text{ tel que}\\ \left|\frac{k}{n}-x\right|>\delta}} \left|f\left(\frac{k}{n}\right) - f(x)\right| \binom{n}{k} x^k\,(1-x)^{n-k} &\leqslant 2\mathrm{M} \sum_{\substack{k \text{ tel que}\\ \left|\frac{k}{n}-x\right|>\delta}} \binom{n}{k} x^k\,(1-x)^{n-k} \\ &= 2\mathrm{M}\,\mathbf{P}\big\{\,|\mathrm{X}_n^x - x| > \delta\big\} \leqslant \varepsilon \qquad (24.3)\end{aligned}$$

Dans la seconde somme, on majore $\big|f(k/n) - f(x)\big|$ par ε, ce qui mène à

$$\begin{aligned}\sum_{\substack{k \text{ tel que}\\ \left|\frac{k}{n}-x\right|\leqslant\delta}} \left|f\left(\frac{k}{n}\right) - f(x)\right| \binom{n}{k} x^k\,(1-x)^{n-k} &\leqslant \varepsilon \sum_{\substack{k \text{ tel que}\\ \left|\frac{k}{n}-x\right|\leqslant\delta}} \binom{n}{k} x^k\,(1-x)^{n-k} \\ &\leqslant \varepsilon \sum_{k=0}^{n} \binom{n}{k} x^k\,(1-x)^{n-k} \\ &= \varepsilon. \qquad (24.4)\end{aligned}$$

Remarque 24.8 Le lecteur trouvera peut-être agréable de remplacer les sommes précédentes par des intégrales selon une mesure de probabilité ; les équations précédentes s'écrivent

$$\big|\mathrm{B}_n(x) - f(x)\big| = \mathbf{E}\Big[\big|f\,(\mathrm{X}_n^x) - f(x)\big|\Big] = \int_\Omega \big|f(\mathrm{X}_n^x) - f(x)\big|\,\mathrm{d}\mathbf{P}, \qquad (24.1')$$

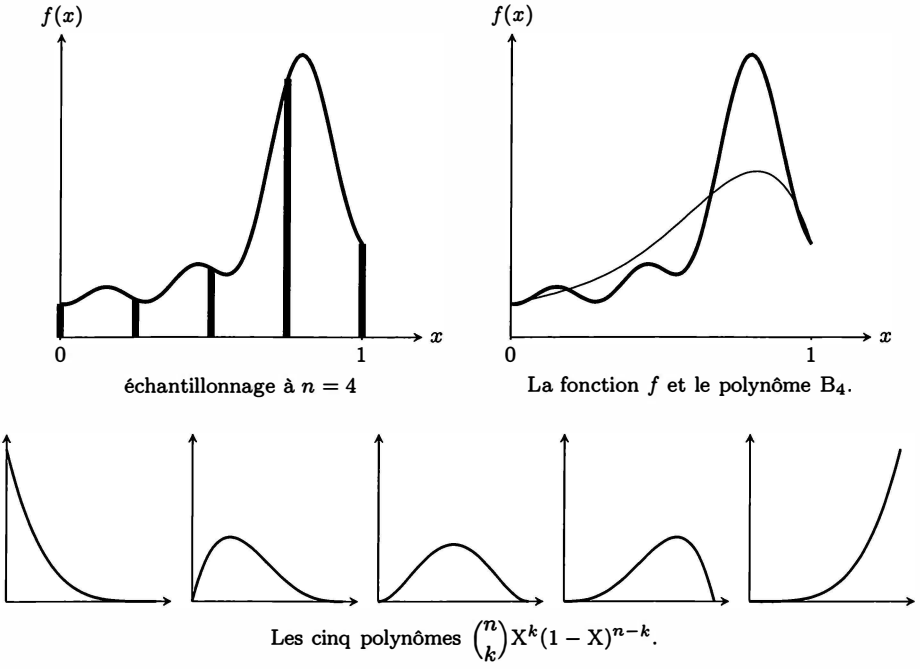


FIGURE 24.2 — *Construction du polynôme de Bernstein* B_n. *En haut, la fonction* f, *dont on échantillonne les valeurs en* $n+1$ *points équidistants. Chacune des valeurs* $f(k/n)$ *multiplie un polynôme* $\binom{n}{k} X^k (1-X)^{n-k}$. *La somme des polynômes ainsi pondérée reconstitue une approximation* B_n *de* f.

où l'on a calculé l'espérance en intégrant sur Ω selon la mesure de probabilité $d\mathbf{P}$. On obtient alors

$$\left|B_n(x) - f(x)\right| = \int_{|X_n^x - x| > \delta} \left|f(X_n^x) - f(x)\right| d\mathbf{P} + \int_{|X_n^x - x| \leqslant \delta} \left|f(X_n^x) - f(x)\right| d\mathbf{P}. \qquad (24.2')$$

Dans la première intégrale, on majore l'intégrande par 2M, ce qui donne, pour tout $n \geqslant N$:

$$\int_{|X_n^x - x| > \delta} \left|f(X_n^x) - f(x)\right| d\mathbf{P} \leqslant 2M \int_{|X_n^x - x| > \delta} d\mathbf{P} \leqslant 2M\, \mathbf{P}\{\, |X_n^x - x| > \delta\} \leqslant \varepsilon. \qquad (24.3')$$

Quant la seconde intégrale, son intégrande est majoré par ε par construction, donc

$$\int_{|X_n^x - x| \leqslant \delta} \left|f(X_n^x) - f(x)\right| d\mathbf{P} \leqslant \varepsilon \int_{|X_n^x - x| \leqslant \delta} d\mathbf{P} \leqslant \varepsilon \int_\Omega d\mathbf{P} = \varepsilon. \qquad (24.4')$$

Ainsi, pour tout $n \geqslant N$ et pour tout $x \in [0\,;\,1]$, on a bien

$$\left|B_n(x) - f(x)\right| \leqslant 2\varepsilon,$$

ce qui achève la démonstration du théorème de Bernstein et, par la même occasion, celui de Weierstrass.

Le lecteur notera à quel point la démonstration ici présentée, et qui reprend l'argument initial de Sergueï Bernstein [8], est naturelle[6], l'échantillonnage de f

(6). L'auteur commence son article de 1913 par :

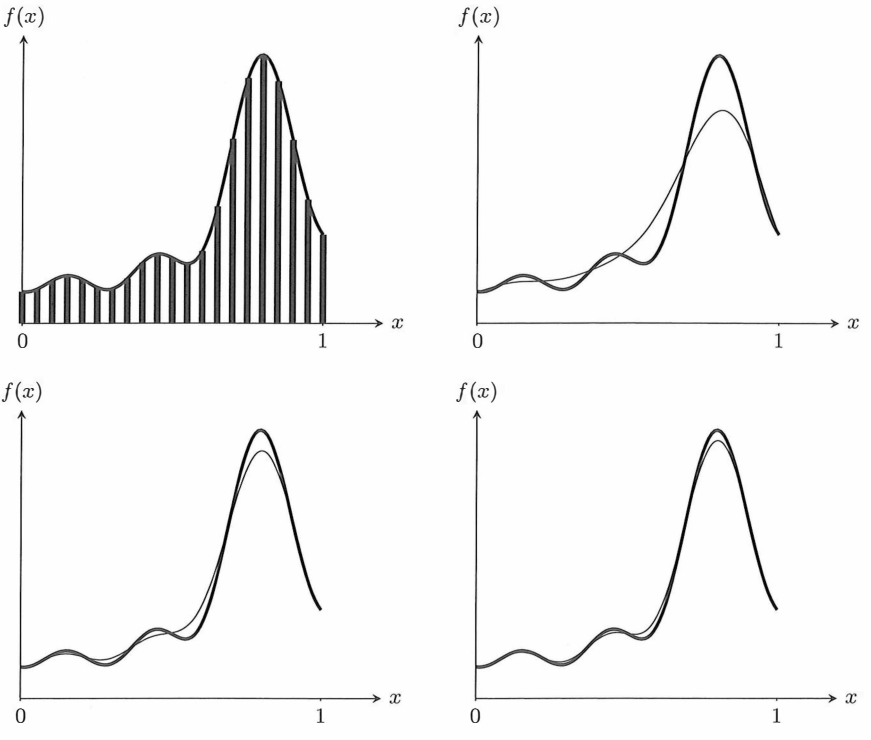


FIGURE 24.3 — *Un échantillonnage plus serré donne une meilleure approximation. En haut, conjointement avec la fonction f, le polynôme* B_{20}. *En bas, les polynômes* B_{70} *puis* B_{150}.

selon une distribution de plus en plus piquée s'obtenant naturellement grâce à la loi des grands nombres.

Je me propose d'indiquer une démonstration fort simple du théorème [...] de Weierstrass.

Si l'on peut *a posteriori* modifier la rédaction de cette preuve pour en masquer totalement le contenu probabiliste, on ne ferait cependant qu'obscurcir le schéma directeur et rendre quasiment magique l'introduction de la famille des polynômes B_n. Il suffit en effet de prouver la formule donnant la variance d'une loi binomiale

$$\sum_{k=0}^{n}(k-nx)^2\binom{n}{k}x^k\,(1-x)^{n-k} = n\,x(1-x),$$

en développant $(k-nx)^2 = k^2-2knx+n^2x^2$ puis en faisant usage de la relation $k\binom{n}{k} = n\binom{n-1}{k-1}$, avant de démontrer l'inégalité de Bienaymé-Tchebychev dans le cas particulier qui nous intéresse :

$$\begin{aligned}\delta^2 \sum_{\substack{k \text{ tel que}\\ |\frac{k}{n}-x|>\delta}} \binom{n}{k} x^k\,(1-x)^{n-k} &\leqslant \sum_{\substack{k \text{ tel que}\\ |\frac{k}{n}-x|>\delta}} \left|\frac{k}{n}-x\right|^2 \binom{n}{k} x^k\,(1-x)^{n-k}\\ &\leqslant \frac{1}{n^2}\sum_{\substack{k \text{ tel que}\\ |\frac{k}{n}-x|>\delta}} |k-nx|^2 \binom{n}{k} x^k\,(1-x)^{n-k}\\ &\leqslant \frac{1}{n^2}\sum_{k=0}^{n} |k-nx|^2 \binom{n}{k} x^k\,(1-x)^{n-k} \leqslant \frac{x(1-x)}{n} \leqslant \frac{1}{4n}.\end{aligned}$$

§ 364 **En guise de conclusion**

Dans la preuve par Bernstein du théorème de Weierstrass, il convient de remarquer que ce n'est pas la théorie des probabilités en tant que telle qui intervient pour fournir une démonstration élégante et relativement courte[7], la preuve en étant qu'on peut tout récrire sans parler le moins du monde de probabilités — d'ailleurs, les théorèmes utilisés sont extrêmement simples et ne vont pas au delà de l'inégalité de Bienaymé-Tchebychev. Ce sont plutôt *les idées, les intuitions et le langage* du monde des probabilités qui ont guidé Bernstein vers cette preuve simple. De même, c'est une intuition probabiliste qui a permis à Pál Turán de proposer une démonstration simplifiée d'un théorème difficile de Hardy et Ramanujan[8] ; par la suite, Hardy a repris la démonstration de Turán en en gommant consciencieusement toute référence aux probabilités [45].

La démonstration de la formule de Bernstein (§ 360) est particulièrement courte, mais fait appel à un résultat nettement plus difficile (le théorème limite central) ; on ne peut dès lors pas espérer rendre cette preuve lisible sans son contenu probabiliste.

EXERCICES

♦ ***Exercice 24.1*** Montrer que, pour tout entier n, on a $\displaystyle\sum_{k=n+1}^{3n-1} \binom{4n}{k} \geqslant \frac{n-1}{n}\, 2^{4n}$.

♦ ***Exercice 24.2*** Un mathématicien se réveille un matin en se demandant comment montrer que

$$\lim_{n\to\infty} \frac{1}{3^n} \sum_{k=0}^{\lfloor n/3 \rfloor} \binom{n}{k} 2^{n-k} = \frac{1}{2}.$$

Pouvez-vous l'aider ?

♦ ***Exercice 24.3 (Utilisation du théorème limite central)*** Montrer que

$$\int_0^n x^{n-1}\, \mathrm{e}^{-x}\, \mathrm{d}x \underset{n\to\infty}{\sim} \frac{(n-1)!}{2}$$

et en déduire un équivalent de

$$\int_n^{+\infty} x^{n-1}\, \mathrm{e}^{-x}\, \mathrm{d}x.$$

♦ ***Exercice 24.4 (Vitesse de convergence des polynômes de Bernstein)***
On cherche à estimer la vitesse de convergence de la suite des polynômes de Bernstein vers f. On supposera, dans tout l'exercice, que f est lipschitzienne, ce qui autorise à prendre δ proportionnel à ε.

i) En utilisant les arguments de la preuve, montrer l'estimation $\|\mathrm{B}_n - f\|_\infty = \mathrm{O}\left(\frac{1}{n^{1/3}}\right)$.

ii) On peut améliorer ce résultat. En supposant que la fonction f est k-lipschitzienne, et en utilisant l'inégalité

$$\left|\mathrm{B}_n(x) - f(x)\right| = \mathbf{E}\left|f(\mathrm{X}_n^x) - f(x)\right| \leqslant k\, \mathbf{E}\,|\mathrm{X}_n^x - x|\,,$$

montrer que $\|\mathrm{B}_n - f\|_\infty \leqslant \dfrac{k}{2\sqrt{n}}$.

---

(7). Pour mémoire, rappelons que d'autres approches existent ; l'une des plus profondes est sans doute celle consistant à effectuer le produit de convolution de la fonction f avec une fonction g_n polynomiale par morceaux, positive, de classe $\mathscr{C}^\infty$ et à support borné, telle que la suite $(g_n)_{n\geqslant 1}$ soit une suite de Dirac (une suite d'approximations de l'unité).

(8). Théorème qui stipule que le nombre de diviseurs premiers d'un entier n est asymptotiquement de l'ordre de $\ln\ln n$.

♦ ***Exercice 24.5*** Soit $f : \mathbb{R} \to \mathbb{C}$ une fonction continue et périodique de période $T > 0$. Montrer que

$$\frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{2k-n}{\sqrt{n}}\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x)\, e^{-x^2/2}\, dx \qquad (*)$$

i) de manière probabiliste, en utilisant le théorème limite central;

ii) de manière purement analytique.

Pour terminer, voici deux applications arithmétiques classiques des probabilités ; une preuve directe est possible à chaque fois, mais la preuve probabiliste est sensiblement plus courte à écrire.

♦ ***Exercice 24.6 (Identité d'Euler)*** Soit $s > 1$. Soit X une variable aléatoire à valeurs entières, de loi définie par

$$\forall n \geqslant 1 \qquad \mathbf{P}\{X = n\} = \frac{n^{-s}}{\zeta(s)}$$

où ζ est la **fonction de Riemann** : $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$.

Notons $(p_k)_{k\geqslant 1} = (2, 3, 5, \dots)$ la suite croissante des entiers premiers. Notons enfin A_k l'événement : « X est divisible par p_k ».

i) Calculer $\mathbf{P}(A_k)$ et étudier l'indépendance des A_k.

ii) En déduire l'identité d'Euler

$$\frac{1}{\zeta(s)} = \prod_{k=1}^{\infty} \left(1 - \frac{1}{p_k^s}\right).$$

♦ ***Exercice 24.7 (Indicatrice d'Euler)*** **L'indicatrice d'Euler** est la fonction $\varphi : \mathbb{N}^* \to \mathbb{N}$ définie par : $\varphi(n) = \text{Card}\,\{k \in [\![1\,;\,n]\!]\ ;\ k \wedge n = 1\}$ (nombre d'entiers de $[\![1\,;\,n]\!]$ qui sont premiers avec n).

Fixons $n \geqslant 1$ un entier, et notons $p_1, \dots, p_q$ ses diviseurs premiers. Soit X une variable aléatoire de loi uniforme sur $[\![1\,;\,n]\!]$.

Pour tout entier $k \in [\![1\,;\,n]\!]$, on note B_k l'événement « k divise X ».

i) Calculer $\mathbf{P}(B_k)$ et étudier l'indépendance des B_k.

ii) En déduire la formule :

$$\varphi(n) = n \prod_{i=1}^{q} \left(1 - \frac{1}{p_i}\right).$$

Indications

◊ *Indication 24.1 :* On pourra appliquer l'inégalité de Bienaymé-Tchebychev à une variable aléatoire bien choisie.

◊ *Indication 24.2 :* Déterminer la loi d'une variable aléatoire somme de n variables de Bernoulli indépendantes de paramètre $1/3$, et utiliser le théorème limite central.

◊ *Indication 24.3 :* On se souviendra (exercice 12.10) que la somme de n variables indépendantes de même loi exponentielle $\mathscr{E}(1)$ suit la loi $\Gamma(n, 1)$, de densité $x^{n-1}\, e^{-x}/(n-1)!$ sur $[0\,;\,+\infty[$, et on appliquera le théorème limite central.

◊ *Indication 24.4 :* Dans la question *ii)*, utiliser l'inégalité de Cauchy-Schwarz.

◊ *Indication 24.5 : De manière probabiliste*, interpréter la somme comme l'espérance de $f(X_n^*)$, où X_n est une variable aléatoire de Bernoulli et X_n^* la variable centrée réduite correspondante, puis appliquer le théorème 19.29 conjointement avec le théorème limite central.

De manière analytique, commencer par vérifier que la propriété cherchée est valable pour toutes les fonctions de la forme $f_\alpha : t \mapsto e^{i\alpha t}$, et donc pour tous les polynômes trigonométriques. On pourra démontrer (ou admettre) la relation

$$h(\alpha) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2}\, e^{i\alpha x}\, dx = e^{-\alpha^2/2},$$

soit par la méthode des résidus, soit en développant $e^{i\alpha x}$ en série et en intégrant terme à terme, soit enfin en dérivant la relation donnant $h(\alpha)$ sous le signe intégral et en formant une équation différentielle vérifiée par h.

SOLUTIONS

♦ **Solution 24.2** On appelle X_n une variable aléatoire somme de n variables de Bernoulli indépendantes de paramètre $1/3$; alors $X_n \rightsquigarrow \mathscr{B}(n, 1/3)$. On note $T_n = (X_n - n/3)/\sigma_{X_n}$ la variable centrée et réduite correspondante, dont on note F_n la fonction de répartition. Alors :

$$\frac{1}{3^n}\sum_{k=0}^{\lfloor n/3\rfloor}\binom{n}{k}2^{n-k} = \mathbf{P}\left\{X_n \leqslant \frac{n}{3}\right\} = \mathbf{P}\{T_n \leqslant 0\} = F_n(0) \xrightarrow[n\to\infty]{} \mathfrak{N}(0) = \frac{1}{2}.$$

♦ **Solution 24.4**

i) Si on lie les réels ε et δ par la relation $\varepsilon = k\,\delta$, on obtient par le raisonnement déjà prouvé la majoration, valable pour tout $\varepsilon > 0$:

$$\|B_n - f\|_\infty \leqslant \varepsilon + \frac{4}{n\,\delta^2} = \varepsilon + \frac{4k^2}{n\varepsilon^2}.$$

On peut alors, à n fixé, *choisir* le réel ε de façon à minimiser le membre de droite, ce qui revient à choisir $\varepsilon = (8k^2/n)^{1/3}$ et donne une majoration de la forme $C^{\text{te}}/n^{1/3}$.

ii) L'inégalité de Cauchy-Schwarz permet d'écrire que

$$\begin{aligned}|B_n(x) - f(x)| &\leqslant k\,\mathbf{E}\,|X_n^x - x| \leqslant k\sqrt{\mathbf{E}\,|X_n^x - x|^2}\\ &\leqslant k\sqrt{\mathbf{V}(X_n^x)} = k\sqrt{\frac{x(1-x)}{n}} \leqslant \frac{k}{2\sqrt{n}}\end{aligned}$$

pour tout $x \in [0\,;\,1]$, et donc $\|B_n - f\|_\infty \leqslant k/2\sqrt{n}$.

Dans tous les cas, la convergence est donc *a priori* lente, ce que confirment les graphes de la figure 24.3. La dernière estimation est en fait optimale : pour la fonction $f : x \mapsto |x - 1/2|$, on peut minorer $\|B_n - f\|_\infty$ par $a/\sqrt{n}$, où a est une constante.

♦ **Solution 24.5** Pour toute fonction f continue et périodique, notons

$$S_n(f) := \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k} f\left(\frac{2k-n}{\sqrt{n}}\right), \qquad \text{et} \qquad I(f) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} f(x)\,e^{-x^2/2}\,dx.$$

i) Par la formule de transfert, on peut interpréter la somme $S_n(f)$ comme

$$S_n(f) = \mathbf{E}\left[f\left(\frac{2X_n - n}{\sqrt{n}}\right)\right] = \mathbf{E}\left[f\left(\frac{X_n - n/2}{\sqrt{n/4}}\right)\right] = \mathbf{E}\big[f(X_n^*)\big],$$

où X_n suit la loi binomiale $\mathscr{B}\left(n\,;\,\frac{1}{2}\right)$ et X_n^* est la variable centrée réduite correspondante.

Or $X_n \xrightarrow{\text{L}} N$ où N est une variable aléatoire de loi normale $\mathscr{N}(0\,;\,1)$. Si f est continue et bornée sur $\mathbb{R}$ (ce qui est notamment le cas si f est continue et périodique), le théorème 19.29 montre que

$$S_n(f) = \mathbf{E}\big[f(X_n^*)\big] \xrightarrow[n\to\infty]{} \mathbf{E}\big[f(N)\big] = I(f),$$

la dernière égalité étant conséquence du théorème de transfert.

ii) Soit α un réel, et notons $f_\alpha : t \mapsto e^{i\alpha t}$. Alors

$$\begin{aligned}S_n(f) &= \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k} f\left(\frac{2k-n}{\sqrt{n}}\right) = \frac{e^{-i\alpha\sqrt{n}}}{2^n}\sum_{k=0}^{n}\binom{n}{k}\left[e^{2i\alpha/\sqrt{n}}\right]^k\\ &= \frac{e^{-i\alpha\sqrt{n}}}{2^n}\left[1 + e^{2i\alpha/\sqrt{n}}\right]^n = \left[\cos\left(\frac{\alpha}{\sqrt{n}}\right)\right]^n \xrightarrow[n\to\infty]{} e^{-\alpha^2/2}.\end{aligned}$$

La relation

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-x^2}\,e^{i\alpha x}\,dx = e^{-\alpha^2/2}$$

montre que la formule de l'énoncé est valide pour la fonction f_α.

Par linéarité, elle est valide pour tout polynôme trigonométrique.

Considérons maintenant f une fonction continue et périodique ; notons T sa période.

Soit $\varepsilon > 0$. D'après le théorème d'approximation de Weierstrass[9], il existe un polynôme trigonométrique P, de période T, qui soit ε-proche de f, au sens où $\|P - f\|_\infty \leqslant \varepsilon$. Alors

$$|S_n(f) - S_n(P)| = \left| \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} \left[f\left(\frac{2k-n}{\sqrt{n}}\right) - P\left(\frac{2k-n}{\sqrt{n}}\right) \right] \right|$$
$$\leqslant \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} \|f - P\|_\infty = \|f - P\|_\infty \leqslant \varepsilon.$$

De même, $|I(f) - I(P)| \leqslant \varepsilon$.

Enfin, il existe un entier N tel que, pour tout $n \geqslant N$, on ait $|I(P) - S_n(P)| \leqslant \varepsilon$ d'après ce qui a été fait précédemment. Alors, pour tout $n \geqslant N$,

$$|S_n(f) - I(f)| \leqslant |S_n(f) - S_n(P)| + |S_n(P) - I(P)| + |I(P) - I(f)| \leqslant 3\varepsilon.$$

♦ **Solution 24.6**

i) On peut écrire

$$\mathbf{P}(A_k) = \mathbf{P}\{p_k \text{ divise } X\} = \sum_{n=1}^{\infty} \mathbf{P}\{X = n p_k\} = \sum_{n=1}^{\infty} \frac{(n p_k)^{-s}}{\zeta(s)} = \frac{1}{p_k^s}.$$

Étudions maintenant l'indépendance mutuelle des A_k. Pour cela, choisissons un entier $q \geqslant 2$ et choisissons des entiers $k_1, \ldots, k_q$ distincts. L'événement $A_{k_1} \cap \cdots \cap A_{k_q}$ se traduit par « X est divisible à la fois par $p_{k_1}, \ldots, p_{k_q}$ », ou ce qui équivaut « X est divisible par le produit $p_{k_1} \ldots p_{k_q}$ ». Ainsi,

$$\mathbf{P}(A_{k_1} \cap \cdots \cap A_{k_q}) = \sum_{n=1}^{\infty} \mathbf{P}\left\{X = n p_{k_1} \ldots p_{k_q}\right\} = \sum_{n=1}^{\infty} \frac{(n p_{k_1} \ldots p_{k_q})^{-s}}{\zeta(s)} = \frac{1}{p_{k_1}^s \cdots p_{k_q}^s}.$$

Ceci étant vrai pour tout choix de q et de $k_1, \ldots, k_q$, on a bien montré l'indépendance mutuelle des A_k.

ii) Enfin, $(1 - p_k^{-s})$ étant la probabilité que X ne soit pas divisible par p_k, on a

$$\prod_{k=1}^{\infty} \left(1 - \frac{1}{p_k^s}\right) = \mathbf{P}\left(\bigcap_{k=1}^{\infty} A_k^c \right)$$
$$= \mathbf{P}\{X \text{ n'est divisible par aucun } p_k\} = \mathbf{P}\{X = 1\} = \frac{1}{\zeta(s)}.$$

♦ **Solution 24.7**

i) L'entier premier p_k est un diviseur de n ; il divise X lorsque X prend l'une des valeurs p_k, $2p_k$, $3p_k$,... jusqu'à $(n/p_k)\,p_k$. Il y a n/p_k telles valeurs, toutes de probabilité $1/n$. Ainsi, $\mathbf{P}(B_k) = 1/p_k$.

Soit $r \in [1\,;\,q]$ et soient $k_1, \ldots, k_r$ des indices distincts de $[1\,;\,q]$. L'événement $B_{k_1} \cap \cdots \cap B_{k_r}$ se traduit par « X est divisible à la fois par $p_{k_1}, \ldots, p_{k_r}$ », ce qui équivaut à « X est divisible par le produit $p_{k_1} \ldots p_{k_r}$ ». Le nombre de multiples de $p_{k_1} \ldots p_{k_r}$ dans $[1\,;\,n]$ est exactement $n/p_{k_1} \ldots p_{k_r}$, donc

$$\mathbf{P}\left(B_{k_1} \cap \cdots \cap B_{k_r}\right) = \frac{1}{p_{k_1} \ldots p_{k_r}} = \prod_{i=1}^{r} \mathbf{P}(B_{k_i}).$$

Ceci étant vrai pour tout choix de r et d'indices $k_1, \ldots, k_r$, on a montré l'indépendance des B_k.

ii) Notons $D := \{k \in [1\,;\,n]\ ;\ k \wedge n = 1\}$. Un entier $1 \leqslant k \leqslant n$ est dans k si et seulement si il n'est multiple ni de p_1, ni de p_2,..., ni de p_q. Ainsi, $D = B_1^c \cap \cdots \cap B_q^c$ et, par indépendance mutuelle des B_k^c, on a

$$\mathbf{P}(D) = \prod_{i=1}^{q} (1 - \mathbf{P}(B_{k_i})) = \prod_{i=1}^{q} \left(1 - \frac{1}{p_i}\right),$$

et puisque $\mathbf{P}(D) = \mathrm{Card}(D)/n$, la formule est démontrée.

---

(9). Voir A36, page 683.

Chapitre 25

Simuler informatiquement une loi de probabilité

GUILDENSTERN *Syllogism the second: One, probability is a factor which operates within natural forces. Two, probability is not operating as a factor. Three, we are now within un-, sub- or supernatural forces. Discuss.* (Rosencrantz is suitably startled. Acidly.) *Not too heatedly.*[1]

Tom STOPPARD, *Rosencrantz & Guildenstern are dead.*

Toutes les méthodes de simulation de lois probabilistes se basent sur l'existence d'une source aléatoire de loi uniforme. Le lecteur ne cherchant qu'à simuler une loi de Poisson, une loi normale, etc., peut admettre l'existence d'un programme informatique, que nous appellerons « `rand` », dont les sorties successives, lorsqu'il est appelé de manière répétée, forment une suite numérique $u_0, u_1, u_2, \ldots$ correspondant à une suite de variables aléatoires indépendantes *et de loi uniforme sur $[0\,;1]$. Une très courte discussion sur les algorithmes permettant de simuler une loi uniforme est proposée en fin de chapitre.*

25.1 SIMULER UNE LOI DISCRÈTE

Dans toute la suite du chapitre, on supposera donnée une fonction `rand` qui retourne un réel de l'intervalle $[0\,;1[$, avec une probabilité uniforme, et telle que des appels successifs à cette fonction donnent des résultats indépendants.

Remarque 25.1 Remarquons d'ores et déjà que l'on aura parfois à considérer que `rand` retourne un réel de $]0\,;1[$, ou de $[0\,;1]$; comme dans tous les cas, la probabilité d'obtenir exactement 0 ou 1 est nulle, il n'y a pas de véritable contradiction.

(1). GUILDENSTERN Syllogisme II : primo, la probabilité est un facteur agissant au sein de forces naturelles. Deuxio, la probabilité n'agit pas en tant que facteur. Tertio, nous sommes les jouets de forces anti-, sous- ou surnaturelles. Discutez. *(Rosencrantz est proprement ébahi. D'un ton aigre :)* Ne nous emportons pas.

§ 365 **Simuler une loi de Bernoulli**

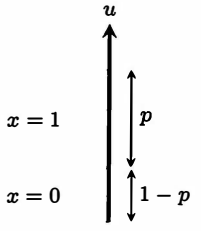
Une variable qui suit une loi de Bernoulli de paramètre p prend la valeur 0 avec la probabilité $1-p$ et la valeur 1 avec la probabilité p. Ainsi, on simule une telle variable en tirant un nombre U selon la loi uniforme sur $[0\,;1]$, puis en posant $X=0$ si $U<1-p$ et $X=1$ si $U\geqslant 1-p$.

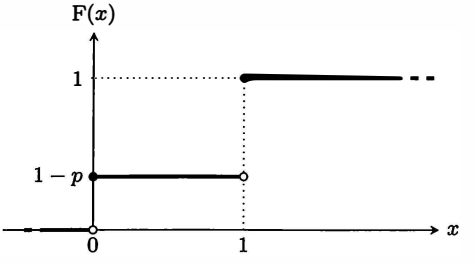
Algorithme 1 : Simuler la loi de Bernoulli $\mathscr{B}(p)$

$u \leftarrow$ `rand()`
si $u<1-p$ **alors** $x\leftarrow 0$

sinon $x\leftarrow 1$

retourner x





§ 366 **Simuler une loi binomiale**

Puisque la somme de n variables indépendantes de même loi de Bernoulli $\mathscr{B}(p)$ suit une loi binomiale $\mathscr{B}(n\,;\,p)$, il suffit de faire appel de manière répétée à l'instruction `rand` pour obtenir une simulation de loi binomiale.

Algorithme 2 : Simuler la loi binomiale $\mathscr{B}(n\,;\,p)$

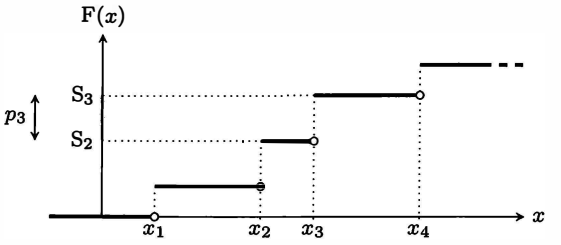
$X\leftarrow 0$
pour $i=1$ **à** n **faire**
  $u\leftarrow$ `rand()`
  si $u\geqslant 1-p$ **alors** $X\leftarrow X+1$
retourner X

§ 367 **Simuler une loi discrète finie**

On cherche à simuler une variable discrète X, prenant des valeurs $x_1, x_2, \ldots, x_n$ supposées ordonnées de manière croissante. On note $p_k := \mathbf{P}\{X=x_k\}$. On tire alors un nombre u selon la loi uniforme sur $[0\,;1[$ et on cherche l'unique entier k tel que

$$\sum_{i=1}^{k-1} p_i =: S_{k-1} < u \leqslant S_k := \sum_{i=1}^{k} p_i$$

(le membre de gauche valant 0 si $k=1$). On pose alors $X=x_k$. Sur la figure ci-dessous, le domaine dans lequel u doit se trouver pour que l'on prenne la valeur x_3 est l'intervalle $]S_2\,;S_3]$, de largeur p_3.



L'algorithme correspondant à cette méthode est donc :

Algorithme 3 : Simuler une loi discrète finie

$k \leftarrow 0$
$S \leftarrow 0$
$u \leftarrow$ `rand()`
tant que $S < u$ **faire**
  $k \leftarrow k+1$
  $S \leftarrow S + p_k$
retourner x_k

Remarque 25.2 (Simuler une loi discrète infinie)
Si l'on peut encore ordonner les valeurs $\{x_k\ ;\ k \in \mathbb{N}\}$ par ordre croissant, l'algorithme donné précédemment fonctionne encore très bien[2]. Ce n'est cependant pas toujours possible, par exemple si les valeurs possibles de X sont $x_k = 1/k$, c'est-à-dire naturellement rangées dans l'ordre décroissant. Dans ce cas, il faut au contraire considérer les sommes $1 - \sum_{i=1}^{k} p_i$, et faire croître k jusqu'à passer *sous* la valeur de u. Cela revient bien sûr à appliquer l'algorithme précédent à $-X$.

Si l'ensemble $\{x_k\ ;\ k \in \mathbb{N}\}$ admet un ou plusieurs points d'accumulation, en revanche, la méthode peut devenir plus compliquée.

Simuler une loi géométrique

§ 368

Nous avons déjà noté qu'une loi géométrique est la version « discrète » d'une loi exponentielle. De fait, on peut passer de l'une à l'autre en prenant la partie entière. Plus précisément :

■ **Proposition 25.3** *Soit* Y *une variable aléatoire de loi exponentielle* $Y \rightsquigarrow \mathscr{E}(\alpha)$. *Alors les variables* $X := \lceil Y \rceil$ *et* $X' := \lfloor Y \rfloor + 1$ *suivent une loi géométrique de paramètre* $p = 1 - e^{-\alpha}$.

DÉMONSTRATION : Tout d'abord, X et X′ ne diffèrent que si Y prend une valeur entière, événement de probabilité nulle ; ainsi, $X = X'$ presque sûrement. Rappelons que la fonction de répartition de Y est $F(y) = 1 - e^{-\alpha y}$ pour tout $y \geqslant 0$. Les valeurs prises par X sont dans $\mathbb{N}^*$; pour tout entier n non nul,

$$\begin{aligned}\mathbf{P}\{X = n\} = \mathbf{P}\{n-1 < Y \leqslant n\} &= F(n) - F(n-1)\\ &= e^{-\alpha(n-1)} - e^{-\alpha n} = e^{-\alpha(n-1)}\left[1 - e^{-\alpha}\right],\end{aligned}$$

ce qui montre le résultat voulu. ∎

On en déduit l'algorithme suivant pour simuler une loi géométrique $\mathscr{G}(p)$ (la simulation d'une loi exponentielle est expliquée dans l'exemple 25.8) :

[2]. Encore qu'il puisse tourner longtemps, dans le cas où la suite des sommes S_k augmente lentement. Si l'on connaît la valeur de certaines sommes partielles, on peut effectuer d'abord un ou plusieurs tests plus grossiers pour savoir « dans quelle région » se situe la valeur de k cherchée.

Algorithme 4 : Simuler la loi géométrique $\mathscr{G}(p)$

$\alpha \leftarrow -\ln(1-p)$ // $p = 1 - \mathrm{e}^{-\alpha}$
$u \leftarrow$ **rand**()
$y \leftarrow -\ln(u)/\alpha$ // $y \rightsquigarrow \mathscr{E}(\alpha)$
retourner $\lceil y \rceil$

Remarque 25.4 Il est également correct de retourner $\lfloor y \rfloor + 1$, qui coïncide avec $\lceil y \rceil$ sauf sur un ensemble négligeable.

§ 369 **Simuler une loi de Poisson**

Il existe un moyen pour simuler une loi de Poisson, basée sur le lien profond qui existe entre lois de Poisson et lois exponentielles : si des instants aléatoires $T_1, T_2, \ldots$ (appelés *points*) sont séparés par des durées $X_i = T_i - T_{i-1}$ de loi exponentielle $X_i \rightsquigarrow \mathscr{E}(1)$, alors le nombre de points durant l'intervalle de temps $[0\,;\lambda]$ suit une loi de Poisson de paramètre λ.

Formalisons ceci. Soit $\lambda > 0$. Soient $X_1, X_2, \ldots$ des variables aléatoires indépendantes, de même loi exponentielle de paramètre 1. On note N le plus grand entier n tel que $X_1 + \cdots + X_n < \lambda$, en convenant que $N = 0$ si $X_1 > \lambda$. En d'autres termes, l'événement $\{N = n\}$ est l'intersection des événements $\{X_1 + \cdots + X_n < \lambda\}$ et $\{X_1 + \cdots + X_n + X_{n+1} \geqslant \lambda\}$.

■ **Proposition 25.5** *La variable aléatoire* N *suit une loi de Poisson de paramètre* λ.

DÉMONSTRATION : Soit $n \in \mathbb{N}$. On a égalité des événements suivants :

$$\{N \geqslant n\} = \{X_1 + X_2 + \cdots + X_n < \lambda\}.$$

Or, la somme $S_n = X_1 + \cdots + X_n$ suit une loi Γ de paramètres $(n, 1)$, de densité

$$f_n(x) = \frac{x^{n-1}}{(n-1)!}\,\mathrm{e}^{-x} \qquad \text{si } x > 0$$

(exercice 12.10). On calcule donc, pour $n \geqslant 1$,

$$\begin{aligned}\mathbf{P}\{N \geqslant n\} = \mathbf{P}\{S_n < \lambda\} &= \int_0^\lambda \frac{t^{n-1}}{(n-1)!}\,\mathrm{e}^{-t}\,\mathrm{d}t \\ &= \int_0^\lambda \frac{t^n}{n!}\,\mathrm{e}^{-t}\,\mathrm{d}t + \frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda} \quad \textit{par intégration par parties} \\ &= \mathbf{P}\{N \geqslant n+1\} + \frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda}.\end{aligned}$$

Notamment,

$$\mathbf{P}\{N = n\} = \mathbf{P}\{N \geqslant n\} - \mathbf{P}\{N \geqslant n+1\} = \frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda},$$

ce qui achève la démonstration. ∎

La méthode est alors la suivante : on considère des variables aléatoires $U_1, U_2, \ldots$ indépendantes et de même loi uniforme sur $]0\,;1[$. On pose $X_k = -\ln U_k$ pour tout entier k, ce qui nous donne des variables aléatoires exponentielles indépendantes. De l'égalité

$$\{X_1 + X_2 + \cdots + X_n < \lambda\} = \left\{U_1\,U_2 \cdots U_n > \mathrm{e}^{-\lambda}\right\},$$

on déduit l'algorithme suivant :

Algorithme 5 : Simuler la loi de Poisson $\mathscr{P}(\lambda)$

$n \leftarrow 0$
$p \leftarrow 1$
répéter
  $u \leftarrow$ rand()
  $p \leftarrow p \cdot u$
  $n \leftarrow n + 1$
jusqu'à $p \leqslant \mathrm{e}^{-\lambda}$

retourner $n - 1$

Remarque 25.6 Cet algorithme n'est réellement efficace que pour les λ « petits » ; en effet, on additionne des temps (de loi exponentielle $\mathscr{E}(1)$) jusqu'à dépasser λ ; tous ces temps intermédiaires sont ensuite oubliés, mais le temps de calcul, lui, est globalement proportionnel à λ. Si λ est très grand, on peut songer à une simulation approchée, la loi de Poisson étant proche d'une loi normale $\mathscr{N}(\lambda\,;\,\lambda)$.

25.2 SIMULER UNE LOI CONTINUE

Méthode de l'inverse

§ 370

Soit X une variable aléatoire réelle, de fonction de répartition F. On note G le pseudo-inverse de F, qui est la fonction définie sur $]0\,;\,1[$ par

$$\mathrm{G}(y) = \inf\big\{x \in \mathbb{R}\,;\ \mathrm{F}(x) \geqslant y\big\}. \qquad (25.1)$$

(On peut étendre cette définition à $[0\,;\,1]$ ou à $\mathbb{R}$ tout entier, voir A29.)

■ **Proposition 25.7** *Si* U *est une variable aléatoire suivant la loi uniforme sur* $]0\,;\,1[$, *alors* G(U) *admet* F *pour fonction de répartition.*

DÉMONSTRATION : Pour tout $x \in \mathbb{R}$ et tout $u \in]0\,;\,1[$, on a l'équivalence

$$\mathrm{G}(u) \leqslant x \Leftrightarrow u \leqslant \mathrm{F}(x)$$

(voir A29 et la proposition A.24 page 680) ; notamment,

$$\mathbf{P}\big\{\mathrm{G(U)} \leqslant x\big\} = \mathbf{P}\,\{\mathrm{U} \leqslant \mathrm{F}(x)\} = \mathrm{F}(x).$$

Ainsi, G(U) a même loi que X. ∎

Cette méthode n'est évidemment efficace que si l'on sait calculer explicitement l'inverse F^{-1} ou le pseudo-inverse G de la fonction de répartition F. ⇗25.5

Exemple 25.8 (Loi exponentielle) La fonction de répartition d'une loi exponentielle $\mathscr{E}(\alpha)$ est la fonction F définie par $\mathrm{F}(x) = 0$ si $x \leqslant 0$ et $\mathrm{F}(x) = 1 - \mathrm{e}^{-\alpha x}$ si $x \geqslant 0$. Le pseudo-inverse de F est la fonction $\mathrm{G} :]0\,;\,1[\to \mathbb{R}$ définie par $\mathrm{G}(y) = -\ln(1-y)/\alpha$. (On peut poser $\mathrm{G}(0) = 0$ et $\mathrm{G}(1) = +\infty$ par souci de complétude, mais cela ne servira à rien dans la suite.)

Par conséquent, si U suit une loi uniforme sur $]0\,;\,1[$, alors $\mathrm{X} = -\ln(1-\mathrm{U})/\alpha$ suit une loi $\mathscr{E}(\alpha)$. Puisque $1 - \mathrm{U}$ et U suivent la même loi $\mathscr{U}(]0\,;\,1[)$, on peut d'ailleurs poser plutôt $\mathrm{X} = -(1/\alpha)\ln \mathrm{U}$.

Ainsi, un algorithme simple pour obtenir une variable aléatoire est le suivant :

Algorithme 6 : Simuler la loi exponentielle $\mathscr{E}(\alpha)$

$u \leftarrow$ rand()
retourner $-\ln(u)/\alpha$

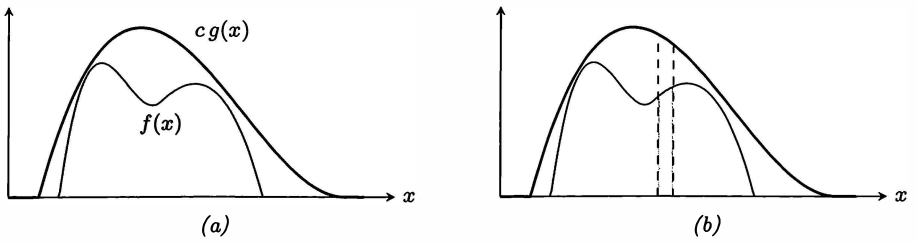


FIGURE 25.1 — À gauche : *la densité $f(x)$ et la fonction $c\,g(x)$*. À droite : *la proportion de points acceptés parmi ceux d'abscisse x est $f(x)/c\,g(x)$.*

§371 **Méthode du rejet**

La première mention de la méthode du rejet apparaît dans une lettre de John VON NEUMANN à Stanisław ULAM en 1947. Le principe est simple : si X est une variable aléatoire admettant une densité f, et si l'on ne sait pas inverser la fonction de répartition F associée, on choisit une densité g qui, à un facteur multiplicatif c près, englobe f (voir figure 25.1.*a*), c'est-à-dire que

$$\forall x \in \mathbb{R} \qquad f(x) \leqslant c\,g(x).$$

Puisque f et g sont normalisées, on a donc $c \geqslant 1$ et même $c > 1$ pour peu que f et g ne soient pas égales presque partout. Si l'on a choisi g de manière astucieuse, on peut simuler, avec la méthode précédente, une suite de variables aléatoires $Y_1, Y_2, \ldots$ indépendantes et de densité g. Heuristiquement, simuler X revient à choisir un point uniformément dans le domaine gris foncé de la figure 25.1.*a*, et à ne retenir que son abscisse ; simuler Y_k revient à choisir un point uniformément dans le domaine gris clair + gris foncé, et à ne retenir que son abscisse. L'idée pour simuler X est de :

— choisir un point dans le domaine total, et
— le *rejeter* avec une probabilité égale à la proportion « gris clair / gris foncé » au-dessus de x (figure 25.1.*b*).

Autrement dit, si l'on obtient une valeur x, aléatoirement selon la densité g, on rejette cette valeur avec une probabilité $1 - f(x)/c\,g(x)$.

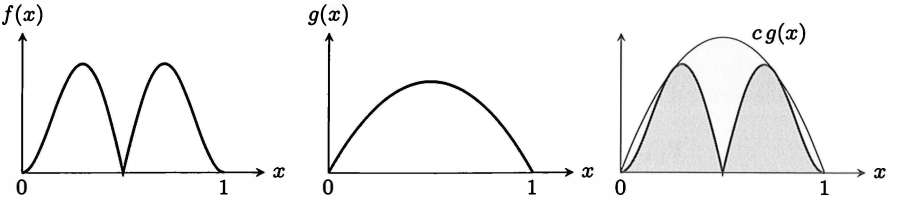
Algorithme 7 : Méthode de rejet

répéter
  tirer un réel x selon une loi de densité g
  $u \leftarrow$ `rand()`
jusqu'à $u \leqslant f(x)/c\,g(x)$
retourner x

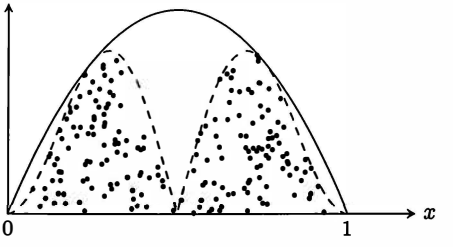
Dans cet algorithme, x admet initialement pour densité g, mais peut être rejeté sous condition ; à la sortie de la boucle, c'est-à-dire *conditionnellement à son non-rejet*, il admet pour densité f.

Exemple 25.9 On souhaite simuler une variable aléatoire de densité $f(x) = \alpha\,x(1-x)\,|\sin 2\pi x|$ sur $[0\,;\,1]$, où α est la constante de normalisation ($\alpha = 4\pi^3/(4+\pi^2)$). Bien qu'une primitive de f soit calculable, il est plus simple de programmer une méthode de rejet, car cette primitive ne s'inverse pas facilement. On pose $g(x) = 6\,x(1-x)$ sur $[0\,;\,1]$, et $c = \alpha/6$, de telle sorte que l'on a bien

$$\alpha\,x(1-x)\,|\sin 2\pi x| = f(x) \leqslant c\,g(x) = \alpha\,x(1-x).$$



On effectue n tirages d'un réel x selon la densité g par la méthode de l'inverse ; pour chacun de ces réels, on tire un réel u uniformément sur $[0\,;\,1]$. On affiche alors le couple $(x, c\,g(x)\,u)$. Si ce point est au-dessus de la fonction f, il est rejeté et on l'affiche en gris. Sinon, il est affiché en noir et comptabilisé. Les valeurs de x comptabilisées suivent bien la loi de densité f. On pouvait bien évidemment choisir pour fonction cg la fonction uniforme égale à α sur $[0\,;\,1]$, ce qui aurait évité de passer par la méthode de l'inverse ; le taux de rejet aurait été un peu plus élevé.



Remarque 25.10 (Justification de la méthode) L'algorithme de rejet est justifié par le théorème suivant :

■ **Théorème 25.11 (Méthode du rejet)** *Soit $(U_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes, de loi uniforme sur $[0\,;\,1]$, et soit $(X_n)_{n\geqslant 0}$ une suite de variables aléatoires indépendantes et indépendantes des U_k, de densité g. On pose*

$$N = \inf\{n \in \mathbb{N}^* \,;\, f(X_n) > U_n \cdot c\,g(X_n)\}$$

et $Y = X_N$ (c'est-à-dire $Y(\omega) = X_{N(\omega)}(\omega)$ pour tout $\omega \in \Omega$). Alors Y admet pour densité f.

DÉMONSTRATION : Il nous faut commencer par montrer que N est une variable aléatoire à valeurs dans $\mathbb{N}^*$. Pour cela notons A_n l'événement

$$A_n = \{f(X_n) > U_n \cdot c\,g(X_n)\}.$$

Les événements A_n sont indépendants, et tous de même probabilité p. (Graphiquement, A_n correspond au fait qu'un point choisi uniformément dans la zone grise tombe en fait dans le gris foncé sur la figure 25.1 ; on devine d'ores et déjà que $\mathbf{P}(A_n) = 1/c$.) La variable N est donc égale au rang du premier succès dans la suite des événements indépendants $A_1, A_2, \ldots$, donc suit la loi géométrique de paramètre p. Utiliser le système quasi-complet d'événements $\big(\{N = n\}\big)_{n\geqslant 1}$ pour calculer

$$\mathbf{P}\{Y \leqslant x\} = \sum_{n=1}^{\infty} \mathbf{P}\{Y \leqslant x, N = n\} = \sum_{n=1}^{\infty} \mathbf{P}\{X_n \leqslant x, N = n\}$$

semble mener à une impasse, car les variables X_n et N ne sont pas indépendantes. Décomposons cependant l'événement $\{N = n\}$ en

$$\{N = n\} = A_n \cap \{N > n-1\} = A_n \cap A_1^c \cap A_2^c \cap \cdots \cap A_{n-1}^c.$$

On a alors

$$\mathbf{P}\{Y \leqslant x\} = \sum_{n=1}^{\infty} \mathbf{P}\{X_n \leqslant x, N = n\} = \sum_{n=1}^{\infty} \mathbf{P}\big(\{X_n \leqslant x\} \cap A_n \cap \{N > n-1\}\big).$$

Or, l'événement $\{N > n-1\}$ ne concerne que les variables $X_1, X_2, \ldots, X_{n-1}$; par le lemme des coalitions, il est indépendant de $\{X_n \leqslant x\} \cap A_n$ qui ne concerne que X_n, et

$$\mathbf{P}\{Y \leqslant x\} = \sum_{n=1}^{\infty} \mathbf{P}\big(\{X_n \leqslant x\} \cap A_n\big) \cdot \mathbf{P}\{N > n-1\} = \sum_{n=1}^{\infty} \mathbf{P}\big(\{X_n \leqslant x\} \cap A_n\big) \cdot (1-p)^{n-1}.$$

Il ne reste qu'à calculer

$$\mathbf{P}\big(\{X_n \leqslant x\} \cap A_n\big) = \mathbf{P}\{X_n \leqslant x, f(X_n) > U_n\, c\, g(X_n)\},$$

sous la forme d'une intégrale double

$$\mathbf{P}\big(\{X_n \leqslant x\} \cap A_n\big) = \int_{-\infty}^{+\infty} \int_0^1 \mathbf{1}_{]-\infty\,;\,x]}(s)\, g(s)\, \chi(s,t)\, dt\, ds$$

où

$$\chi(s,t) \coloneqq \begin{cases} 1 & \text{si } f(s) > tcg(s) \\ 0 & \text{sinon,} \end{cases}$$

ce qui donne[3]

$$\mathbf{P}\big(\{X_n \leqslant x\} \cap A_n\big) = \int_{-\infty}^{x} g(s)\, \frac{f(s)}{c\, g(s)}\, ds = \frac{1}{c} \int_{-\infty}^{x} f(s)\, ds = \frac{1}{c}\, F(x).$$

On aboutit finalement à

$$\mathbf{P}\{Y \leqslant x\} = \frac{1}{c}\, F(x) \cdot \sum_{n=1}^{\infty} (1-p)^{n-1} = \frac{1}{pc}\, F(x)$$

où F est la fonction de répartition associée à la densité f. On en déduit que $p = 1/c$ comme prévu, et que Y a bien pour densité f, ce qui achève la preuve. ∎

Avant de terminer cette remarque : puisque N suit une loi géométrique de paramètre $1/c$, le nombre moyen de tirages pour obtenir une valeur est égal à c. Ainsi, on aura intérêt à choisir c le plus petit possible.

§ 372 **Mélange de fonctions de répartition**

Si l'on sait simuler informatiquement des variables X_1 et X_2 de fonctions de répartition respectives F_1 et F_2, on peut simuler une variable Y dont la fonction de répartition est $\alpha F_1 + (1-\alpha) F_2$, lorsque $0 \leqslant \alpha \leqslant 1$. Il suffit pour cela de créer une variable N suivant la loi de Bernoulli $\mathscr{B}(\alpha)$ et indépendante de X_1 et X_2, puis de poser

$$Y = \begin{cases} X_1 & \text{si } N = 1, \\ X_2 & \text{si } N = 0. \end{cases}$$

Pour tout réel x, on a en effet, par la formule des probabilités totales :

$$\begin{aligned} F(x) &= \mathbf{P}\{Y \leqslant x\} = \mathbf{P}\{Y \leqslant x, N = 1\} + \mathbf{P}\{Y \leqslant x, N = 0\} \\ &= \mathbf{P}\{X_1 \leqslant x, N = 1\} + \mathbf{P}\{X_2 \leqslant x, N = 0\} \\ &= \mathbf{P}\{X_1 \leqslant x\}\, \mathbf{P}\{N = 1\} + \mathbf{P}\{X_2 \leqslant x\}\, \mathbf{P}\{N = 0\} \quad \textit{(indépendance)} \\ &= \alpha F_1(x) + (1-\alpha) F_2(x). \end{aligned}$$

Algorithme 8 : Simuler Y de fonction de répartition $\alpha F_1 + (1-\alpha) F_2$

```
N ← rand()
si N < α alors
  └ tirer X selon F_1
sinon
  └ tirer X selon F_2
retourner X
```

[3]. Lorsque $g(s)$ s'annule, $f(s)$ également et l'intégrande est nul, ce qui justifie l'écriture un peu abusive de cette équation.

Exemple 25.12 (Simuler la loi $\gamma(a)$ pour $0 < a \leqslant 1$) D. Knuth [60] propose un algorithme, suivant J. H. Ahrens, pour simuler une loi γ de paramètre $a \in]0\,;\,1]$ selon la méthode du rejet. On rappelle que la densité associée à cette loi est la fonction f, nulle sur $\mathbb{R}^-$, et valant sinon

$$f(x) = \frac{1}{\Gamma(a)}\, x^{a-1}\, \mathrm{e}^{-x}.$$

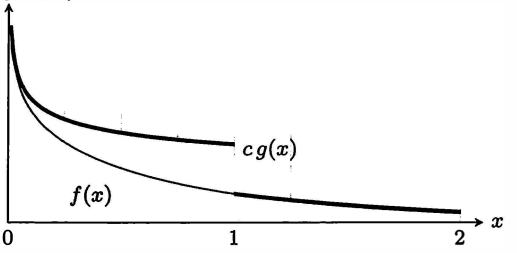
On majore f par la fonction cg définie sur $]0\,;\,+\infty[$ par

$$c\,g(x) = \begin{cases} x^{a-1}/\Gamma(a) & \text{si } 0 < x < 1, \\ \mathrm{e}^{-x}/\Gamma(a) & \text{si } x \geqslant 1. \end{cases}$$

Pour que g soit correctement normalisée à $\int_{-\infty}^{+\infty} g = 1$, il suffit de poser

$$c = \frac{1}{\lambda\,\Gamma(a)} \qquad \lambda = \frac{a\,\mathrm{e}}{a+\mathrm{e}} \qquad \text{et} \qquad g(x) = \begin{cases} \lambda\, x^{a-1} & \text{si } 0 < x < 1, \\ \lambda\, \mathrm{e}^{-x} & \text{si } x \geqslant 1. \end{cases}$$

On représente ces fonctions sur la figure ci-dessous (f et cg coïncident presque pour $x > 1$ et ne se discernent pas sur le graphe).



Reste maintenant à simuler une variable aléatoire de densité g. Comme g est combinaison linéaire $g = \frac{\mathrm{e}}{a+\mathrm{e}}\, g_1 + \frac{a}{a+\mathrm{e}}\, g_2$ de deux fonctions, de même la fonction de répartition associée s'écrit

$$\mathrm{G} = \frac{\mathrm{e}}{a+\mathrm{e}}\, \mathrm{G}_1 + \frac{a}{a+\mathrm{e}}\, \mathrm{G}_2$$

où

$$\mathrm{G}_1(x) = \begin{cases} x^a & \text{si } 0 \leqslant x \leqslant 1, \\ 1 & \text{sinon,} \end{cases} \qquad \text{et} \qquad \mathrm{G}_2(x) = \begin{cases} 0 & \text{si } x < 1, \\ 1 - \mathrm{e}^{1-x} & \text{sinon.} \end{cases}$$

D'après la formule des probabilités totales, c'est la loi d'une variable suivant la loi G_1 avec une probabilité $p = \mathrm{e}/(a+\mathrm{e})$ et la loi G_2 avec la probabilité $1-p$. Les lois G_1 et G_2 ont ceci d'avantageux que l'on sait bien calculer leur pseudo-inverse. On aboutit à l'algorithme suivant :

Algorithme 9 : Simuler la loi $\gamma(a)$ avec $0 < a \leqslant 1$

```
p ← e/(a + e)
répéter
    u ← rand()
    v ← rand()
    w ← rand()
    si u < p alors
        x ← v^{1/a}
        q ← f(x)/c g(x)                          // q = e^{-x}
    sinon
        x ← 1 − ln(1 − v)
        q ← f(x)/c g(x)                          // q = x^{a-1}
jusqu'à w < q
retourner x
```

Avant l'étape de rejet, x admet pour densité g ; conditionnellement à son acceptation, il admet pour densité f.

25.3 SIMULER UNE LOI NORMALE

§ 373 **Méthode découlant du théorème limite central**

On peut laisser travailler le théorème limite central pour nous et tirer n valeurs aléatoires et indépendantes d'une loi quelconque, puis en renormaliser correctement la somme ; si n est suffisamment grand, la loi de la variable obtenue doit être sensiblement la loi normale centrée réduite.

En pratique, l'appel à la fonction `rand` une dizaine de fois suffit pour obtenir une loi normale réaliste ; moyen rapide et peu cher mais qui montrera vite ses limites si on l'utilise un grand nombre de fois, puisque les grandes valeurs n'apparaîtront jamais. Puisque des algorithmes efficaces et plus précis existent, cette méthode d'approximation n'a aucune raison d'être utilisée en pratique.

§ 374 **Méthode de la transformée inverse**

On peut simuler une loi normale en partant d'une loi uniforme et en suivant la procédure du paragraphe § 370. Cela implique cependant de connaître l'inverse de la fonction de répartition $\mathfrak{N}$ de la loi normale $\mathscr{N}(0\,;\,1)$; cette fonction $\mathfrak{N}^{-1}$, bien que tabulée, n'est pas forcément « connue » des logiciels.

Remarque 25.13 (Formule approchée pour $\mathfrak{N}^{-1}$) Il existe cependant une formule d'approximation, que l'on trouve dans l'ouvrage d'Abramowitz et Stegun[4], qui approche $\mathfrak{N}^{-1}(x)$ sur $]\,0\,;\,\frac{1}{2}\,]$ avec une erreur $\varepsilon(x)$ inférieure à $4\cdot 10^{-4}$:

$$\mathfrak{N}^{-1}(x) = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} + \varepsilon(x)$$

avec

$$t = \sqrt{\ln\left(\frac{1}{x^2}\right)}$$

et

$$\begin{array}{lll} c_0 = 2,515517 & c_1 = 0,802853 & c_2 = 0,010328 \\ d_1 = 1,432788 & d_2 = 0,189269 & d_3 = 0,001308 \end{array}$$

Les valeurs de $\mathfrak{N}^{-1}(x)$ pour $x > 1/2$ sont obtenues par

$$\mathfrak{N}(x) = -\mathfrak{N}(1-x).$$

§ 375 **Méthode de Box-Muller**

Voici une seconde approche pour obtenir une loi normale, connue sous le nom de **méthode de Box-Muller** ; elle permet de simuler non seulement une loi normale, mais aussi un couple (X, Y) de loi normales indépendantes[5]. L'idée est la suivante. Supposons un couple (X, Y) de variables aléatoires indépendantes de même loi normale $\mathscr{N}(0\,;\,\sigma)$ connu ; ce couple (X, Y) admet alors pour densité

$$f(x,y) = \frac{1}{2\pi\sigma^2}\,\mathrm{e}^{-(x^2+y^2)/2\sigma^2}.$$

Cette densité admet une symétrie radiale : elle ne dépend que de la distance à l'origine. Un couple de valeurs (x, y) est plus facilement caractérisé par un couple (r, θ), où $r \geqslant 0$ et $0 \leqslant \theta < 2\pi$. L'idée est alors de déterminer la loi de la distance à l'origine, et de prendre pour loi de l'angle θ une loi uniforme sur $]\,0\,;\,2\pi\,]$.

---

(4). Milton Abramowitz et Irene Stegun, *Handbook of mathematical functions* [1, § 26.6.23]

(5). Ce qui est la forme générique d'une loi normale, exprimée dans la base diagonalisant la matrice de covariance.

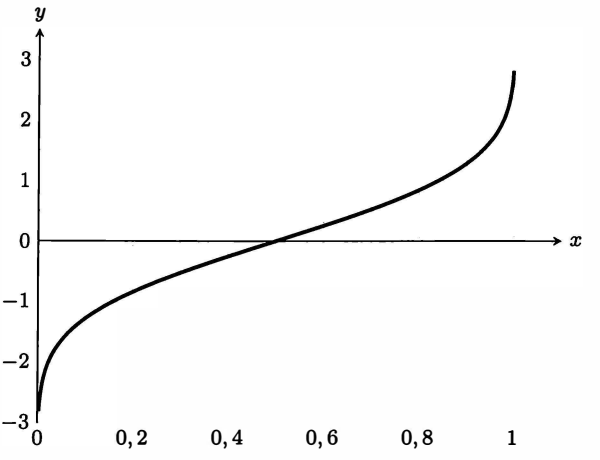


FIGURE 25.2 — *Approximation rationnelle de la fonction de répartition inverse $\mathfrak{N}^{-1}$. À cette échelle, le graphe de $\mathfrak{N}^{-1}$ ne diffère pas de celui de son approximation.*

On pose $\mathrm{R} = \sqrt{\mathrm{X}^2 + \mathrm{Y}^2}$; cette variable suit la **loi de Rayleigh** $\mathscr{R}(\sigma)$. Pour déterminer cette loi, on calcule, pour tout $r \geqslant 0$:

$$\mathbf{P}\{\mathrm{R} \leqslant r\} = \frac{1}{2\pi\sigma^2} \iint_{\mathrm{D}(0,r)} \mathrm{e}^{-(x^2+y^2)/2\sigma^2}\, \mathrm{d}x\, \mathrm{d}y = \frac{1}{\sigma^2} \int_0^r \mathrm{e}^{-\rho^2/2\sigma^2}\, \rho\, \mathrm{d}\rho,$$

ce qui montre que R admet pour densité

$$f_{\mathrm{R}}(r) = \begin{cases} \frac{1}{\sigma^2}\, r\, \mathrm{e}^{-r^2/2\sigma^2} & \text{si } r \geqslant 0 \\ 0 & \text{sinon.} \end{cases}$$

Reste maintenant à simuler une loi de Rayleigh. Un calcul simple montre que :

■ **Lemme 25.14** *Si* E *suit une loi exponentielle* $\mathscr{E}(1/\sigma^2)$, *alors* $\mathrm{R} = \sqrt{2\mathrm{E}}$ *suit la loi* $\mathscr{R}(\sigma)$.

DÉMONSTRATION : Pour tout $r > 0$, $\mathbf{P}\{\mathrm{R} \leqslant r\} = \mathbf{P}\{\mathrm{E} \leqslant r^2/2\} = 1 - \mathrm{e}^{-r^2/2\sigma^2}$ donc R admet pour densité $f_{\mathrm{R}}(r) = (1/\sigma^2)\, r\, \mathrm{e}^{-r^2/2\sigma^2}$. ∎

La méthode est alors la suivante.

■ **Proposition 25.15 (Méthode de Box-Muller)** *Supposons connu un couple* (U, V) *de variables aléatoires indépendantes*[6], *de même loi uniforme sur*]0 ; 1[. *On pose alors* $\Theta = 2\pi\,\mathrm{U}$ *et* $\mathrm{E} = -\ln \mathrm{V}$, *de sorte que* $\mathrm{E} \rightsquigarrow \mathscr{E}(1)$. *On pose enfin* $\mathrm{R} = \sqrt{2\mathrm{E}}$, *de sorte que* $\mathrm{R} \rightsquigarrow \mathscr{R}(1)$. *Les variables* $\mathrm{X} = \mathrm{R}\sin\Theta$ *et* $\mathrm{Y} = \mathrm{R}\cos\Theta$ *sont indépendantes, et suivent la même loi normale centrée réduite* $\mathscr{N}(0\,;\,1)$. *La variable* $\sigma\mathrm{X} + m$ *suit la loi* $\mathscr{N}(m\,;\,\sigma^2)$.

Remarque 25.16 Le calcul numérique étant effectué avec une précision limitée, les grandes valeurs de X et Y sont mal simulées par cet algorithme, car l'aspect discret de V au voisinage de 0 est

(6). En pratique, l'appel répété à un générateur aléatoire permet d'obtenir de telles valeurs aléatoires indépendantes.

très fortement mis en relief par la fonction logarithme. On peut, sans augmenter la précision de la représentation interne des flottants, améliorer la simulation des grandes valeurs en sommant n des variables normales indépendantes et en divisant le résultat par $\sqrt{n}$.

§376 **Simulation d'une loi normale par la méthode du rejet**

Il existe une variante de la méthode précédente permettant d'éviter le calcul d'un sinus et d'un cosinus ; bien que ceux-ci ne soient pas démesurément coûteux(7), il peut être avantageux de les remplacer par une simple méthode de rejet, notamment si l'on a besoin d'obtenir une grande quantité de données aléatoires(8).

Cette variante est fondée sur la remarque suivante : si l'on connaît un couple (U, V) de variables aléatoires dont la loi conjointe est uniforme sur le disque unité

$$\mathrm{D} = \big\{(u,v) \in \mathbb{R}^2 \,;\, u^2+v^2 \leqslant 1\big\},$$

si l'on note $\Delta = \sqrt{\mathrm{U}^2+\mathrm{V}^2}$, $\mathrm{C} = \mathrm{U}/\Delta$ et $\mathrm{S} = \mathrm{V}/\Delta$, alors le couple (C,S) suit la même loi que le couple $(\cos \mathrm{T}, \sin \mathrm{T})$, où T est uniforme sur $[\,0\,;\,2\pi\,]$.

De plus, la variable aléatoire $\mathrm{A} = \Delta^2$ suit une loi uniforme sur $[\,0\,;\,1\,]$, donc $\mathrm{R} = \sqrt{-2\ln \mathrm{A}} = 2\sqrt{-\ln \Delta}$ suit la loi de Rayleigh $\mathscr{R}(1)$. Comme dans le paragraphe précédent, en posant

$$\mathrm{X} = \mathrm{CR} = \frac{2\mathrm{U}}{\Delta}\sqrt{-\ln\Delta} \qquad \text{et} \qquad \mathrm{Y} = \mathrm{SR} = \frac{2\mathrm{V}}{\Delta}\sqrt{-\ln\Delta},$$

le couple (X,Y) vérifie la loi normale centrée réduite sur $\mathbb{R}^2$, donc $\mathrm{X}, \mathrm{Y} \rightsquigarrow \mathscr{N}(0\,;\,1)$ et sont indépendantes.

La difficulté est maintenant de créer un couple (U,V) de loi uniforme sur le disque D. L'astuce est de le faire de manière *algorithmique* : on tire au hasard un couple de réels (u,v), de répartition uniforme sur le carré $[\,-1\,;\,1\,]^2$, et on rejette ceux vérifiant $u^2+v^2 > 1$. L'algorithme de création du couple (x,y) est donc le suivant :

Algorithme 10 : Simuler un couple $(x,y) \rightsquigarrow \mathscr{N}(0\,;\,\mathrm{I}_2)$

répéter
  $u \leftarrow 2 * \mathtt{rand}() - 1$
  $v \leftarrow 2 * \mathtt{rand}() - 1$
  $\delta \leftarrow \sqrt{u^2+v^2}$
jusqu'à $\delta < 1$ **et** $\delta \neq 0$
$x \leftarrow (2u/\delta)\sqrt{-\ln\delta}$
$y \leftarrow (2v/\delta)\sqrt{-\ln\delta}$
retourner (x,y)

Seule une proportion $1-\pi/4 \approx 0,215$ des points est rejetée dans la boucle, ce qui est finalement peu coûteux en terme de calculs : il suffit en moyenne de calculer 127 points pour en avoir 100 non rejetés (voir figure 25.3).

Remarque 25.17 Il existe des généralisations de la méthode du rejet, combinées avec des approximations simples de la fonction réciproque $\mathfrak{N}^{-1}$, permettant de simuler une loi normale. Le lecteur intéressé pourra consulter le grand classique qu'est *The Art of Computer Programming* de D. Knuth [60, § 3.4.1.C.2].

(7). Ils sont bien tabulés, et un recours à une méthode d'approximation rapide, par exemple une méthode de Newton, permet d'accéder à une valeur quelconque en un temps relativement court.

(8). Des résolutions numériques d'équations différentielles stochastiques peuvent faire intervenir des millions de calculs.

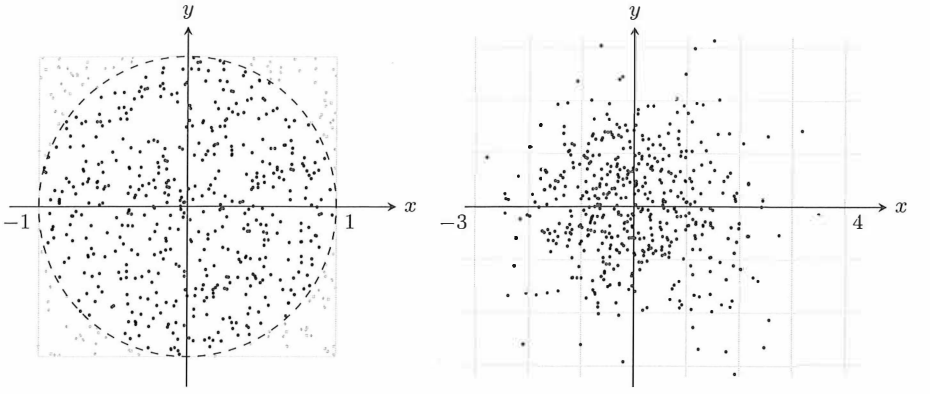


FIGURE 25.3 — À gauche : *620 points, tirés au hasard sur le carré* $[0\,;\,1]^2$, *ont été nécessaires pour obtenir 500 points dans le disque unité.* À droite : *les 500 points précédents sont étalés selon la loi normale* $\mathcal{N}(0\,;\,\mathrm{I}_2)$.

Simulation d'un vecteur gaussien sur $\mathbb{R}^2$ § 377

À partir d'un vecteur gaussien $\mathbf{X} = (\mathrm{X}, \mathrm{Y})$ de loi centrée réduite $\mathcal{N}(\mathbf{0}\,;\,\mathrm{I}_2)$ (c'est-à-dire que X et Y suivent $\mathcal{N}(0\,;\,1)$ et sont indépendantes), on peut créer un vecteur gaussien $\mathbf{X}' = (\mathrm{X}', \mathrm{Y}')$ de caractéristiques données.

Si l'on désire imposer une matrice de covariance V *et une moyenne* $\boldsymbol{m}$, il suffit de déterminer une matrice Λ telle que $\Lambda\,{}^t\Lambda = \mathrm{V}$ et de poser $\mathbf{X}' = \Lambda\mathbf{X} + \boldsymbol{m}$. La matrice Λ peut être obtenue en orthodiagonalisant la matrice symétrique positive V sous la forme $\mathrm{V} = {}^t\Omega\,\mathrm{diag}(\lambda, \mu)\Omega$ et en posant

$$\Lambda = {}^t\Omega \begin{pmatrix} \sqrt{\lambda} & 0 \\ 0 & \sqrt{\mu} \end{pmatrix}.$$

Il est sans doute plus avantageux, du point de vue algorithmique, de considérer la décomposition de Cholesky

$$\mathrm{V} = \mathrm{L}\,{}^t\mathrm{L}$$

où L est une matrice triangulaire inférieure à coefficients diagonaux positifs. Les algorithmes de calcul de cette décomposition sont bien optimisés et de nombreux logiciels de calcul formel la proposent ; dans le cas d'une matrice 2×2, le calcul est presque immédiat :

$$\mathrm{L} = \begin{pmatrix} \sigma_x & 0 \\ \rho\sigma_y & \sqrt{1-\rho^2}\,\sigma_y \end{pmatrix}.$$

On pose alors

$$\mathbf{X}' := \mathrm{L}\mathbf{X} + \boldsymbol{m}.$$

La matrice de covariance de $\mathbf{X}'$ est

$$\mathrm{C}^{\mathbf{X}'} = \mathrm{C}^{\mathrm{L}\mathbf{X}+\boldsymbol{m}} = \mathrm{L}\mathrm{C}^{\mathbf{X}}{}^t\mathrm{L} = \mathrm{L}\,{}^t\mathrm{L} = \mathrm{V}$$

d'après le théorème de changement de base (théorème 15.5 page 334). Si V est définie positive, le vecteur $\mathbf{X}'$ admet pour densité

$$g(\boldsymbol{x}') = f\big(\mathrm{L}^{-1}(\boldsymbol{x}' - \boldsymbol{m})\big) \cdot \left|\frac{\mathrm{D}\boldsymbol{x}}{\mathrm{D}\boldsymbol{x}'}\right|$$

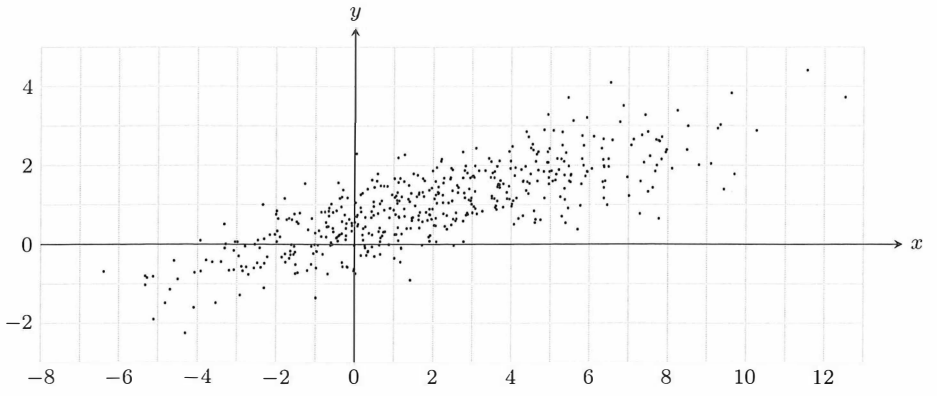


FIGURE 25.4 — *Les 500 points précédents, après transformation affine, simulent un couple* (X', Y') *de coefficient de corrélation* $\rho = 0,8$.

où $D\boldsymbol{x}/D\boldsymbol{x}'$ est le jacobien du $\mathscr{C}^1$-difféomorphisme $\boldsymbol{x}' \mapsto \boldsymbol{x}$, valant $1/\det L$[9], c'est-à-dire

$$g(\boldsymbol{x}') = \frac{1}{2\pi\sqrt{\det V}}\, e^{-\langle \boldsymbol{x}'-\boldsymbol{m} | V^{-1}(\boldsymbol{x}'-\boldsymbol{m})\rangle/2},$$

comme il se doit.

Exemple 25.18 On cherche à modéliser un vecteur gaussien (X', Y') tel que

$$\mathbf{E}(X') = 2 \qquad \mathbf{E}(Y') = 1 \qquad \sigma_{X'} = 3 \qquad \sigma_{Y'} = 1 \qquad \text{et} \qquad \rho(X', Y') = 0,8.$$

La matrice de covariance du couple (X', Y') est alors

$$V = \begin{pmatrix} 9 & 2,4 \\ 2,4 & 1 \end{pmatrix}$$

dont la décomposition de Cholesky est

$$V = L\,{}^tL \qquad L = \begin{pmatrix} 3 & 0 \\ 0,8 & 0,6 \end{pmatrix}.$$

Si (X, Y) est un couple gaussien de loi $\mathscr{N}(\mathbf{0}\,;\, I_2)$, on pose donc

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = L \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 3X + 2 \\ 0,8\,X + 0,6\,Y + 1 \end{pmatrix}.$$

Les tirages normaux de la figure 25.3 sont envoyés par cette transformation affine en les points de la figure 25.4.

Exemple 25.19 Supposons maintenant que l'on cherche un couple (X'', Y'') vérifiant

$$\mathbf{E}(X'') = 2 \qquad \mathbf{E}(Y'') = 1 \qquad \sigma_{X''} = 3 \qquad \sigma_{Y''} = 1 \qquad \text{et} \qquad \rho(X'', Y'') = 1$$

(c'est-à-dire que seul le coefficient de corrélation a changé par rapport à l'exemple précédent), les matrices V et L deviennent

$$V = \begin{pmatrix} 9 & 3 \\ 3 & 1 \end{pmatrix} \qquad \text{et} \qquad L = \begin{pmatrix} 3 & 0 \\ 1 & 0 \end{pmatrix}.$$

Tous les points sont donc désormais sur la droite d'équation $3(x-2) = y-1$, ce qui est le résultat attendu puisque X'' et Y'' sont presque sûrement liés[10] :

$$\mathbf{P}\left\{ \frac{X'' - \mathbf{E}(X'')}{\sigma_{X''}} = + \frac{Y'' - \mathbf{E}(X'')}{\sigma_{Y''}} \right\} = 1.$$

Cette construction se généralise sans difficulté à des vecteurs gaussiens sur $\mathbb{R}^n$.

---

(9). Cette formule se retrouve heuristiquement par invariance de la forme $f(\mathbf{X})\,D\mathbf{X} = g(\mathbf{X}')\,D\mathbf{X}'$; elle a été établie dans le théorème 15.10 page 337.

(10). Revoir le théorème 12.28 page 270.

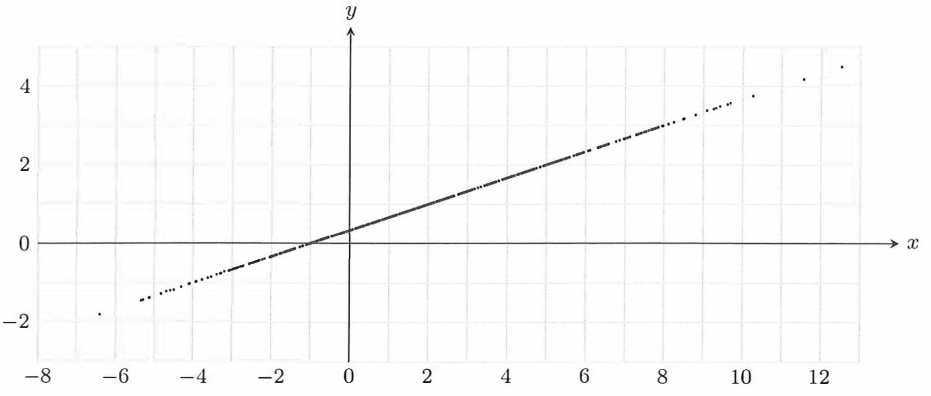


FIGURE 25.5 — *Les mêmes 500 points, après transformation affine non bijective, simulent un couple* (X'', Y'') *de coefficient de corrélation* $\rho = 1$.

25.4 REMARQUES SUR LES GÉNÉRATEURS D'UNE LOI UNIFORME

Remarques préliminaires § 378

Le but d'un générateur uniforme est de simuler une suite « aléatoire » de nombres « indépendants », de loi uniforme sur $[\,0\,;\,1\,[$.

- La contrainte de l'uniformité semble simple, car il est aisé de construire des suites $(x_n)_{n\geqslant 0}$ de nombres dont la densité soit uniforme dans $[\,0\,;\,1\,[$: si α est irrationnel, la suite de terme général $\langle n\alpha\rangle$, où $\langle x\rangle = x - \lfloor x\rfloor$ désigne la partie fractionnaire, vérifie cette propriété[11]. Mais les termes successifs sont, non seulement loin d'être indépendants, mais de plus hautement prévisibles, ce qui fait que ces suites ne peuvent pas servir dans la plupart des applications.
- La simple définition de ce qu'est une « suite aléatoire » est en réalité extrêmement délicate ; c'est une problématique que nous n'aborderons pas, et qui a donné naissance à tout un pan de théorie mathématique [21].

On se contentera de définir des suites $(x_n)_{n\geqslant 0}$ de réels qui vérifient des bonnes propriétés statistiques, comme l'équirépartition de ses valeurs, l'équirépartition des couples (x_n, x_{n+1}) dans $[\,0\,;\,1\,]^2$, mais également celle des couples (x_n, x_{n+2}), etc. En pratique, un « bon » générateur aléatoire est celui qui passe le plus de tests statistiques.

Remarquons d'abord que, dans la pratique, on ne calcule pas une suite de réels mais une suite d'entiers. Un ordinateur stocke une donnée numérique sur un nombre majoré de *bits*. Il existe par conséquent un plus grand entier w représentable par le système, et tout réel de $[\,0\,;\,1\,]$ peut être approché par une fraction k/m, où k est un entier compris entre 0 et un réel $m \leqslant w$. On cherche donc à fabriquer une suite $x_0, x_1, \ldots$ d'entiers aléatoires[12], prenant leurs valeurs entre 0 et m,

(11). On dit qu'elle est *équidistribuée*, cf. § 300.

(12). Passer de la notion de suite de réels aléatoires à celle de suite d'entiers aléatoires n'ôte rien à la difficulté de définir ce que l'on entend par « aléatoire ». S'il est évident pour chacun que la suite « 1010101010101 ⋯ » n'est pas aléatoire, ce l'est peut-être moins pour la suite

0110101000101000101000100000 1 ⋯

... du moins jusqu'à ce qu'on rende compte que le n-ième entier est égal à 1 si et seulement si n est premier. Une suite « aléatoire » ne doit pas — entre autres — recéler de *structure identifiable*. C'est évidemment une propriété difficile à mathématiser ! Une façon de quantifier

et on posera $u_n = x_n/m$ pour simuler une variable de loi uniforme sur $[0\,;\,1[$[13]. En pratique, m est « grand » : de l'ordre de $2^{31} \approx 2,15 \cdot 10^9$, ou de $2^{64} \approx 4,6 \cdot 10^{18}$.

Le principe de tout générateur aléatoire informatisé est le suivant : on part d'une valeur initiale x_0 (la **graine**, ou *seed* en anglais), et on lui applique un algorithme qui transforme x_0 en x_1 ; à ce même algorithme, on fournit alors x_1 en entrée, le résultat retourné étant x_2. De manière générale, le générateur permettra de passer de x_n à x_{n+1}. Des généralisations existent (on part de plusieurs valeurs initiales, la récurrence est d'ordre supérieur à 2,...) mais le principe général reste le même.

Bien évidemment, cette manière de faire est entièrement *déterministe* : si l'on recommence l'expérience en partant de la même valeur x_0, on obtiendra la même suite $x_1, x_2, \ldots$ de valeurs[14] — les valeurs obtenues ne peuvent plus être considérées comme « aléatoires », ni « indépendantes » ! On parle d'ailleurs plutôt de **générateur pseudo-aléatoire**. En pratique, toutefois, ce défaut n'est pas toujours considéré comme grave : si l'algorithme est suffisamment bien étudié, le caractère déterministe ne sera détectable ni par un être humain, ni par une machine. Or, entre le hasard et l'illusion du hasard, il n'y a, peut-être, aucune différence. Un générateur sera réputé correct s'il passe un certain nombre de tests statistiques permettant de déceler des écarts entre les suites qu'il construit et ce que dit la théorie.

Remarque 25.20 (Algorithmes compliqués) Ce n'est pas parce que l'algorithme est compliqué qu'il est performant. Quiconque a essayé de construire lui-même un générateur aléatoire a dû être tenté d'appliquer une procédure complexe pour passer de x_n à x_{n+1}. Donald Knuth [60] propose un exemple d'algorithme tout de son crû, d'une sophistication redoutable... mais qui se révèle très vite extrêmement mauvais (il stationne en un point fixe après un tout petit nombre d'étapes).

Exemple 25.21 (Algorithme middle square *de von Neumann)* En 1946, John (János) von Neumann proposa l'algorithme suivant : on part d'un entier x à $2n$ décimales :

$$x_0 = (x_1\, x_2 \cdots x_{2n}),$$

on l'élève au carré

$$x_0^2 = (y_1\, y_2 \cdots y_n\, \underline{y_{n+1} \cdots y_{3n}\, y_{3n+1}}\, y_{3n+2} \cdots y_{4n})$$

et on ne retient que la moitié du milieu des chiffres de x^2 :

$$x_1 = f(x_0) = (y_{n+1} \cdots y_{3n}\, y_{3n+1}).$$

On utilise alors la séquence des itérés : $x_2 = f(x_1) = f^2(x_0), \ldots, x_k = f^k(x), \ldots$ comme suite « aléatoire ».

En réalité, cet algorithme ne produit pas des suites de bonne qualité, et n'a plus qu'un intérêt historique : pour un grand nombre de valeurs initiales x_0, la suite produite a une période très courte. Un exemple extrême est donné pour la valeur initiale $x_0 = 3792$, qui engendre une suite stationnaire :

$$3792 \quad \to \quad 14\underline{3792}64 \quad \to \quad 3792$$

---

cette imprédictibilité est d'exiger qu'elle ne puisse pas être engendrée par un programme : mais dans ce cas on élimine *de facto* la possibilité de créer une telle suite par un ordinateur. On perçoit la difficulté de définir ce qu'est un générateur « aléatoire » d'entiers.

[13]. Comme variante, on pose $u_n = x_m/(m+1)$ si l'on cherche une loi uniforme sur $[0\,;\,1[$, mais la probabilité que x_m soit égal précisément à m est, si le générateur est réussi, très faible.

[14]. Il peut être utile de programmer un générateur aléatoire pour qu'il conserve en mémoire la dernière valeur donnée, même si l'on éteint (proprement) la machine, afin de ne pas réutiliser la même suite de nombres. À défaut, quelques mésaventures peuvent arriver, comme c'est arrivé au Casino de Montréal, en 1994 : un joueur a pu noter les successions de nombres produites par les « générateurs aléatoires » du keno, qui étaient réinitialisés chaque jour, et empocher — tout à fait légalement — un demi million de dollars.

Certains générateurs peuvent utiliser une source extérieure de hasard pour améliorer leur efficacité et devenir réellement imprédictibles. Nous ne retiendrons que les générateurs autonomes, c'est-à-dire ceux qui sont essentiellement caractérisés par

- le choix de la graine x_0 [15] ;
- l'algorithme (déterministe) permettant de créer x_{n+1} en fonction de x_n.

Générateur à récurrence linéaire § 379

Un des générateurs les plus fréquemment utilisés, décrit en 1948 par un théoricien des nombres du nom de Derrick Henry LEHMER, est basé sur une relation de récurrence linéaire (donc de la forme $f(x) = ax$), mais *modulaire* (c'est-à-dire calculée modulo un certain entier). Le modèle de Lehmer était précisément

$$\begin{cases} x_0 = 47594118 \\ x_{n+1} = 23x_n \mod 10^8 + 1. \end{cases}$$

Grâce au calcul modulaire, on ne « sort pas » de l'ensemble $\{0, 1, 2, \ldots, 10^8\}$ [16]. L'opération « modulo m » est l'équivalent d'un jeu sur une roulette : peu importe le nombre de tours faits par la bille, seul compte sa position (modulo 2π) sur le disque de la roulette ; dans le modèle de Lehmer, la bille effectue plusieurs tours de roulette (entre 0 et 23), seul compte le reste modulo 100 000 001.

Pus généralement, les modèles de récurrence linéaire modulaire couramment utilisés sont de la forme

$$x_{n+1} = (a\,x_n + c) \mod m$$

où m est un entier non nul, a et c des entiers compris entre 0 et $m-1$. Bien sûr, un tel algorithme conduit nécessairement à une suite périodique : puisque l'ensemble des valeurs prises est $\mathrm{K} = \{0, 1, \ldots, m-1\}$, la suite finira par prendre une valeur déjà prise : il existe donc des entiers $0 \leqslant k < \ell < m$ tels que $x_k = x_\ell$. L'entier $p = k - \ell$ est alors une période de la suite : $x_{n+p} = x_n$ pour tout $n \geqslant k$.

Remarque 25.22 Il se peut que la suite ne soit périodique qu'à partir d'un certain rang, et notamment que la valeur x_0 ne soit jamais prise de nouveau, pour peu que l'application de récurrence $f : x \mapsto (ax + x) \mod m$ ne soit pas injective ; par exemple, $x_0 = 0$ et $x_{n+1} = 2x_n + 1 \mod 2q$ ne donnera, à partir de $n = 1$, que des entiers impairs.

Si a est un entier premier avec m, alors a est inversible dans $\mathbb{Z}_m := \mathbb{Z}/m\mathbb{Z}$, et la relation de récurrence est inversible. Sinon, la suite de terme général $a\,x_n$ prend ses valeurs dans un ensemble de cardinal $m/\,\mathrm{PGCD}(a, m)$, et il en est de même de la suite $(x_n)_{n\geqslant 1}$.

---

(15). Comme on l'a dit, le choix de ce premier terme x_0 détermine entièrement la suite pseudo-aléatoire qui sera engendrée par l'algorithme. Lorsque la qualité et l'imprédictibilité absolue d'un générateur aléatoire sont des données cruciales, on a toujours la possibilité de coupler l'ordinateur à une « source externe d'aléatoire ». Une possibilité facile à mettre en œuvre est de choisir comme premier terme le nombre donné par les quelques derniers *bits* de l'horloge interne de l'appareil, qui donne le temps écoulé depuis une date conventionnelle, et ce dans une unité de temps très petite. Des phénomènes naturels imprévisibles ont parfois été employés : désintégration de noyaux radioactifs, bruit atmosphérique (on mesure la déformation de l'image d'une étoile due à la turbulence), etc. Il est même possible d'acheter une clé USB comprenant un détecteur de photons associé à une lame semi-transparente, qui garantit un flux aléatoire de zéros et de uns (coût en 2013 : un milliers d'euros). La société qui le commercialise (IDQ — `idquantique.com`) fournit, entre autres, la Loterie Romande.

(16). En pratique, on confond un élément de $\mathbb{Z}/m\mathbb{Z}$, c'est-à-dire une classe d'équivalence, et l'unique représentant de cette classe modulo m. Ainsi, lorsque l'on écrit $x = 13 \mod 8$, on entend ici $x = 5$.

Une famille répandue de générateurs linéaires est celle où m est un nombre premier (par exemple le plus grand entier premier inférieur à w), ce qui permet de travailler dans l'ensemble $\mathbb{F}_m = \mathbb{Z}/m\mathbb{Z}$ qui est un corps.

La suite définie par la relation $x_{n+1} = a\,x_n + c$ vérifie

$$x_n = \left(a^n\,x_0 + \frac{a^n - 1}{a - 1}\,c\right) \mod m.$$

On aura intérêt à choisir a tel que $a^k \not\equiv 1 \mod m$ pour tout $1 < k < m - 1$ (le théorème de Lagrange imposant de toute façon $a^{m-1} \equiv 1 \mod m$) : le nombre a est ce que l'on appelle une **racine primitive** du corps $\mathbb{F}_m = \mathbb{Z}/m\mathbb{Z}$.

Parmi les générateurs congruentiels linéaires, ceux pour lesquels $c = 0$, comme celui de Lehmer, sont couramment utilisés. Ils vérifient

$$x_n = a^n\,x_0.$$

Bien évidemment, si la valeur 0 est atteinte, la suite reste alors stationnaire en 0, ce qui est une catastrophe. Pour éviter cet écueil, on peut choisir $x_0 \neq 0$ et m premier, ce qui assure que $a^m \neq 0$ si $a \not\equiv 0 \mod m$[17].

De même, dès que a^n prend la valeur $1 \mod m$, une période est atteinte et la suite reprend les mêmes valeurs. Afin d'avoir la période la plus grande possible, on choisit comme valeur de a une racine primitive de $\mathbb{F}_m$, en vertu d'un théorème, dû à Gauss :

■ **Théorème 25.23** *Soit m un entier premier. La période de la suite définie par x_0 et la relation de récurrence $x_{n+1} = a\,x_n$ est égale à $m - 1$ si et seulement si a est une racine primitive du corps $\mathbb{F}_m$.*

Un bon choix peut être $a = 16\,807$ et $m = 2^{31} - 1$, de période effectivement égale à $m - 1$.

§ 380 **Récurrences d'ordre supérieur, ou non linéaires**

On se borne à signaler ici qu'il existe de nombreuses méthodes basées sur des récurrences d'ordre supérieur à 1, c'est-à-dire que x_{n+1} est calculé en fonction, non de x_n seul, mais d'autres termes précédents. Un des avantages de ce genre de méthode est de permettre des périodes plus longues que celles obtenues par des récurrences d'ordre 1, par exemple

$$x_n = (x_{n-2} + x_{n-1}) \mod m$$

ou

$$x_n = (x_{n-24} + x_{n-55}) \mod m.$$

Enfin, on peut tenter des récurrences non linéaires, comme $x_{n+1} = x_n(x_n + 1) \mod 2^\ell$. Nous renvoyons encore le lecteur intéressé à Knuth [60] ainsi qu'à Rousseau & Saint-Aubin [82].

---

(17). Ne pas choisir m premier limite également la période de la suite $(x_n)_{n\geqslant 0}$. Si d est un diviseur commun à x_n et m, alors c'est un diviseur commun à tous les x_i ($i \geqslant n$), ce qui annihile le caractère aléatoire de la suite. On souhaite donc que x_n et m soient premiers entre eux pour tout $n \in \mathbb{N}$, c'est-à-dire que chaque x_n doit vivre dans un sous-ensemble de cardinal $\varphi(m)$: la période de la suite est par conséquent majorée par $\varphi(m)$.

Brève remarque historique § 381

En 1955 est paru un ouvrage, *A Million Random Digits with 100,000 Normal Deviates*, édité par RAND Corporation [18] ; il proposait — comme son titre l'indique — un million de chiffres aléatoires (de 0 à 9), regroupés par paquets de 10, à raison de 5 paquets par ligne. Ces nombres permettent de créer des valeurs aléatoires de la précision voulue. Par exemple, pour obtenir un nombre de l'intervalle $[0\,;\,1]$ avec une précision de 10^{-10}, il suffit de choisir un bloc de 10 chiffres $a_1, \ldots, a_{10}$ et de lui associer le réel

$$x = \sum_{k=1}^{10} \frac{a_k}{10^k}.$$

...	...	...	...	...	...	...	...	...	...	...
13524	28710	46590	21036	61392	38973	65331	06124	28554	86731	82209
13525	32320	42619	42963	95054	51492	47584	09168	55677	43012	04401
13526	30500	54025	80170	49840	15058	90035	31762	28554	97240	28676
13527	33911	90647	00626	56623	20961	53883	79541	40767	23286	69446
13528	75351	72120	40287	99512	26341	29144	76170	98729	42150	15266
...	...	...	...	...	...	...	...	...	...	...

TABLE 25.1 — *Cinq lignes de l'ouvrage* A Million Random Digits.

Avec cet ouvrage, des données numériques aléatoires d'une grande fiabilité devenaient accessibles, par exemple pour des simulations de type Monte Carlo. Ces nombres ont été produits par un générateur aléatoire informatique, et très scrupuleusement vérifiés (par divers tests statistiques) ; bien que des ouvrages de ce type fussent déjà paru quelques décennies plus tôt, celui-ci était le premier à offrir une telle qualité et connut un grand succès. Bien entendu, la collecte de ces données à la main était très fastidieuse ; bientôt, les ordinateurs, devenus plus accessibles aux professionnels, supplantèrent les tables numériques.

Résumé de quelques formules continues

Dans le tableau ci-dessous, on résume les formules permettant de créer une variable de loi voulue, en partant de deux variables u et v indépendantes de loi uniforme sur $]\,0\,;\,1\,[$.

Loi	Formule
Loi uniforme sur $[a\,;\,b]$	$a + (b-a)u$
Loi exponentielle $\mathscr{E}(\alpha)$	$-\ln(u)/\alpha$
Loi normale $\mathscr{N}(0\,;\,1)$	$n = \cos(2\pi u) \cdot \sqrt{-2\ln v}$
Loi normale $\mathscr{N}(m\,;\,\sigma^2)$	$m + \sigma n$

EXERCICES

♦ ***Exercice 25.1*** Avec quelle formule (utilisant `rand`, nombre uniformément compris entre 0 et 1) peut-on obtenir un nombre valant $+1$ avec la probabilité $1/2$, et -1 sinon ?

♦ ***Exercice 25.2*** Simuler informatiquement les variables X et X′ de la remarque 20.19 page 425 et afficher des courbes de moyennes successives semblables à celles proposées.

♦ ***Exercice 25.3*** Déterminer un algorithme permettant de simuler une variable aléatoire X dont la loi est donnée par $\mathbf{P}\{X = (-1)^k/k\} = 2^{-k}$ pour tout entier $k \geqslant 0$.

♦ ***Exercice 25.4 (Middle-square)*** Écrire un programme prenant en entrée deux entiers n et p, ainsi qu'un nombre entier x_0 à n chiffres (les chiffres manquants sont remplacés par des zéros sur la gauche) et qui engendre une suite de p entiers selon la méthode du *middle-square* de von Neumann.

Écrire un programme prenant en entrée n et x_0, et retournant la longueur de la période de la suite ainsi créée.

♦ ***Exercice 25.5 (À vos machines)*** Proposer un algorithme pour simuler une variable aléatoire X à valeurs dans $[0\,;\,2\pi]$, de densité $f(x) \propto |\sin x|$:

i) par la méthode du pseudo-inverse ;

ii) par la méthode du rejet (on estimera alors le nombre de tirages nécessaires pour obtenir n valeurs indépendantes de X).

♦ ***Exercice 25.6 (Simuler la loi $\gamma(a)$)*** On se propose de simuler une loi $\gamma(a)$, pour n'importe quel réel a strictement positif. On admettra que, si $X \rightsquigarrow \gamma(a)$ et $Y \rightsquigarrow \gamma(b)$ sont indépendantes, alors $X + Y \rightsquigarrow \gamma(a + b)$.

i) Soit $k \geqslant 1$ un entier. Proposer un algorithme pour simuler une variable aléatoire de loi $\gamma(k)$.

ii) Conclure.

♦ ***Exercice 25.7*** Simuler la loi du sauteur de haies (exemple 6.14 page 132).

Indications

◊ *Indication 25.3 :* Il suffit de simuler une variable aléatoire de loi $\mathbf{P}\{Y = k\} = 2^{-k}$ puis de poser $X := (-1)^Y/Y$.

◊ *Indication 25.6 :* Une somme de variables aléatoires exponentielles indépendantes suit une loi γ.

SOLUTIONS

♦ **Solution 25.1** La quantité `2 rand()` est un réel de $[0\,;\,2[$, donc $\lfloor$`2 rand()`$\rfloor$ vaut 0 ou 1 avec la probabilité $1/2$. On peut donc poser $x =$ `2` $\lfloor$`2rand()`$\rfloor - 1$ par exemple.

♦ **Solution 25.2** On calcule facilement que la fonction de répartition de la variable X vaut, en un entier k,

$$\mathrm{F}(k) = 1 - \frac{1}{(k+1)^2},$$

On tire alors y au hasard et uniformément dans $[0\,;\,1[$ et on cherche k tel que $1/\sqrt{1-y}$ est compris entre k et $k+1$; il suffit donc de poser $k = \left\lfloor \dfrac{1}{\sqrt{1-y}} \right\rfloor$.

De même, pour la variable X′, on tire y uniformément dans $[0\,;\,1[$ et on pose $k = \lfloor (1-y)^{-1/4} \rfloor$.

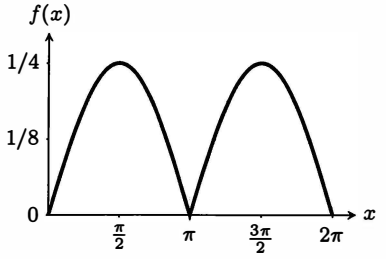
♦ **Solution 25.5**

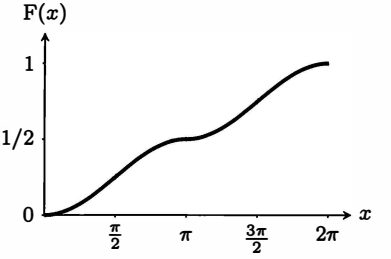
i) La densité de X est $f(x) = \frac{1}{4}|\sin x|$ sur $[0\,;\,2\pi]$; sa fonction de répartition vaut $\mathrm{F}(x) = 0$ si $x < 0$, $\mathrm{F}(x) = 1$ si $x > 2\pi$ et

$$\mathrm{F}(x) = \begin{cases} \frac{1}{4}(1 - \cos x) & \text{si } 0 \leqslant x \leqslant \pi \\ \frac{1}{4}(3 + \cos x) & \text{si } \pi \leqslant x \leqslant 2\pi. \end{cases}$$

Cette fonction établit une bijection de $[0\,;\,2\pi]$ sur $[0\,;\,1]$, que l'on inverse en séparant les cas. On définit alors une variable aléatoire U de loi uniforme sur $[0\,;\,1]$ et on pose

$$\mathrm{X} = \begin{cases} \operatorname{Arc}\cos(1 - 4\mathrm{U}) & \text{si } 0 \leqslant \mathrm{U} \leqslant 1/2 \\ 2\pi - \operatorname{Arc}\cos(4\mathrm{U} - 3) & \text{si } 1/2 \leqslant \mathrm{U} \leqslant 1. \end{cases}$$





ii) On peut bien sûr jouer les paresseux et choisir comme densité g celle de la loi uniforme sur $[0\,;\,2\pi]$! Avec les notations du paragraphe § 371, on pose donc $g(x) = 1/2\pi$ sur $[0\,;\,2\pi]$ et $g(x) = 0$ en dehors, et $c = 2\pi/4 = \pi/2$.
On obtient l'algorithme suivant :

Algorithme 11 : Simuler une loi de densité $\propto |\sin x|$

répéter
  $x \leftarrow 2\pi \cdot$ `rand()`
  $u \leftarrow$ `rand()`
jusqu'à $u < |\sin x|$
retourner x

La proportion de points acceptés est d'environ $2/\pi \approx 0,637$.

♦ **Solution 25.6**

i) Si $\mathrm{X}_1, \ldots, \mathrm{X}_k$ sont des variables aléatoires indépendantes, de même loi exponentielle $\mathscr{E}(1)$, alors $\mathrm{Y} := \mathrm{X}_1 + \cdots + \mathrm{X}_k$ suit la loi $\gamma(k)$. En utilisant l'algorithme de création de lois exponentielles, on peut alors écrire :

Algorithme 12 : Simuler une loi $\gamma(k)$, $k \in \mathbb{N}$

pour $i = 1$ *à* k **faire**
  $u_i \leftarrow$ `rand()`
retourner $-\ln(u_1 u_2 \cdots u_k)$

ii) Dans le cas général, on note $k = \lfloor a \rfloor$ et $a' = a - \lfloor a \rfloor$. L'exemple 25.12 page 519 donne un algorithme pour simuler la loi $\gamma(a')$.

Algorithme 13 : Simuler $\gamma(a)$, $a > 0$

tirer x selon la loi $\gamma(k)$
tirer y selon la loi $\gamma(a')$
retourner $x + y$

Conseils de lecture

Le lecteur intéressé par la notion de « suite aléatoire » (et plus généralement par la théorie de l'information) pourra consulter, par exemple :

- Jean-Paul Delahaye, *Information, complexité et hasard* [21].
- Donald Knuth, *The Art of Computer Programming* [60].
- Bernard Bercu and Djèlil Chafaï, *Modélisation stochastique et simulation* [7].

Chapitre 26

Méthodes de Monte-Carlo

L'analyse est le seul instrument dont on se soit servi jusqu'à ce jour dans la science des probabilités, pour déterminer & fixer les rapports du hasard ; la Géométrie paroissoit peu propre à un ouvrage aussi délié ; cependant si l'on y regarde de près, il sera facile de reconnoître que cet avantage de l'Analyse sur la Géométrie, est tout-à-fait accidentel, & que le hasard, selon qu'il est modifié & conditionné, se trouve du ressort de la géométrie aussi-bien que de celui de l'analyse.

Comte de BUFFON, Essai d'Arithmétique morale, *chap.* XXIII

La méthode de Monte-Carlo a été développée à Los Alamos par John von Neumann, Stanisław Ulam, Enrico Fermi et Nicholas Metropolis (qui en invente le nom) pour effectuer des calculs numériques que les algorithmes déterministes rendaient rédhibitoirement longs [76].

Elle a connu une fortune prodigieuse et ses avatars permettent de résoudre de manière très précise de nombreux problèmes : calcul d'intégrales, résolution d'équations différentielles, recherche de minimas, recherche d'algorithmes optimaux. Le domaine d'application est en réalité si vaste que l'on ne peut qu'à peine en effleurer la surface dans les quelques pages qui suivent.

Des méthodes basées sur des tests aléatoires permettent également de dire, très rapidement et avec un risque très faible d'erreur, si un entier est premier ou non[1] *(§ 386). Des études statistiques permettent également de deviner si une liste de données a été falsifiée ou non (§ 389).*

(1). La commande `isprime` du logiciel Maple, par exemple, effectue un test probabiliste de primalité, et non une étude déterministe, qui serait astronomiquement trop coûteuse en temps pour des grands entiers.

26.1 ESTIMATION D'UNE MOYENNE

§ 382 **L'aiguille de Buffon**

Buffon[2] s'était interrogé, en 1777, sur la probabilité qu'une aiguille de longueur a, lâchée sur un parquet dont les lames ont une largeur $b \geqslant a$, coupe une rainure[3]. Le calcul peut être mené de manière très géométrique. On note y l'ordonnée du centre de l'aiguille, mesurée à partir d'une lame. Pour des raisons de symétrie, on peut toujours se ramener au cas où $0 \leqslant y \leqslant b/2$. On note θ l'angle que fait l'aiguille avec la direction orthogonale aux bords des lames (voir figure 26.1 page suivante) ; pour des raisons de symétrie, on peut encore se ramener au cas où $0 \leqslant \theta \leqslant \pi/2$. On modélise donc le lancer d'une aiguille par la donnée de deux variables aléatoires Y et Θ, de lois respectives

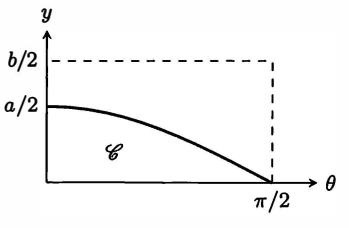
$$\mathrm{Y} \rightsquigarrow \mathscr{U}\big([0\,;\,b/2]\big) \qquad \text{et} \qquad \Theta \rightsquigarrow \mathscr{U}\big([0\,;\,\pi/2]\big).$$

L'événement

$$\mathrm{C} : \text{« l'aiguille coupe le bord de la lame »}$$

a lieu si et seulement si $(a/2)\cos\Theta > \mathrm{Y}$. Si l'on admet que les variables aléatoires Y et Θ sont indépendantes, ce qui semble raisonnable dans une expérience telle que celle décrite[4], alors le couple (Θ, Y) est absolument continu et admet une densité constante égale à $\frac{2}{\pi}\cdot\frac{2}{b} = \frac{4}{\pi b}$ sur le rectangle $\mathrm{R} = \big[0\,;\,\frac{\pi}{2}\big] \times \big[0\,;\,\frac{b}{2}\big]$, et nulle en dehors. Ainsi, en notant

$$\mathscr{C} := \left\{(\theta, y) \in \mathrm{R}\ ;\ y < \frac{a\cos\theta}{2}\right\},$$



on a

$$\mathbf{P}(\mathrm{C}) = \mathbf{P}\left\{\mathrm{Y} < \frac{a}{2}\cos\theta\right\} = \frac{4}{\pi b}\iint_{\mathscr{C}} \mathrm{d}y\,\mathrm{d}\theta = \frac{4}{\pi b}\int_0^{\pi/2}\frac{a}{2}\cos\theta\,\mathrm{d}\theta = \frac{2a}{\pi b}.$$

Pour simplifier, nous envisagerons désormais le cas où $a = b$ et noterons

$$p := \frac{2}{\pi}$$

---

(2). Georges LECLERC, comte de BUFFON (1707–1788), fameux naturaliste et grand écrivain, auteur d'une monumentale *Histoire naturelle*, s'est également intéressé à la théorie des probabilités.

(3). Buffon examinait en fait toute une série de problèmes du même ordre : *Dans une chambre parquetée ou pavée de carreaux égaux, d'une figure quelconque, on jette en l'air un écu ; l'un des joueurs parie que cet écu, après sa chûte, se trouvera à franc-carreau, c'est-à-dire, sur un seul carreau ; le second parie que cet écu se trouvera sur deux carreaux, c'est-à-dire qu'il couvrira un des joints qui les sépare. [...] Je me suis amusé à en faire le calcul [...].* Suivent quelques résultats selon que les carreaux sont carrés, triangulaires, en losange ou hexagonaux. Puis il se penche sur le fameux problème de l'aiguille.

(4). Rappelons encore une fois que l'indépendance du couple (Y, Θ) ne peut pas être prouvée, mais doit être *postulée* lors de l'étape de modélisation de l'expérience. La confrontation des résultats (statistiques) de l'expérience avec les prédictions théoriques peuvent, le cas échéant, discréditer l'hypothèse d'indépendance.

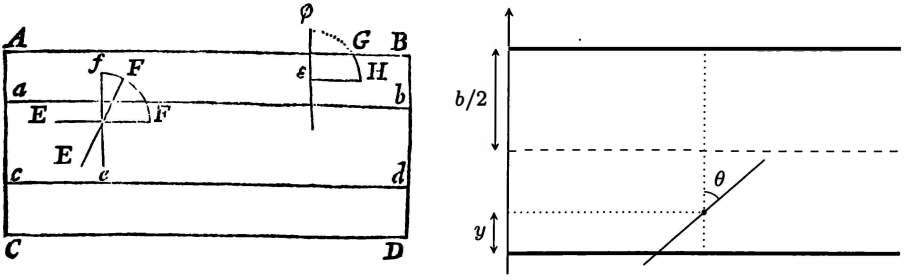


FIGURE 26.1 — À gauche : *la figure explicative de* BUFFON. À droite : *une version modernisée.*

la probabilité qu'une aiguille lancée sur le parquet de Buffon coupe une rainure.

On suppose que le nombre de lancers d'aiguilles est n, et on note X le nombre d'aiguilles coupant une rainure. La variable aléatoire X suit une loi binomiale $\mathscr{B}(n,p)$. Alors le nombre moyen d'aiguilles coupant une rainure est donné par l'espérance $\mathbf{E}(\mathrm{X}) = pn$, et la variance de X est $\sigma^2 = np(1-p)$. Si on lance beaucoup d'aiguilles, le rapport X/n doit être proche de $p = 2/\pi \approx 0,64$, ce qui permet d'accéder « expérimentalement » à une valeur numérique de π.

La question qui vient naturellement à l'esprit est : pour obtenir une estimation satisfaisante de π par ce procédé, combien d'aiguilles dois-je lancer ? Définissons « satisfaisante » par deux données numériques : la précision voulue, et le risque accepté (c'est-à-dire la probabilité maximale que l'évaluation numérique diffère de la véritable valeur d'une quantité supérieure à la précision souhaitée[5]). Par exemple, *si je veux une estimation de* π *à* 10^{-3} *près, combien d'aiguilles dois-je lancer pour être sûr du résultat avec une probabilité de* 99 % *?*

Une première majoration peut être obtenue par l'inégalité de Bienaymé-Tchebychev :

$$\mathbf{P}\left\{\left|\frac{\mathrm{X}}{n} - p\right| > \frac{\varepsilon\sigma}{n}\right\} = \mathbf{P}\{|\mathrm{X} - np| > \varepsilon\sigma\} \leqslant \frac{1}{\varepsilon^2}.$$

On cherche à rendre cette probabilité inférieure à 1% ; on pose donc $\varepsilon = \sqrt{100} = 10$. Par ailleurs, une incertitude $\delta\pi = 10^{-3}$ sur π correspond à une incertitude δp sur p donnée par

$$\delta p = \delta\left(\frac{2}{\pi}\right) \approx \frac{2\,\delta\pi}{\pi^2}.$$

On cherche maintenant n pour que la probabilité de l'événement $\{\,|\mathrm{X}/n - p| > \delta p\}$ soit inférieure à $1/\varepsilon^2$. Il suffit donc — c'est une condition suffisante, mais pas forcément nécessaire — de choisir n pour que $\varepsilon\sigma/n \leqslant \delta p$, c'est-à-dire

$$\frac{\varepsilon\sqrt{n}\sqrt{p(1-p)}}{n} \leqslant \frac{2\,\delta\pi}{\pi^2} \qquad \text{ou encore} \qquad n \geqslant \frac{\pi^4\varepsilon^2}{4\cdot\,\delta\pi^2}\,p(1-p).$$

Une application numérique donne une valeur suffisante de n égale à environ 560 millions ; il ne semble guère raisonnable de chercher à évaluer π ainsi.

Cette estimation n'est pas optimale : l'inégalité de Bienaymé-Tchebychev est, on le sait, généralement bien pessimiste. En tenant compte du fait que la distribu-

(5). En effet, pour une précision donnée, on ne peut pas exclure de tomber sur une série de résultats « peu probables » conduisant à une valeur aberrante de p et donc de π. On cherche donc à limiter l'éventualité d'un manque de chance caractérisé.

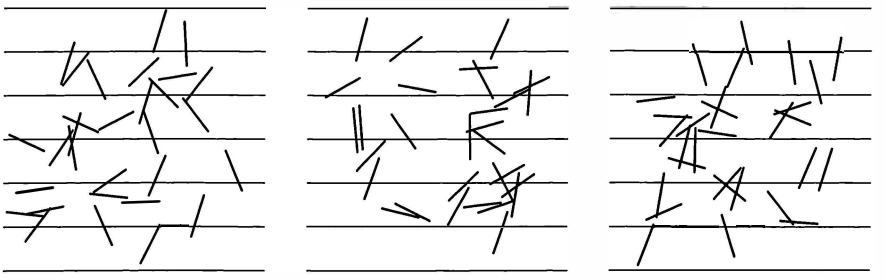


FIGURE 26.2 — *Sur ces simulations informatiques, respectivement 19, 19 et 18 aiguilles sur 30 coupent la rainure. Les estimations de p sont respectivement $p_1 = p_2 = 19/30$ et $p_3 = 3/5$. Les valeurs expérimentales de π sont donc $\pi_1 = \pi_2 \approx 3,1579$ et $\pi_3 \approx 3,3333$.*

tion du nombre d'aiguilles coupant la rainure suit une loi binomiale $\mathscr{B}(n\,;\,p)$, qui peut être approchée par une gaussienne pour les valeurs de n telles que $np(1-p) \gg 1$ (cf. § 325), on arrive à un résultat un peu plus fin et plus sym-
▲26.1 pathique pour l'expérimentateur (exercice 26.1) : une valeur $n \approx 37$ millions de lancers est en réalité suffisante. On économise donc plus de 500 millions de lancers, sans que la méthode devienne raisonnable pour autant.

26.2 INTÉGRATION NUMÉRIQUE

§ 383 **Position du problème**

Soit $\varphi : [\,a\,;\,b\,] \to \mathbb{R}$ une fonction définie sur un segment, et intégrable. On cherche à évaluer numériquement l'intégrale $\mathrm{I} = \int_a^b \varphi$. On a, à notre disposition, tout un panel de méthodes numériques déterministes. Parmi celles-ci, les méthodes dites de quadrature sont simple à programmer : on choisit un entier n, la méthode choisie prescrit des points $\alpha_1, \alpha_2, \ldots, \alpha_n$ dans $[\,a\,;\,b\,]$ et des poids $w_1, w_2, \ldots, w_n$, tels que I soit approchée par la somme $\sum w_k\,\varphi(\alpha_k)$, et ce d'autant mieux que n est grand. Cette méthode se généralise à des intégrales d'une fonction de plusieurs variables. Le nombre de point nécessaires à une approximation satisfaisante dépend de la méthode, de la régularité et bien entendu de la précision souhaitée.

Il existe également des méthodes probabilistes, basées sur la loi des grands nombres. Dans le cas d'une fonction d'une seule variable $\varphi : [\,a\,;\,b\,] \to \mathbb{R}$, on peut distinguer deux points de vue, de prime abord différents.

Dans le premier, on remarque que, si X est une variable aléatoire uniforme sur $[\,a\,;\,b\,]$, alors, par le théorème de transfert, $\varphi(\mathrm{X})$ admet une espérance et

$$\mathbf{E}\big(\varphi(\mathrm{X})\big) = \int_a^b \varphi(x)\,f(x)\,\mathrm{d}x = \frac{1}{b-a}\int_a^b \varphi(x)\,\mathrm{d}x,$$

puisque la densité f de X est constante et égale à $1/(b-a)$ sur $[\,a\,;\,b\,]$, nulle en dehors, ce qui donne

$$\int_a^b \varphi = (b-a)\,\mathbf{E}\big(\varphi(\mathrm{X})\big).$$

La loi des grands nombres permet enfin d'obtenir une estimation de $\mathbf{E}\big(\varphi(\mathrm{X})\big)$ par

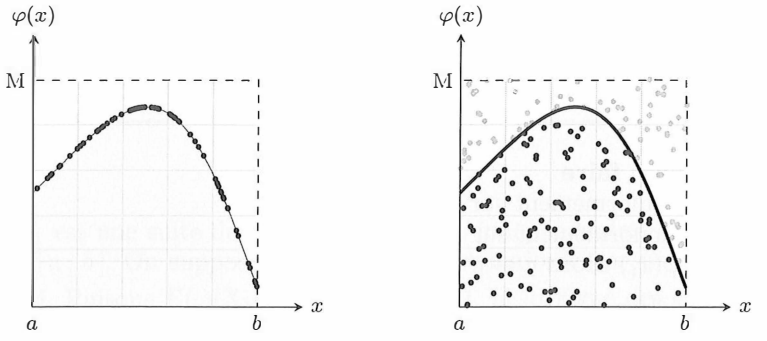


FIGURE 26.3 — À gauche : *méthode de Monte-Carlo simple.* À droite : *méthode de Monte-Carlo avec rejet. Les points gris sont rejetés, les points noirs sont comptabilisés.*

l'évaluation d'une moyenne empirique :

$$\mathbf{E}\big(\varphi(\mathrm{X})\big) \approx \frac{1}{n}\sum \varphi(\mathrm{X}_i)$$

si $(\mathrm{X}_n)_{n\geqslant 1}$ est une suite de variables aléatoires indépendantes, uniformément distribuées sur $[\,a\,;\,b\,]$. Une méthode pour estimer I est de calculer des moyennes empiriques : *en pratique*, quelle que soit la réalisation aléatoire du tirage

$$\big(\mathrm{X}_1(\omega), \mathrm{X}_2(\omega), \ldots, \mathrm{X}_n(\omega), \ldots\big)$$

— obtenue informatiquement ou autre — la suite des moyennes

$$\frac{1}{n}\sum_{k=1}^{n} \varphi\big(\mathrm{X}_k(\omega)\big)$$

converge vers $\mathbf{E}\big(\varphi(\mathrm{X})\big)$ quand n tend vers l'infini ; d'où l'algorithme :

Algorithme 14 : Monte-Carlo simple pour le calcul de $\int_a^b \varphi$

$\mathrm{S} \leftarrow 0$
pour $i = 1$ **à** n **faire**
  tirer x_i selon la loi $\mathscr{U}\big([\,a\,;\,b\,]\big)$
  $\mathrm{S} \leftarrow \mathrm{S} + \varphi(x_i)$
retourner $(b-a)\mathrm{S}/n$

Dans le second point de vue, on suppose que φ est bornée ; disons, pour fixer les idées, que φ est positive et majorée par un réel M. L'intégrale de f est égale à l'aire de la surface sous la courbe de f, c'est-à-dire que

$$\mathrm{I} = \iint_{[\,a\,;\,b\,]\times[\,0\,;\,\mathrm{M}\,]} \mathbf{1}_{\mathrm{A}}(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{26.1}$$

où l'on a noté

$$\mathrm{A} = \big\{(x,y) \in \mathbb{R}^2\,;\ a \leqslant x \leqslant b,\quad 0 \leqslant y \leqslant \mathrm{M},\quad y \leqslant \varphi(x)\big\}$$

(voir figure 26.3). Le rapport entre I et la surface du rectangle $(b-a)\times \mathrm{M}$ est la

probabilité qu'un point tiré au hasard dans le rectangle tombe sous la surface. On obtient donc une évaluation de I selon l'algorithme suivant :

Algorithme 15 : Méthode du rejet pour $\int_a^b \varphi$

$c \leftarrow 0$
pour $i = 1$ **à** n **faire**
  tirer x_i uniformément sur $[a\,;\,b]$
  tirer y_i uniformément sur $[0\,;\,\mathrm{M}]$
  si $y_i \leqslant \varphi(x_i)$ **alors** $c \leftarrow c + 1$
retourner $\dfrac{c(b-a)\mathrm{M}}{n}$

L'intérêt de cette méthode, dite **méthode du rejet**, est double :
— il n'est pas toujours nécessaire de calculer exactement $\varphi(x_i)$ pour savoir s'il faut ou non rejeter le couple (x_i, y_i) : une majoration ou une minoration un peu grossière peut suffire pour conclure pour nombre de points ;
— si φ est la réciproque d'une fonction monotone ($\varphi = \psi^{-1}$), que l'on sait calculer ψ mais pas φ, on remplace le test $y_i \leqslant \varphi(x_i)$ par le test équivalent $x_i \leqslant \psi(y_i)$ si ψ est décroissante, ou $x_i \geqslant \psi(y_i)$ si ψ est croissante.

On peut généraliser la méthode de Monte-Carlo simple au cas d'une fonction vectorielle $\varphi : \mathbb{R}^d \to \mathbb{R}$: l'intégrale de φ sur un pavé $\mathrm{K} = \prod_k [a_k\,;\,b_k]$, de volume $\mathrm{V} = \prod_k (b_k - a_k)$, vaut

$$\int_{\mathrm{K}} \varphi = \mathrm{V} \cdot \mathbf{E}\big(\varphi(\mathbf{X})\big)$$

où $\mathbf{X}$ suit la loi uniforme sur K[6].

Enfin, on peut combiner les deux approches pour évaluer une intégrale sur un domaine K de $\mathbb{R}^d$ qui n'est pas un pavé et est difficilement paramétrable, mais donné par une équation implicite de la forme

$$\mathrm{K} \ : \quad \mathrm{F}(x_1, x_2, \ldots, x_d) \leqslant \alpha.$$

On suppose pour cela que l'on connaît un pavé $\mathrm{Q} = \prod_{k=1}^d [a_k\,;\,b_k]$ contenant K.

Algorithme 16 : Évaluation d'une intégrale multiple sur un domaine K

$c \leftarrow 0$
$\mathrm{S} \leftarrow 0$
répéter
  tirer $(x_1, x_2, \ldots, x_d)$ uniformément dans Q
  si $\mathrm{F}(x_1, x_2, \ldots, x_d) < \alpha$ **alors**
    $\mathrm{S} \leftarrow \mathrm{S} + \varphi(x_1, x_2, \ldots, x_d)$
    $c \leftarrow c + 1$
jusqu'à $c = n$
retourner SV/n

△26.2

(6). La méthode de rejet peut alors être vue comme un cas particulier de la méthode simple grâce à la relation (26.1).

Convergence de la méthode § 384

Concentrons-nous sur la première méthode exposée : φ est une fonction de $[a\,;b]$ dans $\mathbb{R}$, une estimation de $\int_a^b \varphi$ est

$$\widetilde{\mathrm{I}}_n = \frac{b-a}{n}\sum_{k=1}^{n}\varphi(\mathrm{X}_k),$$

où $(\mathrm{X}_k)_{k\geqslant 1}$ est une suite de variables aléatoires indépendantes, de même loi uniforme sur $[a\,;b]$. On suppose de plus que φ est majorée par une constante M : $0 \leqslant \varphi \leqslant \mathrm{M}$. Puisque $\mathbf{E}\big(\varphi(\mathrm{X}_k)\big) = \frac{1}{b-a}\int_a^b \varphi(t)\,\mathrm{d}t$, la loi forte des grands nombres nous apprend que : avec une probabilité 1, la suite de terme général $\widetilde{\mathrm{I}}_n$ converge vers I. Autrement dit, pour la valeur de ω correspondant à la réalisation de l'expérience numérique, la suite des estimations va effectivement converger vers I[7].

Quant à la vitesse de convergence, elle nous est donnée, sous réserve que φ^2 soit intégrable, par le théorème limite central. La variable aléatoire X_k suit la loi uniforme sur $[a\,;b]$, l'espérance et la variance de $\varphi(\mathrm{X}_k)$ vérifient

$$\mathbf{E}\big(\varphi(\mathrm{X}_k)\big) = \frac{1}{b-a}\int_a^b \varphi$$

$$\sigma^2 := \mathbf{V}\left(\varphi(\mathrm{X}_k)\right) \leqslant \frac{1}{b-a}\int_a^b \varphi^2 \leqslant \mathrm{M}^2.$$

Ainsi, pour tout $\varepsilon > 0$, on a

$$\mathbf{P}\left\{\left|\widetilde{\mathrm{I}}_n - \mathrm{I}\right| > \varepsilon\,(b-a)\frac{\sigma}{\sqrt{n}}\right\} \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}}\int_{|x|>\varepsilon} \mathrm{e}^{-x^2/2}\,\mathrm{d}x = 2\big[1-\mathfrak{N}(\varepsilon)\big].$$

Si l'on accepte un risque $2\big[1-\mathfrak{N}(\varepsilon)\big]$ de 5%, on peut donc prendre $\varepsilon \approx 1{,}96$; pour un risque de 1%, on prend $\varepsilon \approx 2{,}58$. Ainsi, au risque choisi et pour n suffisamment grand pour que l'approximation normale soit justifiée, on aura la majoration

$$\left|\widetilde{\mathrm{I}}_n - \mathrm{I}\right| \leqslant \frac{\varepsilon\,\mathrm{M}\cdot(b-a)}{\sqrt{n}}.$$

Une estimation de M permet d'en déduire une valeur de n nous assurant la précision voulue sur le calcul de I (toujours avec le risque accepté).

Effet de la dimension § 385

Le point remarquable dans cette méthode est la vitesse de convergence, en $1/\sqrt{n}$. Est-ce rapide ? Malheureusement, non ! Si φ est continue, le calcul approché de I par une simple méthode des rectangles sera plus performante, l'erreur E_n obtenue pour une subdivision de pas $(b-a)/n$ étant majorée (avec certitude) par $\mathrm{C}^{\underline{\text{te}}}/n$. Si f est plus régulière, d'autres méthodes d'intégration numériques (méthode des trapèzes en $1/n^2$, de Simpson en $1/n^4$, quadratures d'ordre plus élevée...) mènent à des convergences encore plus rapides. La méthode de Monte-Carlo ne semble alors intéressante que pour des fonctions « sans régularité ».

En fait, c'est plutôt dans les dimensions supérieures qu'il faut rechercher l'intérêt d'une méthode stochastique. Supposons que la fonction φ soit définie sur un pavé K de $\mathbb{R}^d$, et que l'on en évalue l'intégrale par la méthode des trapèzes, dont l'erreur E_n en une dimension est de l'ordre de $1/n^2$ lorsque l'on coupe l'intervalle

(7). Sauf, encore une fois, coup de malchance absolu.

en n parties. En dimension d, si le pavé est découpé en N parties selon chaque direction, le nombre de points en lesquels on doit évaluer f est $n = \mathrm{N}^d$; c'est ce nombre n qui détermine le temps de calcul. L'erreur de la méthode n'est donc plus que de

$$\mathrm{E}_n^{\text{trap.}} = \mathrm{O}\left(\frac{1}{\mathrm{N}^2}\right) = \mathrm{O}\left(\frac{1}{n^{2/d}}\right),$$

et la méthode de Monte-Carlo, pour laquelle on a toujours

$$\mathrm{E}_n^{\text{MC}} = \mathrm{O}\left(\frac{1}{n^{1/2}}\right),$$

redevient performante dès que l'on dépasse la dimension 4 ; or, il n'est pas rare de devoir intégrer dans des espaces de dimensions très grandes, pour lesquelles les méthodes classiques d'intégration sont totalement inutilisables[8].

Nous n'en dirons pas davantage ici sur les méthodes de Monte-Carlo ; mentionnons simplement l'existence de quelques raffinements :

— on peut réduire le coefficient devant $1/\sqrt{n}$, proportionnel à la variance, en subdivisant le domaine d'intégration et en mettant en œuvre la méthode de Monte-Carlo pour chacun des petits modèles, ce qui a pour effet de réduire un peu la variance[9] ;

— on peut réduire ce même coefficient en effectuant un changement de variable dans I qui « aplatisse » la fonction φ et réduise donc sa variance (que l'on avait jusque là simplement majorée par M^2, ce qui est évidemment très grossier).

De nombreuses méthodes et raffinement ont été et continuent à être étudiés, et la littérature sur le sujet est très vaste.

26.3 LAISSONS FAIRE LE HASARD !

§ 386 **Un exemple arithmétique : tests de primalité**

Pour déterminer si un entier n est premier, la méthode naïve consiste à tester s'il est divisible par un quelconque des entiers compris entre 2 et $\sqrt{n}$. Cette méthode est trop lente pour être applicable lorsque n est trop grand ; par exemple, si n est de l'ordre de 10^{100}, il faudrait effectuer 10^{50} tests de divisibilité, ce qui est hors de question[10]. Des techniques à base de crible permettent d'accélérer le processus.

D'autres tests existent ; parmi ceux-ci, portons une attention particulière au symbole de Jacobi, qui permet un traitement rapide, mais probabiliste, du problème[11].

Le **symbole de Jacobi** de deux entiers $a, n \in \mathbb{N}$, n impair, noté $\left(\frac{a}{n}\right)$, est un élément de $\{-1, 0, 1\}$ qui est défini par des propriétés arithmétiques que l'on peut ignorer : la seule chose qui nous importe, c'est qu'il est possible de le calculer

(8). Comme déjà mentionné, lorsque de plus le domaine d'intégration est peu régulier, les estimations d'erreurs par les méthodes habituelles ne sont plus valables et la convergence des méthodes déterministes est moins bonne que prévue.

(9). Ce n'est pas anecdotique : vu la lenteur de convergence, réduire l'écart-type d'un facteur 2 permet de diviser le temps de calcul par 4.

(10). À raison d'un million de tests par seconde, le temps nécessaire est encore astronomiquement plus grand que l'âge actuel de l'univers.

(11). Ce n'est pas la seule méthode permettant ce traitement probabiliste. Citons par exemple le test de Miller-Rabin, qui est de même nature que celui que nous présentons ici, et qui ne fait intervenir que des calculs congruentiels ; voir la remarque 26.4 page 543.

relativement rapidement, *via* un algorithme récursif (et déterministe) simple[12]. Son intérêt est le suivant. Si p est un entier premier, alors pour tout $a \in \{1, 2, \ldots, p-1\}$, la condition suivante est réalisée :

$$\begin{cases} a \wedge p = 1 \\ \left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \mod n. \end{cases} \tag{26.2}$$

On dira que a **passe le test de Jacobi** s'il vérifie cette double condition[13]. *Si p est premier, tous les entiers $a \in \{1, 2, \ldots, p-1\}$ passent le test de Jacobi.* On en déduit un **algorithme probabiliste** portant sur la non-primalité d'un nombre :

Algorithme 17 : Test de Jacobi

tirer a uniformément dans $\{1, 2, \ldots, n-1\}$
si *a ne vérifie pas* (26.2) **alors**
  retourner *« n n'est pas premier »*
sinon
  retourner *« il est* possible *que n soit premier »*

Il est crucial de noter que le test de Jacobi, s'il permet de conclure avec certitude que n n'est pas premier, n'offre jamais la certitude qu'il l'est : il ne permet donc pas de fonder un test positif de primalité. Le point clé du test de Solovay-Strassen[14] est que, sous l'hypothèse que n n'est pas un nombre premier, l'ensemble des nombres de

$$\Omega_n := \{1, 2, \ldots, n-1\}$$

qui passent le test de Jacobi forme un sous-groupe propre du groupe $(\mathbb{Z}/n\mathbb{Z})^*$ des inversibles de $(\mathbb{Z}/n\mathbb{Z})$; il est donc de cardinal *au plus la moitié* de celui de Ω_n[15]. En d'autres termes, si l'on choisit un nombre au hasard dans Ω_n (muni de la probabilité uniforme), la probabilité qu'il réussisse le test est au plus égale à $1/2$. Si l'on choisit, indépendamment et au hasard, deux nombres $a_1, a_2 \in \Omega_n$, alors la probabilité qu'ils réussissent le test est au plus $1/4$. Plus généralement, si l'on choisit $a_1, \ldots, a_k$, indépendamment et au hasard dans $\{1, 2, \ldots, n-1\}$, ils ne réussiront tous le test qu'avec une probabilité inférieure à $1/2^k$. En notant T_i l'événement : « le i-ième

(12). Le lecteur pourra les trouver dans tous les bons livres d'arithmétique, par exemple Hindry [49]. Certains logiciels calculent le symbole de Jacobi explicitement et en un temps très court. En voici la définition brute.

Si p est un entier premier, on définit d'abord le symbole de Legendre de a et p par

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{si } a = 0 \mod p, \\ +1 & \text{si } a \text{ est un carré modulo } p, \\ -1 & \text{sinon.} \end{cases}$$

Si n est un entier impair, décomposé en nombres premiers sous la forme $n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_k^{\alpha_k}$, alors le symbole de Jacobi de a et n est

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right)^{\alpha_1} \cdot \left(\frac{a}{p_2}\right)^{\alpha_2} \cdots \left(\frac{a}{p_k}\right)^{\alpha_k}.$$

(13). La condition $a \wedge p = 1$ est équivalente à $\left(\frac{a}{p}\right) \neq 0$; elle se vérifie donc par le même algorithme simple et rapide ou par la méthode d'Euler.

(14). Mis au point par Robert Solovay (1938–), mathématicien américain, et Volker Strassen (1936–), mathématicien allemand. Ils ont publié conjointement, en 1977, un article intitulé *A fast Monte-Carlo test for primality* [92].

(15). Voir Serre [90].

entier tiré au hasard passe le test », on a donc

$$\mathbf{P}\{\mathrm{T}_1 \cap \mathrm{T}_2 \cap \cdots \cap \mathrm{T}_k \,|\, n \notin \mathbb{P}\} \leqslant 2^{-k}.$$

Le taux d'erreur de l'algorithme devient donc rapidement très petit, et c'est ce qui fait son intérêt.

Bien sûr, on aura noté que la probabilité évaluée... n'est pas celle que l'on voudrait. On connaît une majoration de la probabilité que k tests réussissent sachant que n n'est pas premier, mais ce qui nous intéresserait davantage, ce serait de connaître la probabilité que n soit (ou ne soit pas) premier, sachant que les k tests ont été réussis. On reconnaît là une situation typique d'inversion de causalité, et l'on dégaine le théorème de Bayes.

Mais avant cela, modélisons proprement l'expérience. On choisit un entier N « grand », on note Ω l'ensemble des entiers impairs entre 3 et N, que l'on munit de la probabilité uniforme ; on tire ensuite au hasard un entier n dans Ω[16]. C'est cet entier n, qui est le résultat d'une expérience et dépend de l'aléa ω, dont on va évaluer la probabilité qu'il soit premier. On tire donc des entiers $a_1, \ldots, a_k$ dans $\Omega_n = \{1, 2, \ldots, n-1\}$ et on leur fait passer le test de Jacobi.

Notons $\mathrm{A}_k = \mathrm{T}_1 \cap \mathrm{T}_2 \cap \cdots \cap \mathrm{T}_k$ l'événement « $a_1, a_2, \ldots, a_k$ passent le test ». La formule de Bayes pour le système complet d'événements $\big(\{n \in \mathbb{P}\}, \{n \notin \mathbb{P}\}\big)$, mène à

$$\mathbf{P}\{n \in \mathbb{P} \,|\, \mathrm{A}_k\} = \frac{\mathbf{P}\{\mathrm{A}_k \,|\, n \in \mathbb{P}\} \cdot \mathbf{P}\{n \in \mathbb{P}\}}{\mathbf{P}\{\mathrm{A}_k \,|\, n \in \mathbb{P}\} \cdot \mathbf{P}\{n \in \mathbb{P}\} + \mathbf{P}\{\mathrm{A}_k \,|\, n \notin \mathbb{P}\} \cdot \mathbf{P}\{n \notin \mathbb{P}\}}.$$

Si n est premier, tous les entiers de Ω_n passent le test, donc $\mathbf{P}\{\mathrm{A}_k \,|\, n \in \mathbb{P}\} = 1$. La majoration $\mathbf{P}\{\mathrm{A}_k \,|\, n \notin \mathbb{P}\} \leqslant 1/2^k$ conduit à

$$\mathbf{P}\{n \in \mathbb{P} \,|\, \mathrm{A}_k\} \geqslant \frac{\mathbf{P}\{n \in \mathbb{P}\}}{\mathbf{P}\{n \in \mathbb{P}\} + 2^{-k}\,\mathbf{P}\{n \notin \mathbb{P}\}}.$$

Il reste à évaluer $\mathbf{P}\{n \in \mathbb{P}\}$, qui est de l'ordre de $1/\ln \mathrm{N}$ d'après le théorème des nombres premiers, et même $2/\ln \mathrm{N}$ puisque l'on a éliminé *de facto* les entiers pairs. On obtient alors la formule approchée

$$\mathbf{P}\{n \in \mathbb{P} \,|\, \mathrm{A}_k\} \geqslant 1 - 2^{-(k+1)} \ln \mathrm{N}.$$

Algorithme 18 : Test de Solovay-Strassen sur $n \in \{3, 5, \ldots, \mathrm{N}\}$, au risque α

déterminer un entier k tel que $2^{-(k+1)} \ln \mathrm{N} \leqslant \alpha$
$i \leftarrow 1 \quad t \leftarrow 1$
tant que $i \leqslant k$ *et* $t = 1$ **faire**
  tirer a_i uniformément dans $\{1, 3, 5, \ldots, \mathrm{N}\}$
  faire passer le test de Jacobi à a_i
  si *réussi* **alors** $i \leftarrow i + 1$
  sinon $t \leftarrow 0$
si $t = 0$ **alors**
  retourner *« n n'est pas premier »*
sinon
  retourner *« n est probablement premier, au risque α »*

(16). On peut améliorer la méthode en éliminant d'office tous les multiples de 3 ou de 5, ce qui est simple lorsque n est choisi en tirant les chiffres de son écriture décimale au hasard. Mais cela ne change rien au principe général de la méthode

Exemple 26.1 Prenons un nombre n impair à 100 chiffres, tiré aléatoirement avec la probabilité uniforme dans $\{3, 5, \ldots, 10^{100}+1\}$. Alors la probabilité que n soit premier est

$$\mathbf{P}\{n \in \mathbb{P}\} \approx \frac{2}{\ln(10^{100})} \approx 8,7 \cdot 10^{-3}.$$

Imaginons que l'on fasse le test de Jacobi sur une vingtaine de nombres entiers tirés au hasard entre 1 et $n-1$, et que ces vingt tests soient positifs (événement A_{20}). Alors

$$\mathbf{P}\{n \in \mathbb{P} \mid A_{20}\} \geqslant 0,9998902... \approx 1 - 1,1 \cdot 10^{-4}.$$

En d'autres termes, on peut estimer qu'il n'y a qu'une chance sur 10 000 que n ne soit pas un nombre premier ! Si le même test de Jacobi est réussi une quarantaine de fois, alors

$$\mathbf{P}\{n \in \mathbb{P} \mid A_{40}\} \geqslant 1 - 1,04 \cdot 10^{-10}.$$

Ainsi, un petit nombre de tests est suffisant pour obtenir une quasi-certitude de la primalité de n.

Remarque 26.2 (L'approche bayésienne et la vraie vie) La situation est un peu différente si l'on se donne un entier n, et que l'on souhaite déterminer s'il est premier ou non ; en effet, cette fois-ci, n est constant, et il est plus difficile de donner un sens à $\mathbf{P}\{n \in \mathbb{P}\}$. Prenons par exemple $n = 146735287$: ce nombre est premier ou il ne l'est pas, et je ne peux pas recommencer l'expérience et évaluer la proportion du nombre de fois où il est premier ! L'approche « bayésienne » du problème[17] consiste à faire comme si l'on était dans la situation précédente, ce qui revient finalement à ne considérer les probabilités que comme une détermination numérique de la *vraisemblance* d'une propriété. On prend $N = n$, on interprète $\mathbf{P}\{n \in \mathbb{P}\}$ comme une probabilité *a priori* que l'on postule donc comme égale à $2/\ln n$. La formule de Bayes permet alors de calculer une probabilité *a posteriori*, qui est évidemment la même que la précédente, toujours en identifiant n et N.

Les calculs sont les mêmes, c'est l'interprétation du résultat qui est différente ; si l'on refuse le point de vue bayésien, l'expérience aléatoire qui nous permet de parler de probabilité est celle consistant à tirer effectivement n au hasard dans $\{3, 5, \ldots, N\}$ puis effectuer les tests. Cette expérience, et celle-là seulement, se prête à une vérification expérimentale en termes de fréquences.

Remarque 26.3 (Trouver des grands nombres premiers) Munis de cette méthode rapide de test de primalité, nous pouvons également donner une méthode probabiliste de recherche de grands nombres premiers. Si l'on choisit un entier N grand, on tire au hasard un nombre n impair[18] entre 3 et N et on teste s'il est premier. Si ce n'est pas le cas, on en tire un second, etc. Puisque la fraction des nombres premiers dans $\{3, 5, \ldots, N\}$ est de l'ordre de $2/\ln N$, il ne faudra, en moyenne, que $\frac{1}{2}\ln N$ tirages pour obtenir un entier premier.

Remarque 26.4 (Test de Miller-Rabin) Le test de Miller-Rabin est encore plus efficace que celui de Solovay-Strassen ; en effet, la probabilité qu'un nombre $a \in \{1, 2, \ldots, n-1\}$ passe le test lorsque n n'est pas premier est inférieure à $1/4$ (si n est premier, tous les entiers entre 1 et n le passent). Ainsi, il suffit d'un nombre de tests moitié moindre qu'avec Solovay-Strassen pour s'assurer du résultat avec la même probabilité.

Le lecteur intéressé peut facilement programmer ce test. La première chose à faire est d'extraire de $n-1$ toutes les puissances de 2, c'est-à-dire l'écrire sous la forme $n - 1 = 2^m \cdot q$ où q est impair. Si a est un entier compris entre 1 et $n-1$, il suffit de tester si les deux conditions suivantes sont réalisées :

$$a^q \not\equiv 1 \mod n \qquad \text{et} \qquad a^{2^i q} \not\equiv -1 \mod n \qquad \text{pour } i = 0, 1, \ldots, m-1.$$

Si c'est le cas, alors $n \notin \mathbb{P}$. Sinon, n est peut-être premier.

Comme dans la méthode de Solovay-Strassen, on effectue ce test pour un certain nombre d'entiers $a_1, \ldots, a_k$ pris au hasard, et en nombre suffisant pour assurer que n est « très probablement » premier. La seule difficulté de programmation est le calcul de grands nombres comme $a^{2^i q} \mod n$, que l'on effectuera par exponentiation rapide modulo n.

Le lecteur peut consulter avec profit l'ouvrage de Demazure [23, chapitre II, § 2.3.7].

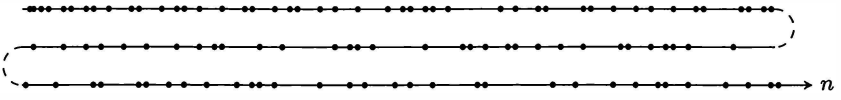
---

(17). Voir chapitre 32.

(18). Comme précédemment, on peut affiner un peu en tirant aléatoirement chacun des d chiffres qui le composent lorsque $N = 10^{d+1} - 1$; on impose que le dernier chiffre soit dans $\{1, 3, 7, 9\}$, et qu'il soit tel que la somme des chiffres ne soit pas un multiple de 3 ; ainsi, n ne sera multiple ni de 3 ni de 5.

§387 **Une machine à deviner des conjectures : le modèle de Cramér**

La répartition des nombres premiers dans l'ensemble $\mathbb{N}$ présente un aspect quelque peu « aléatoire », notamment à petit échelle[19]. À plus grande échelle, au contraire, la répartition est plus harmonieuse — sur un intervalle centré sur n, la fréquence des nombres premiers est asymptotiquement de l'ordre de $1/\ln n$, selon le théorème des nombres premiers.



Cette double constatation (irrégularité à petite échelle, régularité à grande échelle) faisant irrésistiblement penser à une répartition aléatoire, le mathématicien Harald CRAMÉR (1893–1985) a proposé un *modèle stochastique de répartition des nombres premiers*. Le principe est le suivant : on considère une suite $(X_n)_{n\geqslant 2}$ de variables aléatoires *indépendantes* suivant chacune une loi de Bernoulli de paramètre $p_2 = 1$ et $p_n = 1/\ln n$ si $n \geqslant 2$. On note enfin

$$\mathcal{P}(\omega) = \{n \geqslant 2\ ;\ X_n(\omega) = 1\}.$$

L'ensemble aléatoire $\mathcal{P}$ possède presque sûrement la même propriété de densité asymptotique en $1/\ln n$ que le véritable ensemble $\mathbb{P}$ des nombres premiers.

Mais ce qui est fascinant, c'est que les nombres premiers ont tendance à vérifier effectivement un grand nombre de propriétés statistiques prévues par le modèle de Cramér : lorsqu'une certaine propriété sur la distribution de $\mathcal{P}$ est vérifiée presque sûrement, on peut espérer que l'ensemble $\mathbb{P}$ — qui n'est après tout qu'une des réalisations $\mathcal{P}(\omega)$ possibles — la vérifie aussi. L'idée de Cramér est donc de considérer que $\mathbb{P}$ est un *ensemble typique* du point de vue de la répartition des entiers premiers ; l'ensemble aléatoire $\mathcal{P}$ est l'ensemble des « entiers premiers » du modèle.

Par exemple, le modèle de Cramér implique qu'il existe presque sûrement une infinité de « nombres premiers » jumeaux (de la forme n et $n+2$)[20], conjecture
↫26.3 qui avait déjà été formulée auparavant. Dans ce modèle, l'écart $p_{n+1} - p_n$ entre deux nombres premiers successifs, p_n étant le n-ième nombre premier, vérifie

$$p_{n+1} - p_n = O(\ln^2 p_n).$$

Cette propriété, dénommée *conjecture de Cramér*, est pour l'heure encore loin d'être démontrée pour le vrai ensemble $\mathbb{P}$ mais reste une conjecture crédible. Certaines propriétés suggérées par ce modèle ont été prouvées, d'autres *en partie* prouvées, comme l'apparition de la loi de Poisson dans la répartition des nombres premiers (voir section 28.3).

Enfin, pour terminer sur une note moins heureuse, on connaît des résultats qui contredisent les prédictions du modèle de Cramér [41].

---

(19). Modulo le fait, bien sûr, que les entiers pairs plus grands que 2 ne sont jamais premiers.

(20). Bien entendu, il implique aussi qu'il existe une infinité de « nombres premiers » de la forme n et $n+1$, ce qui est grossièrement faux pour les véritables nombres premiers. Comme le disent Tenenbaum et Mendes France [97],

> *les nombres premiers occupent tout l'espace (entendez : le hasard) disponible, [...] tout ce qui n'est pas trivialement interdit est réalisé.*

Le hasard et les algorithmes de recherche de solution optimale § 388

Citons rapidement quelques cas dans lesquels le hasard est utilisé au sein d'un algorithme.

- **Tri rapide.** Les méthodes de tri par comparaison les plus rapides ont une complexité en $O(n \ln n)$, où n est le nombre d'éléments à trier. C'est le cas par exemple du *tri fusion*, qui est un tri récursif. L'inconvénient de ce tri est de ne pouvoir être effectué directement sur le tableau d'origine et de requérir l'utilisation d'un second tableau (si vous avez un disque dur de 1 tera-octet et que vous voulez trier un tera-octet de données, vous n'avez plus qu'à foncer au magasin acheter un deuxième disque dur). Une autre méthode est le **tri rapide**, également récursif, mais qui a l'avantage d'être *en place* (on se contente de permuter des éléments du tableau et le premier disque suffit). Le principe simplifié est le suivant : pour trier une liste d'au moins 2 éléments, on choisit un élément de la liste (un **pivot**, qui pourra être par exemple le premier élément), puis on découpe la liste initiale en deux sous-listes, celle contenant les éléments inférieurs au pivot, et celle contenant les éléments supérieurs au pivot — cette opération se faisant en temps linéaire. Chacune des sous-listes est ensuite traitée récursivement selon la même méthode, jusqu'à obtenir des listes de taille unité. Dans la plupart des cas, cette méthode de tri est effectivement rapide mais, dans le pire des cas, nécessite un nombre d'opérations en $O(n^2)$ — par exemple si la liste est déjà triée — ce qui est dramatiquement plus lent que le tri fusion[21].

  On peut rendre la méthode effectivement rapide en choisissant le pivot de manière aléatoire dans la liste (*tri rapide randomisé*). Alors que dans la vraie vie, les cas pathologiques se révèlent assez fréquents, la « randomisation » nous assure qu'au contraire le temps de calcul sera *très probablement* de l'ordre de $n \ln n$. On peut également mélanger la liste de manière aléatoire, ce qui peut se faire en temps linéaire, avant d'appliquer l'algorithme de tri rapide déterministe.

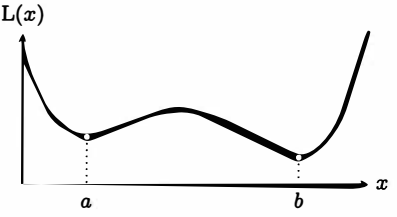
  Le tri rapide randomisé fait partie de la classe des algorithmes probabilistes de **Las Vegas** : alors que les algorithmes de Monte-Carlo terminent toujours mais ont une faible probabilité de retourner un résultat incorrect, ceux de Las Vegas ne retournent que des résutats exacts mais peuvent, avec une faible probabilité, renvoyer un message d'échec, ne pas terminer, ou bien terminer dans un temps plus long que celui espéré — par exemple en $O(n^2)$ au lieu de $O(n \ln n)$. Voir Cormen *et al.* [17].

- **Recuit simulé.** Des problèmes célèbres comme celui du *voyageur de commerce* — trouver le chemin le plus court permettant de relier n villes du plan — n'ont pas d'algorithme déterministe rapide connu[22]. On sait trouver des configurations presque optimales en partant d'un trajet quelconque puis en modifiant légèrement, et de manière aléatoire, une partie du trajet. Si la modification se traduit par un allongement du trajet, on l'oublie, sinon, on la conserve ; puis on recommence. Cette méthode présente un défaut majeur : il est possible que l'on se trouve à un moment donné dans une position de minimum local — point a sur la figure ci-après : toute modification d'une partie du trajet se traduit par un allongement. S'il existe

(21). Pourquoi faire tourner l'algorithme si la liste est déjà ordonnée, me direz-vous ? Tout simplement parce que l'utilisateur ne le sait pas forcément. Il est d'ailleurs fréquent, dans les situations réelles, que les données soient naturellement *presque* ordonnées.

(22). L'algorithme naïf consistant à calculer la longueur de tous les chemins suppose donc de comparer $n!$ nombres. Dès que n est de l'ordre de quelques dizaines, cela devient physiquement impossible.

un *minimum global* (point b), celui-ci est inaccessible.



Pour éviter de se retrouver piégé dans un tel minimum local, l'idée est d'ajouter un *bruit thermique*, c'est-à-dire autoriser les modifications se traduisant par des allongements, avec une probabilité proportionnelle à $e^{-\Delta\ell/T}$, où $\Delta\ell$ est l'allongement du trajet, assimilé à une énergie, et T une température fictive. Un trajet raccourci est donc choisi avec une probabilité plus importante qu'un trajet rallongé, et ce de façon d'autant plus marquée que la température est basse. Au début, on choisit une température élevée, ce qui autorise à modifier le trajet en prenant peu en compte l'allongement ou le raccourcissement. Au fur et à mesure que l'algorithme tourne, on diminue lentement la température, tout en continuant à tester des modifications aléatoires locales. Si l'on est dans une configuration de minimum local, on espère que le bruit thermique est suffisant pour en sortir ; si l'on est au fond d'une cuvette de grande profondeur, on y reste.
Quand la température devient très faible, on espère avoir atteint, avec une grande probabilité, le minimum global ou, du moins, un minimum local de valeur peu éloignée de celle du minimum global. Sous des conditions très générales, on peut montrer qu'une baisse très lente de la température permet d'aboutir presque sûrement au minimum global ; on préférera sans doute une baisse plus rapide, quitte à n'obtenir qu'une solution approchée. Cette méthode est calquée sur le procédé métallurgique consistant à refroidir lentement un métal afin de permettre à ses atomes de se ranger de manière la plus ordonnée possible.

- **Autres algorithmes.** Citons pour terminer deux classes d'algorithmes utilisant le hasard : les *algorithmes de fourmis*, et les *algorithmes de sélection naturelle*, tous deux inspirés de la biologie et permettant de résoudre — imparfaitement — des problèmes en un temps extrêmement réduit.

§ 389 **Falsifier des données comptables et se faire rattraper par la loi de Benford**
L'astronome NEWCOMB[23] avait remarqué que, au sein d'une liste $x_1, x_2, \ldots, x_n$ de nombres issus de catalogues de données numériques, la partie fractionnaire du logarithme décimal[24] de ces données

$$\{\operatorname{Log} x_k\} := \operatorname{Log} x_k - \lfloor \operatorname{Log} x_k \rfloor$$

suivait une loi uniforme ou, ce qui revient au même, que la probabilité que le

(23). Simon NEWCOMB. Note on the frequency of use of the different digits in natural numbers, *American Journal of Mathematics* **4**, 1881, p. 39–40. Simon Newcomb (1835–1909) était un mathématicien, physicien et astronome américain autodidacte.

(24). On notera « Log » le logarithme décimal : $\operatorname{Log} x = \frac{\ln x}{\ln 10}$, et $y = \operatorname{Log} x$ si et seulement si $x = 10^y$.

premier chiffre[25] de cette donnée soit k est

$$p_k \approx \mathrm{Log}\left(\frac{k+1}{k}\right).$$

Les valeurs numériques approchées de p_1 à p_9 sont listées ci-dessous :

k	1	2	3	4	5	6	7	8	9
p_k	0, 301	0, 176	0, 124	0, 097	0, 079	0, 067	0, 058	0, 049	0, 046

La découverte de Newcomb resta dans l'ombre jusqu'à la publication d'un nouvel article du physicien américain Frank BENFORD[26] qui refit la même observation : dans une collection de données numériques, nombreux étaient les nombres dont le premier chiffre était un 1, un peu moins nombreux ceux dont le premier chiffre était un 2,... et ce jusqu'à ceux dont le premier chiffre était un 9, en proportion très faible. C'est ce qui explique l'observation qui a mis la puce à l'oreille de Newcomb, à savoir que dans les tables de logarithmes, les premières pages sont davantage défraîchies que les dernières[27] :

> That the ten digits do not occur with equal frequency must be evident to any one making much use of logarithmic tables, and noticing how much faster the first pages wear out than the last ones. The first significant figure is oftener 1 than any other digit, and the frequency diminishes up to 9.

Remarque 26.5 Cette loi empirique semble être d'autant mieux vérifiée que les données proviennent d'expériences différentes : si vous regardez la longueur des fleuves du monde (dans l'unité de votre choix), la loi est à peu près vérifiée ; si vous y ajoutez les masses moyennes des principales espèces de vertébrés, l'adéquation n'en sera que meilleure. Frank Benford, pour son article, utilisa 20 229 données numériques provenant de vingt sources distinctes : hauteurs d'immeubles, statistiques de base-ball, surface de 335 rivières, numéros de rue des 342 premières personnes de l'annuaire *American Men of Science*, chaleur spécifique de milliers de composés chimiques etc. Certaines listes de données ne suivaient que très approximativement, ou pas du tout, la loi de Benford ; mais la réunion des listes la suivait remarquablement bien.

Puisque $p_1 = \mathrm{Log}(2) \approx 0, 30103$,

- *un nombre pris au hasard dans une collection de données numériques a plus de* 30 % *de chances de commencer par le chiffre* 1 *!*

De même, puisque $p_1 + p_2 = \mathrm{Log}(3) \approx 0, 477$ et $p_1 + p_2 + p_3 = \mathrm{Log}(4) \approx 0, 602$,

- *il y a presque* 48 % *de chances que le premier chiffre soit* 1 *ou* 2,
- *il y a plus de* 60 % *de chances que le premier chiffre soit* 1, 2 *ou* 3,
- *et par conséquent, moins de* 40 % *de chances qu'il soit égale à* 4, 5, 6, 7, 8 *ou* 9.

Quiconque n'a jamais rencontré la loi de Benford ne peut être qu'extrêmement étonné de ce résultat, en apparence paradoxal. L'esprit humain est ainsi fait qu'il s'attend à une loi *uniforme* chaque fois qu'a lieu un événement aléatoire pour lequel la loi est *inconnue* ; c'est un biais psychologique connu, et non un paradoxe mathématique.

---

(25). Le nombre 0, 00734 s'écrit $7, 34 \cdot 10^{-3}$; son « premier chiffre » est donc 7.

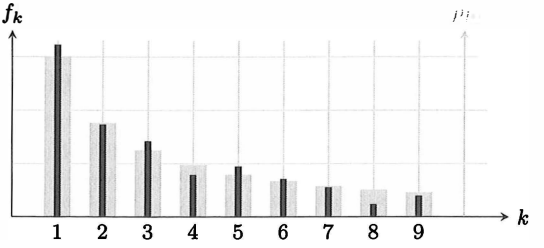
(26). Frank BENFORD. The law of anomalous numbers, *Proceedings of the American Philosophical Society*, **78**, 1938, p. 127–131. Frank BENFORD (1883–1948) était ingénieur et physicien.

(27). Le même phénomène a lieu avec, par exemple, dans un exemplaire de bibliothèque municipale de la *Critique de la raison pure*, bien que pour des raisons différentes.

Exemple 26.6 (Données astronomiques) Dans l'*Astronomie Générale* d'André Danjon[28], on trouve des tables de données numériques concernant les planètes et les principaux satellites du système solaire, notamment leur demi grand axe (en kilomètres), leur diamètre, l'excentricité de leur orbite (un nombre compris entre 0 et 1), et leur période de révolution sidérale (en jours). Collectons ces données et comptons le nombre d'occurrences N_k de chaque « premier chiffre » k de $\{1, 2, \ldots, 9\}$, calculons leurs fréquences f_k puis comparons-les aux probabilités $p_k = \mathrm{Log}(1 + 1/k)$ prévues par la loi de Benford :

k	1	2	3	4	5	6	7	8	9
N_k	41	22	18	10	12	9	7	3	5
f_k	0,323	0,173	0,142	0,079	0,095	0,071	0,055	0,024	0,039
p_k	0,301	0,176	0,124	0,097	0,079	0,067	0,058	0,049	0,046

Voici les histogrammes correspondants (en gris, les probabilités théoriques, en noir les fréquences observées) :



La loi de Benford a été amplement commentée, et bien sûr on ne peut pas *prouver* qu'elle est vérifiée — il existe des bases de données qui ne suivront pas cette loi, par exemple les numéros atomiques des éléments de la table de Mendeleiev ou même tout simplement la liste des entiers de 1 à n. Il existe également de nombreuses bases de données qui ne vérifient la loi de Benford que de façon approchée.

Un argument qui permet de comprendre pourquoi elle est *souvent* valide est le suivant : la loi de Benford est la seule qui soit invariante par tous les changements d'unité, c'est-à-dire par transformation $x_k \mapsto \alpha x_k$ pour tout $\alpha > 0$ [48]. Si des données géographiques sont bien susceptibles de suivre la loi de Benford — les longueurs des fleuves peuvent être exprimées en mètres, en pieds ou en furlongs — en revanche, l'argument n'est plus valide pour la table de Mendeleiev[29].

Ce qui est remarquable, au fond, dans la loi de Benford, n'est pas tant qu'elle soit vérifiée par de nombreuses listes de données[30], mais que l'esprit humain s'attende naturellement à autre chose (comme une loi uniforme[31]). Notamment, une liste de données « aléatoires » créée de toute pièce par un humain ne vérifie *pas* la loi de Benford. Cette particularité est utilisée pour dépister les fausses

(28). André Danjon, *Astronomie Générale*, Librairie Albert Blanchard, 1980.

(29). Pourtant, de manière étonnante, certaines suites purement numériques suivent également la loi de Benford : la suite de terme général $u_n = n!$, ou encore celle de terme général $v_n = n^n$ par exemple.

(30). Une explication naturelle du fait qu'elle est « à peu près vérifiée » sous des conditions très générales est donnée par Nicolas Gauvrit et Jean-Paul Delahaye dans un article d'une grande clarté [37].

(31). Cette loi uniforme existe bien, en ce qui concerne le premier chiffre du *logarithme* des données aléatoires. C'est ainsi que Newcomb a exposé son résultat. Il termine d'ailleurs son article ainsi :

> *It is curious to remark that this law would enable us to decide whether a large collection of independent numerical results were composed of natural numbers or logarithms.*

données, en particulier celles pouvant apparaître sur des livres de comptes truqués : un écart sensible à la loi de Benford incitera le fisc à s'intéresser de près à des transactions possiblement frauduleuses[32]. De même, des fraudes électorales ont pu être soupçonnées.

Remarque 26.7 (Loi du second chiffre) L'uniformité de la loi de la partie fractionnaire du logarithme n'implique pas seulement la loi de Benford pour le premier chiffre significatif X, mais également un effet sur le second chiffre Y, dont les fréquences varient entre 11,39 % pour le chiffre 1 et 8,5 % pour le chiffre 9 :

$$\mathbf{P}\{\mathrm{Y}=k\}=\sum_{i=1}^{9}\mathrm{Log}\left(1+\frac{1}{10\,i+k}\right).$$

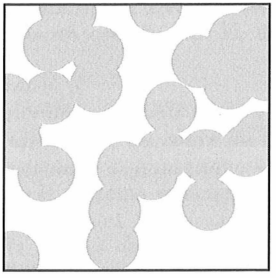
k	1	2	3	4	5	6	7	8	9
$\mathbf{P}\{\mathrm{Y}=k\}$	0, 114	0, 109	0, 104	0, 100	0, 097	0, 093	0, 090	0, 088	0, 085

Concernant le troisième chiffre, l'effet est quasiment négligeable.

EXERCICES

♦ ***Exercice 26.1*** Reprendre le calcul de l'aiguille de Buffon (§ 382), et donner une estimation utilisant une approximation gaussienne du nombre de lancers nécessaires pour calculer π à 10^{-3} près avec une probabilité de 99 %, puis de 95 %. Conclure sur l'efficacité de la méthode.

♦ ***Exercice 26.2 (Recouvrement de disques)*** Dans un rectangle de dimensions $a \times b$, on considère N disques de rayon R dont les centres sont distribués uniformément sur la surface du rectangle (ils peuvent donc notamment dépasser du cadre, se superposer,...). Déterminer, en fonction de n, la surface totale moyenne occupée par ces disques. On utilisera pour cela une méthode de Monte-Carlo (pour chaque configuration), et on effectuera la moyenne sur un nombre suffisant de distributions. On prendra $a = b = 5$, $\mathrm{R} = 1/2$, et on fera varier N entre 0 et 70.



♦ ***Exercice 26.3 (Entiers « premiers » jumeaux)*** Dans le modèle de Cramér (§ 387), montrer qu'il existe presque sûrement une infinité d'« entiers premiers » jumeaux, c'est-à-dire que $\mathcal{P}$ comporte une infinité d'entiers n tels que $n+2$ soit également dans $\mathcal{P}$.

♦ ***Exercice 26.4 (Modèle de Galton-Watson)*** On reprend le modèle de Galton et Watson d'étude de la probabilité d'extinction d'un nom de famille (voir paragraphe § 129). Ici, on supposera que chaque homme a, dans sa vie :

- aucun enfant avec une probabilité 1/8 ;
- un enfant avec une probabilité 3/8 ;
- deux enfants avec une probabilité 3/8 ;
- trois enfants avec probabilité 1/8.

Estimer numériquement la probabilité d'extinction d'un nom de famille porté par un unique homme.

Comparer au résultat théorique : la probabilité d'extinction est de $\sqrt{5}-2 \approx 0,236$.

(32). Il est malheureux de constater que l'on peut également détecter des « données » forgées de toutes pièces dans des articles de revues scientifiques.

SOLUTIONS

♦ **Solution 26.1** On approche la loi de X par une loi gaussienne $\mathcal{N}(np\,;\,np(1-p))$. On écrit alors

$$\mathbf{P}\left\{|\mathrm{X}-np|\geqslant\frac{2n\,\delta\pi}{\pi^2}\right\}=\mathbf{P}\left\{\left|\frac{\mathrm{X}-np}{\sqrt{n\,p(1-p)}}\right|\geqslant\frac{2\sqrt{n}\,\delta\pi}{\pi^2\sqrt{p(1-p)}}\right\}$$
$$=2-2\mathfrak{N}\left(\frac{2\sqrt{n}\,\delta\pi}{\pi^2\sqrt{p(1-p)}}\right)=0,01.$$

ce qui, à la lecture de la table de la loi normale page 720, donne

$$n\geqslant(2,58)^2\,\frac{\pi^4\,p(1-p)}{4\,\delta\pi^2}\approx 37,5\cdot 10^6,$$

c'est-à-dire environ 37 millions de lancers.

Pour une probabilité de 95 %, on obtient

$$n\geqslant(1,96)^2\,\frac{\pi^4\,p(1-p)}{4\,\delta\pi^2}\approx 21,6\cdot 10^6,$$

Soit 21 millions de lancers. On remarque que le nombre de lancers obtenu ne varie qu'assez peu avec le risque que l'on s'autorise (pour des risques petits), mais est proportionnel à l'inverse du carré de la précision demandée. Cette méthode est donc très mauvaise si l'on se fixe pour objectif d'obtenir une valeur *précise* de π .

♦ **Solution 26.3** Notons A_n l'événement « $4n+1$ et $4n+3$ appartiennent à $\mathcal{P}$ » (on les choisit impairs par pure raison esthétique). De l'indépendance des variables X_k, et puisque A_n et A_m dépendent de variables distinctes dès que $n\neq m$, on déduit que les événements A_n sont indépendants entre eux. Enfin,

$$\mathbf{P}(\mathrm{A}_n)=\mathbf{P}\{\mathrm{X}_{4n+1}=\mathrm{X}_{4n+3}=1\}=\frac{1}{\ln(4n+1)}\cdot\frac{1}{\ln(4n+3)}\underset{n\to\infty}{\sim}\frac{1}{\ln^2 n},$$

donc la série $\sum\mathbf{P}(\mathrm{A}_n)$ diverge et le lemme de Borel-Cantelli entraîne $\mathbf{P}\{\mathrm{A}_n\ ;\ \text{i.s.}\}=1$.

Conseils de lecture

Sur les tests de primalité probabiliste, une excellente référence est

- Donald KNUTH, *The Art of Computer Programming* [60], paragraphe 4.5.4.

Sur la loi de Benford, le lecteur pourra consulter encore Knuth ainsi qu'un article de Delahaye.

- Donald KNUTH, *The Art of Computer Programming* [60], paragraphe 4.2.4, et les exercices (parfois difficiles!) liant la loi de Benford à la loi de probabilité logarithmique sur les ensembles d'entiers.
- Jean-Paul DELAHAYE, *L'étonnante loi de Benford*, Pour la Science, janvier 2007, p. 90–95

Sur les algorithmes stochastiques, outre Knuth déjà cité, la grande référence est

- Thomas CORMEN, Charles LEISERSON, Ronald RIVEST, Clifford STEIN, *Introduction aux algorithmes* [17].

Chapitre 27

Chaînes de Markov

De œ très vaste sujet, nous n'aborderons ici qu'un exemple : celui des chaînes de Markov homogènes à nombre fini d'états. On peut voir ces chaînes comme des promenades aléatoires sur un graphe fini où, à chaque pas, le promeneur choisit un des chemins qui lui sont offerts avec une loi de probabilité fixée et ce, sans garder mémoire de son passé.

On étudie un système pouvant se trouver dans un nombre fini d'états, notés $\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_d$, *et qui évolue en temps discret repéré par un entier* n *parcourant* $\mathbb{N}$. *À l'instant* 0, *le système n'est pas forcément dans un état déterminé, mais décrit par une* **loi initiale** $\boldsymbol{\pi_0} = \big(\pi_0(1), \ldots, \pi_0(d)\big)$ *: le système se trouve dans l'état* E_i *avec la probabilité* $\pi_0(i)$. *Des questions se posent naturellement :*

- *Si le système est dans un certain état* E_i *à* $t = 0$, *quels autres états lui sont accessibles ?*
- *Reviendra-t-il à l'état* E_i *?*
- *De manière plus générale, comment décrire la probabilité qu'a le système de se trouver dans un état* E_i *au cours du temps ?*
- *Y a-t-il un « comportement asymptotique » quand* $n \to \infty$ *?*
- *Avec quelle fréquence le système repassera-t-il par chacun des états ?*

Après avoir étudié les propriétés des matrices de transition, on donnera deux types de propriétés des états : celles directement liées aux coefficients de la matrice de transition (périodicité, accessibilité) et celles découlant de propriétés asymptotiques (états récurrents, transitoires).

27.1 MATRICES DE TRANSITION

On se donne un système (physique, mathématique, biologique...) pouvant évoluer, à temps discret (indexé par n), et se trouver dans différents états E_i. On suppose de plus que ce système n'a aucune mémoire de son passé et ne connaît que son état présent à l'instant n pour évoluer jusqu'à l'instant $n+1$. Enfin, on suppose que cette évolution ne dépend pas du temps (ou encore : est invariante par translation dans le temps).

Exemple 27.1 Les systèmes admettant une telle description sont nombreux :

- étude d'une marche aléatoire (et, au-delà, d'un mouvement brownien) ;
- évolution au cours du temps (discrétisé) du nombre d'atomes dont le moment magnétique pointe vers le haut (modèle d'Ising et assimilés) ;
- étude d'une population ;

— dynamique d'un système quantique ;
— étude de jeux de hasard...

et la liste est loin d'être exhaustive : Andreï MARKOV étudiait la répartition statistique des lettres cyrilliques dans le roman en vers *Eugène Onéguine* d'Alexandre POUCHKINE et avait montré qu'il existait un lien probabiliste entre chaque lettre et la suivante.

Plus récemment, l'algorithme *PageRank* utilisé par le moteur de recherche *Google* est essentiellement basé sur les propriétés des chaînes de Markov [82, chapitre 9].

La théorie générale des chaînes de Markov a vu son champ d'application exploser dans les dernières décennies, notamment dans les algorithmes de type Monte-Carlo, permettant de traiter des quantités de données gigantesques, ou des algorithmes de Metropolis pour simuler des lois de probabilités.

Cette théorie générale est complexe, et nous nous limiterons à une présentation succincte de la théorie à nombre d'états finis et à temps discret. (Le cas général fera l'objet d'un ouvrage à paraître, au moins dans certaines réalisations ω de l'univers des possibles.)

§ 390 **Matrices de transition, chaînes de Markov**

On note $\mathscr{E} = \{\mathrm{E}_1, \mathrm{E}_2, \dots, \mathrm{E}_d\}$ un ensemble fini ; par commodité d'écriture, on le prendra le plus souvent sous la forme $\mathscr{E} = [\![1\,;d]\!]$. On considère une matrice $\mathbf{P} = \big(\mathrm{P}(i,j)\big)_{i,j\in\mathscr{E}}$, c'est-à-dire un tableau de nombres doublement indexé par $\mathscr{E}$.

Définition 27.2 (Matrice stochastique) On dit que $\mathbf{P}$ est une **matrice de transition**, ou encore **matrice stochastique**, si

- tous ses coefficients, appelés **probabilités de transition**, sont positifs ;
- la somme des coefficients sur chaque ligne est égale à 1 :

$$\forall i \in \mathscr{E} \qquad \sum_{j\in\mathscr{E}} \mathrm{P}(i,j) = 1.$$

Définition 27.3 (Chaîne de Markov stationnaire) Une suite $(\mathrm{X}_n)_{n\geqslant 0}$ de variables aléatoires est une **chaîne de Markov stationnaire** (ou **homogène**) **de matrice de transition** $\mathbf{P}$ et de **loi initiale** $\boldsymbol{\pi}_0 = \big(\pi_0(1), \dots, \pi_0(d)\big)$ si

- $\mathbf{P}\{\mathrm{X}_0 = i\} = \pi_0(i)$ pour $i = 1, \dots, d$;
- pour tout entier n et tout choix d'indices $i_0, i_1, \dots, i_n$,

$$\mathbf{P}\{\mathrm{X}_0 = i_0, \mathrm{X}_1 = i_1, \dots, \mathrm{X}_n = i_n\} = \pi_0(i_0)\,\mathrm{P}(i_0,i_1)\,\mathrm{P}(i_1,i_2)\cdots\mathrm{P}(i_{n-1},i_n). \quad (*)$$

Dans la suite, *toutes* les chaînes de Markov considérées seront supposées stationnaires, et on omettra donc de le préciser.

Définition 27.4 (Propriété de Markov) Soit $(\mathrm{X}_n)_{n\geqslant 0}$ une suite de variable aléatoire ; on dit qu'elle vérifie la **propriété de Markov** si, pour tout entier n et pour tout choix d'indices $i_0, i_1, \dots, i_{n+1}$ tels que $\mathbf{P}\{\mathrm{X}_0 = i_0, \dots, \mathrm{X}_n = i_n\} > 0$,

$$\mathbf{P}\{\mathrm{X}_{n+1} = i_{n+1} \,|\, \mathrm{X}_0 = i_0, \dots, \mathrm{X}_n = i_n\} = \mathbf{P}\{\mathrm{X}_{n+1} = i_{n+1} \,|\, \mathrm{X}_n = i_n\}. \quad (**)$$

Autrement dit, pour tout ce qui concerne l'évolution de la suite $(\mathrm{X}_n)_{n\geqslant 0}$, les probabilités conditionnées $\mathbf{P}_{\{\mathrm{X}_n = i_n\}}$ et $\mathbf{P}_{\{\mathrm{X}_0 = i_0, \dots, \mathrm{X}_n = i_n\}}$ sont égales.

Cela se traduit, de manière imagée, par : *l'évolution future de la chaîne ne dépend du passé qu'à travers l'instant présent.*

Un point essentiel est que les propriétés $(*)$ et $(**)$ sont intimement liées :

■ **Théorème 27.5 (Chaînes de Markov et propriété de Markov)** *Soit* $(X_n)_{n\geqslant 0}$ *une suite de variables aléatoires à valeurs dans* $\mathscr{E}$.

i) Si $(X_n)_{n\geqslant 0}$ *est une chaîne de Markov, alors elle vérifie la propriété de Markov.*

ii) Réciproquement, si $(X_n)_{n\geqslant 0}$ *vérifie la propriété de Markov et si de plus, pour tout* i *tel que* $\mathbf{P}\{X_n = i\} > 0$, *la quantité* $\mathbf{P}\{X_{n+1} = j \mid X_n = i\}$ *ne dépend que de* i *et* j, *mais pas de* n, *alors il existe une matrice stochastique* $\mathbf{P}$ *telle que* $(X_n)_{n\geqslant 0}$ *soit une chaîne de Markov de matrice de transition* $\mathbf{P}$.

DÉMONSTRATION : Soit $(X_n)_{n\geqslant 0}$ une chaîne de Markov de matrice de transition $\mathbf{P}$ et de loi initiale π_0. Fixons un entier n et des indices $i_0, \ldots, i_{n+1}$, et supposons que

$$\mathbf{P}\{X_0 = i_0, \ldots, X_n = i_n\} > 0$$

afin de pouvoir calculer les probabilités conditionnelles. Le membre de gauche de la relation (**) s'écrit

$$\begin{aligned} G &= \mathbf{P}\{X_{n+1} = i_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n\} \\ &= \frac{\mathbf{P}\{X_0 = i_0, \ldots, X_n = i_n, X_{n+1} = i_{n+1}\}}{\mathbf{P}\{X_0 = i_0, \ldots, X_n = i_n\}} \\ &= \frac{\pi_0(i_0)\,P(i_0,i_1)\,P(i_1,i_2)\cdots P(i_{n-1},i_n)\cdot P(i_n,i_{n+1})}{\pi_0(i_0)\,P(i_0,i_1)\,P(i_1,i_2)\cdots P(i_{n-1},i_n)} = P(i_n,i_{n+1}). \end{aligned}$$

Il reste maintenant à identifier le membre de droite de la relation (**) à cette quantité. Écrivons d'abord

$$\mathbf{P}\{X_{n+1} = i_{n+1} \mid X_n = i_n\} = \frac{\mathbf{P}\{X_n = i_n, X_{n+1} = i_{n+1}\}}{\mathbf{P}\{X_n = i_n\}} = \frac{N}{D}.$$

Le numérateur peut se décomposer en sommant sur toutes les valeurs de $X_0, \ldots, X_{n-1}$:

$$\begin{aligned} N &= \sum_{k_0,\ldots,k_{n-1}} \mathbf{P}\{X_0 = k_0, \ldots, X_{n-1} = k_{n-1}, X_n = i_n, X_{n+1} = i_{n+1}\} \\ &= \sum_{k_0,\ldots,k_{n-1}} \pi_0(k_0)\,P(k_0,k_1)\cdots P(k_{n-1},i_n)\,P(i_n,i_{n+1}) \end{aligned}$$

d'après la relation (*) de la définition 27.3.

De la même manière, le dénominateur vaut

$$\begin{aligned} D &= \sum_{k_0,\ldots,k_{n-1}} \mathbf{P}\{X_0 = k_0, \ldots, X_{n-1} = k_{n-1}, X_n = i_n\} \\ &= \sum_{k_0,\ldots,k_{n-1}} \pi_0(k_0)\,P(k_0,k_1)\cdots P(k_{n-1},i_n), \end{aligned}$$

ce qui fait que $N/D = P(i_n, i_{n+1})$ et la relation (**) est démontrée.

Réciproquement, supposons cette propriété vérifiée par la suite $(X_n)_{n\geqslant 0}$, et que de plus la quantité

$$\varphi(n,i,j) := \mathbf{P}\{X_{n+1} = j \mid X_n = i\}$$

ne dépende pas de n ; notons-la $P(i,j)$. Alors

$$\mathbf{P}\{X_0 = i_0, X_1 = i_1\} = \mathbf{P}\{X_0 = i_0\}\cdot\mathbf{P}\{X_1 = i_1 \mid X_0 = i_0\} = \pi_0(i_0)\cdot P(i_0,i_1),$$

puis

$$\begin{aligned} \mathbf{P}\{X_0 = i_0, X_1 = i_1, X_2 = i_2\} &= \mathbf{P}\{X_0 = i_0, X_1 = i_1\}\cdot\mathbf{P}\{X_2 = i_2 \mid X_0 = i_0, X_1 = i_1\} \\ &= \pi_0(i_0)\cdot P(i_0,i_1)\cdot\mathbf{P}\{X_2 = i_2 \mid X_1 = i_1\} \qquad \textit{par } (**) \\ &= \pi_0(i_0)\,P(i_0,i_1)\,P(i_1,i_2). \end{aligned}$$

Une récurrence permet de conclure. ∎

■ **Corollaire 27.6** *Si* $(X_n)_{n\geqslant 0}$ *est une chaîne de Markov de matrice de transition* $\mathbf{P}$, *et si* $\mathbf{P}\{X_n = i\} > 0$, *alors*

$$\mathbf{P}\{X_{n+1} = j \,|\, X_n = i\} = P(i,j).$$

§391 **Mise en garde : probabilités conditionnelles et mémoire du passé**

La relation $\mathbf{P}\{X_{n+1} = j \,|\, X_n = i\} = P(i,j)$ est utilisée dans de nombreux raisonnements, mais insistons sur un point : elle est beaucoup plus faible que la propriété de définition (∗).

Même si elle signifie peu ou prou « la probabilité d'aller en j sachant qu'on est en i ne dépend pas du temps », l'*oubli* véritable du passé ne s'en déduit pas, contrairement à une intuition répandue.

Contre-exemple 27.7 Choisissons un univers très simple $\Omega = \{\omega_1, \omega_2\}$, muni de la probabilité uniforme, et considérons la suite $(X_n)_{n\geqslant 0}$ définie par

$$X_{2n} = 0 \qquad X_{2n+1}(\omega_1) = 1 \qquad X_{2n+1}(\omega_2) = -1.$$

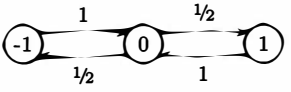
Selon la réalisation ω du hasard, la suite oscille donc entre les états 0 et 1, ou bien entre les états 0 et -1. On peut certes poser

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{pmatrix}$$

et vérifier que toutes les probabilités conditionnelles vérifient la conclusion du corollaire précédent :

$$\mathbf{P}\{X_{2n+1} = 1 \,|\, X_{2n} = 0\} = \tfrac{1}{2} \qquad \mathbf{P}\{X_{2n+1} = -1 \,|\, X_{2n} = 0\} = \tfrac{1}{2}$$
$$\mathbf{P}\{X_{2n} = 0 \,|\, X_{2n-1} = 1\} = 1 \qquad \mathbf{P}\{X_{2n} = 0 \,|\, X_{2n-1} = -1\} = 1.$$

Si X_n représente l'état d'un système à l'instant n, on peut vouloir traduire les relations précédentes par le graphe ci-dessous :



Malgré ces égalités numériques, le système garde la mémoire de son passé : si $X_{2n-1} = 1$, alors on est certain que $X_{2n+1} = 1$! Ainsi, la suite $(X_n)_{n\geqslant 0}$ **n'est pas** une chaîne de Markov (et il serait dangereusement trompeur de la représenter par le graphe précédent).

La situation décrite est celle d'une balle de ping-pong dans une partie opposant d'un côté Alice, et de l'autre Bob et Charlie. Alice renvoie la balle *alternativement* à Bob et à Charlie, en choisissant aléatoirement le premier coup. Le processus n'est pas markovien. Si Alice se fait remplacer par Andreï qui, lui, renvoie aléatoirement la balle à Bob ou Charlie *à chaque coup*, le processus est cette fois markovien.

§392 **Graphe associé à une chaîne de Markov**

On représente une chaîne de Markov par un graphe orienté :

— chaque état (élément de $\mathscr{E}$) est représenté par un sommet du graphe ;

— si la probabilité $P(i,j)$ de transition entre i et j est non nulle, on relie les sommets i et j par une arête orientée de poids $P(i,j)$.

Exemple 27.8 (Mouvement sur un échiquier) On considère un échiquier 3×3, sur la première ligne duquel se trouve, à l'instant $t = 0$, un pion, avec probabilité 1/3 pour chaque case. Ce pion peut, à chaque tour, se déplacer verticalement ou horizontalement, d'une seule case à la fois et ce, avec une probabilité uniforme sur l'ensemble de ses choix possibles. On s'intéresse seulement à la probabilité qu'au tour n, le pion soit au centre, dans un coin ou au milieu d'un côté. La suite $(X_n)_{n\geqslant 0}$ de ses positions forme une chaîne de Markov, dont on représente le graphe, la matrice de transition et la loi initiale ci-après.

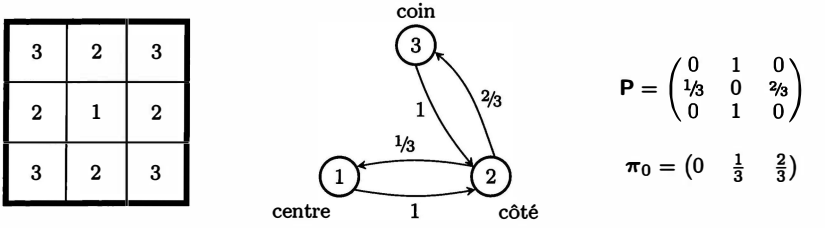


FIGURE 27.1 — *L'échiquier, son graphe de Markov et sa matrice de transition.*

Propriétés des matrices stochastiques § 393

On dira d'une matrice M (de taille quelconque) qu'elle est **positive**, et on écrira $M \geqslant 0$, si tous ses coefficients sont positifs[1]. De même, on dira que M est **strictement positive** et on écrira $M > 0$ si tous ses coefficients sont strictement positifs. On note $\mathbf{1}$ le vecteur colonne dont tous les coefficients valent 1 :

$$\mathbf{1} := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathfrak{M}_{d,1}(\mathbb{R}).$$

■ **Proposition 27.9 (Caractérisation)** *Une matrice* $\mathbf{P}$ *est stochastique si et seulement si*

$$\begin{cases} \mathbf{P} \geqslant 0 \\ \mathbf{P}\mathbf{1} = \mathbf{1}. \end{cases}$$

■ **Corollaire 27.10** *Soient* $\mathbf{P}$ *et* $\mathbf{Q}$ *des matrices stochastiques, alors*
— $\mathbf{P}\mathbf{Q}$ *est également stochastique ;*
— $\mathbf{P}^n$ *est stochastique pour tout* $n \in \mathbb{N}$.

DÉMONSTRATION : Le produit de deux matrices positives est une matrice positive, et si $\mathbf{P}$ et $\mathbf{Q}$ sont stochastiques, alors $\mathbf{P}\mathbf{Q}\mathbf{1} = \mathbf{P}\mathbf{1} = \mathbf{1}$. ∎

Mesures sur l'espace des états § 394

Une probabilité sur l'espace $\mathscr{E} = [\![1 ; d]\!]$ des états peut être représentée sous la forme d'un vecteur ligne

$$\pi = (\pi_1, \pi_2, \ldots, \pi_d)$$

à coefficients positifs ($\pi \geqslant 0$), correctement normalisé : la somme des coefficients vaut 1, ce qui s'écrit

$$\pi\mathbf{1} = 1.$$

Notamment, si π est une probabilité sur $\mathscr{E}$ et $\mathbf{P}$ une matrice stochastique, alors $\pi\mathbf{P}$ est encore une probabilité : $\pi\mathbf{P} \geqslant 0$ et $\pi\mathbf{P}\mathbf{1} = \pi\mathbf{1} = \mathbf{1}$.

---

(1). On prendra garde à ne pas confondre avec la notion de *matrice symétrique positive*, qui est une matrice dont le spectre est positif, les coefficients pouvant être de signe quelconque. La notion de matrice à coefficients positifs n'est pas une notion géométrique, elle n'est pas conservée par exemple par similitude ou congruence : les matrices $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ et $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ sont semblables mais la seconde n'est pas positive.

La propriété de stochasticité n'est pas non plus une propriété géométrique : si $\mathbf{P}$ est stochastique et si $Q \in \mathrm{GL}_k(\mathbb{R})$ est une matrice inversible, alors $Q^{-1}\mathbf{P}Q$ n'a aucune raison d'être stochastique. Par exemple, les matrices $\mathbf{P} := \frac{1}{2}\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ et $D := \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ sont semblables, mais D n'est évidemment pas stochastique.

Une fonction f sur $\mathscr{E}$ peut être représentée par un vecteur

$$\mathbf{F} = \begin{pmatrix} f(1) \\ \vdots \\ f(d) \end{pmatrix}.$$

Si X est une variable aléatoire prenant ses valeurs dans $\mathscr{E}$, de loi

$$\boldsymbol{\pi} = \big(\mathbf{P}\{\mathrm{X}=1\}, \ldots, \mathbf{P}\{\mathrm{X}=d\}\big),$$

alors l'espérance de $f(\mathrm{X})$ vaut

$$\mathbf{E}\big(f(\mathrm{X})\big) = \sum_{k=1}^{d} f(k)\,\mathbf{P}\{\mathrm{X}=k\} = \boldsymbol{\pi}\,\mathbf{F}.$$

Une mesure particulière nous intéressera, celle où le système est, de manière certaine, dans un état i donné à l'instant initial ; la loi initiale est de la forme $\boldsymbol{\pi}_0 = (0, \ldots, 0, 1, 0, \ldots, 0)$. On notera $\mathbf{P}_i$ la loi associée, c'est-à-dire la probabilité conditionnée par l'événement $\{\mathrm{X}_0 = i\}$. Ainsi,

$$\mathbf{P}_i\{\mathrm{X}_n = j\} = \mathbf{P}\{\mathrm{X}_n = j \,|\, \mathrm{X}_0 = i\}.$$

De même, on notera $\mathbf{E}_i$ l'espérance calculée avec cette probabilité, c'est-à-dire l'espérance conditionnée par $\{\mathrm{X}_0 = i\}$.

§ 395 **Évolution à n pas et équation de Chapman-Kolmogorov**

Si $\boldsymbol{\pi}_0$ est la loi initiale de la chaîne, c'est-à-dire la loi de X_0 sur $\mathscr{E}$, la loi de X_1 s'écrit, grâce à la relation de définition $(*)$:

$$\pi_1(j) = \mathbf{P}\{\mathrm{X}_1 = j\} = \sum_{i\in\mathscr{E}} \mathbf{P}\{\mathrm{X}_0 = i, \mathrm{X}_1 = j\} = \sum_{i\in\mathscr{E}} \pi_0(i)\,\mathrm{P}(i,j)$$

c'est-à-dire, sous forme matricielle,

$$\boldsymbol{\pi}_1 = \boldsymbol{\pi}_0\,\mathbf{P}.$$

Définissons $\boldsymbol{\pi}_n$ comme **loi instantanée à l'instant n**, c'est-à-dire le vecteur ligne de coefficients

$$\pi_n(i) := \mathbf{P}\{\mathrm{X}_n = i\}.$$

En sommant la relation $(*)$ sur les valeurs possibles de $i_1, \ldots, i_{n-1}$, on obtient

$$\boldsymbol{\pi}_n = \pi_0\,\mathbf{P}^n.$$

Remarque 27.11 On peut également obtenir ces relations en utilisant plutôt la propriété de Markov $(**)$ et la formule des probabilités totales, qui donnent

$$\mathbf{P}\{\mathrm{X}_{n+1} = j\} = \sum_{i\in\mathscr{E}} \mathbf{P}\{\mathrm{X}_n = i\}\ \mathbf{P}\{\mathrm{X}_{n+1} = j \,|\, \mathrm{X}_n = i\}$$

soit, sous forme matricielle, $\boldsymbol{\pi}_{n+1} = \boldsymbol{\pi}_n\,\mathbf{P}$; une récurrence permet de conclure.

Dans la suite, nous noterons $\mathrm{P}^n(i,j)$ les coefficients de la matrice $\mathbf{P}^n$.

La relation $\mathbf{P}^{m+n} = \mathbf{P}^m\,\mathbf{P}^n$ se réécrit sous la forme développée

$$\mathrm{P}^{m+n}(i,j) = \sum_{k=1}^{d} \mathrm{P}^m(i,k)\,\mathrm{P}^n(k,j),$$

c'est-à-dire

■ **Théorème 27.12 (Chapman-Kolmogorov)** *Soient $i, j \in \mathscr{E}$ et $m, n \in \mathbb{N}$. Alors*

$$\mathbf{P}\{X_{m+n} = j \,|\, X_0 = i\} = \sum_{k=1}^{d} \mathbf{P}\{X_n = j \,|\, X_0 = k\}\ \mathbf{P}\{X_m = k \,|\, X_0 = i\}.$$

Évolution du système aux grandes échelles de temps § 396

La connaissance de la loi initiale $\boldsymbol{\pi}_0$ du système permet de connaître la loi $\boldsymbol{\pi}_n$ à l'instant n par la relation $\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 \mathbf{P}^n$. La question naturelle qui surgit est : que peut-on dire du comportement de cette loi quand $n \to \infty$? Y a-t-il convergence ? Vers quoi ? Et en quel sens ?

Évidemment, le comportement de $\mathbf{P}^n$ est directement relié au spectre de $\mathbf{P}$. Supposons un instant le cas tout à fait favorable où la matrice $\mathbf{P}$ est diagonalisable ; si l'on note $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_d)$ une base de vecteurs propres (à gauche[2]) associée aux valeurs propres $(\lambda_1, \lambda_2, \ldots, \lambda_d)$ et si l'on décompose $\boldsymbol{\pi}_0$ sous la forme

$$\boldsymbol{\pi}_0 = \alpha_1 \boldsymbol{\mu}_1 + \cdots + \alpha_d \boldsymbol{\mu}_d,$$

alors

$$\boldsymbol{\pi}_n = \lambda_1^n \alpha_1 \boldsymbol{\mu}_1 + \cdots + \lambda_d^n \alpha_d \boldsymbol{\mu}_d. \tag{27.1}$$

La première bonne nouvelle est la suivante :

■ **Lemme 27.13** *Pour toute valeur propre λ de $\mathbf{P}$, on a $|\lambda| \leqslant 1$. De plus, $1 \in \operatorname{Sp} \mathbf{P}$ et il existe donc au moins un vecteur ligne $\boldsymbol{\mu}$ tel que $\boldsymbol{\mu}\mathbf{P} = \boldsymbol{\mu}$.*

DÉMONSTRATION : Puisque $\mathbf{P}\mathbf{1} = \mathbf{1}$, 1 est valeur propre de $\mathbf{P}$. Le spectre de $\mathbf{P}$ et celui de ${}^t\mathbf{P}$ étant égaux, on en déduit l'existence d'un vecteur propre à gauche $\boldsymbol{\mu}$ pour $\mathbf{P}$.

Par ailleurs, si $\lambda \in \operatorname{Sp} \mathbf{P}$, et si X est un vecteur propre de $\mathbf{P}$ associé à λ, notons i un indice tel que $|x_i|$ soit maximal (et donc non nul) ; la i-ième ligne de l'équation $\mathbf{P}X = \lambda X$ fournit $\lambda x_i = \sum_j \mathrm{P}(i,j)\, x_j$ donc

$$|\lambda| \leqslant \sum_{j=1}^{d} \mathrm{P}(i,j) \left|\frac{x_j}{x_i}\right| \leqslant \sum_{j=1}^{d} \mathrm{P}(i,j) = 1. \qquad ∎$$

Dans un cas encore plus favorable, les valeurs propres non égales à 1 vont vérifier $|\lambda| < 1$, et dans ce cas la plupart des termes de l'équation (27.1) vont tendre vers 0, ne laissant que les termes constants correspondant à la projection de $\boldsymbol{\pi}_0$ sur le sous-espace propre à gauche[3] $\mathrm{E_g}(1, \mathbf{P})$ associé à 1. Si enfin on est dans un jour particulièrement chanceux et que la dimension de $\mathrm{E_g}(1, \mathbf{P})$ vaut 1, la suite $(\boldsymbol{\pi}_n)_n$ va converger, et sa limite sera l'unique vecteur ligne positif[4] $\boldsymbol{\pi}^*$ de $\mathrm{E_g}(1, \mathbf{P})$ tel que $\boldsymbol{\pi}^* \mathbf{1} = 1$. Un tel vecteur est une **probabilité** (ou **loi**) **invariante**.

Si le lemme précédent nous assure qu'en dimension finie, un vecteur propre pour la valeur propre 1 (autrement appelé **mesure invariante**) existe toujours, ses coefficients peuvent *a priori* être de signes quelconques. Le théorème de Perron-Frobenius (page 565) nous assure heureusement qu'il n'en est rien, et qu'il existe effectivement une *loi de probabilité* invariante.

Quant à savoir si elle est unique, et si la suite $(\boldsymbol{\pi}_n)_{n \geqslant 0}$ converge vers cette loi invariante, c'est plus compliqué, et plusieurs phénomènes peuvent faire obstruction à toutes les étapes exposées :

---

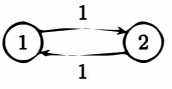
(2). C'est-à-dire des vecteurs lignes non nuls vérifiant la relation $\boldsymbol{\mu}_i \mathbf{P} = \lambda_i \boldsymbol{\mu}_i$; à une transposition près, ce sont donc des vecteurs propres de ${}^t\mathbf{P}$.

(3). On notera $\mathrm{E_g}(\lambda, \mathbf{P})$ le sous-espace des vecteurs lignes propres de $\mathbf{P}$, c'est-à-dire des lignes vérifiant $\mathrm{L}\mathbf{P} = \lambda \mathrm{L}$. La dimension du sous-espace propre à gauche est égale à celle du sous-espace propre à droite.

(4). Nécessairement positif, en tant que limite d'une suite positive.

- **1re obstruction :** il existe une valeur propre non égale à 1, mais de module 1.

Exemple 27.14 Prenons la chaîne associée au graphe et à la matrice suivants :



$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Le spectre de $\mathbf{P}$ est $\{-1, 1\}$. Si l'état initial est $\pi_0 = (1, 0)$, alors on aura pour tout entier n

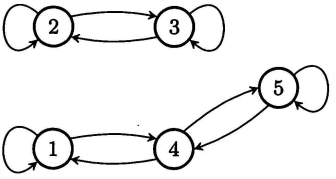
$$\pi_{2n} = (1, 0) \qquad \text{et} \qquad \pi_{2n+1} = (0, 1)$$

c'est-à-dire que la suite des lois instantanées va osciller sans fin, et il n'y a pas d'état limite, bien qu'il existe une probabilité stationnaire $\pi^* = (1/2, 1/2)$. Noter que la suite $(\pi_n)_{n \geqslant 0}$ converge malgré tout au sens des moyennes de Cesàro.

Ici, les états 1 et 2 ont la propriété (désagréable) d'être *périodiques*. La périodicité fera l'objet du paragraphe § 400.

- **2e obstruction** Le sous-espace propre $\mathrm{E_g}(1, \mathbf{P})$ est de dimension strictement supérieure à 1. Il peut alors exister plusieurs lois invariantes et la suite $(\pi_n)_{n \geqslant 0}$ peut converger vers l'une ou l'autre, selon la valeur de la loi initiale.

Exemple 27.15 Considérons une chaîne formée de deux parties ne communiquant pas du tout, et sa matrice de transition associée (où les « $*$ » indiquent des coefficients non nuls, et les « $\cdot$ » des zéros) :



$$\mathbf{P} = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{pmatrix} * & \cdot & \cdot & * & \cdot \\ \cdot & * & * & \cdot & \cdot \\ \cdot & * & * & \cdot & \cdot \\ * & \cdot & \cdot & \cdot & * \\ \cdot & \cdot & \cdot & * & * \end{pmatrix} \end{array} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix}$$

Quitte à réordonner les états et les choisir dans l'ordre $\{2, 3, 1, 4, 5\}$, la matrice de transition s'écrit sous la forme

$$\mathbf{P}' = \begin{array}{c} \begin{matrix} 2 & 3 & 1 & 4 & 5 \end{matrix} \\ \left(\begin{array}{cc|ccc} * & * & \cdot & \cdot & \cdot \\ * & * & \cdot & \cdot & \cdot \\ \hline \cdot & \cdot & * & * & \cdot \\ \cdot & \cdot & * & \cdot & * \\ \cdot & \cdot & \cdot & * & * \end{array}\right) \end{array} \begin{matrix} 2 \\ 3 \\ 1 \\ 4 \\ 5 \end{matrix} = \begin{pmatrix} \mathbf{P}_1 & 0 \\ 0 & \mathbf{P}_2 \end{pmatrix}$$

Si les matrices $\mathbf{P}_1$ et $\mathbf{P}_2$ possèdent une probabilité invariante, n'importe quelle combinaison convexe fournit une probabilité invariante pour $\mathbf{P}'$: on a donc perdu l'unicité. Il est d'ailleurs évident que si le système est initialement dans l'ensemble $\{2, 3\}$ (avec une certaine loi de probabilité), il y restera, et de même pour l'ensemble $\{1, 4, 5\}$.

Ici, la matrice $\mathbf{P}$ est *réductible*, et on a identifié deux parties ne communiquant pas entre elles (elles sont même totalement déconnectées), qui forment, comme on le verra plus tard, deux *classes récurrentes*.

Les notions d'états communicants et de matrice irréductible seront abordées aux paragraphes § 397 et § 398.

- **3e obstruction** La matrice $\mathbf{P}$ n'est pas diagonalisable. Seuls les sous-espaces caractéristiques associés à des valeurs propres $|\lambda| < 1$ sont concernés (voir exercice 27.2), et cela ne gêne finalement pas la convergence de la suite $(\pi_n)_{n \geqslant 0}$ (voir démonstration du théorème ergodique page 566). La notion de diagonalisabilité de $\mathbf{P}$ n'est d'ailleurs pas forcément la meilleure approche, puisque c'est une notion géométrique sans équivalent probabiliste clair.

27.2 CLASSIFICATION IMMÉDIATE DES ÉTATS

Nous étudions ici les propriétés des états que l'on peut établir par les seuls coefficients $P(i,j)$ de la matrice de transition.

États communicants § 397

Soient i et j deux états de $\mathscr{E}$. On dit que i **mène à** j, ou que j **est atteignable par** i s'il existe un entier n tel que $P^n(i,j) > 0$. Si l'on développe ce terme sous la forme d'une somme de termes positifs :

$$P^n(i,j) = \sum_{k_1 \in \mathscr{E}} \cdots \sum_{k_{n-1} \in \mathscr{E}} P(i,k_1)\,P(k_1,k_2)\cdots P(k_{n-1},j),$$

l'un de ces termes doit être strictement positif, ce qui montre qu'il existe un chemin

$$i \to k_1 \to k_2 \to \cdots \to k_{n-1} \to j$$

de probabilité non nulle. On écrit alors (indifféremment) $i \to j$ ou $j \leftarrow i$.

Si $i \to j$ et $j \to i$, on dit que i et j **communiquent**, et on écrit $i \leftrightarrow j$.

Exemple 27.16 On considère les chaînes de Markov dont les graphes sont les suivants :



Dans le graphe de gauche, tous les états communiquent. Dans celui de droite, $1 \leftrightarrow 2$ mais $1 \not\leftrightarrow 3$ et $2 \not\leftrightarrow 3$.

Dire que tous les états communiquent revient donc à dire que, quels que soient les indices i et j, il existe une puissance $k \geqslant 0$ telle que $P^k(i,j) > 0$. Cet entier k dépend de i et j, et rien n'oblige par exemple à avoir $P^{k+1}(i,j) > 0$.

On peut simplifier la vérification en notant $\mathbf{Q} := \frac{1}{2}(\mathbf{I} + \mathbf{P})$, qui est encore une matrice stochastique ; il s'agit de la matrice de transition d'un système construit à partir du système initial en autorisant les hésitations : à chaque instant, le système peut rester inchangé (avec une probabilité $1/2$) ou suivre les règles de transition précédentes.

Si deux états i et j communiquent dans le système initial (régi par $\mathbf{P}$), alors il existe un chemin, de longueur au plus égale à $d-1$, reliant i à j. Ce même chemin est autorisé dans le système modifié, qui de plus peut « rester sur place » jusqu'à ce que le temps écoulé soit égal à $d-1$. Ainsi, il y a équivalence entre les énoncés

i) $i \to j$;

ii) il existe $k \leqslant d-1$ tel que $P^k(i,j) > 0$;

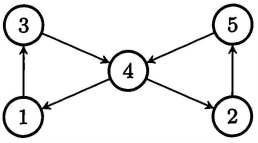
iii) $(\mathbf{I} + \mathbf{P})^{d-1}(i,j) > 0$.

■ **Proposition 27.17** *Tous les états de $\mathscr{E}$ communiquent entre eux (et sont donc accessibles depuis un état initial quelconque) si et seulement si* $(\mathbf{I} + \mathbf{P})^{d-1} > 0$.

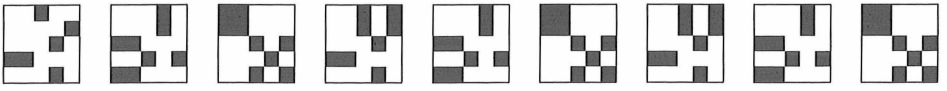
Remarque 27.18 (Sur les puissances successives d'une matrice stochastique) Comme on l'a vu, la non nullité d'un coefficient $\mathrm{P}^k(i,j)$ n'implique pas celle du coefficient $\mathrm{P}^{k+1}(i,j)$. En revanche, la suite des puissances de $\mathbf{Q} := \mathbf{I} + \mathbf{P}$ est croissante, au sens ou $\mathbf{Q}^{n+1} - \mathbf{Q}^n \geqslant 0$. En effet, par positivité des coefficients,

$$\mathrm{Q}^{n+1}(i,j) = \sum_{k=1}^{d} \mathrm{Q}^n(i,k)\big(\overbrace{\mathrm{I}(k,j)}^{\delta_{k,j}} + \mathrm{P}(k,j)\big) \geqslant \mathrm{Q}^n(i,j)\big(1 + \mathrm{P}(j,j)\big) \geqslant \mathrm{Q}^n(i,j).$$

On affiche ci-dessous les neuf premières puissances de la matrice $\mathbf{P}$ associée à la chaîne

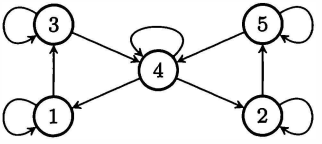


Un coefficient non nul est représenté par un carré noir, un coefficient nul par un carré blanc ; convenons d'appeler *silhouette* de la matrice cette représentation.

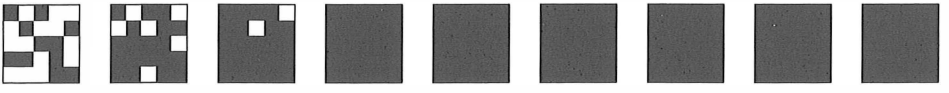


On remarquera qu'à partir de la puissance 2, la suite des silhouettes est 3-périodique. Que lit-on dans la deuxième matrice ? Que depuis l'état 4, par exemple, on peut atteindre 3 et 5 en deux étapes.

Autorisons maintenant le stationnement à chaque état : le graphe devient



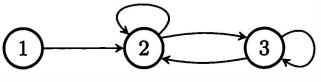
Les puissances successives de $\mathbf{Q} := \frac{1}{2}(\mathbf{I} + \mathbf{P})$ se remplissent plus vite : on atteint tous les états depuis l'état 4 en deux étapes. La puissance quatrième de $\mathbf{Q}$ est déjà strictement positive, tous les états étant accessibles depuis n'importe où en quatre étapes. Voici les silhouettes des $\mathbf{Q}^n$.



Remarque 27.19 (États de non-retour, états transitoires) Puisque $\mathbf{P}^0 = \mathbf{I}$, tout état communique avec lui-même en un temps nul. Cependant, il se peut que, à partir du moment où l'on quitte l'état i, il soit impossible d'y revenir de nouveau, c'est-à-dire que

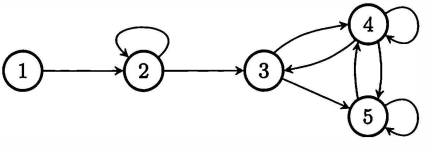
$$\forall n \geqslant 1 \qquad \mathrm{P}^n(i,i) = 0.$$

Un tel état est appelé **état de non-retour**. Dans le graphe ci-dessous, l'état 1 est un état de non-retour :



Plus généralement, un **état inessentiel**, ou **transitoire**, est un état tel que, presque sûrement, on finira par le quitter pour n'y plus jamais revenir, c'est-à-dire qu'il existe un état j acessible depuis i mais tel que i n'est pas accessible depuis j. Plus formellement, il existe un état $j \neq i$ et une puissance n tels que $\mathrm{P}^n(i,j) > 0$ mais $\mathrm{P}^m(j,i) = 0$ pour tout m.

Dans la chaîne suivante, l'état 1 est de non retour, l'état 2 est transitoire : en partant de 1 ou 2, on finira (presque sûrement) par atteindre 3 tôt ou tard ; dès lors, plus de retour possible !



Notion de classes. Graphes et matrices irréductibles. § 398

La relation « $\leftrightarrow$ » définit une relation d'équivalence sur l'ensemble $\mathscr{E}$; ainsi, $\mathscr{E}$ est réunion disjointe des classes d'équivalence selon cette relation. Ces classes sont appelées **classes indécomposables** de la chaîne. Au sein d'une même classe, on peut communiquer entre deux états quelconques. En revanche, si C et C′ sont deux classes distinctes et si l'on peut accéder de C à C′, alors le retour de C′ à C est impossible.

Dans l'exemple 27.16, pour le graphe de droite, l'ensemble $\mathscr{E} = \{1, 2, 3\}$ se décompose en deux classes : $\{1, 2\}$ et $\{3\}$.

Dans la représentation matricielle d'un graphe par une matrice de transition, il peut être intéressant de regrouper les états selon leurs classes, c'est-à-dire de renuméroter $\mathscr{E}$, que l'on avait ordonné implicitement en le notant $\{1, 2, \ldots, d\}$, grâce à une permutation σ. Après cette permutation, la matrice de transition prend la forme

$$\mathbf{P}' = \mathrm{Q}_\sigma^{-1}\, \mathbf{P}\, \mathrm{Q}_\sigma,$$

où Q_σ est la matrice représentative de la permutation σ ; ses coefficients vérifient $\mathrm{Q}_\sigma(i, j) = \delta_{j,\sigma(i)}$. Cette matrice est orthogonale ($\mathrm{Q}^{-1} = {}^t\mathrm{Q}$), et agit sur $\mathbf{P}$ en effectuant une permutation simultanément sur les lignes et sur les colonnes. La matrice $\mathbf{P}'$ représente les transitions des états classés dans l'ordre $\sigma(1), \ldots, \sigma(d)$.

On peut faire encore mieux que regrouper les éléments par classe. Sur l'ensemble $\mathcal{C} = \{\mathrm{C}_1, \ldots, \mathrm{C}_m\}$ des classes, définissons la relation « est accessible depuis », notée « $\leftarrow$ », par :

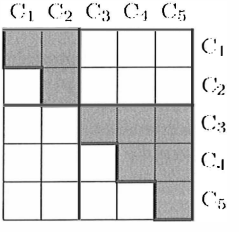
$$\mathrm{C} \leftarrow \mathrm{C}' \text{ s'il existe un état } i \in \mathrm{C} \text{ et un état } j \in \mathrm{C}' \text{ tels que } i \leftarrow j.$$

Puisque tous les états de C (et de C′) communiquent, cela équivaut à :

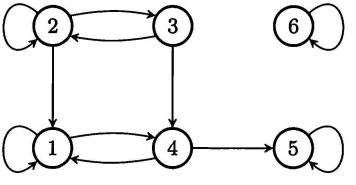
tout état de C est accessible depuis tout état de C′.

La relation $\leftarrow$ est une relation d'ordre non-total. On range alors les classes dans un ordre $\mathrm{C}_1, \ldots, \mathrm{C}_m$ de telle manière que C_i ne soit jamais accessible depuis C_{i+1}. Pour ce faire, lorsque des classes sont comparables, on les ordonne ; par exemple, $\mathrm{C}_2 \leftarrow \mathrm{C}_1$ et $\mathrm{C}_5 \leftarrow \mathrm{C}_4 \leftarrow \mathrm{C}_3$, mais C_2 et C_3 ne communiquent ni dans un sens, ni dans l'autre. On choisit une permutation de l'espace des états qui respecte cet ordre de classes et, dans cet exemple, la matrice $\mathbf{P}'$ a par blocs la forme suivante, où les cases en blanc n'ont que des coefficients nuls :

$\mathbf{P}' =$



Exemple 27.20 Considérons la chaîne associée au graphe suivant, et sa matrice de transition $\mathbf{P}$:



$$\mathbf{P} = \begin{pmatrix} * & \cdot & \cdot & * & \cdot & \cdot \\ * & * & * & \cdot & \cdot & \cdot \\ \cdot & * & \cdot & * & \cdot & \cdot \\ * & \cdot & \cdot & \cdot & * & \cdot \\ \cdot & \cdot & \cdot & \cdot & * & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & * \end{pmatrix}$$

On peut visuellement identifier quatre classes : $C_1 = \{2, 3\}$, $C_2 = \{1, 4\}$, $C_3 = \{5\}$ et $C_4 = \{6\}$. Les deux premières seront appelées *classes transitoires*, car tous leurs éléments sont transitoires. Elles communiquent ainsi :

$$C_1 \to C_2 \to C_3 \qquad C_4$$

La matrice de transition, en considérant les états dans l'ordre 2, 3, 1, 4, 5, 6, est

$$\mathbf{P}' = \begin{pmatrix} * & * & * & \cdot & \cdot & \cdot \\ * & \cdot & \cdot & * & \cdot & \cdot \\ \cdot & \cdot & * & * & \cdot & \cdot \\ \cdot & \cdot & * & \cdot & * & \cdot \\ \cdot & \cdot & \cdot & \cdot & * & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & * \end{pmatrix}$$

Définition 27.21 (Matrice irréductible) On dit que la chaîne $(X_n)_{n\geqslant 0}$ (ou que la matrice $\mathbf{P}$ associée) est **irréductible** si l'une des propriétés équivalentes suivantes est vérifiée :

- $\mathscr{E}$ est la seule classe d'équivalence.
- Il n'existe aucune matrice de permutation Q_σ telle que $\mathbf{P}' := {}^tQ_\sigma \, \mathbf{P} \, Q_\sigma$ soit triangulaire par blocs, c'est-à-dire de la forme

$$\mathbf{P}' = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \tag{27.2}$$

- Pour tout choix de deux états i et j, il existe une puissance n (dépendant de ce choix) de la matrice telle que $P^n(i, j) > 0$; autrement dit, tous les états communiquent.
- $(\mathbf{I} + \mathbf{P})^{d-1} > 0$.

Une matrice (ou la chaîne associée) est dite **réductible** lorsqu'elle n'est pas irréductible, c'est-à-dire lorsqu'il existe une permutation mettant $\mathbf{P}$ sous la forme (27.2) précédente.

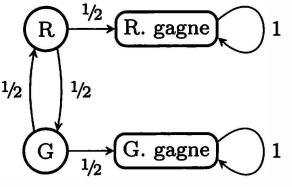
§ 399 **Parties fermées. États absorbants**

Une partie $\mathscr{F}$ de $\mathscr{E}$ est dite **fermée** ou **absorbante** si on ne peut plus en sortir une fois qu'on y est entré, c'est-à-dire

$$\forall x \in \mathscr{F} \quad \forall y \in \mathscr{E} \qquad (x \to y) \implies y \in \mathscr{F}.$$

Un état j est dit **état final** ou **état absorbant** si $\{j\}$ est une partie absorbante, c'est-à-dire si $P(j, j) = 1$.

Remarque 27.22 Il peut être intéressant, dans la modélisation d'un système, d'ajouter un état absorbant particulier ; par exemple, dans la description d'une partie infinie de *pile* ou *face*, on déclare vainqueur celui qui, pour la première fois, obtient *pile*. Si l'on note R et G les états correspondants à « Rosencrantz joue » et « Guildenstern joue », on peut ajouter les états « Rosencrantz gagne » et « Guildenstern gagne ». Le graphe de la chaîne de Markov est alors



Les deux états de gain sont absorbants et signalent la fin de la partie.

Une partie $\mathscr{F}$ est absorbante s'il existe une permutation des états telle que la matrice de transition soit de la forme

$$\mathbf{P}' = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix}_{\leftarrow \mathscr{F}} \quad \mathscr{F} \downarrow$$

Le lecteur se convaincra que

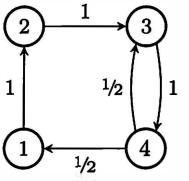
■ **Proposition 27.23** *Une chaîne est réductible si et seulement si elle admet au moins une classe absorbante non égale à $\mathscr{E}$.*

Remarque 27.24 Le fait qu'une partie soit absorbante se lit « horizontalement » sur la matrice de transition, au contraire de la stabilité d'un sous-espace vectoriel que l'on lit « verticalement » en algèbre linéaire.

Les questions naturelles à se poser quand on a identifié une classe absorbante sont : avec quelle probabilité finira-t-on dans cette classe ? Combien de temps s'écoulera-t-il avant que l'on s'y retrouve ?

Périodicité § 400

Considérons la chaîne de Markov à quatre états définie par le graphe suivant :



Supposons maintenant que l'état de départ soit l'état 1. Il est possible de revenir à l'état 1 en 4 étapes (chemin $1 \to 2 \to 3 \to 4 \to 1$), ou bien en 6 étapes (chemin $1 \to 2 \to 3 \to 4 \to 3 \to 4 \to 1$)... et finalement, en n'importe quel nombre de l'ensemble $\{4, 6, 8, 10, \ldots, 2n, \ldots\}$ — il suffit de piétiner entre les états 3 et 4 suffisamment de fois avant de revenir en 1.

De manière générale, pour un état initial i donné, notons $\mathscr{J}_i$ l'ensemble des nombres d'étapes en lequel on peut revenir à l'état i :

$$\mathscr{J}_i = \{n \geqslant 1 \ ; \ \mathrm{P}^n(i,i) > 0\}.$$

Alors cet ensemble est stable par addition (et donc par passage aux multiples) : si l'on peut partir de i et y revenir en m étapes, mais également en n étapes, alors on peut certainement faire se succéder les deux chemins et y revenir en $m+n$ étapes. Plus formellement, si $\mathrm{P}^m(i,i) > 0$ et $\mathrm{P}^n(i,i) > 0$, alors, par positivité de tous les termes,

$$\mathrm{P}^{m+n}(i,i) = \sum_{k \in \mathscr{E}} \mathrm{P}^m(i,k)\,\mathrm{P}^n(k,i) \geqslant \mathrm{P}^m(i,i)\,\mathrm{P}^n(i,i) > 0.$$

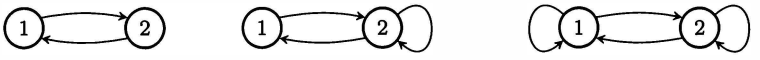
Notons δ le PGCD de l'ensemble $\mathscr{J}_i$. Si $\delta > 1$, on dit que l'état i est **périodique**, de période δ. Si $\delta = 1$, on dit que l'état i est **apériodique**.

Remarque 27.25 Dans l'exemple précédent, l'état 1 est périodique, de période 2 ; on remarquera que, cependant, $\mathrm{P}^2(1,1) = 0$.

La notion d'apériodicité ou de périodicité (et la valeur de la période dans ce cas) est une notion de classe : si $i \leftrightarrow j$, alors i et j sont simultanément périodiques (et de même période), ou simultanément apériodiques. On parlera donc de **classe apériodique** et de **période d'une classe**. Enfin, on dira qu'**une chaîne est apériodique** si elle ne possède aucune classe périodique. Notamment, si une chaîne est irréductible, alors tous les états sont équivalents pour la périodicité. La suite nous montrera que les chaînes irréductibles et apériodiques sont les chaînes les plus « sympathiques » pour l'étude du comportement asymptotique.

Remarque 27.26 (Un cas simple d'apériodicité) Supposons que, dans une classe donnée il existe un état i tel que $\mathrm{P}(i,i) > 0$, c'est-à-dire que i boucle sur lui-même. Alors pour tout état j de cette classe, l'ensemble $\mathscr{J}_j$ contient une partie de la forme $\{k, k+1, k+2, \cdots\}$, donc son PGCD vaut 1 et j est apériodique. (Quant à l'ensemble $\mathscr{J}_i$ lui-même, il est égal à $\mathbb{N}^*$.) La classe de i est apériodique.

Exemple 27.27 (Un état apériodique) La chaîne de gauche est périodique de période 2, celles de droite sont apériodiques.



Remarque 27.28 (Caractérisation matricielle d'une classe périodique) Au sein d'une classe indécomposable périodique, des **sous-classes périodiques** apparaissent, qui communiquent de la façon suivante : $\mathrm{C}_1 \to \mathrm{C}_2 \to \cdots \to \mathrm{C}_\delta \to \mathrm{C}_1$, et dans lesquelles on ne peut pas rester plus d'un tour. Matriciellement, la restriction de la matrice $\mathbf{P}$ à cette classe a la forme suivante (après permutation adéquate des états)

$$\mathbf{Q} = \begin{pmatrix} & \mathbf{P}_1 & & & \\ & & \mathbf{P}_2 & & \\ & & & \ddots & \\ & & & & \mathbf{P}_{\delta-1} \\ \mathbf{P}_\delta & & & & \end{pmatrix}$$

§401 **Ergodicité (irréductibilité forte)**

Une chaîne de Markov, ou sa matrice stochastique $\mathbf{P}$ associée, est dite **ergodique** (ou **fortement irréductible**, ou encore **régulière**) s'il existe un entier n tel que $\mathbf{P}^n > 0$, c'est-à-dire

$$\forall i, j \in \mathscr{E} \qquad \mathrm{P}^n(i,j) > 0.$$

À partir de cet entier, les coefficients de $\mathbf{P}^k$ sont tous strictement positifs et, plus précisément, si l'on note $m(\mathbf{P}^n) = \min_{i,j} \mathrm{P}^n(i,j)$,

$$\forall k \geqslant n \quad \forall i, j \in [\![1\,;\,d]\!] \qquad \mathrm{P}^k(i,j) \geqslant m(\mathbf{P}^n) > 0.$$

Toute matrice ergodique est bien entendu irréductible, mais la réciproque n'est pas vraie ; un exemple simple est donné par une chaîne possédant une classe périodique.

Nous admettrons (voir [91]) que :

■ **Théorème 27.29 (Caractérisation des matrices ergodiques)** *Une matrice stochastique* $\mathbf{P}$ *est ergodique si et seulement si elle est irréductible et apériodique.*

Une matrice stochastique irréductible est ergodique si et seulement si 1 *est sa seule valeur propre de module* 1.

Théorème de Perron-Frobenius et théorème ergodique § 402

Le premier théorème que nous présentons est un classique de l'analyse matricielle ; appliqué aux probabilités, il donne ce premier résultat, assurant l'existence d'une loi invariante, mais pas la convergence classique (revoir l'exemple 27.14 page 558).

■ **Théorème 27.30 (Perron-Frobenius)** *Soit* $\mathbf{P}$ *une matrice stochastique irréductible. Alors*

i) 1 *est valeur propre de multiplicité* 1 *; notamment, le sous-espace propre associé est de dimension* 1. *Toutes les valeurs propres de* $\mathbf{P}$ *sont de module inférieur ou égal à* 1.

ii) Il existe une unique loi strictement positive et invariante $\boldsymbol{\pi}^*$, *c'est-à-dire un vecteur ligne tel que* $\boldsymbol{\pi}^* > 0$, $\boldsymbol{\pi}^*\mathbf{1} = \mathbf{1}$ *et* $\boldsymbol{\pi}^*\mathbf{P} = \boldsymbol{\pi}^*$.

iii) Pour toute loi initiale $\boldsymbol{\pi}_0$, *la suite des moyennes de Cesàro* $\frac{1}{n}\sum_{k=1}^{n} \boldsymbol{\pi}_0 \mathbf{P}^k$ *converge vers* $\boldsymbol{\pi}^*$.

Notez qu'il peut tout à fait exister des valeurs propres de module 1, mais différentes de 1, comme dans le cas de la matrice $\mathbf{P} = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ (qui est 2-périodique).

Dans le cas d'une matrice ergodique (et donc apériodique), toutes les obstructions à la convergence que nous avons présentées au paragraphe § 396 tombent. En plus des propriétés précédentes, on a convergence de la suite des mesures instantanées :

■ **Théorème 27.31 (Ergodicité)** *Supposons* $\mathbf{P}$ *ergodique, c'est-à-dire qu'il existe un entier* n *tel que* $\mathbf{P}^n > 0$. *Alors il existe une unique loi invariante*

$$\boldsymbol{\pi}^* = (\pi_1^*, \pi_2^*, \ldots, \pi_d^*)$$

et celle-ci vérifie :

i) Ses composantes sont strictement positives : $\boldsymbol{\pi}^* > 0$.

ii) La suite $(\mathbf{P}^n)_{n\geqslant 0}$ *des puissances de* $\mathbf{P}$ *converge vers la matrice*

$$\mathbf{P}^\infty = \begin{pmatrix} \pi_1^* & \pi_2^* & \cdots & \pi_d^* \\ \pi_1^* & \pi_2^* & \cdots & \pi_d^* \\ \vdots & \vdots & & \vdots \\ \pi_1^* & \pi_2^* & & \pi_d^* \end{pmatrix}.$$

iii) Quel que soit $\boldsymbol{\pi}_0$, *la suite* $\boldsymbol{\pi}_n = \boldsymbol{\pi}_0\mathbf{P}^n$ *converge vers* $\boldsymbol{\pi}^*$ *; autrement dit :*

$$\mathrm{P}^n(i,j) = \mathbf{P}\{\mathrm{X}_n = j \,|\, \mathrm{X}_0 = i\} \xrightarrow[n\to\infty]{} \pi_j^*$$

pour tout j *et ce, indépendamment de* i. *De plus, cette convergence est exponentiellement rapide.*

Ces propriétés caractérisent d'ailleurs les matrices stochastiques.

Nous proposons ci-après la démonstration des propriétés de ces deux théorèmes, dans le cas ergodique. Le lecteur en quête d'une démonstration générale pourra se reporter à de nombreux ouvrages : Shiryaev [91, p. 118] ou, pour une approche plus algébrique, Mansuy [73, p. 175] par exemple.

DÉMONSTRATION : Commençons par un cas particulier, celui où $\mathbf{P}$ est une matrice stochastique strictement positive : $\mathbf{P} > 0$.

- On sait que 1 est valeur propre et que le vecteur $\mathbf{1}$ est vecteur propre associé. *Montrons que* 1 *est la seule valeur propre de module* 1 *de* $\mathbf{P}$.
  Soit $\lambda \in \mathrm{Sp}\,\mathbf{P}$ telle que $|\lambda| = 1$. Soit $X = {}^t(x_1, x_2, \dots, x_d)$ un vecteur propre associé, et notons i un indice tel que $|x_i|$ soit maximal ; alors $x_i \neq 0$ et, quitte à diviser X par x_i, on peut supposer que $x_i = 1$. La i-ième ligne du système $\mathbf{P}X = \lambda X$ s'écrit $\sum_j P(i,j)\,x_j = 1$ donc, par inégalité triangulaire :
  $$1 = \left|\sum_{j=1}^{d} P(i,j)\,x_j\right| \underset{(*)}{\leqslant} \sum_{j=1}^{d} P(i,j)\,|x_j| \underset{(**)}{\leqslant} \sum_{j=1}^{d} P(i,j) = 1.$$
  L'inégalité $(**)$ est donc une égalité, ce qui implique, puisque $\mathbf{P} > 0$ que $|x_j| = 1$ pour tout j ; on reconnaît en $(*)$ le cas d'égalité de l'inégalité triangulaire, d'où l'on déduit que $x_j = 1$ pour tout j. Ainsi, $X = \mathbf{1}$ et $\lambda = 1$.
- Le raisonnement précédent montre de plus que le sous-espace propre (à droite) associé à 1 est la droite engendrée par $\mathbf{1}$. Cela implique également que 1 est racine simple (voir exercice 27.2). Enfin, le sous-espace propre à gauche $E_g(1, \mathbf{P})$ est également de dimension 1.
- Réduisons maintenant la matrice $\mathbf{P}$ sous sa forme de Jordan : il existe des blocs de Jordan $J_1, \dots, J_r$, associés à des valeurs propres $\lambda_1, \dots, \lambda_r$ de module strictement inférieur à 1 :
  $$\mathbf{P} = Q \begin{pmatrix} 1 & & \\ & J_1 & \\ & & \ddots \\ & & J_r \end{pmatrix} Q^{-1} \qquad J_k = \begin{pmatrix} \lambda_k & 1 & & \\ & \ddots & & \\ & & \ddots & 1 \\ & & & \lambda_k \end{pmatrix} \qquad |\lambda_k| < 1.$$
  La suite $(J_k^n)_n$ des puissances de J_k converge vers 0, et ses coefficients sont tous de la forme : un polynôme en n, multiplié par λ_k^n. Si l'on choisit un réel α tel que $|\lambda_k| < \alpha < 1$ pour $k = 1, \dots, r$, alors $J_k^n = o(\alpha^n)$ pour tout k ; notamment,
  $$\mathbf{P}^n \xrightarrow[n\to\infty]{} \mathbf{P}^\infty := Q \begin{pmatrix} 1 & & \\ & 0 & \\ & & \ddots \\ & & 0 \end{pmatrix} Q^{-1}.$$
  Les sous-espaces propres *à gauche* associés à 1 pour $\mathbf{P}$ et $\mathbf{P}^\infty$ sont égaux, et de dimension 1 : $E_g(1, \mathbf{P}) = E_g(1, \mathbf{P}^\infty)$.
  Puisque $\mathbf{P}^n$ est stochastique pour tout n, la limite $\mathbf{P}^\infty$ est également stochastique et, puisqu'elle est de rang 1, on peut l'écrire sous la forme
  $$\mathbf{P}^\infty = \begin{pmatrix} \pi_1^* & \pi_2^* & \cdots & \pi_d^* \\ \pi_1^* & \pi_2^* & \cdots & \pi_d^* \\ \vdots & \vdots & & \vdots \\ \pi_1^* & \pi_2^* & & \pi_d^* \end{pmatrix} = \mathbf{1}\,\boldsymbol{\pi}^* \qquad \boldsymbol{\pi}^* = (\pi_1^*, \pi_2^*, \dots, \pi_n^*)$$
  où $\boldsymbol{\pi}^* \geqslant 0$ et $\boldsymbol{\pi}^* \mathbf{1} = 1$.
- La relation $\mathbf{P}^n\mathbf{P} = \mathbf{P}\mathbf{P}^n = \mathbf{P}^{n+1}$ donne, à la limite $n \to \infty$, $\mathbf{P}\mathbf{P}^\infty = \mathbf{P}^\infty\mathbf{P} = \mathbf{P}^\infty$. De plus,
  $$\boldsymbol{\pi}^*\mathbf{P}^\infty = \boldsymbol{\pi}^*\,\mathbf{1}\,\boldsymbol{\pi}^* = \boldsymbol{\pi}^*,$$
  donc $\boldsymbol{\pi}^* \in E_g(1, \mathbf{P}^\infty)$, donc de $\boldsymbol{\pi}^* \in E_g(1, \mathbf{P})$: autrement dit, $\boldsymbol{\pi}^*$ est une l'unique loi invariante de $\mathbf{P}$.
- S'il existait un indice j tel que $\pi_j^* = 0$, alors l'égalité $\boldsymbol{\pi}^*\mathbf{P} = \boldsymbol{\pi}^*$ impliquerait $0 = \sum_i \pi_i^* P(i,j)$; or dans le membre de droite, l'un des π_i est strictement positif, ainsi que $P(i,j)$, ce qui est contradictoire. Ainsi, $\boldsymbol{\pi}^* > 0$.
- Pour tout choix de loi $\boldsymbol{\pi}_0$, on a $\boldsymbol{\pi}_n = \boldsymbol{\pi}_0\mathbf{P}^n \to \boldsymbol{\pi}_0\,\mathbf{P}^\infty = \boldsymbol{\pi}_0\,\mathbf{1}\,\boldsymbol{\pi}^* = \boldsymbol{\pi}^*$; de plus, $\boldsymbol{\pi}_n - \boldsymbol{\pi}^* = \boldsymbol{\pi}_0(\mathbf{P}^n - \mathbf{P}^\infty)$ décroît exponentiellement vite, en $o(\alpha^n)$.

Il reste à généraliser au cas où $\mathbf{P}$ est ergodique, c'est-à-dire qu'il existe un entier n tel que $\mathbf{P}^n > 0$. Si λ est valeur propre de $\mathbf{P}$ de module 1, alors λ^n est valeur propre de $\mathbf{P}^n$ de module 1 donc, d'après ce qui précède, $\lambda^n = 1$. Or $\mathbf{P}^{n+1} > 0$ donc, pour la même raison, $\lambda^{n+1} = 1$, ce qui implique $\lambda = 1$. Ainsi, $E_g(1, \mathbf{P}^n)$ est une droite. Le sous-espace propre $E_g(1, \mathbf{P})$ étant inclus dans $E_g(1, \mathbf{P}^n)$, il est encore de dimension 1. On peut alors reprendre le raisonnement avec les matrices de Jordan et conclure de même. ∎

27.3 CLASSIFICATION ASYMPTOTIQUES DES ÉTATS

États transitoires, états récurrents § 403

Soient i et j des états ; rappelons que $\mathbf{P}_i$ est la probabilité conditionnée par $\{X_0 = i\}$ et définissons la **probabilité de premier retour à l'état i à l'instant n** et la **probabilité de première arrivée à l'état j à l'instant n** par

$$R^{(n)}(i,i) = \mathbf{P}_i\{X_1 \neq i, X_2 \neq i \ldots, X_{n-1} \neq i, X_n = i\}$$
$$R^{(n)}(i,j) = \mathbf{P}_i\{X_1 \neq j, X_2 \neq j \ldots, X_{n-1} \neq j, X_n = j\}.$$

Ces formules se justifient intuitivement : aller de l'état i à l'état j en n étapes peut se faire en allant, pour la première fois, à l'état j en k étapes, puis en allant de j à j en $n-k$ étapes, et ce pour toutes les valeurs $k = 1, \ldots, n$. Cela nous donne la formule suivante :

$$P^n(i,j) = \sum_{k=1}^{n} R^{(k)}(i,j)\, P^{n-k}(j,j). \tag{27.3}$$

Notons enfin

$$R(i,i) := \sum_{n=1}^{\infty} R^{(n)}(i,i)$$

la probabilité que le système, partant de l'état i, y retourne un jour ou l'autre. Si l'on note

$$T_i := \inf\{n \geqslant 1 \ ; \ X_n = i\},$$

on a donc $R^{(n)}(i,j) = \mathbf{P}_i\{T_j = n\}$ et $R(i,i) = \mathbf{P}_i\{T_i < +\infty\}$. La variable aléatoire T_i est le **temps de retour à l'état i** (on parle également de **temps d'atteinte**).

Définition 27.32 On dit que l'état i est **récurrent** si $R(i,i) = 1$, c'est-à-dire si, presque sûrement, le système partant de l'état i finira par y revenir. Lorsque $R(i,i) < 1$, l'état est dit **transitoire** (ou **transient**).

Le calcul des nombres $R(i,i)$ pouvant être passablement compliqué, voici une deuxième caractérisation de la propriété de récurrence. Commençons par écrire que $P^n(i,i) = \mathbf{E}_i(\mathbf{1}_{\{X_n = i\}})$; par sommation de variables positives, $\sum_{n=1}^{\infty} P^n(i,i)$ est l'espérance du nombre de retours en i.

■ **Proposition 27.33** *Un état est récurrent si la série $\sum_n P^n(i,i)$ diverge, et transitoire si elle converge.*

■ **Proposition 27.34** *La propriété de récurrence est une propriété de classe : si $i \leftrightarrow j$, alors i et j sont simultanément récurrents, ou simultanément transitoires.*

On peut alors parler de **classe récurrente** et de **classe transitoire**.

Enfin, les dénominations « état récurrent » et « état transitoire » se justifient ainsi : si le système, partant de i, y revient presque sûrement, alors l'oubli du passé fait qu'il y reviendra presque sûrement une infinité de fois.

Remarque 27.35 Cette dernière propriété ne peut se démontrer rigoureusement qu'avec la *propriété de Markov forte*. Rappelons qu'une variable aléatoire à valeurs entières T est appelée **temps d'arrêt** pour la chaîne $(X_n)_{n \geqslant 0}$ si l'événement $\{T = n\}$ est dans la tribu engendrée par $X_0, \ldots, X_n$, c'est-à-dire s'il est « fonction » uniquement des valeurs prises par $X_0, \ldots, X_n$ (voir § 251). La **propriété de Markov forte** s'écrit alors

$$\mathbf{P}\{X_{T+1} = i_1, \ldots, X_{T+k} = i_k \,|\, X_T = i \text{ et } T < \infty\} = \mathbf{P}\{X_1 = i_1, \ldots, X_k = i_k \,|\, X_0 = i\}.$$

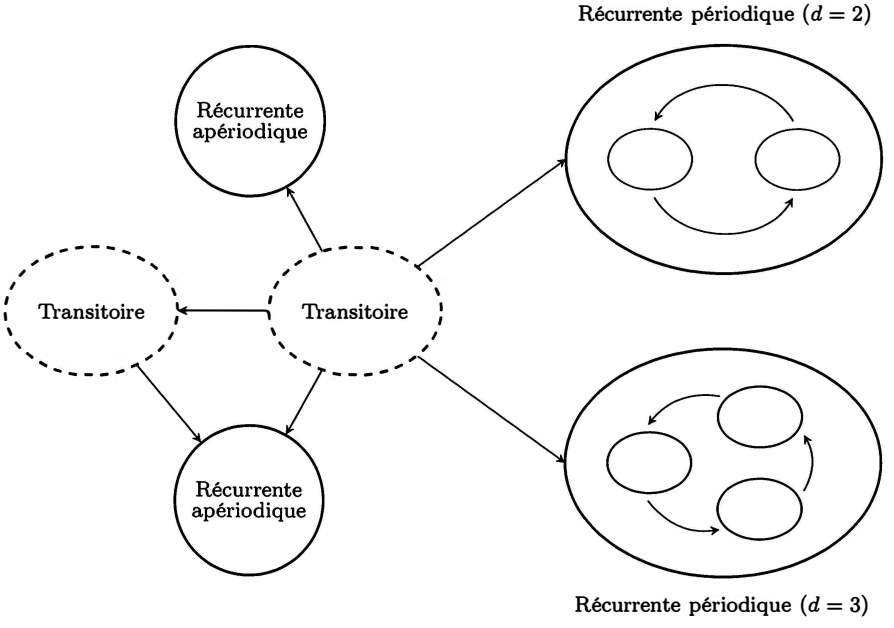
Autrement dit, l'absence de mémoire pour une chaîne de Markov vaut également si on considère un instant aléatoire défini uniquement par la connaissance du passé et du présent de la chaîne (et non de son futur)..

■ **Théorème 27.36** *Soit i un état de $\mathscr{E}$. Alors :*
- *si i est récurrent,* $\mathbf{P}_i\{X_n = i \text{ ; } i.s.\} = 1$;
- *si i est transitoire,* $\mathbf{P}_i\{X_n = i \text{ ; } i.s.\} = 0$.

Notons qu'il n'y a pas d'autre choix : ou bien (presque sûrement) le système reviendra une infinité de fois à l'état i, ou bien (presque sûrement) il finira par quitter l'état i et n'y jamais plus revenir. Cela permet de relier la définition que l'on vient de donner d'un état transitoire à la remarque 27.19 page 560.

Notamment, si C et C′ sont deux classes récurrentes, aucune ne peut mener à l'autre. En effet, si C menait à C′, alors C′ ne mènerait pas à C (sinon ce serait la même classe), donc C serait transitoire.

Sur le schéma ci-dessous, on donne une allure générale que peut avoir le graphe d'une chaîne de Markov ; dans chaque classe, tous les états communiquent.



§ 404 **Matrice réduite**

Notons $\mathcal{T}$ l'ensemble de tous les états transitoires (cet ensemble peut être formé de plusieurs classes transitoires). On décompose ensuite l'espace des états en

$$\mathscr{E} = \mathcal{T} \cup \mathcal{R}_1 \cup \mathcal{R}_2 \cup \cdots \cup \mathcal{R}_k$$

où chaque classe $\mathcal{R}_1, \ldots, \mathcal{R}_k$ est une classe récurrente. On réordonne les états pour s'adapter à cette écriture ; la matrice de transition de la chaîne prend alors la forme générale

$$\mathbf{P}' = \left(\begin{array}{cccc|c} \multicolumn{4}{c|}{\multirow{4}{*}{}} & \\ & & & & \\ & & & & \\ & & & & \\ \hline & & & & \end{array} \right)$$

Donnons des noms aux matrices carrées sur la diagonale :

$$\mathbf{P}' = \begin{pmatrix} \boxed{\mathrm{T}} & & & & \\ & \boxed{\mathrm{R}_1} & & & \\ & & \boxed{\mathrm{R}_2} & & \\ & & & \ddots & \\ & & & & \boxed{\mathrm{R}_k} \end{pmatrix}$$

alors

- toutes les valeurs propres λ de T vérifient $|\lambda| < 1$;
- les matrices R_i admettent des valeurs propres de module 1 ;
- plus précisément, si R_i est ergodique, alors seul 1 est valeur propre de module 1 ;
- si R_i est de période d, alors en posant $\omega := \mathrm{e}^{2\mathrm{i}k\pi/d}$, les nombres $1, \omega, \ldots, \omega^{d-1}$ sont valeurs propres, .

27.4 MÉTHODE « UN PAS EN AVANT »

Un certain nombre de propriétés découlent de l'invariance dans le temps d'une chaîne de Markov : sachant $\{\mathrm{X}_1 = i\}$, la chaîne $(\mathrm{X}_n)_{n\geqslant 1}$ a les mêmes propriétés que la chaîne $(\mathrm{X}_n)_{n\geqslant 0}$ sachant $\{\mathrm{X}_0 = i\}$, et est indépendante de X_0.

Temps moyen de retour § 405

Notons μ_i le **temps moyen de retour à l'état** i :

$$\mu_i := \sum_{n=1}^{\infty} n\, \mathrm{R}^{(n)}(i,i)$$

et plus généralement

$$\mu_{i,j} := \sum_{n=1}^{\infty} n\, \mathrm{R}^{(n)}(i,j) = \mathbf{E}_i(\mathrm{T}_j)$$

où on rappelle que $\mathrm{T}_j = \inf\{n \geqslant 1\ ;\ \mathrm{X}_n = j\}$ est le **temps d'atteinte** de j. La quantité $\mu_{i,j}$ est le **temps d'atteinte moyen de j à partir de i**. Elle vérifie la relation interne

$$\mu_{i,j} = 1 + \sum_{k\neq j} \mathrm{P}(i,k)\, \mu_{k,j} \tag{27.4}$$

qui permet parfois, par des considérations de symétrie, d'accéder aux valeurs de ces temps. Si l'on note $\mathbf{\Lambda} = (\mu_{i,j})_{i,j}$, $\mathbf{D} = \mathrm{diag}(\mu_1, \ldots, \mu_d)$ et $\mathbf{J} = (1)$ la matrice carrée dont tous les coefficients valent 1, cette relation s'écrit $\mathbf{\Lambda} = \mathbf{J} + \mathbf{P}\mathbf{\Lambda} - \mathbf{P}\mathbf{D}$.

DÉMONSTRATION : Décomposons selon le système complet d'événements fourni par les valeurs de X_1 :

$$\mu_{i,j} = \mathbf{E}_i(\mathrm{T}_j) = \mathbf{E}_i(\mathrm{T}_j \,|\, \mathrm{X}_1 = j)\cdot \mathbf{P}_i\{\mathrm{X}_1 = j\} + \sum_{k\neq j} \mathbf{E}_i(\mathrm{T}_j \,|\, \mathrm{X}_1 = k)\cdot \mathbf{P}_i\{\mathrm{X}_1 = k\}$$

$$= 1\cdot \mathrm{P}(i,j) + \sum_{k\neq j}(1+\mu_{kj})\,\mathrm{P}(i,k) = 1 + \sum_{k\neq j} \mathrm{P}(i,k)\,\mu_{k,j}$$

où le « pas en avant » a permis d'écrire $\mathbf{E}_i(\mathrm{T}_j \,|\, \mathrm{X}_1 = k) = 1 + \mu_{k,j}$. ∎

§406 **Temps moyen d'atteinte d'une partie**

On peut généraliser ce raisonnement pour rechercher le temps nécessaire à rejoindre une partie A de $\mathscr{E}$.

Si i est en dehors de A, définissons le **temps d'atteinte de A depuis l'état i**

$$\mu_{i,\mathrm{A}} = \mathbf{E}_i(\mathrm{T_A}) \qquad \mathrm{T_A} = \inf\{n \geqslant 1\ ;\ \mathrm{X}_n \in \mathrm{A}\}\,.$$

Si i est élément de A, convenons de poser $\mu_{i,\mathrm{A}} = 0$. Ce temps d'atteinte vérifie la relation fermée, appelée **règle de la valeur moyenne** pour les temps d'atteinte :

↗27.4 ↗27.6

■ **Théorème 27.37 (Règle de la valeur moyenne pour les temps d'atteinte)** *Soit* A *une partie de* $\mathscr{E}$. *Alors*

$$\mu_{i,\mathrm{A}} = 1 + \sum_{k\in\mathscr{E}} \mathrm{P}(i,k)\,\mu_{k,\mathrm{A}}. \tag{27.5}$$

Le lecteur trouvera, chez Engel [30], de très nombreux exemples simples et profonds d'utilisation de cette relation.

§407 **Temps de retour et loi stationnaire**

Nous avons vu que, dans le cas d'une chaîne fortement irréductible, le théorème ergodique nous fournissait des conclusions fortes. Si la matrice est irréductible mais non ergodique, elle admet des valeurs propres de module 1, qui sont des racines de l'unité, et dont la suite des puissances ne converge pas mais converge, au sens des moyennes de Cesàro, vers 0 ; on a donc le résultat plus faible :

■ **Proposition 27.38** *La suite des moyennes de Cesàro de la suite* $(\mathbf{P}^n)_n$ *converge vers une matrice stochastique* $\mathbf{P}^\infty$, *vérifiant* $\mathbf{P}^\infty\mathbf{P} = \mathbf{P}\,\mathbf{P}^\infty = \mathbf{P}^\infty$ *et* $(\mathbf{P}^\infty)^2 = \mathbf{P}^\infty$.

Les coefficients de cette matrice $\mathbf{P}^\infty$ ont une interprétation simple. Notons

$$\Phi_n(j) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{\mathrm{X}_k = j\}}$$

la fréquence de séjour dans l'état j sur la durée $[\![1\,;\,n]\!]$. Par linéarité et en utilisant la relation $\mathbf{E}_i(\mathbf{1}_{\{\mathrm{X}_n=j\}}) = \mathbf{P}\{\mathrm{X}_n = j \,|\, \mathrm{X}_0 = i\} = \mathrm{P}^n(i,j)$, on déduit

$$\mathbf{E}_i\big(\Phi_n(j)\big) = \frac{1}{n}\big[\mathrm{P}(i,j) + \mathrm{P}^2(i,j) + \cdots + \mathrm{P}^n(i,j)\big] \xrightarrow[n\to\infty]{} \mathrm{P}^\infty(i,j).$$

Ainsi, les $\mathrm{P}^\infty(i,j)$ sont les **fréquences moyennes (asymptotiques) de séjour dans l'état j**, sachant qu'on est parti de $\mathrm{X}_0 = i$.

■ **Théorème 27.39 (Temps de retour pour une chaîne irréductible)** *Si* $\mathbf{P}$ *est irréductible, alors son unique loi stationnaire* $\boldsymbol{\pi}^*$ *vérifie*

$$\pi_j^* = \mathrm{P}^\infty(i,j) = \frac{1}{\mu_j}.$$

Ce théorème peut s'interpréter en disant que, *dans une chaîne irréductible, le temps moyen de retour à un état est l'inverse de la probabilité que le système à l'équilibre se trouve dans cette configuration.*

DÉMONSTRATION : Multiplions la relation $\mathbf{\Lambda} = \mathbf{J} + \mathbf{P}\mathbf{\Lambda} - \mathbf{P}\mathbf{D}$ par $\mathbf{P}^\infty$ à gauche ; avec la relation $\mathbf{P}^\infty\mathbf{P} = \mathbf{P}^\infty$, on obtient $\mathbf{P}^\infty\mathbf{\Lambda} = \mathbf{P}^\infty\mathbf{J} + \mathbf{P}^\infty\mathbf{\Lambda} - \mathbf{P}^\infty\mathbf{D}$; la relation $\mathbf{P}^\infty\mathbf{J} = \mathbf{J}$ permet de conclure à $\mathbf{P}^\infty\mathbf{D} = \mathbf{J}$, ce qui donne $\pi_j^*\mu_j = 1$. ∎

Probabilité d'absorption § 408

Si la chaîne de Markov possède plusieurs classes récurrentes, il est naturel de se demander, à partir d'un état i fixé, quelle est la probabilité d'absorption dans telle ou telle classe.

Soit B une partie absorbante de $\mathscr{E}$. Notons $a_{i,\mathrm{B}}$ la probabilité que le système, partant de i, finisse dans la partie B, c'est-à-dire (en utilisant le fait que B est absorbante)

$$a_{i,\mathrm{B}} = \mathbf{P}_i\left(\bigcup_{n=0}^{\infty} \{X_n \in \mathrm{B}\}\right).$$

Ces quantités peuvent souvent se calculer, de proche en proche, grâce à une règle de la valeur moyenne très semblable à celle concernant les temps moyens d'absorption, et par leurs conditions aux limites : ☞27.6

$$\begin{cases} a_{i,\mathrm{B}} = 1 & \text{si } i \in \mathrm{B}, \\ a_{i,\mathrm{B}} = 0 & \text{si } i \text{ est dans une partie absorbante disjointe de B.} \end{cases}$$

■ **Théorème 27.40 (Règle de la valeur moyenne pour la probabilité d'absorption)** *Soit* B *une partie absorbante de* $\mathscr{E}$. *Alors*

$$a_{i,\mathrm{B}} = \sum_{k \in \mathscr{E}} \mathrm{P}(i,k)\, a_{k,\mathrm{B}}$$

L'interprétation intuitive est simple : d'après la formule des probabilités totales, la probabilité d'être absorbé par B lorsque l'on est en i à un instant donné se calcule en sommant, sur tous les états k, la probabilité d'être en k à l'instant suivant multipliée par la probabilité d'absorption sachant que l'on part de k.

Remarque 27.41 (Problème de Dirichlet) Une fonction $i \mapsto a_i$ vérifiant la relation précédente est parfois appelée **fonction harmonique** sur $\mathscr{E}$ par analogie avec les fonctions harmoniques sur $\mathbb{R}^2$ ou $\mathbb{C}$, qui vérifient la propriété de la moyenne

$$f(a) = \frac{1}{2\pi}\int_0^{2\pi} f(a + r\,\mathrm{e}^{\mathrm{i}t})\,\mathrm{d}t.$$

(Le lien est plus facile à voir si l'on considère une marche aléatoire dans le carré $[\![1\,;n]\!]^2$, par exemple, où les transitions possibles se font, de manière équiprobable vers chaque voisin ; la matrice $\mathbf{I} - \mathbf{P}$ est la matrice du laplacien discret bien connu en analyse numérique.)

Déterminer la fonction $i \mapsto a_i$ ou le vecteur colonne $\mathbf{A}$ correspondant revient à résoudre $(\mathbf{I} - \mathbf{P})\mathbf{A} = 0$ avec des *conditions aux bords* (valeurs des a_i sur B et sur les autres parties absorbantes), ce que l'on appelle, toujours dans le même esprit, un **problème de Dirichlet**. Ce résultat est le pendant discret du théorème de Kakutani présenté au paragraphe § 449, page 630.

Remarque 27.42 (Marches aléatoires et théorie du potentiel) Considérons une marche aléatoire sur un carré $[\![1\,;n]\!]^2$ comme une chaîne de Markov à $d = n^2$ états ; les transitions possibles se font, de manière équiprobable vers chaque voisin, sauf pour les états au bord du carré, qui sont absorbants. La matrice $\mathbf{P}$ est alors symétrique, mais bien sûr fortement non irréductible (elle possède $4n - 2$ états absorbants).

Fixons un point de départ $(i_0, j_0) \in [\![2\,;n-1]\!]^2$; notons

$$\mathrm{G}(i,j) = \mathbf{E}_{(i_0,j_0)}\left(\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=(i,j)\}}\right) = \sum_{n=0}^{\infty} \mathrm{P}^n\big((i_0,j_0),(i,j)\big)$$

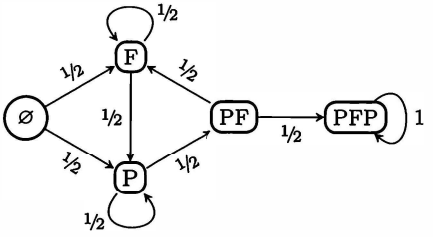
le nombre moyen de passage de la marche aléatoire issue de (i_0, j_0) (la marche finit presque sûrement par être absorbée par le bord et la série converge car l'état est transitoire). Alors la fonction G vérifie $\mathrm{G}\mathbf{P} = \mathrm{G} + \delta_{(i_0,j_0)}$, c'est-à-dire $\sum_{k,\ell} \mathrm{G}(k,\ell)\,\mathrm{P}\big((k,\ell),(i,j)\big) = \mathrm{G}(i,j) + \delta_{(i,j),(i_0,j_0)}$. Cette équation est l'analogue discret de l'équation de Poisson $\Delta \mathrm{G} = \delta$. On peut donc évaluer numériquement la fonction de potentiel G par méthode de Monte-Carlo, en tirant un grand nombre de marches aléatoires issues de (i_0, j_0) et en incrémentant un compteur en chaque point (i,j) du domaine, et ce à chaque passage du mobile.

27.5 TROIS EXEMPLES

§ 409 **Attente d'un motif dans le jeu de *pile* ou *face***

Reprenons le jeu auquel se livrent Rosencrantz et Guildenstern à l'exercice 9.2 page 220 et qui consiste à tirer indéfiniment à *pile* ou *face*, dans l'attente du motif « PFP ».

On peut modéliser la partie par la chaîne de Markov à cinq états suivante, où un état « vide » a été ajouté (essentiellement dans le but de faire apparaître un état de non-retour), et sa matrice de transition associée :

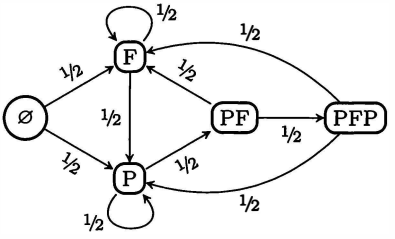


$$\begin{array}{c} \\ \varnothing \\ \mathrm{F} \\ \mathrm{P} \\ \mathrm{PF} \\ \mathrm{PFP} \end{array} \begin{array}{c} \begin{array}{ccccc} \varnothing & \mathrm{F} & \mathrm{P} & \mathrm{PF} & \mathrm{PFP} \end{array} \\ \begin{pmatrix} \cdot & 1/2 & 1/2 & \cdot & \cdot \\ \cdot & 1/2 & 1/2 & \cdot & \cdot \\ \cdot & \cdot & 1/2 & 1/2 & \cdot \\ \cdot & 1/2 & \cdot & \cdot & 1/2 \\ \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \end{array} = \mathbf{P}$$

Dans cette description, la partie commence à l'état « ∅ ».

- La classe $\{\varnothing\}$ est de non retour, donc transitoire (on la quitte au bout d'un tour), la classe $\{\mathrm{F}, \mathrm{P}, \mathrm{PF}\}$ est transitoire (on finit presque sûrement par en sortir en temps fini, pour n'y jamais retourner), et la classe $\{\mathrm{PFP}\}$ est absorbante, donc récurrente.
- La chaîne n'est donc pas irréductible (elle possède trois classes).
- Son unique mesure invariante est
$$\boldsymbol{\pi}^* = (0, 0, 0, 0, 1)$$
c'est-à-dire que la loi d'équilibre est évidemment lorsque le système est, avec une probabilité 1, dans l'état final PFP.
- 1 est valeur propre simple, toutes les autres étant de module strictement inférieur à 1.
- On voit que toutes les conclusions du théorème d'ergodicité ne s'appliquent pas, puisque $\boldsymbol{\pi}^*$ n'est pas strictement positif. En revanche, $\mathbf{P}^n$ converge vers la matrice $\mathbf{P}^\infty$ de coefficients $\mathrm{P}^\infty(i,j) = \pi^*(j)$.
- Par contre, on n'accède pas aux temps moyen d'atteinte de l'état PFP.

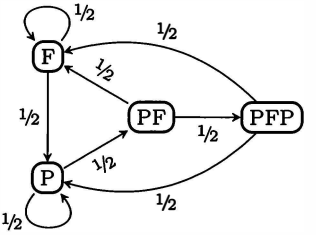
Il est aisé de remédier à ce problème. Tout d'abord, éliminons la classe absorbante $\{\mathrm{PFP}\}$ en modifiant le point de vue sur le jeu. Au lieu de s'arrêter dès que le motif PFP est apparu, considérons que l'apparition de ce motif est régénératif, c'est-à-dire qu'une fois qu'il est apparu, on reprend à zéro le décompte. Par exemple, dans la séquence « PPPFPFP », on considère que le motif cherché n'apparaît qu'une seule fois, et pas deux. On commence la partie à ∅ et on attend le prochain retour du motif. La chaîne et sa matrice sont désormais



$$\begin{array}{c} \\ \varnothing \\ \mathrm{F} \\ \mathrm{P} \\ \mathrm{PF} \\ \mathrm{PFP} \end{array} \begin{array}{c} \begin{array}{ccccc} \varnothing & \mathrm{F} & \mathrm{P} & \mathrm{PF} & \mathrm{PFP} \end{array} \\ \begin{pmatrix} \cdot & 1/2 & 1/2 & \cdot & \cdot \\ \cdot & 1/2 & 1/2 & \cdot & \cdot \\ \cdot & \cdot & 1/2 & 1/2 & \cdot \\ \cdot & 1/2 & \cdot & \cdot & 1/2 \\ \cdot & 1/2 & 1/2 & \cdot & \cdot \end{pmatrix} \end{array} = \mathbf{P}_2$$

- Cette fois-ci, la chaîne ne comporte plus que deux classes : la classe $\{\varnothing\}$ qui est toujours de non retour, et la classe $\{\mathrm{F}, \mathrm{P}, \mathrm{PF}, \mathrm{PFP}\}$ qui est la seule classe récurrente.
- La chaîne n'est toujours pas irréductible ; cependant, la classe de non retour, on le devine, ne joue aucun rôle dans la dynamique du processus.
- Cela se traduit par le fait que l'unique mesure invariante est strictement positive si on se limite aux états essentiels :
$$\boldsymbol{\pi}_2^* = \tfrac{1}{10}(0, 3, 4, 2, 1).$$
- La suite de terme général $\mathbf{P}_2^n$ converge vers la matrice
$$\mathbf{P}_2^\infty = \frac{1}{10}\begin{pmatrix} 0 & 3 & 4 & 2 & 1 \\ 0 & 3 & 4 & 2 & 1 \\ 0 & 3 & 4 & 2 & 1 \\ 0 & 3 & 4 & 2 & 1 \\ 0 & 3 & 4 & 2 & 1 \end{pmatrix}$$

Pour être certain de pouvoir appliquer le théorème 27.39 sur les temps de retour, on va encore modifier la chaîne de façon à la rendre irréductible. Pour cela, on fait disparaître la classe inessentielle $\{\varnothing\}$ et l'on fait débuter le jeu à l'état $X_0 = PFP$, qui remet les compteurs à zéro, et qui nous assure que la dynamique est bien la même que dans le cas précédent. Le graphe et la matrice de transition sont maintenant

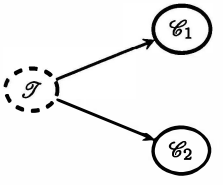


$$\begin{array}{c} \\ F \\ P \\ PF \\ PFP \end{array} \begin{array}{c} \begin{array}{cccc} F & P & PF & PFP \end{array} \\ \begin{pmatrix} 1/2 & 1/2 & \cdot & \cdot \\ \cdot & 1/2 & 1/2 & \cdot \\ 1/2 & \cdot & \cdot & 1/2 \\ 1/2 & 1/2 & \cdot & \cdot \end{pmatrix} \end{array} = \mathbf{P}_3$$

- Cette fois-ci, $\mathbf{P}_3$ est irréductible, la chaîne ne possède qu'une seule classe (récurrente, par la force des choses...) ; elle est de plus apériodique (certains états bouclent sur eux-mêmes). Elle est donc ergodique.
- Son unique loi invariante est
$$\boldsymbol{\pi}_3{}^* = \tfrac{1}{10}(3,4,2,1)$$
et, si $\boldsymbol{\pi}_0$ est une loi initiale quelconque, $\boldsymbol{\pi}_n = \mathbf{P}_3{}^n \boldsymbol{\pi}_0 \to \boldsymbol{\pi}_3^*$.
- La loi $\boldsymbol{\pi}_3^*$ est bien à valeurs strictement positives.
- Le temps moyen de retour à l'état PFP, qui est donc également le temps moyen d'attente du motif PFP à partir du début de la partie, est donc
$$\mu_4 = \frac{1}{\pi_3(4)} = 10.$$
On retrouve le résultat de l'exercice 9.2 page 220.

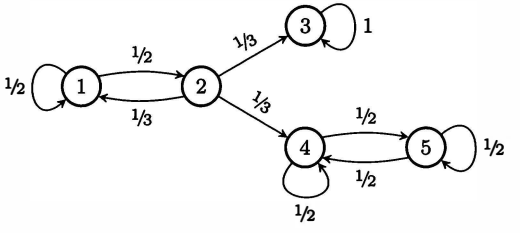
Une chaîne réductible § 410

Lorsqu'une chaîne possède deux classes récurrentes distinctes, aucune ne peut mener à l'autre. Elles peuvent en revanche être atteintes depuis une classe transitoire, selon le schéma global suivant :



$$\mathbf{P} = \begin{pmatrix} \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \end{pmatrix}$$

Considérons par exemple la chaîne et la matrice suivantes :



$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 & \cdot & \cdot & \cdot \\ 1/3 & \cdot & 1/3 & 1/3 & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1/2 & 1/2 \\ \cdot & \cdot & \cdot & 1/2 & 1/2 \end{pmatrix}$$

- La classe $\mathscr{T} := \{1, 2\}$ est transitoire : on en sort presque sûrement après un nombre fini d'étapes.
- En revanche, en sortant de cette classe, on peut pénétrer dans l'une ou l'autre des deux classes absorbantes
$$\mathscr{C}_1 := \{3\} \qquad \text{et} \qquad \mathscr{C}_2 := \{4, 5\}.$$
- Ainsi, en fonction de la loi initiale, la suite de terme général $\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 \mathbf{P}^n$ pourra converger vers des valeurs distinctes.
- Cela se traduit par le fait que le sous-espace vectoriel (à gauche) $E_g(1, \mathbf{P})$ associé à la valeur propre 1 est de dimension 2, et engendré par les vecteurs
$$\boldsymbol{\pi}_1^* = (0,0,1,0,0) \qquad \text{et} \qquad \boldsymbol{\pi}_2^* = (0,0,0,\tfrac{1}{2},\tfrac{1}{2}).$$
Toute combinaison linéaire convexe de ces lois est encore une loi invariante.

- Si $\pi_0 = (0,0,1,0,0)$, alors $\pi_n \to \pi_1^*$. Si en revanche $\pi_0 = (0,0,0,x,1-x)$, alors $\pi_n \to \pi_2^*$.
- La suite de terme général $\mathbf{P}^n$ est convergente, de limite

$$\mathbf{P}^\infty = \begin{pmatrix} 0 & 0 & 1/2 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 1/4 & 1/4 \end{pmatrix}$$

§ 411 **Le modèle des urnes d'Ehrenfest**

La mathématisation de la dynamique moléculaire des gaz, et notamment le problème de l'asymétrie temporelle des lois macroscopiques de la thermodynamique, censées résulter de lois microscopiques fondamentalement réversibles, a donné lieu entre autres à un modèle célèbre, proposé en 1907 par Paul et Tatiana EHRENFEST.

Deux urnes A et B contiennent au total N particules. On répète, à intervalles de temps réguliers, l'opération : *tirer au hasard (uniformément) une des* N *particules, et la changer d'urne.*

Ce modèle, bien que simpliste, illustre la dynamique d'un gaz de particules dans une boîte formée de deux compartiments (urnes) communicants :

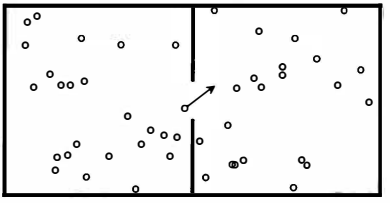


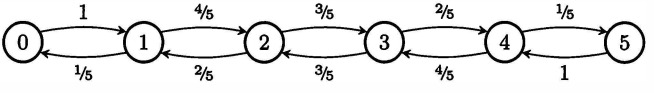
FIGURE 27.2 — *Urnes d'Ehrenfest : passage de* $X_n = k$ *à* $X_{n+1} = k-1$.

Au lieu d'une description microscopique de ce système (où est située chacune des particules ?), adoptons un point de vue macroscopique (combien y a-t-il de particules dans chaque urne ?).

On note X_n le nombre de particules dans le compartiment A à l'instant n. En description macroscopique, le processus revient donc à passer de $X_n = i$ à $X_{n+1} = i+1$ avec la probabilité $1 - i/N$ et à $X_{n+1} = i-1$ avec la probabilité i/N. Ainsi, le modèle d'Ehrenfest est décrit par une chaîne de Markov à $N+1$ états, dont l'espace des états est $\mathscr{E} = \{0,1,2,\ldots,N-1,N\}$, et dont la matrice de transition a les coefficients

$$P(i,i+1) = \frac{N-i}{N} \qquad P(i,i-1) = \frac{i}{N},$$

tous les autres coefficients étant nuls. Par exemple, pour $N = 5$ particules, le graphe est



Cette chaîne est irréductible (tous les états communiquent évidemment) : tous les états sont récurrents. Cependant, elle est n'est pas ergodique, car elle est périodique de période 2. On peut montrer que 1 et -1 sont valeurs propres simples de P. Le théorème de Perron-Frobenius s'applique, mais pas le théorème ergodique. (Il serait d'ailleurs facile de remédier à ce point en autorisant, même avec une probabilité très faible, les particules à ne pas changer d'urne, c'est-à-dire de considérer la matrice $\mathbf{P}' = \alpha\mathbf{I} + (1-\alpha)\,\mathbf{P}$.)

Une première implication de l'irréductibilité est la suivante. On part d'une configuration initiale où *toutes* les particules sont dans le compartiment A, c'est-à-dire $\pi_0 = (0,0,\ldots,0,1)$. Dès que l'on met le processus en route, les particules vont commencer à se répartir à peu près uniformément entre les compartiments. Cependant, presque sûrement, le système reviendra, une infinité de fois, dans cette configuration où toutes les particules sont du côté A. La contradiction apparente avec les lois de la thermodynamique (l'entropie redevient anormalement basse lorsqu'on retourne à cette configuration) est en partie levée par le calcul du temps moyen de retour à l'aide du théorème 27.39 et qui, comme le calcul suivant le justifiera, vaut :

$$\mu_N = \frac{1}{\pi^*(N)} = 2^N.$$

Or, en thermodynamique, on a fréquemment affaire à des nombres de l'ordre de grandeur de $N = 6 \cdot 10^{23}$ (nombre d'Avogadro) ; quelle que soit la rapidité des échanges microscopiques, ce temps de récurrence dépasse *de très loin* l'âge de l'Univers[5].

Quant à la loi invariante, quelques lignes de calcul sur les premiers termes suivies d'une récurrence permettent de montrer qu'elle vaut

$$\pi^*(i) = \frac{1}{2^N}\binom{N}{i},$$

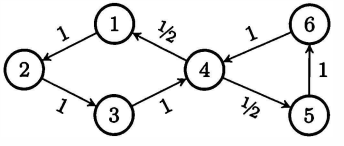
en quoi l'on reconnaît une distribution binomiale $\mathscr{B}(N, \frac{1}{2})$; cette distribution a une moyenne $N/2$ (donc de l'ordre de 10^{23}) et un écart-type de $\sqrt{N}/2$: elle est donc extrêmement piquée, puisque sa largeur est trois cents milliards de fois inférieure à sa valeur moyenne!

EXERCICES

♦ ***Exercice 27.1 (Urnes d'Ehrenfest)*** Avec $N = 6 \cdot 10^{23}$ (soit environ un gramme d'hydrogène), quelle est la probabilité que le nombre de particules dans une urne dépasse d'au moins 10^{12} la valeur $N/2$? On pourra utiliser une approximation normale de la loi binomiale et employer l'équivalent (A.2) (voir page 676) de la fonction de survie de la loi normale.

♦ ***Exercice 27.2*** Montrer qu'il n'existe aucune matrice stochastique 2×2 semblable au bloc de Jordan $J = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Montrer plus généralement qu'une matrice stochastique ne peut pas être semblable à une matrice triangulaire par blocs, dont un bloc serait de la forme $K = \begin{pmatrix} e^{i\theta} & 1 \\ 0 & e^{i\theta} \end{pmatrix}$; autrement dit, il ne peut pas exister de valeur propre de module 1 pour laquelle le sous-espace caractéristique soit strictement plus grand que le sous-espace propre.

♦ ***Exercice 27.3 (Plus petite puissance strictement positive d'une matrice ergodique)*** On considère la chaîne de Markov suivante :

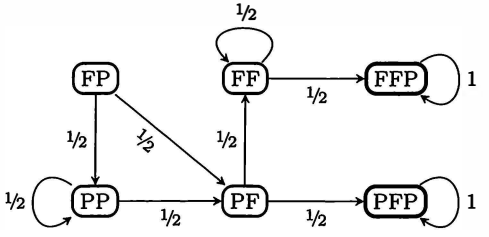


$$\mathbf{P} = \begin{pmatrix} \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ 1/2 & \cdot & \cdot & \cdot & 1/2 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \end{pmatrix}$$

Est-elle irréductible ? Périodique ? Ergodique ? Si elle est ergodique, quelle est la plus petite puissance n telle que $\mathbf{P}^n > 0$?

♦ ***Exercice 27.4 (Paradoxe de Walter Penney)*** On s'intéresse ici à une série infinie de tirages à pile ou face, en s'arrêtant dès qu'on rencontre la séquence PFP ou bien la séquence FFP. Pour un tirage aléatoire de 3 pièces, chaque séquence a la même probabilité de présence ($p = 1/8$), mais l'arrivée de la *première* occurrence n'est pas équiprobable (voir § 468, page 661).

On représente la partie sous la forme d'une chaîne de Markov. À l'issue des deux premiers tirages, on peut se trouver dans l'un des quatre états FP, PP, PF et FF. On note FFP et PFP les états de gain de chacune des séquences, considérés comme absorbants.



$$\mathbf{P} = \begin{array}{c} \begin{matrix} \text{FP} & \text{PP} & \text{PF} & \text{FF} & \text{FFP} & \text{PFP} \end{matrix} \\ \begin{pmatrix} \cdot & 1/2 & 1/2 & \cdot & \cdot & \cdot \\ \cdot & 1/2 & 1/2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1/2 & \cdot & 1/2 \\ \cdot & \cdot & \cdot & 1/2 & 1/2 & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{pmatrix} \end{array} \begin{matrix} \text{FP} \\ \text{PP} \\ \text{PF} \\ \text{FF} \\ \text{FFP} \\ \text{PFP} \end{matrix}$$

[5]. L'âge de l'Univers s'écrit, en secondes, avec dix-huit chiffres ; le temps de récurrence s'écrirait avec cent quatre-vingt mille milliards de milliards de chiffres... Même en considérant que le rapport entre le temps physique et le « temps de la chaîne de Markov » est proportionnel à N, le temps physique de récurrence est à peine réduit.

On numérote les états dans l'ordre FP, PP, PF, FF, FFP, PFP.

i) Identifier les différentes classes de ce graphe.

ii) Que vaut π_0 ? Écrire la matrice de transition.

iii) Y a-t-il unicité de la loi invariante de cette chaîne ?

iv) On se propose de déterminer la limite de π_n.

(a) On va montrer que la suite de terme général $\mathbf{P}^n$ converge vers une matrice $\mathbf{P}^\infty$. Tout d'abord, que va-t-il advenir du système s'il est initialement dans l'état PFP ou FFP ?

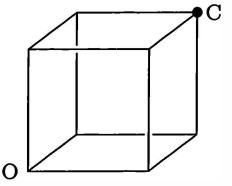
(b) Répondre de même lorsque le système est initialement en FF, puis en PF, puis en FP ou PP.

(c) En déduire que $\mathbf{P}^n$ converge vers une matrice $\mathbf{P}^\infty$ que l'on précisera.

v) En déduire la probabilité que le motif FFP arrive avant le motif PFP.

vi) En utilisant la règle de la moyenne (27.5) page 570, calculer le temps moyen de la partie avant qu'un vainqueur soit déclaré.

♦ ***Exercice 27.5 (Chaîne avec barrière réfléchissante)*** On reprend l'exemple de la chenille se promenant sur les arêtes d'un cube (exercice 1.6 page 50). Son point de départ est O. Le point C est enduit de colle forte. À chaque étape, la chenille choisit, de façon équiprobable, une arête parmi les trois qu'elle a à disposition, indépendamment de son passé. Quel est le temps moyen de trajet de la chenille ? *(On pourra modéliser le problème par une chaîne de Markov puis utiliser la règle de la moyenne.)*



♦ ***Exercice 27.6 (Ruine d'un joueur)*** Rosencrantz et Guildenstern jouent à *pile* ou *face*. Au début de la partie, Rosencrantz possède 30 pièces d'or, tandis que Guildenstern en possède 70. Ils lancent une pièce équilibrée et parient une pièce d'or à chaque lancer.

i) Quelle est la durée moyenne de la partie avant que l'un des deux joueurs ne se retrouve ruiné ?

ii) Quelle est la probabilité que Rosencrantz gagne ?

♦ ***Exercice 27.7 (Génétique végétale)*** Une plante possède un gène pouvant s'exprimer sous les formes A et a, appelées *allèles*. Ce gène apparaît par paire dans son génome, c'est-à-dire que chaque individu est caractérisé par l'une des trois formes AA, aa ou aA (identique à Aa). Lors de l'auto-reproduction de la plante, les paires se séparent durant la méiose, chaque moitié se combine avec une moitié d'une seconde cellule du même type. Ainsi, un individu AA ne donne naissance qu'à des AA, tandis qu'un aA peut donner naissance à des aa, des aA et des AA.

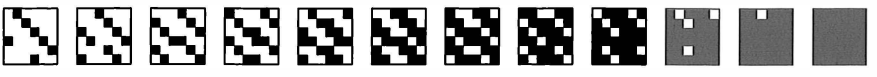
Modéliser le passage d'un individu de la génération n à un individu de la génération $n+1$ sous la forme d'une chaîne de Markov. Quel est le comportement asymptotique de la suite des lois instantanées ?

SOLUTIONS

♦ **Solution 27.2** La suite $(\mathrm{J}^n)_n$ est non bornée tandis que la suite $(\mathbf{P}^n)_n$ l'est (quelle que soit la norme choisie). De même, pour la suite $(\mathrm{K}^n)_n$, puisque

$$\mathrm{K}^n = \begin{pmatrix} \mathrm{e}^{in\theta} & n\,\mathrm{e}^{i(n-1)\theta} \\ 0 & \mathrm{e}^{in\theta} \end{pmatrix}.$$

♦ **Solution 27.3** En raisonnant sur la chaîne, on devine que la puissance minimale est $4 \wedge 3 = 12$, ce que confirme un calcul numérique des puissances successives de $\mathbf{P}$, dont voici les silhouettes :



♦ **Solution 27.4**

i) {FP} et {PF} sont des classes de non retour, {PP} et {FF} des classes transitoires, tandis que {FFP} et {PFP} sont des classes absorbantes.

ii) Après les deux premiers tirages, on se trouve avec une probabilité uniforme dans chacun des quatre états FP, PP, PF ou FP, donc

$$\boldsymbol{\pi}_0 = (\tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, 0, 0).$$

iii) Il n'y a pas unicité de la loi invariante puisqu'il existe deux classes récurrentes. Un peu de pivot de Gauss (ou l'utilisation d'un logiciel de calcul formel) montre que le sous-espace propre associé à 1 est de dimension 2, et engendré par les lois invariantes suivantes :

$$\boldsymbol{\pi}_1^* = (0, 0, 0, 0, 1, 0) \qquad \text{et} \qquad \boldsymbol{\pi}_2^* = (0, 0, 0, 0, 0, 1).$$

iv) (a) Si le système est dans un des états PFP ou FFP, c'est-à-dire si $\boldsymbol{\pi}_0 = (0, 0, 0, 0, 1, 0)$ ou $(0, 0, 0, 0, 0, 1)$, il y reste indéfiniment. Ainsi,

$$(0, 0, 0, 0, 1, 0) \cdot \mathbf{P}^n \xrightarrow[n\to\infty]{} (0, 0, 0, 0, 1, 0)$$

ce qui montre que l'avant-dernière ligne de $\mathbf{P}^n$ converge vers $(0, 0, 0, 0, 1, 0)$. De même, la dernière ligne de $\mathbf{P}^n$ converge vers $(0, 0, 0, 0, 0, 1)$.

(b) Si le système est en FF, alors après avoir hésité quelques temps, il finira presque sûrement en FFP (le temps nécessaire suivant une loi géométrique). De même, si le système est en PF, il ira au coup suivant en PFP (probabilité 1/2) ou en FF, auquel cas il finira en FFP (probabilité 1/2).

Enfin, s'il est en PP, il ira presque sûrement en PF, et finira donc dans un des deux états finaux avec une probabilité 1/2 chacun ; s'il est en FP de même.

Il en résulte que

$$(0, 0, 0, 1, 0, 0) \cdot \mathbf{P}^n \to (0, 0, 0, 0, 1, 0) \qquad (0, 0, 1, 0, 0, 0) \cdot \mathbf{P}^n \to (0, 0, 0, 0, \tfrac{1}{2}, \tfrac{1}{2})$$
$$(0, 1, 0, 0, 0, 0) \cdot \mathbf{P}^n \to (0, 0, 0, 0, \tfrac{1}{2}, \tfrac{1}{2}) \qquad (1, 0, 0, 0, 0, 0) \cdot \mathbf{P}^n \to (0, 0, 0, 0, \tfrac{1}{2}, \tfrac{1}{2})$$

(c) Cela permet de conclure que

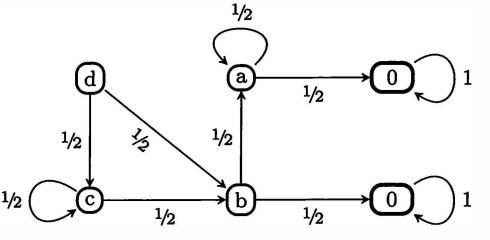
$$\mathbf{P}^n \xrightarrow[n\to\infty]{} \mathbf{P}^\infty := \begin{pmatrix} 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

v) On en déduit enfin

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 \, \mathbf{P}^n \xrightarrow[n\to\infty]{} \boldsymbol{\pi}_0 \, \mathbf{P}^\infty = (0, 0, 0, 0, \tfrac{5}{8}, \tfrac{3}{8}),$$

c'est-à-dire que le système termine dans l'état FFP avec la probabilité 5/8 = 0,625, et dans l'état PFP avec la probabilité 3/8 = 0,375.

vi) Donnons des noms aux temps moyens d'atteinte de la partie A = {PFP, FFP} (deux d'entre eux étant déjà nuls...) :



La règle de la moyenne donne successivement

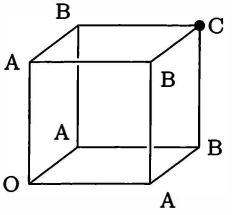
$$a = 1 + \frac{a}{2} \qquad b = 1 + \frac{a}{2} \qquad c = 1 + \frac{c}{2} + \frac{b}{2} \qquad d = 1 + \frac{b}{2} + \frac{c}{2}$$

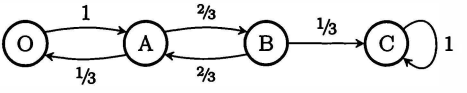
d'où $(a, b, c, d) = (2, 2, 4, 4)$. À partir du début de la partie, il faut 2 coups avant d'arriver à l'un des états FP, FF, PF ou PP, le temps moyen d'atteinte d'un état final est donc

$$m = 2 + \frac{a + b + c + d}{4} = 5.$$

♦ **Solution 27.5** Le graphe du cube représente très bien la chaîne de Markov, si l'on admet que chaque arête porte un probabilité 1/3 dans les deux sens, à l'exception des arêtes issues de B (voir graphe ci-dessous), qui mènent à C avec une probabilité 1/3 mais n'en repartent pas. Le point C boucle sur lui-même avec une probabilité 1.

On peut également utiliser les symétries du système pour se ramener à un graphe plus simple : les points sont identifiés à un état A s'ils sont à distance 1 de O, et à un état B s'ils sont à distance 1 de C. Cela permet de donner un graphe réduit :



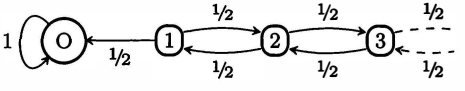


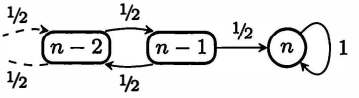
Notons z, a, b et c les temps moyens d'atteinte de l'état C depuis les points O, A, B et C. On a bien sûr $c = 0$. De plus la règle de la moyenne donne les équations

$$b = 1 + \tfrac{2}{3}\,a \qquad a = 1 + \tfrac{2}{3}b + \tfrac{1}{3}z \qquad z = 1 + a$$

ce qui, après résolution, donne $a = 9$, $b = 7$ et $z = 10$. Le temps moyen de trajet de la chenille est donc de 10 étapes.

♦ **Solution 27.6** On modélise la partie par une chaîne de Markov à $n = 100$ états, chacun étant étiqueté par la somme que possède Rosencrantz en cours de partie. Le graphe associé est donc de la forme





La partie s'arrête lors de la ruine d'un des joueurs, c'est-à-dire à l'un des états (absorbants) 0 (ruine de Rosencrantz) et 100 (ruine de Guildenstern).

i) Pour chaque état i, notons m_i le temps moyen d'absorption par $\{0, n\}$. On a tout d'abord $m_0 = m_n = 0$; quant aux autres coefficients, ils vérifient une relation de récurrence qui découle de la règle de la moyenne :

$$\forall i \in [\![1\,;\,n]\!] \qquad m_i = 1 + \tfrac{1}{2}\,[m_{i-1} + m_{i+1}]\,.$$

ou encore

$$\forall i \in [\![1\,;\,n]\!] \qquad m_{i+1} - 2m_i + m_{i-1} = -2.$$

On peut reconnaître la version discrétisée de la dérivée seconde, par exemple en écrivant

$$\forall i \in [\![1\,;\,n]\!] \qquad (m_{i+1} - m_i) - (m_i - m_{i-1}) = -2.$$

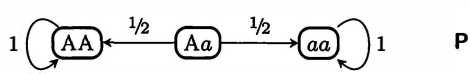
ou encore $\delta_i - \delta_{i-1} = -2$, donc $\delta_i = \alpha - 2i$, ce qui donne m de la forme $-i^2 + \alpha i + \beta$. Avec les conditions aux bords, on trouve $m_i = i(n - i)$.
Ainsi, la durée moyenne de la partie est $m_{30} = 30 \cdot 70 = 2\,100$.

ii) Pour chaque état i, notons a_i la probabilité d'absorption par l'état n (gain de Rosencrantz). On a $a_0 = 0$ et $a_n = 1$. Cette fois, la relation de récurrence découlant de la règle de la moyenne pour les probabilités d'absorption est

$$\forall i \in \mathscr{E} \qquad a_i = \tfrac{1}{2}\,[a_{i+1} + a_i]\,.$$

Une technique similaire à la question précédente montre que les $a_{i+1} - a_i$ sont constants, donc on cherche a_i sous la forme $a_i = \alpha\, i + \beta$; avec les conditions aux bords, $a_i = i/n$.
Ainsi, la probabilité de gain de Rosencrantz est $a_{30} = 0{,}3$, soit 30 % de chances de gagner.

♦ **Solution 27.7**



$$\mathbf{P} = \begin{pmatrix} 1 & \cdot & \cdot \\ 1/4 & 1/2 & 1/4 \\ \cdot & \cdot & 1 \end{pmatrix} \qquad \mathbf{P}^{\infty} = \begin{pmatrix} 1 & \cdot \\ 1/2 & 1/2 \\ \cdot & 1 \end{pmatrix}.$$

Si l'on part d'une loi $\boldsymbol{\pi}_0 = (\alpha, 1 - \alpha - \beta, \beta)$, la suite $(\boldsymbol{\pi}_n)_n$ converge vers la loi invariante $\boldsymbol{\pi}^*_{\alpha,\beta} = \frac{1}{2}(1+\alpha-\beta, 0, 1-\alpha+\beta)$. Autrement dit, les individus porteurs de la paire Aa disparaissent rapidement dans ce mode de reproduction non sexué.

Chapitre 28

Arithmétique et probabilités : une randonnée

Par Emmanuel Kowalski

La théorie des probabilités est basée sur un certain nombre d'idées et de concepts abstraits qui ont été modélisés par Kolmogorov en particulier. Depuis lors, il est naturel de s'intéresser à ces idées indépendamment de toute réalisation concrète, en s'appuyant sur des résultats généraux qui établissent, a priori, *que les hypothèses que l'on peut faire (comme l'existence de suites* (X_n) *de variables aléatoires indépendantes dont les lois sont données à priori) ne sont pas contradictoires.*

Dans ce chapitre, le point de vue est un peu différent : nous allons considérer des objets parfaitement concrets et, apparemment, « déterministes », et nous allons pourtant expliquer qu'ils peuvent incarner la plupart des constructions probabilistes importantes. Ces objets seront, tout simplement, les entiers naturels strictement positifs, qui sont l'objet de la théorie des nombres.

On peut aussi voir cette approche comme parallèle à celle des simulations à la Monte-Carlo : ici, on va faire apparaître des lois de probabilités usuelles (comme la loi normale, ou les lois de Poisson, et l'on pourrait donner beaucoup d'autres exemples) ou inhabituelles (la mesure qui contrôle l'écart entre valeurs propres entre matrices unitaires de grande taille) non pas à l'aide de processus physiques que l'on peut aisément (!) créer dans le monde réel, mais à l'aide de la suite des entiers naturels.

Du point de vue des mathématiques pures, ce chapitre est aussi une illustration de l'intérêt des idées et des raisonnements probabilistes, et surtout de la convergence en loi, qui — cachée souvent sous le nom « équirépartition » — est l'une des idées les plus utiles qui soit...

28.1 EXEMPLE ÉLÉMENTAIRE

Un exemple clarifiera bien ce que nous souhaitons expliquer. À la question « Quelle est la probabilité qu'un entier naturel soit pair ? », la réponse naturelle qui vient aux lèvres est « 1/2 ». Mais cette réponse n'a pas de sens du point de vue mathématique si l'on considère les probabilités telles que Kolmogorov les a formalisées : il n'existe pas de mesure sur l'ensemble des entiers naturels qui donne la même probabilité à chaque entier, et qui permette de parler de la probabilité de l'ensemble des entiers pairs.

Il y a cependant une façon de poser la question différemment qui permet d'obtenir la réponse demandée. Comme cela arrive souvent, ce changement de point de vue ouvre de nombreuses perspectives.

Considérons, plutôt que tous les entiers, l'ensemble fini

$$\Omega_N = \{1, \dots, N\}$$

pour tout entier $N \geqslant 1$. On peut munir Ω_N de la mesure de probabilité uniforme $\mathbf{P}_N$ telle que

$$\mathbf{P}_N\{n\} = \frac{1}{N}$$

pour tout entier tel que $1 \leqslant n \leqslant N$. Pour cette mesure $\mathbf{P}_N$, la probabilité que n soit pair est

$$\mathbf{P}_N\{n \text{ est pair}\} = \begin{cases} \frac{1}{2} & \text{si N est pair,} \\ \frac{N-1}{2N} = \frac{1}{2} - \frac{1}{2N} & \text{si N est impair.} \end{cases}$$

On voit que la limite quand $N \to +\infty$ de cette probabilité existe, et qu'elle vaut 1/2. C'est une *première* manière de donner un sens à la probabilité qu'un entier soit pair. Mais on peut faire mieux. En effet, si la probabilité qu'un entier n soit pair est légitime, il en est de même de la probabilité qu'il soit impair, ou que le chiffre des unités en base 10 soit 1, ou 2, ..., 9, ou toute autre question concernant le reste de la division euclidienne de n après division par un entier q fixé ($q = 2$ correspondant à la parité de n, et $q = 10$ au chiffre des unités).

On considère les mesures ci-dessus comme définies sur tout $\mathbb{N}$, mais supportées sur $\{1, \dots, N\}$: $\mathbf{P}_N(\{n\}) = 0$ si $n > N$. On remarque que cette suite de mesures de probabilité sur $\mathbb{N}$ ne converge pas en loi vers une mesure de probabilité. Par contre, étant donné l'entier $q \geqslant 1$, considérons l'application

$$r_q : \mathbb{N} \longrightarrow \{0, \dots, q-1\}$$

qui envoie un entier n vers le reste de la division euclidienne de n par q. Par exemple, $r_2(n) = 0$ si et seulement si n est pair, $r_{10}(n) = 5$ si et seulement si le chiffre des unités de n est 5, etc.

D'après les principes généraux, chaque mesure $\mathbf{P}_N$ induit une mesure image sur l'ensemble fini $\{0, \dots, q-1\}$, que nous noterons $\mathbf{P}_N^{(q)}$. Par définition, on a

$$\mathbf{P}_N^{(q)}\{i\} = \frac{1}{N}(\text{nombre d'entiers } 1 \leqslant n \leqslant N \text{ tels que } r_q(n) = i),$$

pour tout $i \in \{0, \dots, q-1\}$.

Alors on a le résultat suivant : pour tout $q \geqslant 1$, la suite de mesures de probabilités $(\mathbf{P}_N^{(q)})_N$ converge en loi quand $N \to +\infty$ vers la loi de probabilité uniforme sur l'ensemble fini $\{0, \dots, q-1\}$.

Cela permet donc de donner un sens raisonnable à toute assertion du type « La probabilité qu'un entier pris au hasard soit congru à a modulo q est $1/q$ », ou bien : pour tout sous-ensemble $A \subset \mathbb{Z}/q\mathbb{Z}$, « la probabilité que le reste de la division euclidienne de n par q soit dans A est $|A|/q$. »

Exemple 28.1 Les deux exemples ci-dessous sont très classiques.

- Sur Ω_N, notons s_N la variable aléatoire définie par

$$s_N(n) = \begin{cases} 1 & \text{si } n \text{ n'est pas divisible par le } \textit{carré} \text{ d'un nombre premier,} \\ 0 & \text{sinon.} \end{cases}$$

(Par exemple, $s_N(4) = 0$ si $N \geqslant 4$, $s_N(p) = 1$ si p est premier et $N \geqslant p$, $s_{1000}(486) = 0$ car 9 divise 486, etc.)
Alors on peut montrer que $(s_N)_N$ converge en loi vers une loi de Bernoulli B_1 avec probabilité de succès

$$\mathbf{P}\{B_1 = 1\} = \frac{6}{\pi^2}$$

(on peut donc calculer π en factorisant des entiers au hasard et en regardant s'ils sont divisibles par un carré d'un nombre premier...)

- Sur $\Omega_N \times \Omega_N$, notons c_N la variable aléatoire définie par

$$c_N(n_1, n_2) = \begin{cases} 1 & \text{si } n_1 \text{ et } n_2 \text{ sont premiers entre eux,} \\ 0 & \text{sinon.} \end{cases}$$

Là aussi (!), on montre que $(c_N)_N$ converge en loi... vers la même loi de Bernoulli B_1 ! (Numériquement, si $N = 1000$, on trouve que $\mathbf{P}\{c_N = 1\} = 0,608383$, alors que $6/\pi^2 \approx 0,60792710...$).

28.2 LE THÉORÈME D'ERDŐS-KAC ET LA LOI NORMALE

Sur le modèle de l'exemple précédent, nous considérons maintenant le comportement statistique d'autres fonctions définies sur les entiers que les restes de divisions euclidiennes. Il est frappant que la plupart des lois de probabilités usuelles peuvent apparaître de cette manière.

L'exemple le plus célèbre est fourni par le théorème d'Erdős-Kac qui fait apparaître la loi normale à partir du nombre de facteurs premiers des entiers. Précisément, pour tout entier $n \geqslant 1$, on note $\omega(n)$ le nombre de nombres premiers distincts divisant n, de sorte que $\omega(1) = 0$, et $\omega(n) = 1$ si et seulement si n est une puissance p^k d'un nombre premier (avec $k \geqslant 1$), et par exemple

$$\omega(30) = \omega(2 \cdot 3 \cdot 5) = 3 = \omega(120).$$

On a alors le résultat suivant, qui montre que la loi normale peut apparaître très simplement :

■ **Théorème 28.2 (Erdős-Kac)** *Soit ω_N la variable aléatoire définie sur Ω_N par $\omega_N(n) = \omega(n)$ pour $1 \leqslant n \leqslant N$. Pour $N \to +\infty$, la suite de variables aléatoires*

$$\frac{\omega_N - \ln\ln N}{\sqrt{\ln\ln N}}$$

converge en loi vers la loi normale centrée de variance 1.

Il existe de nombreuses démonstrations de ce théorème (et de nombreuses généralisations !), dont certaines très récentes (Granville et Soundararajan 2006 [42], Harper 2009 [46]). Bien qu'aucune ne soit complètement élémentaire — il faut connaître quelques propriétés fondamentales concernant la répartition des nombres

premiers, qui ne sont nullement évidentes —, on peut justifier heuristiquement le résultat de manière convaincante, de la manière suivante.

Pour tout nombre premier p, soit χ_p la fonction définie pour tous les entiers $n \geqslant 1$ par

$$\chi_p(n) = \begin{cases} 1 & \text{si } p \mid n, \text{ c'est-à-dire, si } p \text{ divise } n \\ 0 & \text{sinon.} \end{cases}$$

On peut alors écrire

$$\omega(n) = \sum_p \chi_p(n)$$

pour tout entier $n \geqslant 1$ (la somme est finie pour tout entier n fixé, donc ne présente aucun problème de définition).

Intuitivement (comme dans la section précédente) la probabilité qu'un entier soit divisible par p doit être $1/p$. Si on interprète χ_p comme une variable aléatoire, définie sur Ω_N pour tout entier $N \geqslant 1$, on vérifie alors que lorsque N tend vers l'infini, χ_p (restreint à Ω_N) converge en loi vers une loi de Bernoulli B_p telle que

$$\mathbf{P}\{B_p = 1\} = \frac{1}{p}.$$

Puisque un entier $n \leqslant N$ ne peut être divisible par aucun nombre premier $> N$, on peut donc penser que ω_N se comportera comme la somme

$$X_N = \sum_{p \leqslant N} B_p.$$

De plus, on peut penser intuitivement que χ_p et χ_q sont « indépendantes » quand $p \neq q$ car le fait d'être divisible par un nombre premier p est indépendant de la divisibilité par un autre nombre premier. La somme ci-dessus ressemble donc à une somme de variables de Bernoulli *indépendantes* (mais qui ne sont pas identiquement distribuées, puisque la probabilité $\mathbf{P}\{B_p = 1\}$ dépend de p). La forme du théorème limite central de Markov (ou de Lindeberg) implique néanmoins que

$$\frac{X_N - \mathbf{E}(X_N)}{\sqrt{\mathbf{V}(X_N)}}$$

doit converger vers une loi normale centrée de variance 1. Il reste alors à reconnaître l'espérance et la variance de X_N: la première est

$$\mathbf{E}(X_N) = \sum_{p \leqslant N} \frac{1}{p}$$

et la seconde

$$\mathbf{V}(X_N) = \sum_{p \leqslant N} \frac{1}{p}\left(1 - \frac{1}{p}\right).$$

Or, il est connu que

$$\sum_{p \leqslant N} \frac{1}{p} \sim \ln \ln N$$

quand $N \to +\infty$ (ce résultat, qui n'est pas très difficile, a été démontré d'abord par Mertens, basé sur des idées de Tchebychev...) On voit donc que le modèle heuristique d'indépendance de la divisibilité des entiers par des nombres premiers suggère bien le théorème d'Erdős-Kac.

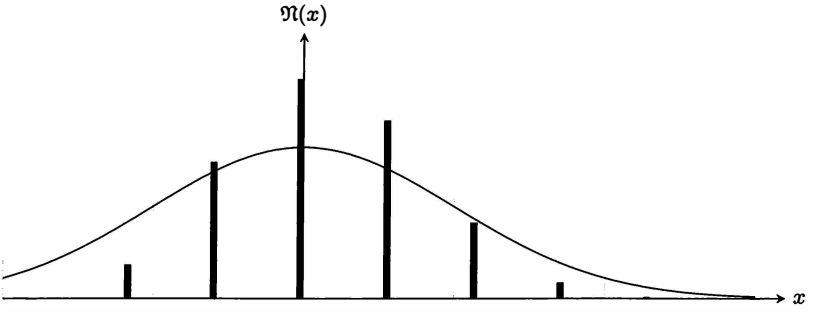


FIGURE 28.1 — *Théorème d'Erdős-Kac : comparaison entre la loi de* ω_N *(une fois centrée et réduite) et la loi normale.*

La figure 28.1 représente la comparaison entre la densité gaussienne et la répartition de $(\omega_N - \ln\ln N)/\sqrt{\ln\ln N}$ pour $N = 1000000$. On voit que l'accord avec le théorème d'Erdős-Kac est loin d'être évident ! C'est en fait un phénomène bien compris : en réalité, il est préférable d'approcher la variable aléatoire $\omega_N - 1$ par une loi de Poisson de paramètre $\ln\ln N$. En effet, on peut démontrer que pour tout $k \geqslant 0$ entier, on a

$$\frac{1}{N}\,\mathrm{Card}\,\{n \leqslant N\ ;\ \omega(n) - 1 = k\} \underset{N\to\infty}{\sim} \frac{1}{k!}\frac{1}{\ln N}(\ln\ln N)^k,$$

ce qui est exactement $\mathbf{P}\{P_{\ln\ln N} = k\}$ pour une variable aléatoire $P_{\ln\ln N}$ répartie comme une loi de Poisson de paramètre $\ln\ln N$ (voir figure 28.2). Le théorème d'Erdős-Kac est alors similaire au fait que si P_λ est une loi de Poisson de paramètre $\lambda > 0$, on a la convergence en loi

$$\frac{P_\lambda - \lambda}{\sqrt{\lambda}} \overset{\mathrm{L}}{\Longrightarrow} \mathcal{N}(0\ ;\ 1)$$

quand $\lambda \to +\infty$. Et on peut en effet majorer de manière convenable la « distance » entre ω_N et une loi de Poisson de paramètre $\ln\ln N$, et en déduire le théorème d'Erdős-Kac. Voir Kowalski & Nikeghbali [66] pour plus de détails.

28.3 ÉCARTS ENTRE NOMBRES PREMIERS ET LOIS DE POISSON

Nous considérons maintenant un tout autre problème, qui concerne la répartition des nombres premiers dans des intervalles « courts ». Plus précisément, on souhaite regarder des intervalles

$$X < n \leqslant X + Y$$

dont on s'attend, en moyenne au moins, à ce qu'ils recèlent un nombre fixé (disons $\lambda > 0$) de nombres premiers. Pour déterminer la forme de ces intervalles, on peut faire appel au théorème des nombres premiers, et à un raisonnement heuristique. Gauss avait conjecturé que la « probabilité » qu'un grand entier n soit premier est approximativement $1/(\ln n)$. Plus précisément, cette conjecture suggère que le nombre $\pi(X)$ de nombres premiers $p \leqslant X$ devrait satisfaire

$$\pi(X) \sim \sum_{2\leqslant n\leqslant X} \frac{1}{\ln n},$$

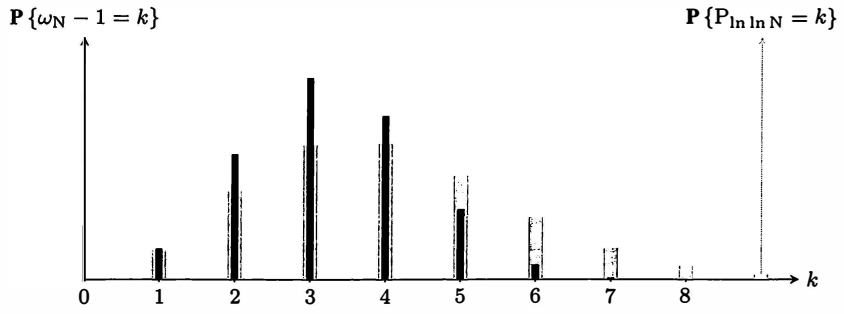


FIGURE 28.2 — *Théorème d'Erdős-Kac : comparaison entre la loi de* ω_N *et la loi de Poisson de paramètre* $\ln\ln N$.

une somme qui peut être évaluée asymptotiquement assez aisément : par comparaison avec une intégrale, on a

$$\sum_{2\leqslant n\leqslant X} \frac{1}{\ln n} \sim \int_2^X \frac{dt}{\ln t},$$

une fonction spéciale habituellement appelée le *logarithme intégral* de X, et notée Li(X).

La formule asymptotique

$$\pi(X) \sim \mathrm{Li}(X)$$

a été finalement démontrée en 1896 par Jacques HADAMARD et Charles de LA VALLÉE-POUSSIN, indépendamment. Elle suggère (mais ne démontre pas du tout !) que, lorsque p est un nombre premier de taille approximativement N, le nombre premier suivant devrait être d'ordre de grandeur proche de $N+\ln N$. Cela n'est pas forcément vrai pour tout entier N, mais le théorème des nombres premier prouve que c'est le cas en moyenne sur N, au sens suivant :

$$\frac{1}{N}\sum_{2\leqslant n\leqslant N} \{\pi(n+\ln n) - \pi(n)\} \sim \frac{1}{N}\sum_{2\leqslant n\leqslant N} \frac{\ln n}{\ln N} \to 1.$$

Pour la même raison, on a

$$\frac{1}{N}\sum_{2\leqslant n\leqslant N} \{\pi(n+\lambda\ln n) - \pi(n)\} \to \lambda,$$

et donc, en moyenne sur X, un intervalle du type

$$[X; X+\lambda(\ln X)]$$

contient λ nombres premiers.

La question qui apparaît alors est la suivante : lorsque n varie, quelle est la répartition de la quantité

$$P_\lambda(n) = \pi(n+\lambda(\ln n)) - \pi(n) \ ?$$

Si l'on prend $n \in \Omega_N$, et que l'on voit P_λ comme une variable aléatoire, y a-t-il une répartition limite bien définie quand $N \to +\infty$?

On ne connaît pas la réponse à cette question, mais on a la conjecture suivante :

■ **Conjecture 28.3** *Soit $\lambda > 0$ fixé. Lorsque $N \to +\infty$, la variable aléatoire $n \mapsto P_\lambda(n)$ sur Ω_N converge en loi vers une loi de Poisson de paramètre λ, c'est-à-dire, pour tout $j \geqslant 0$, on a*

$$\frac{1}{N}\operatorname{Card}\{n \leqslant N\,;\ \pi(n+\lambda\ln n)-\pi(n)=j\} \xrightarrow[N\to\infty]{} \frac{\lambda^j}{j!}e^{-\lambda}.$$

Cette conjecture a une justification théorique plus profonde que celle basée sur l'heuristique ci-dessus. P. X. Gallagher a prouvé, en effet, que l'on pourrait démontrer cet énoncé si l'on savait résoudre de manière suffisamment précise les conjectures dites *des k-uplets* de Hardy et Littlewood [35]. Celles-ci prédisent que si un k-uplet $(h_1, \ldots, h_k)$ d'entiers $\geqslant 0$ est donné, et si aucune obstruction évidente ne se présente, il doit exister une infinité d'entiers $n \geqslant 1$ tels que

$$n + h_1, \ldots, n + h_k$$

soient tous premiers. Par exemple, si l'on prend $k = 2$, $h_1 = 0$, $h_2 = 2$, on obtient la conjecture des nombres premiers jumeaux, et le cas général en est une généralisation considérable. Comme, de plus, l'argument de Gallagher demande une version très précise et uniforme, on peut voir que la conjecture semble très difficile...

Comme illustration de cela, notons que l'on peut montrer que *si* la méthode de Gallagher s'applique... elle s'applique aussi bien pour compter les nombres premiers jumeaux dans des intervalles de la bonne longueur ! Précisément, on peut conjecturer [65] :

■ **Conjecture 28.4** *Soit $\lambda > 0$ fixé. Posons*

$$c_2 := 2\prod_{p\geqslant 3}\frac{1-2/p}{(1-1/p)^2} \approx 1,320336593\ldots$$

la constante des nombres premiers jumeaux. *Notons $\pi_2(N)$ le nombre de nombres premiers jumeaux $p \leqslant N$, c'est-à-dire le nombre de $p \leqslant N$ premiers tels que $p+2$ soit aussi premier.*

Lorsque $N \to +\infty$, la variable aléatoire

$$n \mapsto \pi_2\big(n + c_2^{-1}\lambda(\ln n)^2\big) - \pi_2(n)$$

sur Ω_N converge en loi vers une loi de Poisson de paramètre λ.

28.4 VALEURS PROPRES DE MATRICES ALÉATOIRES

Notre dernier exemple est plus hypothétique, et beaucoup plus sophistiqué. Il s'agit cette fois de modéliser les zéros, ou les valeurs, de la fonction zêta de Riemann, définie par

$$\zeta(s) = \sum_{n=1}^{\infty}\frac{1}{n^s}$$

pour tout s complexe de partie réelle > 1. Cette fonction admet un prolongement méromorphe au plan complexe tout entier, et Riemann a découvert que la distribution des nombres premiers est largement contrôlée par les zéros complexes de $\zeta(s)$ qui se trouvent dans la *bande critique*, définie par

$$0 \leqslant \mathscr{R}e\,(s) \leqslant 1.$$

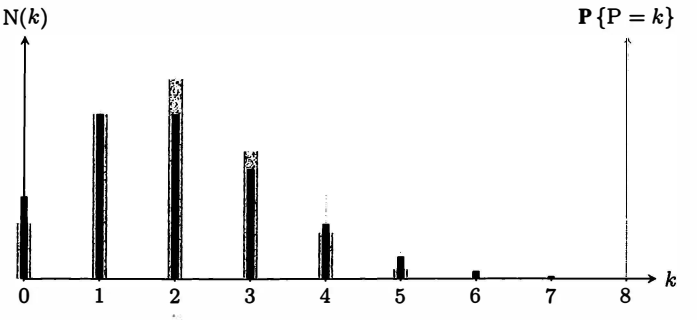


FIGURE 28.3 — *Comparaison entre la proportion* $\mathrm{N}(k)$ *d'entiers compris entre* 2 *et* $10^7 - 2\ln(10^7)$ *vérifiant* $\pi(n + 2\ln n) - \pi(n) = k$ (en noir) *et une loi de Poisson* P *de paramètre* 2 (en gris).

Concrètement, cette correspondance prend par exemple la forme suivante : si on note $\theta \leqslant 1$ le supremum des parties réelles β des nombres complexes $\rho = \beta + \mathrm{i}\gamma$ tels que $\zeta(\rho) = 0$ et $0 \leqslant \beta \leqslant 1$, on a alors

$$\pi(x) = \int_2^x \frac{\mathrm{d}t}{\ln t} + \mathrm{O}(x^{\theta+\epsilon})$$

pour $x \geqslant 2$ et pour tout $\epsilon > 0$. La célèbre *hypothèse de Riemann* peut s'énoncer comme l'égalité $\theta = 1/2$, et cette conjecture est toujours fort mystérieuse (à vrai dire, on ne sait pas à l'heure actuelle si $\theta < 1$, et la formule ci-dessus n'apporte aucune information lorsque $\theta = 1$...)

On sait en revanche (depuis Riemann) qu'il existe dans la bande critique une infinité de zéros $\rho = \beta + \mathrm{i}\gamma$. Afin de tenter de comprendre la fonction zêta, H. Montgomery a commencé (dans les années 70) l'étude de la distribution des écarts entre les zéros de $\zeta(s)$. Pour cela, on ordonne les parties imaginaires des zéros en suite croissante

$$0 < \gamma_1 \leqslant \gamma_2 \leqslant \cdots \leqslant \gamma_n \leqslant \cdots$$

(on sait que le premier zéro est $1/2 + \mathrm{i}\gamma$ avec $\gamma \approx 14$), et on regarde les écarts successifs

$$\delta_1 = \gamma_2 - \gamma_1, \ \ldots, \ \delta_n = \gamma_{n+1} - \gamma_n, \ldots$$

On sait également que ces écarts tendent vers 0, mais que les écarts normalisés

$$\Delta_n = \delta_n \cdot \ln \frac{\gamma_n}{2\pi}$$

ont une moyenne égale à 1, au sens où

$$\frac{1}{\mathrm{N}} \sum_{n \leqslant \mathrm{N}} \Delta_n \to 1$$

quand $\mathrm{N} \to +\infty$. L'idée de Montgomery est d'étudier si ces écarts normalisés ont une répartition asymptotique précise quand $\mathrm{N} \to +\infty$. La conjecture à laquelle Montgomery a abouti, fondée sur des résultats théoriques partiels (basés sur l'Hypothèse de Riemann) est la suivante : ces écarts se comportent asymptotiquement comme l'écart entre les angles propres successifs de matrices unitaires aléatoires de taille $(\ln \frac{\mathrm{N}}{2\pi})$.

Nous allons maintenant rendre précis cet énoncé. Soit $\mathrm{M} \geqslant 1$ un entier, et soit $\mathrm{U}(\mathrm{M})$ le groupe des matrices unitaires de taille M, c'est-à-dire les matrices carrées

A de taille M à coefficients complexes telles que $A^t\bar{A} = 1$ (les isométries de $\mathbb{C}^M$ pour son produit hermitien usuel). Ce groupe de matrices, vu dans l'espace des matrices complexes de taille M, est compact (fermé et borné). Il existe par ailleurs une mesure de probabilité canonique μ_M sur U(M), qui est déterminée de manière unique par la condition

$$\mu_M(X) = \mu_M(AX)$$

pour tout $A \in U(M)$ et tout sous-ensemble mesurable $X \subset U(M)$.

Exemple 28.5 Pour M = 1, U(1) n'est autre que l'ensemble des nombres complexes de module 1, c'est-à-dire le cercle unité dans $\mathbb{C}$. La mesure μ_1 est tout simplement la mesure de Lebesgue, ou mesure angulaire, donnée par

$$\frac{d\vartheta}{2\pi}$$

pour $z = e^{i\vartheta} \in U(1)$.

On peut donc voir U(M), avec la mesure μ_M, comme un espace probabilisé, et parler alors des propriétés de « matrices unitaires aléatoires » de taille M. Parmi les questions les plus naturelles se posent celles concernant la répartition des valeurs propres des éléments de U(M).

Chaque matrice $A \in U(M)$ peut effectivement être diagonalisée, et ses valeurs propres sont de module 1. Écrivons celles-ci sous la forme

$$e^{i\vartheta_1}, \dots, e^{i\vartheta_M}$$

avec la convention

$$0 \leqslant \vartheta_1 \leqslant \vartheta_2 \leqslant \cdots \leqslant \vartheta_M < 2\pi.$$

Les écarts successifs sont alors

$$\eta_1 = \vartheta_2 - \vartheta_1, \ \dots, \ \eta_{M-1} = \vartheta_M - \vartheta_{M-1},$$

et là encore, ces écarts sont généralement petits (car il faut placer M nombres dans un intervalle de longueur fixée 2π), mais les écarts normalisés

$$\Phi_n = \frac{M}{2\pi}\eta_n$$

ont une moyenne

$$\frac{1}{M}\sum_{1\leqslant n\leqslant M-1} \Phi_n = \frac{1}{2\pi}\sum_{1\leqslant n\leqslant M-1} \eta_n = \frac{1}{2\pi}(\vartheta_M - \vartheta_1)$$

qui est (avec grande probabilité) proche de 1 si M est grand. Ici, nous considérons chaque angle propre ϑ_i comme une *fonction* de $A \in U(M)$; cette fonction est mesurable, et l'on peut parler de sa moyenne par rapport à U(M), etc.

Diverses propriétés de matrices aléatoires (pas seulement du type ci-dessus) ont été étudiées par des statisticiens, des physiciens, et bien d'autres.

En particulier, Freeman Dyson[1] savait (au moment où Montgomery étudiait les zéros de la fonction zêta) que les Φ_n ont une propriété limite remarquable quand M tend vers l'infini. On a en effet, pour tous réels

$$0 \leqslant a < b,$$

(1). Freeman Dyson (1923–), physicien et mathématicien anglais, naturalisé américain, est un des théoriciens de l'électrodynamique quantique et de la théorie quantique des champs.

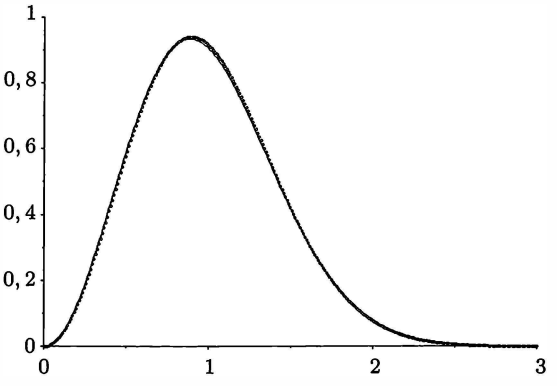


FIGURE 28.4 — *Expériences de Xavier* GOURDON *[39].*

la limite

$$\lim_{M\to+\infty} \mathbf{E}_{U(M)}\Big(\frac{1}{M}\,\mathrm{Card}\,\{n\,;\ a<\Phi_n(A)<b\}\Big)=\int_a^b d\nu(t)$$

pour une certaine mesure ν sur $[0\,;+\infty[$. Cette mesure a une densité $f(x)$ par rapport à la mesure de Lebesgue, supportée sur $[0\,;+\infty[$, qui est représentée sur la figure 28.4. On remarque en particulier que $f(x)$ s'annule (à l'ordre 2) en $x=0$, ce qui signifie intuitivement que les valeurs propres *se repoussent* : si on a une valeur propre en $e^{i\vartheta}$, en général, l'angle propre suivant sera plus loin que ce ne serait le cas si on avait des angles ϑ_n choisis indépendamment et uniformément dans $[0\,;2\pi[$.

La conjecture de Montgomery, indépendamment de toute considération de matrices aléatoires, est l'assertion suivante :

■ **Conjecture 28.6** *Pour tout $a<b$, avec les notations ci-dessus concernant les zéros de la fonction zêta de Riemann, on a*

$$\lim_{N\to+\infty}\frac{1}{N}\,\mathrm{Card}\,\{n\leqslant N\,;\ a<\Delta_n<b\}=\int_a^b d\nu(t).$$

Bien qu'elle reste complètement ouverte, cette conjecture peut être étudiée numériquement. Les résultats, dus en grande partie à Odlyzko, sont spectaculaires ! (Bien plus, d'ailleurs que ce que l'on peut voir pour le théorème d'Erdős-Kac, ou pour la conjecture, bien étayée, concernant la distribution de $\pi(n+\lambda\ln n)-\pi(n)$...)

Il existe de nombreuses variantes de cette conjecture, et des généralisations où l'on utilise les valeurs du *polynôme caractéristique* de matrices aléatoires pour modéliser certaines propriétés de la fonction zêta...

Pour plus de détails et d'autres références, voir par exemple Katz & Sarnak [56].

Chapitre 29

Bosons, fermions et boltzmannions

Dans ce court chapitre, nous montrons deux applications simples du dénombrement à la mécanique statistique. La première consiste à montrer que, dans un système isolé constitué d'un grand nombre de particules, la probabilité de trouver l'une d'elles dans un état individuel d'énergie ε est proportionnel à $e^{-\beta\varepsilon}$, où β est une constante que l'on peut identifier à $1/k_B T$, où T est la température du système et k_B la constante de Boltzmann. Dans la seconde, nous montrons comment l'hypothèse d'indiscernabilité des particules — bosons ou fermions — affecte leur répartition et, notamment, le nombre moyen d'états occupés.

29.1 LE FACTEUR DE BOLTZMANN

Théorie cinétique des gaz et statistiques non quantiques §412

Dans une série de travaux qui ont profondément bouleversé la physique moderne, le physicien allemand Ludwig BOLTZMANN (1844–1906) a posé les principes probabilistes de la thermodynamique statistique. Il s'agissait de construire, de façon mathématique, un modèle de gaz dont les particules, régies par les lois usuelles de la mécanique, sont en nombre si grand que la description individuelle doit être remplacée par une description statistique[1]. À l'époque de la fondation de la mécanique statistique, il s'agissait d'ailleurs surtout de décrire des systèmes à l'équilibre[2].

Le principe de base de la mécanique statistique est de considérer que les positions et les vitesses des molécules d'un gaz peuvent être considérées comme des

(1). Rappelons qu'un litre de gaz à pression et température ambiante contient environ $3 \cdot 10^{22}$ molécules ; la description mécanique du système est donc la donnée d'autant d'équations différentielles d'ordre 2 pour chaque dimension d'espace.

(2). Les mouvements des particules étant rapides, une transformation « lente » à l'échelle humaine peut être décrite comme une succession de positions d'équilibre ; les transformations violentes (comme le cas d'un gaz emprisonné dans une bouteille de verre, elle-même contenue dans une enceinte hermétique, au moment où la bouteille se brise) ne sont pas concernées par la mécanique statistique à l'équilibre : tout au plus peut-on décrire l'état initial et l'état final après relaxation du système.

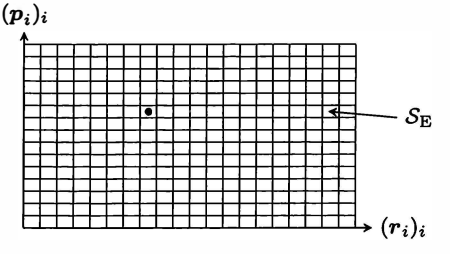
variables aléatoires, dont les espérances sont des grandeurs macroscopiquement mesurables. Le fait que ces particules soient en nombre extrêmement grand entraîne que l'écart de la valeur moyenne observée d'une quantité à cette espérance est, relativement au nombre n de particules mises en jeu, extrêmement petit, puisque d'ordre $1/\sqrt{n}$. De plus, sous des conditions très générales, les espérances des différentes variables aléatoires sont égales aux moyennes temporelles de ces mêmes grandeurs : c'est l'**hypothèse ergodique** introduite par James Clerk MAXWELL (1831–1879) et Willard GIBBS (1839–1903).

On considère un gaz constitué de n particules (des molécules d'un gaz rare, comme l'argon par exemple), évoluant librement et sans interactions dans une enceinte de volume V fixé. Les parois de cette enceinte sont supposées rigides, de sorte que les chocs des particules contre elles se traduisent par des simples rebonds, sans perte d'énergie. Chacune des n particules est caractérisée par une position $\boldsymbol{r}_i = (x_i, y_i, z_i)$ et une impulsion[3] $\boldsymbol{p}_i = (p_i^x, p_i^y, p_i^z)$. L'**espace des phases** est l'espace $\mathcal{E}_n = \mathbb{R}^{6n}$ dont les vecteurs sont les données de toutes les positions et toutes les impulsions des n particules.

Un tel système isolé, à énergie fixée, est représenté, à un instant t donné, par un point de l'espace des phases $\mathcal{E}_n$. Au cours du temps, ce point se déplace sur la surface $\mathcal{S}_\mathrm{E}$ de l'espace des phases d'énergie E, surface de dimension $6n-1$. Pour simplifier les raisonnements, on considère habituellement que l'énergie, au lieu d'être strictement fixée à E, peut en réalité varier entre E et $\mathrm{E}+\delta\mathrm{E}$, où $\delta\mathrm{E} \ll \mathrm{E}$.

Pour rendre compte de l'absence de connaissance de la description microscopique du système, on décrit un gaz à l'équilibre thermodynamique non pas par un unique point, mais par un ensemble de systèmes répartis de façon équiprobable sur la couche d'énergie $[\mathrm{E}\,;\,\mathrm{E}+\delta\mathrm{E}]$. Cet ensemble est appelé **ensemble microcanonique**. Son évolution dynamique est régie par les équations de Hamilton ; d'après le théorème de Liouville, la densité — constante sur la couche d'énergie, nulle ailleurs — est préservée au cours du temps.

Afin d'éviter les problème liés aux probabilités continues, on *discrétise* l'espace des phases, c'est-à-dire que les positions et les impulsions de chaque particule ne peuvent prendre qu'un nombre dénombrable de valeurs[4]. Ce nombre sera même fini puisque l'extension spatiale est limitée, et l'énergie totale étant fixée, cela majore *de facto* la valeur de chaque $\|\boldsymbol{p}_i\|$. Ainsi, l'**état microscopique** du système est entièrement décrit par le choix d'une des cellules de l'espace $\mathcal{E}_n$, et toutes ces cellules sont équiprobables.



(3). En mécanique non relativiste et en l'absence de champ électromagnétique, cette impulsion est identique à la quantité de mouvement $\boldsymbol{p}_i = m\,\boldsymbol{v}_i$ de la particule.

(4). Cette discrétisation change évidemment la physique du système ; cependant, lorsque la taille d'une cellule élémentaire de l'espace des phases tend vers 0, les différentes quantités physiques calculées par cette méthode convergent vers une limite fixée, ce qui assure déjà que le procédé n'est pas complètement idiot. Mieux, la mécanique quantique préconise justement l'existence d'une échelle physique pour ces cellules ; pour chacune des trois dimensions, on doit avoir $\delta x_i \cdot \delta p_i \geqslant \hbar$, où $\hbar$ est la constante de Planck réduite, $\hbar = h/2\pi$.

« Nombre de complexions » d'un état § 413

Voici une autre formulation, équivalente, du même problème. Chaque particule i a accès à des états individuels, qui sont des cellules de l'espace $\mathcal{E} = \mathbb{R}^6$. (L'espace des phases est donc le produit des n espaces associés à chaque particule, qui sont tous des copies de $\mathbb{R}^6$.) Ces états individuels accessibles sont en nombre fini, nous les nommerons $A_1, A_2, \ldots, A_q$.

On peut alors décrire l'état physique du système en indiquant comment les particules sont réparties dans ces états. Il y a deux façons de voir les choses :

- *Dans le point de vue macroscopique*, on donne, pour chaque état A_k, le nombre de particules dans cet état, sous la forme d'une liste de nombres $(n_1, n_2, \ldots, n_q)$ ou d'un schéma où chaque état est représenté par une case. Par exemple, si le système est composé de $n = 5$ particules ayant accès à $q = 4$ états individuels, un état macroscopique peut être représenté par

$$(2,0,3,0) \qquad \text{ou par} \qquad \boxed{\bullet\bullet}\,\boxed{}\,\boxed{\bullet\bullet\bullet}\,\boxed{} \tag{29.1}$$

- *Dans le point de vue microscopique*, on donne, pour chaque état A_k, la liste des particules qui se trouvent dans cet état, par exemple

$$\boxed{①②}\,\boxed{}\,\boxed{⑤③④}\,\boxed{}$$

Reste à déterminer comment se définit la loi de probabilité associée et, notamment, comment compter ces états. Boltzmann postule que :

(P_1) : *les particules sont discernables*,

ce qui revient à dire qu'il est possible de leur attribuer (fictivement) un numéro. Sous ces hypothèses, les états suivants sont bien distincts et doivent être comptés séparément :

$$\boxed{①②}\,\boxed{}\,\boxed{⑤③④}\,\boxed{} \neq \boxed{①③}\,\boxed{}\,\boxed{⑤②④}\,\boxed{} \tag{29.2}$$

En revanche, il faut compter comme un seul état physique microscopique les deux configurations suivantes :

$$\boxed{①②}\,\boxed{}\,\boxed{⑤③④}\,\boxed{} = \boxed{②①}\,\boxed{}\,\boxed{⑤④③}\,\boxed{} \tag{29.3}$$

Les deux graphes de (29.3) correspondent à la même cellule de l'espace des phases, tandis que les deux graphes de (29.2) correspondent à deux cellules distinctes.

Pour traduire l'équiprobabilité sur l'espace des phases, Boltzmann appelle **complexion** la description microscopique de l'état individuel de chaque particule. Ainsi, dans (29.3) apparaît une seule et même complexion, tandis que (29.2) montre deux complexions distinctes. Le second postulat de Boltzmann est alors :

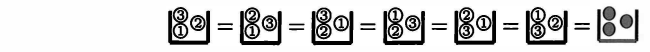
(P_2) : *toutes les complexions sont équiprobables.*

Comme, à un état physique macroscopique donné, peuvent correspondre plusieurs complexions, la probabilité de cet état macroscopique est simplement proportionnelle au nombre de complexions qui le représentent.

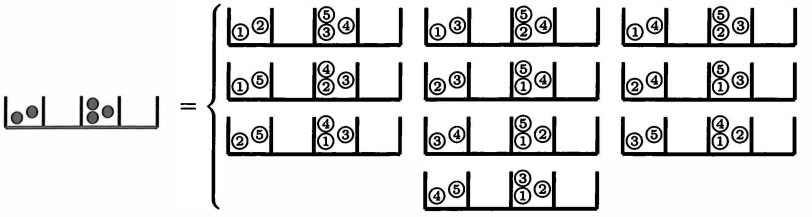
Exemple 29.1 Comptons par exemple le nombre de complexions distinctes décrivant le même état physique (29.1). Tout d'abord, il s'agit de répartir les 5 particules dans les 5 « emplacements » que représentent les disques noirs de la figure, ce qui donne $5! = 120$ possibilités. Plusieurs de ces possibilités correspondent à une même complexion puisque, comme déjà dit,

$$\boxed{①②} = \boxed{②①} = \boxed{\bullet\bullet}$$

Il convient donc de diviser le nombre trouvé par le nombre de permutations à deux objets, c'est-à-dire $2! = 2$. De même,



et il convient de diviser le nombre trouvé par le nombre de permutations à trois objets, c'est-à-dire $3! = 6$. Ainsi, le nombre total de complexions de l'état physique (29.1) est $5!/2!\,3! = 10$, les complexions étant :



De manière générale, un état physique $\omega = (n_1, n_2, \dots, n_q)$ a, en notant $n = n_1 + n_2 + \cdots + n_q$, le nombre de complexions suivant

$$\mathrm{C}(\omega) = \frac{n!}{n_1!\, n_2! \cdots n_q!}.$$

Remarque 29.2 Si l'on ne prend pas en compte la contrainte énergétique, la somme des nombres de complexions des états microscopiques est égale au nombre de possibilités de ranger n particules discernables dans q états, c'est-à-dire q^n :

$$\sum_{\omega} \mathrm{C}(\omega) = q^n.$$

§ 414 **Un exemple (très) élémentaire**

Toujours suivant Boltzmann, considérons un système de $n = 7$ particules, pouvant se trouver dans l'un des $q = 8$ états $\mathrm{A}_0, \mathrm{A}_1, \dots, \mathrm{A}_7$ possibles. On suppose que l'état A_k est caractérisé par une énergie $k\varepsilon$, où ε est une énergie fondamentale. On suppose de plus que l'énergie totale du système est $\mathrm{E} = 7\varepsilon$. On cherche les états possibles du système, c'est-à-dire les états $(n_0, n_1, \dots, n_7)$ satisfaisant la double contrainte

$$\begin{cases} n = n_0 + n_1 + n_2 + \cdots + n_7 \\ \mathrm{E} = (0)\, n_0 + (\varepsilon)\, n_1 + (2\varepsilon)\, n_2 + \cdots + (7\varepsilon)\, n_7. \end{cases}$$

Il suffit de s'armer d'un crayon et d'un tout petit peu de patience pour retrouver les 15 états annoncés par Boltzmann, et présentés dans le tableau 29.1 page suivante. Pour chaque état, le nombre de complexions est calculée en dernière ligne. Il apparaît par exemple que l'état le plus probable est l'état $\omega_{10} = (3, 2, 1, 1, 0, 0, 0, 0)$, qui a un nombre de complexions de 420, et donc une probabilité d'apparition

$$\mathbf{P}(\omega_{10}) = \frac{\mathrm{C}(\omega_{10})}{\mathrm{C}(\omega_1) + \mathrm{C}(\omega_2) + \cdots + \mathrm{C}(\omega_{15})} = \frac{420}{1716} = \frac{35}{143} \approx 0,245.$$

Dans l'avant-dernière colonne, on fait apparaître la population moyenne de chaque niveau d'énergie $k\varepsilon$ (la population de l'état A_k dans la configuration ω_i est multipliée par $\mathrm{C}(\omega_i)$ et on somme sur les indices i) ; une fois normalisée, cette colonne donne la proportion de particules dans l'état A_k.

On constate une très nette décroissance de la population lorsque l'énergie du niveau correspondant augmente. La raison se comprend aisément : lorsqu'une ou plusieurs particules sont dans un niveau d'énergie élevée, elles contraignent les autres à se tasser dans les niveaux de basse énergie, diminuant donc énormément le nombre de complexions des états compatibles.

Éner.	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}	ω_{11}	ω_{12}	ω_{13}	ω_{14}	ω_{15}	$\langle n_k \rangle$	p_k
7ε	•															7	≈ 0
6ε		•														42	0,003
5ε			•	•												147	0,012
4ε					•	•	•									392	0,033
3ε					•			••	•	•	•					882	0,073
2ε			•			•			••	•		•••	••	•		1764	0,147
ε		•		••		•	•••	•		••	•••	•	•••	•••	•••	3234	0,269
0	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	••	•••	••	•		5544	0,462
C(ω)	7	42	42	105	42	210	140	105	105	420	105	140	210	42	1		

FIGURE 29.1 — *Les 15 états physiques accessibles à un système de 7 particules et d'énergie totale* $\mathrm{E} = 7\,\varepsilon$. *En dernière ligne, le nombre de complexions. En dernière colonne, la probabilité qu'une particule soit dans le niveau d'énergie* $k\varepsilon$.

Le facteur de Boltzmann

§ 415

Il est très difficile de calculer la population moyenne $\langle n_k \rangle$ d'un niveau d'énergie E_k donné lorsque le nombre de particules du système devient grand. Au lieu de mener ce calcul, nous allons chercher quel est l'*état le plus probable.* Plusieurs méthodes permettent d'obtenir ce résultat : celle de Boltzmann, que nous allons plus ou moins suivre, ne fait intervenir que des raisonnements purement statistiques[(5)].

Remarque 29.3 Sur le tableau 29.1, on a calculé la probabilité p_k qu'une particule soit dans un état d'énergie $k\varepsilon$ donnée, en tenant compte de toutes les complexions de tous les états. Si l'on calcule la probabilité d'occupation $\tilde{p}_k$ des niveaux d'énergie en ne considérant que l'état le plus probable, à savoir ω_{10}. On trouve

$$\tilde{p}_0 = \frac{3}{7} \qquad \tilde{p}_1 = \frac{2}{7} \qquad \tilde{p}_3 = \tilde{p}_4 = \frac{1}{7} \qquad \text{et} \qquad \tilde{p}_5 = \tilde{p}_6 = \tilde{p}_7 = 0.$$

Les valeurs numériques coïncident plus ou moins :

Énergie	0	ε	2ε	3ε	4ε	5ε	6ε	7ε
p_k	0,462	0,269	0,147	0,073	0,033	0,012	0,003	≈ 0
$\tilde{p}_k$	0,429	0,286	0,143	0,143	0	0	0	0

Lorsque le nombre de particules est grand, les distributions de probabilités sont tellement piquées que la différence entre les valeurs calculées au moyen de tous les états, et les valeurs calculées au moyen de l'état le plus probable, coïncident avec une remarquable précision.

Considérons de nouveau un système physique composé de n particules, ayant accès à des états physiques $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_q$ d'énergies respectives $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_q$. On suppose que le système est isolé et n'échange donc pas d'énergie avec l'extérieur ; son énergie totale est constante et égale à E. Un état $\omega = (n_1, n_2, \ldots, n_q)$, dans lequel n_k particules sont dans l'état A_k, est caractérisé par un nombre de complexions

$$\mathrm{C}(\omega) = \frac{n!}{n_1!\, n_2! \cdots n_q!}.$$

Le problème est donc de déterminer l'état ω qui maximise la quantité $\mathrm{C}(\omega)$ ou, de façon plus pratique, son logarithme

$$f(\omega) := \ln \mathrm{C}(\omega) = \ln(n!) - \ln(n_1!) - \ln(n_2!) - \cdots - \ln(n_q!),$$

(5). D'autres raisonnements, faisant davantage intervenir la physique, sont présentés dans la plupart des ouvrages de mécanique statistique, par exemple Diu *et al.* [25] ou Thompson [98].

sous la double contrainte

$$n = n_1 + n_2 + \cdots + n_q \tag{29.4}$$

$$\mathrm{E} = n_1 \varepsilon_1 + n_2 \varepsilon_2 + \cdots + n_q \varepsilon_q. \tag{29.5}$$

Ici, nous faisons une première approximation, en supposant que le nombre de particules est tellement grand que les n_k peuvent être considérées comme des variables *continues*. On doit donc maximiser une fonction de q variables sous les contraintes (29.4) et (29.5). Le plus simple est d'utiliser la méthode des multiplicateurs de Lagrange[6] : on introduit deux réels α et β et on cherche $\tilde{\omega} = (\tilde{n}_1, \tilde{n}_2, \ldots, \tilde{n}_q)$ tel que

$$g := f - \alpha \sum_{k=1}^{q} n_k - \beta \sum_{k=1}^{q} \varepsilon_k n_k$$

soit maximale en $\tilde{\omega}$.

Pour commencer, on a besoin d'estimer les dérivées partielles de g par rapport à chacun des n_k. On récrit $n_k! = \Gamma(n_k + 1)$, ce qui permet de traiter n_k comme une variable continue, et on utilise le développement asymptotique suivant[7] :

$$\frac{\mathrm{d} \ln \Gamma}{\mathrm{d}x}(x+1) = \ln x + \underset{x \to \infty}{\mathrm{O}}\left(\frac{1}{x}\right).$$

La différentielle de g étant

$$\mathrm{d}g = \sum_{k=1}^{q} \left\{ -\frac{\mathrm{d} \ln \Gamma}{\mathrm{d}n_k}(n_k + 1) - \alpha - \beta \varepsilon_k \right\} \mathrm{d}n_k,$$

elle s'annule donc sous les conditions

$$\forall k \in [\![1\,;\,q]\!] \qquad -\frac{\mathrm{d} \ln \Gamma}{\mathrm{d}n_k}(\tilde{n}_k + 1) - \alpha - \beta \varepsilon_k = 0,$$

ce qui donne

$$\forall k \in [\![1\,;\,q]\!] \qquad \ln \tilde{n}_k = -\alpha - \beta \varepsilon_k + \mathrm{O}\left(\frac{1}{\tilde{n}_k}\right)$$

et donc

$$\tilde{n}_k \approx \mathrm{e}^{-\alpha - \beta \varepsilon_k}.$$

---

(6). Rappelons brièvement le principe de cette méthode. On cherche un extremum d'une fonction $f : \mathbb{R}^q \to \mathbb{R}$, soumise aux contraintes $\mathrm{C}_1(x_1, x_2, \ldots, x_q) = \mathrm{C}_2(x_1, x_2, \ldots, x_q) = 0$. D'après le théorème des extrema liés, en un point $\tilde{x}$ satisfaisant ces conditions, le noyau de la forme linéaire $\mathrm{d}f(\tilde{x})$ contient l'intersection des noyaux de $\mathrm{dC}_1(\tilde{x})$ et de $\mathrm{dC}_2(\tilde{x})$, ou encore que $\mathrm{d}f(\tilde{x})$ est combinaison linéaire de $\mathrm{dC}_1(\tilde{x})$ et de $\mathrm{dC}_2(\tilde{x})$. Ainsi, il existe α et β tels que $\mathrm{d}f(\tilde{x}) - \alpha\,\mathrm{dC}_1(\tilde{x}) - \beta\,\mathrm{dC}_1(\tilde{x}) = 0$. Le résultat général est présenté en annexe (A47, page 687).

(7). En suivant Rényi [87], on utilise la représentation d'Euler

$$\ln \Gamma(x+1) = \left(x + \frac{1}{2}\right) \ln x - x + \ln \sqrt{2\pi} - \int_0^{+\infty} \frac{\varphi(t)}{x+t}\,\mathrm{d}t,$$

où φ est la fonction définie par $\varphi(t) := t - \lfloor t \rfloor - \frac{1}{2}$. On peut dériver cette égalité sous le signe somme (il n'est en revanche pas possible de dériver directement la formule asymptotique de Stirling) :

$$\frac{\mathrm{d} \ln \Gamma}{\mathrm{d}x}(x+1) = \frac{\Gamma'(x+1)}{\Gamma(x+1)} = \ln x + \frac{1}{2x} + \int_0^{+\infty} \frac{\varphi(t)}{(x+t)^2}\,\mathrm{d}t = \ln x + \mathrm{O}\left(\frac{1}{x}\right).$$

La première condition (29.4) permet de déterminer explicitement α :

$$\tilde{n}_k \approx n \frac{e^{-\beta\varepsilon_k}}{\sum_i e^{-\beta\varepsilon_i}}. \tag{29.6}$$

Le multiplicateur de Lagrange β a la valeur permettant à la seconde condition de (29.5) d'être satisfaite. Cette valeur dépend évidemment de l'énergie totale du système, mais ne se détermine pas exactement[8]. Des raisonnements physiques permettent de l'identifier avec la quantité $1/k_B T$, où T est la température (absolue) du système et k_B la constante de Boltzmann:

$$\beta = \frac{1}{k_B T}.$$

Enfin, la stricte convexité de g permet de confirmer que $g(\tilde{\omega})$ est bien un maximum.

L'équation (29.6) donne donc la distribution *la plus probable*, que l'on appelle **distribution de Maxwell-Boltzmann.**

Remarque 29.4 Bien sûr, un système physique réel pourra s'écarter plus ou moins de cette distribution. Cependant, lorsque le nombre de particules du système est très grand, les configurations voisines deviennent rapidement extrêmement improbables; la variance de n_k est en effet de l'ordre de $\tilde{n}_k$, et dès que l'un des nombres n_k/n diffère de $\tilde{n}_k/n$ d'une quantité de l'ordre de $1/\sqrt{n}$, la probabilité de la configuration devient absolument négligeable : rappelons qu'en mécanique statistique, on a affaire à des nombres aussi grands que $n = 10^{23}$.

Ainsi, il n'est pas déraisonnable d'identifier, par exemple, les valeurs moyennes des populations dans chaque état avec la quantité $\tilde{n}_k$[9].

Remarque 29.5 Le postulat d'équiprobabilité des états n'est pas compatible avec l'hypothèse d'un gaz absolument sans interactions. Il suffit d'imaginer un gaz dans une enceinte parallélépipédique, telle que, à $t = 0$, les vitesses des particules soient toutes parallèles à l'une des parois; un tel gaz ne « thermalisera » jamais et ne gardera que des configurations qui sont, en principe, hautement improbables. En revanche, une interaction entre particules, aussi faible soit-elle, pourra suffire à ce que le système atteigne la distribution de Maxwell-Boltzmann, ou l'une des distributions extrêmement voisines [10, 50]. Le temps nécessaire pour atteindre un tel équilibre est cependant difficile à estimer.

Remarque 29.6 (De la description microcanonique à la description canonique) La température (ou son coefficient $\beta = 1/k_B T$) d'un système est, on l'a vu, entièrement déterminée par les données « énergie totale, nombre de particules, description des états accessibles ».

Au sein d'un tel système S, considérons un sous-système S′ de particules donné. Ce système S′ est libre d'échanger de l'énergie avec $S'' := S - S'$: son énergie n'est donc plus fixée. À l'équilibre thermodynamique, S′ possède une température qui est, par définition, égale à la température microcanonique de S; la répartition de ses particules dans les différents états accessibles suit la loi de Maxwell-Boltzmann (la probabilité qu'une particule soit dans un état d'énergie ε_k est proportionnelle à $e^{-\beta\varepsilon_k}$). C'est ce que l'on appelle la description **canonique** d'un système (la température est fixée, mais pas l'énergie).

(8). Encore qu'il soit possible de montrer que, si les niveaux d'énergie sont nombreux et distants d'une quantité très inférieure à l'énergie moyenne par particule $\langle E \rangle = E/n$, alors $\beta \approx 1/\langle E \rangle$.

(9). En d'autres termes : avec une distribution très piquée, le mode et la moyenne ont des valeurs très proches.

29.2 STATISTIQUES QUANTIQUES

§ 416 **L'entropie et le paradoxe de Gibbs**

L'une des plus profondes innovations de Boltzmann fut d'identifier l'entropie S d'un gaz — dans un état macroscopique donné — au logarithme du nombre W de configurations microscopiques compatibles avec les données macroscopiques :

$$S = k_B \ln W. \qquad \textit{(formule de Boltzmann)}$$

Comme nous l'avons dit, déterminer W requiert de découper l'espace des phases en cellules élémentaires. Dans le cas d'un gaz de particules sans interactions et d'encombrement nul, le calcul de W sa factorise en deux contributions : une contribution *configurationelle*, provenant d'un comptage de la répartition spatiale des particules, et une contribution venant du comptage de la répartition des vitesses des particules. Cette dernière contribution, qui dépend de l'énergie totale du gaz et de la masse des particules, n'apporte rien au raisonnement qui va suivre et nous la passons sous silence[10], pour ne nous concentrer que sur l'entropie configurationelle.

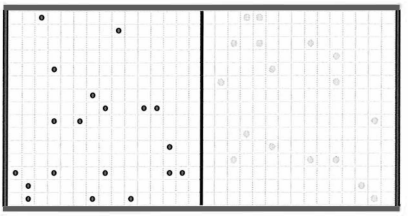
Considérons en effet une enceinte de volume V contenant un gaz de n particules, supposées sans interactions, et calculons l'entropie de ce système. Divisons, de manière un peu arbitraire, le volume V en N cellules spatiales. Chaque particule peut se trouver dans l'une des N cellules possibles et, par indépendance des choix,

$$W = N^n.$$

L'entropie associée est donc, *en oubliant le facteur* k_B :

$$S = n \ln N.$$

Considérons maintenant une enceinte de volume 2V, séparée en deux parties égales par une cloison étanche. Chacune des deux petites enceintes, divisée en N cellules élémentaires, contient n particules identiques, de type A pour la première, de type B pour la seconde.

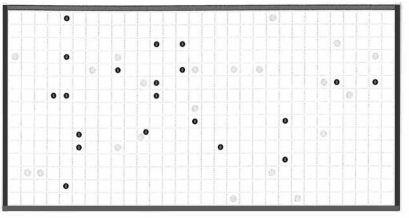


Le nombre de configurations possible est le produit W × W du nombre de configuration dans chaque petite enceinte, ce qui fait que l'entropie du système total est

$$S = 2n \ln N.$$

Retirons maintenant la paroi, les particules sont libres de parcourir le volume 2V tout entier :

(10). La contribution passée sous silence est celle qui est responsable des propriétés thermodynamiques du gaz parfait, elle est donc, pour le physicien, absolument essentielle. Le lecteur intéressé par son calcul pourra consulter Diu *et al.* [25] par exemple.



Le nombre de configurations pour le nouveau système est

$$W' = (2N)^n \cdot (2N)^n = (2N)^{2n}$$

et l'entropie vaut

$$S' = 2n \ln N + 2n \ln 2.$$

Retirer la paroi a donc provoqué un accroissement d'entropie

$$\delta S := S' - S = (2 \ln 2)\, n, \qquad (*)$$

qui est une quantité strictement positive. Lorsque les espèces A et B sont distinctes (par exemple si l'on place de l'hélium à gauche et de l'argon à droite), le résultat est expérimentalement correct : l'entropie d'un mélange hélium/argon est bien plus élevée que la somme des entropies des gaz séparés, conformément aux principes de la thermodynamique du XIX^{e} siècle, qui voulait que cela caractérisât les transformations irréversibles.

Tout se gâte si les espèces A et B sont identiques. Placer ou retirer une paroi au milieu de l'enceinte ne devrait *a priori* pas changer l'entropie du système[(11)]. L'équation $(*)$ est alors une véritable catastrophe puisque :

— l'entropie n'est plus une grandeur extensive ;
— ce n'est plus une fonction d'état car elle dépend non seulement de l'état du gaz, mais aussi de son histoire ;
— en fait, l'entropie n'est même plus définie correctement : on peut toujours imaginer que l'enceinte était séparée en deux, ou en quatre, en huit... permettant de donner à l'entropie une valeur arbitrairement grande !

La résolution de ce paradoxe a conduit à changer la manière de compter les configurations. Au lieu de considérer les particules comme des entités discernables, comptons comme une seule et même configuration toutes celles qui se déduisent les unes des autres par une permutation des particules (c'est le *principe de symétrisation*). Le nombre de configurations possibles pour un système de n particules dans une enceinte de N cellules est maintenant

$$W_* = \frac{N^n}{n!}$$

et l'entropie $\ln W_*$ peut être estimée en utilisant l'approximation de Stirling :

$$\ln n! \approx n \ln n - n,$$

dans laquelle les termes négligés sont extrêmement petits[(12)]. Dans le cas d'une

(11). Du moins dans la limite où n est « grand ». Lorsque la paroi est retirée, il est possible que le nombre de particules à gauche et à droite ne soit pas tout à fait n ; en pratique, le théorème limite central montre que les fluctuations autour de n sont extrêmement petites ; n'oublions pas que $\delta n/n \approx 1/\sqrt{n}$ est typiquement de l'ordre de 10^{-11} pour une expérience macroscopique.

(12). De l'ordre de $\ln n$, qui vaut quelques dizaines d'unités, tandis que $n \ln n$ est de l'ordre de 10^{25}...

enceinte séparée en deux, l'entropie est maintenant[13]

$$S_* = 2\ln\left(\frac{N^n}{n!}\right) \approx 2\big[n\ln N - n\ln n + n\big].$$

tandis que, dans l'enceinte sans paroi, elle vaut

$$S'_* = \ln\left(\frac{(2N)^{2n}}{(2n)!}\right) \approx 2n\,\ln N + 2n\ln 2 - 2n\ln n - 2n\ln 2 + 2n = S_*$$

c'est-à-dire que l'augmentation d'entropie est nulle (ou du moins d'ordre au plus $\ln n$, négligeable devant n) :

$$\delta S \approx 0.$$

Avec cette manière de compter, l'entropie est bien extensive et l'on peut à loisir ajouter ou retirer une paroi sans changer la valeur de S.

Remarque 29.7 Gibbs avait compris que l'introduction du facteur 1/N! permettait de sauver la situation, mais avait dû se contenter de postuler cet ajout, à défaut de pouvoir le justifier [38]. Ce n'est qu'avec la révolution de la mécanique quantique que l'abandon des propriétés « usuelles » des objets, et notamment leur discernabilité intrinsèque, a été possible. La validité du facteur 1/N! a fait l'objet de nombreuses discussions entre Planck, Einstein, Ehrenfest et Schrödinger [13].

§417 **Fermions et bosons**

Les particules élémentaires de même nature sont, par essence, indiscernables : le comptage des états doit en tenir compte et, comme vu avec le paradoxe de Gibbs, cela a une influence directe sur les propriétés macroscopiques des gaz. Le procédé de comptage mis en œuvre dans la section précédente est donc tout bonnement erroné — ainsi, *a priori*, que ses conclusions, le facteur de Boltzmann entre autres. Mais, dans les conditions usuelles de pression et de température, et pour des systèmes à grand nombre de particules, il se trouve que les résultats précédents constituent d'excellentes approximations, essentiellement parce que le nombre total d'états accessibles est encore plus gigantesque que celui du nombre de particules et que la probabilité qu'il se trouve plus d'une particule par état est négligeable (voir Brillouin [13] et la remarque 29.9 à la fin de ce chapitre).

Lorsque la température devient très basse, ou lorsque la pression devient très importante, le nombre d'états accessibles n'est plus assez grand et les effets statistiques de l'indiscernabilité deviennent visibles, parfois même à l'échelle macroscopique.

Nous pouvons mener quelques calculs explicites avec un (très) petit nombre de particules pour rendre compte de la différence entre statistique classique (Maxwell-Boltzmann, ou MB) et les statistiques quantiques. « Les » statistiques quantiques, car les particules peuvent se regrouper en deux grandes familles :

— celle des **bosons**, qui n'obéissent à d'autre règle que celle de l'indiscernabilité, et dont la statistique porte le nom de **Bose-Einstein**[14] ;

(13). On remarquera que le terme supplémentaire $-2n\ln n + 2n$ ne dépend ni du volume de l'enceinte, ni de l'énergie du gaz, et ne modifie donc pas l'équation d'état, ni les propriétés thermodynamiques, d'un système fixé.

(14). Satyendranath Bose (1894–1974), physicien indien, qui découvrit la statistique des photons. Albert Einstein (1879–1955) étendit ses idées à d'autres particules, et prévit l'existence du phénomène de condensation de Bose-Einstein.

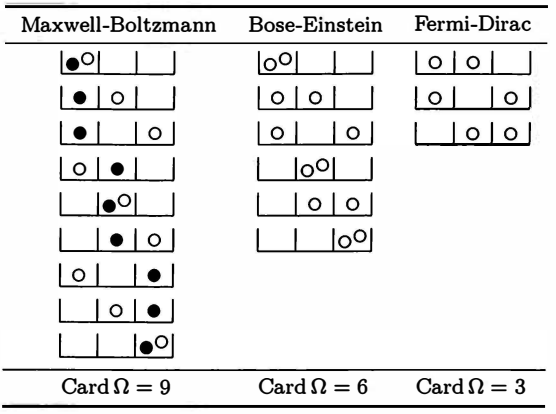


FIGURE 29.2 — *Nombre de possibilités de placer $n = 2$ particules dans $q = 3$ cases.*

— celle des **fermions**, qui obéissent de plus au *principe d'exclusion de Pauli* : un même état ne peut être occupé par plus d'une particule ; leur statistique porte le nom de **Fermi-Dirac**[15].

Sur la figure 29.2 apparaissent les différentes configurations microscopiques possibles pour ranger $n = 2$ particules dans $q = 3$ états individuels.

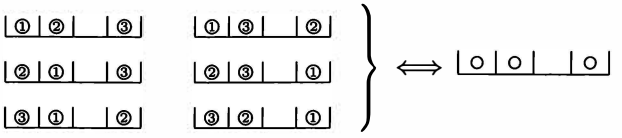
Il apparaît sur cet exemple que le nombre de configurations microscopiques en statistique de Maxwell-Boltzmann est plus grand que celui en statistique de Bose-Einstein, lui même plus grand que celui en statistique de Fermi-Dirac. De fait, c'est une règle générale, et le nombre de configuration dans chacune de ces statistiques est donné dans le tableau suivant :

Boltzmannions	Bosons	Fermions
q^n	$\binom{q+n-1}{n}$	$\binom{q}{n}$

- Le premier résultat est immédiat, puisque pour chaque particule, que l'on peut numéroter de 1 à n, nous avons q choix possibles sans restriction d'aucune sorte.
- Dans le cas des fermions, commençons par numéroter les particules (c'est-à-dire qu'on les rend momentanément discernables) ; nous pouvons placer la première particule dans une case au choix ; pour la suivante, nous n'avons plus que $q - 1$ choix, le premier étant exclu ; et ainsi de suite jusqu'à la dernière pour laquelle il ne reste plus que $q - n + 1$ choix. Le nombre de configurations trouvé est alors $q(q-1)\cdots(q-n+1)$. Cependant, si l'on considère désormais que les particules sont indiscernables, on voit que chaque configuration réelle se trouve représentée $n!$ fois, $n!$ étant le nombre

(15). Enrico FERMI (1901–1954), physicien italien, auteur de contributions majeures dans des domaines allant de la physique théorique à la physique nucléaire. Paul DIRAC (1902–1984), physicien britannique, un des plus grands théoriciens du vingtième siècle, est l'un des pères fondateurs de la mécanique quantique, le découvreur de l'équation quantique relativiste de l'électron et de l'antimatière.

de façon de numéroter n particules :

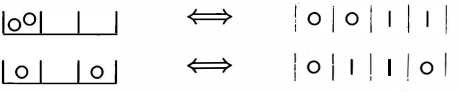


Ainsi, le nombre total de configurations fermioniques distinctes est

$$\frac{q(q-1)\cdots(q-n+1)}{n!} = \binom{q}{n}.$$

On pouvait également dire qu'il suffisait de choisir n cases parmi les q pour caractériser exactement un état fermionique macroscopique, en raison de l'indiscernabilité des particules.

- Enfin, pour démontrer la formule concernant les bosons, il est avantageux de transformer le problème P du rangement des particules en un problème P′ équivalent. On considère n particules indiscernables et $q-1$ barrières également indiscernables, et on cherche le nombre de façons d'ordonner les particules et les barrières. Chaque configuration du problème P peut être mise en correspondance avec une configuration du problème P′, la correspondance étant bijective.



Ce problème est lui-même équivalent au problème P″ consistant à placer n particules indiscernables (les ronds) dans $n+q-1$ cases (les emplacements pouvant recevoir un rond ou un trait) ; le nombre de telles combinaisons est $\binom{n+q-1}{n}$.

§ 418 **Nombre d'états occupés**

Le système physique étudié étant toujours constitué de n particules ayant accès à q états individuels distincts, on s'intéresse au nombre d'états occupés par au moins une particule ou, ce qui sera plus facile, au nombre d'états vides de toute particule. Pour cela, on note A_i^k l'événement

$$\mathrm{A}_i^k : \quad \text{« l'état } i \text{ est occupé par } k \text{ particules »}$$

et I_i^k la fonction indicatrice de A_i^k. Le nombre d'états comportant k particules est donc

$$\mathrm{M}_k = \sum_{i=1}^{q} \mathrm{I}_i^k$$

ce qui permet de calculer l'espérance de M_k par linéarité.

- Dans un premier temps, supposons que les particules obéissent à la statistique de Maxwell-Boltzmann. Notons X_i le nombre de particules dans l'état i : c'est une variable aléatoire suivant une loi binomiale de paramètres n et $p = 1/q$. Comme $\mathrm{A}_i^k = \{\mathrm{X}_i = k\}$,

$$\mathbf{E}(\mathrm{I}_i^k) = \mathbf{P}(\mathrm{A}_i^k) = \mathbf{P}\{\mathrm{X}_i = k\} = \binom{n}{k}\left(\frac{1}{q}\right)^k\left(1-\frac{1}{q}\right)^{n-k}$$

et

$$\mathbf{E}(\mathrm{M}_k) = q\binom{n}{k}\left(\frac{1}{q}\right)^k\left(1-\frac{1}{q}\right)^{n-k}$$

Notamment, le nombre moyen d'états vides est

$$\mathbf{E}(\mathrm{M}_0) = q\left(1 - \frac{1}{q}\right)^n. \qquad \textit{Maxwell-Boltzmann}$$

Remarque 29.8 On peut d'ailleurs déterminer aisément la loi de M_0 dans la statistique de Maxwell-Boltzmann : on a $\mathrm{M}_0 = i$ quand les n particules sont réparties dans $q - i$ états exactement, le nombre de possibilités étant $\binom{q}{i}(q-i)!\, s_n^{q-i}$, où s_n^{q-i} est le nombre de Stirling de seconde espèce. Ainsi,

$$\mathbf{P}\{\mathrm{M}_0 = i\} = \frac{\binom{q}{i}(q-i)!\, s_n^{q-i}}{q^n}.$$

- Cherchons maintenant, en statistique de Bose-Einstein, la loi de M_0. Il existe $\binom{q}{i}$ façons de choisir i états qui seront vides parmi les q. Pour chacun de ces choix, il faut commencer par mettre une particule dans chacune des $q - i$ cases ; ensuite, on peut librement répartir les $n - q + i$ particules restantes parmi les $q-i$ cases, ce qui, en statistique bosonique, laisse $\binom{n-1}{n-q+i}$ choix[16]. Au total, le nombre de configurations laissant exactement i cases vides est

$$\binom{q}{i} \cdot \binom{n-1}{n-q+i} = \binom{q}{i}\binom{n-1}{q-i-1},$$

et la probabilité d'avoir l'événement $\{\mathrm{M}_0 = i\}$ (« il y a exactement i états vides ») est le rapport de ce nombre et du nombre total de configurations bosoniques :

$$\mathbf{P}\{\mathrm{M}_0 = i\} = \frac{\binom{q}{i} \cdot \binom{n-1}{q-i-1}}{\binom{q+n-1}{n}}.$$

Le nombre moyen d'états vides est alors

$$\mathbf{E}(\mathrm{M}_0) = \sum_{i=0}^{q-1} i \cdot \frac{\binom{q}{i} \cdot \binom{n-1}{q-i-1}}{\binom{q+n-1}{n}}. \qquad \textit{Bose-Einstein}$$

- Pour terminer, lorsque l'on place n fermions dans q états, le nombre d'états vides est exactement $n - q$:

$$\mathbf{E}(\mathrm{M}_0) = n - q. \qquad \textit{Fermi-Dirac}$$

On trace, sur la figure 29.3 page suivante, le nombre moyen d'états occupés, c'est-à-dire $q - \mathbf{E}(\mathrm{M}_0)$, en fonction de n, pour chacune des statistiques et pour une même valeur $q = 50$. On s'aperçoit que les bosons occupent moins vite les espaces vacants que les « boltzmannions » ; ce qui veut dire qu'ils vont avoir tendance à davantage occuper un plus petit nombre d'états. Cette propriété, de nature purement combinatoire, est caractéristique des bosons, et devient d'autant plus flagrante que l'on travaille avec des grands nombres. Elle est responsable du comportement « grégaire » des bosons : lorsqu'un grand nombre de bosons occupent un même état individuel, et qu'un nouveau boson entre dans le système, il a, comparativement à un « boltzmannion », beaucoup plus de chances d'aller dans l'état déjà bien peuplé que dans un autre. Ce comportement grégaire, lorsqu'il a lieu à l'échelle macroscopique, explique[17] des phénomènes divers comme la supraconductivité, la superfluidité, les rayons laser, etc.

(16). Le raisonnement est le même que celui qui a permis de dresser le petit tableau de la page 599.

(17). Hmm... un peu de travail est nécessaire quand même !

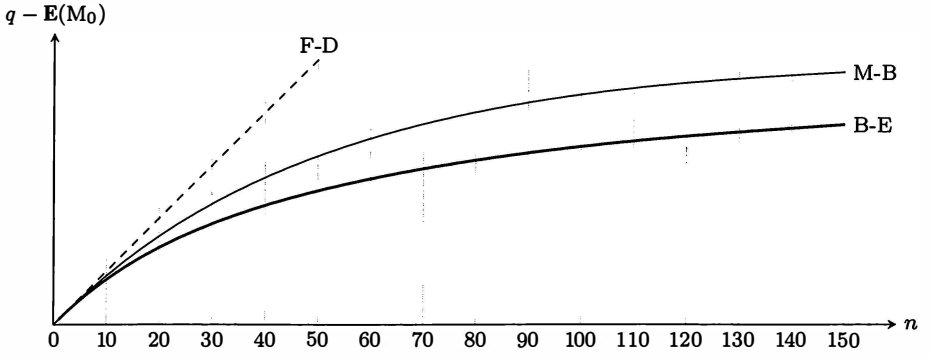


FIGURE 29.3 — *Nombre moyen* $q - \mathbf{E}(M_0)$ *d'états occupés par n particules parmi* $q = 50$ *états accessibles, selon la statistique : Fermi-Dirac (F-D), Bose-Einstein (B-E) ou Maxwell-Boltzmann (M-B).*

Remarque 29.9 (Grandes dilutions) Aux grandes dilutions (qui sont les conditions usuelles d'un gaz), on remarque que les trois courbes sont asymptotiquement les mêmes ; en effet, le nombre de particules est très petit devant le nombre d'états accessibles ($n \ll q$) et le nombre d'occupation moyen devient très inférieur à l'unité. Heuristiquement, le principe d'exclusion de Pauli joue peu (les bosons et les fermions diffèrent peu), et le nombre de complexions de la grande majorité des états macroscopiques est égal à 1 (les bosons et les boltzmannions diffèrent peu).

Chapitre 30

Processus de Poisson

(brève introduction)

Dans ce court chapitre, nous présentons un premier exemple simple de processus stochastique (un second exemple plus compliqué, le mouvement brownien, fait l'objet du chapitre suivant). Un processus de Poisson est un cas particulier de processus de comptage : c'est une famille $\{N(t)\,;\, t \geqslant 0\}$ de variables aléatoires à valeurs entières, qui « compte » le nombre d'occurrences d'un certain type d'événement avant l'instant t. Le lien, déjà aperçu plusieurs fois, entre variables exponentielles et variables de Poisson, trouve ici son expression la plus achevée. La théorie montre de plus que, sous des hypothèses générales, un processus de Poisson est une modélisation convenable de situations naturelles, ce qui explique la grande faveur dans laquelle est tenue ce type de processus.

En guise d'introduction : comment placer des points uniformément sur $\mathbb{R}^+$? § 419

Comment distribuer aléatoirement, et indépendamment les uns des autres, des points sur la demi-droite réelle $\mathbb{R}^+$, de telle sorte que leur densité soit (en moyenne) constante sur chaque intervalle ? Par indépendant, on entend ceci : que le nombre de points dans un intervalle $[a\,;\,b]$ soit indépendant du nombre de points dans un intervalle $[c\,;\,d]$ disjoint du premier.

La méthode naïve (placer les points les uns après les autres) échoue avant même de débuter : comment diable pourrait-on placer un point uniformément sur $[0\,;\,+\infty[$?

Une autre approche possible est la suivante. Notons λ la densité désirée. Il est bien entendu que, par densité, on entend : le nombre de points sur un intervalle $[a\,;\,b]$ est une variable aléatoire, d'espérance $(b-a)\lambda$. Si $n \geqslant 1$ est un entier, plaçons, uniformément et indépendamment, n points sur l'intervalle $[0\,;\,n/\lambda]$. À la limite $n \to \infty$, on devrait obtenir la modélisation désirée.

Le problème de cette deuxième méthode est qu'elle ne converge qu'en loi : chaque fois que l'on change n, il faut recalculer tous les points. Et, à n fixé, aucune condition n'est vraiment respectée : la densité n'est bien sûre égale à λ que sur $[0\,;\,n/\lambda]$, mais de plus la répartition des points ne satisfait pas à la condition d'indépendance. Par construction, le nombre total de points sur l'intervalle

$[0\,;\,n/\lambda]$ est égal à n, ce qui implique que s'il y en a beaucoup sur $[0\,;\,n/2\lambda]$, il y en aura peu sur $]n/2\lambda\,;\,n/\lambda]$.

Avant de continuer, on se souviendra de deux propriétés déjà démontrées :

- si $[a\,;\,b]$ est un intervalle fixé, le nombre de points situés dans cet intervalle lors de la construction précédente suit une loi binomiale de paramètres n et $(b-a)\lambda/n$; nous savons que, en faisant tendre n vers l'infini, la loi limite obtenue est une loi de Poisson $\mathscr{P}\big(\lambda(b-a)\big)$;
- la distance entre deux points suit une loi exponentielle de paramètre λ (exercice 10.6) ; réciproquement, si la distance entre deux points suit $\mathscr{E}(\lambda)$, alors le nombre de points dans un intervalle suit une loi de Poisson de paramètre $(b-a)\lambda$ (voir § 369).

§ 420 **Processus de comptage**

Nous étudions ici un **processus stochastique réel**, c'est-à-dire une famille

$$\{\mathrm{N}(t)\ ;\ t \geqslant 0\}$$

de variables aléatoires, définies sur un même espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ et à valeurs réelles, permettant de compter, au cours du temps, le nombre de réalisations d'un événement donné. Le mot *événement* reprend ici son sens usuel (et non son sens probabiliste d'élément de la tribu $\mathfrak{T}$) ; ce peut être par exemple le déclenchement d'un détecteur de rayons cosmiques, une panne dans un réseau informatique, la traversée d'un péage autoroutier par une voiture etc.

Un tel processus est appelé **processus de comptage** si les variables $\mathrm{N}(t)$ sont toutes à valeurs dans $\mathbb{N}$ et si, pour tout $\omega \in \Omega$, la fonction $t \mapsto \mathrm{N}(t,\omega)$ est croissante. La quantité $\mathrm{N}(t)$ compte le nombre d'événements, qu'on appelle également **points**, ayant eu lieu durant l'intervalle de temps $[0\,;\,t]$. Ainsi, $\mathrm{N}(t') - \mathrm{N}(t)$ est le nombre de points ayant lieu dans l'intervalle $]t\,;\,t']$.

Dans le paragraphe suivant, nous fabriquons un processus de comptage qui soit tel que $\mathrm{N}(t') - \mathrm{N}(t)$ suive une loi de Poisson de paramètre $(t'-t)\lambda$, de telle sorte que le nombre moyen de points dans l'intervalle $[t\,;\,t']$ est bien $(t'-t)\lambda$.

§ 421 **Processus de Poisson**

Soit $(\mathrm{T}_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes de même loi exponentielle de paramètre $\lambda > 0$; on note $\mathrm{S}_n = \mathrm{T}_1 + \mathrm{T}_2 + \cdots + \mathrm{T}_n$ pour tout $n \geqslant 1$, et $\mathrm{S}_0 = 0$. On définit enfin la famille de variables aléatoires $\{\mathrm{N}(t)\ ;\ t \geqslant 0\}$ comme le processus de comptage associé aux *points* $\mathrm{S}_1, \mathrm{S}_2, \ldots, \mathrm{S}_n, \ldots$, c'est-à-dire que l'on pose

$$\mathrm{N}(t) := \sum_{n=1}^{\infty} \mathbf{1}_{\{\mathrm{S}_n \leqslant t\}}.$$

La famille $\{\mathrm{N}(t)\ ;\ t \geqslant 0\}$ est appelée **processus de Poisson d'intensité λ**. Notamment, pour tout $t > 0$ et tout entier $n \geqslant 0$ on a (voir figure 30.1) :

$$\{\mathrm{S}_n \leqslant t\} = \{\mathrm{N}(t) \geqslant n\}\,.$$

La relation $\mathbf{P}\{\mathrm{N}(t) = n\} = \mathbf{P}\{\mathrm{S}_n \leqslant t\} - \mathbf{P}\{\mathrm{S}_{n+1} \leqslant t\}$ permet alors d'obtenir la loi de $\mathrm{N}(t)$ grâce au résultat classique suivant (démontré dans l'exercice 12.10):

■ **Lemme 30.1** *La variable S_n suit une loi $\Gamma(n,\lambda)$, c'est-à-dire qu'elle admet pour densité la fonction f nulle sur $\mathbb{R}^-$ et valant*

$$f_n(x) = \frac{(\lambda x)^{n-1}}{(n-1)!}\,\lambda\,\mathrm{e}^{-\lambda x}.$$

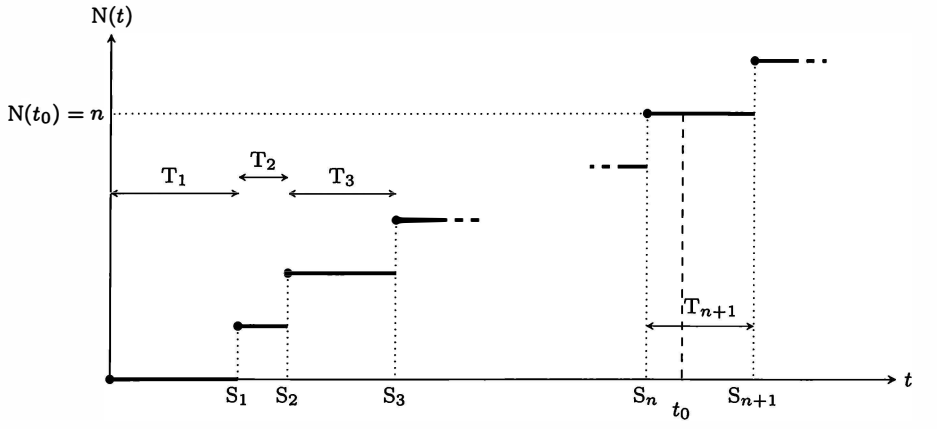


FIGURE 30.1 — *Relations entre les points* S_n, *les durées inter-arrivées* T_n *et les variables* $N(t)$.

■ **Proposition 30.2** *Pour tout* $t_0 > 0$, *la variable* $N(t_0)$ *suit la loi de Poisson de paramètre* λt_0 *(voir figure 30.2).*

DÉMONSTRATION : On a

$$\mathbf{P}\{N(t_0) = n\} = \int_0^{t_0} \left[\frac{(\lambda x)^{n-1}}{(n-1)!} - \frac{(\lambda x)^n}{n!}\right] \lambda e^{-\lambda x}\, dx$$
$$= \left[\frac{(\lambda x)^n}{n!} e^{-\lambda x}\right]_0^{t_0} = \frac{(\lambda t_0)^n}{n!} e^{-\lambda t_0}.$$

∎

La propriété essentielle du processus de Poisson est qu'il est **à accroissements indépendants** :

■ **Théorème 30.3** *Si* $I_1, I_2, \ldots, I_n$ *sont des intervalles disjoints, alors les nombres d'événements ayant lieu dans ces intervalles sont indépendants.*

Exemple 30.4 Si $I_k =]\, a_k\, ;\, b_k\,]$ pour $k = 1, 2, 3$, les variables

$$N(b_1) - N(a_1), \qquad N(b_2) - N(a_2) \qquad \text{et} \qquad N(b_3) - N(a_3)$$

sont indépendantes.

Cette propriété revient à dire que, si $t < t'$, alors $N(t') - N(t)$ est indépendant de $N(s)$ pour tout $s < t$. En d'autres termes, le processus de Poisson est un processus sans mémoire. De plus, il est à **accroissements stationnaires** :

■ **Théorème 30.5** *Si* $s, t > 0$, *alors* $N(s+t) - N(t)$ *a même loi que* $N(s)$.

Notamment, *la loi du nombre de points dans un intervalle donné ne dépend que de la longueur de cet intervalle.*

Ces propriétés découlent plus ou moins rapidement des définitions ; le lecteur intéressé pourra en trouver des démonstrations par exemple dans Grimmett & Stirzaker [44] ou, en français, dans Foata & Fuchs [34].

Caractérisation des processus de Poisson § 422

Les deux propriétés précédentes (accroissements indépendants et stationnaires)

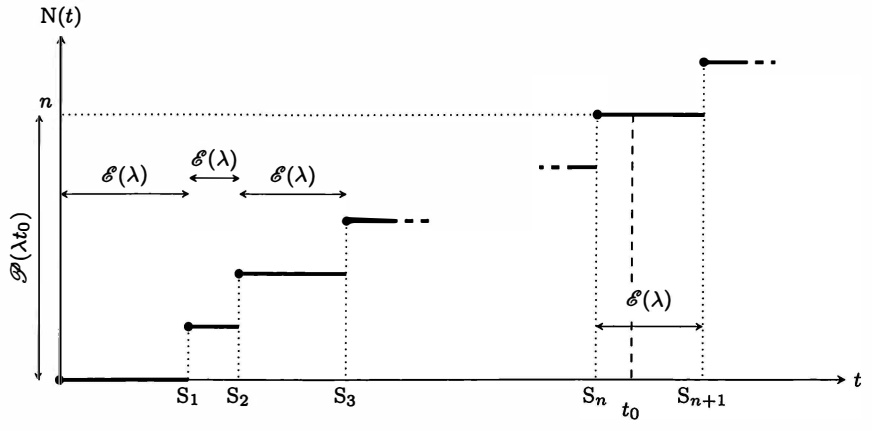


FIGURE 30.2 — *Loi exponentielle et loi de Poisson dans un processus de Poisson.*

résultent essentiellement du mode de construction proposé pour le processus de Poisson. Ce qui est remarquable, c'est qu'elles caractérisent les processus de Poisson.

■ **Théorème 30.6** *Soit* $\{N(t), t \geqslant 0\}$ *un processus de comptage. On suppose que ce processus est à accroissements indépendants et stationnaires :*

$$\forall s, t > 0 \qquad N(t+s) - N(s) \rightsquigarrow N(s)$$

et $N(t+s) - N(t)$ *est indépendant de tous les* $N(r)$, *pour* $r \in [0\,;\, t]$. *Alors c'est un processus de Poisson, de paramètre* $\mathbf{E}N(1)$.

§ 423 **Processus marqués**

La propriété précédente permet par exemple d'étudier les **processus marqués**. L'idée est la suivante : parmi tous les points d'un processus de Poisson, on n'en sélectionne que quelques uns, avec une probabilité p, et de manière indépendante du processus lui-même (on peut imaginer qu'on les marque en rouge, par exemple). Formellement, on choisit une suite $(Y_n)_{n\geqslant 1}$ de variables aléatoires, indépendantes entre elles et indépendantes de la famille $\{N(t), t \geqslant 0\}$, de même loi de Bernoulli $\mathscr{B}(p)$, et on pose

$$N^*(t) = \sum_{n=1}^{\infty} \mathbf{1}_{\{S_n \leqslant t, Y_n = 1\}}.$$

La quantité $N^*(t)$ compte bien le nombre de points marqués en rouge.

Le résultat obtenu est encore un processus à accroissements indépendants et stationnaires, c'est donc un processus Poisson. Pour calculer son intensité, on écrit

$$\begin{aligned}
\mathbf{E}N^*(1) &= \sum_{n=1}^{\infty} \mathbf{E}\mathbf{1}_{\{S_n \leqslant 1, Y_n = 1\}} = \sum_{n=1}^{\infty} \mathbf{E}\big(\mathbf{1}_{\{S_n \leqslant 1\}}\, \mathbf{1}_{\{Y_n = 1\}}\big) \\
&= \sum_{n=1}^{\infty} \mathbf{E}\big(\mathbf{1}_{\{S_n \leqslant 1\}}\big)\, \mathbf{E}\big(\mathbf{1}_{\{Y_n = 1\}}\big) \qquad \textit{par indépendance} \\
&= p \sum_{n=1}^{\infty} \mathbf{E}\big(\mathbf{1}_{\{S_n \leqslant 1\}}\big) = p \sum_{n=1}^{\infty} \mathbf{P}\{N_1 \geqslant n\} = p\,\mathbf{E}(N_1) = p\lambda.
\end{aligned}$$

■ **Théorème 30.7 (Poissons rouges)** *Soit* $\{N(t), t \geqslant 0\}$ *un processus de Poisson, d'intensité* λ*, que l'on marque au moyen d'une suite de variables de Bernoulli indépendantes entre elles et indépendantes de* $\{N(t), t \geqslant 0\}$*. Alors le processus marqué* $\{N^*(t), t \geqslant 0\}$ *est un processus de Poisson d'intensité* $p\lambda$*.*

Distribution des instants d'arrivée § 424

Le résultat suivant justifie la discussion du début du chapitre (« comment distribuer des points uniformément ? »)

■ **Théorème 30.8** *Conditionnellement à* $\{N(b) - N(a) = n\}$*, la loi de distribution des points apparaissant dans l'intervalle* $[a\,;\,b]$ *est la loi de* n *variables aléatoires indépendantes de même loi uniforme sur* $]a\,;\,b]$*.*

Ce résultat, pris à rebours, permet de simuler localement un processus de Poisson d'intensité λ sur $]a\,;\,b]$: on tire d'abord un nombre n selon une loi de Poisson de paramètre $(b-a)\lambda$, puis on distribue aléatoirement, et indépendamment, des points $(x_1, x_2, \ldots, x_n)$ sur $]a\,;\,b]$.

Effet physique : le bruit de grenaille § 425

Puisque $N(t)$ suit la loi de Poisson $\mathscr{P}(\lambda t)$, on en déduit que :

■ **Théorème 30.9** *Soit* $\{N(t), t \geqslant 0\}$ *un processus de Poisson d'intensité* λ*. Alors*

$$\frac{\sigma_{N(t)}}{\mathbf{E}N(t)} = \frac{1}{\sqrt{\lambda t}}.$$

On en tire deux conséquences, dans les limites respectives $t \to 0$ et $t \to \infty$:

— Lorsque t est grand, les fluctuations du nombre d'événements dans l'intervalle $[0\,;\,t]$ sont très petites devant le nombre total. Autrement dit, on accède à λ par l'estimateur $N(t)/t$, ce qui est la conclusion banale de la loi des grands nombres.

— Plus intéressant, lorsque $t \to 0$, les fluctuations deviennent au contraire très importantes devant le nombre moyen attendu. Autrement dit, aux temps courts devant $1/\lambda$, les fluctuations de $N(t)$ (le *bruit*) deviennent très grandes devant la valeur de $\mathbf{E}N(t)$ (le *signal*). C'est ce qui cause l'effet connu sous le nom de *bruit de grenaille*[1] : le courant passant à travers un conducteur n'est pas dû à un phénomène continu, mais au passage d'électrons. Le nombre d'électrons traversant le conducteur pendant un intervalle de temps Δt suit une loi de Poisson de paramètre $I\Delta t/e$, où I est l'intensité du courant et e la charge de l'électron. Lorsque Δt est petit devant e/I, les fluctuations de $N(\Delta t)$ sont grandes devant ce nombre moyen.

Âge et temps de vie résiduel § 426

On considère un processus de Poisson $\{N(t)\ ;\ t \geqslant 0\}$ et on choisit un instant particulier t. Le nombre de points ayant eu lieu avant t est $N(t)$, le dernier point en date ayant eu lieu à l'instant $S_{N(t)}$; le point suivant est $S_{N(t)+1}$. Ainsi, t est dans l'intervalle $[S_{N(t)}\,;\,S_{N(t)+1}[$.

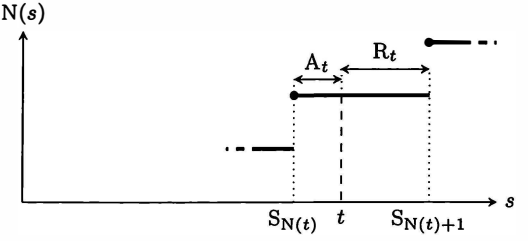
Appelons **temps de vie écoulé**, ou plus simplement **âge** à l'instant t, la quantité

$$A_t := t - S_{N(t)},$$

(1). Ou bruit Schottky, du nom du physicien Walter Schottky qui l'a découvert en 1918.

et **temps de vie résiduel**

$$R_t := S_{N(t)+1} - t.$$



Remarque 30.10 Ces dénominations viennent de la modélisation suivante : considérons une lampe que l'on munit d'une ampoule ayant une durée de vie suivant la loi exponentielle $\mathscr{E}(\lambda)$. Dès que l'ampoule claque, on la remplace par une neuve, de durée de vie suivant la même loi et indépendante de la première. Les changements d'ampoule sont les points d'un processus de Poisson. La variable A_t représente l'âge de l'ampoule présente à l'instant t, et R_t le temps qui lui reste à vivre avant de claquer.

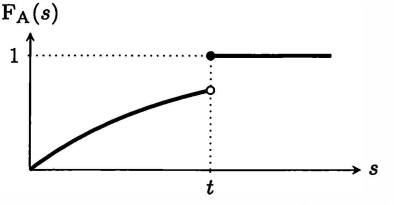
On connaît les propriétés d'absence de mémoire de la loi exponentielle : on en déduit que R_t suit la même loi $\mathscr{E}(\lambda)$ que les variables T_k. Ce qui est remarquable, c'est que la loi de l'âge est également une loi exponentielle — ou presque : par construction, l'âge à l'instant t ne saurait dépasser t.

■ **Proposition 30.11** *La variable* A_t *suit la loi de* $T_1 \wedge t$.

DÉMONSTRATION : On sait déjà que, si $s \geqslant t$, alors $\mathbf{P}\{A_t > s\} = 0$. Soit donc $s \in [0\,;\, t[$. L'événement $\{A_t > s\}$ est égal à l'événement $\{S_{N(t)} < t - s\}$, qui est encore égal à l'événement « il n'y a aucun point entre $t-s$ et t. Puisque le processus de Poisson est à accroissements indépendants, cet événement a même probabilité que l'événement « il n'y a aucun point entre 0 et s », à savoir $\{S_1 > s\} = \{T_1 > s\}$: $\mathbf{P}\{A_t > s\} = \mathbf{P}\{T_1 > s\} = e^{-\lambda s}$.

La fonction de répartition de A_t est donc

$$F_A(s) = \begin{cases} 1 - e^{-\lambda s} & \text{si } 0 \leqslant s < t, \\ 1 & \text{si } s > t. \end{cases}$$



△30.2 Cette fonction de répartition est exactement celle de $T_1 \wedge t$ (voir exemple 13.14). ▮

§ 427 **Processus de Poisson sur $\mathbb{R}$**

Dans le processus de Poisson tel qu'il a été décrit, on a pu « parsemer » les points, indépendamment les uns des autres, sur $\mathbb{R}^+$. On peut étendre cette méthode à $\mathbb{R}$ tout entier. Mais d'abord une question :

♦ ***Exercice 30.1*** On définit un processus de Poisson $\{N(t)\,;\, t \geqslant 0\}$ sur $\mathbb{R}^+$, puis un second, indépendant du premier, sur $\mathbb{R}^-$, par une simple opération de symétrie : si $\{M(t)\,;\, t \geqslant 0\}$ est un processus de Poisson, on pose $N(-t) = -M(t)$ pour tout $t \geqslant 0$. Le processus $(N(t)\,;\, t \in \mathbb{R})$ est-il un processus à accroissements indépendants ?

On simule ainsi un processus de Poisson sur $\mathbb{R}$:

— On part de zéro et l'on crée un processus de Poisson sur $\mathbb{R}^+$ selon la méthode du paragraphe § 421.

— On part du premier point du processus précédent, et on crée un processus de Poisson en remontant le temps à partir de ce point.

Processus de Poisson sur $\mathbb{R}^d$ § 428

On généralise la notion de processus de Poisson à l'espace $\mathbb{R}^d$: comme dans le cas de la droite réelle, on désire disséminer des points sur $\mathbb{R}^d$, à densité constante, et indépendamment les uns des autres.

On considère donc une fonction aléatoire $\Pi : \Omega \to \mathfrak{P}(\mathbb{R}^d)$ qui, à chaque réalisation ω du hasard, associe une partie $\Pi(\omega)$ de $\mathbb{R}^d$. On dira qu'une telle fonction Π est un **processus de Poisson sur $\mathbb{R}^d$ d'intensité λ** si $\Pi(\omega)$ est au plus dénombrable pour tout $\omega \in \mathbb{R}$, et si :

- pour tout borélien $B \in \mathfrak{B}^d$, la variable aléatoire
$$N(B) := \mathrm{Card}\,(\Pi \cap B)$$
suit la loi de Poisson de paramètre $\lambda\mu(B)$ (où $\mu(B)$ est le volume de B, obtenu avec la mesure de Lebesgue sur $\mathbb{R}^d$) ;
- pour tout entier n et pour tous boréliens $B_1, B_2, \ldots, B_n$ deux à deux disjoints, les variables $N(B_1), N(B_2), \ldots, N(B_n)$ sont deux à deux indépendantes.

Remarque 30.12 (Simuler informatiquement un processus de Poisson sur $\mathbb{R}^2$)
Pour simuler une loi approchée, limitée à un domaine borné A du plan, il est toujours possible de choisir un domaine B de $\mathbb{R}^2$ « beaucoup plus grand que A », et de forme simple (par exemple un rectangle) et de disséminer au hasard sur B un nombre de points égal à $n = \lambda\mu(B) \gg 1$ (ou, du moins, la partie entière de ce nombre). Le nombre de points dans A suivra une loi binomiale $\mathscr{B}(n\,;\,p)$, où $p = \mu(A)/\mu(B)$. On sait (section 22.1) que cette loi est proche de la loi de Poisson $\mathscr{P}(np) \approx \mathscr{P}(\lambda\mu(A))$.

Une simulation *exacte* d'un processus de Poisson peut se faire ainsi. On commence par créer un processus de Poisson sur $\mathbb{R}^+$, d'intensité $\mu = \pi\lambda$, dont on note $S_1, S_2, \ldots$ les points. On fabrique ensuite une suite $(\Theta_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, de même loi uniforme sur $[0\,;\,2\pi[$. On pose enfin
$$\mathbf{X}_n = (X_n, Y_n) = \left(\sqrt{S_n}\cos\Theta_n, \sqrt{S_n}\sin\Theta_n\right).$$
L'idée derrière cette construction est que, presque sûrement, $S_n/(n/\mu) \to 1$, c'est-à-dire que $\sqrt{S_n}$ est presque sûrement équivalente à $\sqrt{n/\mu}$. C'est géométriquement ce que l'on désire : les n premiers points tiennent sur un disque de rayon $\sqrt{S_n}$, d'aire $\approx \pi n/\mu = n/\lambda$.

Un calcul relativement élémentaire montre que l'ensemble des points $\{\mathbf{X}_n\ ;\ n \in \mathbb{N}\}$ vérifie bien les deux propriétés d'un processus de Poisson plan.

EXERCICES

♦ ***Exercice 30.2*** Que dire de la loi de la durée de vie et de l'âge pour un processus de Poisson sur $\mathbb{R}$?

♦ ***Exercice 30.3 (Paradoxe de l'autobus, ou de l'inspection (2/2))***
Montrer que les variables A_t et R_t définies dans le paragraphe § 426 sont indépendantes. Calculer la longueur moyenne de l'intervalle $\left[S_{N(t)}\,;\,S_{N(t)+1}\right]$ comprenant t. Commenter. *(On pourra comparer à l'exercice 2.1 page 72).*

Indications

◊ *Indication 30.1 :* La distance entre le premier point positif et le premier point négatif ne suit pas la loi exponentielle $\mathscr{E}(\lambda)$, mais la loi $\mathscr{E}(2\lambda)$.

SOLUTIONS

♦ **Solution 30.2** Elles sont toutes deux égales à la loi de X_1, c'est-à-dire la loi exponentielle $\mathscr{E}(\lambda)$.

♦ **Solution 30.3** On peut directement calculer l'espérance de A_t avec sa fonction de répartition :

$$\mathbf{E}(A_t) = \int_0^{+\infty} [1 - F_{A_t}(s)]\, ds = \int_0^t e^{-\lambda s}\, ds = \frac{1 - e^{-\lambda t}}{\lambda}.$$

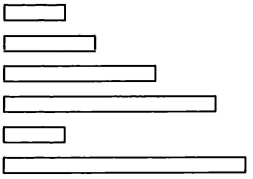
La variable R_t suivant la loi géométrique $\mathscr{E}(\lambda)$, son espérance est $1/\lambda$. Par linéarité,

$$\mathbf{E}\big(S_{N(t)+1} - S_{N(t)}\big) = \mathbf{E}(A_t + R_t) = \mathbf{E}(A_t) + \mathbf{E}(R_t) = \frac{2 - e^{-\lambda t}}{\lambda}.$$

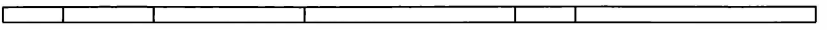
On note que cette espérance est strictement supérieure à $1/\lambda$, qui est la « longueur moyenne » d'un intervalle entre deux points — résultat qui semble évidemment paradoxal. Si l'on effectue l'« inspection » de la longueur d'un intervalle à un instant t éloigné de l'origine (c'est-à-dire tel que $\lambda t \gg 1$), les effets de bord dûs à cette origine s'estompent et l'espérance de la largeur de l'intervalle est pratiquement égale à $2/\lambda$.

On se retrouve dans le cas des inspecteurs d'ampoules de l'exercice 2.1 : la durée de vie de l'ampoule en fonction à l'instant (aléatoire) de l'inspection est supérieure à la durée de vie moyenne d'une ampoule.

Graphiquement, imaginons quelques intervalles typiques, équiprobables :



et mettons-les bout à bout :



Choisissons maintenant un instant t (*instant d'inspection*) de l'intervalle total : on a plus de chances de tomber à l'intérieur d'un grand intervalle que d'un petit[(2)]. Ainsi, une inspection induit un biais dans la mesure de la longueur moyenne observée des intervalles. Dans le cas d'une loi de Poisson, on double presque cette quantité. (Si l'on avait crée un processus de Poisson sur $\mathbb{R}$, ce qui revient à faire tendre t vers l'infini, la longueur observée serait exactement le double de la longueur moyenne.)

Remarque : ce résultat paradoxal (du moins, qui heurte le sens commun), est parfois présenté comme le *paradoxe de l'autobus*. On imagine que les instants d'arrivée d'un autobus forment un processus de Poisson ; on sait qu'un autobus passe en moyenne toutes les demi-heures. Si on rejoint l'arrêt à un moment quelconque de la journée, alors on attendra en moyenne une demi-heure (et non pas un quart d'heure), et le bus précédent sera passé, en moyenne, une demi-heure auparavant. Ainsi, en moyenne, on arrive durant un intervalle durant lequel il n'y a pas de bus pendant une heure. Ce modèle a l'agréable propriété de justifier nos sentiments paranoïaques ; signalons tout de même que les instants d'arrivée d'un autobus ne sont que très médiocrement modélisés par un processus de Poisson.

Conseils de lecture

- Carl GRAHAM et Denis TALAY, *Simulations numériques et méthodes de Monte-Carlo* [40].
- Geoffrey GRIMMETT et David STIRZAKER, *Probability and random processes* [44].

(2). Bien sûr, dans l'exercice proposé, t n'est pas choisi au hasard ; à cause de l'absence de mémoire des lois exponentielles, et si l'on est assez éloigné de l'origine, tout se passe cependant de la même manière.

Chapitre 31

Mouvements browniens

La théorie du mouvement brownien s'est, depuis un peu plus d'un siècle, développée de façon spectaculaire. De manière tout à fait remarquable, elle est au cœur de l'intérêt de nombreux mathématiciens et de physiciens ; si elle a fourni une modélisation satisfaisante du mouvement observé de fines particules et permis une détermination expérimentale du nombre d'Avogadro, son intérêt n'a cessé de croître quand d'éminents physiciens ont proposé une formulation entièrement nouvelle de la théorie quantique des champs par des intégrales dont les variables étaient des chemins browniens ; ces connexions très profondes entre mouvement brownien et théorie quantique des champs ont été étendues à la mécanique statistique, la gravitation quantique, la théorie de la percolation, etc. Le champ de recherche du mouvement brownien est gigantesque et en perpétuelle expansion, diverses disciplines (probabilités, analyse harmonique, physique théorique, mathématiques financières) s'y rencontrant et s'enrichissant mutuellement.

À cause de la très grande technicité des outils mis en jeu, toutefois, nous devrons nous contenter de proposer une simple introduction à ce vaste domaine, le but (avoué) étant d'abord de donner envie au lecteur de s'y plonger entièrement ; et puisque l'eau est fraîche, on commence par une partie heuristique simple.

31.1 UNE APPROCHE HEURISTIQUE

L'observation de Robert Brown

§ 429

Le botaniste anglais Robert Brown fit paraître, en 1828, un opuscule intitulé : *A brief account of microscopical observations [...] on the particles contained in the pollen of plants ; and on the general existence of active molecules.* Cherchant à élucider, durant l'été 1827, le problème du transport de particules jusqu'à l'ovule d'une plante, il observa au microscope, de très fines particules oblongues contenues *dans* les grains de pollen de *Clarkia pulchella* (une petite fleur rose découverte durant la fameuse expédition de Lewis & Clark, choisie parce que ces particules étaient de tailles inhabituellement grosses, et donc plus faciles à observer), et les trouva animées de mouvements désordonnés, indépendants les uns des autres et, surtout, perpétuels ; l'origine de ces mouvements lui était inconnue. Étaient-ils liés à la nature *vivante* d'une plante ? Il effectua des observations similaires sur d'autres espèces de plantes, puis chercha à savoir si cette propriété perdurait après la mort. Il constata successivement que :

- le mouvement était identique pour les particules du pollen de plantes séchées pendant plusieurs jours, une fois ces particules immergées dans une goutte d'eau ;
- de même pour des plantes séchées depuis plus d'un siècle ;
- de même pour d'autres tissus, animaux ou végétaux, une fois réduits en poudre et mis en suspension dans de l'eau ;
- de même pour des fossiles d'âges divers, dont les plus anciens semblaient ne plus contenir que des éléments minéraux.

Doutant alors du fait que ce mouvement soit une manifestation vitale, il multiplia ses expériences, utilisant de nombreux minéraux et métaux (dont un fragment du Sphinx, des tubes de silice formés par la foudre tombant sur du sable, et même des météorites de provenances diverses...)

Le mémoire de Brown (suivi, un an plus tard, d'un second, *Additional remarks on active molecules*), bien que faisant allusion à de nombreuses querelles de chapelles sur les doctrines vitalistes professées par les savants de l'époque et qui paraissent bien obscures au lecteur moderne, est un témoignage remarquable, analysant avec clairvoyance le processus de réflexion et de déduction d'un savant en interaction constante avec l'observation.

Précisons, avant de terminer cette brève note historique, que Brown ne fut pas le premier découvreur du mouvement brownien ; il cite notamment Leeuwenhoek[1] qui le confondait avec un mouvement d'origine vivante (*animalcular motion*), Needham[2] et Buffon ; il rapporte également les observations de John Bywater, qui publia en 1819 un compte-rendu de ses observations microscopiques

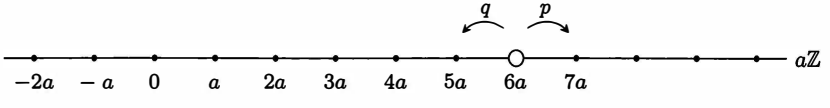
> *...in which it is stated that not only organic tissues, but also innorganic substances, consist of what he terms animated or irritable particles.*

Bywater conjecturait que le mouvement observé était d'origine thermique, la source d'énergie invoquée étant la lumière solaire.

Au XIX^e^ siècle, plusieurs séries d'expériences méticuleuses, dont celles de Gouy[3] prouvent que ce mouvement ne provient ni de l'éclairage, ni des gradients de température, ni de mouvements de convection dans le liquide. L'idée progresse qu'il pourrait avoir son origine dans la nature fondamentalement atomique de la matière ; la théorie cinétique de la chaleur permet alors d'expliquer le mouvement des particules par les chocs incessants des molécules d'eau environnantes.

§ 430 **Marche aléatoire discrète et passage à la limite continue**

Dans sa version la plus simple, à une dimension, la marche aléatoire décrit un marcheur (ou une particule), astreint à se déplacer sur un réseau régulier $\mathcal{R} = a\mathbb{Z}$ de pas $a > 0$, en temps discret. Il part de l'abscisse 0 à l'instant $t = 0$, et fait un pas de longueur a toutes les τ secondes : vers la droite avec une probabilité p, et vers la gauche avec une probabilité $q = 1 - p$.



(1). Antoni LEEUWENHOEK (1632–1723) fut — outre un marchand d'étoffes — un précurseur de la biologie cellulaire ; il améliora le microscope, dont il fabriqua lui-même plusieurs centaines d'exemplaires, et découvrit l'existence des spermatozoïdes.

(2). John Turberville NEEDHAM (1713–1781), prêtre et physicien anglais, effectua lui aussi de nombreuses observations microscopiques.

(3). Louis Georges GOUY, (1854–1926), physicien français. *« Le point le plus important est la généralité du phénomène ; des milliers de particules ont été observées, et dans aucun cas on n'a vu une particule en suspension qui n'offrît pas le mouvement habituel. »*

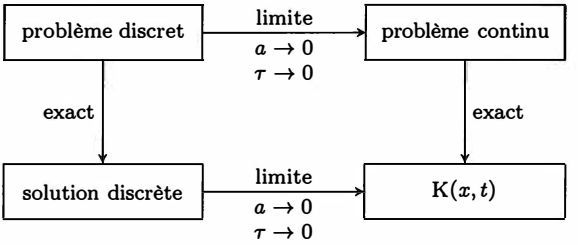
La marche est dite **symétrique** lorsque $p = q = \frac{1}{2}$, et **biaisée** dans le cas contraire.

Il existe deux manières de définir une « limite continue » pour une marche aléatoire.

— La première est de résoudre exactement le problème de la marche aléatoire discrète, de déterminer la probabilité de présence $\mathrm{K}(\ell a, n\tau)$ du marcheur en chaque point ℓa et pour chaque instant $n\tau$, puis de faire tendre le pas spatial a et le pas temporel τ vers 0 ; on en déduit une (densité de) probabilité continue $\mathrm{K}(x, t)$.

— La seconde est de transformer l'équation maîtresse sur la fonction K par une équation continue (qui approche donc l'équation exacte dans la limite choisie), puis de résoudre exactement cette équation continue.

On voit que les deux processus diffèrent fondamentalement. Dans le premier, on part de la solution exacte d'un problème discret, et on cherche la limite continue de cette solution. Dans le second, on transforme le problème discret en un problème continu, dont on cherche ensuite la solution exacte. Que les deux procédés (limite continue de la solution discrète à un problème discret, et solution d'un problème continu, limite d'un problème discret) mènent à un même résultat est tout à fait remarquable.



Limite continue — première approche § 431

Notons Y_n la variable aléatoire décrivant le déplacement élémentaire lors du n-ième pas, égale à 1 si le pas est vers la droite, et à -1 s'il est vers la gauche ; la position du marcheur à l'instant $n\tau$ est la variable $\mathrm{Z}_n := \mathrm{Y}_1 + \mathrm{Y}_2 + \cdots + \mathrm{Y}_n$, centrée et de variance n. D'après le théorème limite central, la suite de variables aléatoires centrées réduites, de terme général $\mathrm{Z}_n/\sqrt{n}$, converge en loi vers la loi normale $\mathscr{N}(0\ ;\ 1)$.

Introduisons maintenant les facteurs d'échelle. On se fixe une unité de temps $t^* = 1$, et on modifie l'échelle de temps discrets en notant $\tau := 1/n$; de même, on impose au pas a de tendre vers 0 lorsque n tend vers l'infini (de manière que nous établirons prochainement). Pour un entier n donné, la position du marcheur au bout de n pas (c'est-à-dire à l'instant *fixe* $t^* = n\tau = 1$) est

$$\mathrm{X}_n = a\,\mathrm{Z}_n$$

et la suite des variables centrées réduites $\mathrm{X}_n^* := \mathrm{X}_n/a\sqrt{n}$ converge en loi[4] vers la loi normale $\mathscr{N}(0\ ;\ 1)$.

Ici intervient le choix de l'échelle spatiale : afin d'obtenir un résultat intéressant, on impose que $a\sqrt{n} = a/\sqrt{\tau}$ soit une constante ; pour des raisons qui apparaîtront

(4). On se souvient que la différence de normalisation des variables X_n (le préfacteur a dépend de n) n'interfère pas avec le théorème limite central, puisque le processus impose le passage par une suite de variables centrées et réduites, ce qui élimine *de facto* tous les choix de normalisation.

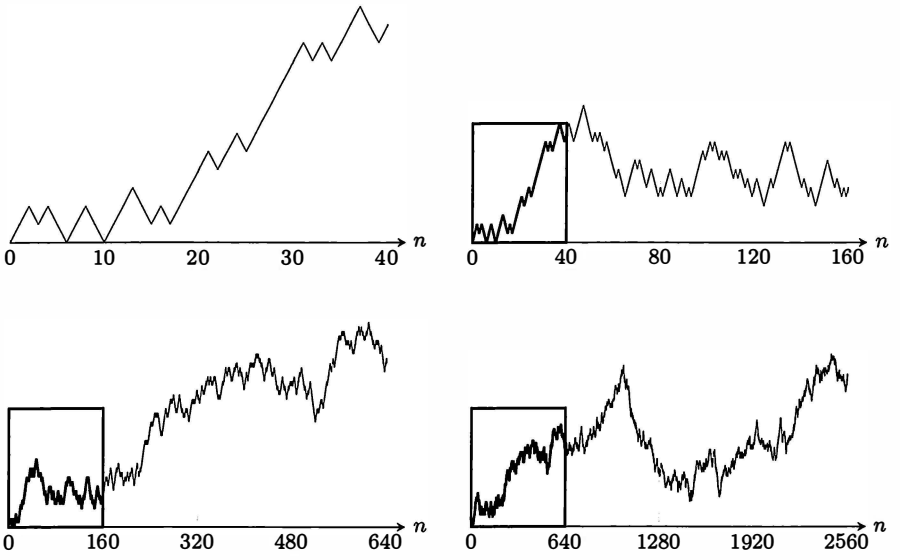


FIGURE 31.1 — *Une même marche aléatoire à $n = 40$, $n = 160$, $n = 640$ et $n = 2560$ pas. Dans les trois dernières figures, on a encadré d'un rectangle la partie de la marche aléatoire qui était visible dans la précédente. L'échelle des abscisse est multipliée par 4 à chaque étape, celle des ordonnées par 2. L'allure générale de la courbe change peu entre la 3e et la 4e figure.*

dans le paragraphe suivant, on note $\sqrt{2D}$ cette constante, c'est-à-dire que l'on impose la relation $a = \sqrt{2D/n}$, ou encore

$$D = \frac{a^2}{2\tau}.$$

La variable X_n suit donc approximativement la loi $\mathcal{N}(0\,;\,2D)$. Notamment, sous l'effet d'un changement d'échelle spatiale $a \leftarrow \sqrt{\lambda}a$ et temporel $t \leftarrow \lambda t$, cette loi n'est pas modifiée. Sur la figure 31.1, la loi limite est bien approchée pour des grandes valeurs de n, l'allure de la courbe est très peu modifiée entre le troisième graphe et le quatrième.

Une fois ce travail effectué, on peut, en maintenant le lien entre a et τ fixé, regarder la position du marcheur à un instant fixé t quelconque ; les calculs sont identiques, la position X_t du marcheur à l'instant t, dans l'échelle spatiale renormalisée, converge en loi vers la loi $\mathcal{N}(0\,;\,2Dt)$, donc admet une densité de probabilité

$$\forall x \in \mathbb{R} \quad \forall t > 0 \qquad K(x,\,t) = \frac{1}{\sqrt{4\pi D t}}\,\mathrm{e}^{-x^2/4Dt}. \tag{31.1}$$

§ 432 **Limite continue — seconde approche**

On revient à une marche aléatoire symétrique, discrète, à une dimension sur un réseau de pas a ; l'intervalle de temps entre deux sauts est noté τ. Notons $K(x,t)$ la probabilité que le marcheur soit à l'abscisse x à l'instant t. Pour tout $N \geqslant 0$ et tout $\ell \in \mathbb{Z}$, on a donc

$$K\big(\ell a\,,\,(N+1)\tau\big) = \tfrac{1}{2}K\big((\ell-1)a\,,\,N\tau\big) + \tfrac{1}{2}K\big((\ell+1)a\,,\,N\tau\big).$$

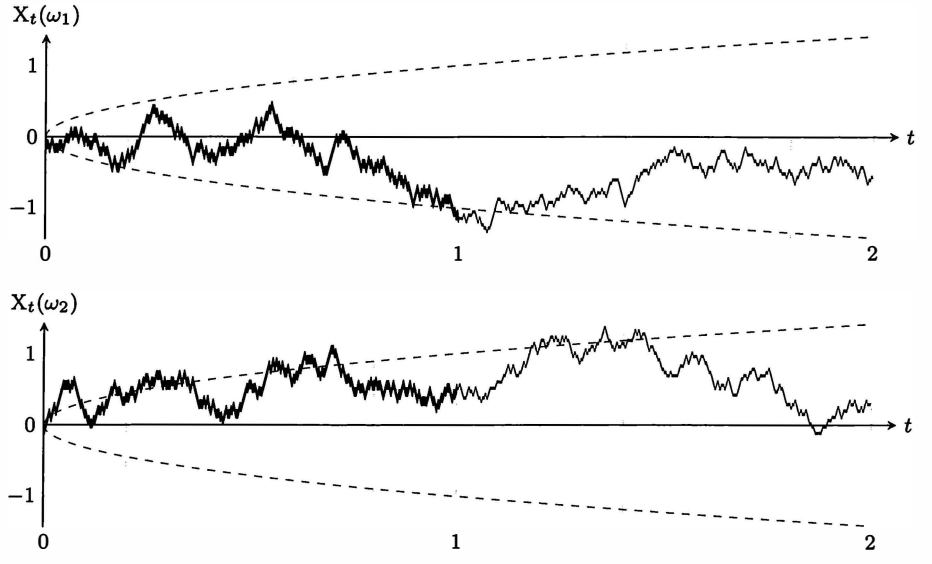


FIGURE 31.2 — *Pour d'autres choix de ω, la marche aléatoire a une allure très différente. Pour chacune de ces marches, n est grand ($n = 2000$), et t varie entre 0 et 1 (en gras) puis entre 1 et 2 (en maigre). La parabole d'équation $y^2 = t$ est représentée conjointement ; elle représente la valeur de l'écart-type de X_t au cours du temps. À chaque instant, le mouvement a environ 68 % de chances d'être entre les deux branches de la parabole.*

Effectuons maintenant un changement d'échelle pour obtenir la limite continue de ces équations : on suppose donc que a et τ tendent vers zéro (mais pas forcément de façon indépendante !) et on remplace l'équation précédente par une équation approchée, en supposant simplement que la fonction K est suffisamment régulière pour admettre un développement de Taylor à l'ordre 1 en t, et à l'ordre 2 en x :

$$\mathrm{K}\big((\ell+1)a\,,\,\mathrm{N}\tau\big) = \mathrm{K}\big(\ell a\,,\,\mathrm{N}\tau\big) + a\,\frac{\partial \mathrm{K}}{\partial x}\big(\ell a,\ \mathrm{N}\tau\big) + \frac{a^2}{2}\,\frac{\partial^2 \mathrm{K}}{\partial x^2}\big(\ell a,\ \mathrm{N}\tau\big) + \cdots$$

$$\mathrm{K}\big(\ell a\,,\,(\mathrm{N}+1)\tau\big) = \mathrm{K}\big(\ell a\,,\,\mathrm{N}\tau\big) + \tau\,\frac{\partial \mathrm{K}}{\partial t}\big(\ell a,\ \mathrm{N}\tau\big) + \cdots$$

ce qui mène classiquement à

$$\frac{\partial \mathrm{K}}{\partial t} \approx \frac{a^2}{2\tau}\,\frac{\partial^2 \mathrm{K}}{\partial x^2}.$$

Une limite continue non triviale n'est obtenue qu'en faisant tendre simultanément a et τ vers 0, et en supposant que la quantité $a^2/2\tau$ converge vers une limite finie

$$\mathrm{D} = \lim_{\substack{a\to 0\\ \tau\to 0}} \frac{a^2}{2\tau}$$

non nulle. L'équation limite

$$\frac{\partial \mathrm{K}}{\partial t} - \mathrm{D}\,\frac{\partial^2 \mathrm{K}}{\partial x^2} = 0 \tag{31.2}$$

est l'**équation de la chaleur** à une dimension d'espace. Quant aux conditions initiales nécessaires pour en donner une solution, elles s'énoncent ainsi : à l'instant $t = 0$, le marcheur est à l'abscisse $x = 0$. Cette condition ne peut pas s'énoncer en terme d'une *fonction* $x \mapsto \mathrm{K}(x, 0)$, mais nécessite l'emploi de distributions : on écrit donc que

$$\mathrm{K}(x,t) = \delta(x), \tag{31.3}$$

où δ est la distribution de Dirac. La résolution de l'équation (31.2) assortie de la condition initiale (31.3) conduit à[5]

$$\forall x \in \mathbb{R} \quad \forall t > 0 \qquad \mathrm{K}(x,\, t) = \frac{1}{\sqrt{4\pi \mathrm{D} t}}\, \mathrm{e}^{-x^2/4\mathrm{D}t}.$$

La fonction K est appelée **noyau de la chaleur**. On vérifie aisément qu'elle est correctement normalisée : $\int_{-\infty}^{+\infty} \mathrm{K}(x,\, t)\, \mathrm{d}x = 1$ pour tout $t \geqslant 0$.

Remarque 31.1 Pour un choix de $t > 0$, aussi petit soit il, la fonction $x \mapsto \mathrm{K}(x,\, t)$ est à valeurs strictement positives sur $\mathbb{R}$ tout entier. Cela veut dire que, lors d'un mouvement brownien, la vitesse du marcheur peut prendre des valeurs arbitrairement grandes. Ce n'est pas une erreur, car les processus de normalisation de l'espace et du temps se font, l'un en $1/\sqrt{n}$, l'autre en $1/n$, ce qui veut dire que la vitesse d'un saut est proportionnel à n. À la « limite continue », elle peut donc être infinie — ou, plus précisément, non définie. La probabilité pour le marcheur d'être à une distance x de l'origine décroît cependant comme $\mathrm{e}^{-x^2/2\mathrm{D}t}$, c'est-à-dire extrêmement rapidement. Cela est dû au fait qu'il est nécessaire de faire de nombreux pas dans la même direction, ce qui est évidemment de plus en plus improbable avec la distance.

Dans une situation physique comme la propagation de la chaleur, une telle propriété n'est évidemment pas souhaitable, et ne traduit qu'une lacune de modélisation ; lacune d'ailleurs sans grande importance dans la plupart des situations.

§ 433 Une propriété du noyau de la chaleur

Notons

$$\mathrm{W}(x,\, t\, ;\, x_0,\, t_0) = \mathrm{K}(x - x_0,\, t - t_0) = \frac{1}{\left[4\pi \mathrm{D}(t - t_0)\right]^{1/2}}\, \mathrm{e}^{-(x-x_0)^2/4\mathrm{D}(t-t_0)}.$$

La fonction W est la solution fondamentale d'une marche aléatoire de condition initiale $\mathrm{W}(x, t_0) = \delta(x - x_0)$.

■ **Proposition 31.2 (Équation de Chapman-Kolmogorov)** *Si* $t_0 \leqslant t_1 \leqslant t$, *alors*

$$\mathrm{W}(x,\, t\, ;\, x_0,\, t_0) = \int_{-\infty}^{+\infty} \mathrm{W}(x,\, t\, ;\, x_1,\, t_1)\ \mathrm{W}(x_1,\, t_1\, ;\, x_0,\, t_0)\ \mathrm{d}x_1.$$

Cette formule se démontre par un calcul explicite d'intégrale gaussienne (voir par exemple la démonstration du théorème 12.18). Elle a également une interprétation probabiliste : la distribution de probabilité est obtenue en sommant sur tous les états intermédiaires à l'instant t_1 — c'est donc un équivalent continu de la formule des probabilités totales[6]. Elle est caractéristique des *processus markoviens* (sans mémoire).

---

(5). Une méthode permettant de déterminer ce résultat est présenté en annexe à la fin du chapitre, page 637.

(6). On utilise le caractère *sans mémoire* de la marche aléatoire, qui permet d'écrire que la probabilité de trouver le marcheur en (x,t), sachant qu'il était en (x_1,t_1) *et* en (x_0,t_0), est égal à celle de la trouver en (x,t) sachant qu'il était en (x_1,t_1).

Limite continue d'une marche aléatoire asymétrique § 434

Considérons une marche au hasard asymétrique : la probabilité de faire un pas vers la droite est notée $p = \frac{1}{2} - \varepsilon$, celle de faire un pas vers la gauche est $q = \frac{1}{2} + \varepsilon$. La quantité ε est *a priori* quelconque, mais nous verrons que nous serons obligés de la faire varier en même temps que a et τ tendent vers 0. L'équation maîtresse est maintenant

$$\mathrm{K}\big(\ell a\,,\,(\mathrm{N}+1)\tau\big) = \left(\frac{1}{2}+\varepsilon\right)\mathrm{K}\big((\ell-1)a\,,\,\mathrm{N}\tau\big) + \left(\frac{1}{2}-\varepsilon\right)\mathrm{K}\big((\ell+1)a\,,\,\mathrm{N}\tau\big).$$

Comme précédemment, on retranche, à gauche et à droite, la quantité $\mathrm{K}\big(\ell a\,,\,\mathrm{N}\tau\big)$ et on effectue un développement de Taylor à l'ordre 1 en t et à l'ordre 2 en x :

$$\frac{\partial \mathrm{K}}{\partial t} \approx \frac{2\varepsilon a}{\tau}\,\frac{\partial \mathrm{P}}{\partial x} + \frac{a^2}{2\tau}\,\frac{\partial^2 \mathrm{K}}{\partial x^2}.$$

Ici encore, on fait tendre les quantités a et τ vers 0, de façon que $a^2/2\tau$ tende vers une constante $\mathrm{D} > 0$. Quant à la quantité $2\varepsilon a/\tau$, on va lui imposer qu'elle converge également vers une constante v, non nulle pour que l'asymétrie de la marche aléatoire ne disparaisse pas de la nouvelle équation ; il suffit pour cela de prendre $\varepsilon = v\tau/2a$. L'équation maîtresse devient alors

$$\frac{\partial \mathrm{K}}{\partial t} = v\,\frac{\partial \mathrm{K}}{\partial x} + \mathrm{D}\frac{\partial^2 \mathrm{K}}{\partial x^2}.$$

Le premier terme du membre de droite est un terme de dérive, dont la signification s'éclaire lorsque l'on résout cette équation[7] :

$$\mathrm{K}^{(v)}(x,t) = \frac{1}{\sqrt{4\pi \mathrm{D}\,t}}\,\mathrm{e}^{-(x-vt)^2/4\mathrm{D}t} = \mathrm{K}\,(x - vt,\,t)$$

En d'autres termes, on trouve la solution de la marche symétrique mais dans un repère en translation uniforme à la vitesse v.

Marche aléatoire en dimension supérieure § 435

On peut généraliser les constructions précédentes pour définir une marche aléatoire sur réseau en dimension d quelconque. Le processus de passage à la limite continue à partir d'une solution exacte est de prime abord plus délicat car cette solution est plus difficile à calculer explicitement[8] ; toutefois, le théorème limite central s'applique encore et permet de déterminer la densité limite :

$$\mathrm{K}_d(\boldsymbol{x},\,t) = \frac{1}{(4\pi \mathrm{D}t)^{d/2}}\,\mathrm{e}^{-\|\boldsymbol{x}\|^2/4\mathrm{D}t}. \qquad (31.4)$$

Dans la deuxième approche, le passage à l'équation continue est le même qu'à une dimension, et mène à l'équation aux dérivées partielles

$$\frac{\partial \mathrm{K}}{\partial t} = \mathrm{D}\ \triangle_d \mathrm{K} \qquad \text{où} \qquad \triangle_d := \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \cdots + \frac{\partial^2}{\partial x_d^2}$$

(en dimension $d = 1$, on retrouve donc l'équation (31.2)), et dont la solution est encore la fonction (31.4).

Les figures 31.3 et 31.4 présentent quelques marches aléatoires planes ($d = 2$).

---

(7). On utilise la même technique que précédemment : transformation de Laplace temporelle et de Fourier spatiale, résolution de l'équation algébrique, transformations inverses.

(8). On pourra consulter par exemple l'ouvrage de Itzykson et Drouffe [51, chapitre 1] pour une résolution physique de ce problème.

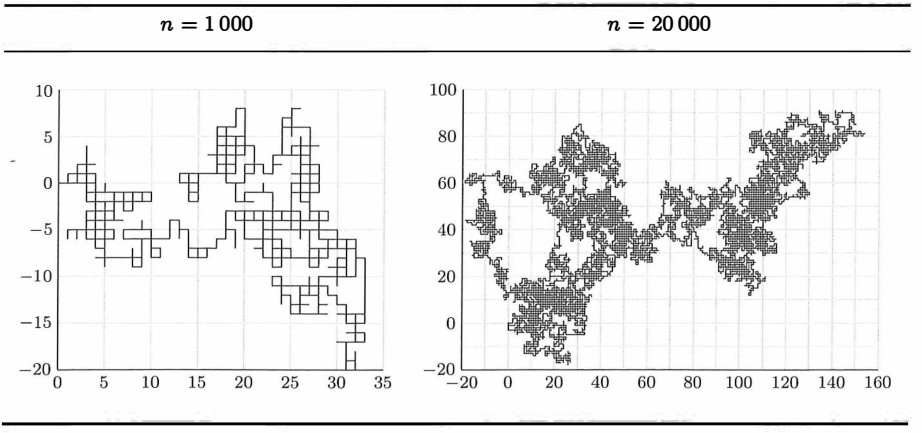


FIGURE 31.3 — *Figures de marches aléatoires discrètes planes : on représente respectivement* 1 000 *et* 20 000 *pas sur le réseau* $\mathbb{Z}^2$.

§ 436 **Temps moyen de visite**

Si A est un domaine de $\mathbb{R}^d$, la quantité

$$\int_{\mathrm{A}} \mathrm{K}_d(\boldsymbol{x}, t)\, \mathrm{d}^d\boldsymbol{x}$$

représente la probabilité de présence de la particule dans A à l'instant t. Si l'on intègre par rapport à t sur $\mathbb{R}^+$, on obtient le temps moyen de présence dans le domaine A :

$$\mathrm{T}(\mathrm{A}) = \int_0^{+\infty} \int_{\mathrm{A}} \mathrm{K}_d(\boldsymbol{x}, t)\, \mathrm{d}^d\boldsymbol{x}\, \mathrm{d}t.$$

On peut alors définir une densité de temps moyen de présence à la position $\boldsymbol{x}$:

$$\tau(\boldsymbol{x}) = \int_0^{+\infty} \mathrm{K}_d(\boldsymbol{x}, t)\, \mathrm{d}t.$$

Pour $d \leqslant 2$, cette quantité est infinie (l'intégrale diverge en $+\infty$), ce qui laisse deviner que la particule brownienne passera, dans tout domaine A, un temps infini, et que donc elle passera et repassera une infinité de fois par A. Cette propriété sera démontrée ultérieurement (théorème de Pólya).

Pour $d \geqslant 3$, le changement de variable $u = \|\boldsymbol{x}\|^2/2t$ conduit à

$$\tau(\boldsymbol{x}) = \frac{\Gamma(d/2-1)}{2^{d-1}\,\pi^{d/2}}\, \frac{1}{\|x\|^{d-2}}.$$

Il est à noter que cette quantité est égale au potentiel newtonien créé par une charge ponctuelle égale à -2 située à l'origine : $\triangle_d \tau = -2\delta$.

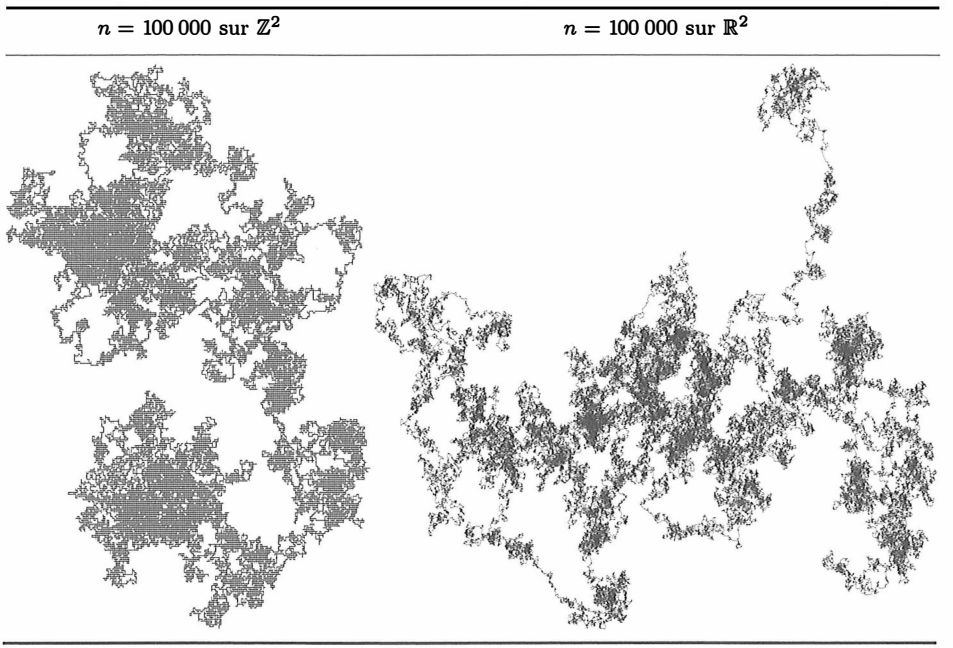


FIGURE 31.4 — *Comparaison de marches aléatoires sur $\mathbb{Z}^2$ et sur $\mathbb{R}^2$, chacune de 100 000 pas.* À gauche : *marche aléatoire sur le réseau $\mathbb{Z}^2$.* À droite : *marche dans $\mathbb{R}^2$; la longueur d'un pas est toujours égale à 1, mais la direction est tirée aléatoirement selon une loi uniforme sur $[0; 2\pi]$. La différence d'allure n'est qu'à peine perceptible, seule apparaît une différence de densité, due au fait que sur $\mathbb{Z}^2$ de nombreux tronçons se superposent, et n'apparaissent donc qu'une seule fois.*

Dernières remarques

§ 437

- Un *vrai* mouvement brownien est aussi éloigné des graphes de la figure 31.2 que ceux-ci des graphes de la figure 31.1. La limite continue de la marche aléatoire a une allure encore plus accidentée ; on peut montrer que, presque sûrement, elle n'est dérivable en aucun point. Cela se comprend intuitivement de la manière suivante : on a fait tendre le pas temporel vers 0 en le divisant par n, tandis que le pas spatial n'était divisé que par $1/\sqrt{n}$. Ainsi, les sauts se font à vitesse de plus en plus élevée au fur et à mesure que l'on diminue ces échelles. À la limite, la « vitesse » du marcheur est infinie.
  Jean Perrin, qui à la suite de la publication par Einstein d'une série d'articles sur le mouvement brownien et la détermination des tailles moléculaires, avait pu déterminer expérimentalement le nombre d'Avogadro (ce qui lui valut le prix Nobel), avait remarqué cette irrégularité absolue du mouvement brownien [78, chap IV, § 64] :

  > *L'enchevêtrement de la trajectoire est tel que la trajectoire notée est toujours infiniment plus simple et plus courte que la trajectoire réelle. En particulier, quand la durée qui sépare deux pointés d'un même grain décroît, la vitesse moyenne de ce grain durant cette durée, loin de tendre vers une limite, grandit sans cesse, et varie follement en direction [...]. On voit du même coup que l'on ne peut fixer une tangente en aucun point de la trajectoire, et c'est un cas où il est vraiment naturel de penser à ces fonctions continues sans dérivées que les mathématiciens ont imaginées, et que l'on regardait à tort comme de simples curiosités mathématiques, puisque l'expérience peut les suggérer.*

- En dimension d, le mouvement brownien est une « courbe » de dimension de Hausdorff égale à 2. Cela implique notamment que deux mouvements browniens en dimension supérieure à 4 ne vont que très exceptionnellement se rencontrer (cf. théorème 31.17).
- Dans le cas d'un mouvement brownien de dimension 2, si l'on part d'un réseau non pas carré, mais triangulaire, la limite continue obtenue est la même [51, chapitre 1]. En fait, cette limite continue est totalement dépendante de la structure du réseau, dont l'influence disparaît avec le processus de renormalisation des échelles ; c'est une conséquence directe et tangible du théorème limite central qui gomme les détails des lois des variables aléatoires sommées, et n'en garde que la moyenne et la variance.
- La frontière extérieure d'un mouvement brownien plan est de dimension 4/3. Ce résultat, conjecturé par Mandelbrot à la fin des années 70 [72], a été démontré au tournant du millénaire [70, 69]. Cette frontière a la même géométrie que celle d'un mouvement brownien auto-évitant, qui est un mouvement brownien dont la trajectoire ne se recoupe jamais. De manière remarquable, des considérations physiques, issues de la théorie des champs conformes, avaient permis de donner une démonstration heuristique de ce résultat par analyse multifractale [27].
- Le mouvement brownien sert d'outil de base dans de nombreuses branches de la physique théorique. Notamment, en théorie des champs, la dynamique d'une particule massive est décrite par des superpositions de mouvements browniens ; le propagateur d'une particule massive est la fonction de Green $K(x,t)$ calculée plus haut [52].
  De même, en mécanique statistique des plasmas, la nature quantique des particules peut être simulée dans un modèle semi-classique où les électrons sont remplacés par des filaments browniens classiques (formalisme de Feynman-Kac). La modélisation de polymères par des chemins browniens permet d'étudier efficacement leurs propriétés thermodynamiques [74].

31.2 MODÉLISATION PROBABILISTE

En inspectant la figure 31.3, on conçoit qu'on ne puisse pas définir une trajectoire brownienne comme une limite de marches aléatoires ; une famille de trajectoires, dont on modifie les échelles d'espace et de temps comme indiqué, ne converge pas au sens « géométrique » du terme.

Ce qui caractérise le mouvement brownien, ce ne sont pas les propriétés géométriques d'*une* trajectoire, mais les propriétés probabilistes de l'*ensemble* des trajectoires. Une ligne droite, par exemple, ou un arc régulier, sont des trajectoires browniennes ; simplement, elles sont exceptionnelles[(9)]. La (très grosse) majorité des courbes browniennes ne ressemble ni à une droite, ni même à un arc régulier.

C'est donc par le biais purement probabiliste qu'il est fécond d'appréhender le mouvement brownien, en le définissant par une loi sur un ensemble de chemins du plan.

(9). Si l'on réfléchit au cas d'une marche aléatoire à une dimension, les trajectoires correspondantes sont les deux trajectoires pour lesquelles les n premiers pas se font tous dans le même direction (droite ou gauche) ; l'ensemble des trajectoires « rectilignes » est donc de probabilité $2 \cdot 2^{-n} = 2^{1-n}$. Après le processus de limite $n \to \infty$, on a donc un ensemble de trajectoires de probabilité nulle.

Définition §438

Un mouvement brownien est une trajectoire *aléatoire* dans l'espace $\mathbb{R}^d$ à d dimensions, c'est-à-dire une variable aléatoire à valeurs dans l'espace des trajectoires de l'espace $\mathbb{R}^d$, trajectoires paramétrées par un temps $t \in [0\,;+\infty[$.

Plus précisément, pour chaque réalisation du hasard ω, on se donne une application $t \mapsto \mathrm{B}_t(\omega)$ définie sur $[0\,;+\infty[$, qui est une *réalisation de la trajectoire brownienne*, encore appelée **chemin brownien**. L'objet B possède donc deux variables :

— la variable ω, qui représente le hasard ;

— la variable t, qui est le temps.

À t fixé, l'application $\omega \mapsto \mathrm{B}_t(\omega)$ est une variable aléatoire à valeurs dans $\mathbb{R}^d$. Le mouvement brownien peut donc être vu comme une famille $(\mathrm{B}_t)_{t\geqslant 0}$ de variables aléatoires, indexée par l'ensemble continu $\mathbb{R}^+$. La définition classique du mouvement brownien prescrit, non pas une loi sur la variable $t \mapsto \mathrm{B}_t$, dont la valeur en chaque ω est une trajectoire, mais des lois sur les variables B_t, ainsi que des prescriptions d'indépendance[10] ; ces lois sont celles qui se déduisent de l'étude heuristique qui précède. Pour chaque valeur de ω, on demande enfin que $t \mapsto \mathrm{B}_t(\omega)$ soit continue.

Définition 31.3 On dira qu'une famille $(\mathrm{B}_t)_{t\geqslant 0}$ de variables aléatoires $\mathrm{B}_t : \Omega \to \mathbb{R}^d$ est un **mouvement brownien** sur un univers $(\Omega, \mathfrak{T}, \mathbf{P})$ si elle vérifie les propriétés suivantes, que nous commenterons aussitôt après :

- ⋆ **Axiome MB1 :** $\mathrm{B}_0 = 0$;
- ⋆ **Axiome MB2 :** les variations de B sont indépendantes : pour tout choix d'un entier $n \geqslant 1$ et d'instants $0 < t_1 < t_2 < \cdots < t_n$, les variables aléatoires

$$\mathrm{B}_{t_1},\quad \mathrm{B}_{t_2} - \mathrm{B}_{t_1},\qquad ,\quad \mathrm{B}_{t_n} - \mathrm{B}_{t_{n-1}}$$

  sont indépendantes ;
- ⋆ **Axiome MB3 :** pour tous $s, t \geqslant 0$, $\mathrm{B}_{t+s} - \mathrm{B}_t$ est un vecteur gaussien centré, de variance s (ou s fois la matrice identité I_d si $d \geqslant 2$) ;
- ⋆ **Axiome MB4 :** pour tout $\omega \in \Omega$, la trajectoire $t \mapsto \mathrm{B}_t(\omega)$ est continue.

Remarque 31.4 Justifions ces axiomes.

— L'axiome MB1 est clair : toute les trajectoires partent de l'origine. L'usage est parfois d'assouplir cette que la condition en n'exigeant que : $\mathrm{B}_0 = 0$ presque sûrement.

— Considérons une marche aléatoire à pas fini $(\mathrm{S}_n)_{n\geqslant 0}$, où $\mathrm{S}_0 = 0$ et $\mathrm{S}_n = \mathrm{X}_1 + \cdots + \mathrm{X}_n$, chacune des variables X_k prenant, de manière équiprobable, les valeurs $+1$ et -1. Fixons un instant n_0. Puisque les X_k sont indépendants, la valeur de $\mathrm{S}_n - \mathrm{X}_1$ est indépendante de la valeur de X_1. C'est une propriété de perte de mémoire, ou encore **markovienne** : la marche aléatoire, de position initiale X_{n_0} et d'instant initial n_0, a les mêmes propriétés qu'une marche aléatoire de position initiale 0 à l'instant $n = 0$, à une simple translation près. L'axiome MB2 ne fait que traduire cette propriété pour la « loi limite ».

— Toujours en une dimension, nous avons vu que la probabilité limite de trouver le marcheur en (x, t), sachant qu'il était parti de $(0, 0)$, était donnée par la loi normale centrée de variance t (équation (31.1) avec la convention que $\mathrm{D} = 1/2$). Ceci, et l'invariance par translation dans le temps et dans l'espace, justifie l'axiome MB3.
C'est cet axiome qui oblige (presque sûrement) une trajectoire brownienne à « explorer » l'espace, à une vitesse moyenne d'autant plus importante que l'on regarde à petite échelle. Il est donc la cause de la structure si accidentée d'une trajectoire brownienne (du moins d'une trajectoire brownienne « typique » !).

(10). Là encore, un théorème général de Kolmogorov permet de justifier l'existence d'une telle famille *non dénombrable* de variables aléatoires, possédant le type de propriétés que l'on va exiger.

— La dernière condition est cruciale : sans elle, on pourrait ne trouver que des apparitions aléatoires de points, et pas ce qu'on appelle communément une « trajectoire » ! Cependant, on remplace parfois l'axiome MB4 par un axiome légèrement plus faible, qui n'impose la continuité des trajectoires que presque sûrement.

La question peut légitimement se poser : un mouvement brownien est-il possible ? (suspense.... !)

Remarquons que, *a priori*, les axiomes MB2-MB3[11] peuvent sembler difficilement compatibles avec l'axiome MB4. Momentanément, nous admettrons que :

■ **Théorème 31.5 (Le mouvement brownien existe !)**
Il existe un espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ *sur lequel une famille* $(B_t)_{t\geqslant 0}$ *de variables aléatoires vérifie les axiomes MB1 à MB4. De plus, le mouvement brownien est « unique en loi » : deux mouvements browniens ont même loi.*

On peut citer au moins deux démonstrations possibles de ce théorème :
— la construction canonique utilisant la mesure de Wiener sur l'espace des trajectoires continues (§ 441) ;
— une construction plus formelle par les fonctions de Schauder, utilisant les propriétés de l'espace $L^2([0\,;1])$ (§ 447).
Nous ne ferons qu'évoquer la structure générale de ces démonstrations.

§ 439 **La loi du mouvement brownien est la loi limite de marches aléatoires**
Considérons une marche aléatoire sur le réseau $\mathbb{Z}^d$, que nous modélisons par exemple ainsi : pour tout entier k, X_k est une variable aléatoire, prenant avec une probabilité égale à $1/2^d$ les valeurs $\pm e_1, \pm e_2, \dots, \pm e_d$, où $(e_1, \dots, e_d)$ sont les vecteurs de la base canonique de $\mathbb{R}^d$. La position du marcheur après n pas est

$$S_n = X_1 + X_2 + \cdots + X_n.$$

C'est donc une marche où chaque pas se fait, de façon équiprobable, vers un plus proche voisin. Comme précédemment, normalisons cette marche en posant, pour tout $t \geqslant 0$ et tout $n \in \mathbb{N}$:

$$S_t^{(n)} = \frac{1}{\sqrt{n}}\, S_{\lfloor nt \rfloor},$$

où $\lfloor\cdot\rfloor$ désigne la fonction partie entière. Alors, pour tout $t \geqslant 0$, le théorème limite central montre que la suite $(S_t^{(n)})_{n\in\mathbb{N}}$ converge en loi vers la loi de B_t, qui est $\mathcal{N}(0\,;\, t\,I_d)$. De même, si $t_1 < t_2 < \cdots < t_p$, alors

$$(S_{t_1}^{(n)}, S_{t_2}^{(n)}, \dots, S_{t_p}^{(n)}) \overset{L}{\Longrightarrow} (B_{t_1}, B_{t_2}, \dots, B_{t_p}).$$

Autrement dit, le mouvement brownien est bien la « limite » (en loi) d'une suite de marches aléatoires proprement normalisées.

Ce résultat peut d'ailleurs être étendu au cas où les variables X_n sont à valeurs dans $\mathbb{R}^d$ (et non plus simplement $\mathbb{Z}^d$), indépendantes, centrées et de même variance finie σ^2 :

$$(S_{t_1}^{(n)}, S_{t_2}^{(n)}, \dots, S_{t_p}^{(n)}) \overset{L}{\Longrightarrow} (\sigma B_{t_1}, \sigma B_{t_2}, \dots, \sigma B_{t_p}).$$

En d'autres termes, la modélisation par des marches aléatoires de pas fini fonctionne, quel que soit le détail du mouvement (sur réseau ou non, à pas constant ou non...). C'est ce qui explique la similitude d'aspect des deux dernières marches aléatoires de la figure 31.3.

(11). Qui satisfont les conditions de compatibilité du théorème d'extension de Kolmogorov (théorème C.3 page 714).

Loi du mouvement brownien ; mesure de Wiener § 440

On peut aborder le mouvement brownien de manière totalement différente. Plutôt que d'essayer de construire des chemins possibles (en les approchant, *globalement*, par des marches aléatoires), on considère l'ensemble $\Omega = \mathscr{C}^0(\mathbb{R}^+, \mathbb{R}^d)$ de *toutes les trajectoires possibles*, c'est-à-dire des fonctions ω continues sur $\mathbb{R}^+$, à valeurs dans $\mathbb{R}^d$ et vérifiant $\omega(0) = 0$; on munit alors Ω d'une mesure de probabilité $\mathbf{P}$ qui vérifie les « bonnes » propriétés[12]. C'est la voie empruntée par Wiener[13].

Puisque Ω est non dénombrable, on ne peut pas, de manière efficiente, affecter de probabilité à chaque trajectoire (typiquement, chacune est de probabilité nulle) ; on n'affecte une probabilité qu'à des ensembles de trajectoires, c'est-à-dire des événements de Ω. Selon l'idée de Wiener, on simplifie encore le problème en ne cherchant pas à définir cette probabilité pour *tous* les événements — c'est une tâche titanesque — mais on se contente d'une classe plus restreinte, celle des *cylindres*. Choisissons des parties (boréliennes) $A_1, \dots, A_n$ de $\mathbb{R}^d$, et des instants $t_1 < t_2 < \cdots < t_n$. À l'ensemble

$$\mathscr{A}^{t_1,\dots,t_n}_{A_1,\dots,A_n} = \{\omega \in \Omega\,;\ \omega(t_1) \in A_1, \quad \omega(t_2) \in A_2, \quad \dots, \quad \omega(t_n) \in A_n\}$$

des trajectoires passant par A_i à l'instant t_i (pour $i = 1, \dots, n$), on associe la probabilité

$$\mathbf{P}\big(\mathscr{A}^{t_1,\dots,t_n}_{A_1,\dots,A_n}\big) = \int_{A_1\times\cdots\times A_n} K_{t_1}(x_1)\, K_{t_2-t_1}(x_2 - x_1)\cdots \cdots K_{t_n-t_{n-1}}(x_n - x_{n-1})\, dx_1\, dx_2 \cdots dx_n \quad (31.5)$$

où l'on a noté

$$K_t(x) = \frac{1}{(2\pi t)^{d/2}}\, e^{-\|x\|^2/2t}. \quad (31.6)$$

Remarque 31.6 La signification de cette probabilité est la suivante. Prenons pour simplifier $n = 3$; le marcheur aléatoire doit tout d'abord arriver en t_1 dans A_1 en partant de 0, puis de A_1 il doit arriver en A_2 à l'instant t_2, avant finalement d'arriver dans A_3 à l'instant t_3. On considère, pour tous les points x_1 de A_1, tous les points x_2 de A_2 et tous les points x_3 de A_3, la probabilité d'aller

- de 0 à x_1 en un temps t_1, puis
- d'aller de x_1 à x_2 en un temps $t_2 - t_1$, puis
- d'aller de x_2 à x_3 en un temps $t_3 - t_2$.

On somme enfin : ces probabilités (formule des probabilités totales) :

$$\mathbf{P}\big(\mathscr{A}^{t_1,t_2,t_3}_{A_1,A_2,A_3}\big) = \int_{A_1\times A_2\times A_3} K_{t_1}(x_1)\, K_{t_2-t_1}(x_2 - x_1)\, K_{t_3-t_2}(x_3 - x_2)\, dx_1\, dx_2\, dx_3.$$

Le procédé est illustré sur la figure 31.5.

Le choix de cette mesure de probabilité va grandement favoriser les chemins très irréguliers : pour des durées $t_{i+1} - t_i$ très courtes, la formule (31.6) autorise une « vitesse moyenne » de l'ordre de $1/\sqrt{t_2 - t_1}$ et de direction quelconque.

---

(12). Notamment, les chemins « réguliers » seront négligeables pour cette probabilité ; un chemin typique sera au contraire très accidenté.

(13). Norbert WIENER (1894–1964), mathématicien américain, a travaillé dans différents domaines : logique mathématique, analyse harmonique, analyse de Fourier ; il est également considéré comme le père de la cybernétique. Sa distraction était légendaire. Une légende (sans doute apocryphe) lui attribue l'aventure suivante. Un soir, en rentrant chez lui, il trouve sa maison vide. Il avise une fillette du quartier et lui demande ce qui s'est passé. La fillette lui répond que la famille a déménagé plusieurs jours auparavant. Il la remercie et elle lui répond : « Je me doutais que tu reviendrais ici, papa, c'est pourquoi je suis venue te chercher. »

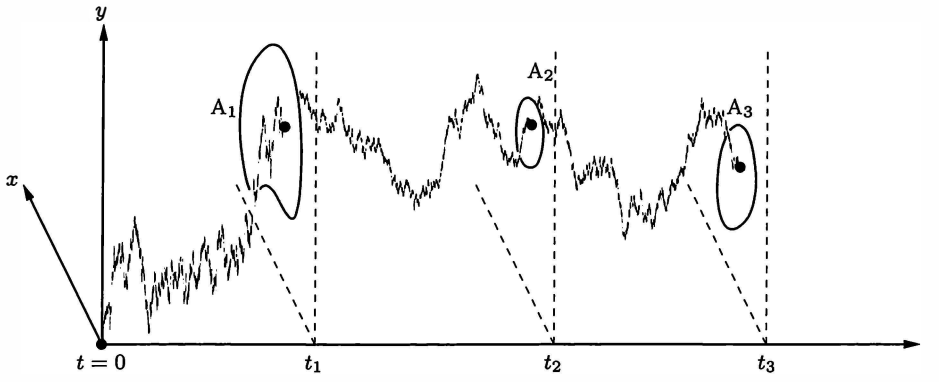


FIGURE 31.5 — *Une trajectoire du cylindre défini par les couples* (t_1, A_1), (t_2, A_2) *et* (t_3, A_3).

Les ensembles de la forme $\mathscr{A}^{t_1,\dots,t_n}_{\mathrm{A}_1,\dots,\mathrm{A}_n}$, où les A_k sont des boréliens de $\mathbb{R}^d$, sont appelés des **cylindres** de l'espace Ω. La formule précédente donne donc une mesure de probabilité sur la classe des cylindres. Celle-ci est stable par union croissante et par différence, et contient l'ensemble Ω tout entier ; le théorème de classe monotone, nous assure qu'il existe une (unique) manière d'étendre cette mesure à la tribu engendrée par la classe des cylindres. La mesure ainsi obtenue est appelée **mesure de Wiener** ; on la notera $\mathrm{d}\mathcal{W}$ ou $\mathcal{W}(\mathrm{dB})$.

§ 441 **Construction canonique du mouvement brownien**

Cette mesure une fois définie, on peut construire un mouvement brownien, c'est-à-dire une famille $(\mathrm{B}_t)_{t\geqslant 0}$ vérifiant les axiomes MB1 à MB4. On choisit comme univers $\Omega = \mathscr{C}^0(\mathbb{R}^+, \mathbb{R}^d)$, muni de la tribu engendrée par les cylindres et de la probabilité $\mathbf{P} = \mathrm{d}\mathcal{W}$ égale à la mesure de Wiener. On définit alors, pour tout $t \in \mathbb{R}$, l'application $\mathrm{B}_t : \Omega \to \mathbb{R}^d$ par $\mathrm{B}_t(\omega) = \omega(t)$ pour toute trajectoire $\omega \in \Omega$. Les axiomes MB1 et MB4 sont vérifiés par construction. Les axiomes MB2 et MB3 découlent de l'équation (31.5) (qui n'est à vrai dire qu'une reformulation de ces axiomes). Le théorème 31.5 en découle.

Remarque 31.7 (Cylindres (*)) Si T est un ensemble quelconque, il est possible de fabriquer une tribu sur l'espace $\mathbb{R}^{\mathrm{T}}$ des fonctions de T dans $\mathbb{R}$. Le cas qui nous préoccupe ici est celui pour lequel $\mathrm{T} = [\,0\,;+\infty\,[$. On note $\mathfrak{B}(\mathbb{R}^{\mathrm{T}})$ la tribu engendrée par les cylindres de la forme

$$\mathscr{C}^{t_1,\dots,t_n}(\mathrm{I}_1 \times \cdots \times \mathrm{I}_n) = \{\omega \in \mathbb{R}^{\mathrm{T}} \,;\, \omega(t_1) \in \mathrm{I}_1, \dots, \omega(t_n) \in \mathrm{I}_n\},$$

où les I_k sont des intervalles de la forme $]\,a_k\,;\,b_k\,]$ et où n est un entier quelconque. On peut montrer[14] que $\mathfrak{B}(\mathbb{R}^{\mathrm{T}})$ est également la tribu engendrée par les cylindres de la forme

$$\mathscr{C}^{t_1,\dots,t_n}(\mathrm{B}_1 \times \cdots \times \mathrm{B}_n) = \{\omega \in \mathbb{R}^{\mathrm{T}} \,;\, \omega(t_1) \in \mathrm{B}_1, \dots, \omega(t_n) \in \mathrm{B}_n\}$$

où cette fois les B_k sont des boréliens de $\mathbb{R}$. C'est enfin la tribu engendrée par les fonctions coordonnées $\pi_t : \omega \mapsto \omega(t)$, c'est-à-dire la plus petite tribu rendant ces fonctions mesurables.

La tribu $\mathfrak{B}(\mathbb{R}^{\mathrm{T}})$ a une structure un peu particulière, puisque *tous* ses éléments sont de la forme

$$\mathrm{A} = \{\omega \in \mathbb{R}^{\mathrm{T}} \,;\, (\omega_{t_1}, w_{t_2}, \dots) \in \mathrm{B}\}$$

où $(t_n)_{n\geqslant 1}$ est un ensemble *dénombrable* de points de T, et B un borélien de $\mathbb{R}^\infty$ (cf. A67). Cela signifie que tout élément de la tribu $\mathfrak{B}(\mathbb{R}^{\mathrm{T}})$ est entièrement déterminé par des conditions sur ses chemins ω, en un nombre *au plus dénombrable* d'instants[15].

Dans le cas qui nous intéresse, la situation est un peu plus compliquée car on exige également que les fonctions soient continues ; or l'ensemble des fonctions continues n'est pas dans $\mathfrak{B}(\mathbb{R}^{\mathrm{T}})$.

(14). Voir par exemple Shiryaev [91, chap. II,§ 2].

(15). Voir par exemple Billingsley [9, section 36].

Du coup, certaines constructions du mouvement brownien se font en deux étapes : on construit d'abord un pré-mouvement brownien sur $\mathbb{R}^{\mathrm{T}}$ satisfaisant les axiomes MB1 à MB3, puis on le rend continu en le modifiant « très peu », afin qu'il garde les mêmes propriétés de distribution : cela nécessite de prouver que le pré-mouvement brownien a une certaine régularité, par exemple qu'il soit uniformément continu sur les rationnels dyadiques de tout segment.

Covariance

§ 442

Certains ouvrages définissent un mouvement brownien à partir de l'expression de sa covariance. Le résultat, pour un mouvement brownien sur $\mathbb{R}$, est simple :

■ **Proposition 31.8** *Soit $(\mathrm{B}_t)_{t\geqslant 0}$ un mouvement brownien à valeurs dans $\mathbb{R}$. Pour tous réels positifs s et t,*

$$\mathbf{Cov}(\mathrm{B}_s, \mathrm{B}_t) = \mathbf{E}(\mathrm{B}_s\,\mathrm{B}_t) = \min(s,t).$$

DÉMONSTRATION : Supposons par exemple que $0 < s < t$. Notons déjà que $\mathbf{E}(\mathrm{B}_s) = \mathbf{E}(\mathrm{B}_t) = 0$. Ainsi, le calcul de la covariance se réduit à celui de l'espérance de $\mathrm{B}_s\,\mathrm{B}_t$. En écrivant $\mathrm{B}_t = \mathrm{B}_s + \mathrm{N}$, où l'on a simplement défini $\mathrm{N} = \mathrm{B}_t - \mathrm{B}_s$ et en utilisant la linéarité de l'espérance, on obtient

$$\mathbf{E}(\mathrm{B}_s\,\mathrm{B}_t) = \mathbf{E}(\mathrm{B}_s^2) + \mathbf{E}\,(\mathrm{B}_s\,\mathrm{N})\,.$$

Le premier terme vaut $\mathbf{E}(\mathrm{B}_s^2) = s$ (axiome MB3). Quant au second terme, il vaut

$$\mathbf{E}\,(\mathrm{B}_s\,\mathrm{N}) = \mathbf{E}(\mathrm{B}_s)\,\mathbf{E}(\mathrm{N}) = 0,$$

puisque $\mathrm{N} = \mathrm{B}_t - \mathrm{B}_s$ est indépendante de B_s (c'est l'axiome MB2) et que chacun des deux facteurs est nul (axiome MB3). Ainsi, $\mathbf{Cov}(\mathrm{B}_s, \mathrm{B}_t) = s$, ce qui montre le résultat annoncé. ▮

Remarque 31.9 Le résultat précédent peut paraître surprenant : la covariance de $(\mathrm{B}_s, \mathrm{B}_t)$ ne dépend que du plus petit des deux temps s et t, et nullement de l'autre. On le comprend intuitivement ainsi : pour les temps s très courts, la variable B_s est sans doute proche de 0 ; aussi, nous n'avons quasiment aucun renseignement sur B_t, dont la moyenne sera quasiment nulle. Au contraire, si s est plus grand, alors B_s aura sensiblement bougé de l'origine : cela nous donne donc une plus grande connaissance sur ce que sera B_t, puisque quoi qu'il arrive, il sera en moyenne centré sur B_s.

Plus formellement, dans le calcul de l'espérance de $\mathrm{B}_s\,\mathrm{B}_t$, la variable B_t est, « en moyenne », à la même position que B_s c'est-à-dire que l'on peut écrire $\mathrm{B}_t = \mathrm{B}_s + \mathrm{N}$, où N est indépendante de B_s, et suit une loi gaussienne centrée et de variance $t - s$. Graphiquement, cela signifie que la trajectoire brownienne, d'instant d'origine s et issue du point B_s, est de moyenne égale à B_s. Notamment, l'espérance de $\mathrm{B}_s\,\mathrm{N}$ est nulle : le temps t n'intervient plus. Nous n'avons d'ailleurs fait que paraphraser la démonstration du théorème.

Remarque 31.10 On peut proposer une seconde preuve de ce résultat, utilisant la mesure de Wiener pour calculer l'espérance de $\mathrm{B}_s\,\mathrm{B}_t$ (toujours en supposant $0 < s < t$). La densité de probabilité du couple $(\mathrm{B}_s, \mathrm{B}_t)$ est

$$f(x,y) = \frac{1}{\sqrt{2\pi s}}\,\exp\left\{-\frac{x^2}{2s}\right\}\cdot\frac{1}{\sqrt{2\pi(t-s)}}\,\exp\left\{-\frac{(y-x)^2}{2(t-s)}\right\}$$

c'est-à-dire que, si l'on note

$$g_t(x) := \frac{1}{\sqrt{2\pi t}}\,\mathrm{e}^{-x^2/2t},$$

on a

$$f(x,y) = g_s(x)\;g_{t-s}(y-x).$$

On obtient alors

$$\mathbf{Cov}(\mathrm{B}_s, \mathrm{B}_t) = \mathbf{E}(\mathrm{B}_s\,\mathrm{B}_t) = \iint xy\cdot g_s(x)\;g_{t-s}(y-x)\,\mathrm{d}x\,\mathrm{d}y.$$

Le changement de variable $u = y - x$ dans l'intégrale sur y permet d'écrire

$$\begin{aligned}\mathbf{Cov}(\mathrm{B}_s, \mathrm{B}_t) &= \iint x(u+x)\cdot g_s(x)\;g_{t-s}(u)\,\mathrm{d}x\,\mathrm{d}u\\ &= \iint xu\cdot g_s(x)\;g_{t-s}(u)\,\mathrm{d}x\,\mathrm{d}u + \iint x^2\cdot g_s(x)\;g_{t-s}(u)\,\mathrm{d}x\,\mathrm{d}u\end{aligned}$$

Le premier terme est nul par parité ; dans le second terme, l'intégration sur u se résume à un facteur 1, et l'intégrale restante est $\mathbf{E}(\mathrm{B}_s^2) = \mathbf{V}(\mathrm{B}_s) = s$.

Ce résultat se généralise pour un mouvement brownien à valeurs dans $\mathbb{R}^d$, la démonstration étant identique, à ceci près que $\mathbf{E}(\mathrm{B}_s^k \mathrm{B}_s^\ell) = \delta_{k,\ell}\, s\, \mathrm{I}_d$ pour un vecteur gaussien.

■ **Corollaire 31.11** *Soit $(\mathbf{B}_t)_{t\geqslant 0}$ un mouvement brownien à valeurs dans $\mathbb{R}^d$. On note $\mathbf{B}_t = (\mathrm{B}_t^1, \mathrm{B}_t^2, \ldots, \mathrm{B}_t^d)$. Pour tous $k, \ell \in \{1, \ldots, d\}$ et tous $s, t \geqslant 0$,*

$$\mathbf{Cov}(\mathrm{B}_s^k, \mathrm{B}_t^\ell) = \mathbf{E}(\mathrm{B}_s^k, \mathrm{B}_t^\ell) = \delta_{k,\ell} \min(s,t).$$

31.3 PROPRIÉTÉS DES CHEMINS BROWNIENS

§ 443 **Invariances**
Sans démonstration, indiquons que le mouvement brownien admet un certain nombre d'invariances.

■ **Théorème 31.12 (Invariances)** *Soit $\mathrm{B} = (\mathrm{B}_t)_{t\geqslant 0}$ un mouvement brownien à valeurs dans $\mathbb{R}^d$.*

*i) **(invariance par isométries)** $-\mathrm{B}$ est encore un mouvement brownien ; plus généralement, si φ est une isométrie vectorielle de $\mathbb{R}^d$, $\varphi(\mathrm{B})$ est un mouvement brownien.*

*ii) **(auto-similarité)** B est invariant par changement d'échelle quand on multiplie les dimensions d'espace par $\sqrt{\alpha}$ et celle de temps par $1/\alpha$. Ainsi, pour tout $\alpha > 0$, $t \mapsto \sqrt{\alpha}\, \mathrm{B}_{t/\alpha}$ est un mouvement brownien.*

*iii) **(oubli du passé)** Pour tout $s \geqslant 0$, $(\mathrm{B}_{s+t} - \mathrm{B}_s)_{t\geqslant 0}$ est un mouvement brownien ; il est indépendant du passé de s*[16].

*iv) **(retournement du temps)** $t \mapsto \frac{1}{t}\mathrm{B}_{1/t}$ est encore un mouvement brownien.*

Les trois premiers points se démontrent facilement. Pour le dernier, signalons que

$$\lim_{t\to 0^+} \tfrac{1}{t}\, \mathrm{B}_{1/t} = 0$$

presque sûrement, et que donc l'axiome MB1 n'est vérifié que dans le sens « presque sûr ».

§ 444 **Propriétés de régularité**
La caractère fortement irrégulier pressenti par Jean Perrin voit sa justification dans le théorème suivant, qui affirme que, pour *presque tous* les ω, le chemin brownien $t \mapsto \mathrm{B}_t(\omega)$ est très irrégulier.

■ **Théorème 31.13 (Paley-Wiener-Zygmund, 1933)**
Soit $(\mathrm{B}_t)_{t\geqslant 0}$ un mouvement brownien. Alors l'ensemble des $t \geqslant 0$ en lesquels la trajectoire est dérivable, est presque sûrement vide : si l'on note, pour tout $\omega \in \Omega$,

$$\mathrm{D}(\omega) = \big\{ t \in \mathbb{R} \,;\, s \longmapsto \mathrm{B}_s(\omega) \text{ est dérivable en } t \big\},$$

alors l'ensemble $\big\{\omega \in \Omega \,;\, \mathrm{D}(\omega) = \varnothing\big\}$ est de probabilité 1.

---

(16). C'est-à-dire qu'il est indépendant de la tribu $\sigma(\mathrm{B}_t \,;\, t \leqslant s)$.

Remarque 31.14 (Interprétation heuristique de la non-différentiabilité)
Le mouvement brownien est doté d'une propriété importante : c'est un **processus markovien**, ce qui signifie que la loi de B_t connaissant B_s (avec $s < t$) est la même que la loi de B_t connaissant toute l'histoire $\{B_u \,;\, 0 \leqslant u \leqslant s\}$. Une définition exacte de cette propriété requiert d'utiliser les probabilités conditionnelles sachant une tribu[17]. Ce qu'elle signifie, comme pour les marches aléatoires discrètes, c'est que B_t dépend uniquement de la position B_s, mais « ne se souvient pas de la façon dont on est arrivé en B_s ». Notamment, le chemin $t \mapsto B_t$ quitte la position B_s sans savoir comment y assurer une tangente.

Il existe une démonstration relativement élémentaire de ce résultat, que nous reportons en annexe (§ 455).

Récurrence : le théorème de Pólya § 445

Nous donnons, sans démonstration, les principaux résultats concernant les propriétés de passage du mouvement brownien dans un domaine donné du plan. L'idée principale de la démonstration de ce joli théorème dû à Pólya[18] sera donnée plus loin, en relation avec la théorie du potentiel.

■ **Théorème 31.15 (Pólya)** *Soit* B *un mouvement brownien à valeurs dans* $\mathbb{R}^d$.

i) Si $d = 1$, *tout point de* $\mathbb{R}$ *est presque sûrement visité une infinité de fois par le mouvement brownien.*

ii) Si $d = 2$, *tout point qui n'est pas le point de départ du mouvement brownien n'est presque sûrement jamais visité.*

iii) Si $d = 1$ *ou* $d = 2$, *le mouvement brownien est* **récurrent** *: pour tout ouvert* $U \subset \mathbb{R}^d$ *non vide,* $t \mapsto B_t$ *revient une infinité de fois dans* U *presque sûrement. Plus précisément, l'ensemble* $\{t \geqslant 0 \,;\, B_t \in U\}$ *est presque sûrement non borné.*

iv) Si $d \geqslant 3$, *au contraire,* $|B_t|$ *tend vers l'infini presque sûrement, et le nombre d'entrées dans* U *est donc presque sûrement fini : le mouvement brownien est* **transient**.

Autres propriétés (∗) § 446

Citons pour finir deux propriétés supplémentaires du mouvement brownien :

■ **Proposition 31.16 (Régularité)** *La courbe brownienne* $t \mapsto B_t$ *est presque sûrement höldérienne de rapport* $1/2 - \varepsilon$, *et ce pour tout* $\varepsilon > 0$ *:*

$$\forall \varepsilon > 0 \quad \forall T \geqslant 0 \qquad \sup_{0 \leqslant t' < t \leqslant T} \frac{|B_t - B_{t'}|}{|t - t'|^{1/2-\varepsilon}} < +\infty \quad p.s.$$

Sur tout intervalle de temps fini $[0\,;T]$, *la courbe brownienne est presque sûrement à variation infinie. En revanche, sa variation quadratique est presque sûrement finie : la borne supérieure des*

$$\sum_{k=0}^{n-1} \Big(B_{t_{k+1}}(\omega) - B_{t_k}(\omega)\Big)^2$$

où $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$ *est une subdivision quelconque*[19] *de* $[0\,;T]$ *est presque sûrement égale à* T.

---

(17). Si l'on note $\sigma(s)$ la tribu engendrée par la famille $\{B_u \,;\, 0 \leqslant u \leqslant s\}$, on a $\mathbf{E}\big(B_t \,\|\, \sigma(s)\big) = \mathbf{E}(B_t \,\|\, B_s)$.

(18). Győrgy (George) Pólya, (1887–1985), mathématicien hongrois naturalisé américain, s'est intéressé à la théorie des nombres et aux probabilités. Il fut également un pédagogue talentueux, auteur de nombreux ouvrages de problèmes et d'exercices, parfois redoutables, comme les « Pólya-Szegő » [79].

(19). En particulier, n n'est pas fixé.

■ **Proposition 31.17 (Points multiples)** *Soit $t \mapsto B_t$ un mouvement brownien.*

i) Le mouvement brownien plan ($d = 2$) présente presque sûrement une infinité de points doubles dans tout intervalle de temps ; le cardinal de ces points doubles a presque sûrement la puissance du continu.

Ce résultat se généralise : $t \mapsto B_t$ présente presque sûrement une infinité de points de multiplicité m arbitraire.

ii) Si $d = 3$, le mouvement brownien présente encore presque sûrement une infinité de points doubles dans tout intervalle de temps ; en revanche, il n'a presque sûrement pas de points triples.

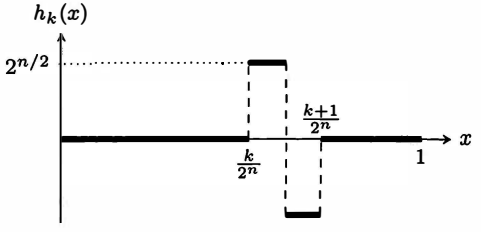
iii) Si $d \geqslant 4$, le mouvement brownien n'a presque sûrement pas de points doubles (ni de multiplicité supérieure).

Remarque 31.18 L'invariance par changement d'échelle ($\sqrt{\alpha} B_{t/\alpha}$ est encore un mouvement brownien) permet de prouver que, s'il y a presque sûrement un point double dans un intervalle de temps donné, alors il y en a presque sûrement une infinité : on découpe l'intervalle en une infinité d'intervalles sur lesquels les petits mouvements browniens ont les mêmes propriétés que le mouvement brownien global.

§ 447 **Construction de Lévy-Ciesielski[21] du mouvement brownien**

Nous ne pouvons pas présenter en détail les constructions explicites du mouvement brownien ; il en existe au moins quatre ou cinq, fondamentalement différentes, sans compter les variations. Nous n'en citerons qu'une, en plus de la construction canonique déjà évoquée, utilisant les *fonctions de Schauder*, dont voici un squelette de démonstration.

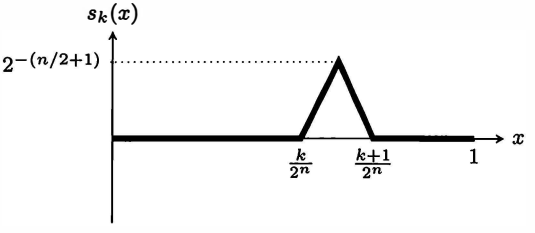
- On commence par construire la famille des **fonctions de Haar** : on pose $h_0 = 1$ puis, pour tout entier $n \geqslant 1$ et tout entier k compris entre 2^{n-1} et $2^n - 1$, on définit la fonction h_k par le graphe suivant :



- Ces fonctions forment une base hilbertienne de l'espace $L^2([0;1],\mathbb{R})$.
- On construit ensuite les primitives s'annulant en 0 de ces fonctions

$$s_k(t) := \int_0^t h_k(x)\,dx.$$

On les appelle **fonctions de Schauder**. Le graphe de s_k est alors



(21). D'après Paul Lévy (1886–1971) et Zbigniew Ciesielski (1934–), mathématicien polonais.

- Ces fonctions de Schauder vérifient la relation fondamentale[22]

$$\sum_{n=0}^{\infty} s_n(s)\, s_n(t) = \min(s,t)$$

pour tous réels s et t de $[0\,;\,1]$. C'est cette relation qui permettra d'obtenir la propriété $\mathbf{E}(B_s\, B_t) = \min(s,t)$.

- On construit ensuite une suite $(A_n)_{n\geqslant 1}$ de variables aléatoires indépendantes de même loi $\mathcal{N}(0\,;\,1)$, et on pose

$$B_t(\omega) := \sum_{k=0}^{\infty} A_k(\omega)\, s_k(t)$$

pour tout $0 \leqslant t \leqslant 1$ et tout $\omega \in \Omega$. Alors la série converge, uniformément en t, pour presque tout ω ; notamment, $t \mapsto B_t(\omega)$ est continue pour presque tout ω. ⇗31.3

- On prouve que $\mathbf{E}(e^{iu(B_t - B_s)}) = e^{-u^2(t-s)/2}$ lorsque $t > s$ par indépendance des A_k et grâce à la relation fondamentale des fonctions de Schauder, ce qui montre que $B_t - B_s$ suit la loi $\mathcal{N}(0\,;\,t-s)$.
- On fabrique ensuite un mouvement brownien B_t pour tout $t \geqslant 0$ en « recollant » des mouvements browniens du type précédent, dont on fabrique des exemplaires indépendants $B^{(n)}$:

$$B_t = B_n + B^{(n+1)}_{t-n}$$

pour tout $t \in [n\,;\,n+1]$.

Le fait que les fonctions de Schauder fassent des pics de plus en plus minces (la hauteur est proportionnelle à $1/\sqrt{2^n}$ tandis que la largeur est égale à $1/2^n$) est évidemment un élément crucial dans le fait que la fonction $(B_t)_{t\geqslant 0}$ est un mouvement brownien.

Remarque 31.19 (Construction par les polynômes de Bernstein) On trouvera, dans un article d'Emmanuel Kowalski [64], une construction originale du mouvement brownien utilisant les polynômes de Bernstein[23].

---

[22]. Cette relation provient uniquement du fait que la famille $(h_n)_{n\geqslant 0}$ des fonctions de Haar est une base hilbertienne de l'espace $L^2[0\,;\,1]$. On remarque en effet que $s_n(t) = \langle I_t | h_n \rangle$, où I_t est la fonction caractéristique du segment $[0\,;\,t]$ et où le produit scalaire est bien entendu $\langle g|h\rangle = \int_0^1 gh$. Par conséquent (théorème A.41 page 688)

$$\sum_{n=0}^{\infty} s_n(s)\, s_n(t) = \sum_{n=0}^{\infty} \langle I_s | s_n\rangle \langle s_n | I_t\rangle = \langle I_s | I_t\rangle = \min(s,t).$$

[23]. Voir page 502 pour la définition de ces polynômes.

31.4 MOUVEMENT BROWNIEN, PHYSIQUE ET POTENTIELS

§448 **Le modèle d'Einstein et la détermination du nombre d'Avogadro**

Dans un article historique de 1905, Albert Einstein expose une théorie quantitative du mouvement brownien ; il relie notamment le déplacement quadratique moyen $\langle x^2 \rangle$ d'une particule sphérique de rayon a, plongée dans un liquide de viscosité η, au nombre d'Avogadro $\mathcal{N}_A$:

$$\langle x^2 \rangle = 2\mathrm{D}t \qquad \mathrm{D} = \frac{\mathrm{RT}}{6\pi \mathcal{N}_A \eta a} = \frac{k_B \mathrm{T}}{6\pi \eta a}, \tag{31.7}$$

où R est la constante des gaz parfaits, $k_B = \mathrm{R}/\mathcal{N}_A$ est la constante de Boltzmann, T la température et t le temps. Il importe de comprendre ce que cette formule avait d'absolument novateur : elle relie des quantités macroscopiques, toutes mesurables indépendamment en laboratoire (la température T, la viscosité η, la constante de diffusion D, la constante R et le diamètre a des particules) à la constante $\mathcal{N}_A$ qui, elle, est de nature fondamentalement microscopique[24]. De nos jours, il paraît banal d'affirmer que, dans un gramme d'hydrogène, on compte $\mathcal{N}_A \approx 6{,}023 \cdot 10^{23}$ atomes ; on oublie facilement qu'un tel nombre ne se détermine pas comme on compte des patates. Le physicien Jean Perrin[25], aidé de son étudiant Chaudesaigues, mène en 1908 des expériences longues et délicates, basées sur le relation d'Einstein ; ils obtiennent $\mathcal{N}_A$ avec moins de 20 % d'erreur[26] ce qui, compte-tenu d'hypothèses simplificatrices entrant en jeu dans la formule d'Einstein, est excellent. Dans *Les atomes*, Jean Perrin s'enthousiasme de ce résultat qui, pour la première fois, faisait passer les atomes du statut d'hypothèse (et donc combattus par les adeptes du positivisme) à celui de réalité tangible :

> *La théorie atomique a triomphé. Nombreux encore naguère, ses adversaires enfin conquis renoncent l'un après l'autre aux défiances qui longtemps furent légitimes et sans doute utiles.*

Avant de fermer cette parenthèse historique, on ne saurait passer sous silence les travaux remarquables de Smoluchowski[27] qui obtient, indépendamment d'Einstein, et en se basant sur la théorie cinétique, l'équation (31.7) — à un facteur géométrique $64/27 = (4/3)^3$ près.

§449 **Problème de Dirichlet et méthode de Monte-Carlo**

On se donne un contour Γ fermé, délimitant une portion bornée du plan. Ce contour est supposé formé de deux arcs : $\Gamma = \gamma \cup \gamma'$. On se pose la question de savoir, lorsque l'on initie une marche aléatoire d'un point $\mathrm{A} = (x_0, y_0)$ (situé à l'intérieur de Γ), quelle est la probabilité qu'il sorte en franchissant γ, plutôt que γ'[28]. La probabilité G(A) cherchée vérifie l'équation de Laplace

$$\triangle \mathrm{G} = \frac{\partial^2 \mathrm{G}}{\partial x^2} + \frac{\partial^2 \mathrm{G}}{\partial y^2} = 0,$$

---

(24). Pour mémoire, $1/\mathcal{N}_A$ est la masse, en grammes, d'*un* atome d'hydrogène (plus précisément, $12/\mathcal{N}_A$ est celle d'un atome de carbone 12).

(25). Jean Perrin (1870–1942), physicien français, prix Nobel en 1926.

(26). La première estimation précise de $\mathcal{N}_A$, due au physicien et chimiste autrichien Johann Loschmidt (1865), donnait $\mathcal{N}_A \approx 10^{24}$, soit presque 70 % de trop ; ce résultat était cependant en tous points remarquable.

(27). Marian Smoluchowski (1872–1917), physicien polonais, fut le premier à expliquer l'opalescence critique par les fluctuations de densité à grande échelle. Le lecteur intéressé par une analyse des travaux de Smoluchowski peut lire, par exemple, l'article de Duplantier [28].

(28). Il est implicitement admis que, presque sûrement, le mouvement brownien finira par franchir Γ, ce que l'on peut démontrer rigoureusement.

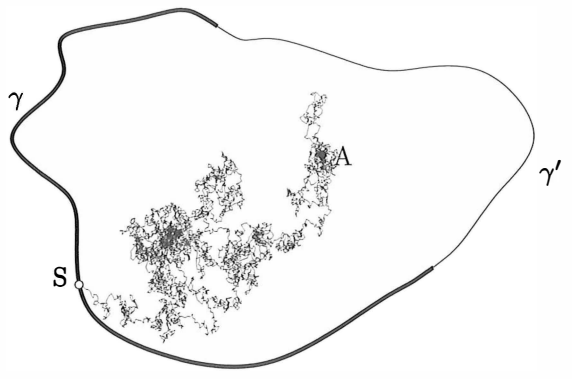


FIGURE 31.6 — *Méthode de Monte-Carlo pour déterminer la solution d'un problème de Dirichlet : trouver la fonction* $G(x,y)$ *vérifiant* $\triangle G = 0$, *et prenant la valeur* 1 *sur la courbe* γ, *et la valeur* 0 *sur* γ'. *Ici, la marche sort en* S, *situé sur* γ, *et est donc comptabilisée.*

(c'est une fonction harmonique, on le montrera un peu plus loin) et la condition au bord

$$G(x,y) = \begin{cases} 1 & \text{si } (x,y) \in \gamma, \\ 0 & \text{si } (x,y) \in \gamma'. \end{cases}$$

Déterminer la probabilité $G(x_0, y_0)$ revient donc à résoudre un problème de Dirichlet. Réciproquement, déterminer une solution (numérique) de ce problème de Dirichlet peut se faire en déterminant *expérimentalement* la densité de probabilité G. Pour cela, on utilise une méthode de Monte-Carlo. On simule (informatiquement) des marches aléatoires depuis un point $A(x_0, y_0)$, et on compte comme valides, parmi toutes ces marches, celles qui sortent par γ. La proportion des marches valides fournit une approximation de $G(x_0, y_0)$.

De manière plus générale, si D est un domaine ouvert de $\mathbb{R}^d$, délimité par une hypersurface $\Sigma = \partial D$, et si $f : \Sigma \to \mathbb{R}$ est une fonction mesurable et bornée, la valeur de la fonction G répondant au problème de Dirichlet

$$\begin{cases} \triangle_d G(\boldsymbol{x}) = 0 & \text{pour tout } \boldsymbol{x} \in D \\ G(\boldsymbol{x}) = f(\boldsymbol{x}) & \text{pour tout } \boldsymbol{x} \in \Sigma \end{cases}$$

peut être déterminée en tout point P de D, de la manière suivante. On tire un chemin brownien $t \mapsto B_t(\omega)$, d'origine P, dont le *point de première sortie* est noté $S(\omega)$; ce point est la position $B_{t^*}(\omega)$ à l'instant $t^*(\omega)$ de première sortie, défini par

$$t^*(\omega) = \inf \{ t \geqslant 0 \,;\, B_t(\omega) \notin D \}.$$

On calcule ensuite la moyenne sur la valeur de $f(S)$, la « moyenne » se faisant avec la loi des chemins browniens issus de P (c'est la mesure de Wiener $\mathcal{W}_P(dB)$, translatée de la mesure $\mathcal{W}(dB)$). On peut écrire, formellement,

$$G(P) = \int f(B_{t^*})\, \mathcal{W}_P(dB). \tag{31.8}$$

(Une version discrète de ce résultat est présentée dans le chapitre sur les chaînes de Markov, paragraphe § 408, page 571.) Cette formule a été démontrée par le

mathématicien japonais Kakutani[29] en 1944, en utilisant la mesure de Wiener pour obtenir une limite continue dans une formule discrète équivalente, obtenue dans les années 1920. On peut l'interpréter ainsi : G(P) est l'espérance $\mathbf{E}_{\mathrm{P}}\big(f(\mathrm{B}_{t^*})\big)$ de $f(\mathrm{B}_{t^*})$, sous la probabilité $\mathrm{d}\mathbf{P}_{\mathrm{P}} = \mathcal{W}_{\mathrm{P}}(\mathrm{dB})$ du mouvement brownien issu de P.

Elle nous fournit un algorithme pour calculer numériquement une approximation de G : on approche un mouvement brownien par un tirage d'un grand nombre de marches aléatoires $\mathrm{M}_1, \mathrm{M}_2, \ldots, \mathrm{M}_n$ de pas a « petit » ; pour chacune de ces marches, on cherche le point de sortie S_k ; enfin, on évalue la moyenne

$$\frac{1}{n}\sum_{k=1}^{n} f(\mathrm{S}_k).$$

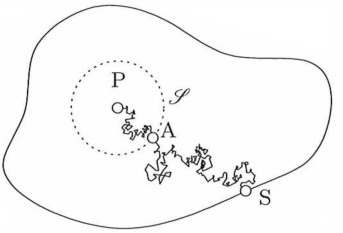
Cette moyenne est bien arithmétique, sans pondération ; dans ce procédé, la loi du mouvement brownien (celle que l'on retrouve dans le symbole $\mathbf{E}_{\mathrm{P}}$ ou dans la mesure $\mathcal{W}_{\mathrm{P}}(\mathrm{dB})$) est automatiquement prise en compte : les marches arrivant en un point S donné seront en nombre très variables selon la localisation de S.

Remarque 31.20 (Démonstration heuristique de l'harmonicité de G*)*
Pour montrer que G est harmonique, il suffit de montrer qu'elle vérifie la propriété de la moyenne : la valeur de G en un point P quelconque est égale à la valeur moyenne de G sur toute sphère centrée en P et incluse dans le domaine D.

Pour cela, traçons une sphère $\mathscr{S}$ de rayon r, centrée sur un point P de D. Tout chemin brownien issu de P et atteignant Σ traverse nécessairement $\mathscr{S}$; notons A le point de première traversée. Écrivons ceci formellement :

$$\mathrm{G}(\mathrm{P}) = \frac{1}{\sigma(r)}\int_{\mathscr{S}} \varphi_{\mathrm{A}}(\mathrm{P})\,\mathrm{dA},$$

où l'on a noté $\sigma(r)$ la surface de la sphère de rayon r, et $\varphi_{\mathrm{A}}(\mathrm{P})$ l'espérance de $f(\mathrm{B}_{t^*})$ sous la loi du mouvement brownien passant par P *et* par A.



Or, un mouvement brownien partant de P et passant par A a, ultérieurement, le même comportement qu'un mouvement brownien partant de A[30]. Ainsi, la quantité $\varphi_{\mathrm{A}}(\mathrm{P})$ est égale à G(A). La formule précédente s'écrit alors

$$\mathrm{G}(\mathrm{P}) = \frac{1}{\sigma(r)}\int_{\mathscr{S}} \mathrm{G}(\mathrm{A})\,\mathrm{dA},$$

ce qui est exactement la propriété de la moyenne. Cela prouve que G est harmonique.

Une démonstration formelle et complète est nettement plus technique, et fait appel, comme on l'a dit, à la propriété de Markov forte. Le lecteur pourra la trouver dans de nombreux ouvrages [55].

---

(29). KAKUTANI Shizuo (1911–2004), connu pour ses travaux en analyse fonctionnelle.

(30). C'est la **propriété de Markov forte**, ainsi nommée car l'instant de passage par A n'est pas un instant fixe, mais une variable aléatoire (un **temps d'arrêt**). Soit τ une telle variable aléatoire, à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$ et ne dépendant que du passé du mouvement brownien (i.e. $\{\tau \leqslant t\} \in \sigma(\mathrm{B}_s\ ;\ s \leqslant t)$: pour connaître la valeur de τ, il suffit de connaître les valeurs de B_s pour $s \leqslant \tau$) ; alors, conditionnellement à $\{\tau < \infty\}$, $t \mapsto \mathrm{B}_{t+\tau} - \mathrm{B}_\tau$ est encore un mouvement brownien.

Mouvement brownien et potentiel capacitif § 450

Le théorème de Kakutani permet d'établir un lien entre la théorie du mouvement brownien et l'équation de Poisson ($\Delta\varphi = \rho$) qui est à la base de la théorie du potentiel[31].

En dimension $d \geqslant 3$, le potentiel créé par une distribution ponctuelle décroît à l'infini. Soit K un domaine fermé de l'espace $\mathbb{R}^d$, délimité par une hypersurface Σ suffisamment régulière[32]. On définit le potentiel capacitif de K comme le potentiel électrostatique créé par une distribution de charges assurant un potentiel égal à 1 sur Σ, et s'annulant en l'infini ; c'est donc également le potentiel coulombien (harmonique) créé par une distribution de charges sur un conducteur parfait.

Ce potentiel capacitif en un point A s'obtient, d'après la formule de Kakutani, comme la probabilité qu'un mouvement brownien issu de A entre ultimement dans K. Pour le voir, il suffit d'entourer K d'une sphère $\mathcal{C}_R$ de rayon R et de résoudre le problème de Dirichlet $\Delta G = 0$, $G(M) = 1$ pour tout $M \in \Sigma$ et $G(M) = 0$ sur le cercle $\mathcal{C}_R$: la quantité G(A) est égale à la probabilité que la marche brownienne entre dans K avant de sortir par $\mathcal{C}_R$. On fait alors tendre R vers l'infini pour obtenir le potentiel capacitif de K.

Dans le paragraphe suivant, c'est au contraire en utilisant des solutions connues de la théorie du potentiel que l'on va enrichir la compréhension des propriétés de récurrence du mouvement brownien.

Théorie du potentiel et propriétés de récurrence ($d = 1$) § 451

La formule de Kakutani (31.8) permet de comprendre simplement les propriétés de récurrence du mouvement brownien.

Commençons par le cas unidimensionnel $d = 1$. On assimile un segment [OA] de longueur a à l'intervalle $[0\,;\,a]$; lançons un mouvement brownien d'un point P d'abscisse x et notons $G(x)$ la probabilité que son point de première sortie soit A plutôt que O. La fonction $G : [0\,;\,a] \to \mathbb{R}$ vérifie donc le problème de Dirichlet unidimensionnel

$$\frac{\partial^2 G}{\partial x^2} = 0 \qquad G(0) = 0 \qquad G(a) = 1,$$

dont la résolution est immédiate :

$$\forall x \in [0\,;\,a] \qquad G(x) = \frac{x}{a}.$$

Ainsi, la probabilité de sortir par A en partant de x vaut x/a ; la probabilité de sortir par O vaut $1 - x/a$ (on rappelle que, presque sûrement, le mouvement brownien est non borné et donc finit par sortir par un côté ou par l'autre).

Notons génériquement T(B) ou T_B la variable aléatoire égale à l'instant de première entrée dans le domaine fermé B pour un mouvement brownien issu de x. La propriété précédente peut alors s'écrire

$$\mathbf{P}\{T_{]-\infty\,;\,0]} > T_{[a\,;\,+\infty[}\} = \frac{x}{a}.$$

---

(31). Rappelons l'origine physique de cette équation : le potentiel électrostatique φ (également appelé **potentiel coulombien**) créé par une densité électrique de charges ρ vérifie l'équation $\Delta\varphi = -\rho/\varepsilon_0$, où ε_0 est une constante physique (permittivité du vide). En dimension 3, le potentiel coulombien créé par une charge ponctuelle est proportionnel à $1/r$; en dimension 2, à $\ln r$; en dimension 1, à r. Ces différents comportements réapparaissent dans le calcul de la fonction G. En dimension deux, on peut profiter de toutes les propriétés liées à l'invariance conforme.

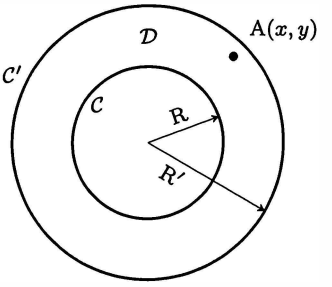
(32). Les conditions de régularité requises sont assez peu contraignantes : en chaque point de Σ on doit pouvoir placer la pointe d'un demi-cône entièrement inclus dans le complémentaire de K. En pratique, si Σ est une surface de classe $\mathscr{C}^1$, tout se passe bien.

Tout en gardant fixé le point x, faisons tendre a vers $+\infty$; en faisant usage de l'axiome 3', on obtient que *la probabilité de sortir par le point* 0 *avant de partir vers l'infini est égale à* 1. Puisque x est quelconque, l'invariance des propriétés browniennes par translation permet de déduire qu'un mouvement brownien passera (presque sûrement) par tous les points de $\mathbb{R}$.

♦ ***Exercice 31.1*** Montrer qu'il passera, presque sûrement, une infinité de fois par chacun des points de $\mathbb{R}$.

§ 452 **Théorie du potentiel et propriétés de récurrence ($d = 2$)**

Abordons maintenant le problème du mouvement brownien plan ($d = 2$). On considère la couronne ouverte $\mathcal{D}$ délimitée par les cercles $\mathcal{C}$ et $\mathcal{C}'$ de rayons respectifs R et R'.



Si A(x,y) est un point de la couronne, notons $\mathrm{G}_{\mathrm{R}}^{\mathrm{R}'}(x,y)$ la probabilité qu'un mouvement brownien issu de A sorte de $\mathcal{D}$ par le cercle extérieur plutôt que par le cercle intérieur. La formule de Kakutani montre que la fonction $\mathrm{G}_{\mathrm{R}}^{\mathrm{R}'} : \mathcal{D} \to \mathbb{R}$ vérifie le problème de Dirichlet

$$(\mathrm{S}) : \quad \begin{cases} \triangle \mathrm{G}_{\mathrm{R}}^{\mathrm{R}'} = 0 & \text{sur } \mathcal{D} \\ \mathrm{G}_{\mathrm{R}}^{\mathrm{R}'}|_{\mathcal{C}} = 0 & \mathrm{G}_{\mathrm{R}}^{\mathrm{R}'}|_{\mathcal{C}'} = 1. \end{cases}$$

La symétrie du problème permet de chercher une solution radiale $\mathrm{G}_{\mathrm{R}}^{\mathrm{R}'}(x,y) = \widetilde{\mathrm{G}}(r)$, ce qui transforme l'équation au laplacien en l'équation différentielle ordinaire

$$\frac{\mathrm{d}^2}{\mathrm{d}r^2}\widetilde{\mathrm{G}}(r) + \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\widetilde{\mathrm{G}}(r) = 0.$$

Tous calculs effectués, l'unique solution au problème (S) est

$$\widetilde{\mathrm{G}}(r) = \mathbf{P}\{\mathrm{T}(\overline{\mathrm{B}}_{\mathrm{R}}) > \mathrm{T}(\mathrm{B}_{\mathrm{R}'}^{\mathrm{c}})\} = \frac{\ln(r/\mathrm{R})}{\ln(\mathrm{R}'/\mathrm{R})}.$$

En passant au complémentaire,

$$\mathbf{P}\{\mathrm{T}(\overline{\mathrm{B}}_{\mathrm{R}}) < \mathrm{T}(\mathrm{B}_{\mathrm{R}'}^{\mathrm{c}})\} = \frac{\ln(\mathrm{R}'/r)}{\ln(\mathrm{R}'/\mathrm{R})}.$$

Cette quantité, qui est la probabilité de traverser $\mathcal{C}$ avant de traverser $\mathcal{C}'$, admet une limite quand $\mathrm{R}' \to \infty$: la probabilité de traverser $\mathcal{C}$ avant de partir à l'infini est égale à 1. En d'autres termes, *le mouvement brownien entrera dans le cercle* $\mathcal{C}$ *avec une probabilité* 1.

Ce résultat étant indépendant du choix de la position initiale (x,y), on a montré que le mouvement brownien sur $\mathbb{R}^2$ est récurrent, c'est-à-dire qu'il visite tout voisinage de tout point de $\mathbb{R}^2$, et ce une infinité de fois.

Remarque 31.21 À cause de la divergence de la fonction logarithme à la fois en 0 et en l'infini, on observe en dimension deux un comportement d'apparence paradoxale. Faisons tendre R vers 0 en gardant r et R′ fixés, la quantité $\widetilde{G}(r)$ admet alors pour limite 1 : presque sûrement, le mouvement brownien sortira par C' avant de passer par $\{0\}$. Ceci étant vrai pour toute valeur de R′, la probabilité que le mouvement brownien (issu d'un point différent de l'origine) passe par 0 est nulle.

Conclusion : presque sûrement le chemin brownien ne passera pas par 0 ; en revanche, il repassera une infinité de fois dans tout voisinage de 0, c'est-à-dire que 0 est presque sûrement point d'accumulation du chemin brownien.

Théorie du potentiel et propriétés de récurrence ($d \geqslant 3$) § 453

Traitons rapidement le cas des dimensions $d \geqslant 3$. On définit toujours le domaine $\mathcal{D}$ délimité par les sphères de rayons R et R′, et on considère un mouvement brownien issu d'un point A de ce domaine, situé à la distance r du centre ($R < r < R'$).

L'équation au laplacien correspondante se résout génériquement en

$$\widetilde{G}(r) = \alpha r^{2-d} + \beta$$

soit, en prenant en compte les conditions aux bords,

$$\widetilde{G}(r) = \frac{R^{2-d} - r^{2-d}}{R^{2-d} - R'^{2-d}}.$$

En passant au complémentaire, on a donc

$$\mathbf{P}\{T(\overline{B}_R) < T(B^c_{R'})\} = \frac{r^{2-d} - R'^{2-d}}{R^{2-d} - R'^{2-d}}.$$

Cette fois-ci, le potentiel ne diverge qu'en 0. En faisant tendre R′ vers $+\infty$, on obtient que

$$\mathbf{P}\{T(\overline{B}_R) < +\infty\} = \left(\frac{R}{r}\right)^{d-2}$$

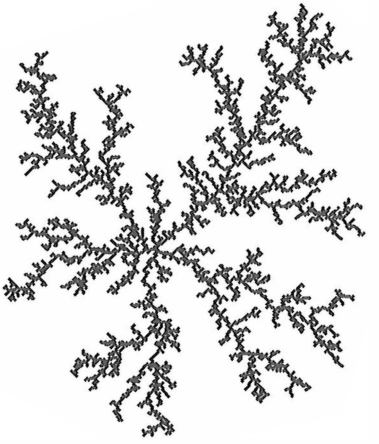
et donc que : la probabilité que le mouvement brownien aille visiter la boule de centre R en temps fini est strictement inférieure à 1.

Remarque 31.22 (Divergence presque sûre du mouvement brownien lorsque $d \geqslant 3$) Pour terminer, on peut montrer que, presque sûrement, $\lim_{t \to +\infty} |B_t| = +\infty$, mais une démonstration rigoureuse utilise la propriété de Markov forte pour les mouvements browniens ; signalons simplement que le cœur de la démonstration est d'utiliser la formule précédente pour montrer que

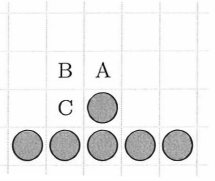
$$\mathbf{P}\left\{\inf_{t \geqslant T(\overline{B}^c_{2^n})} |B_t| \leqslant n\right\} = \left(\frac{n}{2^n}\right)^{d-2}.$$

Le lemme de Borel-Cantelli permet de conclure que les événements $\left\{\inf_{t \geqslant T(\overline{B}^c_{2^n})} |B_t| > n\right\}$ sont (presque sûrement) toujours réalisés à partir d'un certain rang, ce qui entraîne le résultat cherché.

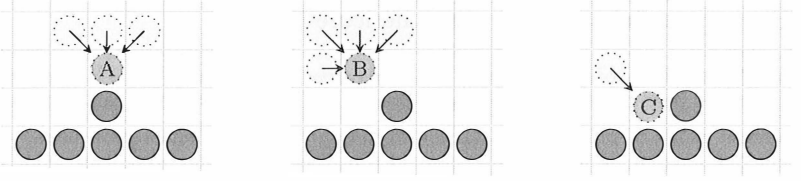
Remarque 31.23 (Agrégation fractale (DLA)) Le lien entre marche aléatoire et théorie du potentiel est illustré par la croissance d'agrégats selon le protocole suivant. Une particule se meut selon une marche aléatoire plane (sur un réseau dont le pas est petit) jusqu'à ce qu'elle se « colle » à un germe fixe ; l'agrégat ainsi formé est à même de fixer de nouvelles particules, et on réitère le processus. À première vue, on pourrait croire que la croissance de l'agrégat serait plus ou moins régulière et amènerait à la formation d'un patatoïde. Or, il n'en est rien, et la forme obtenue est arborescente !



Dans un article célèbrissime [101], Thomas Witten et Leonard Sander décrivent la formation d'un tel agrégat (appelé DLA : *Diffusion-limited aggregation*) et déterminent sa dimension fractale approximative (elle vaut environ $1,68$)(33). Un raisonnement simple permet de comprendre pourquoi les « branches » de l'agrégat ont tendance à grandir. Supposons existante la structure suivante :



Lorsqu'une nouvelle particule arrive à proximité de cette construction, on imagine que la probabilité qu'elle vienne se coller en A (ou en B) est environ trois (ou quatre) fois plus grande que celle qu'elle vienne se coller en C, compte tenu du nombre de possibilités de terminer en A ou B plutôt qu'en C.



À la lumière du paragraphe § 449, nous pouvons également interpréter ce phénomène comme étant un *effet de pointe*, bien connu en théorie du potentiel ; c'est cet effet qui attire la foudre par temps d'orage, ou crée des feux de Saint-Elme sur les mâts des bateaux. En dimension deux, il s'explique par le fait que la transformation conforme $f : z \mapsto w := z^\alpha$ (avec $\alpha < 1$) envoyant le demi-plan supérieur sur le complémentaire d'un angle aigu a une dérivée divergente au voisinage de l'origine(34).

---

(33). La méthode utilisée par Witten et Sander est la suivante. On définit la fonction de corrélation de densité expérimentale

$$f(r) = \frac{1}{\mathrm{N}} \sum_i \sum_j \rho(x_i)\, \rho(x_j),$$

où $\rho(x_i) = 1$ si le site i est occupé, et $\rho(x_i) = 0$ sinon, et où les indices i et j vérifient $|x_i - x_j| \approx r$. L'entier N est le nombre total de particules sur le réseau. Le comportement de la fonction f est une loi de puissance : $f(r) \propto r^{d-1}$, où d est la dimension fractale de l'agrégat.

(34). Sur les transformations conformes, voir Lang [67], Greene & Krantz [43], ou Appel [2] pour leur utilisation en physique.

31.5 ANNEXES

Détermination du noyau de la chaleur § 454

On cherche une fonction (ou une distribution) $\mathrm{K}(x, t)$, de condition initiale $\mathrm{K}(x, 0) = \delta_x$, et vérifiant l'équation de la chaleur

$$\forall x \in \mathbb{R} \quad \forall t > 0 \qquad \left(\frac{\partial}{\partial t} - \mathrm{D} \frac{\partial^2}{\partial x^2} \right) \mathrm{K}(x, t) = 0.$$

Prenons la transformée de Laplace des deux membres de cette équation, en prenant en compte la condition initiale :

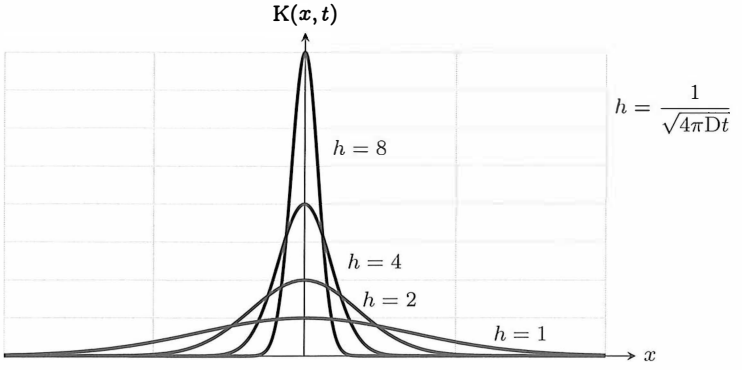
$$\left(p - \mathrm{D} \frac{\partial^2}{\partial x^2} \right) \mathscr{K}(x, p) = \delta(x).$$

En prenant la transformée de Fourier de cette égalité, $(p + 4\pi^2 \mathrm{D}\, k^2)\, \widetilde{\mathscr{K}}(k, p) = 1$, on obtient

$$\widetilde{\mathscr{K}}(k, p) = \frac{1}{(p + 4\pi^2 \mathrm{D}\, k^2)}.$$

La transformation de Laplace inverse mène à $\widetilde{\mathrm{K}}(k, t) = \mathrm{e}^{-4\pi^2 \mathrm{D} k^2 t}$ et la transformée de Fourier inverse à

$$\mathrm{K}(x, t) = \frac{1}{\sqrt{4\pi \mathrm{D} t}}\, \mathrm{e}^{-x^2/4\mathrm{D}t}.$$



Comme l'illustre la figure précédente, la famille de fonctions $x \mapsto \mathrm{K}(x, t)$ converge, au sens des distributions, vers une distribution de Dirac lorsque $t \to 0$.

Le même calcul peut se mener en dimension d ; on trouve alors

$$\mathrm{K}_d(x, t) = \frac{1}{(4\pi \mathrm{D} t)^{d/2}}\, \mathrm{e}^{-\|x\|^2/4\mathrm{D}t}.$$

Démonstration du théorème de Paley-Wiener-Zygmund § 455

La démonstration qui suit est essentiellement due à Dvoretsky, Erdős et Kakutani.

Notons A l'ensemble des fonctions localement lipschitziennes en au moins un point :

$$\mathrm{A} = \left\{ \exists s \geqslant 0 \qquad \limsup_{t \to s} \left| \frac{\mathrm{B}_t - \mathrm{B}_s}{t - s} \right| < +\infty \right\}.$$

(Attention, il s'agit de « lim sup » au sens analytique, non ensembliste.) Si un chemin $\mathrm{B}(\omega)$ est dérivable en au moins un point s, alors il est dans A ; pour prouver le théorème, nous allons prouver que A est négligeable, c'est-à-dire, puisque A n'est pas *a priori* un événement, que A inclus dans un ensemble N mesurable et de mesure nulle : $\mathrm{A} \subset \mathrm{N}$ et $\mathbf{P}(\mathrm{N}) = 0$.

Fixons un entier $p > 0$. Pour tout entier $n \geqslant 1$, posons

$$\mathrm{C}_n(p) = \left\{ \exists s \in [\,0\,;\, n\,[\ ;\ \text{pour tout } t \geqslant 0 \text{ tel que } |t - s| \leqslant \frac{5}{n},\ |\mathrm{B}_t - \mathrm{B}_s| \leqslant p\,|t - s| \right\}.$$

(La raison du choix du majorant $5/n$ apparaîtra plus tard.) La suite $(C_n(p))_{n\geqslant 1}$ est croissante ; on pose

$$C(p) = \lim_n \uparrow C_n(p) = \bigcup_{n=1}^{\infty} C_n(p)$$
$$= \{\exists s \geqslant 0 \quad \exists \delta > 0 \quad \forall t \geqslant 0 \qquad |t-s| \leqslant \delta \implies |B_t - B_s| \leqslant p\,|t-s|\},$$

ensemble des fonctions localement p-lipschitziennes en au moins un point. On a alors

$$A = \bigcup_{p=1}^{\infty} C(p).$$

Hélas, ni les $C_n(p)$ ni $C(p)$ ne sont *a priori* des événements(35). On pose alors

$$D_n(p) = \left\{\text{il existe } 1 \leqslant k \leqslant n^2 \text{ ; pour } j = 0, \dots, 4, \quad \left|B_{(k+j)/n} - B_{(k+j-1)/n}\right| \leqslant \frac{9p}{n}\right\}.$$

Les $D_n(p)$ sont bien des événements. De plus :

■ **Lemme 31.24** *Pour tout n entier, $C_n(p) \subset D_n(p)$.*

DÉMONSTRATION : Soit $\omega \in C_n(p)$; choisissons s comme dans la définition de $C_n(p)$. On encadre s grâce à un unique $k \leqslant n^2$ vérifiant $\frac{k-1}{n} \leqslant s < \frac{k}{n}$. Alors, pour tout $j = 0, \dots, 4$, on a

$$\left|\frac{k+j}{n} - s\right| \leqslant \frac{5}{n} \qquad \text{et} \qquad \left|\frac{k+j-1}{n} - s\right| \leqslant \frac{4}{n},$$

donc par inégalité triangulaire

$$\left|B_{(k+j)/n}(\omega) - B_{(k+j-1)/n}(\omega)\right| \leqslant \left|B_{(k+j)/n}(\omega) - B_s(\omega)\right| + \left|B_{(k+j-1)/n}(\omega) - B_s(\omega)\right| \leqslant \frac{9p}{n},$$

ce qui montre que $\omega \in D_n(p)$. ∎

La propriété d'invariance par changement d'échelle montre que B_t a même loi que $\sqrt{t}\,B_1$; l'axiome MB3 montre que $B_{(k+j)/n} - B_{(k+j-1)/n}$ a même loi que $B_{1/n}$, et donc que $B_1/\sqrt{n}$. On en déduit, par indépendance des $B_{(k+j)/n} - B_{(k+j-1)/n}$, que

$$\mathbf{P}\big(D_n(p)\big) \leqslant \sum_{k=1}^{n^2} \left[\mathbf{P}\left\{|B_{1/n}| \leqslant \frac{9p}{n}\right\}\right]^5 = n^2 \left[\mathbf{P}\left\{|B_{1/n}| \leqslant \frac{9p}{n}\right\}\right]^5 = n^2 \left[\mathbf{P}\left\{|B_1| \leqslant \frac{9p}{\sqrt{n}}\right\}\right]^5.$$

Enfin, puisque B_1 suit une loi normale $\mathcal{N}(0\,;\,I_d)$, si l'on note V le volume de la boule unité de $\mathbb{R}^d$,

$$\mathbf{P}\big(D_n(p)\big) \leqslant n^2 \left[\frac{1}{(2\pi)^{d/2}} \int_{|x|\leqslant 9p/\sqrt{n}} e^{-|x|^2/2}\, d^d(x)\right]^5$$
$$\leqslant \left(\frac{V}{(2\pi)^{d/2}}\right)^5 n^2 \left(\frac{9p}{\sqrt{n}}\right)^{5d} = O\left(\frac{1}{n^{(5d-4)/2}}\right)$$

et donc $\mathbf{P}\big(D_n(p)\big) \to 0$ car $5d - 4 > 0$ quelle que soit la dimension. Si l'on note maintenant

$$D(p) = \liminf D_n(p) = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} D_n(p),$$

alors par l'axiome 3′ (continuité monotone croissante), on a

$$\mathbf{P}\big(D(p)\big) = \lim_{N\to\infty} \mathbf{P}\left(\bigcap_{n=N}^{\infty} D_n(p)\right) \leqslant \lim_{N\to\infty} \mathbf{P}\big(D_N(p)\big) = 0$$

donc $D(p)$ est négligeable. Puisque $A = \bigcup_{p=1}^{\infty} C(p)$, il suffit maintenant de poser $N = \bigcup_{p=1}^{\infty} D(p)$ pour obtenir les relations voulues : $A \subset N$ et $\mathbf{P}(N) \leqslant \sum_{p=1}^{\infty} \mathbf{P}\big(D(p)\big) = 0$.

(35). Ils ne peuvent pas se ramener à des conditions sur les trajectoires portant sur un nombre dénombrable d'instants.

EXERCICES

♦ ***Exercice 31.2 (Variance d'un pont brownien)*** Soit $(B_t)_{t\geqslant 0}$ un mouvement brownien. On rappelle que $\mathbf{V}(B_t) = t$ pour tout $t \geqslant 0$. On définit $P_t = B_t - t\,B_1$ pour tout $t \geqslant 0$, c'est-à-dire

$$P_t(\omega) = B_t(\omega) - t\,B_1(\omega)$$

pour tout $t \geqslant 0$ et tout $\omega \in \Omega$. Calculer l'espérance et la variance de P_t.

♦ ***Exercice 31.3 (Une étape de la construction du MB)*** On reprend la construction du mouvement brownien par les fonctions de Schauder (§ 447).

i) Soit $(a_n)_{n\geqslant 0}$ une suite réelle ; on suppose qu'il existe un réel β suffisamment proche de 0, par exemple $\beta \in [0\,;\,1/2\,[$, tel que $|a_k| = O(k^\beta)$. Montrer que $\sum a_k\, s_k$ converge uniformément sur $[0\,;\,1]$.

ii) Montrer que, pour tout $x > 0$,

$$\mathbf{P}\{\,|A_k| > x\} \leqslant e^{-x^2/2}$$

iii) En déduire que, presque sûrement,

$$|A_k| = \underset{k\to\infty}{O}\left(\sqrt{\ln k}\right).$$

iv) Conclure que, presque sûrement, la série $\sum A_k\, s_k$ converge uniformément sur $[0\,;\,1]$, et donc que $t \mapsto B_t$ est presque sûrement continu sur $[0\,;\,1]$.

♦ ***Exercice 31.4 (Intégrale de Paley-Wiener-Zygmund)***
Soit $(B_t)_{t\geqslant 0}$ un mouvement brownien. Pour toute fonction $f : [0\,;\,1] \to \mathbb{R}$ de classe $\mathscr{C}^1$ et vérifiant $f(0) = f(1) = 0$, on définit l'intégrale « $\int_0^1 f\,dB$ » par

$$\int_0^1 f\,dB := -\int_0^1 f'(t)\,B_t\,dt,$$

le second membre étant bien défini puisque $t \mapsto B_t$ est continue. (Attention, cette intégrale est, non pas un nombre, mais une variable aléatoire.)

i) Montrer que $\mathbf{E}\left(\int_0^1 f\,dB\right) = 0$ (on admettra que l'on peut intervertir les intégrations sur t et sur Ω).

ii) Montrer que $\mathbf{E}\left[\left(\int_0^1 f\,dB\right)^2\right] = \int_0^1 f^2$.

L'histoire ne s'arrête pas là ; si f est une fonction de $L^2[0\,;\,1]$, alors il existe une suite $(\varphi_n)_{n\geqslant 0}$ de fonctions de classe $\mathscr{C}^1$, nulles en 0 et en 1, qui converge vers f dans L^2, c'est-à-dire telle que $\int_0^1 |\varphi_n - f|^2 \to 0$. En utilisant la deuxième propriété, montrer que la suite de terme général $\int_0^1 \varphi_n\,dB$ est de Cauchy dans $L^2(\Omega)$. On définit alors

$$\int_0^1 f\,dB := \lim_{n\to\infty} \int_0^1 \varphi_n\,dB,$$

intégrale de Paley-Wiener-Zygmund pour une fonction $f \in L^2$.

Indications

◊ *Indication 31.3 :* Noter que les fonctions s_k ont des supports disjoints lorsque $2^n \leqslant k < 2^{n+1}$. En déduire que

$$\sum_{k=2^n}^{\infty} |a_k\, s_k(t)| \leqslant C^{\text{te}} \sum_{k=n}^{\infty} 2^{(k+1)\beta}\, 2^{-k/2-1}$$

pour tout entier p et tout $t \in [0\,;\,1]$.

Pour la question *iii)*, utiliser le lemme de Borel-Cantelli (sens trivial).

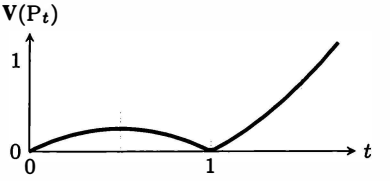
◊ *Indication 31.4 :* On rappelle que $\mathbf{E}(B_s\,B_t) = \min(s,t)$.

SOLUTIONS

♦ **Solution 31.2** Par linéarité, $\mathbf{E}(\mathrm{P}_t) = \mathbf{E}(\mathrm{B}_t) - t\,\mathbf{E}(\mathrm{B}_1) = 0 - 0 = 0$. Pour la variance, on développe

$$\mathbf{V}(\mathrm{B}_t) = \mathbf{V}(\mathrm{B}_t - t\,\mathrm{B}_1) = \mathbf{V}(\mathrm{B}_t) + t^2\,\mathbf{V}(\mathrm{B}_1) - 2t\,\mathbf{Cov}(\mathrm{B}_t, \mathrm{B}_1) = t + t^2 - 2t\,\min(t,1).$$

Il ne reste qu'à séparer les cas $t \leqslant 1$ et $t \geqslant 1$, et l'on trouve finalement $\mathbf{V}(\mathrm{P}_t) = |t^2 - t|$.



♦ **Solution 31.3**

i) Il existe un entier N et un réel $\mathrm{M} \geqslant 0$ tels que, pour tout $k \geqslant 2^{\mathrm{N}}$, on ait $|a_k| \leqslant \mathrm{M}\,k^\beta$. Pour tout $k \geqslant \mathrm{N}$, notons n l'entier vérifiant $2^n \leqslant k < 2^{n+1}$; on a alors

$$\forall t \in [0\,;1] \qquad |a_k\,s_k(t)| \leqslant \frac{\mathrm{M}}{2}\,\frac{k^\beta}{2^{n/2}} \leqslant \frac{\mathrm{M}}{2}\,\frac{2^{(n+1)\beta}}{2^{n/2}} = \mathrm{M}\,2^{\beta-1}\,2^{n(\beta-1/2)} =: \gamma_n.$$

Les supports des fonctions s_k pour $k \in [\![2^n\,;\,2^{n+1}-1]\!]$ étant disjoints à un point près, la majoration précédente reste vraie en sommant toutes ces fonctions s_k :

$$\left\| \sum_{k=2^n}^{2^{n+1}-1} a_k\,s_k \right\|_\infty \leqslant \gamma_n.$$

Cette égalité reste également vraie si l'on ne somme que sur $[\![p\,;\,2^{n+1}-1]\!]$ pour $2^n \leqslant p < 2^{n+1}$. Notamment, pour tout $p \geqslant 2^{\mathrm{N}}$, on peut majorer la norme infinie du reste R_p de la série $\sum a_k\,s_k$:

$$\|\mathrm{R}_p\|_\infty \leqslant \left\| \sum_{k=p}^{2^{n+1}-1} a_k\,s_k \right\|_\infty + \sum_{i=n+1}^{\infty} \gamma_i \leqslant \sum_{i=n}^{\infty} \gamma_i,$$

où $n = \lfloor \ln_2(p) \rfloor$. Cela prouve la convergence uniforme de la suite des restes vers 0, et donc la convergence uniforme de la série.

ii) Pour tout $x > 0$,

$$\begin{aligned}\mathbf{P}\{\,|\mathrm{A}_k| > x\} &= \frac{2}{\sqrt{2\pi}} \int_x^\infty \mathrm{e}^{-t^2/2}\,\mathrm{d}t = \frac{2}{\sqrt{2\pi}} \int_0^\infty \mathrm{e}^{-(t+x)^2/2}\,\mathrm{d}t \\ &\leqslant \mathrm{e}^{-x^2/2} \cdot \frac{2}{\sqrt{2\pi}} \int_0^\infty \mathrm{e}^{-t^2/2}\,\mathrm{d}t = \mathrm{e}^{-x^2/2}.\end{aligned}$$

iii) Si l'on évalue par exemple

$$\mathbf{P}\{\,|\mathrm{A}_k| > \sqrt{\ln k}\} \leqslant \mathrm{e}^{-(\ln k)/2} = \frac{1}{\sqrt{k}},$$

on ne peut rien conclure car la série diverge. En revanche,

$$\mathbf{P}\{\,|\mathrm{A}_k| > 2\sqrt{\ln k}\} \leqslant \mathrm{e}^{-2\ln k} = \frac{1}{k^2},$$

et cette fois-ci, la convergence de la série $\sum 1/k^2$ permet de déduire par le lemme de Borel-Cantelli que

$$\mathbf{P}\{\,|\mathrm{A}_k| > 2\sqrt{\ln k}\ ;\ \text{i.s.}\} = 0.$$

Cela montre que, pour tout ω appartenant à un ensemble B de probabilité 1, il n'existe qu'un nombre fini d'entiers k tels que $|\mathrm{A}_k| > 2\sqrt{\ln k}$, ou encore qu'il existe un entier N_ω (dépendant de ω) tel que, pour tout $k \geqslant \mathrm{N}_\omega$, on ait $|\mathrm{A}_k(\omega)| \leqslant 2\sqrt{\ln k}$.

En d'autres termes, on a bien montré que, presque sûrement, $\mathrm{A}_k = \mathrm{O}\left(\sqrt{\ln k}\right)$.

Cette condition implique par exemple que, presque sûrement $\mathrm{A}_k = \mathrm{O}(k^{1/4})$; on peut appliquer le résultat de la question *i)* et conclure que, presque sûrement, la série $\sum \mathrm{A}_n\,s_n$ converge uniformément sur $[0\,;1]$ et que sa somme est continue.

♦ **Solution 31.4** Par linéarité de l'espérance, et en intervertissant sans scrupule l'espérance et l'intégrale,

$$\mathbf{E}\left(\int_0^1 f\,\mathrm{dB}\right) = \mathbf{E}\left(-\int_0^1 f'(t)\,\mathrm{B}_t\,\mathrm{d}t\right) = -\int_0^1 f'(t)\,\mathbf{E}(\mathrm{B}_t)\,\mathrm{d}t = 0$$

puisque $\mathrm{B}_t \rightsquigarrow \mathcal{N}(0\,;\,t)$. Pour la seconde formule,

$$\begin{aligned}\mathbf{E}\left[\left(\int_0^1 f\,\mathrm{dB}\right)^2\right] &= \mathbf{E}\left(\int_0^1 f'(t)\,\mathrm{B}_t\,\mathrm{d}t\cdot\int_0^1 f'(s)\,\mathrm{B}_s\,\mathrm{d}s\right)\\ &= \int_0^1\int_0^1 f'(t)\,f'(s)\,\mathbf{E}(\mathrm{B}_t\,\mathrm{B}_s)\,\mathrm{d}t\,\mathrm{d}s = \int_0^1\int_0^1 f'(t)\,f'(s)\,\min(t,s)\,\mathrm{d}t\,\mathrm{d}s\end{aligned}$$

On découpe l'intégrale sur t en deux intégrales : sur $[0\,;\,s]$ et sur $[s\,;\,1]$; deux intégrations par parties mènent alors à

$$\begin{aligned}\mathbf{E}\left[\left(\int_0^1 f\,\mathrm{dB}\right)^2\right] &= \int_0^1 f'(s)\left(\int_0^s f'(t)\,t\,\mathrm{d}t + s\int_s^1 f'(t)\,\mathrm{d}t\right)\mathrm{d}s\\ &= \int_0^1 f'(s)\left(-\int_0^s f(t)\,\mathrm{d}t\right)\mathrm{d}s = \int_0^1 f(s)^2\,\mathrm{d}s.\end{aligned}$$

En utilisant cette propriété, on obtient immédiatement

$$\mathbf{E}\left[\left(\int_0^1 \varphi_p\,\mathrm{dB} - \int_0^1 \varphi_q\,\mathrm{dB}\right)^2\right] = \int_0^1 |\varphi_p - \varphi_q|^2\,\mathrm{d}t.$$

La suite de terme général $\varphi_n - f$ est convergente dans $\mathrm{L}^2[0\,;\,1]$, elle est donc de Cauchy ; la majoration précédente prouve que la suite de terme général $\int_0^1 \varphi_n\,\mathrm{dB}$ est de Cauchy dans $\mathrm{L}^2(\Omega)$; cet espace étant complet, elle converge.

Conseils de lecture

Le lecteur désireux d'un vaste panorama (mathématisé au minimum) sur le mouvement brownien sera sans doute passionné par la lecture de l'article de Bertrand Duplantier, *Le mouvement brownien, "divers et ondoyant"* [28].

Pour un point de vue mathématique et de nombreuses applications, la référence incontournable est Revuz & Yor [81].

Chapitre 32

Estimation bayésienne

Lors de la modélisation probabiliste d'une situation, il arrive qu'une probabilité p, constitutive du modèle, soit inconnue, par manque d'observations, par impossibilité de répéter l'expérience ne serait-ce qu'une seconde fois, ou toute autre raison.

L'approche bayésienne consiste alors à construire une méta-modélisation, en considérant tous les modèles possibles fondés sur le nombre p, et à postuler une loi de distribution a priori, *pour la valeur de p — en général une densité $f(p)$ — qui modélise, plus ou moins précisément, les informations d'ordre psychologique que nous pensons détenir sur p (est-il proche de 0, de 1, a-t-on une bonne confiance en sa valeur ou l'imagine-t-on au contraire bien étalé dans $[0\,;1]$?). Les observations effectuées permettent ensuite de modifier cette distribution, et fournissent une nouvelle évaluation de p.*

Cette approche est assez fréquente mais, malgré des succès certains, parfois critiquée, notamment par les tenants d'une interprétation statistique des probabilités.

Remarque 32.1 Il convient de ne pas mélanger le *théorème de Bayes* (§ 61, page 104), qui est un théorème tout ce qu'il y a de plus solide, et nulle part contesté, et l'*approche bayésienne* des probabilités, qui consiste à utiliser des distributions *a priori* de p.

Distribution *a priori* uniforme, et distribution *a posteriori* § 456

Considérons une situation simple : un mathématicien décide d'aller visiter l'Écosse ; il a entendu dire que les moutons, noirs ou blancs, y sont nombreux et, durant son premier trajet en train dans la campagne, il scrute le paysage par la fenêtre en guettant les ovins. La proportion p de moutons blancs lui est inconnue. Une attitude possible pour lui est d'attribuer, en son for intérieur, une distribution uniforme à p, c'est-à-dire choisir une densité

$$f(p) = \begin{cases} 1 & \text{si } 0 \leqslant p \leqslant 1, \\ 0 & \text{sinon.} \end{cases}$$

Avec cette modélisation, la variable p admet pour valeur moyenne $\hat{p} = \frac{1}{2}$. (Nous ne la notons pas « $\mathbf{E}(p)$ », réservant la notion d'espérance au procédé de calcul effectué dans un univers probabilisé $(\Omega, \mathfrak{T}, \mathbf{P}_p)$, où $\mathbf{P}_p$ est une mesure de probabilité

dépendant du paramètre p.) Si A est un événement quelconque, sa probabilité « moyennée » est donc

$$\mathbf{P}(\mathrm{A}) = \int_0^1 \mathbf{P}_t(\mathrm{A})\,\mathrm{d}t.$$

Remarque 32.2 Cette formulation n'est pas sans engendrer des problèmes : même si A et B sont indépendants dans $(\Omega, \mathfrak{T}, \mathbf{P}_p)$ pour toute valeur de p, on n'a pas forcément $\mathbf{P}(\mathrm{A}\cap\mathrm{B}) = \mathbf{P}(\mathrm{A})\cdot\mathbf{P}(\mathrm{B})$. Par exemple, si $\mathbf{P}_p(\mathrm{A}) = \mathbf{P}_p(\mathrm{B}) = p$ pour tout $p \in [0;1]$ et $\mathbf{P}_p(\mathrm{A}\cap\mathrm{B}) = p^2$, on trouve $\mathbf{P}(\mathrm{A}) = \mathbf{P}(\mathrm{B}) = \frac{1}{2}$ mais $\mathbf{P}(\mathrm{A}\cap\mathrm{B}) = \frac{1}{3}$.

Lorsque des informations nouvelles sont apportées, c'est-à-dire lorsqu'un événement A est réalisé, notre mathématicien voyageur peut chercher une nouvelle densité pour p, conditionnée par l'événement A, en utilisant la formule de Bayes de probabilité des causes. Si par exemple le premier mouton qu'il voit est blanc, il peut estimer que p a plus de chances d'être dans l'intervalle $[1/2;1]$ que dans l'intervalle $[0;1/2]$. Cette impression lui paraîtra confirmée s'il observe un second mouton blanc[1].

On peut obtenir la densité conditionnelle de p, sachant l'événement A, en fonction de la loi *a priori* grâce à la formule de Bayes. Pour cela, écrivons la probabilité que p se trouve dans l'intervalle $[x;\, x+\mathrm{d}x]$ sous la forme $f(x)\,\mathrm{d}x$; on obtient

$$f_{\mathrm{A}}(x)\,\mathrm{d}x = \mathbf{P}\,(x \leqslant p \leqslant x+\mathrm{d}x \,|\, \mathrm{A}) = \frac{\mathbf{P}\big(\{x \leqslant p \leqslant x+\mathrm{d}x\}\cap \mathrm{A}\big)}{\mathbf{P}(\mathrm{A})}$$
$$= \frac{\mathbf{P}\,(\mathrm{A} \,|\, x \leqslant p \leqslant x+\mathrm{d}x)\ \mathbf{P}\,\{x \leqslant p \leqslant x+dx\}}{\mathbf{P}(\mathrm{A})}$$

Si la probabilité de l'événement A dépend continûment de p, on peut identifier les probabilités $\mathbf{P}\,(\mathrm{A} \,|\, x \leqslant p \leqslant x+\mathrm{d}x)$ et $\mathbf{P}(\mathrm{A} \,|\, p = x)$, ce qui nous donne

$$f_{\mathrm{A}}(x) = \frac{\mathbf{P}(\mathrm{A} \,|\, p=x)\ f(x)}{\mathbf{P}(\mathrm{A})} = \frac{\mathbf{P}(\mathrm{A} \,|\, p=x)\ f(x)}{\int_0^1 \mathbf{P}(\mathrm{A} \,|\, p=t)\ f(t)\,\mathrm{d}t.}$$

On reconnaît dans cette formule une extension continue du théorème de Bayes. Dans le dénominateur, on a fait usage de la règle de calcul d'une probabilité moyennée, et noté $\mathbf{P}(\mathrm{A} \,|\, p=t)$ pour $\mathbf{P}_t(\mathrm{A})$.

Avec cette nouvelle densité, on calcule la valeur moyenne de p sachant A :

$$\hat{p} = \int_0^1 x\, f_{\mathrm{A}}(x)\,\mathrm{d}x,$$

qui fournit une estimation *a posteriori* de p, prenant en compte l'observation de l'événement A. Selon l'idée des « bayésiens », cette estimation est meilleure que la précédente, et prend en compte l'observation A.

Remarque 32.3 Que cette estimation soit meilleure ou non est une affaire de confiance, non de mathématiques, nous l'avons dit. Montrons à un observateur bayésien une succession de tirages à *pile* ou *face* avec une pièce qui est réellement équilibrée, mais sans que l'observateur le sache. Au bout de trois tirages, on obtient forcément un décompte inégal entre les *pile* et les *face*, ce qui conduira l'observateur à formuler une estimation de p qui sera différente de $1/2$, et donc en réalité moins bonne que l'estimation *a priori* $p = 1/2$.

(1). Bien sûr, s'il multiplie les observations, une bonne estimation de p lui sera fournie par la valeur limite de la fréquence de moutons blancs, valeur limite dont on sait qu'elle existe si les observations sont indépendantes.

Supposons, pour reprendre notre exemple, que sur les n premières observations, r moutons soient blancs ; on note A l'événement $A = \{X_n = r\}$, où X_n est la variable aléatoire donnant le nombre de moutons blancs après n observations. Sous l'hypothèse d'une fréquence p de moutons blancs, X_n suit une loi binomiale $\mathscr{B}(n\,;\,p)$, c'est-à-dire que

$$\mathbf{P}_p\{X_n = r\} = \binom{n}{r} p^r (1-p)^{n-r}.$$

La densité *a posteriori* de p est donc

$$f_A(p) = \frac{\binom{n}{r} p^r (1-p)^{n-r}}{\int_0^1 \binom{n}{r} t^r (1-t)^{n-r}\,dt},$$

qui est la densité d'une loi $\beta(r+1, n-r+1)$. Notamment, la valeur moyenne de p est[(2)]

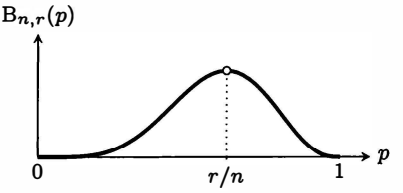
$$\hat{p} = \frac{r+1}{n+2}.$$

Ainsi, l'observation de r moutons blancs parmi les n aperçus conduit à remplacer la valeur *a priori* $p = 1/2$ par la valeur $p = (r+1)/(n+2)$, conçue comme plus représentative du résultat observé.

Remarque 32.4 (Estimation bayésienne et maximum de vraisemblance)
D'autres méthodes existent pour estimer le paramètre p d'après observation. Selon la méthode du **maximum de vraisemblance**, on cherchera plutôt la valeur de p qui maximise $\mathbf{P}_p(A)$; dans l'exemple précédent,

$$\mathbf{P}_p(A) = B_{n,r}(p) := \binom{n}{r} p^r (1-p)^{n-r}$$

est maximisé pour la valeur $p = r/n$.



Lorsque n et r sont grands, c'est-à-dire lorsque les observations sont nombreuses, alors les valeurs estimées de p selon la méthode bayésienne et selon la méthode du maximum de vraisemblance sont quasiment les mêmes. Dans le cas d'un faible nombre d'observations, elles diffèrent sensiblement. Le choix de l'une ou l'autre méthode relève davantage de la psychologie que de la méthodologie scientifique, et donne lieu à des situations amusantes : après l'observation d'un unique mouton, qui se révèle être noir, un personnage utilisant la méthode du maximum de vraisemblance conclura que « en Écosse, tous les moutons sont noirs », tandis qu'un bayésien conclura que deux tiers sont noirs[(3)].

---

(2). Rappelons, et cela se montre par récurrence, que pour tous entiers k et ℓ,

$$B(k+1, \ell+1) := \int_0^1 t^k (1-t)^\ell\,dt = \frac{k!\,\ell!}{(k+\ell+1)!}.$$

(3). Chacun connaît le blague classique, faisant intervenir trois personnages A, B et C dans un train en Écosse, dont les professions supposent (censément) un degré de rigueur croissant (selon l'auditoire, on prend un Chimiste, un Physicien et un Mathématicien, ou encore un Physicien, un Mathématicien et un Logicien). Après observation d'un premier mouton noir, A s'exclame : « Oh, en Écosse, les moutons sont noirs ! » ; ce à quoi B rétorque : « Non, cher collègue, en Écosse, il existe au moins un mouton noir ». C soupire et lâche avec dédain : « En Écosse, il existe au moins un mouton dont au moins un côté est noir. »

Exemple 32.5 (Loi de succession de Laplace) Un exemple des plus célèbres de méthode bayésienne a été formulé par Laplace, dans sa *Théorie analytique des probabilités*[4] . Il s'agissait d'estimer la probabilité qu'un événement se produise à la $(n+1)$-ième expérience, sachant qu'il s'était réalisé les n premières fois. En reprenant l'exemple précédent et en posant $r = n$, on obtient

$$\hat{p} = \frac{n+1}{n+2}.$$

Ainsi, en sachant que le soleil s'est levé tous les jours depuis cinq mille ans[5], la « probabilité » qu'il se lève demain est, selon cette méthode,

$$p = \frac{1\,826\,212}{1\,826\,213} \approx 1 - 5,48 \cdot 10^{-7}.$$

Le lecteur concevra qu'il convient de rester extrêmement critique par rapport à ces méthodes qui, encore une fois, ne traduisent pas une réalité numérique (vérifiable, même de manière approximative, par répétition de l'expérience), mais une simple tentative de quantification du degré de confiance que l'on peut avoir en un événement, et qui est donc *d'ordre purement psychologique.*

§ 457 **Choix d'une autre distribution *a priori***

Dans de nombreuses situations, on n'a pas de raison de supposer une distribution uniforme de p ; on peut alors choisir comme distribution *a priori* une loi bêta, en raison de la grande variété de ses allures (§ 160, page 235). Une méthode pour choisir les bons paramètres est d'évaluer la valeur moyenne p_0 de p que l'on estime la meilleure, ainsi que sa variance V_0 (d'autant plus petite que l'on est confiant sur cette valeur moyenne *a priori*). Les paramètres a et b de la loi bêta sont alors reliés à p_0 et V_0 par

$$a = p_0\left(\frac{p_0(1-p_0)}{V_0} - 1\right) \qquad \text{et} \qquad b = (1-p_0)\,a.$$

Exemple 32.6 Dans le paragraphe § 386, nous avions donné une probabilité *a priori* $p = 2/\ln n$ pour qu'un entier donné n impair soit un nombre premier. Cette probabilité *a priori*, même si elle s'appuie sur un théorème précis (théorème des nombres premiers), n'est *pas* le résultat d'un véritable calcul probabiliste. Nous avions fait comme si nous prenions un entier au hasard dans l'ensemble $\{3, 5, 7, \ldots, n\}$; dans ce cas, la probabilité qu'un tel entier soit premier est bien de l'ordre de $2/\ln n$. Mais bien entendu, pour un entier donné, la question ne se pose pas. Les tests probabilistes (Solovay-Strassen ou Miller-Rabin) permettent de majorer rigoureusement la probabilité que l'entier n passe k tests de Jacobi s'il n'est pas premier ; le point de vue bayésien permet d'évaluer la *vraisemblance* que n soit premier sachant qu'il a passé k tests de Jacobi.

EXERCICES

♦ ***Exercice 32.1*** Donnons à la variable p une loi de distribution *a priori* égale à une loi bêta $B(a, b)$; une variable aléatoire X, censée suivre la loi binomiale $\mathscr{B}(n\,;\,p)$, prend la valeur r. Calculer la loi *a posteriori* de p selon la méthode de l'inférence bayésienne. Montrer que cette loi *a posteriori* est encore une loi bêta. Que se passera-t-il si on réutilise cette loi comme loi *a priori* et que l'on réitère le processus ?

Indications

◊ *Indication 32.1 :* On trouve une loi bêta de paramètres $a + r$ et $b + n - r$. Après k itérations du procédé, on trouve une loi $B\big(a + kr, b + k(n-r)\big)$. Celle-ci converge vers la loi certaine $\delta_{r/n}$ lorsque $k \to \infty$.

---

(4). LAPLACE [68], Livre II, chapitre VI, *De la probabilité des causes et des événements futurs, tirée des événements observés*, paragraphe § 32. Il semblerait que Laplace n'ait pas connu les travaux de Bayes à cette époque ; il a démontré indépendamment la formule intégrale de Bayes.

(5). Date proposée par Laplace comme la plus ancienne époque de l'histoire.

Chapitre 33

Retour à pile ou face : loi de l'Arc sinus et séries aléatoires

Dans ce court chapitre, nous revenons sur les successions infinies de tirages à pile *ou* face.

- *Dans une première partie, nous montrons que, si l'on est momentanément vainqueur dans une partie endiablée de* pile *ou* face, *alors on a de fortes chances de l'être pour longtemps !*
- *Puis, nous revenons à la nature des séries* $\sum \pm 1/n$ *— séries harmoniques avec une distribution aléatoire de signes, et nous montrons que : si les signes sont répartis équitablement, la série est presque sûrement convergente, mais qu'elle diverge presque sûrement dans le cas contraire.*
- *Enfin, nous montrons que le rayon d'une série entière* $\sum A_n z^n$*, dont les coefficients* A_n *sont aléatoires mais indépendants et identiquement distribués, ne saurait prendre (presque sûrement) que trois valeurs :* 0, 1 *et* $+\infty$.

33.1 L'INIQUITÉ FLAGRANTE D'UN JEU ÉQUITABLE

La loi de l'Arc sinus

§ 458

Considérons une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires de Bernoulli indépendantes de loi $\mathscr{B}(1/2)$, et notons $Y_n = 2X_n - 1$, de telle sorte que les Y_n sont des variables de Rademacher indépendantes

$$\mathbf{P}\{Y_n = +1\} = \mathbf{P}\{Y_n = -1\} = \frac{1}{2}.$$

Les variables Y_n modélisent le gain de l'un des joueurs (Rosencrantz) lors d'une partie de *pile* ou *face* : gain de $1\,€$ si la pièce tombe sur *pile*, gain de $-1\,€$ si elle tombe sur *face*. On note $S_n = Y_1 + Y_2 + \cdots + Y_n$ le gain total à l'issue de n parties. Enfin, on note K_n le nombre de sommes $S_1, \ldots, S_n$ positives ou nulles : K_n est donc

la durée, dans l'intervalle $[1\,;\,n]$, durant laquelle Rosencrantz est satisfait d'être globalement gagnant.

■ **Théorème 33.1 (Lévy)** *La suite de variables aléatoires* K_n/n *converge en loi vers la* **loi de l'Arc sinus**, *de fonction de répartition et de densité respectivement égales à*

$$\mathrm{F}(x) = \frac{2}{\pi}\,\mathrm{Arc}\sin\sqrt{x} \qquad f(x) = \frac{1}{\pi}\,\frac{1}{\sqrt{x\,(1-x)}}$$

pour tout $x \in [\,0\,;\,1\,]$.

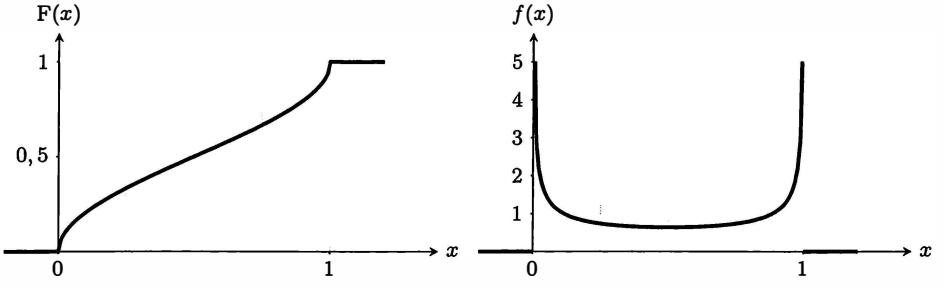


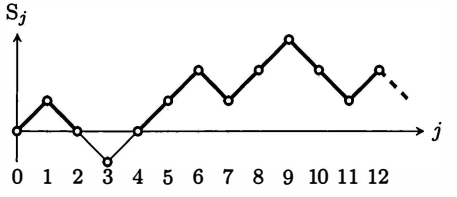
FIGURE 33.1 — *Fonctions de répartition et de densité de la loi de l'Arc sinus.*

Ce comportement peut paraître, de prime abord, assez paradoxal : à l'inspection de la densité, on voit que les valeurs les *moins* probables prises par K_n/n sont celles qui sont proches de $1/2$, alors que les *plus* probables sont celles qui sont proches de 0 et de 1. Le jeu de *pile* ou *face*, archétype du jeu équilibré, a donc tendance à favoriser grandement un des joueurs et à privilégier les situations déséquilibrées. Par exemple, sur une longue partie, Rosencrantz a plus de 20% de chances d'être en tête 90% du temps, et 10% de chances d'être en tête 97,6% du temps !

Cela est dû au fait que, si S_n est positif et de quelque ampleur, il a de fortes chances de le rester un long moment (n'oublions pas que $\mathbf{E}(\mathrm{S}_m \,\|\, \mathrm{S}_n) = \mathrm{S}_n$ pour $m \geqslant n$ et que l'écart-type de $\mathrm{S}_m - \mathrm{S}_n$ augmente, lentement, comme $\sqrt{m-n}$).

Remarque 33.2 (Sur la démonstration du théorème de Lévy) Le théorème de Lévy sur la loi de l'Arc sinus est quelque peu technique à démontrer ; la preuve repose essentiellement sur du dénombrement et des propriétés astucieuses de symétrie des marches aléatoires. Le lecteur intéressé est vivement encouragé à lire une preuve détaillée dans l'ouvrage de Rényi [87] ou, à défaut, dans celui de Shiryaev [91]. En voici un résumé ultra-rapide.

Le principal résultat combinatoire utilisé est le suivant. Durant les $2n$ premiers lancers, traçons un graphe en lignes brisées des gains cumulés :



Alors la probabilité pour que le graphe possède exactement $2k$ segments au-dessus de l'axe des abscisses (en gras ; le lecteur vérifiera que le nombre de tels segments ne saurait être impair) est

$$P_{2n,2k} = \frac{1}{2^{2n}} \binom{2k}{k} \cdot \binom{2n-2k}{n-k}.$$

La formule de Stirling permet de donner un équivalent de cette quantité quand n et $n-k$ sont tous deux grands :

$$P_{2n,2k} \underset{\substack{k\to\infty \\ n-k\to\infty}}{\sim} \frac{1}{\sqrt{\pi k(n-k)}}.$$

Une majoration propre permet alors de trouver le résultat cherché, à savoir :

$$\mathbf{P}\left\{a \leqslant \frac{K_{2n}}{2n} \leqslant b\right\} \underset{n\to\infty}{\sim} \frac{1}{\pi n} \sum_{k=\lfloor na\rfloor+1}^{\lfloor nb\rfloor} \frac{1}{\sqrt{\frac{k}{n}\left(1-\frac{k}{n}\right)}} \xrightarrow[n\to\infty]{} \frac{1}{\pi}\int_a^b \frac{dt}{\sqrt{t(1-t)}}.$$

33.2 SÉRIES HARMONIQUES DE SIGNE ALÉATOIRE

Revenons à la discussion du paragraphe § 307. On cherche à connaître la probabilité que la série $\sum \varepsilon_n/n$ converge lorsque les ε_n sont des variables aléatoires indépendantes, prenant les valeurs $+1$ et -1 uniquement, c'est-à-dire si l'on tire à *pile* ou *face* le signe de chaque terme.

La suite des $1/n$ décroît suffisamment vite pour que, les signes étant choisis au hasard (à *pile* ou *face*), les termes se compensent bien. Cela transparaît si l'on effectue une transformation d'Abel dans cette série :

$$S_n = \sum_{k=1}^{n} \frac{\varepsilon_k}{k} = \sum_{k=1}^{n} \frac{T_k - T_{k-1}}{k} = \sum_{k=1}^{n-1} \frac{T_k}{k(k+1)} + \frac{T_n}{n}.$$

La valeur des T_k augmente lentement et le dénominateur, équivalent à k^2, fait bien converger la série dans la plupart des cas.

Cas d'une pièce déséquilibrée § 459

Si la proportion d'un des signes (« + » ou « − ») l'emporte, on peut se douter que la série $\sum \varepsilon_n/n$ diverge.

■ **Proposition 33.3** *Si la pièce est déséquilibrée, c'est-à-dire si*

$$\mathbf{P}\{\varepsilon_n = 1\} \neq \mathbf{P}\{\varepsilon_n = -1\},$$

alors la série $\sum \varepsilon_n/n$ diverge presque sûrement.

DÉMONSTRATION : Supposons $p = \mathbf{P}\{\varepsilon_n = +1\} \neq 1/2$. Si l'on note $T_n = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n$, alors d'après la loi des grands nombres, on a, avec une probabilité 1, $T_n \sim n(2p-1)$. La transformation d'Abel

$$S_n = \sum_{k=1}^{n} \frac{\varepsilon_k}{k} = \sum_{k=1}^{n} \frac{T_k - T_{k-1}}{k} = \sum_{k=1}^{n-1} \frac{T_k}{k(k+1)} + \frac{T_n}{n}.$$

montre que, toujours avec probabilité 1, $S_n \sim (2p-1)\ln n \to \pm\infty$ (selon que $p > 1/2$ ou $p < 1/2$), et la série diverge. ∎

§ 460 **Cas équilibré**

Le cas du *pile* ou *face* déséquilibré étant réglé, on suppose désormais que les ε_n sont des variables aléatoires de Rademacher :

$$\mathbf{P}\{\varepsilon_n = +1\} = \mathbf{P}\{\varepsilon_n = -1\} = \frac{1}{2},$$

et on va montrer que $\sum \varepsilon_n/n$ converge presque sûrement. C'est un peu plus délicat, la technique employée dans la proposition précédente ne fonctionnant plus.

Tout d'abord, un résultat préliminaire[1].

■ **Proposition 33.4 (Inégalité de Kolmogorov)** *Soient* $X_1, X_2, \dots, X_n$ *des variables aléatoires indépendantes, d'espérance nulle et admettant une variance. On note* $S_k = X_1 + \cdots + X_k$ *pour* $k = 1, \dots, n$. *Alors, pour tout* $\alpha > 0$,

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant n} |S_k| \geqslant \alpha\right\} \leqslant \frac{1}{\alpha^2}\ \mathbf{V}(S_n). \tag{33.1}$$

DÉMONSTRATION : Pour $k = 1, \dots, n$, notons A_k l'événement

$$A_k = \{\, |S_1| < \alpha, \dots, |S_{k-1}| < \alpha, |S_k| \geqslant \alpha\}.$$

Les événements $A_1, A_2, \dots, A_n$ sont disjoints, mais leur union n'est pas Ω tout entier, ce qui permet d'écrire que

$$\mathbf{V}(S_n) = \mathbf{E}(S_n^2) \geqslant \sum_{k=1}^{n} \mathbf{E}(S_n^2\, \mathbf{1}_{A_k}).$$

L'inégalité $(a-b)^2 \geqslant 0$, développée, donne $a^2 \geqslant 2ab - b^2 = b^2 + 2b(a-b)$; on l'applique à $a = S_n$ et $b = S_k$, ce qui donne

$$\mathbf{V}(S_n) \geqslant \sum_{k=1}^{n} \mathbf{E}\Big[\big(S_k^2 + 2S_k(S_n - S_k)\big)\, \mathbf{1}_{A_k}\Big].$$

Maintenant, les variables aléatoires $\mathbf{1}_{A_k}$ et S_k dépendent de $X_1, \dots, X_k$ uniquement. Elles sont donc indépendantes de $S_n - S_k = X_{k+1} + \cdots + X_n$ (lemme des coalitions), ce qui permet d'écrire que

$$\mathbf{E}\big[S_k(S_n - S_k)\, \mathbf{1}_{A_k}\big] = \mathbf{E}(S_k\, \mathbf{1}_{A_k})\, \mathbf{E}(S_n - S_k) = 0.$$

Ainsi,

$$\mathbf{V}(S_n) \geqslant \sum_{k=1}^{n} \mathbf{E}(S_k^2\, \mathbf{1}_{A_k}) \geqslant \sum_{k=1}^{n} \alpha^2\, \mathbf{E}(\mathbf{1}_{A_k}) = \alpha^2 \sum_{k=1}^{n} \mathbf{P}(A_k) = \alpha^2\, \mathbf{P}\left\{\max_{1\leqslant k\leqslant n} |S_k| \geqslant \alpha\right\}. \quad ∎$$

Remarque 33.5 Billingsley interprète cette inégalité ainsi : dans la suite des sommes partielles de variables aléatoires indépendantes, si $\max_{k\leqslant n} |S_k|$ est grand, alors $|S_n|$ est probablement grand lui aussi.

La réponse à la question posée est un cas particulier d'un résultat un peu plus général :

■ **Proposition 33.6** *Soit* $(X_n)_{n\geqslant 0}$ *une suite de variables aléatoires toutes d'espérance nulle et telle que* $\sum \mathbf{V}(X_n)$ *converge. Alors la série* $\sum X_n$ *converge presque sûrement.*

Dans le cas où l'on pose $X_n = \varepsilon_n/n$, on a $\mathbf{V}(X_n) = 1/n^2$ et donc :

■ **Corollaire 33.7 (Séries $\sum \pm 1/n$)** *Presque sûrement, la série* $\sum \varepsilon_n/n$ *converge.*

---

(1). Nous suivons ici la démonstration de Patrick Billingsley [9]. Le lecteur pourra en trouver une autre dans Stein et Shakarchi [93].

DÉMONSTRATION DE LA PROPOSITION 33.6 : Nous allons montrer que, avec une probabilité 1, la série est de Cauchy.

Fixons momentanément un réel $\varepsilon > 0$. Pour tous entiers n et k, on a $S_{n+k} - S_n = X_{n+1} + X_{n+2} + \cdots + X_{n+k}$ et donc, d'après l'inégalité de Kolmogorov (33.1) et pour tout entier $q \geqslant 1$:

$$\mathbf{P}\left\{\max_{1\leqslant k\leqslant q} |S_{n+k} - S_n| \geqslant \varepsilon\right\} \leqslant \frac{1}{\varepsilon^2}\,\mathbf{V}(S_{n+q} - S_n) = \frac{1}{\varepsilon^2}\sum_{k=1}^{q}\mathbf{V}(X_{n+k}).$$

L'axiome 3′ permet de passer à la limite $q \to \infty$ puisqu'à gauche, les événements forment une suite croissante :

$$\mathbf{P}\left\{\sup_{k\geqslant 1} |S_{n+k} - S_n| > \varepsilon\right\} \leqslant \frac{1}{\varepsilon^2}\sum_{k=1}^{\infty}\mathbf{V}(X_{n+k}) = \frac{R_n}{\varepsilon^2},$$

où R_n est le reste d'ordre n de la série convergente $\sum \mathbf{V}(X_i)$. On en déduit immédiatement que

$$\lim_{n\to\infty} \mathbf{P}\left\{\sup_{k\geqslant 1} |S_{n+k} - S_n| > \varepsilon\right\} = 0. \qquad (*)$$

Notons $\mathcal{C}$ l'événement « la série $\sum X_i$ est de Cauchy » :

$$\mathcal{C} = \{\forall \varepsilon > 0 \quad \exists n \in \mathbb{N} \quad \forall p, q \geqslant n \quad |S_p - S_q| \leqslant \varepsilon\} = \bigcap_{\varepsilon>0}\bigcup_{n=1}^{\infty} C(\varepsilon, n)$$

où

$$C(\varepsilon, n) = \{\forall p, q \geqslant n \quad |S_p - S_q| \leqslant \varepsilon\}.$$

L'union est bien croissante. L'intersection est décroissante *mais* porte sur un ensemble non dénombrable ; on préférera donc écrire

$$\mathcal{C} = \bigcap_{N=1}^{\infty}\bigcup_{n=1}^{\infty} C(2^{-N}, n). \qquad (**)$$

Or, chaque ensemble $C(\varepsilon, n)$ contient l'événement $\{\sup_{k\geqslant n} |S_{n+k} - S_n| \leqslant \varepsilon/2\}$, complémentaire de celui de la formule $(*)$, moyennant le fait que ε est maintenant $\varepsilon/2$; on en déduit que

$$\mathbf{P}\left[\bigcup_{n=1}^{\infty} C(\varepsilon, n)\right] = \lim_{n\to\infty} \mathbf{P}(C(\varepsilon, n)) = 1.$$

La formule $(**)$ montre que $\mathcal{C}$ est intersection dénombrable d'événements de probabilité 1, donc est de probabilité 1. ∎

33.3 RAYON D'UNE SÉRIE ENTIÈRE ALÉATOIRE

Puisque nous en sommes dans les séries, voici une autre question : étant donnée une série entière $\sum A_n z^n$, dont les coefficients A_n sont des variables aléatoires, que peut-on dire de la loi du rayon ?

Le problème, en toute généralité, est difficile, et nous nous contentons d'étudier un cas particulier : celui où $(A_n)_{n\geqslant 0}$ est une suite de variables aléatoires indépendantes et de même loi, à valeurs réelles ou complexes.

Le rayon est presque sûrement constant § 461

On cherche à déterminer le rayon de la série entière $\sum A_n z^n$, donné par la formule de Cauchy :

$$\frac{1}{R} = \limsup |A_n|^{1/n}.$$

Ce rayon est une variable aléatoire, qui de plus est mesurable par rapport à la tribu asymptotique

$$\mathfrak{Q}_\infty = \bigcap_{n=0}^{\infty} \sigma(A_n, A_{n+1}, \ldots).$$

D'après la loi du 0–1 de Kolmogorov, elle est presque sûrement constante (voir exemple 20.6 page 420). Nous allons montrer que ce rayon ne peut être égal qu'à 0, à 1 ou à $+\infty$ (du moins « presque sûrement »).

§ 462 **Trois rayons sont possibles**

Rappelons d'abord que, si $\sum a_n z^n$ est une série entière de rayon R, et si c est un réel positif, alors

i) si $|a_n| \leqslant c^n$ à partir d'un certain rang, alors $R_a \geqslant 1/c$;

ii) si $|a_n| \leqslant c^n$ pour une infinité d'indices, alors la suite de terme général $|a_n|\, c^{-n}$ ne tend pas vers 0, donc $R_a \leqslant 1/c$.

Nous allons appliquer ce double critère à notre série aléatoire. Remarquons d'abord que

$$\mathbf{P}\{\, |A_n| \geqslant c^n \} = \mathbf{P}\{\, |A_1| \geqslant c^n \}$$

puisque les A_n ont tous même loi que A_1. Pour tout réel $c > 0$, on a donc

i) si $\sum \mathbf{P}\{|A_1| \geqslant c^n\}$ converge, alors par le lemme de Borel-Cantelli (sens facile), on a

$$\mathbf{P}\{\, |A_n| \geqslant c^n \ ;\ \text{i.s.}\} = 0,$$

ce qui revient à dire que, presque sûrement, $|A_n| < c^n$ à partir d'un certain rang et donc $R \geqslant 1/c$ (presque sûrement).

ii) si au contraire $\sum \mathbf{P}\{|A_1| \geqslant c^n\}$ diverge, c'est-à-dire si $\sum \mathbf{P}\{|A_n| \geqslant c^n\}$ diverge, alors par indépendance des événements $\{|A_n| \geqslant c^n\}$ et par le sens difficile de Borel-Cantelli, on a

$$\mathbf{P}\{\, |A_n| \geqslant c^n \ ;\ \text{i.s.}\} = 1$$

et donc, presque sûrement, $R \leqslant 1/c$.

Reste à déterminer sous quelle(s) condition(s) sur la loi de A_1 on a convergence ou divergence de la série. Tout d'abord, transformons son expression :

$$\mathbf{P}\{\, |A_1| \geqslant c^n\} = \mathbf{P}\{\, \ln|A_1| \geqslant n \ln c\}.$$

Rappelons maintenant le résultat de l'exercice 13.3 page 304 :

■ **Lemme 33.8** *Soit* X *une variable aléatoire, et notons* $X^+ := \max(X, 0)$ *sa partie positive. Soit* $\alpha > 0$. *Alors* X^+ *est intégrable si et seulement si la série*

$$\sum \mathbf{P}\{X > n\alpha\}$$

converge.

On l'applique à la variable $Y = (\ln|A_1|)^+$.

— **1er cas : $(\ln|A_1|)^+$ n'est pas intégrable.** Alors la série

$$\sum \mathbf{P}\{\, \ln|A_1| \geqslant n \ln c\}$$

diverge pour tout $c > 1$, donc $R \leqslant 1/c$ presque sûrement pour tout $c > 1$. Notamment, $R \leqslant 1/p$ presque sûrement pour tout entier $p \geqslant 1$; l'intersection dénombrable d'événements presque sûrs étant presque sûre, on en déduit que $R = 0$ presque sûrement.

— **2e cas : $(\ln|A_1|)^+$ est intégrable.** Alors la série $\sum \mathbf{P}\{\ln|A_1| \geqslant n \ln c\}$ converge pour tout réel $c > 1$, donc $R \geqslant 1/c$ presque sûrement pour tout $c > 1$ et donc, en suivant un raisonnement similaire, $R \geqslant 1$ presque sûrement. Éliminons d'abord le cas où A_1 est une variable presque sûrement égale à 0 (auquel cas le rayon de la série est évidemment $R = +\infty$). Il existe un réel $\eta > 0$ tel que $\mathbf{P}\{|A_1| > \eta\}$ est strictement positif. Par le sens difficile de Borel-Cantelli, la divergence (grossière) de la série $\sum \mathbf{P}\{|A_n| > \eta\}$ implique que $\mathbf{P}\{|A_n| \geqslant \eta \text{ ; i.s.}\} = 1$ et donc $R \leqslant 1$ presque sûrement. On a donc finalement $R = 1$ presque sûrement.

En conclusion, on a montré que :

- si $A_1 = 0$ presque sûrement, alors $R = +\infty$ presque sûrement.
- sinon, et si $(\ln|A_1|)^+$ admet une espérance, alors $R = 1$ presque sûrement ;
- si $(\ln|A_1|)^+$ n'admet pas d'espérance, alors $R = 0$ presque sûrement.

Exemple 33.9 Si A_1 suit la loi normale $\mathcal{N}(0\,;\,1)$, l'espérance de $(\ln|A_1|)^+$, définie dans $[0\,;+\infty]$, vaut

$$\mathbf{E}\big[(\ln|A_1|)^+\big] = \int_0^{+\infty} \mathbf{P}\{\ln|A_1| > t\}\,dt = \int_0^{+\infty} \mathbf{P}\{|A_1| > e^t\}\,dt = \int_0^{+\infty} 2\,G(e^t)\,dt,$$

où G est la fonction de survie de la loi normale. L'équivalent

$$2\,G(e^t) \underset{t\to+\infty}{\sim} \sqrt{\frac{2}{\pi}}\,e^{-t-e^{2t}/2}$$

(voir proposition A.19 page 676) montre que l'intégrale précédente converge et que $\sum A_n z^n$ a pour rayon $R = 1$ presque sûrement.

Exemple 33.10 Considérons maintenant une variable pouvant prendre facilement des valeurs très grandes : on suppose que A_1 admet une densité valant

$$f(t) = \frac{\ln 2}{t \ln^2 t}$$

pour tout $t \geqslant 2$, nulle ailleurs. La fonction de répartition de A_1 est

$$F(x) = 1 - \frac{\ln 2}{\ln x}.$$

L'espérance de $(\ln A_1)^+$ est alors

$$\mathbf{E}\big[(\ln A)^+\big] = \int_0^{+\infty} \big[1 - F(e^t)\big]\,dt = \int_{\ln 2}^{+\infty} \frac{\ln 2}{t}\,dt = +\infty,$$

donc $R = 0$ presque sûrement.

Chapitre 34

En guise de conclusion : à propos de quelques paradoxes

Crois-moi : il y a peu de raisons plus valables de faire des cauchemars que de découvrir, caché au cœur des mathématiques, le germe de la folie.

Daniel KEHLMANN, La Nuit de l'Illusionniste *[58].*

La plupart des paradoxes ont pour cause l'ambiguïté inhérente à toutes les langues ; le passage difficile d'un énoncé en français à un énoncé mathématique censément équivalent engendre aussi bien des résultats amusants que d'épineux problèmes. D'autres paradoxes proviennent d'un résultat rigoureusement démontré, mais non conforme à l'intuition, parfois trompeuse, comme le paradoxe des anniversaires(1) *; un des cas les plus classiques résulte de la confusion entre* ignorance *de la réalisation ou non d'un événement, et* croyance *en l'équiprobabilité des éventualités.*

Enfin, le fameux paradoxe de Saint-Pétersbourg est typique de ce qui arrive lorsque l'on manipule des quantités infinies pour modéliser une situation qui, par nature, ne connaît que des quantités finies.

34.1 BIAIS D'ÉQUIPROBABILITÉ

Biais d'équiprobabilité : le paradoxe des deux cassettes § 463

On appelle *biais d'équiprobabilité* la tendance (fautive) à traduire l'absence de connaissance sur la probabilité de chacune des issues possibles d'une expérience par l'équiprobabilité entre ces issues.

(1). Présenté dans l'exercice 1.1 page 50.

Si le raisonnement suivant paraît idiot :

Je ne sais pas quel temps il fera demain, donc il y a une chance sur deux qu'il pleuve,

les biais d'équiprobabilité peuvent être plus subtils (voir la discussion sur la loi de Benford au paragraphe § 389). Voici un paradoxe classique, proposé sous forme d'une énigme.

*Un génie vous présente deux cassettes. L'une des deux contient un trésor ; l'autre contient également un trésor, deux fois plus important. On vous propose de choisir une des deux cassettes, dont l'aspect extérieur est identique ; vous finissez par en choisir une. Le génie l'ouvre — vous pouvez constater qu'elle contient effectivement un trésor — et vous donne maintenant la possibilité de revenir sur votre décision, et de choisir l'*autre *cassette, ou bien de conserver votre premier choix et de garder le trésor visible. Que faites-vous ?*

La plupart des gens sensés répondront que changer ou non sa décision est globalement sans incidence, et qu'il n'y a pas de stratégie gagnante. Pour les déstabiliser, on peut leur proposer le raisonnement suivant :

Notons x la valeur (connue) du trésor contenu dans la cassette ouverte, et Y *la variable aléatoire égale à la valeur de celui contenu dans l'autre. La deuxième cassette contient soit $2x$, soit $x/2$. Puisque le choix de la première cassette était équiprobable, la probabilité que* Y *soit égale à $2x$ est $1/2$, et la probabilité que* Y *soit égale à $x/2$ est $1/2$. Ainsi, l'espérance de* Y *vaut*

$$\mathbf{E}(\mathrm{Y}) = \frac{1}{2}\cdot 2x + \frac{1}{2}\cdot\frac{x}{2} = \frac{5x}{4} > x.$$

Conclusion : *il est préférable de revenir sur sa décision, et de choisir la 2^{e} cassette, qui contient en moyenne un plus gros trésor.*

Quel est le raisonnement juste ? Il serait dommage de donner immédiatement des éléments de la résolution de ce paradoxe ; nous invitons le lecteur à y réfléchir calmement avant de lire le dernier paragraphe de cette section (§ 465).

§ 464 **Le problème de Monty Hall**

Ce problème est devenu tellement classique que ce n'est plus guère un paradoxe ; mais en son temps, il a dérouté des esprits parmi les plus grands et même — selon la légende — le grand Erdős lui-même. Il est tiré d'un jeu télévisé, *Let's make a Deal*, animé par un présentateur du nom de Monty Hall, dont la dernière épreuve est la suivante.

On présente au candidat trois portes ; derrière deux d'entre elles se trouve une chèvre, derrière la troisième une belle voiture. Le candidat est amené à choisir une des trois portes. Une fois ce choix effectué, le présentateur, plutôt que de lui ouvrir la porte choisie et de lui dévoiler le lot qui lui échoit, lui offre un indice en ouvrant une des deux autres portes — *derrière laquelle se trouve évidemment une chèvre. Le candidat peut alors, s'il le souhaite, revenir sur sa décision et choisir la troisième porte, ou bien au contraire conserver son choix et faire ouvrir la porte initiale.*

Quelle est la meilleure stratégie pour le candidat ?

Un argument souvent donné par ceux qui ne connaissent pas le paradoxe est le suivant :

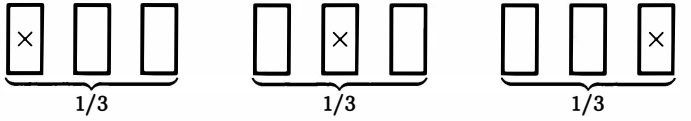
La voiture se trouve obligatoirement derrière l'une des deux portes encore fermées, et avec une probabilité $1/2$ pour chacune. Par conséquent, conserver son choix ou revenir sur sa décision ne modifie en rien la probabilité de gagner.

Comme nous allons le voir, imaginer que la voiture est derrière chacune des portes avec la même probabilité 1/2 est typiquement un biais d'équiprobabilité.

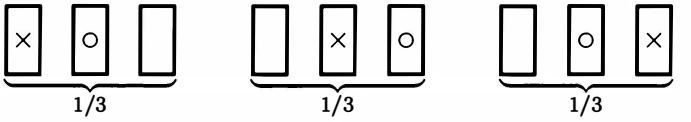
La position de la voiture est choisie aléatoirement avant le début du jeu, et le candidat n'en a pas connaissance. La porte gagnante sera représentée en première position sur nos schémas, indépendamment de sa position physique sur le plateau télévisé. Le joueur choisit (selon sa propre psychologie) une porte, qui peut être

- la porte gagnante avec une probabilité 1/3,
- ou l'une des deux portes perdantes avec la probabilité 1/3 chacune.

Si l'on marque d'une croix la porte choisie, les trois situations suivantes sont *équiprobables* :



Le présentateur dévoile maintenant une des portes, *mais il ne le fait pas au hasard*, parmi les portes non choisies par le candidat, *il choisit toujours* une porte derrière laquelle se trouve une chèvre. Marquons d'un cercle la porte qu'ouvre le présentateur : si le joueur avait deviné la porte gagnante, il a le choix d'ouvrir une des deux portes équivalentes restantes, mais dans les deux autres cas, la porte qu'il ouvre lui est imposée.



Dans le premier cas (probabilité 1/3), le joueur a intérêt à garder sa porte. Dans les deux autres cas (probabilité 2/3), le joueur a intérêt à revenir sur sa décision et choisir la troisième porte. Ainsi : *le joueur gagnera avec une probabilité* 2/3 *s'il revient sur sa décision, mais avec une probabilité* 1/3 *s'il s'entête et conserve son premier choix.*

Retour au problème des deux cassettes § 465

Plusieurs erreurs de raisonnement ont été commises. Tout d'abord, celle de noter la valeur du trésor comme une constante x, alors que c'est une variable aléatoire. Même si l'on ignore quels choix ont présidé à la formation des trésors (le génie peut être pingre, ou libéral, ou peut-être doit-il suivre un règlement précis...), on peut toujours envisager la modélisation du problème au moment où le génie a préparé ses deux cassettes ; l'une présente une somme a, l'autre une somme $2a$. La valeur du trésor est donc $\mathrm{X} = a$ ou $\mathrm{X} = 2a$.

On propose ensuite de considérer que les événements $\{\mathrm{Y} = 2\mathrm{X}\}$ et $\{\mathrm{Y} = \mathrm{X}/2\}$ sont équiprobables — ce qui revient à dire que les événements $\{\mathrm{X} = a\}$ et $\{\mathrm{X} = 2a\}$ sont équiprobables — et c'est là la véritable erreur : on a implicitement admis que *la non connaissance d'un paramètre se traduit par une équiprobabilité — une chance sur deux que ce soit vrai, une chance sur deux que ce soit faux.* Or, il n'y a aucune raison à cela. Pour s'en convaincre, supposons que le génie, familier de cette épreuve et passablement radin, garnisse systématiquement une cassette de 10 € et l'autre de 20 €. Vous choisissez une cassette, vous l'ouvrez, elle contient, disons, 20 €. Y a-t-il une chance sur deux que l'autre cassette contienne 40 € ? Non, bien sûr. *Conditionnellement à l'événement* $\{\mathrm{X} = 20\}$, *on a* $\mathbf{P}\{\mathrm{X} = 2a\} = 1$.

De manière plus générale, postuler l'équiprobabilité (la 2e cassette contient le double ou la moitié de la somme que l'on a observée) implique que les valeurs

possibles d'un trésor soient en nombre infini et que la loi de probabilité de cette somme sur $\mathbb{R}^+$ soit invariante par homothétie de rapport 1/2. Une telle mesure de probabilité n'existe pas. ⊿34.2

34.2 PROBLÈMES DUS À L'INTERPRÉTATION DE L'ESPÉRANCE

§ 466 **L'espérance d'un jeu n'est pas un critère suffisant**

Des paradoxes apparents surgissent lorsque l'on s'imagine l'espérance comme une valeur typique. Ces paradoxes sont davantage psychologiques que mathématiques. Deux exemples simples illustrent notre propos, et montrent que, lorsqu'il s'agit de jouer de l'argent, l'espérance n'est pas un critère suffisant pour décider de la validité d'une stratégie.

Le premier est familier, il s'agit du jeu de Loto. Il est bien connu que l'espérance de ce jeu est *négative*, c'est-à-dire que, en moyenne, on doit perdre de l'argent. Cela n'empêche pas de très nombreux joueurs de s'y adonner, occasionnellement ou régulièrement. Preuve s'il en était besoin que nous n'agissons pas seulement de manière rationnelle.

Dans le second exemple, on considère un jeu de style roulette, dont la probabilité de gain est de une sur mille ; si l'on gagne, on emporte dix mille fois sa mise. L'espérance de gain à ce jeu est donc, pour une mise de M, de

$$\mathbf{E}(\mathrm{G}) = \frac{10\,000}{1000}\,\mathrm{M} - \mathrm{M} = 9\mathrm{M}.$$

Rosencrantz vient jouer à cette roulette ; il dispose des économies de toute une vie, la somme de $\mathrm{N} = 100\,000$ pistoles. Que va-t-il faire ? Jouer d'un coup toutes ses économies ? Jouer pistole par pistole ? Si les deux stratégies ont le même gain moyen — c'est-à-dire la même espérance de gain — la première mène potentiellement à un gain énorme (presque un milliard de pistoles), mais avec une probabilité de 999/1000 à une perte des économies de toute une vie : seul un flambeur complet pourrait la tenter. La seconde stratégie mène, avec une probabilité proche de 1, à un gain net d'environ 900 000 pistoles.

§ 467 **Paradoxe de Saint-Pétersbourg**

Abordons maintenant le jeu de *pile* ou *face* dont l'étude a nourri une correspondance entre des mathématiciens de Paris et de Saint-Pétersbourg, où l'impératrice Catherine II avait réuni à sa cour de nombreux savants et philosophes — au rang desquels Diderot et Euler. Son principe a été énoncé durant une correspondance entre Daniel et Nicolas Bernoulli, et publié dans les *Transactions de l'Académie de Saint-Pétersbourg.*

Un meneur de jeu vous propose le jeu suivant :

- *Vous payez un certain droit d'entrée, égal à* D *euros, avant de commencer la partie.*
- *Le meneur lance une pièce honnête ; s'il obtient* pile*, il vous verse* 2 *euros et la partie s'arrête.*
- *Dans le cas contraire, il relance la pièce ; s'il obtient* pile*, il vous verse* 4 *euros et la partie s'arrête.*
- *Dans le cas contraire, il recommence ; si l'on note* N *le rang du premier* pile *obtenu, il vous verse une somme égale à* 2^{N} *euros.*

Ainsi, si vous acceptez de jouer, après avoir payé le droit d'entrée, vous avez une chance sur deux de gagner 2 euros ; une chance sur quatre d'en gagner 4 ; une chance sur huit d'en gagner 8, etc. La question que l'on pose, et qui est à la base du paradoxe de Saint-Pétersbourg, est la suivante :

Quel droit d'entrée D *êtes-vous prêt à payer pour pouvoir participer au jeu ?*

Si l'on note G le gain brut obtenu, on peut l'écrire $\mathrm{G} = 2^{\mathrm{N}}$, où N est la variable aléatoire égale au rang du premier *pile* et qui suit la loi géométrique $\mathscr{G}(1/2)$. Ainsi,

$$\mathbf{P}\{\mathrm{G} = 2^n\} = \frac{1}{2^n}$$

pour tout entier n supérieur ou égal à 1, et l'espérance de G est infinie :

$$\mathbf{E}(\mathrm{G}) = \sum_{n=1}^{\infty} 2^n \cdot \frac{1}{2^n} = +\infty.$$

Le gain net est $\mathrm{G}' = \mathrm{G} - \mathrm{D}$, d'espérance égale à $+\infty$ quelle que soit la valeur de D. Ainsi, *un joueur rationnel devrait accepter de payer une somme arbitrairement grande pour avoir le droit de participer au jeu.*

Et pourtant, posez-vous la question : *quelle somme accepteriez-vous, vous, de payer ?*

C'est là que réside le paradoxe : vous trouverez facilement des joueurs prêts à payer quelques euros, mais personne de sensé n'acceptera d'en payer mille ou dix mille !

Encore une fois, la valeur de l'espérance d'une variable aléatoire n'est pas une donnée suffisante pour fixer une stratégie. Plusieurs raisons peuvent être évoquées.

- Tout d'abord, il y aurait une sorte d'escroquerie à pousser un joueur à accepter de payer un fort droit d'entrée en argumentant que : « En moyenne, on gagne une infinité d'euros ». En effet, aucun tirage ne mène à un gain infini ou, énoncé différemment, *avec une probabilité* 1, *on va gagner moins, beaucoup moins que l'espérance de gain !* La situation est extrêmement différente de celles où l'espérance de gain est une quantité finie[(2)].
- Si l'on accepte de payer un droit d'entrée $\mathrm{D} = 2^k$ euros, alors la probabilité de sortir bénéficiaire de la partie est $\mathbf{P}\{\mathrm{N} \geqslant k\} = 1/2^k$. Si l'on paye donc 1024 euros, on a moins d'une chance sur mille de sortir bénéficiaire. Même si l'espérance de gain est grande, la probabilité de gagner est tellement faible qu'elle peut paraître négligeable au regard de la somme perdue.
- La loi des grands nombres nous apprend que, si l'on joue une infinité de fois au jeu de Saint-Pétersbourg, en payant toujours un droit d'entrée fixe D, alors, inévitablement, on finira par gagner des sommes si importantes qu'elles assureront un gain effectif tendant vers l'infini.
  Or, ce qui arrivera, c'est que, selon toute probabilité, on commencera par perdre de grandes sommes d'argent : on ne peut espérer gagner au jeu de Saint-Pétersbourg que si l'on a la possibilité de jouer, *à crédit*, durant un temps indéterminé. Avant que la fortune n'arrive, on aura peut-être une dette dépassant la valeur marchande totale existant sur Terre, voire dépassant le prix d'une masse d'or équivalente à la galaxie tout entière ; sans compter qu'on aura sans doute dû passer le jeu à ses enfants, petits-enfants et autres descendants.
  Moralité, les gains exceptionnellement gros, que l'on risque d'attendre un temps plus long que sa propre vie, ne « comptent » pas dans l'espérance de gain obtenu sur un grand nombre de parties, puisqu'en pratique on a toutes les chances de ne jamais les voir.
- Heuristiquement : si l'on décide par exemple d'arrêter de jouer, quoi qu'il arrive, après un million de parties, c'est-à-dire à peu près $2^{20} = 1\,048\,576$ parties — ce qui est déjà une perte de temps considérable — on a de fortes chances de ne pas obtenir de gain supérieur à quelques millions d'euros, et le gain que l'on peut espérer *en pratique* est plutôt de l'ordre de
$$\mathbf{E}(\mathrm{G} \cdot \mathbf{1}_{\{\mathrm{G} \leqslant 2^{20}\}}) = \sum_{k=1}^{20} 2^{-k} \cdot 2^k = 20,$$
  résultat tout à fait ridicule.
- Plus mathématiquement, si l'on note S_n le gain après la n-ième partie (sans compter la mise), alors S_n est de l'ordre de $n \ln_2 n$ et, plus précisément, on a convergence en probabilité de $\mathrm{S}_n / n \ln_2 n$ vers 1 (la notation $\ln_2(x) = \ln x / \ln 2$ désigne le logarithme en base 2). Cela signifie que la suite $(\mathrm{S}_n)_{n \geqslant 1}$ croît très lentement. Ainsi, si le joueur doit miser une somme de $\mathrm{D} = 100$ euros pour chaque partie jouée, alors son gain ne dépasse ses pertes que lorsque $\ln_2 n$ est de l'ordre de 100, c'est-à-dire au bout de 2^{100} parties. À raison d'une partie par seconde, ou même d'un milliard, le temps requis dépasse — de nombreux ordres de grandeur — l'âge de l'univers.

(2). Lorsque l'espérance de gain est une quantité finie, la possibilité de gagner moins que l'espérance entraîne automatiquement la possibilité de gagner strictement plus.

- Enfin, sur un plan moins mathématique, les gains astronomiques sont illusoires au niveau de l'utilité : si gagner un milliard d'euros ($\approx 2^{30}$) peut être considéré agréable, en gagner mille milliards risque fort de n'être aucunement plus avantageux. On peut donc, *sans être moralement lésé*, écrêter le gain, en remplaçant le gain effectif G par un *gain utile* $\varphi(G)$. La fonction φ dépend de la psychologie du joueur, mais doit saturer aux grandes valeurs, par exemple $\varphi(x) = \min(x, 2^{30})$. L'espérance du gain utile est alors

$$\sum_{n=1}^{\infty} \varphi(2^n) \cdot \frac{1}{2^n} = 31,$$

ce qui encore une fois est *très* faible, et justifie que personne ne s'aventure à payer de droit d'entrée dépassant quelques euros, voire quelques dizaines d'euros pour les têtes brûlées.
Cette approche du paradoxe a été proposée par Daniel Bernoulli et reprise par Laplace dans sa *Théorie analytique des probabilités* [68, page 439].

La figure suivante illustre bien le côté erratique du gain : quelques montées spectaculaires du gain moyen, suivi d'une baisse régulière. Pour observer l'augmentation logarithmique de la moyenne des gains

$$M_n = \frac{G_1 + G_2 + \cdots + G_n}{n},$$

il faudrait moyenner un grand nombre de courbes de ce style.

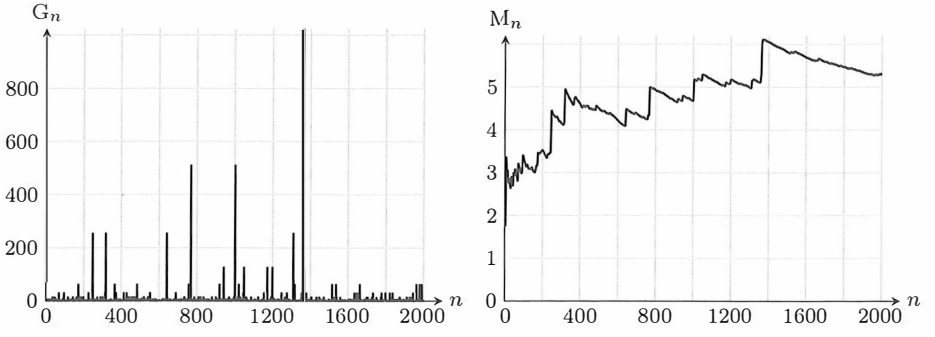
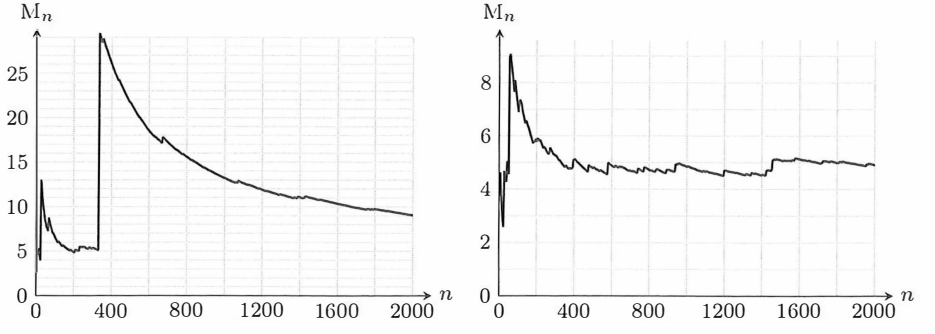


FIGURE 34.1 — *Paradoxe de Saint-Pétersbourg.* (À gauche :) *les gains successifs* G_n *au cours de 2000 parties.* (À droite :) *les moyennes de gains* $M_n = (G_1 + \cdots + G_n)/n$ *au cours des mêmes 2000 parties.*

Ci-après, deux autres courbes des gains moyens illustrent la diversité des allures possibles de l'évolution de M_n.



34.3 PSEUDO-PARADOXES

Certains résultats heurtent l'intuition, ou le sens commun, et sont de ce fait appelés paradoxes, comme le paradoxe des anniversaires ou le paradoxe de l'autobus (ou de l'inspection, exercices 2.1 et 30.3). Insistons bien sur le fait qu'ils ne sont, justement, en rien de véritables paradoxes (pas plus que le paradoxe de Banach-Tarski n'en est un), mais simplement des résultats *étonnants*. Un des plus curieux concerne les jeux non transitifs.

Le paradoxe de Walter Penney § 468

Walter PENNEY (1953–2002) était un ingénieur américain, passionné de mathématiques, qui énonça en 1969 le paradoxe qui suit dans le *Journal of Recreational Mathematics*.

On considère un jeu dans lequel deux joueurs s'affrontent. On tire une suite infinie à *pile* ou *face* (avec une pièce équilibrée) ; chaque joueur doit, avant le début des tirages, proposer une séquence de trois résultats (comme « PFP » ou « FFF »). Le joueur qui, le premier, voit sa séquence apparaître dans la suite des tirages est déclaré vainqueur[3].

> *Existe-t-il une stratégie (un choix parmi les* $2^3 = 8$ *motifs existants) qui soit toujours gagnante, c'est-à-dire meilleure que toutes les autres ?*

Ce qu'a montré Penney, c'est que la réponse est négative : *il n'existe pas de stratégie optimale*. Quelle que soit la stratégie choisie par le joueur n° 1, le joueur n° 2, s'il en a connaissance, peut choisir un motif qui apparaîtra avant le motif adverse avec une probabilité strictement supérieure à $1/2$[4]. On peut donc notamment trouver des « chemins » dont chaque élément est battu par le suivant, et qui bouclent. Par exemple : la stratégie FPP est battue par FFP, qui est battue par PFF, qui est battue par PPF, qui est battue par FPP !

(Un calcul explicite de probabilité de gain entre deux motifs est exposé à l'exercice 27.4 page 575, dans le chapitre consacré aux chaînes de Markov.)

L'aspect paradoxal de ce résultat, authentiquement étonnant, vient de ce que l'on a tendance à associer, au vocable « meilleur que », une propriété de transitivité : si A est meilleur que B et que B est meilleur que C, alors on s'attend à ce que A soit meilleur que C. Ici, l'erreur commune est de penser, par exemple, que si le temps moyen d'attente d'un motif M_1 est inférieur au temps d'attente d'un motif M_2, alors la probabilité que M_1 apparaisse avant M_2 est plus grande que $1/2$. Ce n'est évidemment pas le cas, et le résultat de Penney montre qu'il n'y a pas d'ordre naturel entre les stratégies : le jeu de Walter Penney fait partie de la classe des **jeux non transitifs**[5]. ⇗34.3

(3). La section 9.2 présente le temps d'attente moyen associé à quelques stratégies.

(4). Pour ceux qui désirent le mettre en pratique, il y a une recette simple pour trouver un tel motif. Si vous laissez votre adversaire choisir un motif ABC de son choix, il vous suffit de parier sur le motif $\overline{B}AB$, où $\overline{B}$ est le tirage opposé à B. Ainsi, le motif FPP est battu, selon cette méthode, par $\overline{F}PP = FFP$. Le calcul est détaillé dans l'exercice 27.4 page 575.

(5). Un exemple minimal de jeu non transitif : aux échecs, Eugène bat toujours la vieille gouvernante Filippievna, et la gouvernante bat toujours sa jeune protégée Tatiana. Mais quand Eugène se mesure à Tatiana, il perd ses moyens et ses parties.

EXERCICES

♦ ***Exercice 34.1 (Deux autres cassettes)*** On vous propose le jeu suivant : deux coffres fermés sont présentés, le premier contenant une somme X_1 de loi uniforme sur $[0\,;\,1]$, le second une somme X_2 de même loi uniforme sur $[0\,;\,1]$. Les variables aléatoires X_1 et X_2 sont supposées indépendantes. On vous ouvre le premier coffre, dévoilant ainsi la valeur de X_1. Deux possibilités s'offrent à vous : garder pour vous la valeur X_1, ou bien ouvrir le deuxième coffre et empocher la valeur X_2.

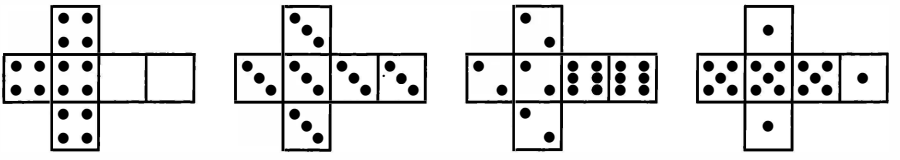
i) Quel est votre gain attendu si vous gardez systématiquement X_1 ? Si vous ouvrez systématiquement le deuxième coffre et empochez X_2 ?

ii) Vous optez pour une stratégie mixte : vous choisissez un nombre c fixé, et au moment où l'on vous présente la somme X_1, vous comparez X_1 à c. Si $X_1 \leqslant c$, vous dédaignez cette somme et demandez à empocher la somme X_2. Si au contraire $X_1 > c$, vous vous contentez de X_1.

Quelle est votre espérance de gain ? Quelle est la valeur de c vous permettant de maximiser votre gain moyen ?

♦ ***Exercice 34.2*** Montrer qu'il n'existe pas de mesure de probabilité sur $]0\,;+\infty[$ qui soit invariante par homothétie de rapport 2.

♦ ***Exercice 34.3 (Des dés non transitifs[6])*** Guildenstern présente à Rosencrantz les dés que lui a offerts le statisticien Bradley EFRON, de l'université de Stanford, et propose une partie selon le protocole suivant : Rosencrantz prend un dé de son choix, puis Guildenstern peut choisir le sien. Ensuite, ils lancent tous deux leur dé : le gagnant est celui qui obtient le plus grand nombre de points. Ces dés sont un peu spéciaux et présentent l'allure suivante une fois le cube « déplié » : le nombre de points varie entre 0 et 6.



dé A dé B dé C dé D

Montrer que : A bat B avec une probailité $p = 2/3$; de même, B bat C avec la probabilité $p = 2/3$; de même, C bat D, et D bat A, toujours avec la probabilité $p = 2/3$. Le jeu n'est donc pas transitif, et Guildenstern pourra se refaire une santé.

♦ ***Exercice 34.4 (Paradoxe de Galton)*** On lance simultanément trois pièces. Nécessairement, deux (au moins) présentent une face identique. La troisième pièce a une chance sur deux de présenter la même face. *Ergo*, il y a une chance sur deux que ces trois pièces présentent la même face.

(6). Ce jeu de dés non transitifs est présenté dans un article de Michel Criton [19].

SOLUTIONS

♦ **Solution 34.1**

i) Que l'on choisisse X_1 ou X_2 de manière déterministe, le gain moyen est le même est vaut $\mathbf{E}(X_1) = \mathbf{E}(X_2) = 1/2$.

ii) Le gain vaut maintenant

$$G(\omega) = \begin{cases} X_2(\omega) & \text{si } X_1(\omega) \leqslant c, \\ X_1(\omega) & \text{si } X_1(\omega) > c. \end{cases}$$

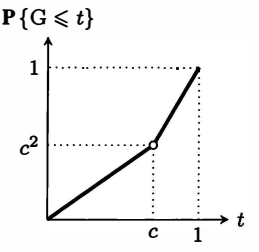
La formule des probabilités totales permet de calculer la fonction de répartition de G : pour tout $t \in [0\,;\,1]$,

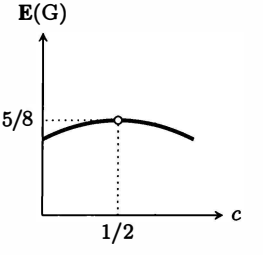
$$\begin{aligned} \mathbf{P}\{G \leqslant t\} &= \mathbf{P}\{G \leqslant t, X_1 \leqslant c\} + \mathbf{P}\{G \leqslant t, X_1 > c\} \\ &= \mathbf{P}\{G \leqslant t \,|\, X_1 \leqslant c\}\, \mathbf{P}\{X_1 \leqslant c\} + \mathbf{P}\{G \leqslant t \,|\, X_1 > c\}\, \mathbf{P}\{X_1 > c\} \\ &= \begin{cases} ct & \text{si } t \leqslant c, \\ ct + (t - c) & \text{si } t > c. \end{cases} \end{aligned}$$

ce qui montre que G admet une densité $g'(t) = c$ sur $[0\,;\,c[$ et $g'(t) = c + 1$ sur $]c\,;\,1]$. L'espérance de G se calcule enfin par l'une des formules

$$\mathbf{E}(G) = \int_0^1 t\, g(t)\, dt = \int_0^1 \mathbf{P}\{G > t\}\, dt = \frac{1}{2} + \frac{c(1-c)}{2} \geqslant \frac{1}{2}.$$

Les valeurs extrêmes $c = 0$ et $c = 1$ correspondent aux deux stratégies de la première question. Si $0 < c < 1$, le gain moyen est strictement supérieur, avec une valeur optimale en $c = 1/2$, valant $\mathbf{E}(G) = 5/8$.





♦ **Solution 34.2** Supposons qu'une telle probabilité **P** existe. Notons $I_0 = [1\,;\,2[$ et $m := \mathbf{P}(I_0)$. Pour tout entier relatif n, notons $I_n := [2^n\,;\,2^{n+1}[$. Par hypothèse, chaque intervalle I_n est de probabilité m. Comme $]0\,;\,+\infty[$ s'écrit comme l'union disjointe $]0\,;\,+\infty[= \sum_{n\in\mathbb{Z}} I_n$, on en déduit

$$1 = \mathbf{P}(]0\,;\,+\infty[) = \sum_{n\in\mathbb{Z}} m.$$

Si $m > 0$, on obtient $1 = +\infty$ (ce qui faux) ; si $m = 0$, on obtient $1 = 0$ (ce qui est également faux).

Une probabilité sur $]0\,;\,+\infty[$ invariante par homothétie de rapport 2 ne saurait exister.

Conseils de lecture

Sur le paradoxe de Saint-Pétersbourg, de nombreuses analyses ont été proposées ; nous suggérons la lecture de Borel.

- Émile Borel, *Probabilités et certitude*, coll. Que sais-je ? [11].

Une variante instructive du paradoxe de Saint-Pétersbourg y est également discutée.

Sur d'autres aspects paradoxaux, et pour le plaisir de la lecture :

- Paul Deheuvels, *La probabilités, le hasard et la certitude*, coll. Que sais-je ? [20, chapitre XIV, notamment].

Annexes

Annexe A

Éléments d'analyse et d'algèbre

On résume ici quelques points d'analyse et d'algèbre utilisés dans l'ouvrage : suites, séries, familles sommables, suites de fonctions, séries entières, fonctions dérivables, convexité, pseudo-inverse, calcul intégral, calcul dans les espaces de Hilbert, transformation de Fourier, théorème spectral notamment.

A.1 SUITES

Fini, dénombrable, indénombrable A1

Un ensemble A est dit **(strictement) dénombrable** s'il existe une bijection $\varphi : \mathbb{N} \to A$ (que l'on appelle un **dénombrement**). En d'autres termes, on a pu *numéroter* tous les éléments de A par des entiers, c'est-à-dire écrire $A = \{a_n \ ; \ n \in \mathbb{N}\}$, en notant plus commodément $a_n := \varphi(n)$. ✦ On dit souvent d'un ensemble qu'il est **dénombrable** pour indiquer qu'il est *fini* ou *strictement dénombrable*. ✦ Une réunion dénombrable d'ensembles dénombrables est dénombrable. ✦ Si A est un ensemble et D un ensemble dénombrable, et s'il existe une injection $\varphi : A \to D$, alors A est dénombrable. ✦ De même, s'il existe une surjection $\psi : D \to A$ et si D est dénombrable, alors A est dénombrable. ✦ Si A et B sont dénombrables, alors $A \times B$ l'est également. ✦ Les ensembles $\mathbb{N}$ et $\mathbb{Z}$ sont évidemment dénombrables. En revanche, $[0\,;1]$, $\mathbb{R}$, $\mathbb{C}$, $\mathfrak{P}(\mathbb{N})$,... ne le sont pas. Deux cas particuliers interviennent fréquemment :

■ **Proposition A.1** $\mathbb{Q}$ *est dénombrable.*

DÉMONSTRATION : À tout couple (p,q) de $\mathbb{Z}\times\mathbb{N}^*$, on peut associer un rationnel r ; l'application $(p,q) \mapsto r$ ainsi définie est une *surjection* de l'ensemble dénombrable $\mathbb{Z}\times\mathbb{N}^*$ dans $\mathbb{Q}$, donc $\mathbb{Q}$ est dénombrable. ∎

■ **Proposition A.2** *L'ensemble* $\{0,1\}^{\mathbb{N}}$ *n'est pas dénombrable.*

DÉMONSTRATION : Suivons l'argument classique de la *diagonale de Cantor* et supposons qu'il existe une bijection $\varphi : \mathbb{N} \to \{0,1\}^{\mathbb{N}}$. Pour chaque entier n, notons

$$\varphi(n) =: (\varepsilon_0(n), \varepsilon_1(n), \ldots, \varepsilon_k(n), \ldots).$$

Rangeons ces valeurs dans un tableau, en faisant apparaître la diagonale :

n	$\varphi(n)$						
0	$\varepsilon_0(0)$	$\varepsilon_1(0)$	$\varepsilon_2(0)$	$\varepsilon_3(0)$	$\cdots$	$\varepsilon_k(0)$	$\cdots$
1	$\varepsilon_0(1)$	$\varepsilon_1(1)$	$\varepsilon_2(1)$	$\varepsilon_3(1)$	$\cdots$	$\varepsilon_k(1)$	$\cdots$
2	$\varepsilon_0(2)$	$\varepsilon_1(2)$	$\varepsilon_2(2)$	$\varepsilon_3(2)$	$\cdots$	$\varepsilon_k(2)$	$\cdots$
3	$\varepsilon_0(3)$	$\varepsilon_1(3)$	$\varepsilon_2(3)$	$\varepsilon_3(3)$	$\cdots$	$\varepsilon_k(3)$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	
k	$\varepsilon_0(k)$	$\varepsilon_1(k)$	$\varepsilon_2(k)$	$\varepsilon_3(k)$		$\varepsilon_k(k)$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	

On considère enfin la suite $a = (a_0, a_1, a_2, \dots, a_k, \dots)$ de terme général $a_k = 1 - \varepsilon_k(k)$. Alors a ne saurait être l'image d'aucun entier par φ : en effet, a et $\varphi(n)$ diffèrent de leur n-ième terme. Cela contredit la surjectivité de φ. ∎

Comme l'ensemble $[0;1]$ est en bijection[1] avec $\{0,1\}^{\mathbb{N}}$, on en déduit que :

■ **Proposition A.3** *L'intervalle* $[0;1]$ *n'est pas dénombrable. Plus généralement, aucun intervalle de* $\mathbb{R}$ *d'intérieur non vide n'est dénombrable.*

A2 **Limite supérieure, limite inférieure**

Toute suite réelle monotone est convergente dans $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$. Si $u = (u_n)_{n\geqslant 1}$ est une suite réelle, on définit les suites $(s_n)_{n\geqslant 1}$ et $(i_n)_{n\geqslant 1}$ par

$$s_n = \sup_{k\geqslant n} u_k \qquad \text{et} \qquad i_n = \inf_{k\geqslant n} u_k.$$

La suite s est décroissante, tandis que la suite i est croissante. On définit la **limite supérieure** et la **limite inférieure** de u comme

$$u^* = \limsup_{n\to\infty} u_n := \lim_{n\to\infty} s_n \qquad \text{et} \qquad u_* = \liminf_{n\to\infty} u_n := \lim_{n\to\infty} i_n.$$

On a toujours $u_* \leqslant u^*$, avec égalité si et seulement si la suite u converge vers cette valeur commune. ✦ u^* est également la borne supérieure de l'ensemble des valeurs d'adhérence de u (qui est un fermé), et u_* sa borne inférieure. Si l'ensemble de ces valeurs d'adhérence est majoré (minoré), u^* en est donc le maximum (u_* en est le minimum). ✦ Pour tout réel x, si $x > \limsup u_n$, alors $x > u_n$ à partir d'un certain rang (c'est-à-dire pour tout $n \geqslant N$, pour un certain N) ; si $x < \limsup u_n$, alors $x < u_n$ pour une infinité d'indices n. ✦ Ce qui rend cette notion si pratique est la possibilité de « passer à la limite supérieure/inférieure » (ce que l'on ne peut pas faire avec la limite ordinaire tant que l'on n'a pas prouvé son existence) : si $u_n \leqslant v_n$ pour tout entier n (ou au moins à partir d'un certain rang), alors

$$\limsup_{n\to\infty} u_n \leqslant \limsup_{n\to\infty} v_n \qquad \text{et} \qquad \liminf_{n\to\infty} u_n \leqslant \liminf_{n\to\infty} v_n.$$

A3 **Suites de Cauchy**

Une suite $(u_n)_{n\geqslant 1}$ à valeurs dans un espace vectoriel normé $\big(E, \|\cdot\|\big)$ vérifie le **critère de Cauchy** ou, plus simplement, est une **suite de Cauchy** si, pour tout $\varepsilon > 0$, il existe un entier N tel que

$$\forall p, q \in \mathbb{N} \qquad p, q \geqslant N \implies \|u_p - u_q\| \leqslant \varepsilon.$$

---

[1]. Pour établir la bijection susdite, définissons d'abord l'application s qui à une suite $(a_n)_{n\geqslant 1}$ à valeurs dans $\{0,1\}$ associe le réel $x = \sum a_n 2^{-n}$. C'est une surjection de $\Omega := \{0,1\}^{\mathbb{N}^*}$ vers $[0;1]$, mais pas une injection (les suites $(0,0,1,1,1,1,\dots)$ et $(0,1,0,0,0,0,\dots)$ définissent le même réel). On définit l'ensemble Ω' en retranchant à $\{0,1\}^{\mathbb{N}^*}$ l'ensemble *dénombrable* des suites stationnaires en 1 ; alors la restriction de s à Ω' est une bijection de Ω' sur $[0;1[$. Enfin, Ω et Ω' sont équipotents car $\Omega - \Omega'$ est dénombrable tandis que Ω ne l'est pas.

✦ Toute suite de Cauchy est bornée. ✦ Toute suite convergente est de Cauchy. ✦ Toute suite de Cauchy admettant une valeur d'adhérence est convergente et sa limite est cette valeur d'adhérence. ✦ Dans un espace vectoriel normé complet ($\mathbb{R}$, $\mathbb{C}$ et les espaces vectoriels normés de dimension finie notamment), toute suite de Cauchy est convergente.

Comparaison de suites A4

Soient $(u_n)_{n\geqslant 1}$ et $(v_n)_{n\geqslant 1}$ deux suites réelles ou complexes. On définit les notions de **domination** et de **négligeabilité** :

$$u_n = \mathrm{O}(v_n) \iff \exists \mathrm{M} \geqslant 0 \quad \exists \mathrm{N} \in \mathbb{N} \quad \forall n \geqslant \mathrm{N} \qquad |u_n| \leqslant \mathrm{M}\,|v_n|$$
$$u_n = \mathrm{o}(v_n) \iff \forall \varepsilon > 0 \quad \exists \mathrm{N} \in \mathbb{N} \quad \forall n \geqslant \mathrm{N} \qquad |u_n| \leqslant \varepsilon\,|v_n|\,.$$

On dit que $(u_n)_{n\geqslant 1}$ et $(v_n)_{n\geqslant 1}$ sont **équivalentes**, et on écrit $u_n \sim v_n$, si $u_n - v_n = \mathrm{o}(u_n)$. La relation « $\sim$ » est une relation d'équivalence. ✦ Dans le cas où la suite $(v_n)_{n\geqslant 1}$ ne s'annule pas, les relations de domination, de négligeabilité et d'équivalence se récrivent plus simplement :

$$u_n = \mathrm{O}(v_n) \iff \left(\frac{u_n}{v_n}\right)_n \text{ est bornée} \qquad u_n = \mathrm{o}(v_n) \iff \frac{u_n}{v_n} \to 0$$
$$u_n \sim v_n \iff \frac{u_n}{v_n} \to 1.$$

Formule de Stirling A5

On rappelle l'équivalent classique de Stirling ainsi que le développement plus précis

$$n! \underset{n\to\infty}{\sim} \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n} \qquad n! = \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}\left(1 + \mathrm{O}\left(\frac{1}{n}\right)\right).$$

A.2 SÉRIES NUMÉRIQUES

Définitions A6

La **suite des sommes partielles** associée à une suite $u = (u_n)_{n\geqslant 1}$ est la suite de terme général

$$\mathrm{S}_n(u) := u_1 + u_2 + \cdots + u_n.$$

On dit que **la série** $\sum \boldsymbol{u_n}$ converge si $\big(\mathrm{S}_n(u)\big)_{n\geqslant 1}$ converge ; dans le cas de la convergence, la **somme** de la série $\sum u_n$ est

$$\sum_{n=1}^{\infty} u_n := \lim_{n\to\infty} \mathrm{S}_n(u).$$

✦ Si $\sum u_n$ converge, alors $u_n \to 0$. Par contraposée, on en déduit que si $u_n \not\to 0$, alors la série $\sum u_n$ diverge (on dit qu'elle diverge **grossièrement**).

Séries positives A7

Si $u_n \geqslant 0$ pour tout $n \in \mathbb{N}$, la série $\sum u_n$ converge si et seulement si la suite $(\mathrm{S}_n)_{n\geqslant 1}$ est majorée. ✦ On en déduit que si $0 \leqslant u_n \leqslant v_n$ et si $\sum v_n$ converge, alors $\sum u_n$ converge. ✦ La série $\sum 1/n^\alpha$ converge si et seulement si $\alpha > 1$. ✦ La **série de Bertrand** $\sum 1/(n^\alpha \ln^\beta n)$ converge si et seulement si $\alpha > 1$ ou ($\alpha = 1$ et $\beta > 1$). ✦ Le cas de la série harmonique $\sum 1/n$ est important : la série diverge, mais *lentement* :

$$\sum_{k=1}^{n} \frac{1}{k} = \ln n + \gamma + \mathrm{o}(1)$$

où $\gamma \approx 0,5772157$ est la constante d'Euler-Mascheroni. ✦ Dans $[0\,;+\infty] = \mathbb{R}^+ \cup \{+\infty\}$, la suite $(\mathrm{S}_n)_{n\geqslant 0}$ des sommes partielles est toujours convergente et l'écriture $\sum_{n=0}^{\infty} u_n$ a un sens. On a alors l'équivalence

$$\sum u_n \text{ converge} \iff \sum_{n=0}^{\infty} u_n < +\infty.$$

A8 **Théorème de comparaison de séries positives**
Soient $\sum a_n$ et $\sum b_n$ deux séries à termes positifs. Si $\sum b_n$ converge et si l'une des conditions suivantes est vérifiée :

$$\forall n \in \mathbb{N} \quad 0 \leqslant a_n \leqslant b_n, \qquad a_n = O(b_n), \qquad a_n = o(b_n), \qquad a_n \sim b_n,$$

alors $\sum a_n$ converge.

A9 **Produits infinis**
Soit $(u_n)_{n\geqslant 1}$ une suite de réels de $[0\,;1[$. La suite de terme général $P_n = \prod_{k=1}^{n}(1-u_k)$ est positive et décroissante, elle admet donc une limite, et on note

$$\prod_{k=1}^{\infty}(1-u_k) := \lim_{n\to\infty}\prod_{k=1}^{n}(1-u_k).$$

■ **Proposition A.4** *Soit $(u_n)_{n\geqslant 1}$ une suite de réels de $[0\,;1[$; alors $\prod_{k=1}^{\infty}(1-u_k)=0$ si et seulement si $\sum u_n$ diverge.*

DÉMONSTRATION : Le résultat est immédiat si la suite $(u_n)_{n\geqslant 1}$ ne tend pas vers 0. Si elle tend vers 0, on écrit

$$-\ln P_n = -\sum_{k=0}^{n}\ln(1-u_k)$$

et puisque $-\ln(1-u_k)\sim u_k$, les séries *positives* $\sum -\ln(1-u_k)$ et $\sum u_k$ ont même nature. Dans le cas de la divergence, on en déduit que $-\ln P_n \to +\infty$ et donc $P_n \to 0$; dans le cas de la convergence, on en déduit que $-\ln P_n$ admet une limite finie, et donc P_n admet une limite non nulle. ∎

A10 **Convergence absolue**
On dit que $\sum u_n$ **converge absolument** si $\sum |u_n|$ converge, ou encore $\sum_{n=0}^{\infty}|u_n| < +\infty$.

■ **Proposition A.5 (Inégalité triangulaire)** *Si la série $\sum u_n$ converge absolument, alors elle converge et*

$$\left|\sum_{n=1}^{\infty}u_n\right| \leqslant \sum_{n=1}^{\infty}|u_n|.$$

DÉMONSTRATION : On décompose la suite $(u_n)_{n\geqslant 1}$ en ses parties positive et négative :

$$u_n^+ := \max(u_n, 0) \qquad \text{et} \qquad u_n^- := \max(-u_n, 0).$$

Pour tout entier n, on a

$$0 \leqslant u_n^+ \leqslant |u_n| \qquad 0 \leqslant u_n^- \leqslant |u_n| \qquad u_n^+ - u_n^- = u_n \qquad \text{et} \qquad u_n^+ + u_n^- = |u_n|$$

La convergence de $\sum|u_n|$ implique alors la convergence de $\sum u_n^+$ et $\sum u_n^-$, et donc la convergence de $\sum(u_n^+ - u_n^-) = \sum u_n$. Enfin, on a les inégalités

$$|S_n(u)| = |S_n(u^+) - S_n(u^-)| \leqslant S_n(u^+) + S_n(u^-) = S_n(|u|)$$

d'où, par passage à la limite, $S(u) = S(u^+) - S(u^-)$ et $|S(u)| \leqslant S(|u|)$. ∎

Cette propriété reste vraie si $(u_n)_{n\geqslant 1}$ est à valeurs dans $\mathbb{C}$, ou dans un espace vectoriel normé de dimension finie comme $\mathbb{R}^d$ ou $\mathbb{C}^d$ (de manière générale, la convergence absolue implique la convergence si et seulement si l'espace vectoriel normé est complet).

A11 **Exponentielle complexe**
Si z est un complexe, son exponentielle est définie comme la somme de la série absolument convergente $\sum z^n/n!$:

$$e^z := \sum_{n=0}^{\infty}\frac{z^n}{n!},$$

■ **Théorème A.6** *Pour tout complexe z, $\lim_{n\to\infty}\left(1+\frac{z}{n}\right)^n = e^z$.*

DÉMONSTRATION : La propriété est bien connue si $z = x$ est un réel, et se démontre par un simple développement limité :

$$\left(1+\frac{x}{n}\right)^n = \exp\left[n\ln\left(1+\frac{x}{n}\right)\right] = \exp\left[n\left(\frac{x}{n}+o\left(\frac{1}{n}\right)\right)\right] = \exp\left[x+o(1)\right] \to e^x.$$

Reste à examiner le cas complexe.

Pour k fixé, la suite de terme général $a_{n,k} = \binom{n}{k}/n^k$ croît vers $1/k!$ quand $n \to \infty$. Or, pour tout entier n,

$$\left|e^z - \left(1+\frac{z}{n}\right)^n\right| \leqslant \left|\sum_{k=0}^{\infty}\frac{z^k}{k!} - \sum_{k=0}^{n} a_{n,k}\,z^k\right| \leqslant \sum_{k=0}^{n}\left|\frac{1}{k!}-a_{n,k}\right||z|^k + \sum_{k=n+1}^{\infty}\frac{|z|^k}{k!}$$
$$\leqslant \sum_{k=0}^{n}\left(\frac{1}{k!}-a_{n,k}\right)|z|^k + \sum_{k=n+1}^{\infty}\frac{|z|^k}{k!} = e^{|z|} - \left(1+\frac{|z|}{n}\right)^n$$

et cette dernière quantité converge vers 0 quand $n \to \infty$ d'après la propriété établie précédemment sur $\mathbb{R}$. ∎

D'autres démonstrations sont possibles (par exemple en écrivant $1 + z/n$ sous forme polaire), mais celle-ci a l'avantage de se généraliser à d'autres ensembles que le corps $\mathbb{C}$ (en fait, à toute algèbre normée complète), par exemple pour montrer que $\left(I + \frac{1}{n}A\right)^n \to \exp(A)$ pour toute matrice carrée $A \in \mathfrak{M}_p(\mathbb{R})$.

■ **Proposition A.7** *Si $u_n \sim z/n$, alors*

$$\lim_{n\to\infty}(1+u_n)^n = e^z. \tag{A.1}$$

Là encore, la propriété est immédiate si les u_n sont réels ; la démonstration dans le cas complexe est identique à la précédente.

Changement d'ordre de sommation A12

La somme d'une série n'est pas une addition ordinaire — elle résulte de sommations et d'un processus de limite. Pour cette raison, l'addition n'est plus commutative. Si φ est une bijection de $\mathbb{N}^*$ sur lui-même (c'est-à-dire une permutation de $\mathbb{N}^*$), la série $\sum u_{\varphi(n)}$ peut être de nature différente de celle de $\sum u_n$; même en cas de convergence des deux, les sommes peuvent différer.

Exemple A.8 (Changer l'ordre change la somme) Posons $u_n = (-1)^{n-1}/n$ pour tout $n \geqslant 1$; on sait que la série $\sum u_n$ converge et que sa somme vaut $S = \ln 2$. Notons S'_n les sommes partielles de la série $\sum u_{\varphi(n)}$, où l'on a défini $\varphi : \mathbb{N}^* \to \mathbb{N}^*$ par

$$\varphi(3k) = 4k \qquad \varphi(3k+1) = 2k+1 \qquad \varphi(3k+2) = 4k+2.$$

Ainsi, quand on regarde la succession des $u_{\varphi(n)}$, on a un terme positif, suivi de deux termes négatifs, puis un positif, deux négatifs, etc. :

n	1	2	3	4	5	6	7	8	
u_n	1	$-\frac{1}{2}$	$\frac{1}{3}$	$-\frac{1}{4}$	$\frac{1}{5}$	$-\frac{1}{6}$	$\frac{1}{7}$	$-\frac{1}{8}$	⋯
$u_{\varphi(n)}$	1	$-\frac{1}{2}$	$-\frac{1}{4}$	$\frac{1}{3}$	$-\frac{1}{6}$	$-\frac{1}{8}$	$\frac{1}{5}$	$-\frac{1}{10}$	⋯

Le lecteur pourra vérifier que $S'_{3n} = \frac{1}{2}S_{2n}$ pour tout n, et que par conséquent la série $\sum u_{\varphi(n)}$ converge et a pour somme $S/2 = \frac{1}{2}\ln 2$. Plus généralement, si l'on se fixe α réel (ou même égal à $\pm\infty$), il existe une permutation σ telle que $\sum u_{\sigma(n)}$ a pour somme α.

On dira que $\sum u_n$ est **commutativement convergente** si elle converge et si, pour toute permutation φ de $\mathbb{N}^*$, la série $\sum u_{\varphi(n)}$ converge et a même somme que $\sum u_n$.

■ **Théorème A.9 (Dirichlet)** *Une série est commutativement convergente si et seulement si elle est absolument convergente.*

On retiendra notamment que la somme d'une série absolument convergente n'est pas modifiée par une permutation de ses termes, et que les séries *positives* ne peuvent être que divergentes ou commutativement convergentes.

A13 **Familles sommables positives**

La notion de famille sommable généralise celle de série ; on se donne une famille $A = (a_i)_{i\in I}$ de réels positifs, et on cherche à les sommer. Cela présuppose que l'ensemble des indices i pour lesquels $a_i \neq 0$ est au plus dénombrable : si ce n'était pas le cas, il existerait un entier n_0 tel que $a_i \geqslant 1/n_0$ pour une infinité (non dénombrable) d'indices, ce qui interdirait la sommation en tout sens raisonnable. Dans la suite, on supposera pour plus de facilité que l'ensemble I d'indexation est dénombrable. La différence entre la théorie des séries et celle des familles sommables est qu'ici, l'ensemble I n'est pas supposé ordonné, c'est-à-dire que *l'on n'impose pas l'ordre dans lequel on va (tenter de) sommer les* a_i.

Si tous les a_i sont positifs, on dit que la famille $(a_i)_{i\in I}$ est **sommable** si l'on peut en sommer un nombre fini aussi grand que voulu sans jamais dépasser une valeur finie — c'est-à-dire s'il existe un réel $M \geqslant 0$ tel que, pour toute sous-famille finie $J \subset I$, $\sum_{j\in J} a_j \leqslant M$. La **somme** de la famille $(a_i)_{i\in I}$ est alors

$$\sum_{i\in I} a_i := S(A) := \sup_{\substack{J\subset I \\ J \text{ fini}}} \sum_{j\in J} a_j.$$

C'est une quantité réelle si la famille est sommable, et qui vaut $+\infty$ dans le cas contraire. Toute manière de sommer la famille conduit au même résultat :

■ **Proposition A.10** *Si* $(i_n)_{n\geqslant 1}$ *est un dénombrement de* I, *alors* $\lim_{n\to\infty} \sum_{k=1}^{n} a_{i_k} = S(A)$, *cette relation étant valable dans* $\mathbb{R}^+ \cup \{+\infty\}$.

Notamment, si $(J_n)_{n\geqslant 1}$ est une suite croissante de parties de I dont l'union vaut I, c'est-à-dire si $J_n \uparrow I$, alors

$$S(A) = \lim_{n\to\infty} \sum_{j\in J_n} a_j.$$

A14 **Familles sommables de signes quelconques**

Si les a_i sont des réels de signe quelconque, ou complexes, on dit que $(a_i)_{i\in I}$ est **sommable** si la famille de réels positifs $(|a|_i)_{i\in I}$ est sommable. Le résultat essentiel est alors que :

■ **Théorème A.11** *Si la famille* $(|a|_i)_{i\in I}$ *est sommable, alors la série* $\sum a_{i_k}$ *est commutativement convergente pour tout dénombrement* $(i_n)_{n\geqslant 1}$ *de* I, *et on peut définir la* **somme** *de la famille* $(a_i)_{i\in I}$ *comme*

$$S(A) := \lim_{n\to\infty} \sum_{k=1}^{n} a_{i_k}.$$

De même que précédemment, si la famille est sommable, et si $(J_n)_{n\geqslant 1}$ est une suite de parties de I telle que $J_n \uparrow I$, alors $S(A) = \lim_{n\to\infty} \sum_{j\in J_n} a_j$.

A15 **Théorème de Fubini pour les séries doubles**

Si $(u_{m,n})_{m,n\geqslant 1}$ est une famille de réels à double indice, on a, quels que soient les entiers p et q :

$$\sum_{m=1}^{p} \sum_{n=1}^{q} u_{m,n} = \sum_{n=1}^{q} \sum_{m=1}^{p} u_{m,n},$$

puisque les deux sommes se font sur un nombre fini d'indices. En revanche, on n'a pas forcément égalité entre les quantités suivantes (si elles existent !)

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} u_{m,n} \qquad \text{et} \qquad \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} u_{m,n}.$$

On peut s'en convaincre grâce à l'exemple suivant, où l'indice i est celui des lignes et l'indice j celui des colonnes :

$$(u_{m,n}) = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & -1 & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & & & \ddots & \end{pmatrix}$$

Le lecteur vérifiera que

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} u_{m,n} = 0 \qquad \text{mais} \qquad \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} u_{m,n} = 1.$$

On peut même donner un exemple plus frappant, en posant

$$(u_{m,n}) = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 2 & -2 & 0 & 0 & 0 & \dots \\ 0 & 0 & 3 & -3 & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & & \ddots & \ddots & \end{pmatrix}$$

ce qui conduit à

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} u_{m,n} = \sum_{m=1}^{\infty} 0 = 0 \qquad \text{mais} \qquad \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} u_{m,n} = \sum_{n=1}^{\infty} 1 = +\infty.$$

Comme pour les séries simples, la convergence absolue autorise la sommation dans n'importe quel ordre :

■ **Théorème A.12 (Fubini)** *Soit $(u_{m,n})_{(m,n)\in\mathbb{N}^2}$ une famille de réels ou de complexes.*

i) Si cette famille est sommable, alors sa somme peut se calculer en sommant dans n'importe quel ordre, notamment

$$S = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} u_{m,n} = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} u_{m,n} = \sum_{k=0}^{\infty}\sum_{\ell=0}^{k} u_{k,k-\ell}.$$

Plus généralement, si $(I_n)_{n\geqslant 0}$ est une partition de $\mathbb{N}^2$, alors

$$S = \sum_{n=0}^{\infty}\sum_{(m,n)\in I_n} u_{m,n}.$$

ii) Si, pour tout $n \in \mathbb{N}$, la série $\sum_m |u_{m,n}|$ converge, notons $A_n = \sum_{m=0}^{\infty} |u_{m,n}|$; si de plus la série positive $\sum A_n$ converge, alors la famille $(u_{m,n})_{m,n\in\mathbb{N}}$ est sommable, et la somme peut se calculer comme précédemment.

En d'autres termes, la condition $\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} |u_{m,n}| < +\infty$ assure la sommabilité de la famille.

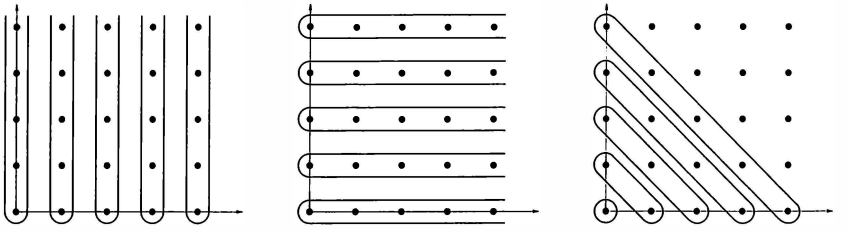


FIGURE A.1 — *Trois méthodes équivalentes de sommation lorsque la série double converge absolument.*

En pratique, on retiendra le corollaire suivant :

■ **Corollaire A.13** *Si $(u_{m,n})_{(m,n)\in\mathbb{N}^2}$ est une famille de réels positifs, on peut toujours intervertir les symboles de sommation, les sommes étant toujours définies dans la demi-droite achevée $\mathbb{R}^+ \cup \{+\infty\}$ et égales.*

A16 **Sommation par paquets**
Le théorème de Fubini pour les séries doubles se généralise :

■ **Théorème A.14 (Théorème de Fubini par paquets)** *Soit* $\mathrm{A} = (a_i)_{i\in \mathrm{I}}$ *une famille de réels ou de complexes, et soit* $(\mathrm{I}_n)_{n\geqslant 0}$ *une partition de* I. *On définit, pour tout* $n \in \mathbb{N}$ *:*

$$u_n := \sum_{i\in \mathrm{I}_n} |a_i|$$

Alors la famille $\mathrm{A} = (a_i)_{i\in \mathrm{I}}$ *est sommable si et seulement si* $\sum u_n$ *converge ; en cas de convergence, elle a pour somme*

$$\mathrm{S}(\mathrm{A}) := \sum_{n=0}^{\infty} \sum_{i\in \mathrm{I}_n} a_i.$$

A17 **Théorème de convergence dominée discrète**
Le théorème de convergence dominée permet de déduire le résultat discret suivant, souvent utile :

■ **Lemme A.15 (Convergence dominée discrète)** *Soient* $v_n(k)$ $(n, k = 1, 2, \dots)$ *des réels, tels que* $v_n(k) \to v(k)$ *quand* n *tend vers l'infini, pour tout* k. *On suppose de plus qu'il existe une suite positive* $\big(w(k)\big)_{k\geqslant 1}$ *telle que* $\big|v_n(k)\big| \leqslant w(k)$ *pour tous* k *et* $n \geqslant 1$, *et telle que* $\sum w(k)$ *converge. Alors*

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} v_n(k) = \sum_{k=1}^{\infty} v(k).$$

A.3 FONCTIONS

A18 **Continuité, continuité à droite**
Soit I un intervalle de $\mathbb{R}$. Une application $f : \mathrm{I} \to \mathbb{R}$ est **continue en un point** $x \in \mathrm{I}$ si et seulement si, pour toute suite $(x_n)_{n\geqslant 0}$ à valeurs dans I et de limite x, la suite de terme général $f(x_n)$ converge vers $f(x)$. La continuité en x équivaut à :

$$\forall \varepsilon > 0 \quad \exists \eta > 0 \quad \forall y \in \mathrm{I} \qquad |y - x| \leqslant \eta \implies \big|f(y) - f(x)\big| \leqslant \varepsilon.$$

La fonction f est **continue sur I** si elle est continue en chaque point de I :

$$\forall x \in \mathrm{I} \quad \forall \varepsilon > 0 \quad \exists \eta > 0 \quad \forall y \in \mathrm{I} \qquad |y - x| \leqslant \eta \implies \big|f(y) - f(x)\big| \leqslant \varepsilon.$$

Si $\mathrm{I}^+ := \mathrm{I} \cap]\, x\,;\, +\infty\,[$ n'est pas réduit à un point, on dit que f est **continue à droite** en x si la restriction de f à I^+ est continue en x. Alors f est continue à droite en x si et seulement si, pour toute suite $(x_n)_{n\geqslant 0}$ à valeurs dans I^+ et de limite x, la suite de terme général $f(x_n)$ converge vers $f(x)$.

Une fonction $f : \mathrm{I} \to \mathbb{R}$ est **uniformément continue** si

$$\forall \varepsilon > 0 \quad \exists \eta > 0 \quad \forall (x, y) \in \mathrm{I}^2 \qquad |y - x| \leqslant \eta \implies \big|f(y) - f(x)\big| \leqslant \varepsilon.$$

La différence avec la définition de la continuité est que le réel η peut être choisi en fonction de ε seulement, et non en fonction de ε *et* de x.

Exemple A.16 La fonction $x \mapsto \sqrt{x}$ est uniformément continue sur $\mathbb{R}$, mais pas la fonction $x \mapsto x^2$.

■ **Théorème A.17 (Heine)** *Soit* $f : \mathrm{K} \to \mathbb{R}$ *une application définie sur un compact* K *de* $\mathbb{R}$ *ou de* $\mathbb{C}$ *(par exemple, un segment* $[\, a\,;\, b\,]$*). Alors* f *est uniformément continue sur* K.

Fonctions à variation bornée A19

Soit $F : [a ; b] \to \mathbb{C}$ une fonction définie sur un segment. Si $\sigma = (a_1, a_2, \dots, a_n)$ est une subdivision de $[a ; b]$ (l'entier n étant quelconque), c'est-à-dire un n-uplet tel que

$$a = a_1 < a_2 < a_3 < \cdots < a_{n-1} < a_n = b,$$

on note

$$\Delta F(\sigma) := \sum_{k=0}^{n-1} \left|F(a_{k+1}) - F(a_k)\right|.$$

On dit que F est à **variation bornée** sur $[a ; b]$ s'il existe un réel $M \geqslant 0$ tel que $\Delta F(\sigma) \leqslant M$ pour toute subdivision σ. ✦ Une fonction $F : \mathbb{R} \to \mathbb{C}$ est dite **à variation bornée sur $\mathbb{R}$** si elle est à variation bornée sur tout segment de $\mathbb{R}$.

■ **Proposition A.18** *Une fonction est à variation bornée si et seulement si elle est différence de deux fonctions monotones. Notamment (théorème A.21), toute fonction à variation bornée est dérivable presque partout.*

Fonctions absolument continues A20

Selon la définition originelle, une fonction $F : I \to \mathbb{C}$ définie sur un intervalle I est dite **absolument continue** si, pour tout $\varepsilon > 0$, il existe un réel $\eta > 0$ vérifiant : pour toute famille de segments $[a_i ; b_i]$, deux à deux disjoints sauf éventuellement en un point, telle que $\sum_i |b_i - a_i| \leqslant \eta$, on ait $\sum_i \left|F(b_i) - F(a_i)\right| \leqslant \varepsilon$. Une fonction absolument continue est donc uniformément continue, et donc continue.

Si $f : I \to \mathbb{C}$ est une fonction intégrable, si a est un point quelconque de I et si l'on pose

$$F(x) := \int_a^x f(t)\,dt$$

pour tout réel x, alors F est absolument continue.

La réciproque est également vraie : toute fonction F absolument continue est à variation bornée, donc dérivable presque partout (théorème A.18), et si l'on note f la dérivée de F (définie presque partout), alors pour tous $a, b \in I$ on a

$$F(b) - F(a) = \int_a^b f(t)\,dt.$$

Fonction Γ d'Euler A21

On définit, pour tout $x > 0$:

$$\Gamma(x) := \int_0^{+\infty} t^{x-1}\,e^{-t}\,dt.$$

La fonction Γ est de classe $\mathscr{C}^\infty$ sur $]0 ; +\infty[$ et vérifie, pour tout $x > 0$, la relation $\Gamma(x+1) = x\,\Gamma(x)$. ✦ Notamment, $\Gamma(1) = 1$ et $\Gamma(n+1) = n!$ pour tout entier n. ✦ La fonction Γ admet également les deux représentations suivantes, dues à Euler :

$$\Gamma(x) = \lim_{n\to\infty} \frac{n^x\,n!}{x(x+1)\cdots(x+n)}$$

et

$$\ln\Gamma(x+1) = \left(x + \frac{1}{2}\right)\ln x - x + \ln\sqrt{2\pi} - \int_0^{+\infty} \frac{\varphi(u)}{x+u}\,du,$$

où $\varphi : u \mapsto u - \lfloor u \rfloor - 1/2$.

Fonction d'erreur, fonction de survie de la loi normale A22

Notons

$$G(x) = 1 - \mathfrak{N}(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-t^2/2}\,dt,$$

la **fonction de survie de la loi normale**. Alors

■ **Proposition A.19 (Équivanlent de la fonction de survie normale)** *On a l'équivalent*

$$\mathrm{G}(x) \underset{x\to+\infty}{\sim} \frac{1}{\sqrt{2\pi}\,x}\,\mathrm{e}^{-x^2/2}, \tag{A.2}$$

cet équivalent étant d'ailleurs un majorant strict de G *sur* $]0\,;+\infty[$.

Pour démontrer cette formule, on effectue une intégration par parties :

$$\sqrt{2\pi}\,\mathrm{G}(x) = \int_x^{+\infty} \mathrm{e}^{-t^2/2}\,\mathrm{d}t = \int_x^{+\infty} \frac{1}{t}\left(t\,\mathrm{e}^{-t^2/2}\right)\,\mathrm{d}t = \frac{\mathrm{e}^{-x^2/2}}{x} - \int_x^{+\infty} \frac{\mathrm{e}^{-t^2/2}}{t^2}\,\mathrm{d}t$$

qui montre déjà la majoration stricte annoncée. Enfin, on majore le dernier terme par

$$0 \leqslant \int_x^{+\infty} \frac{\mathrm{e}^{-t^2/2}}{t^2}\,\mathrm{d}t \leqslant \frac{1}{x^2}\int_x^{+\infty} \mathrm{e}^{-t^2/2}\,\mathrm{d}t = \frac{\sqrt{2\pi}\,\mathrm{G}(x)}{x^2},$$

ce qui montre qu'il est négligeable devant $\mathrm{G}(x)$ et nous fournit l'équivalent.

A.4 FONCTIONS CROISSANTES

Dans toute cette section, F et G sont des fonctions définies sur $\mathbb{R}$, à valeurs réelles, et croissantes (au sens large).

A23 **Continuité, continuité à gauche, à droite**
La fonction F est **continue à droite** en un réel x si, pour toute suite $(x_n)_{n\geqslant 0}$ décroissante et de limite x, la suite de terme général $f(x_n)$ converge vers $f(x)$. Elle est **continue à gauche** en x si cette propriété est vérifiée pour les suites croissantes. Elle est **continue** en x si elle est à la fois continue à gauche et à droite.

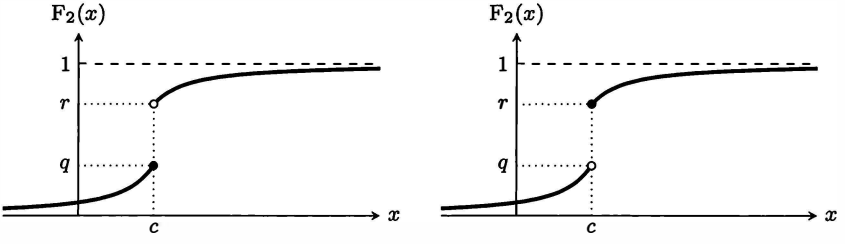


FIGURE A.2 — *Fonctions continues à gauche en c (resp. à droite en c).*

A24 **Discontinuités de 1re et 2e espèce.**
Une fonction de la variable réelle discontinue en un point x peut admettre en ce point une limite à gauche et une limite à droite : on parle alors de **discontinuité de première espèce**. Si elle n'admet pas de limite à gauche ou à droite, c'est une **discontinuité de deuxième espèce** (voir figure A.3).

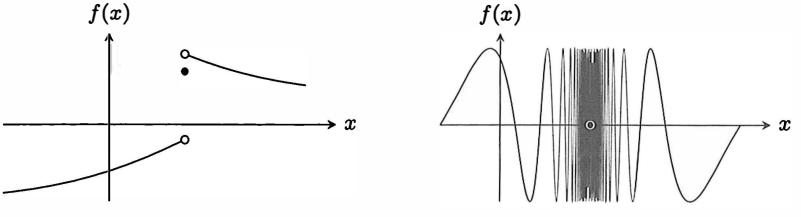


FIGURE A.3 — *Discontinuités de 1re et de 2e espèce.*

Sauts de discontinuité d'une fonction croissante A25

Si F est une fonction croissante, alors elle n'admet que des discontinuités de 1re espèce. En un point a de discontinuité, c'est-à-dire tel que $F(a^-) < F(a^+)$, on appelle **saut de discontinuité** de F la quantité $\sigma_F(a) := F(a^+) - F(a^-)$ (voir figure A.4).

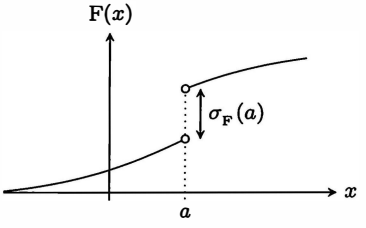


FIGURE A.4 — *Saut de discontinuité de* F *en a.*

■ **Proposition A.20** *Les points de discontinuité d'une fonction croissante sont dénombrables.*

DÉMONSTRATION : Soit $n \in \mathbb{Z}$, notons $I_n = [\,n\,;\,n+1\,[$ et montrons que les points de discontinuité de f sur I_n sont dénombrables. Pour tout $k \in \mathbb{Z}$, notons $D_{n,k}$ l'ensemble des discontinuités en lesquelles le saut de f est compris entre 2^{-k} et $2^{-(k+1)}$. Chaque $D_{n,k}$ est un ensemble fini, puisque la somme des sauts ne saurait excéder $f(n+1) - f(n)$. Ainsi, l'ensemble des discontinuités sur I_n, qui est

$$D_n = \bigcup_{k \in \mathbb{Z}} D_{n,k}$$

est dénombrable. L'ensemble

$$D = \bigcup_{n \in \mathbb{Z}} D_n$$

des discontinuités de f est encore dénombrable. ∎

Régularité d'une fonction croissante A26

Les résultats suivants concernent la théorie de l'intégration selon Lebesgue. On en trouvera des démonstrations dans Billingsley [9] ou dans Rudin [83].

■ **Proposition A.21** *Toute fonction croissante* $F : \mathbb{R} \to \mathbb{R}$ *est presque partout dérivable ; sa dérivée* F' *est positive et* $\int_a^b F'(t)\,dt \leqslant F(b) - F(a)$ *pour tous* a, b *tel que* $a \leqslant b$.

On remarquera qu'il n'y a qu'une *inégalité*, et pas forcément l'égalité, qui n'est réalisée que pour les fonctions absolument continues.

■ **Proposition A.22** *Si* $f : \mathbb{R} \to \mathbb{R}$ *est positive et intégrable, et si l'on pose*

$$F(x) := \int_{-\infty}^{x} f(t)\,dt$$

pour tout réel x, *alors* F *est presque partout dérivable et sa dérivée* F' *coïncide avec* f *sauf sur un ensemble de mesure de Lebesgue nulle.*

En résumé : si f est positive et intégrable, alors intégrer puis dériver redonne la fonction f (à un ensemble de mesure nulle près). En revanche, si F est croissante (et donc presque partout dérivable), dériver puis intégrer ne permet pas de retrouver la fonction initiale.

Décomposition de Lebesgue A27

Les théorèmes précédents permettent de décomposer toute fonction croissante en la somme de trois fonctions : l'une discontinue, la seconde absolument continue, et une troisième. Dans le cas de fonctions de répartition, ce théorème, dû à Lebesgue, prend la forme suivante :

■ **Théorème A.23 (Lebesgue)** *Soit* F *une fonction de répartition. Alors il existe trois fonctions croissantes* F_d, F_{ac} *et* F_s *telles que* $F = F_d + F_{ac} + F_s$ *et*

- F_d *est étagée (à valeurs dans un ensemble dénombrable) ;*
- F_{ac} *est absolument continue ;*
- F_s *est de dérivée presque partout nulle (***fonction singulière***).*

Si l'on note $\{a_k\ ;\ k \in K\}$ *l'ensemble (dénombrable) des points de discontinuité de* f*, la fonction* F_d *vaut*

$$F_d(x) = \sum_{k\ ;\ a_k \leqslant x} F(a_k) - F(a_k^-)$$

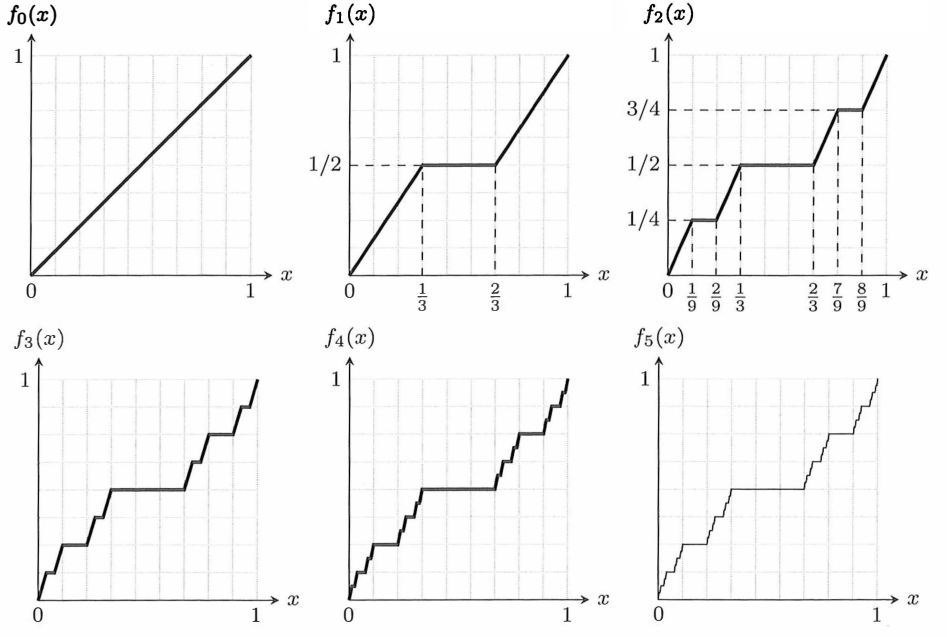
en tout réel x*. La fonction* $F - F_d$ *est continue et croissante ; elle est donc presque partout dérivable et on définit*

$$F_{ac}(x) := \int_{-\infty}^{x} (F - F_d)',$$

fonction absolument continue, et de dérivée presque partout égale à $(F - F_d)'$*. Enfin,* $F_s := F - F_d - F_{ac}$ *est singulière puisque de dérivée presque partout nulle.*

A28 **Une fonction singulière : la fonction de Cantor**

Voici un exemple analytique simple de fonction singulière non constante. Définissons la suite $(f_n)_{n\geqslant 0}$ de fonctions de $[0\,;1]$ dans $\mathbb{R}$ de la façon suivante.



Cette suite converge uniformément vers une fonction f (**fonction de Cantor**). La conver-
⇗A.2 gence étant uniforme, la fonction f est continue sur $[0\,;1]$. Une fois qu'une fonction f_n a établi un nouveau « palier » (il y en a 2^n, de largeur $1/3^n$ chacun), toutes les fonctions f_p $(p \geqslant n)$ restent constantes sur ce palier. Ainsi, f est dérivable sur l'union des intervalles ouverts correspondant à chaque palier, et sa dérivée y est nulle. La longueur totale de cette union est

$$\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots = \sum_{n=1}^{\infty} \frac{2^{n-1}}{3^n} = \frac{1}{3}\frac{1}{1-2/3} = 1,$$

ce qui montre que f est presque partout dérivable et de dérivée presque partout nulle : c'est une fonction singulière.

Pseudo-inverse d'une fonction de répartition A29

Dans ce paragraphe, on se limite aux fonctions F qui sont des fonctions de répartition, c'est-à-dire que l'on suppose que F est croissante, continue à droite, et admettant une limite nulle en $-\infty$ et une limite égale à 1 en $+\infty$. Le problème que l'on cherche à résoudre est de trouver une fonction G qui soit aussi proche que possible d'un « inverse » de F.

Si F est continue et strictement croissante, on sait qu'elle établit une bijection de $\mathbb{R}$ sur $]0\,;\,1[$ et sa bijection réciproque, $G = F^{-1}$, existe. On a alors l'équivalence

$$t = F(x) \iff x = G(t)$$

c'est-à-dire que l'on a les deux relations $F \circ G = \mathrm{Id}_{]0;1[}$ et $G \circ F = \mathrm{Id}_{\mathbb{R}}$.

De manière générale, si $K : A \to B$ et $L : B \to A$ sont des fonctions, la relation $L \circ K = \mathrm{Id}_A$ implique que L est surjective ; la condition $K \circ L = \mathrm{Id}_B$ implique que L est injective. Dans le cas qui nous intéresse, il peut arriver que F ne soit pas bijective, et ceci pour deux types de raisons :

- *défaut d'injectivité*, lorsque F présente un « palier », c'est-à-dire que F est constante sur un intervalle $[a\,;\,b]$ de $\mathbb{R}$;
- *défaut de surjectivité*, lorsque F admet un saut de discontinuité en un point c : aucun point de l'intervalle $[f(c^-)\,;\,f(c)[$ n'est atteint — voir la figure A.5.

Pour contourner ces problèmes, on définit le **pseudo-inverse** G de F par :

$$G(y) = \inf\{x \in \mathbb{R}\,;\ F(x) \geqslant y\}. \tag{A.3}$$

Cette formule est prise pour toutes les valeurs y dans $[0\,;\,1]$, étant entendu que : si l'ensemble du membre de droite de l'équation (A.3) est vide, alors sa borne inférieure est $G(y) = +\infty$; s'il est non borné inférieurement, il est alors égal à $\mathbb{R}$ et sa borne inférieure est $G(y) = -\infty$. On pose, de même, $F(-\infty) = 0$ et $F(+\infty) = 1$. (Voir figure A.6.)

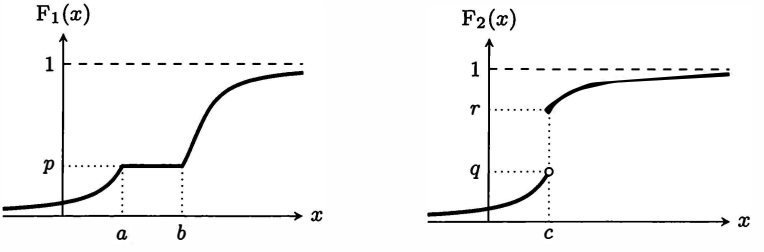


FIGURE A.5 — *Deux défauts de bijectivité pour une fonction de répartition* F. À gauche, *défaut d'injectivité*. À droite, *défaut de surjectivité*.

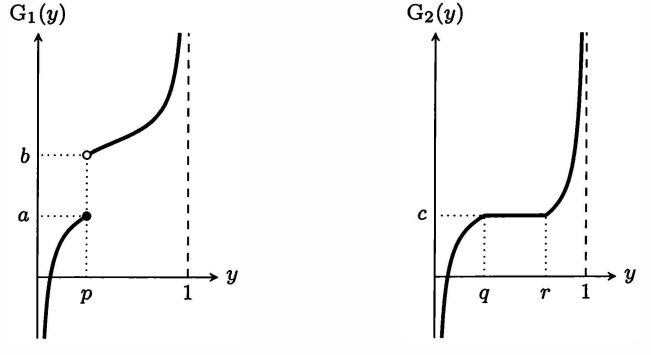


FIGURE A.6 — *Pseudo-inverses* G_1 *et* G_2 *des deux fonctions précédentes.*

Sur la figure A.6, on représente les pseudo-inverses G_1 et G_2 des fonctions F_1 et F_2 de la figure A.5. Le défaut d'injectivité de F_1 fait que, pour tout $x \in [a; b]$, on a $G_1(F_1(x)) = a$. Le défaut de surjectivité de F_2 fait que, pour tout $y \in [q; r]$, on a $F_2(G_2(y)) = r$.

■ **Proposition A.24** *Soit* $F : \overline{\mathbb{R}} \to [0; 1]$ *une fonction de répartition. On note* G *son pseudo-inverse, défini par la relation* (A.3).

i) Pour tout $y \in]0; 1[$ *et tout* $x \in \mathbb{R}$, *on a l'équivalence*

$$y \leqslant F(x) \iff G(y) \leqslant x.$$

ii) Si F *est strictement croissante, alors* $G \circ F = Id_{\overline{\mathbb{R}}}$.

iii) Si F *est continue, alors* $F \circ G = Id_{[0;1]}$.

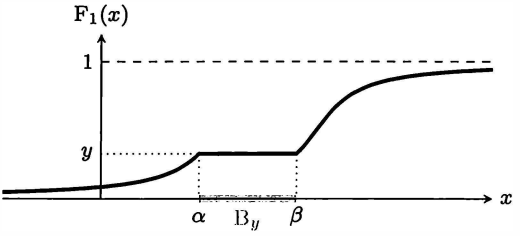
DÉMONSTRATION : *i)* Le sens « $\Rightarrow$ » découle immédiatement de la définition.
Montrons le sens « $\Leftarrow$ ». Soient x et y tels que $G(y) \leqslant x$. Posons

$$A_y := \{x \in \mathbb{R} \,;\, F(x) \geqslant y\},$$

c'est un intervalle de la forme $A_y =]G(y); +\infty[$ ou $A_y = [G(y); +\infty[$ a priori. Si $G(y) < x$, alors x est élément de A_y, donc $F(x) \geqslant y$ et c'est fini. Le seul point délicat est lorsque $G(y) = x$. On choisit une suite $(x_n)_{n\geqslant 0}$ strictement décroissante de limite x ; notamment, $x_n \in A_y$ pour tout n, donc $F(x_n) \geqslant y$ et, par continuité à droite de F, on conclut que $F(x) \geqslant y$.
On notera au passage que ce raisonnement prouve que A_y est fermé et de la forme $A_y = [G(y); +\infty[$.

ii) Pour tout élément x de $[-\infty; +\infty]$, notons $A_x := \{t \in \mathbb{R} \,;\, F(t) \geqslant F(x)\}$, de telle sorte que $G(F(x)) = \inf A_x$. La stricte croissance de F montre que $A_x = [x; +\infty[$, et donc $G(F(x)) = x$.

iii) La fonction F établit une surjection de $[-\infty; +\infty]$ sur $[0; 1]$.
Soit $y \in [0; 1]$. Notons $B_y := \{x \in [-\infty; +\infty] \,;\, F(x) = y\} = F^{-1}(\{y\})$. Cet ensemble est non vide par surjectivité, et par continuité de F c'est un intervalle fermé $[\alpha; \beta]$. Par croissance de F, si $F(x) > y$, alors $x > \beta$, donc $G(y) = \inf B_y = \alpha$; on conclut alors $F(G(y)) = F(\alpha) = y$.



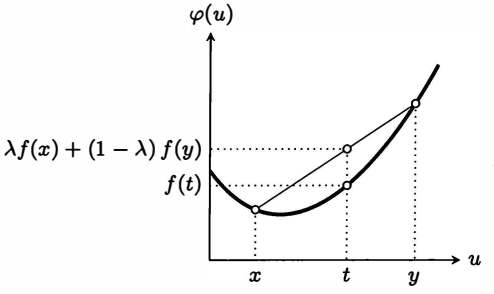
■

A.5 FONCTIONS CONVEXES

A30 **Fonctions convexes, strictement convexe, concaves**
Soit I un intervalle de $\mathbb{R}$. Si φ est une fonction définie sur I et à valeurs dans $\mathbb{R}$, on dit que φ est une fonction **convexe** si

$$\forall (x, y) \in I^2 \quad \forall \lambda \in]0; 1[\qquad \varphi(\lambda x + (1-\lambda)y) \leqslant \lambda\varphi(x) + (1-\lambda)\varphi(y).$$

On dit que φ est **strictement convexe** si l'inégalité ci-dessus est stricte dès que $x \neq y$. On dit que φ est **concave** (resp. **strictement concave**) si $-\varphi$ est une fonction convexe (resp. strictement convexe).

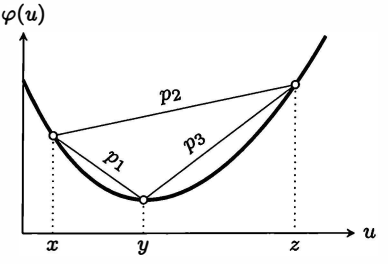


Caractérisation par la pente A31

$\varphi : \mathrm{I} \to \mathbb{R}$ est convexe si, et seulement si, pour tout $x \in \mathrm{I}$, la fonction $p_x : \mathrm{I} \setminus \{x\} \to \mathbb{R}$ qui à t associe la pente de la corde entre les points d'abscisses x et t :

$$p_x : t \longmapsto \frac{\varphi(t) - \varphi(x)}{t - x}$$

est croissante. ✦ On en déduit, pour toute fonction convexe, la double inégalité des pentes illustrée ci-dessous : si $x < y < z$, alors on a les inégalités des pentes $p_1 \leqslant p_2 \leqslant p_3$.

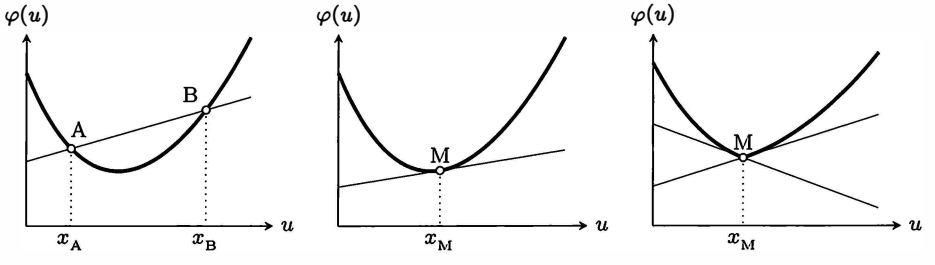


Caractérisation par la dérivée A32

Soit $\varphi : \mathrm{I} \to \mathbb{R}$ une fonction dérivable. Alors φ est convexe si et seulement si φ' est croissante ; φ est strictement convexe si et seulement si φ' est strictement croissante. ✦ Si φ est deux fois dérivable, alors φ est convexe si et seulement si $\varphi'' \geqslant 0$.

Cordes, tangentes et demi-tangentes A33

Soit φ une fonction convexe sur I. Soient A et B deux points du graphe de φ. Alors, le graphe est au-dessous de la corde (AB) entre les abscisses x_{A} et x_{B}, et au-dessus de la corde (AB) en dehors de l'intervalle $[\,x_{\mathrm{A}}\,;\,x_{\mathrm{B}}\,]$. ✦ Soit M un point du graphe d'une fonction φ convexe sur I ; on suppose de plus que φ est dérivable en x_{M}, ce qui signifie que son graphe admet une tangente en M. Alors le graphe de φ est partout au-dessus de la tangente en M.



✦ Si I est ouvert et si $\varphi : I \to \mathbb{R}$ est convexe, alors elle est continue sur I. ✦ Si, de plus, a est un point intérieur à I, alors φ admet en a une dérivée à gauche $\varphi'_g(a)$ et une dérivée à droite $\varphi'_d(a)$, vérifiant $\varphi'_g(a) \leqslant \varphi'_d(a)$. Le graphe de φ est au-dessus de chacune des tangentes à gauche et à droite en M. ✦ Les points en lesquels cette inégalité est stricte sont au plus dénombrables (il suffit de considérer les sauts $\sigma(x) = f'_d(x) - f'_g(x)$ et d'appliquer une technique similaire à celle de la démonstration de la proposition A.20). Ainsi, φ est dérivable en tout point de I, mis à part un ensemble dénombrable de points.

A34 **Inégalités de Jensen**

Inégalité discrète de Jensen — Soit φ une fonction convexe définie sur un intervalle I. Pour tout $n \in \mathbb{N}$ tel que $n \geqslant 2$, pour tout n-uplet $(x_1, \dots, x_n) \in I^n$ de réels, et pour tout n-uplet $(\lambda_1, \dots, \lambda_n) \in [0\,;\,1]^n$ tel que $\lambda_1 + \cdots + \lambda_n = 1$,

$$\varphi\left(\sum_{i=1}^{n} \lambda_i\, x_i\right) \leqslant \sum_{i=1}^{n} \lambda_i\, \varphi(x_i).$$

Inégalité intégrale de Jensen — Soit φ une fonction convexe. Soit f une fonction positive, intégrable et telle que $\int_{\mathbb{R}} f = 1$. Soit g une fonction continue et à valeurs réelles. Alors

$$\varphi\left(\int_{-\infty}^{+\infty} g(x)\, f(x)\, \mathrm{d}x\right) \leqslant \int_{-\infty}^{+\infty} \varphi\big(g(x)\big)\, f(x)\, \mathrm{d}x.$$

A.6 MODES DE CONVERGENCE DE SUITES ET SÉRIES DE FONCTIONS

A35 **Convergence simple et convergence uniforme**

Soit A un ensemble. Si f est une application de A valeurs dans $\mathbb{R}$[2], on note

$$\|f\|_\infty = \sup_{x \in A} \big|f(x)\big|.$$

Soit $(f_n)_{n \geqslant 1}$ une suite d'applications $A \to \mathbb{R}$ et soit $f : A \to \mathbb{R}$. La suite $(f_n)_{n \geqslant 1}$ **converge simplement vers f sur A** si, pour tout $x \in A$, la suite $\big(f_n(x)\big)_{n \geqslant 1}$ converge vers $f(x)$; la fonction f est appelée **limite simple** de la suite $(f_n)_{n \geqslant 1}$. ✦ La suite $(f_n)_{n \geqslant 1}$ **converge uniformément vers f** si $f_n - f$ est bornée à partir d'un certain rang et $\lim\limits_{n \to \infty} \|f_n - f\|_\infty = 0$. Cela revient à écrire

$$\forall \varepsilon > 0 \quad \exists N \in \mathbb{N} \quad \forall n \geqslant N \quad \forall x \in A \qquad \big|f_n(x) - f(x)\big| \leqslant \varepsilon.$$

La convergence uniforme implique la convergence simple, mais la réciproque est fausse. ✦ L'intérêt principal de la convergence uniforme est qu'elle transporte la continuité à la limite :

■ **Théorème A.25** *Si les fonctions $f_n : I \to \mathbb{R}$ sont continues et si $(f_n)_{n \geqslant 1}$ converge uniformément vers $f : I \to \mathbb{R}$, alors f est continue.*

Le cas des fonctions croissantes — et notamment des fonctions de répartition — mérite qu'on s'y arrête.

■ **Théorème A.26 (Dini)** *Soit K un segment de $\mathbb{R}$. Soit $(F_n)_{n \geqslant 1}$ une suite de fonctions croissantes de K dans $\mathbb{R}$, convergeant simplement vers une fonction F continue. Alors la convergence est uniforme.*

Plus généralement :

(2). Ou, plus généralement, à valeurs dans un espace vectoriel normé E.

■ **Théorème A.27 (Convergence uniforme de fonctions de répartition)** *Si* $(F_n)_{n\geqslant 1}$ *est une suite de fonctions croissantes de* $\overline{\mathbb{R}}$ *dans* $\mathbb{R}$ *qui converge simplement vers une fonction* $F : \overline{\mathbb{R}} \to \mathbb{R}$ *continue et bornée, alors la convergence est uniforme.*

Notamment, si les F_n *ainsi que* F *sont des fonctions de répartition, et si* F *est continue, alors la convergence simple de* $(F_n)_{n\geqslant 1}$ *vers* F *entraîne la convergence uniforme.*

DÉMONSTRATION : Remarquons d'abord que, par passage à la limite, la fonction F est croissante. L'image de $\overline{\mathbb{R}}$ par F est un segment $[a\,;\,b]$.

Soit $\varepsilon > 0$. On choisit un entier q tel que $1/q < \varepsilon/(b-a)$. Par continuité F, les points $\beta_k := a + (b-a)k/q$ $(k = 0, \dots, q)$ sont atteints : il existe $-\infty \leqslant a_0 < a_1 < a_2 < \cdots < a_{q-1} < a_q \leqslant +\infty$ tels que $F(a_k) = \beta_k$ pour tout $k \in [\![0\,;\,q]\!]$. La convergence simple en chacun de ces points permet alors de trouver un rang N tel que, pour tout $n \geqslant N$,

$$\forall k \in [\![0\,;\,q]\!] \qquad \left|F_n(a_k) - F(a_k)\right| \leqslant \varepsilon.$$

Par croissance des fonctions, sur chaque intervalle $[a_k\,;\,a_{k+1}]$ on a

$$F(a_{k+1}) - \varepsilon \leqslant F(a_k) \leqslant F(x) \leqslant F(a_{k+1}) \leqslant F(a_k) + \varepsilon$$

et

$$F(a_k) - \varepsilon \leqslant F_n(a_k) \leqslant F_n(x) \leqslant F_n(a_{k+1}) \leqslant F(a_{k+1}) + \varepsilon$$

ce qui fait que $F(x)$ et $F_n(x)$ diffèrent d'au plus 2ε. Ceci étant vrai pour chaque intervalle $[a_k\,;\,a_{k+1}]$, on a donc $\|F_n - F\|_\infty \leqslant 2\varepsilon$ pour tout $n \geqslant N$. ∎

Théorèmes de Weierstrass A36

Ce sont deux théorèmes de densité d'espaces de fonctions dans l'espace de fonctions continues.

■ **Théorème A.28 (1er th. de Weierstrass)** *Soit* $f : [a\,;\,b] \to \mathbb{K}$ *une application continue ; alors il existe une suite* $(P_n)_n$ *de polynômes qui converge uniformément vers* f *sur* $[a\,;\,b]$.

■ **Théorème A.29 (2e th. de Weierstrass)** *Soit* $T > 0$ *un réels. Soit* f *une application continue et* T*-périodique sur* $\mathbb{R}$. *Alors il existe une suite* $(P_n)_{n\geqslant 0}$ *de polynômes trigonométriques, de période* T *qui converge uniformément vers* f *sur* $\mathbb{R}$.

Séries de fonctions A37

Soient $(f_n)_{n\geqslant 1}$ une suite de fonctions définies sur un même intervalle I et à valeurs réelles. On dit que la série $\sum f_n$ **converge simplement** si la suite des sommes partielles $S_n := f_1 + f_2 + \cdots + f_n$ converge simplement ; on note alors $S := \sum_{n=1}^{\infty} f_n$ sa limite. On dit qu'elle **converge uniformément** si cette même suite converge uniformément vers S ; il est équivalent de dire que la suite des restes $R_n := \sum_{k=n+1}^{\infty} f_k$ converge uniformément vers 0.

✦ Si toutes les fonctions f_n sont continues et si la série $\sum f_n$ converge uniformément, alors sa somme S est encore continue.

On dit que la série $\sum f_n$ **converge normalement** si la série $\sum \|f_n\|_\infty$ converge ; dans ce cas, $\sum f_n$ converge uniformément.

A.7 SÉRIES ENTIÈRES

Rayon A38

Soit $\sum a_n z^n$ une série entière. On appelle **rayon de** $\sum a_n z^n$ l'élément de $\mathbb{R}^+ \cup \{+\infty\}$:

$$R := \sup\left\{r \geqslant 0\,;\ (a_n r^n)_n \text{ est bornée}\right\}.$$

On remarquera que $I := \left\{r \geqslant 0\,;\ (a_n r^n)_n \text{ est bornée}\right\}$ est un intervalle, de la forme $[0\,;\,R]$ ou $[0\,;\,R[$. Alors, pour tout $z \in \mathbb{C}$:

$$\begin{cases} |z| < R & \Longrightarrow \ (\sum a_n z^n \quad \text{converge absolument}) \\ |z| > R & \Longrightarrow \ ((a_n z^n)_{n\in\mathbb{N}} \quad \text{n'est pas bornée}). \end{cases}$$

A39 **Fonctions définies par une série entière**

On se limite ici aux fonctions définies sur la droite réelle. Soit f une fonction définie au voisinage de 0 comme la somme d'une série entière de rayon R :

$$\forall x \in]-\mathrm{R}\,;+\mathrm{R}[\qquad f(x) = \sum_{n=0}^{\infty} a_n\, x^n.$$

Alors f est de classe $\mathscr{C}^\infty$ sur $]-\mathrm{R}\,;+\mathrm{R}[$ et on peut récupérer ses coefficients en n'utilisant que le comportement local de f à l'origine :

$$a_n = \frac{f^{(n)}(0)}{n!}. \qquad \textit{récupération locale}$$

Si l'on étend la fonction f au disque ouvert de rayon R du plan complexe, les coefficients a_n sont également reliés aux valeurs de f sur tout cercle de rayon $r < \mathrm{R}$ par la **formule de Cauchy** :

$$a_n = \frac{1}{2\pi\, r^n} \int_0^{2\pi} f(r\,\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{e}^{-\mathrm{i}n\theta}\,\mathrm{d}\theta. \qquad \textit{récupération intégrale}$$

Par ailleurs, les dérivées successives de f s'obtiennent en dérivant terme à terme :

$$\forall k \in \mathbb{N} \quad \forall x \in]-\mathrm{R}\,;+\mathrm{R}[\qquad f^{(k)}(x) = \sum_{n=0}^{\infty} \frac{(n+k)!}{n!}\, a_{n+k}\, x^n = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!} a_n\, x^{n-k}.$$

✦ Le rayon des séries dérivées successives est toujours égal à R. ✦ La convergence de la série est normale, donc uniforme, sur tout segment inclus dans $]-\mathrm{R}\,;+\mathrm{R}[$.

A40 **Lemme radial d'Abel**

On suppose que f est somme d'une série entière de rayon $\mathrm{R} > 0$: $f(x) = \sum_{n=0}^{\infty} a_n\, x^n$.

■ **Lemme A.30 (Continuité radiale)** *Si* $\sum a_n\,\mathrm{R}^n$ *converge, alors* $\lim_{x\to\mathrm{R}^-} f(x) = f(\mathrm{R})$.

■ **Corollaire A.31** *Si* $\sum a_n$ *converge, alors* $\mathrm{R} \geqslant 1$ *et* $\lim_{x\to 1^-} \sum_{n=0}^{\infty} a_n\, x^n = \sum_{n=0}^{\infty} a_n$.

Il n'y a pas de réciproque : si f admet une limite en 1^-, cela n'implique pas que la série $\sum a_n$ converge[(3)]. Il existe une exception notable, si les a_n sont tous positifs :

■ **Proposition A.32 (Réciproque partielle au lemme d'Abel)** *Si* $a_n \geqslant 0$ *pour tout* $n \in \mathbb{N}$, *alors* f *est croissante sur* $[0\,;\mathrm{R}[$. *Elle admet donc une limite* ℓ *en* R^-, *finie ou infinie.*

- *Si* ℓ *est finie, alors la série* $\sum a_n\,\mathrm{R}^n$ *converge et a pour somme* ℓ.
- *Si* $\ell = +\infty$, *alors la série* $\sum a_n\,\mathrm{R}^n$ *diverge.*

A41 **Produit de Cauchy de deux séries entières**

Soient $\sum u_n\, z^n$ et $\sum v_n\, z^n$ deux séries entières de rayons respectifs R_u et R_v. La série $\sum w_n\, z^n$, où

$$\forall n \in \mathbb{N} \qquad w_n = \sum_{k=0}^{n} u_k\, v_{n-k}$$

est appelé **série produit de Cauchy** des séries $\sum u_n\, z^n$ et $\sum v_n\, z^n$; son rayon R_w est au moins égal à $\min(\mathrm{R}_u, \mathrm{R}_v)$ et

$$\forall\, |z| < \mathrm{R}_w \qquad \sum_{n=0}^{\infty} w_n\, z^n = \left(\sum_{n=0}^{\infty} u_n\, z^n\right)\left(\sum_{n=0}^{\infty} v_n\, z^n\right).$$

(3). Par exemple, posons $a_n = (-1)^n$ pour tout n, alors $f(x) = 1/(1+x)$ pour tout $x \in]-1\,;+1[$; la fonction f admet pour limite 1/2 en 1^-, alors que la série $\sum(-1)^n$ diverge grossièrement. Le **procédé de resommation d'Abel** consiste justement à attribuer à la série divergente $\sum(-1)^n$ la « somme » 1/2... mais ceci est une autre histoire !

Quelques sommes A42

La fonction $f : x \mapsto 1/(1-x)$ est développable en série entière autour de 0, et

$$\forall x \in]-1\,;\,+1[\qquad f(x) = \frac{1}{1-x} = \sum_{n=0}^{\infty} x^n.$$

Par dérivations terme à terme successives, on obtient notamment

$$f'(x) = \frac{1}{(1-x)^2} = \sum_{n=1}^{\infty} n\,x^{n-1} = \sum_{n=0}^{\infty} (n+1)\,x^n$$

et

$$f''(x) = \frac{2}{(1-x)^3} = \sum_{n=2}^{\infty} n(n-1)\,x^{n-2} = \sum_{n=0}^{\infty} (n+1)(n+2)\,x^n,$$

dont on déduit les formules suivantes :

$$\sum_{n=0}^{\infty} n\,x^n = \frac{x}{(1-x)^2} \qquad \text{et} \qquad \sum_{n=0}^{\infty} n^2\,x^n = \frac{x^2+x}{(1-x)^3}. \tag{A.4}$$

Par intégration terme à terme, on montre également que

$$\forall x \in [-1\,;\,+1[\qquad \ln(1-x) = -\sum_{n=1}^{\infty} \frac{x^n}{n}. \tag{A.5}$$

A.8 CALCUL INTÉGRAL

Théorème de Fubini A43

Soit $f : \mathbb{R}^2 \to \mathbb{R}$ une fonction mesurable (pour la tribu $\mathfrak{B}^2$ des boréliens de $\mathbb{R}^2$). Alors, pour tout réel x, la **section** $f(x,\cdot)$, c'est-à-dire l'application $y \mapsto f(x,y)$, est $\mathfrak{B}$-mesurable ; de même pour la section $f(\cdot,y)$.

On suppose que f est *positive*, ou *intégrable* sur $\mathbb{R}^2$; alors les fonctions

$$\varphi : x \longmapsto \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y \qquad \text{et} \qquad \psi : y \longmapsto \int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}x$$

sont définies presque partout, et

$$\iint_{\mathbb{R}^2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{+\infty} \varphi(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} \psi(y)\,\mathrm{d}y,$$

cette égalité étant prise dans $[0\,;+\infty]$ si f est positive, et dans $]-\infty\,;+\infty[$ si f est de signe quelconque mais intégrable.

Plus généralement, si $\mathbf{P}$ et $\mathbf{Q}$ sont des mesures positives sur $\mathbb{R}$ (par exemple des probabilités), si $f : \mathbb{R}^2 \to \mathbb{R}$ est mesurable et si elle est *positive*, ou *intégrable* sur $\mathbb{R}^2$ muni de la probabilité produit $\mathbf{P} \otimes \mathbf{Q}$, alors les fonctions

$$x \longmapsto \int_{-\infty}^{+\infty} f(x,y)\,\mathbf{Q}(\mathrm{d}y) \qquad \text{et} \qquad y \longmapsto \int_{-\infty}^{+\infty} f(x,y)\,\mathbf{P}(\mathrm{d}x)$$

sont mesurables et on peut intégrer dans le sens voulu :

$$\begin{aligned}\iint_{\mathbb{R}^2} f(x,y)\,(\mathbf{P}\otimes\mathbf{Q})(\mathrm{d}x,\mathrm{d}y) &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x,y)\,\mathbf{Q}(\mathrm{d}y) \right) \mathbf{P}(\mathrm{d}x) \\ &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x,y)\,\mathbf{P}(\mathrm{d}x) \right) \mathbf{Q}(\mathrm{d}y).\end{aligned}$$

A44 **Changement de variable dans $\mathbb{R}^n$**

On considère un changement de variable $\boldsymbol{y} = \varphi(\boldsymbol{x})$, où $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ est une application continûment différentiable (classe $\mathscr{C}^1$), dont on note $\varphi = (\varphi_1, \varphi_2, \ldots, \varphi_n)$ les composantes. On définit la **matrice jacobienne** $\mathscr{J}_\varphi(\boldsymbol{x})$ de φ en $\boldsymbol{x}$, qui est la matrice de coefficients

$$\big[\mathscr{J}_\varphi(\boldsymbol{x})\big]_{i,j} = \frac{\partial \varphi_i}{\partial x_j}(\boldsymbol{x}).$$

Le **jacobien** de φ en $\boldsymbol{x}$ est le déterminant de cette matrice. On le note symboliquement

$$\mathrm{J}_\varphi(\boldsymbol{x}) = \text{dét}\ \mathscr{J}_\varphi(\boldsymbol{x}) =: \frac{\mathrm{D}\boldsymbol{y}}{\mathrm{D}\boldsymbol{x}}.$$

■ **Théorème A.33 (Changement de variable dans $\mathbb{R}^n$)**
Soit $\varphi : \mathrm{D} \to \Delta$ un $\mathscr{C}^1$-difféomorphisme d'un ouvert D de $\mathbb{R}^n$ sur un ouvert $\Delta = \varphi(\mathrm{D})$. Soit $f : \Delta \to \mathbb{R}$ une fonction intégrable sur Δ. Alors

$$\int_\Delta f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \int_\mathrm{D} f\big(\varphi(\boldsymbol{x})\big)\,\big|\mathrm{J}_\varphi(\boldsymbol{x})\big|\,\mathrm{d}\boldsymbol{x}.$$

Si l'on note $\tilde{f} = f \circ \varphi$, de telle sorte que $f(\boldsymbol{y}) = \tilde{f}(\boldsymbol{x})$ lorsque $\boldsymbol{y} = \varphi(\boldsymbol{x})$ est l'image de $\boldsymbol{x}$, la formule de changement de variable s'écrit sous la forme agréable

$$\int_\Delta f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \int_\mathrm{D} \tilde{f}(\boldsymbol{x})\,\left|\frac{\mathrm{D}\boldsymbol{y}}{\mathrm{D}\boldsymbol{x}}\right|\,\mathrm{d}\boldsymbol{x}.$$

Un cas particulier important est celui de la transformation entre coordonnées polaires (r, θ) et coordonnées cartésiennes $(x, y) = \varphi(r, \theta) = (r\cos\theta, r\sin\theta)$; le jacobien de φ est $\mathrm{J}_\varphi(r, \theta) = r$ et ne s'annule qu'en $r = 0$. Si f est une fonction intégrable sur $\mathbb{R}^2$, un simple procédé de passage à la limite donne

$$\iint_{\mathbb{R}^2} f(x, y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^{2\pi}\int_0^{+\infty} f(r\cos\theta, r\sin\theta)\,r\,\mathrm{d}r\,\mathrm{d}\theta.$$

Fréquemment dans la mise en œuvre pratique du théorème de changement de variable, on détourne pudiquement le regard des problèmes de bord.

A45 **Produit de convolution**

Soient f et g deux fonctions de la variable réelle. En tout réel x pour lequel cela a un sens, on définit

$$f * g(x) = \int_{-\infty}^{+\infty} f(t)\,g(x - t)\,\mathrm{d}t.$$

La fonction $f * g$ est le **produit de convolution** de f et g. ✦ Le produit de convolution est commutatif : $f * g = g * f$ en tout point où il est défini. ✦ Si f et g sont mesurables, si f est intégrable et si g est bornée, alors $f * g$ est définie sur $\mathbb{R}$. ✦ Si f et g sont toutes deux à support borné à gauche[(4)], alors $f * g$ est bien définie sur $\mathbb{R}$. ✦ Si f et g sont intégrables, alors $f * g$ est défini presque partout et intégrable, et de plus

$$\int_{-\infty}^{+\infty} f * g = \left(\int_{-\infty}^{+\infty} f\right)\cdot\left(\int_{-\infty}^{+\infty} g\right).$$

Remarque A.34 Il y a une forte analogie entre le produit de convolution et le produit de Cauchy de deux séries, qui est un cas particulier de produit de convolution pour une mesure discrète. Cela explique la ressemblance formelle entre les théorèmes 8.9 page 182 et 12.15 page 265.

(4). Par exemple si ce sont des densités des variables aléatoires à valeurs positives.

Quelques intégrales utiles A46

On donne sans démonstration la valeur de quelques intégrales utiles ; les deux premières s'obtiennent par des moyens élémentaires (intégration par parties, changement de variables), pour $\alpha > 0$ et $n \in \mathbb{N}$, la dernière par intégrations par parties successives.

$$\int_0^{+\infty} \mathrm{e}^{-\alpha t}\,\mathrm{d}t = \frac{1}{\alpha} \qquad \int_0^{+\infty} t^n\,\mathrm{e}^{-\alpha t}\,\mathrm{d}t = \frac{n!}{\alpha^{n+1}}$$

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \mathrm{e}^{-t^2/2}\,\mathrm{d}t = 1 \qquad \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} t^{2n}\,\mathrm{e}^{-t^2/2}\,\mathrm{d}t = \frac{(2n)!}{2^n\,n!}.$$

Méthode des multiplicateurs de Lagrange A47

Considérons une partie $\mathscr{S}$ de $\mathbb{R}^n$, définie par un système de d équations

$$\mathscr{S} = \big\{x \in \mathbb{R}^n \;;\; f_1(x) = f_2(x) = \cdots = f_d(x) = 0\big\}.$$

Pour que $\mathscr{S}$ soit bien une « surface » de dimension $n-d$, on suppose que les fonctions f_k sont toutes différentiables et que leurs différentielles en tous points de $\mathscr{S}$ sont linéairement indépendantes. Soit $g : \mathbb{R}^n \to \mathbb{R}$ une fonction numérique différentiable. Si la restriction de g à $\mathscr{S}$ admet un extremum en un point $\boldsymbol{a} \in \mathscr{S}$, alors la différentielle de g en $\boldsymbol{a}$ est nulle sur l'espace tangent à $\mathscr{S}$ en a. Cela revient à dire que $\mathrm{d}g(\boldsymbol{a})$ est combinaison linéaire des $\mathrm{d}f_k(\boldsymbol{a})$, c'est-à-dire qu'il existe des scalaires $\lambda_1, \lambda_2, \ldots, \lambda_d$ tels que

$$\mathrm{d}g(\boldsymbol{a}) = \sum_{k=1}^{d} \lambda_k\,\mathrm{d}f_k(\boldsymbol{a}).$$

Les scalaires $\lambda_1, \lambda_2, \ldots, \lambda_d$ sont appelés **multiplicateurs de Lagrange**.

Remarque A.35 Certaines quantités physiques apparaissent naturellement comme des multiplicateurs de Lagrange. C'est le cas de la température en thermodynamique (ou, plus précisément, de la quantité $\beta = 1/k_{\mathrm{B}}\mathrm{T}$, où k_{B} est la constante de Boltzmann, voir § 415).

A.9 ESPACES DE HILBERT

Espaces préhilbertiens A48

Un **espace préhilbertien** est un espace vectoriel E muni d'un produit scalaire $\langle\cdot|\cdot\rangle$. Il est alors naturellement muni de la norme associée à ce produit scalaire : $\|x\|_2 := \langle x|x\rangle^{1/2}$.

✦ On dit que $x \perp y$ si $\langle x|y\rangle = 0$. Si F est un sous-espace vectoriel de E, on note

$$\mathrm{F}^{\perp} = \big\{x \in \mathrm{E} \;;\; \forall y \in \mathrm{F} \quad \langle x|y\rangle = 0\big\}.$$

■ **Théorème A.36 (Inégalité de Cauchy-Schwarz)** *Pour tous $x, y \in \mathrm{E}$, on a*

$$\big|\langle x|y\rangle\big| \leqslant \|x\|_2\,\|y\|_2\,.$$

■ **Théorème A.37 (Théorème de projection orthogonale)** *Soit* E *un espace préhilbertien de dimension quelconque. Soit* V *un sous-espace vectoriel de dimension finie de* E. *Soit $x \in \mathrm{E}$. Il existe un unique élément x_{V} de* V *tel que $x - x_{\mathrm{V}} \perp \mathrm{V}$, et il est donné par la* **projection orthogonale** *de x sur* V, *définie par*

$$x_{\mathrm{V}} := \sum_{i=1}^{p} \langle e_i|x\rangle\; e_i.$$

où $(e_1, e_2, \ldots, e_p)$ est une base orthonormée quelconque de V. *De plus, x_{V} est l'unique vecteur de* V *tel que $d(x\,;\,\mathrm{V}) = \|x - x_{\mathrm{V}}\|$.*

A49 **Espaces de Hilbert, bases hilbertiennes**

On appelle **espace de Hilbert** un espace vectoriel H réel ou complexe, muni d'un produit scalaire, et *complet* pour la norme associée à ce produit scalaire : toute suite de Cauchy de H est convergente, toute série absolument convergente est convergente.

Exemple A.38 L'espace $\ell^2(\mathbb{N})$ des suites $u = (u_n)_{n\geqslant 0}$ telles que $\sum |u_n|^2$ converge est un espace de Hilbert pour le produit scalaire $\langle u|v\rangle := \sum_{n=0}^{\infty} \overline{u_n}\, v_n$. ✦ L'espace L^2 des fonctions de carré intégrables, définies à une égalité presque partout près, est un espace de Hilbert pour le produit scalaire $\langle f|g\rangle := \int_{-\infty}^{+\infty} \bar{f} g$.

Soit $(e_i)_{i\in I}$ une famille de vecteurs de H indexée par un ensemble I, dénombrable ou non. La famille $(e_i)_{i\in I}$ est appelée **système total** si $\mathrm{Vect}\{e_i\ ;\ i \in I\}$ est dense.

Soit H un espace de Hilbert. On dit que la famille $(e_i)_{i\in I}$ est une **base hilbertienne** si c'est une famille orthonormée totale[5] dans H.

Exemple A.39 L'espace $\ell^2(\mathbb{N})$ admet pour base hilbertienne la famille dénombrable $(e_n)_{n\geqslant 0}$ où l'on a défini $e_n := (0, \dots, 0, 1, 0, \dots)$, le nombre de zéros avant l'unique 1 valant n. ✦ L'espace $L^2([0\,;1], \mathbb{R})$ admet pour base hilbertienne la famille $(h_n)_{n\geqslant 0}$ des fonctions de Haar définies page 628. ✦ Une base hilbertienne bien connue de $L^2([0\,;1], \mathbb{C})$ est la famille $(e_n)_{n\in\mathbb{Z}}$ où $e_n : t \mapsto \mathrm{e}^{2\mathrm{i}\pi n x}$; cette famille permet de fonder la théorie des séries de Fourier. ✦ Si l'on définit les **polynômes d'Hermite** par la formule

$$\mathrm{H}_n(t) := (-1)^n \mathrm{e}^{t^2} \frac{\mathrm{d}^n}{\mathrm{d}t^n} \mathrm{e}^{-t^2},$$

et les fonctions $\mathscr{H}_n : t \mapsto \mathrm{H}_n(t)\, \mathrm{e}^{-t^2/2}$ pour tout $n \geqslant 0$, la famille $(\mathscr{H}_n)_{n\geqslant 0}$ est une base hilbertienne de $L^2(\mathbb{R})$, dont tous les éléments sont de plus des vecteurs propres pour la transformation de Fourier.

■ **Théorème A.40 (Caractérisations d'une base hilbertienne)** *Soit* H *un espace de Hilbert et soit* $(e_n)_{n\geqslant 0}$ *un système orthonormé. Alors on a équivalence entre les assertions :*

i) $(e_n)_{n\geqslant 0}$ *est une base hilbertienne ;*

ii) pour tout $x \in \mathrm{H}$, *si l'on pose* $x_p := \sum_{n=0}^{p} \langle e_n|x\rangle\, e_n$, *alors* $\lim_{p\to\infty} \|x - x_p\| = 0$;

iii) pour tous $x, y \in \mathrm{H}$, *on a* $\langle x|y\rangle = \sum_{n=0}^{\infty} \langle x|e_n\rangle\, \langle e_n|y\rangle$;

iv) pour tout $x \in \mathrm{H}$, *on a l'***égalité de Parseval** $\|x\|^2 = \sum_{n=0}^{\infty} \big|\, \langle e_n|x\rangle \,\big|^2$;

v) $\mathrm{Vect}\{e_n\ ;\ n \in \mathbb{N}\}^{\perp} = \{0\}$ *c'est-à-dire que, pour tout* $x \in \mathrm{H}$, *on a*

$$\big(\langle e_n|x\rangle = 0 \text{ pour tout } n \in \mathbb{N}\big) \iff \big(x = 0\big).$$

On en déduit les règles de calcul suivantes, qui généralisent le cas euclidien.

■ **Théorème A.41 (Calcul dans des bases hilbertiennes)** *Soit* H *un espace de Hilbert, muni d'une base hilbertienne* $(e_n)_{n\geqslant 0}$. *Alors, pour tous vecteur* x, y, *on a*

$$x = \sum_{n=0}^{\infty} \langle e_n|x\rangle\, e_n \qquad \langle x|y\rangle = \sum_{n=0}^{\infty} \langle x|e_n\rangle\, \langle e_n|y\rangle \qquad \|x\|^2 = \sum_{n=0}^{\infty} \big|\, \langle e_n|x\rangle \,\big|^2.$$

Remarque A.42 Attention, la première égalité a lieu dans l'espace H, au sens de la norme induite par le produit scalaire : c'est une réécriture de la propriété *ii)* du théorème précédent. Notamment, deux fonctions égales dans l'espace L^2 ne sont pas « égales en tous points », mais seulement presque partout.

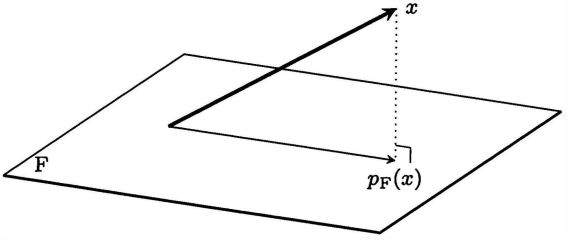
---

[5]. En d'autres termes, si $(e_i)_{i\in I}$ est une base hilbertienne, alors tout vecteur de H peut être approché, avec une précision arbitrairement choisie, par une combinaison linéaire (finie) des vecteurs e_i. Cette propriété généralise donc la notion d'« être génératrice ».

Théorème de projection orthogonale A50

Un des principaux intérêts de la structure hilbertienne est la possibilité de projeter un élément sur un ensemble fermé et convexe, et notamment sur un sous-espace vectoriel fermé de E.

■ **Théorème A.43** *Soit* F *un sous-espace vectoriel fermé d'un espace de Hilbert* H. *Alors, pour tout* $x \in \mathrm{H}$, *il existe un unique élément* $p_{\mathrm{F}}(x)$ *de* F *tel que* $x - p_{\mathrm{F}}(x)$ *soit orthogonal à* F. *L'application linéaire* p_{F} *est continue, et c'est le projecteur orthogonal sur* F *parallèlement à* $\mathrm{F}^{\perp}$. *On a donc notamment* $\mathrm{H} = \mathrm{F} \oplus \mathrm{F}^{\perp}$.



A.10 TRANSFORMÉE DE FOURIER

Transformée de Fourier d'une fonction intégrable A51

Soit $f : \mathbb{R} \to \mathbb{C}$ une fonction intégrable. Sa **transformée de Fourier** est l'application $\tilde{f} : \mathbb{R} \to \mathbb{C}$ définie par

$$\tilde{f}(x) = \int_{-\infty}^{+\infty} f(t)\, \mathrm{e}^{-\mathrm{i}xt}\, \mathrm{d}t. \qquad (\text{A.6})$$

(Selon la convention choisie, le facteur exponentiel peut être pris égal à $\mathrm{e}^{-2\mathrm{i}\pi xt}$.) La transformée de Fourier d'une fonction intégrable n'est pas nécessairement intégrable elle-même. Toutefois, lorsque c'est le cas, on a la **formule d'inversion**

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{f}(x)\, \mathrm{e}^{\mathrm{i}xt}\, \mathrm{d}x \qquad (\text{A.7})$$

valable pour presque tout t (et valable notamment en tous les points de continuité de f).

Il est ennuyeux que l'espace $\mathrm{L}^1(\mathbb{R})$ ne soit pas stable par transformée de Fourier[6]. Il est d'autant plus remarquable que l'espace $\mathrm{L}^2(\mathbb{R})$, lui, soit stable : la transformée de Fourier d'une fonction de carré intégrable est encore une fonction de carré intégrable[7]. La formule d'inversion (A.7) est alors valable.

Distributions. Distributions tempérées A52

Une **distribution** est une forme linéaire continue sur un certain espace de fonctions, appelées **fonctions-test**. Un choix possible est l'espace $\mathscr{D}$ des fonctions de classe $\mathscr{C}^{\infty}$ à support borné ; l'espace des distributions correspondant est noté $\mathscr{D}'$.

Un second choix est l'espace $\mathscr{S}$ de Schwartz des fonctions φ de classe $\mathscr{C}^{\infty}$ à décroissance rapide — c'est-à-dire que φ et toutes ses dérivées successives $\varphi^{(k)}$ décroissent, en $\pm\infty$, plus vite que toute fonction puissance. On dira qu'une suite $(\varphi_n)_{n\geqslant 0}$ **converge vers 0 dans** $\mathscr{S}$ si la suite de fonctions de terme général $x \mapsto x^k\, \varphi_n^{(j)}(x)$ converge uniformément vers 0 sur $\mathbb{R}$, et ce pour toutes valeurs de k et de ℓ. Cela permet de définir la

(6). Exercice : calculer la transformée de Fourier de la fonction f indicatrice de $[-1\,;+1]$, et vérifier qu'elle n'est pas intégrable.

(7). Si f n'est pas intégrable mais seulement de carré intégrable, on ne peut pas définir sa transformée de Fourier par la formule (A.6) ; on la définit *via* un procédé de passage à la limite, en approchant f par une suite de fonctions à décroissance rapide et en faisant usage de la complétude de l'espace $\mathrm{L}^2(\mathbb{R})$ [36].

notion de continuité d'une forme linéaire sur $\mathscr{S}$: c'est une forme linéaire $T : \mathscr{S} \to \mathbb{C}$ telle que, pour toute suite $(\varphi_n)_{n\geqslant 0}$ tendant vers 0 dans $\mathscr{S}$, on ait $T(\varphi_n) \to 0$.

Les formes linéaires continues sur $\mathscr{S}$ (au sens précédent) sont appelées **distributions tempérées**. On note habituellement $\langle T, \varphi\rangle := T(\varphi)$.

La **distribution de Dirac** δ est la forme linéaire définie par

$$\forall \varphi \in \mathscr{S} \qquad \langle \delta, \varphi\rangle := \varphi(0).$$

Si a est un réel quelconque, on définit par translation la **distribution de Dirac en a** par

$$\forall \varphi \in \mathscr{S} \qquad \langle \delta_a, \varphi\rangle := \varphi(a).$$

Toute fonction f localement sommable et à croissance « modérée[8] » définit une distribution (également notée f) par la formule

$$\forall \varphi \in \mathscr{S} \qquad \langle f, \varphi\rangle := \int_{-\infty}^{+\infty} f(x)\,\varphi(x)\,\mathrm{d}x.$$

Le **produit tensoriel** $S \otimes T$ de deux distributions tempérées est défini comme la forme linéaire sur l'espace des fonctions-tests à deux variables par

$$\forall \varphi \in \mathscr{S}(\mathbb{R}^2) \qquad \langle S \otimes T, \varphi\rangle = \big\langle S(x)\,T(y), \varphi(x,y)\big\rangle := \Big\langle S(x), \big\langle T(y), \varphi(x,y)\big\rangle\Big\rangle$$

(on fait agir T sur $y \mapsto \varphi(x,y)$ à x fixé ; le résultat est une fonction de x sur lequel on fait agir S). On le note parfois (abusivement) $S(x)\,T(y)$. Le **produit de convolution** $S * T$ de deux distributions est

$$\forall \varphi \in \mathscr{S} \qquad \langle S * T, \varphi\rangle := \big\langle S(x)\,T(y), \varphi(x+y)\big\rangle.$$

Le produit de convolution de δ_a par une fonction localement sommable g est la translatée $x \mapsto g(x-a)$ de cette fonction ; ce résultat se retrouve par la formule *purement mnémotechnique*

$$\delta_a * f(x) = \int_{-\infty}^{+\infty} \delta_a(t)\,f(x-t)\,\mathrm{d}t = f(x-a).$$

A53 **Transformée de Fourier d'une distribution**
Si T est une distribution tempérée, on définit sa transformée de Fourier par

$$\forall \varphi \in \mathscr{S} \qquad \big\langle \widetilde{T}, \varphi\big\rangle := \big\langle T, \widetilde{\varphi}\big\rangle.$$

On peut montrer que toute distribution tempérée admet une transformée de Fourier. Notamment, la transformée de Fourier de δ est la fonction constante égale à 1, et la transformée de Fourier de 1 est $2\pi\,\delta$.

A.11 DÉNOMBREMENT

A54 **Coefficients binomiaux**
Soient $n \in \mathbb{N}$ et $k \in [\![0\,;\,n]\!]$; le **coefficient binomial** « k parmi n » est

$$C_n^k = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

✦ Le nombre $\binom{n}{k}$ est égal au nombre de parties à k éléments d'un ensemble à n éléments. C'est donc le nombre de choix possibles de k éléments distincts parmi n, étant entendu que l'on ne tient pas compte de l'ordre. ✦ Si $k > n$ ou $k < 0$, on pose par définition

(8). C'est-à-dire telle que, si φ est à décroissance rapide, $f\varphi$ est intégrable.

$\binom{n}{k} := 0$. ✦ On définit enfin des **coefficients binomiaux généralisés** : si x est un entier (positif ou négatif), voire un complexe quelconque, et si q est un entier naturel, on pose

$$\binom{x}{q} = \frac{1}{q!}\prod_{i=0}^{q-1}(x-i) = \frac{x(x-1)(x-2)\cdots(x-q+1)}{q!}.$$

Cette définition permet d'écrire la formule du binôme généralisée : pour tout complexe z tel que $|z| < 1$ et pour tout a,

$$(1+z)^a = 1 + \binom{a}{1}z + \binom{a}{2}z^2 + \binom{a}{3}z^3 + \cdots = \sum_{n=0}^{\infty}\binom{a}{n}z^n.$$

Si a est un entier naturel, la somme du membre de droite est en réalité une somme finie (les coefficients binomiaux généralisés s'annulant dès que $n > a$) ; sinon, on a une série convergente lorsque $t \in]-1\,;\,+1[$.

Formulaire pour les coefficients binomiaux A55

Si $n \geqslant 0$ et $k \in [\![0\,;\,n]\!]$, on a les formules et propriétés suivantes :

$$\binom{n}{k} = \binom{n}{n-k} \quad \text{(symétrie)} \qquad (x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k} \quad \text{(binôme)}$$

$$k\binom{n}{k} = n\binom{n-1}{k-1} \qquad \binom{n+m}{r} = \sum_{k=0}^{r}\binom{n}{k}\binom{m}{r-k} \quad \text{(Vandermonde)}$$

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1} \quad \text{(Pascal)} \qquad \frac{1}{(1-x)^{k+1}} = \sum_{n=k}^{\infty}\binom{n}{k}x^{n-k} \quad |x| < 1$$

Nombres de Stirling A56

Pour tout $n \in \mathbb{N}$, la famille $\mathcal{C}_n := (1, X, X^2, \ldots, X^n)$ est la base canonique du $\mathbb{R}$-espace vectoriel $\mathbb{R}_n[X]$ des polynômes à coefficients réels et de degré au plus égal à n. La famille $\mathcal{B}_n := (T_0, T_1, T_2, \ldots, T_n)$, où l'on a noté

$$T_0 = 1 \qquad \text{et} \qquad T_k = X(X-1)(X-2)\cdots(X-k+1),$$

est échelonnée en degré et de cardinal $n+1$, c'est donc encore une base de $\mathbb{R}_n[X]$. Les nombres de Stirling sont les coefficients permettant d'exprimer une des familles en fonction de l'autre. Plus précisément, dans l'unique décomposition

$$T_n = \sum_{k=0}^{n} S_n^k X^k,$$

les coefficients S_n^k sont appelés **nombres de Stirling de première espèce**. La matrice de passage de la base canonique $\mathcal{C}_n$ à la base $\mathcal{B}_n$ est

$$P = \begin{array}{c} \begin{matrix} T_0 & T_1 & T_2 & & T_n \end{matrix} \\ \begin{pmatrix} S_0^0 & S_1^0 & S_2^0 & & S_n^0 \\ & S_1^1 & S_2^1 & & S_n^1 \\ & & S_2^2 & & S_n^2 \\ & & & & \vdots \\ & & & & S_n^n \end{pmatrix} \end{array} \begin{matrix} X^0 \\ X^1 \\ X^2 \\ \\ X^n \end{matrix}$$

On peut de même exprimer les polynômes de la base canonique au moyen des T_k ; dans l'unique décomposition

$$X^n = \sum_{k=0}^{n} s_n^k T_k,$$

les coefficients s_n^k sont appelés **nombres de Stirling de seconde espèce**. La matrice de passage de la base $\mathcal{B}_n$ à la base canonique $\mathcal{C}_n$ est

$$\mathrm{P}^{-1} = \begin{array}{c} \begin{array}{cccc} X^0 & X^1 & X^2 & \cdots & X^n \end{array} \\ \begin{pmatrix} s_0^0 & s_1^0 & s_2^0 & \cdots & s_n^0 \\ & s_1^1 & s_2^1 & \cdots & s_n^1 \\ & & s_2^2 & \cdots & s_n^2 \\ & & & & \vdots \\ & & & & s_n^n \end{pmatrix} \end{array} \begin{array}{c} \\ T_0 \\ T_1 \\ T_2 \\ \vdots \\ T_n \end{array}$$

Nous donnons ici quelques propriétés des nombres de Stirling ; leurs démonstrations sont laissées en exercice.

■ **Proposition A.44 (Propriétés des nombres de Stirling)**
Les nombres s_n^k et S_n^k sont nuls si $k > n$. De plus, on a les propriétés suivantes :

i) ***Récurrence*** *: Pour tous $n \geqslant k \geqslant 1$, $\mathrm{S}_{n+1}^k = \mathrm{S}_n^{k-1} - n\mathrm{S}_n^k$ et $s_{n+1}^k = s_n^{k-1} + k s_n^k$*

ii) ***Initialisation*** *: Pour tout $n \geqslant 1$, $s_n^0 = \mathrm{S}_n^0 = 0$ $\quad s_0^0 = \mathrm{S}_0^0 = 1$.*

iii) ***Règle de somme*** *: Pour tout $n \geqslant 2$, $\sum_{k=0}^{n} \mathrm{S}_n^k = 0$.*

iv) ***P et P^{-1} sont inverses l'une de l'autre*** *: $\sum_{k=0}^{\infty} \mathrm{S}_p^k \, s_k^q = \delta_{p,q}$.*

v) ***Valeurs particulières*** *:*

$$\mathrm{S}_n^1 = (-1)^{n-1}(n-1)! \qquad \mathrm{S}_n^{n-1} = -\binom{n}{2} \qquad \mathrm{S}_n^n = 1$$

$$s_n^1 = 1 \qquad s_n^{n-1} = \binom{n}{2} \qquad s_n^n = 1.$$

vi) ***Valeur explicite des nombres de seconde espèce*** *: $s_n^k = \frac{1}{k!} \sum_{j=0}^{k} (-1)^{j+k} \binom{k}{j} j^n$.*

■ **Proposition A.45 (Autre interprétation combinatoire)** *s_n^k est le nombre de partitions d'un ensemble de n éléments en k classes non vides. Ainsi, $k!\, s_n^k$ est le nombre de surjections d'un ensemble à n éléments vers un ensemble à k éléments.*

Nombres de Stirling de 1re espèce S_n^k							
$n \setminus k$	0	1	2	3	4	5	6
0	1						
1	0	1					
2	0	−1	1				
3	0	2	−3	1			
4	0	−6	11	−6	1		
5	0	24	−50	35	−10	1	
6	0	−120	274	−225	85	−15	1

Nombres de Stirling de 2de espèce s_n^k							
$n \setminus k$	0	1	2	3	4	5	6
0	1						
1	0	1					
2	0	1	1				
3	0	1	3	1			
4	0	1	7	6	1		
5	0	1	15	25	10	1	
6	0	1	31	90	65	15	1

TABLE A.1 — *Valeurs des premiers nombres de Stirling.*

A.12 TOPOLOGIE ÉLÉMENTAIRE

Ouverts, fermés A57

Un **espace vectoriel normé** est un couple $(E, \|\cdot\|)$ où E est un espace vectoriel et $\|\cdot\|$ une norme, c'est-à-dire une application de E dans $\mathbb{R}^+$ vérifiant les propriétés :

- pour tout $x \in E$, $\|x\| = 0 \implies x = 0$; (*séparation*)
- pour tout $x \in E$ et tout $\lambda \in \mathbb{K}$, on a $\|\lambda x\| = |\lambda| \, \|x\|$; (*homogénéité*)
- pour tout $x, y \in E$, on a $\|x + y\| \leqslant \|x\| + \|y\|$ (*inégalité triangulaire*).

Soient $a \in E$ et $r \geqslant 0$; on appelle **boule ouverte de centre a et de rayon r** l'ensemble

$$\mathcal{B}(a\,;\,r) := \{x \in E\,;\ \|x - a\| < r\}.$$

Soient $x \in E$ et V une partie de E. On dit que V est un **voisinage de x dans E** si et seulement si il contient une boule ouverte non vide et de centre x, c'est-à-dire s'il existe $r > 0$ tel que $\mathcal{B}(x\,;\,r) \subset A$.

Une partie $A \subset E$ est dite **ouverte** dans E (on dit aussi que A est un **ouvert**) si A est un voisinage de chacun de ses points :

$$\forall a \in A \quad \exists r > 0 \qquad \mathcal{B}(a\,;\,r) \subset A.$$

Une partie $F \subset E$ est dite **fermée** dans E (on dit aussi que F est un **fermé**) si le complémentaire F^c de F dans E est ouvert. On peut caractériser séquentiellement les fermés : F est fermée si et seulement si pour toute suite $(u_n)_{n\geqslant 0}$ à valeurs dans F et convergente (dans E), on a $\lim\limits_{n\to\infty} a_n \in F$ (penser par exemple aux intervalles ouverts et fermés de $\mathbb{R}$).

Espaces complets A58

Une suite $(u_n)_{n\geqslant 1}$ d'un espace vectoriel normé $(E, \|\cdot\|)$ est une **suite de Cauchy** si :

$$\forall \varepsilon > 0 \quad \exists N \in \mathbb{N} \quad \forall p, q \geqslant N \qquad \|u_p - u_q\| \leqslant \varepsilon.$$

Toute suite convergente est de Cauchy. Toute suite de Cauchy est bornée. Toute suite de Cauchy admettant une sous-suite convergente est convergente.

L'espace $(E, \|\cdot\|)$ est dit **complet** s'il vérifie l'une des propriétés équivalentes suivantes :

i) toute suite de Cauchy est convergente ;

ii) toute série $\sum u_n$ absolument convergente[9] est convergente.

On dit également que $(E, \|\cdot\|)$ est un **espace de Banach**. ✦ Les espaces vectoriels normés de dimension finie sont tous complets.

Exemple A.46 L'espace $\mathscr{C}^0([0\,;\,1], \mathbb{R})$, muni de la norme $\|\cdot\|_\infty$, est complet. ✦ L'espace $L^1(I, \mathbb{R})$, muni de la norme $\|f\|_1 := \int_0^1 |f|$, est complet. ✦ En revanche, l'espace $\mathscr{C}^0([0\,;\,1], \mathbb{R})$ muni de $\|f\|_1 := \int_0^1 |f|$, n'est pas complet : la suite de terme général $f_n : x \mapsto x^n$ est de Cauchy, mais ne converge pas.

A.13 RÉDUCTION, THÉORÈME SPECTRAL

Soit E un **espace euclidien**, c'est-à-dire un espace vectoriel réel de dimension finie muni d'un produit scalaire. On note $\mathcal{B} = (e_1, e_2, \ldots, e_n)$ une base orthonormée de E.

Adjoint A59

Si u est un endomorphisme de E, de matrice représentative A dans la base orthonormée $\mathcal{B}$, on définit l'**adjoint** u^* de u comme l'unique endomorphisme dont la matrice représentative dans $\mathcal{B}$ est ${}^tA$ (transposée de A). C'est également l'unique endomorphisme vérifiant

$$\forall x, y \in E \qquad \langle u^*(x) | y \rangle = \langle x | u(y) \rangle.$$

---

[9]. C'est-à-dire telle que $\sum \|u_n\|$ converge

Cette relation, écrite matriciellement

$$\forall X, Y \in \mathfrak{M}_{n,1}(\mathbb{R}) \qquad {}^t(A^*X) \cdot Y = {}^tX \cdot AX$$

conduit d'ailleurs directement à la relation $A^* = {}^tA$.

Un endomorphisme u tel que $u = u^*$ est dit **autoadjoint** ou **symétrique**. Cette propriété équivaut au fait que sa matrice représentative *dans une base orthonormée* soit symétrique ($A = {}^tA$).

Un endomorphisme $u \in \mathscr{L}(E)$ est dit **orthogonal** si $uu^* = u^*u = \mathrm{Id}$. Cette propriété équivaut au fait que sa matrice représentative *dans une base orthonormée* soit une **matrice orthogonale**, c'est-à-dire $A{}^tA = {}^tAA = I_n$. ✦ La matrice représentative d'une base orthonormée dans une autre est une matrice orthogonale.

A60 **Théorème spectral**

• *Version géométrique :* soit $u \in \mathscr{L}(E)$ un endomorphisme symétrique d'un espace vectoriel euclidien E ; alors il existe une base orthonormée propre pour u, c'est-à-dire dans laquelle la matrice représentative de u est diagonale.

• *Version matricielle :* soit $A \in \mathfrak{M}_n(\mathbb{R})$ une matrice symétrique ; alors il existe une matrice orthogonale U telle que ${}^tUAU$ soit diagonale.

A61 **Matrices symétriques positives, définies positives**

Une matrice S symétrique est dite **symétrique positive** si ${}^tXSX \geqslant 0$ pour tout vecteur colonne $X \in \mathfrak{M}_{n,1}(\mathbb{R})$. Elle est dite **symétrique définie positive** si ${}^tXSX > 0$ pour tout vecteur colonne X non nul. ✦ Une matrice est positive si et seulement si son spectre est inclus dans $[0 ; +\infty[$, et définie positive si et seulement si son spectre est inclus dans $]0 ; +\infty[$. ✦ Soit S une matrice symétrique positive. Si l'on ne garde qu'un certain nombre de lignes de S, et uniquement colonnes de mêmes numéros, le résultat est une matrice extraite qui est elle-même symétrique positive. ✦ Enfin, si A est une matrice carrée, alors ${}^tA\,A$ est une matrice symétrique positive : pour tout vecteur colonne,

$${}^tX({}^tA\,A)X = {}^t(AX)\,(AX) = \|AX\|^2 \geqslant 0.$$

Si A est de plus inversible, alors ${}^tA\,A$ est symétrique définie positive.

A.14 MATRICES POSITIVES

A62 **Rayon spectral**

Soit $A \in \mathfrak{M}_n(\mathbb{C})$. Le **rayon spectral** de A est défini comme le plus grand module des ses valeurs propres :

$$\rho(A) = \max_{\lambda \in \mathrm{Sp}\,A} |\lambda| .$$

A63 **Matrices à coefficients positifs**

Une matrice A (de taille quelconque) est **positive** si tous ses coefficients sont positifs[10] et on note $A \geqslant 0$. Elle est **strictement positive** si tous ses coefficients sont strictement positifs, et on note $A > 0$.

■ **Théorème A.47 (Matrices positive)** *Soit $A \in \mathfrak{M}_n(\mathbb{R})$ une matrice positive. Alors $\rho(A)$ est valeur propre de A, et il existe un vecteur propre positif (à droite ou à gauche) associé à cette valeur propre.*

Une matrice positive A est dite **stochastique** si la somme des coefficients sur chaque ligne vaut 1 :

$$\forall i \in [1 ; n] \qquad \sum_{j=1}^{n} a_{ij} = 1.$$

(10). On prendra garde à ne pas confondre avec la notion de matrice *symétrique positive*, qui est une matrice réelle carrée symétrique vérifiant ${}^tXMX \geqslant 0$ pour tout vecteur X, voir A61).

En notant $\mathbf{1}$ le vecteur colonne de taille n dont tous les coefficients valent 1, A est donc stochastique si et seulement si $A > 0$ et $A\mathbf{1} = \mathbf{1}$.

Dans ce contexte, on appellera **loi** un vecteur ligne $\boldsymbol{\pi} \in \mathbb{R}^n$ telle que $\boldsymbol{\pi} \geqslant 0$ et $\boldsymbol{\pi} \cdot \mathbf{1} = 1$.

■ **Théorème A.48 (Matrices stochastiques)** *Soit* $A \in \mathfrak{M}_n(\mathbb{R})$ *une matrice stochastique. Alors*

i) $\rho(A) = 1$;

ii) 1 *est valeur propre de* A ;

iii) de plus, il existe un vecteur ligne propre $\boldsymbol{\pi} \geqslant 0$ *associé à cette valeur propre, c'est-à-dire un vecteur ligne tel que* $\boldsymbol{\pi} A = \boldsymbol{\pi}$.

En normalisant ce vecteur ligne, on peut obtenir une loi ; il n'y a cependant pas unicité de cette loi (le sous-espace propre associé à 1 peut être de dimension $\geqslant 2$).

Matrices irréductibles, matrices ergodiques A64

Une matrice positive est dite **irréductible** s'il n'existe aucune partie I de $[\![1 ; n]\!]$ (à part $\varnothing$ et I tout entier), telle que

$$\forall i \in I \qquad \forall j \notin I \qquad a_{i,j} = 0.$$

Autrement dit, A est irréductible s'il n'existe pas de matrice de permutation Q_σ telle que $Q_\sigma^{-1} A Q_\sigma$ soit de la forme

$$Q_\sigma^{-1} A Q_\sigma = \left(\begin{array}{|c|c|} \hline * & * \\ \hline 0 & * \\ \hline \end{array} \right).$$

■ **Proposition A.49** *Une matrice* $A \geqslant 0$ *est irréductible si et seulement si* $(I + A)^{n-1} > 0$.

■ **Théorème A.50 (Perron-Frobenius)** *Soit* $A \in \mathfrak{M}_n(\mathbb{R})$ *une matrice positive irréductible. Alors*

- $\rho(A)$ *est valeur propre ;*
- *le sous-espace propre (à gauche ou à droite) associé à* $\rho(A)$ *est de dimension* 1, *et dirigé par un vecteur strictement positif ; il existe donc une unique loi* $\boldsymbol{\pi}^* > 0$ *telle que* $\boldsymbol{\pi}^* A = \boldsymbol{\pi}^*$.

Dans le contexte probabiliste, $\boldsymbol{\pi}^*$ est l'unique **loi invariante** de A.

Une matrice positive $A \in \mathfrak{M}_n(\mathbb{R})$ est dite **ergodique**, ou **fortement irréductible**, ou **régulière**, s'il existe un entier m tel que $A^m > 0$. Une telle matrice est alors irréductible. Le théorème principal (dont la démonstartion se trouve page 566) est le suivant :

■ **Théorème A.51 (Matrices ergodiques)** *Soit* $A \in \mathfrak{M}_n(\mathbb{R})$ *une matrice positive ergodique. Alors toute valeur propre* λ *de* A *telle que* $\lambda \neq 1$ *vérifie* $|\lambda| < 1$. *Notons* $\boldsymbol{\pi}^*$ *l'unique loi invariante. Alors pour tout vecteur* $\boldsymbol{\pi}_0 \geqslant 0$ *normalisé à* $\boldsymbol{\pi}_0 \mathbf{1} = 1$, *on a la convergence* $\boldsymbol{\pi}_n := \boldsymbol{\pi}_0 A^n \to \pi^*$.

Notez un cas particulier : si A est stochastique et $A > 0$, alors A est irréductible et même ergodique.

Voici un résumé des quelques théorèmes précédents :

matrice	$\rho(A) \in \operatorname{Sp} A$ $\overrightarrow{\text{v.p.}}\ \boldsymbol{\pi} \geqslant 0$	$\rho(A) = 1 \in \operatorname{Sp} A$ $\overrightarrow{\text{v.p.}}\ \boldsymbol{\pi} \geqslant 0$	$\dim E_1(A) = 1$ $\boldsymbol{\pi}^* > 0$	$\boldsymbol{\pi}_n \to \boldsymbol{\pi}^*$
positive	✓			
stochastique	✓	✓		
irréductible	✓	✓	✓	
ergodique	✓	✓	✓	✓

EXERCICES

♦ ***Exercice A.1 (Contre-exemple au théorème de Fubini discret)***
Pour tout couple $(n,p) \in \mathbb{N}^2$, on note $a_{n,p} = 1/(n^2 - p^2)$ si $n \neq p$, et $a_{n,p} = 0$ sinon. Calculer $S_n = \sum_{p=0}^{\infty} a_{n,p}$. En déduire $S = \sum_{n=0}^{\infty} \sum_{p=0}^{\infty} a_{n,p}$. Par des arguments de symétrie, que vaut $T = \sum_{p=0}^{\infty} \sum_{n=0}^{\infty} a_{n,p}$? Conclure.

♦ ***Exercice A.2 (Continuité de la fonction de Cantor)*** Montrer que la suite $(f_n)_{n\geqslant 0}$ définie dans le paragraphe A28 converge uniformément.

Indications

◊ *Indication A.1 :* Décomposer en éléments simples. On trouve $S_n = 1/4n^2$ si $n \neq 0$, et $\pi^2/6$ sinon. $S = -\pi^2/8$. Par symétrie, $T = \pi^2/8$. La famille n'est donc pas sommable.

◊ *Indication A.2 :* Montrer que $\|f_{n+1} - f_n\|_\infty = \frac{1}{2}\|f_n - f_{n-1}\|_\infty$ et en déduire que la série $\sum(f_{n+1} - f_n)$ converge uniformément sur $[0\,;1]$.

Annexe B

Davantage sur la mesure

Donnons ici quelques compléments sur les tribus et les fonctions mesurables.

- *Nous commençons par passer en revue les divers boréliens utiles en théorie des probabilités : boréliens de* $\mathbb{R}$ *et de* $\mathbb{R}^n$*, puis boréliens de* $\mathbb{R}^T$ *où* T *est un ensemble dénombrable (*A 67*, utile pour le jeu de* pile *ou* face *par exemple) ou non dénombrable (*A 68*, utile pour les processus à temps continus comme le mouvement brownien).*
- *Après quelques définitions générales sur les fonctions mesurables, nous démontrerons le théorème de factorisation des fonctions mesurables par rapport à la tribu* $\sigma(X)$.
- *Nous terminons le chapitre par un rapide tour d'horizon des techniques liées au lemme de classe monotone, et qui sont d'une très grande puissance. Une description complète de la mesure* **P** *sur chacun des éléments de* $\mathcal{T}$ *étant, en pratique, impossible, on montre qu'il suffit de la définir sur un ensemble* $\mathcal{A}$ *plus réduit — le but étant évidemment de chercher l'ensemble* $\mathcal{A}$ *le plus petit possible :*
  - *— pour définir* **P** ex nihilo*, on la définit d'abord sur une algèbre* $\mathcal{A}$ *engendrant* $\mathcal{T}$ *;*
  - *— si au contraire on suppose que* **P** *existe, alors elle est caractérisée par ses valeurs prises sur une structure encore plus simple (un* π*-système).*

B.1 BORÉLIENS

Boréliens de $\mathbb{R}$ A65

Tout d'abord, une petite remarque d'ordre topologique. Soit U un ouvert de $\mathbb{R}$. On définit la relation d'équivalence « $\sim$ » par : $x \sim y$ si et seulement si le segment $[x\,;y]$ est inclus dans U. La classe d'équivalence d'un élément x de U est donc un intervalle ouvert. L'ensemble U est la réunion disjointe de ses classes d'équivalence modulo $\sim$; si l'on choisit, dans chaque classe d'équivalence, un représentant *rationnel*, on a donc montré que :

■ **Lemme B.1** *Tout ouvert* U *de* $\mathbb{R}$ *est union dénombrable d'intervalles ouverts disjoints.*

La **tribu $\mathfrak{B}$ des boréliens de $\mathbb{R}$** est la tribu engendrée par la classe $\mathcal{I}$ des intervalles ouverts.

■ **Proposition B.2** *La tribu $\mathfrak{B}$ est :*

- *la tribu engendrée par les ouverts ;*
- *la tribu engendrée par les fermés ;*
- *la tribu engendrée par les intervalles de $\mathbb{R}$;*
- *la tribu engendrée par les intervalles de la forme $]-\infty\,;\,a]$ pour $a \in \mathbb{R}$;*
- *la tribu engendrée par les intervalles de la forme $]-\infty\,;\,a]$ pour $a \in \mathbb{Q}$.*

DÉMONSTRATION : Notons $\mathcal{I}$ la classe des intervalles ouverts de $\mathbb{R}$, $\mathcal{O}$ celle des ouverts, $\mathcal{H}$ celle des intervalles de la forme $]-\infty\,;\,a]$ pour a réel et $\mathcal{H}'$ celle des intervalles $]-\infty\,;\,a[$.

- Tout intervalle $]-\infty\,;\,a]$ peut s'écrire comme l'intersection dénombrable d'intervalles ouverts $]-\infty\,;\,a+1/n[$, donc $\mathcal{H} \subset \sigma(\mathcal{H}')$.
- De même, tout intervalle $]-\infty\,;\,a[$ est union dénombrable d'intervalles $]\,a\,;\,a-1/n\,]$, donc $\mathcal{H}' \subset \sigma(\mathcal{H})$. Par croissance, on en déduit que $\sigma(\mathcal{H}) = \sigma(\mathcal{H}')$.
- Un intervalle ouvert $]\,a\,;\,b\,[$ s'écrit comme l'intersection de $]-\infty\,;\,b\,[$ et du complémentaire de $]-\infty\,;\,a]$, il est donc élément la tribu $\sigma(\mathcal{H})$. Cela prouve l'inclusion $\mathcal{I} \subset \sigma(\mathcal{H})$. On a de plus $\mathcal{H}' \subset \mathcal{O}$ et on a montré eu début du paragraphe que $\mathcal{O} \subset \sigma(\mathcal{I})$; on en déduit
$$\mathfrak{B} := \sigma(\mathcal{I}) \subset \sigma(\mathcal{H}) = \sigma(\mathcal{H}') \subset \sigma(\mathcal{O}) \subset \sigma(\mathcal{I}) =: \mathfrak{B},$$
ce qui montre l'égalité entre toutes ces tribus.
- Par passage au complémentaire, les tribus engendrées par les ouverts et par les fermés sont identiques.
- Enfin, tout intervalle $]-\infty\,;\,a]$ est intersection dénombrable d'intervalles $]-\infty\,;\,q_n\,]$, où $(q_n)_{n\geqslant 1}$ est une suite décroissante de rationnels de limite a, ce qui fait que la tribu engendrée par les intervalles rationnels est également $\sigma(\mathcal{H}) = \mathfrak{B}$.

∎

L'avant-dernière des caractérisations de $\mathfrak{B}(\mathbb{R})$ est la plus utile :

— elle permet de montrer certaines propriétés sur les intervalles $]-\infty\,;\,a]$ et, par « passage à la limite » (utilisant un théorème de monotonie), d'en déduire que ces propriétés sont vraies pour tous les boréliens ;

— elle permet également de justifier que : la fonction de répartition d'une variable aléatoire caractérise entièrement sa loi (ou, ce qui revient au même : la fonction de répartition d'une probabilité $\mathbf{Q}$ sur $\mathbb{R}$ caractérise entièrement cette probabilité).

Remarque B.3 Les singletons $\{x\}$ sont tous éléments de la tribu $\mathfrak{B}(\mathbb{R})$. Cependant, ils sont loin d'engendrer $\mathfrak{B}(\mathbb{R})$: ils n'engendrent que la tribu formée des parties dénombrables de $\mathbb{R}$ et de leur complémentaire (voir exercice 3.5).

Remarque B.4 (La tribu borélienne est indescriptible) On pourrait croire que la tribu $\mathfrak{B}$, qui est la plus petite tribu contenant les intervalles et stable par union dénombrable et passage au complémentaire, pourrait se décrire en termes d'unions ou intersections dénombrables d'intervalles.

C'est hélas (très) loin d'être le cas. Notons $\mathfrak{A}_0$ l'algèbre des unions dénombrables d'intervalles, de telle sorte que $\mathfrak{B} = \sigma(\mathfrak{A}_0)$. On définit alors $\mathfrak{A}_1$ comme l'algèbre des parties de $\mathbb{R}$ qui sont union dénombrables ou intersections dénombrables d'éléments de $\mathfrak{A}_0$. C'est bien une algèbre, mais — c'est plus délicat — ce n'est pas une tribu. Définissons de la même manière $\mathfrak{A}_2$ comme l'algèbre des parties de $\mathbb{R}$ qui sont unions dénombrables ou intersections dénombrables d'éléments de $\mathfrak{A}_1$, et de même par récurrence $\mathfrak{A}_{n+1}$ à partir de $\mathfrak{A}_n$. Alors la suite $(\mathfrak{A}_n)_{n\geqslant 0}$ est strictement croissante, aucun de ses membres n'est une tribu. Pire, même l'union $\mathfrak{A}_\infty = \bigcup_{n=0}^{\infty} \mathfrak{A}_n$ n'est pas une tribu[1], et il faut poursuivre la construction au-delà du dénombrable, en s'aventurant dans le domaine du transfini[2].

(1). Voir par exemple Kechris [57, chap. 22], ou bien Billingsley [9, section 2, p.30].

(2). Là encore, cela nous entraînerait bien trop loin ; le lecteur peut se reporter à l'ouvrage de Daniel Saada [88, chap. 7].

Boréliens de $\mathbb{R}^n$ A66

Si $\mathfrak{F}$ et $\mathfrak{G}$ sont des tribus, on note $\mathfrak{F} \otimes \mathfrak{G}$ la tribu engendrée par leur produit cartésien :

$$\mathfrak{F} \otimes \mathfrak{G} := \sigma(\mathfrak{F} \times \mathfrak{G}).$$

On note $\mathfrak{B}(\mathbb{R}^2)$ — ou plus simplement $\mathfrak{B}^2$ — la tribu engendrée par les ouverts de $\mathbb{R}^2$, qu'on appelle **tribu des boréliens de $\mathbb{R}^2$**.

De même que précédemment, on montre que tout ouvert de $\mathbb{R}^2$ est union dénombrable de rectangles ouverts $]a\,;b[\times]c\,;d[$[(3)], on en déduit que $\mathfrak{B}^2$ est également la tribu engendrée par les rectangles. En suivant un raisonnement similaire à celui effectué dans $\mathbb{R}$, on montre que c'est également la tribu engendrée par les domaines

$$\Delta(a,b) :=]-\infty\,;a] \times]-\infty\,;b].$$

Enfin, on peut montrer que c'est la tribu engendré par les rectangles généralisés $B_1 \times B_2$, où B_1, B_2 sont des boréliens de $\mathbb{R}$, c'est-à-dire que

$$\mathfrak{B}^2 = \mathfrak{B} \otimes \mathfrak{B}.$$

Tous ces résultats se généralisent bien entendu à $\mathbb{R}^n$:

■ **Théorème B.5** *La tribu $\mathfrak{B}^n$ des boréliens de $\mathbb{R}^n$ est engendrée par*

- *les rectangles (ouverts ou fermés) ;*
- *les domaines $\Delta(a_1, a_2, \ldots, a_n)$;*
- *les ouverts (ou les fermés) ;*
- *les rectangles généralisés $B_1 \times B_2 \times \cdots \times B_n$.*

Boréliens de $\mathbb{R}^\infty$ A67

Notons $\mathbb{R}^{\mathbb{N}^*}$ ou $\mathbb{R}^\infty$ l'ensemble des suites $x = (x_n)_{n \geqslant 1}$ de réels. Un **cylindre** est un ensemble de la forme

$$C(B_1, B_2, \ldots, B_r) = \{(x_n)_{n \geqslant 1}\ ;\ x_1 \in B_1, x_2 \in B_2, \ldots, x_r \in B_r\},$$

où l'entier r est quelconque, et $B_1, B_2, \ldots, B_r$ sont des boréliens de $\mathbb{R}$. La **tribu des boréliens de $\mathbb{R}^\infty$**, notée $\mathfrak{B}^\infty$, est la tribu engendrée par la classe de tous les cylindres. On peut montrer que c'est également la tribu engendrée par, au choix

- les cylindres de bases rectangulaires

$$C(I_1, I_2, \ldots, I_r) := \{(x_n)_{n \geqslant 1}\ ;\ x_1 \in I_1, x_2 \in I_2, \ldots, x_r \in I_r\},$$

  où les I_k sont des intervalles et r un entier quelconque ;
- les cylindres de la forme

$$C(B_r) := \{(x_n)_{n \geqslant 1}\ ;\ (x_1, x_2, \ldots, x_r) \in B_r\},$$

  où B_r est un borélien de $\mathbb{R}^r$ et r un entier quelconque.

Tribus produits, espaces produits A68

On généralise d'abord la notion de tribu produit. Soit $(\Omega_i, \mathcal{T}_i)_{i \in I}$ une famille d'espaces probabilisables, l'ensemble I étant quelconque (en particulier, il n'est pas nécessairement dénombrable). On note $\Omega = \prod_{i \in I} \Omega_i$, produit cartésien des Ω_i. Les **cylindres** de Ω sont les éléments de la forme

$$C = \prod_{i \in I} B_i,$$

où $B_i \in \mathcal{T}_i$ pour tout $i \in I$, et $B_i = \Omega_i$ *sauf pour un nombre fini d'indices*. On note alors $\mathcal{T} = \bigotimes_{i \in I} \mathcal{T}_i$ la tribu engendrée par la classe $\mathcal{C}$ des cylindres de Ω.

(3). On ne peut pas calquer exactement la démonstration précédente ; voici donc une approche légèrement différente. Écrivons que, pour tout $x \in U$, il existe un rectangle ouvert R_x dont les sommets ont des coordonnées rationnelles, contenant x et inclus dans U (les rectangles ainsi construits ne sont pas disjoints). On peut alors écrire que $U = \bigcup_{x \in U} R_x$. L'union est *a priori* non dénombrable — elle porte sur les éléments de U — mais les rectangles rationnels sont en quantité dénombrable : l'union se restreint donc en réalité à une union dénombrable.

Exemple B.6 (Boréliens de $\mathbb{R}^T$) Soit T un ensemble non dénombrable, par exemple $T = [0\,;\,1]$ ou bien $T = [0\,;\,+\infty[$. On munit l'univers $\mathbb{R}^T$ des fonctions de T dans $\mathbb{R}$ de la tribu produit $\bigotimes_{t\in T} \mathfrak{B}_t$, où tous les $\mathfrak{B}_t$ sont égaux à la tribu des boréliens de $\mathbb{R}$.

Alors, chaque élément A de $\mathfrak{B}(\mathbb{R}^T)$ est spécifié par un nombre *dénombrable* de contraintes : il existe une suite $(t_n)_{n\geqslant 1}$ d'éléments de T et un borélien B de $\mathfrak{B}^\infty$ tels que

$$A = \left\{ x \in \mathbb{R}^T \,;\, (x_{t_1}, x_{t_2}, \dots) \in B \right\}.$$

(voir par exemple Shiryaev [91] pour une démonstration).

Si chacun des espaces est muni d'une probabilité $\mathbf{P}_i$, on définit la **probabilité produit** $\mathbf{P} = \bigotimes_{i\in I} \mathbf{P}_i$ par

$$\mathbf{P}(C) = \prod_{i\in I} \mathbf{P}_i(B_i)$$

lorsque $C = \prod_i B_i$ est un cylindre (le produit est bien défini puisque, à l'exception d'un nombre fini d'entre eux, tous les facteurs valent 1). La formule précédente s'étend, de manière unique, sur la tribu $\mathfrak{T}$ engendrée par la classe des cylindres ; nous l'admettons ici dans le cas général[4] (un cas particulier est démontré dans la section suivante). On note

$$\bigotimes_{i\in I} (\Omega_i, \mathfrak{T}_i, \mathbf{P}_i)$$

l'espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ ainsi défini.

Exemple B.7 Notons $\Omega = \{P, F\}$, $\mathfrak{T} = \{\varnothing, \{P\}, \{F\}, \{P, F\}\}$ et $\mathbf{P}$ la probabilité caractérisée par $\mathbf{P}(\{P\}) = p$. Alors l'espace probabilisé utilisé pour modéliser une succession infinie de tirages à *pile* ou *face* (et décrit au début du chapitre 9), n'est autre que l'espace produit $\bigotimes_{n\geqslant 1} (\Omega, \mathfrak{T}, \mathbf{P})$.

B.2 FONCTIONS MESURABLES

A69 **Définition et premières propriétés**

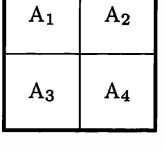
Soient $(E, \mathfrak{E})$ et $(F, \mathfrak{F})$ deux espaces probabilisables (ou espaces mesurés). Une **application mesurable** entre ces espaces est une application $f : E \to F$ telle que :

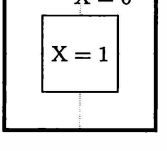
$$\forall B \in \mathfrak{F} \qquad f^{-1}(B) \in \mathfrak{E}.$$

Les variables aléatoires sont des fonctions mesurables de $(\Omega, \mathfrak{T})$ dans $(\mathbb{R}, \mathfrak{B})$.

Exemple B.8 (Fonction mesurable par rapport à une partition) Soit $\mathfrak{D} = \{A_k \,;\, k \in K\}$ une partition dénombrable de Ω. Soit Y une variable aléatoire qui soit $\sigma(\mathfrak{D})$-mesurable, c'est-à-dire que $\{Y \in B\} \in \sigma(\mathfrak{D})$ pour tout borélien B. Notamment, pour tout réel x, l'événement $\{X = x\}$ appartient à $\sigma(\mathfrak{D})$: il est donc vide ou réunion dénombrable de A_k. En d'autres termes, la fonction Y est constante sur chacun des atomes A_k (les constantes pouvant être égales sur des A_k distincts). Réciproquement, si Y est constante sur chaque A_k, alors elle est $\sigma(\mathfrak{D})$-mesurable.

Sur le schéma ci-dessous, on représenta à gauche une partition finie $\mathfrak{D}$. À droite, la variable aléatoire X prend deux valeurs ; elle n'est pas $\sigma(\mathfrak{D})$-mesurable.





Dans la suite, nous ne considérerons pratiquement que des applications à valeurs dans l'espace mesuré $(\mathbb{R}, \mathfrak{B})$. Un cas particulier doit retenir notre attention : une application mesurable f de $(\mathbb{R}, \mathfrak{B})$ dans $(\mathbb{R}, \mathfrak{B})$ est appelée **fonction borélienne**.

(4). Voir par exemple [9, chapitre 7].

Tribu engendrée par une variable aléatoire A70

Si X est une variable aléatoire, on rappelle que la **tribu engendrée par X** est la classe des événements de la forme $\{X \in A\}$, où A décrit l'ensemble des boréliens de $\mathbb{R}$[5]. On la note $\sigma(X)$. C'est donc la plus petite tribu qui rende X mesurable.

Si X, Y sont des variables aléatoires, la **tribu engendrée par (X, Y)**, notée $\sigma(X, Y)$, est la plus petite tribu qui rende X et Y mesurables : c'est la classe des ensembles $\{(X, Y) \in A\}$, où A parcourt l'ensemble $\mathfrak{B}^2$ des boréliens de $\mathbb{R}^2$. Puisque $\mathfrak{B}^2 = \mathfrak{B} \otimes \mathfrak{B}$, c'est également la tribu engendrée par les événements $\{X \in A\}$ et $\{Y \in B\}$, pour A et B parcourant $\mathfrak{B}$.

De même, la **tribu engendrée par une famille $(X_i)_{i\in I}$** de variables aléatoires est la plus petite tribu rendant chacune des variables X_i mesurable. C'est également la tribu engendré par les événements liés à chaque variable :

$$\sigma(X_i \ ; \ i \in I) = \sigma(\{X_i \in A\} \ ; \ i \in I, A \in \mathfrak{B}).$$

Si $\mathfrak{F}$ est une sous-tribu de $\mathfrak{T}$, on dira qu'une variable aléatoire $Y : \Omega \to \mathbb{R}$ est **mesurable par rapport à la tribu $\mathfrak{F}$** ou, en abrégé, **$\mathfrak{F}$-mesurable**, si, pour tout borélien B, $\{Y \in B\}$ est dans $\mathfrak{F}$, c'est-à-dire si $\sigma(Y) \subset \mathfrak{F}$. On écrit alors $\mathfrak{F} \succcurlyeq \sigma(Y)$ (attention au sens !).

Le lemme de factorisation généralise l'exemple B.8 et donne une interprétation pour qu'une fonction Y soit $\sigma(X)$ mesurable :

■ **Lemme B.9 (Doob-Dynkin : factorisation d'une variable aléatoire)** *Soient* X *et* Y *des variables aléatoires ; alors* Y *est* $\sigma(X)$*-mesurable si et seulement si il existe une fonction borélienne* f *telle que* $Y = f(X)$.

Démonstration du théorème B.9 : Un des sens est immédiat : si $Y = f(X)$ avec f borélienne, alors Y est $\sigma(X)$-mesurable.

Réciproquement, supposons que Y est $\sigma(X)$-mesurable, c'est-à-dire que $\sigma(Y) \subset \sigma(X)$, et laissons-nous guider par la Cavalerie Légère (§ 215).

1e cas : Y est une variable aléatoire simple, c'est-à-dire qu'il existe une partition

$$\Omega = A_1 + A_2 + \cdots + A_n$$

et des réels distincts $\alpha_1, \alpha_2, \ldots, \alpha_n$ tels que $Y = \sum_k \alpha_k \mathbf{1}_{A_k}$. Pour $k = 1, \ldots, n$, l'élément $A_k = Y^{-1}(\{\alpha_k\})$ est élément de $\sigma(X)$, il existe donc un borélien B_k tel que $A_k = X^{-1}(B_k)$. En d'autres termes : sur A_k, X prend ses valeurs dans B_k. Attention, les B_k ne sont pas forcément disjoints, mais une intersection du type $B_k \cap B_\ell$ ne contient *que des valeurs non atteintes par* X.

Si l'on définit f par ses restrictions aux B_k, constantes et égales à α_k :

$$f = \sum_{k=1}^{n} \alpha_k \mathbf{1}_{B_k},$$

alors $Y = f(X)$, et f est naturellement borélienne.

2e cas : Y est une variable aléatoire positive. On sait (théorème 13.25) qu'on peut l'approcher par une suite croissante de variables aléatoires étagées : $Y = \lim\limits_{n\to\infty} \uparrow Y_n$. On factorise chaque variable aléatoire Y_k selon la méthode précédente : $Y_k = f_k(X)$. La suite $(f_n)_{n\geqslant 0}$ est donc croissante sur $X(\Omega)$: on pose $f := \lim\limits_{k\to\infty} f_k$. On a alors, pour tout $\omega \in \Omega$:

$$Y(\omega) = \lim_{k\to\infty} Y_k(\omega) = \lim_{k\to\infty} f_k\big(X(\omega)\big) = f\big(X(\omega)\big),$$

c'est-à-dire $Y = f(X)$.

Enfin, le théorème 13.7 page 293 prouve que la limite simple de fonctions mesurables est mesurable, donc f est bien borélienne.

3e cas : Y est une variable aléatoire quelconque. On décompose $Y = Y^+ - Y^-$, on factorise chaque partie $Y^+ = f^+(X)$, $Y^- = f^-(Y)$, puis on pose $f = f^+ - f^-$. ∎

(5). On peut bien sûr généraliser ce résultat au cas où $X : E \to F$ est une application d'un ensemble mesuré $(E, \mathfrak{E})$ dans un ensemble mesuré $(F, \mathfrak{F})$. Le lecteur imaginera sans peine comment modifier les théorèmes présentés pour englober ce cas de figure.

B.3 UN COMPLÉMENT AU JEU DE PILE OU FACE

Nous démontrons ici que l'application $\mathbf{P}_0$ définie sur la classe des cylindres de l'univers $\Omega = \{\mathrm{P},\mathrm{F}\}^{\mathbb{N}^*}$ par l'équation (9.1) page 204 est en réalité σ-additive.

A71 **Additivité dénombrable de $\mathbf{P}_0$ sur la classe $\mathfrak{C}$ des cylindres (*)**
Pour le prouver, nous allons montrer d'abord montrer un résultat plus fort : *si* $(\mathrm{C}_n)_{n\geqslant 1}$ *est une suite décroissante de cylindres non vides, alors leur intersection* $\bigcap_n \mathrm{C}_n$ *est non vide.*

On suppose donc donnée une suite $(\mathrm{C}_n)_{n\geqslant 1}$ de cylindres non vides. Chacun des cylindres C_k s'écrit

$$\mathrm{C}_k = \big\{\omega\in\Omega\,;\ (\omega_1,\omega_2,\dots,\omega_{r_k})\in \mathrm{H}_k\big\},$$

où r_k est un entier et H_k une partie de $\{\mathrm{P},\mathrm{F}\}^{r_k}$. Chacun de ces cylindres étant non vide, on peut choisir un élément ω^k dans C_k, pour tout $k\geqslant 1$. On peut écrire les coordonnées de ces vecteurs sous la forme d'un tableau :

$$\begin{array}{llll}\omega^1 = (\omega_1^1 & \omega_2^1 & \omega_3^1 & \cdots)\\ \omega^2 = (\omega_1^2 & \omega_2^2 & \omega_3^2 & \cdots)\\ \omega^3 = (\omega_1^3 & \omega_2^3 & \omega_3^3 & \cdots)\\ \vdots & \vdots & \vdots & \vdots\end{array}$$

Au moins l'un des deux résultats P ou F apparaît une infinité de fois dans la première *colonne* du tableau — appelons-le R_1. Il apparaît dans une infinité de lignes, dont nous noterons les indices sous la forme d'une suite strictement croissante $\big(n_1(k)\big)_{k\geqslant 1}$. Ainsi, on a

$$\omega_1^{n_1(k)} = \mathrm{R}_1 \quad \text{pour tout } k\geqslant 1.$$

Regardons maintenant les éléments de la deuxième colonne, en ne retenant que ceux dont l'indice de ligne est dans la suite précédente, c'est-à-dire la suite des $\omega_2^{n_1(k)}$; parmi ceux-ci, il y en a une infinité prenant un même résultat R_2 (valant P ou F), en des indices que nous écrirons sous la forme $\big(n_2(k)\big)_{k\geqslant 1}$. La suite $\big(n_2(k)\big)_{k\geqslant 1}$ vérifie

$$\omega_2^{n_2(k)} = \mathrm{R}_2 \quad \text{pour tout } k\geqslant 1.$$

Comme c'est une une sous-suite de la suite $\big(n_1(k)\big)_{k\geqslant 1}$, elle vérifie également

$$\omega_1^{n_2(k)} = \mathrm{R}_1 \quad \text{pour tout } k\geqslant 1.$$

On continue le processus en examinant la troisième colonne, puis la quatrième etc. On construit ainsi par récurrence des suites $\big(n_i(k)\big)_{k\geqslant 1}$ strictement croissantes, ainsi qu'une suite $(\mathrm{R}_k)_{k\geqslant 1}$ à valeurs dans $\{\mathrm{P},\mathrm{F}\}$, telles que

- $\big(n_{i+1}(k)\big)_{k\geqslant 1}$ est une sous-suite de $\big(n_i(k)\big)_{k\geqslant 1}$;
- $\omega_i^{n_j(k)} = \mathrm{R}_i$ pour tout entier k et tout indice $i\leqslant j$.

On pose alors

$$\omega^* := (\mathrm{R}_1,\mathrm{R}_2,\mathrm{R}_3,\dots).$$

Un petit argument diagonal maintenant : on pose $\mathrm{N}_k = n_k(k)$ pour tout $k\geqslant 1$. On a donc

$$\omega_i^{\mathrm{N}_k} = \mathrm{R}_i \quad \text{pour tous } i,k \text{ tels que } k\geqslant i.$$

Notamment, pour tous entiers $i\leqslant k$,

$$(\omega_1^{\mathrm{N}_k},\omega_2^{\mathrm{N}_k},\dots,\omega_i^{\mathrm{N}_k}) = (\mathrm{R}_1,\mathrm{R}_2,\dots,\mathrm{R}_i) = (\omega_1^*,\omega_2^*,\dots,\omega_i^*). \qquad (*)$$

Il ne nous reste plus qu'à montrer que ω^* est élément de tous les cylindres C_p.

Soit donc p un entier arbitraire. Le cylindre C_p s'écrit

$$\mathrm{C}_p = \big\{\omega\in\Omega\,;\ (\omega_1,\omega_2,\dots,\omega_{r_p})\in \mathrm{H}_p\big\}.$$

On choisit alors un entier $k \geqslant r_p$. Par stricte croissance de la suite $(N_i)_{i\geqslant 1}$, on a $N_k \geqslant p$. L'élément ω^{N_k} appartient (par construction) à C_{N_k}, lui-même inclus dans C_p, ce qui prouve que l'ensemble H_p contient le point $(\omega_1^{N_k}, \omega_2^{N_k}, \dots, \omega_{r_p}^{N_k})$, qui d'après la relation $(*)$ n'est autre que $(\omega_1^*, \omega_2^*, \dots, \omega_{r_p}^*)$. Cela prouve bien que ω^* est élément de C_p.

La démonstration est sur le point de s'achever. Si $(C_n)_{n\geqslant 0}$ est une suite décroissante de cylindres telle que $C_n \downarrow \varnothing$, alors le raisonnement précédent montre, par contraposition, qu'au moins un des cylindres C_k est vide; ainsi $\mathbf{P}(C_n) = 0$ pour tout $n \geqslant k$, ce qui implique bien sûr que $\mathbf{P}(C_n) \to 0$. Ainsi, nous avons montré la *continuité monotone en* $\varnothing$ — axiome $3'''$ — donc nous savons qu'elle équivaut à la σ-additivité.

B.4 LE LEMME DE CLASSE MONOTONE

Comme on l'a déjà dit, les tribus sont généralement des objets terriblement compliqués, parfois au-delà de toute description possible. On s'intéresse aux classes plus simples d'événements que sont les π-systèmes et qui contiennent en essence une partie des propriétés de la tribu qu'ils engendrent. La notion de classe monotone est celle qui complète la précédente : une tribu est exactement un π-système qui est également une classe monotone.

Le lemme de classe monotone, dû au mathématicien polonais Wacław Sierpiński[6] mais parfois connu sous le nom de *lemme de Dynkin*, ou théorème π-λ chez les anglo-saxons, permet de prouver que deux probabilités, coïncidant sur un « petit » ensemble de parties, et plus précisément sur un π-système, sont en fait égales. En conséquence directe, on prouve que : la loi est entièrement caractérisée par la fonction de répartition, qui est donc bien, comme on l'a déjà dit, l'*outil fondamental* de la théorie des probabilités.

π-systèmes et classes monotones A72

On appelle **π-système** une famille $\mathfrak{C}$ de parties de Ω, qui est stable par intersection finie : si $A_1, A_2, \dots, A_n$ sont des éléments de $\mathfrak{C}$, alors $A_1 \cap A_2 \cap \cdots \cap A_n \in \mathfrak{C}$.

Exemple B.10 (de π-système) La famille des intervalles $]a\,;b]$, pour $a \leqslant b$, est un π-système de $\mathbb{R}$ (il est à noter qu'il contient $\varnothing =]0\,;0]$). ✦ De même, la famille des segments $[a\,;b]$ (pour $a \leqslant b$), à laquelle on adjoint l'ensemble vide, est un π-système. ✦ Enfin, exemple le plus important, l'ensemble des demi-droites $]-\infty\,;x]$ (pour $x \in \mathbb{R}$), est un π-système.

On appelle **classe monotone**[7] toute famille $\mathfrak{M}$ de parties de Ω stable par union dénombrable croissante et par différence, et qui contient Ω :

- $\Omega \in \mathfrak{M}$;
- si $(A_n)_{n\geqslant 1}$ est une suite d'éléments de $\mathfrak{M}$ telle que $A_n \uparrow A$, alors A est élément de $\mathfrak{M}$;
- pour tous A et B de $\mathfrak{M}$ tels que $A \subset B$, la différence $B - A$ est dans $\mathfrak{M}$. ⇗B.3

Remarque B.11 Toute tribu est une classe monotone. La réciproque n'est pas vraie. Si $\Omega = \{a,b,c,d\}$, notons $\mathfrak{C} = \big\{\varnothing, \Omega, \{a,b\}, \{a,c\}, \{a,d\}, \{b,c\}, \{b,d\}, \{c,d\}\big\}$. Alors $\mathfrak{C}$ est une classe monotone, mais pas une tribu.

Exemple B.12 (de classe monotone) Soient $\mathbf{P}$ et $\mathbf{Q}$ deux mesures de probabilités sur un même espace probabilisable $(\Omega, \mathcal{T})$. La classe $\mathfrak{M}$ des éléments A de $\mathcal{T}$ vérifiant $\mathbf{P}(A) = \mathbf{Q}(A)$ est une classe monotone. En effet, l'espace Ω est élément de $\mathfrak{M}$, puisque $\mathbf{P}(\Omega) = \mathbf{Q}(\Omega) = 1$. Si $A, B \in \mathfrak{M}$ et $A \subset B$, alors

$$\mathbf{P}(B - A) = \mathbf{P}(B) - \mathbf{P}(A) = \mathbf{Q}(B) - \mathbf{Q}(A) = \mathbf{Q}(B - A)$$

donc $B - A$ est encore dans $\mathfrak{M}$. Enfin, si les A_n sont dans $\mathfrak{M}$ et $A_n \uparrow A$, alors

$$\mathbf{P}(A) = \mathbf{P}(\lim\uparrow A_n) = \lim_{n\to\infty} \mathbf{P}(A_n) = \lim_{n\to\infty} \mathbf{Q}(A_n) = \mathbf{Q}(\lim\uparrow A_n) = \mathbf{Q}(A)$$

en vertu de l'axiome $3'$ de continuité monotone croissante.

---

(6). Wacław Sierpiński, *Un théorème général sur les familles d'ensembles*, 1928.

(7). En anglais : *d-system*, ou *Dynkin-system*, ou *λ-system*.

■ **Proposition B.13** *Une partie $\mathfrak{A}$ de $\mathfrak{P}(\Omega)$ est une tribu si et seulement si c'est à la fois un π-système et une classe monotone.*

DÉMONSTRATION : Si $\mathfrak{A}$ est une tribu, alors c'est évidemment un π-système et une classe monotone. Réciproquement, soit $\mathfrak{A}$ une classe de parties de A qui soit à la fois un π-système et une classe monotone. Alors $\mathfrak{A}$ contient Ω et est stable par passage au complémentaire par définition d'une classe monotone ; il ne reste qu'à prouver qu'elle est stable par union dénombrable.

Tout d'abord, vérifions qu'elle est stable par union. Soient A et B deux éléments de $\mathfrak{A}$, alors

$$A \cup B = \Omega - (A^c \cap B^c)$$

est élément de $\mathfrak{A}$ qui est stable par passage au complémentaire et par intersection. Par récurrence immédiate, $\mathfrak{A}$ est stable par union finie.

Soit $(A_n)_{n\geqslant 1}$ une suite d'éléments de $\mathfrak{A}$. Posons $B_n = \bigcup_{k=1}^{n} A_k$, qui est élément de $\mathfrak{A}$. La suite $(B_n)_{n\geqslant 1}$ est croissante, et composée d'éléments de $\mathfrak{A}$: sa limite $B = \bigcup_{k=1}^{\infty} A_k$ est donc dans $\mathfrak{A}$, ce qui achève la preuve. ∎

Toute algèbre étant un π-système, *une algèbre $\mathfrak{A}$ est une tribu si et seulement si c'est une classe monotone.*

■ **Lemme B.14 (de classe monotone / théorème π–λ de Dynkin)**
Toute classe monotone $\mathfrak{M}$ contenant un π-système $\mathfrak{C}$ contient la tribu engendrée par $\mathfrak{C}$. En d'autres termes, la plus petite classe monotone contenant un π-système $\mathfrak{C}$ est la tribu $\sigma(\mathfrak{C})$ engendrée par ce système.

DÉMONSTRATION : L'ensemble des classes monotones étant stable par intersection, si l'on note

$$\mathfrak{m}(\mathfrak{C}) := \bigcap_{\substack{\mathfrak{M} \text{ classe monotone} \\ \mathfrak{C}\subset\mathfrak{M}}} \mathfrak{M}$$

l'intersection de toutes les classes monotones contenant $\mathfrak{C}$, c'est encore une classe monotone, et c'est donc la plus petite classe monotone contenant $\mathfrak{C}$; on l'appelle **classe monotone engendrée par $\mathfrak{C}$.**

Puisque $\sigma(\mathfrak{C})$ est l'intersection de toutes les tribus contenant $\mathfrak{C}$ et que toute tribu est une classe monotone, on en déduit que $\mathfrak{C} \subset \mathfrak{m}(\mathfrak{C}) \subset \sigma(\mathfrak{C})$. En vertu de la proposition B.13, il suffit donc de prouver que $\mathfrak{m}(\mathfrak{C})$ est stable par intersection finie — ce sera l'équation (**) — pour établir que $\mathfrak{m}(\mathfrak{C}) = \sigma(\mathfrak{C})$, ce qui prouvera le théorème.

La preuve se fait en deux étapes.

- Fixons momentanément un élément A de $\mathfrak{C}$, et notons

$$\mathfrak{D}_A := \{B \in \mathfrak{m}(\mathfrak{C}) \ ; \ A \cap B \in \mathfrak{m}(\mathfrak{C})\}.$$

Par hypothèse, $\mathfrak{C}$ est stable par intersections finies, donc $\mathfrak{C} \subset \mathfrak{D}_A$. Or, $\mathfrak{D}_A$ est une classe monotone, car elle contient Ω et

— $\mathfrak{D}_A$ est stable par intersection : si B, B' sont dans $\mathfrak{m}(\mathfrak{C})$ et $B \subset B'$, alors

$$A \cap (B - B') = (A \cap B) - (A \cap B')$$

est différence de deux éléments de $\mathfrak{m}(\mathfrak{C})$, donc est dans $\mathfrak{m}(\mathfrak{C})$;

— $\mathfrak{D}_A$ est stable par union croissante : $A \cap \bigcup_n B_n = \bigcup_n (A \cap B_n)$ est dans $\mathfrak{m}(\mathfrak{C})$ lorsque les B_n le sont.

Par minimalité, $\mathfrak{m}(\mathfrak{C}) \subset \mathfrak{D}_A$; comme par construction $\mathfrak{D}_A \subset \mathfrak{m}(\mathfrak{C})$, on a montré que $\mathfrak{D}_A = \mathfrak{m}(\mathfrak{C})$. Ceci étant vrai pour tout $A \in \mathfrak{C}$, on a donc prouvé :

$$\forall A \in \mathfrak{C} \quad \forall B \in \mathfrak{m}(\mathfrak{C}) \qquad A \cap B \in \mathfrak{m}(\mathfrak{C}). \tag{*}$$

- Le résultat précédent est assez proche de ce que l'on veut, mais doit être maintenant démontré pour tout $A \in \mathfrak{m}(\mathfrak{C})$. Ce peut être fait selon une technique similaire. On commence par fixer un élément B de $\mathfrak{m}(\mathfrak{C})$, puis on pose

$$\mathfrak{D}'_B := \{A \in \mathfrak{m}(\mathfrak{C}) \ ; \ A \cap B \in \mathfrak{m}(\mathfrak{C})\}.$$

De la même façon, on vérifie que $\mathfrak{D}'_B$ est une classe monotone. La relation (*) montre que $\mathfrak{C} \subset \mathfrak{D}'_B$, et donc par minimalité que $\mathfrak{m}(\mathfrak{C}) \subset \mathfrak{D}'_B$; l'inclusion réciproque étant vérifiée par construction, on conclut que $\mathfrak{m}(\mathfrak{C}) = \mathfrak{D}'_B$. Enfin, ce le résultat étant vrai pour tout choix de $B \in \mathfrak{m}(\mathfrak{C})$, on a montré :

$$\forall A \in \mathfrak{m}(\mathfrak{C}) \quad \forall B \in \mathfrak{m}(\mathfrak{C}) \qquad A \cap B \in \mathfrak{m}(\mathfrak{C}). \tag{**}$$

Cette dernière relation permet de conclure que $\mathfrak{m}(\mathfrak{C})$ est la tribu $\sigma(\mathfrak{C})$. ∎

Avant de voir les applications — cruciales ! — de ce lemme, rappelons que

■ **Proposition B.15** *La tribu engendrée par le π-système $\mathcal{C} = \{]-\infty\,;\,a]\ ;\ a \in \mathbb{R}\}$ est la tribu $\mathcal{B}$ des boréliens de $\mathbb{R}$.*

Première application : le lemme des coalitions A73

Le premier résultat présenté stipule que, pour tester l'indépendance d'une famille de tribus, on peut restreindre la vérification de la formule du produit aux éléments de π-systèmes qui engendrent ces tribus.

■ **Lemme B.16 (π-systèmes et indépendance)** *Soit $(\Omega, \mathcal{T})$ un espace probabilisé. Soient $\mathcal{A}_1, \dots, \mathcal{A}_n$ des sous-tribus de $\mathcal{T}$. On suppose que chaque tribu $\mathcal{A}_k$ est engendrée par une classe $\mathcal{C}_k$ qui est un π-système. Si $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_n$ sont indépendantes, c'est-à-dire si, pour tout choix de $(C_1, \dots, C_n) \in \mathcal{C}_1 \times \cdots \times \mathcal{C}_n$, on a*

$$\mathbf{P}(C_1 \cap C_2 \cap \cdots \cap C_n) = \mathbf{P}(C_1)\,\mathbf{P}(C_2)\,\dots\,\mathbf{P}(C_n),$$

alors les tribus $\mathcal{A}_1, \dots, \mathcal{A}_n$ sont indépendantes.

DÉMONSTRATION : On met en œuvre la *méthode des ensembles adéquats.* ⏢B.2

L'indépendance de $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_n$ est équivalente à celle de $\mathcal{C}'_1, \mathcal{C}'_2, \dots, \mathcal{C}'_n$, où l'on a défini $\mathcal{C}'_k = \mathcal{C}_k \cup \{\Omega\}$; on peut donc supposer dans la suite que chaque $\mathcal{C}_k$ contient Ω.

On commence par fixer des événements $(C_2, \dots, C_n) \in \mathcal{C}_2 \times \cdots \times \mathcal{C}_n$; posons

$$\mathcal{M} = \{A_1 \in \mathcal{A}_1\ ;\ \mathbf{P}(A_1 \cap C_2 \cap \cdots \cap C_n) = \mathbf{P}(A_1)\,\mathbf{P}(C_2) \cdots \mathbf{P}(C_n)\},$$

ensembles des événements A_1 tels que $A_1, C_2, \dots, C_n$ sont indépendants. La classe $\mathcal{M}$ est une classe monotone qui, par hypothèse, contient tout $\mathcal{C}_1$; d'après le lemme de classe monotone, elle contient la tribu $\mathcal{A}_1$ engendrée par $\mathcal{C}_1$, c'est-à-dire que

$$\forall A_1 \in \mathcal{A}_1 \qquad \mathbf{P}(A_1 \cap C_1 \cap \cdots \cap C_n) = \mathbf{P}(A_1)\,\mathbf{P}(C_2) \cdots \mathbf{P}(C_n).$$

Le raisonnement est valable pour tout choix de $C_2, \dots, C_n$:

$$\forall A_1 \in \mathcal{A}_1 \quad \forall C_2, \dots, C_n \in \mathcal{C}_2 \times \mathcal{C}_n \qquad \mathbf{P}(A_1 \cap C_1 \cap \cdots \cap C_n) = \mathbf{P}(A_1)\,\mathbf{P}(C_2) \cdots \mathbf{P}(C_n),$$

ce qui prouve que $\mathcal{A}_1, \mathcal{C}_2, \dots, \mathcal{C}_n$ sont indépendantes.

On poursuit le raisonnement en fixant $A_1 \in \mathcal{A}_1$ et $C_3, \dots, C_n$ et en montrant de même que $\mathcal{A}_1, \mathcal{A}_2, \mathcal{C}_3, \dots, \mathcal{C}_n$ sont indépendants ; et ainsi de suite jusqu'au résultat final. ∎

Le second résultat formalise une propriété d'« associativité » de l'indépendance (si X, Y, Z, T sont indépendantes, alors toute fonction de X et Y est indépendante de toute fonction de Z et de T...)

■ **Lemme B.17 (des coalitions)**

i) Soient $\mathcal{A}_1, \dots, \mathcal{A}_n, \mathcal{A}_{n+1}, \dots, \mathcal{A}_{n+k}$ des tribus indépendantes. Alors les tribus

$$\mathcal{B} = \sigma(\mathcal{A}_1, \dots, \mathcal{A}_n) \qquad \text{et} \qquad \mathcal{B}' = \sigma(\mathcal{A}_{n+1}, \dots, \mathcal{A}_{n+k})$$

sont indépendantes.

ii) Si $X_1, X_2, \dots, X_n, X_{n+1}, \dots, X_{n+k}$ sont des variables aléatoires indépendantes, si $\varphi : \mathbb{R}^n \to \mathbb{R}$ et $\psi : \mathbb{R}^k \to \mathbb{R}$ sont des applications mesurables, alors les variables $Y = \varphi(X_1, \dots, X_n)$ et $Z = \psi(X_{n+1}, \dots, X_{n+k})$ sont indépendantes.

iii) Plus généralement, si $(X_n)_{n \geqslant 0}$ est une suite de variables aléatoires indépendantes, si l'on se donne une partition $\mathbb{N} = I \cup I'$ de $\mathbb{N}$, alors les tribus

$$\mathcal{B} = \sigma(X_k\ ;\ k \in I) \qquad \text{et} \qquad \mathcal{B}' = \sigma(X_k\ ;\ k \in I')$$

sont indépendantes.

DÉMONSTRATION : Notons $\mathfrak{C}$ la classe des parties de Ω de la forme $A_1 \cap \cdots \cap A_n$, pour $A_k \in \mathfrak{A}_k$. Alors la classe $\mathfrak{C}$ engendre $\sigma(\mathfrak{A}_1, \ldots, \mathfrak{A}_n) = \mathfrak{B}$. De même, notons $\mathfrak{C}'$ la classe des parties de la forme $A_{n+1} \cap \cdots \cap A_{n+k}$, classe qui engendre $\mathfrak{B}'$. Ces classes $\mathfrak{C}$ et $\mathfrak{C}'$ sont indépendantes, et il ne reste qu'à appliquer le lemme B.16.

De même, on applique ce lemme aux classes

$$\mathfrak{C} = \bigcup_{n=0}^{\infty} \sigma(X_k \,;\, k \in J, k \leqslant n) \qquad \text{et} \qquad \mathfrak{C}' = \bigcup_{n=0}^{\infty} \sigma(X_k \,;\, k \in J', k \leqslant n).$$

pour montrer le troisième point. Le deuxième en découle immédiatement. ∎

A74 **Deuxième application : rôle central de la fonction de répartition**

Le lemme de classe monotone, comme on l'a dit, permet notamment de montrer l'unicité d'une probabilité déterminée sur un π-système.

■ **Théorème B.18 (Unicité des mesures)** *Soient* $\mathbf{P}_1$ *et* $\mathbf{P}_2$ *deux probabilités sur un même espace probabilisable* $(\Omega, \mathfrak{T})$. *Soit* $\mathfrak{C}$ *un* π*-système engendrant* $\mathfrak{T}$, *tel que* $\mathbf{P}_1$ *et* $\mathbf{P}_2$ *coïncident sur* $\mathfrak{C}$. *Alors* $\mathbf{P}_1 = \mathbf{P}_2$.

DÉMONSTRATION : Notons $\mathfrak{M}$ l'ensemble des événements A de $\mathfrak{T}$ tels que $\mathbf{P}_1(A) = \mathbf{P}_2(A)$. L'ensemble $\mathfrak{M}$ est une classe monotone (voir exemple B.12) ; il contient $\mathfrak{C}$: il est donc égal à la tribu $\mathfrak{T}$ tout entière. ∎

■ **Corollaire B.19 (La fonction de répartition caractérise la loi)** *Deux probabilités sur l'espace probabilisable* $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ *sont égales si et seulement si elles ont même fonction de répartition.*

DÉMONSTRATION : Soient $\mathbf{P}$ et $\mathbf{P}'$ deux probabilités sur $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, ayant même fonction de répartition.

Notons $\mathfrak{C}$ l'ensemble des intervalles de la forme $]-\infty\,;\,a]$, pour a parcourant $\mathbb{R}$. Alors $\mathfrak{C}$ est un π-système, et les probabilités $\mathbf{P}$ et $\mathbf{P}'$ coïncident sur $\mathfrak{C}$; elles coïncident donc sur la tribu engendrée par $\mathfrak{C}$, qui est exactement la tribu $\mathfrak{B}(\mathbb{R})$ des boréliens. ∎

■ **Corollaire B.20** *Deux variables aléatoires réelles* X *et* Y *ont même loi si et seulement si elles ont même fonction de répartition.*

La fonction de répartition permet également de caractériser l'indépendance de deux variables aléatoires : X et Y sont indépendantes si et seulement si leur fonction de répartition conjointe $F_{X,Y}$ est le produit tensoriel $F_X \otimes F_Y$ des fonctions de répartition marginales.

■ **Corollaire B.21 (Les fonctions de répartition caractérisent l'indépendance)** *Soient* X *et* Y *deux variables aléatoires réelles, définies sur le même espace* $(\Omega, \mathfrak{T}, \mathbf{P})$. *Alors* X *et* Y *sont indépendantes si et seulement si, pour tout couple* $(x, y) \in \mathbb{R}^2$, *on a*

$$\mathbf{P}\big(\{X \leqslant x\} \cap \{Y \leqslant y\}\big) = \mathbf{P}\{X \leqslant x\} \cdot \mathbf{P}\{Y \leqslant y\},$$

ou encore

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y).$$

■ **Corollaire B.22 (Densité et calcul des probabilités)** *Si* X *est une variable aléatoire de fonction de répartition* F, *et admettant une densité* f, *c'est-à-dire que*

$$\forall x \in \mathbb{R} \qquad F(x) = \int_{-\infty}^{x} f(t)\,dt,$$

alors, pour tout borélien $A \in \mathfrak{B}(\mathbb{R})$, *on a*

$$\mathbf{P}\{X \in A\} = \int_A f(t)\,dt.$$

DÉMONSTRATION : La mesure de probabilité $\mathbf{P}_X(A) = \mathbf{P}\{X \in A\}$ et celle définie par $\mathbf{Q}(A) = \int_A f$ coïncident sur le π-système $\mathcal{C}$ des intervalles de la forme $]-\infty\,;\,x]$; par conséquent, elles coïncident sur $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$. ∎

Le théorème d'extension de Carathéodory A75

Le lemme de classe monotone permettait de prouver que, si deux probabilités coïncident sur un même π-système, alors elles sont égales. Mais il ne prouve pas, par exemple, que de telles probabilités existent.

Au contraire, le théorème de Carathéodory permet de conclure à l'existence d'une probabilité sur une tribu lorsqu'on impose ses valeurs sur une algèbre et que l'on réussit à prouver qu'elle est σ-additive sur cette algèbre. Nous l'admettons ici[8].

■ **Théorème B.23 (Carathéodory)** *Soit Ω un ensemble, muni d'une algèbre $\mathfrak{A}$ de parties de Ω. On note $\mathcal{T} = \sigma(\mathfrak{A})$ la tribu engendrée par $\mathfrak{A}$. Soit $\mathbf{P}_0$ une mesure σ-additive sur $(\Omega, \mathfrak{A})$. Alors, il existe une unique mesure $\mathbf{P}$ sur $(\Omega, \mathcal{T})$ dont la restriction à $\mathfrak{A}$ soit égale à $\mathbf{P}_0$, c'est-à-dire telle que*

$$\forall A \in \mathfrak{A} \qquad \mathbf{P}(A) = \mathbf{P}_0(A).$$

B.5 LE THÉORÈME DE RADON-NIKODÝM

Mesure absolument continue par rapport à une autre A76

Soient $\mathbf{P}$ et $\mathbf{Q}$ deux mesures finies[9]. On dit que $\mathbf{Q}$ est **absolument continue** par rapport à $\mathbf{P}$ si, pour tout événement $A \in \mathcal{T}$, $\mathbf{P}(A) = 0$ implique $\mathbf{Q}(A) = 0$.

Exemple B.24 On note $\mathbf{P}$ la probabilité uniforme sur $[0\,;\,1]$ — c'est-à-dire la mesure de Lebesgue sur la tribu des boréliens de $[0\,;\,1]$. Si $f : [0\,;\,1] \to \mathbb{R}$ est une densité, on note $\mathbf{Q}(A) = \int_A f\,d\mathbf{P}$. Alors $\mathbf{Q}$ est absolument continue par rapport à $\mathbf{P}$.

Le théorème de Radon-Nikodým A77

S'il existe une fonction $\mathcal{T}$-mesurable $f : \Omega \to \mathbb{R}^+$ telle que pour tout élément A de $\mathcal{T}$ on ait $\mathbf{Q}(A) = \int_A f\,d\mathbf{P}$, alors $\mathbf{Q}$ est absolument continue par rapport à $\mathbf{P}$. La théorème de Radon-Nikodým affirme que la réciproque est également vraie.

■ **Théorème B.25 (Radon-Nikodým)** *Soient $\mathbf{P}$ et $\mathbf{Q}$ deux mesures de probabilité, ou plus généralement deux mesures σ-finies, sur un espace mesurable $(\Omega, \mathcal{T})$. Si $\mathbf{Q}$ est absolument continue par rapport à $\mathbf{P}$, alors il existe une fonction $f : \Omega \to \mathbb{R}$, $\mathcal{T}$-mesurable telle que*

$$\forall A \in \mathcal{T} \qquad \mathbf{Q}(A) = \int_A f\,d\mathbf{P}.$$

La fonction f est appelée **densité de $\mathbf{Q}$** par rapport à $\mathbf{P}$. On la note parfois $f = d\mathbf{Q}/d\mathbf{P}$. Cette densité est unique au sens où, si g est encore une densité, alors $\mathbf{P}\{f \neq g\} = 0$.

---

(8). C'est un classique de la théorie de la mesure, que le lecteur pourra trouver par exemple dans Billingsley [9, section 3]. D'autres démonstrations existent, notamment à base de martingales [100].

(9). Ou, plus généralement, σ-finies. Une mesure sur $(\Omega, \mathcal{T})$ est dite **σ-finie** s'il existe une famille dénombrable $\{A_k\,;\,k \in K\}$ d'événements de mesure finie telle que $\Omega = \bigcup_{k\in K} A_k$. La mesure de Lebesgue sur $\mathbb{R}$ est σ-finie.

EXERCICES

♦ ***Exercice B.1 (Deux cas importants de boréliens)*** Soit $(f_n)_n$ une suite d'applications continues de $\mathbb{R}$ dans $\mathbb{R}$. Montrer que les ensembles

$$Z = \left\{x \in \mathbb{R} \,;\, \lim_{n\to\infty} f_n(x) = 0\right\} \qquad \text{et} \qquad C = \left\{x \in \mathbb{R} \,;\, \left(f_n(x)\right)_{n\in\mathbb{N}} \text{converge}\right\}$$

sont des boréliens de $\mathbb{R}$.

♦ ***Exercice B.2*** Soit $\mathfrak{C}$ une classe de parties de Ω. Soit $A \in \Omega$. Convenons de noter

$$\mathfrak{C} \cap A := \{C \cap A \,;\, C \in \mathfrak{C}\}.$$

Montrer que la tribu engendrée par $\mathfrak{C} \cap A$ est $\sigma(\mathfrak{C} \cap A) = \sigma(\mathfrak{C}) \cap A$.

♦ ***Exercice B.3 (Autre caractérisation des classes monotones)***
Soit $\mathfrak{C}$ une classe de parties de Ω. Montrer que $\mathfrak{C}$ est une classe monotone si et seulement si

- $\Omega \in \mathfrak{C}$;
- $\mathfrak{C}$ est stable par passage au complémentaire ;
- si $(A_n)_{n\geqslant 1}$ est une suite d'éléments disjoints de $\mathfrak{C}$, alors $\bigcup_n A_n \in \mathfrak{C}$.

Indications

◊ *Indication B.1 :* Vérifier que $A_k(\varepsilon) = \left\{x \in \mathbb{R} \,;\, \left|f_k(x)\right| < \varepsilon\right\}$ est un ouvert de $\mathbb{R}$, donc un borélien. Exprimer Z en terme d'unions et intersections dénombrables d'ensembles $A_k(\varepsilon)$. Pour l'ensemble C, utiliser le fait qu'une suite réelle est convergente si et seulement si elle est de Cauchy.

◊ *Indication B.2 :* Seule l'inclusion « $\supset$ » est non triviale. Définir, selon la méthode des ensembles adéquats, $\mathfrak{A} = \left\{B \in \sigma(\mathfrak{C}) \,;\, B \cap A \in \sigma(\mathfrak{C} \cap A)\right\}$, vérifier que c'est une tribu et en déduire que c'est $\sigma(\mathfrak{C})$. Conclure.

SOLUTIONS

♦ **Solution B.1** On rappelle qu'un réel x vérifie la propriété « $\lim_{n\to\infty} f_n(x) = 0$ » si et seulement si :

$$\forall \varepsilon > 0 \quad \exists n \in \mathbb{N} \quad \forall k \geqslant n \qquad \left|f_k(x)\right| < \varepsilon.$$

En ayant en vue la toute fin de la phrase mathématique précédente, on définit, pour tout $\varepsilon > 0$ et pour tout $k \in \mathbb{N}$: $A_k(\varepsilon) = \left\{x \in \mathbb{R} \,;\, \left|f_k(x)\right| < \varepsilon\right\} = f_k^{-1}\left(\left]\,\varepsilon\,;\,+\varepsilon\,\right[\right)$; en tant qu'image réciproque d'un ouvert par une application continue, c'est un ouvert de $\mathbb{R}$, et donc un borélien. Posons alors, pour tout $n \in \mathbb{N}$:

$$L_n(\varepsilon) = \bigcap_{k=n}^{\infty} A_k(\varepsilon),$$

qui est encore un borélien, puisque $\mathfrak{B}(\mathbb{R})$ est stable par intersection dénombrable. Quelques instants de réflexion montrent que cet ensemble est égal à

$$L_n(\varepsilon) = \left\{x \in \mathbb{R} \,;\, \forall k \geqslant n \quad \left|f_k(x)\right| < \varepsilon\right\}.$$

Pour traduire la partie « $\exists n \in \mathbb{N}$ », on pose

$$Z(\varepsilon) = \bigcup_{n=0}^{\infty} L_n(\varepsilon) = \left\{x \in \mathbb{R} \,;\, \exists n \in \mathbb{N} \quad \forall k \geqslant n \quad \left|f_k(x)\right| < \varepsilon\right\}.$$

C'est encore un borélien (union dénombrable de boréliens). Enfin, pour traduire le « $\forall \varepsilon > 0$ », on peut écrire

$$Z = \bigcap_{\varepsilon > 0} Z(\varepsilon),$$

mais cette écriture ne permet pas de conclure (la tribu des boréliens n'est pas stable par intersection non dénombrable) ; en revanche, on a également

$$Z = \bigcap_{m=0}^{\infty} Z\left(\frac{1}{2^m}\right)$$

et cette écriture sous forme d'une intersection dénombrable permet d'affirmer que Z est bien un borélien. Pour résumer, on a écrit Z sous la forme

$$Z = \bigcap_{m=0}^{\infty} \bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k \left(\frac{1}{2^m} \right).$$

Quant à l'ensemble C, en suivant l'indication, on l'écrit

$$\begin{aligned} C &= \{x \in \mathbb{R} \,;\, (f_n(x))_{n\in\mathbb{N}} \text{ est de Cauchy}\} \\ &= \{x \in \mathbb{R} \,;\, \forall \varepsilon > 0 \quad \exists n \in \mathbb{N} \quad \forall p, q \geqslant n \qquad |f_p(x) - f_q(x)| < \varepsilon\}. \end{aligned}$$

Comme précédemment, on définit, pour tout $\varepsilon > 0$ et pour tous $p, q \in \mathbb{N}$:

$$A_{p,q}(\varepsilon) = \{x \in \mathbb{R} \,;\, |f_p(x) - f_q(x)| < \varepsilon\},$$

qui est un ouvert de $\mathbb{R}$ (image réciproque de $]-\varepsilon\,;+\varepsilon[$ par l'application continue $f_p - f_q$), et donc un borélien. On conclut alors avec les mêmes arguments que

$$C = \bigcap_{m=0}^{\infty} \bigcup_{n=0}^{\infty} \bigcap_{p,q\geqslant n} A_{p,q} \left(\frac{1}{2^m} \right)$$

est bien un borélien.

♦ **Solution B.2** Notons que $\sigma(\mathfrak{C}) \cap A$ est une tribu. Tout élément de $\mathfrak{C} \cap A$ est évidemment dans $\sigma(\mathfrak{C}) \cap A$, ce qui montre que $\sigma(\mathfrak{C} \cap A) \subset \sigma(\sigma(\mathfrak{C}) \cap A) = \sigma(\mathfrak{C}) \cap A$.

Définissons maintenant $\mathfrak{A} = \{B \in \sigma(\mathfrak{C}) \,;\, B \cap A \in \sigma(\mathfrak{C} \cap A)\}$; on vérifie immédiatement que c'est bien une tribu. De plus, on a les inclusions $\mathfrak{C} \subset \mathfrak{A} \subset \sigma(\mathfrak{C})$ dont on déduit, par croissance, que

$$\sigma(\mathfrak{C}) \subset \sigma(\mathfrak{A}) = \mathfrak{A} \subset \sigma(\mathfrak{C})$$

et donc égalité de toutes ces tribus. Cela signifie que, pour tout $B \in \sigma(\mathfrak{C})$, $B \in \mathfrak{A}$ et donc $B \cap A \in \sigma(\mathfrak{C} \cap A)$. C'est bien la seconde inclusion que l'on voulait montrer.

Conseils de lecture

Une des meilleures références concernant la théorie de la mesure et ses liens avec les probabilités est sans doute

- Patrick Billingsley, *Probability and Measure* [9].

Une approche synthétique et en français peut se trouver dans

- Jean Jacod et Philip Protter, *L'essentiel en théorie des probabilités* [53].

Annexe C

Théorèmes d'existence

En d'innombrables occasions, nous avons formulé des hypothèses du genre : « soit X *une variable aléatoire suivant telle loi », ou encore « soit* $(X_n)_{n\geqslant 0}$ *une suite de variables aléatoires indépendantes et de même loi ». Diantre ! A-t-on effectivement le droit de le faire ? Après avoir donné des exemples élémentaires de théorèmes d'existence, nous présentons le théorème de Kolmogorov qui les généralise et qui permet, notamment, de fabriquer des suites de variables aléatoires de lois conjointes imposées, ou bien de suites de variables aléatoires indépendantes de lois marginales imposées.*

C.1 QUELQUES EXEMPLES SIMPLES DE THÉORÈMES D'EXISTENCE

Existence d'une variable aléatoire de loi donnée A78

Le problème le plus simple que l'on se pose est : si l'on se donne une fonction $F : \mathbb{R} \to [0\,;1]$ convenable, existe-t-il *effectivement* une variable aléatoire qui admette F comme fonction de répartition ?

■ **Théorème C.1 (Théorème de réalisation)** *Soit* F *une fonction croissante, continue à droite et admettant les limites* 0 *et* 1 *en* $-\infty$ *et* $+\infty$ *respectivement. Alors il existe un espace probabilisé* $(\Omega, \mathfrak{T}, \mathbf{P})$ *et une variable aléatoire* $X : \Omega \to \mathbb{R}$ *tels que* F *soit la fonction de répartition de* X.

En d'autres termes, si l'on se donne une « loi de variable aléatoire », alors on peut réaliser cette loi au moyen d'une variable aléatoire.

Le démonstration est aisée dans son principe : on choisit pour espace $\Omega := \mathbb{R}$, et on définit la probabilité **P** par sa valeur sur les intervalles $]-\infty\,;\,x]$ par $\mathbf{P}\big(]-\infty\,;\,x]\big) := F(x)$. On choisit enfin comme tribu la tribu $\mathfrak{B}(\mathbb{R})$ des boréliens. Le fait d'avoir défini **P** sur les intervalles $]-\infty\,;\,x]$ permet d'étendre, de manière unique, **P** à toute la tribu engendrée par ces intervalles, à savoir justement la tribu $\mathfrak{B}(\mathbb{R})$ des boréliens[1]. Il ne reste plus qu'à définir la variable X comme l'identité :

$$\forall \omega \in \mathbb{R} \qquad X(\omega) = \omega.$$

(1). Rappelons qu'il s'agit essentiellement d'utiliser le théorème de Carathéodory, après avoir défini **P** sur l'algèbre $\mathfrak{A}$ des unions finies disjointes d'intervalles $]a\,;\,b]$, et après avoir prouvé que **P** ainsi étendue est σ-additive sur $\mathfrak{A}$.

Par construction, on a bien

$$\forall x \in \mathbb{R} \qquad \mathbf{P}\{X \leqslant x\} = \mathbf{P}\big(]-\infty\,;\,x]\big) = F(x).$$

L'idée essentielle à retenir est que l'on a pu modéliser une variable aléatoire — c'est-à-dire la donnée d'un espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ et une application mesurable $X : \Omega \to \mathbb{R}$ — non pas en se plaçant sur un espace naturel et en se démenant pour fabriquer X, mais au contraire en construisant une bonne probabilité sur $\mathbb{R}$ et en se contentant de la fonction X la plus simple, à savoir l'identité.

Remarque C.2 (Représentation selon Skorokhod) D'autres constructions sont possibles, notamment en choisissant $\Omega := [0\,;1]$ comme espace de départ, muni de la tribu des boréliens, et en posant

$$Y_n(\omega) := \inf\{x \in \mathbb{R}\,;\, F(x) \geqslant \omega\}$$

pour tout $\omega \in [0\,;1]$. Le théorème A.24 page 680 montre que Y admet la fonction F comme fonction de répartition (voire également § 370).

Le même type de raisonnement permet de construire un couple (X, Y) admettant pour fonction de répartition une fonction $F : \mathbb{R}^2 \to [0\,;1]$ croissante par rapport à chacune de ses variables, continue à droite[2] et telle que

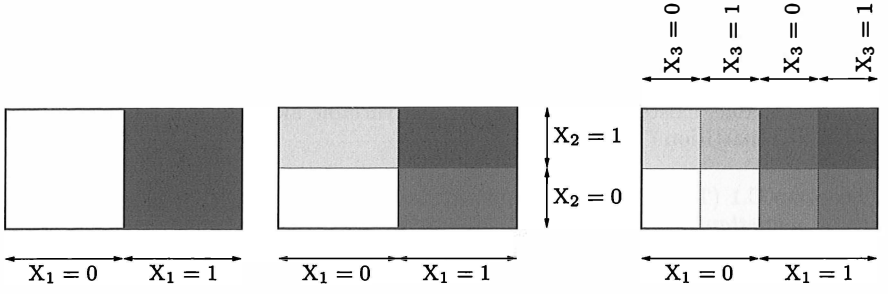
$$F(-\infty, y) = F(x, -\infty) = 0 \qquad F(+\infty, +\infty) = 1.$$

A79 **Un exemple de suite $(X_n)_{n\geqslant 0}$ de variables aléatoires indépendantes**

Le second type de problème que l'on peut se poser est l'existence de suites infinies de variables aléatoires indépendantes et identiquement distribuées, suites dont l'importance est évidemment primordiale[3].

Nous donnons ici un exemple de construction effective d'une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, de même loi de Bernoulli $\mathscr{B}(\frac{1}{2})$.

L'idée est la suivante. Sur un espace Ω, la première variable doit prendre les valeurs 0 et 1, avec une probabilité 1/2 pour chaque ; on obtient donc une partition de Ω en deux espaces, à savoir $\{X_1 = 0\}$ et $\{X_1 = 1\}$. Sur chacun de ces espaces, on permet à X_2 de prendre les deux valeurs 0 et 1 avec une probabilité 1/2 chacune ; ainsi, chaque partie de Ω est encore séparée en deux. On continue le procédé « à l'infini ».



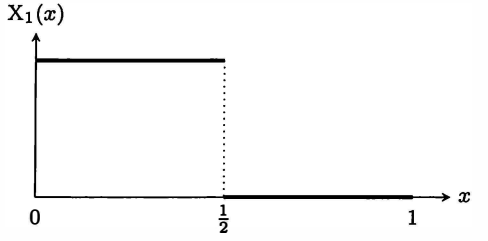
Comme on le voit, l'idée est simple mais elle oblige l'univers Ω à être « gros ». En effet, à l'étape n, on a obtenu une partition de Ω en 2^n parties. À la limite, on s'attend à ce que Ω soit de cardinal $2^{\aleph_0}$ (où $\aleph_0$ est le cardinal de $\mathbb{N}$), c'est-à-dire que Ω a la puissance du continu.

Pour mettre l'idée précédente en pratique, et construire effectivement la suite voulue, plusieurs approches sont possibles. Outre le choix « évident » $\Omega = \{0,1\}^{\mathbb{N}^*}$, muni de la tribu engendrée par les cylindres (§ 132 et A67), il peut être intéressant de choisir pour espace probabilisé l'ensemble $\Omega = [0\,;1]$ (qui a bien le cardinal du continu), muni de la tribu habituelle des boréliens et de la probabilité uniforme (mesure de Lebesgue sur $\mathfrak{B}([0\,;1])$). On pose ensuite $X_1 = \mathbf{1}_{[0\,;1/2]}$:

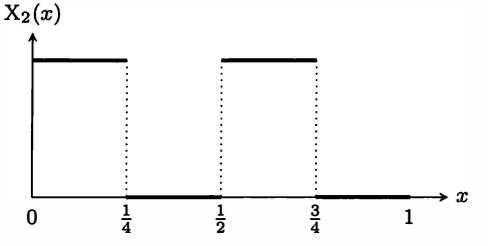
---

(2). Au sens où : si $(x_n)_{n\geqslant 1}$ décroît vers x et $(y_n)_{n\geqslant 1}$ décroît vers y, alors $F(x_n, y_n)$ tend vers $F(x, y)$.

(3). Ces suites sont apparues dès les travaux de Bernoulli et n'ont plus quitté le champ d'étude des mathématiciens.



On veut construire une variable aléatoire X_2 qui suive la même loi et soit indépendante de X_1. Il faut notamment que, si l'on se limite à l'intervalle $\left[0\,;\,\frac{1}{2}\right]$ sur lequel X_1 vaut 1, la fonction X_2 puisse prendre les valeurs 0 et 1 avec une probabilité $1/2$; et de même sur l'intervalle $\left]\frac{1}{2}\,;\,1\right]$. Le lecteur vérifiera sans peine que la fonction définie par le graphe suivant remplit les bonnes conditions, et définit une variable aléatoire de loi $\mathscr{B}(\frac{1}{2})$ et indépendante de X_1 :



On poursuit la construction, c'est-à-dire que l'on pose, pour tout $n \in \mathbb{N}$ et tout $x \in [0\,;\,1]$:

$$X_n(x) = \frac{1}{2}\left[1 + \operatorname{sgn}(\sin 2^n \pi x)\right]$$

où sgn est la fonction signe. Comme dans le cas des fonctions de Rademacher (§ 142), on vérifie explicitement que les fonctions X_n sont indépendantes, de même loi de Bernoulli $\mathscr{B}(\frac{1}{2})$.

Une suite de variables uniformes indépendantes A80

Un point essentiel de la construction qui précède est que l'on a effectué le parallèle :

$$\textit{une} \text{ variable uniforme} \iff (X_n)_{n\geqslant 0} \text{ v.a.i.i.d. de Bernoulli.}$$

Une autre manière de considérer la construction précédente passe par le développement en base 2 d'un réel de $[0\,;\,1]$ (voir paragraphe § 141).

Tous les ingrédients sont prêts pour créer une suite $(U_n)_{n\geqslant 1}$ de variables aléatoires indépendantes de même loi $\mathscr{U}\left([0\,;\,1]\right)$:

- on part d'une variable aléatoire U de loi uniforme ;
- on fabrique alors une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires de Bernoulli $\mathscr{B}(1/2)$ indépendantes, où X_n est la n-ième décimale dans la représentation de U en base 2 ;
- on choisit sa bijection préférée $\varphi : \mathbb{N}^{*2} \to \mathbb{N}^*$;
- on pose $Y_{i,j} := X_{\varphi(i,j)}$ pour tous $i, j \geqslant 1$;
- avec chacune des suites $(Y_{i,j})_{j\geqslant 1}$, on construit une variable

$$U_i = \sum_{j=1}^{\infty} \frac{Y_{i,j}}{2^j}.$$

Les $Y_{i,j}$ sont évidemment indépendantes, donc les U_i le sont, grâce au lemme des coalitions.

A81 **Une suite de variables indépendantes de lois quelconques données**
Une fois obtenue une suite $(U_n)_{n\geqslant 1}$ de variables aléatoires indépendantes de même loi uniforme, on peut fabriquer une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires indépendantes, telles que X_n suive une loi donnée, par exemple par sa fonction de répartition F_n. Pour cela, en utilisant la méthode présentée au paragraphe § 370, on note G_n le pseudo inverse de F_n et on pose $X_n := G_n(U_n)$. Les variables X_n sont bien indépendantes et X_n a pour fonction de répartition F_n.

C.2 LE THÉORÈME FONDAMENTAL DE KOLMOGOROV

A82 **Conditions de compatibilité et théorème d'extension**
Si l'on se donne une fonction de répartition F_1, nous savons en déduire un espace probabilisé $(\mathbb{R}, \mathfrak{B}, \mathbf{P}_1)$ où l'on a noté $\mathfrak{B} = \mathfrak{B}(\mathbb{R})$, tel que F_1 soit la fonction de répartition de X_1 :

Donnons-nous maintenant une fonction de répartition $F_2 : \mathbb{R}^2 \to \mathbb{R}$, qui soit **compatible** avec la première, c'est-à-dire que

$$\forall x_1 \in \overline{\mathbb{R}} \qquad F_2(x_1, +\infty) = F_1(x_1).$$

Alors, en suivant la même construction, on peut munir $(\mathbb{R}^2, \mathfrak{B}^2)$ d'une probabilité $\mathbf{P}_2$ dont F_2 est la fonction de répartition.

Continuons en nous donnant une fonction de répartition $F_3 : \mathbb{R}^2 \to \mathbb{R}$ compatible avec F_2 :

$$\forall x_1, x_2 \in \overline{\mathbb{R}} \qquad F_3(x_1, x_2, +\infty) = F_2(x_1, x_2).$$

Cela implique notamment que F_3 est compatible avec F_1 :

$$\forall x_1 \in \overline{\mathbb{R}} \qquad F_3(x_1, +\infty, +\infty) = F_1(x_1).$$

De la même manière, on munit $(\mathbb{R}^3, \mathfrak{B}^3)$ d'une probabilité $\mathbf{P}_3$ dont F_3 est la fonction de répartition. En généralisant, supposons donnée une famille $\{F_n \ ; \ n \in \mathbb{N}\}$ de fonctions de répartition, $F_n : \mathbb{R}^n \to [0\,;\,1]$, vérifiant les conditions de compatibilité

$$F_{n+p}(x_1, \ldots, x_n, +\infty, +\infty, \ldots, +\infty) = F_n(x_1, \ldots, x_n).$$

Peut-on alors munir l'espace $\Omega = \mathbb{R}^{\mathbb{N}^*}$ (qui est donc l'espace des suites réelles indexées par $\mathbb{N}^*$), muni d'une tribu $\mathfrak{B}^\infty$ convenable, d'une probabilité $\mathbf{P}$ qui soit représentée par ces fonctions de répartition ? « Représentée », au sens où, pour tout entier $k \geqslant 1$ et tous $a_1, \ldots, a_k \in \overline{\mathbb{R}}$

$$\mathbf{P}\big\{(x_n)_{n\geqslant 1} \ ; \ x_1 \leqslant a_1, \ldots, x_k \leqslant a_k\big\} = F_k(x_1, \ldots, x_k).$$

La réponse, positive, constitue le théorème d'extension de Kolmogorov

■ **Théorème C.3 (Théorème d'extension de Kolmogorov)**
Soit $(F_n)_{n\geqslant 1}$ une suite compatible de fonctions de distributions. Alors l'espace $\Omega = \mathbb{R}^{\mathbb{N}^*}$, peut être muni d'une tribu $\mathfrak{B}^\infty$ et d'une probabilité $\mathbf{P}$ telles que, pour tout entier k et tous $a_1, \ldots, a_k \in \overline{\mathbb{R}}$, l'ensemble

$$C_{a_1,\ldots,a_k} := \big\{(x_n)_{n\geqslant 1} \in \Omega \ ; \ x_1 \leqslant a_1, \ldots, x_k \leqslant a_k\big\}$$

est un événement et

$$\mathbf{P}\big\{(x_n)_{n\geqslant 1} \ ; \ x_1 \leqslant a_1, \ldots, x_k \leqslant a_k\big\} = F_n(x_1, \ldots, x_k).$$

DÉMONSTRATION : Nous ne donnons ici que les grandes lignes de la démonstration. On pose $\Omega := \mathbb{R}^{\mathbb{N}^*}$ et on définit pour tout entier $k \geqslant 1$ la projection $\pi_k : \Omega \to \mathbb{R}^k$ qui, à toute suite $\omega = (\omega_n)_{n\geqslant 1}$, associe

$$\pi_k(\omega) = (\omega_1, \omega_2, \ldots, \omega_k).$$

On munit de plus les espaces $\Omega_k := \mathbb{R}^k$ de leur tribu $\mathfrak{B}^k$ des boréliens. Si $A \in \mathfrak{B}^k$ est un événement de Ω_k, on définit le **cylindre de base A** comme

$$C_A = \pi_k^{-1}(A) = \{\omega \in \Omega\ ;\ \pi_k(\omega) \in A\}.$$

L'ensemble des cylindres forme une algèbre $\mathfrak{A}$ sur Ω. On note alors

$$\mathfrak{B}^\infty := \sigma(\mathfrak{A})$$

la tribu engendrée par tous les cylindres[(4)]. On définit ensuite sur $\mathfrak{A}$ une application $\mathbf{P}_0$ à valeurs dans $[0\,;\,1]$ par

$$\mathbf{P}_0\big(\pi_n^{-1}(A)\big) = \mathbf{P}_n(A), \qquad (*)$$

où $\mathbf{P}_n$ est l'unique probabilité sur $(\mathbb{R}^n, \mathfrak{B}^n)$ de fonction de répartition F_n.

C'est ici qu'interviennent les conditions de compatibilité. En effet, un événement de Ω_k peut définir le même cylindre qu'un événement de Ω_{k+p} ; par exemple, si B_1 et B_2 sont deux boréliens de $\mathbb{R}$, on peut identifier par exemple les événements $A = B_1 \times B_2 \in \mathfrak{B}^2$ et $A' = B_1 \times B_2 \times \mathbb{R} \in \mathfrak{B}(\mathbb{R}^3)$, au sens où

$$\pi_2^{-1}(A) = \pi_3^{-1}(A') = \{(x_n)_{n\geqslant 1}\ ;\ x_1 \in B_1,\ x_2 \in B_2\}.$$

La condition de compatibilité entre F_2 et F_3 assure que $\mathbf{P}_2(A) = \mathbf{P}_3(A')$. Cet exemple se généralise et montre que la définition $(*)$ est bien cohérente. On montre assez facilement que $\mathbf{P}_0$ est additive ; la seule difficulté est de montrer qu'elle est en fait σ-additive[(5)]. Le théorème de Carathéodory permet alors d'étendre (de manière unique !) la probabilité $\mathbf{P}_0$ sur $\mathfrak{A}$ en une probabilité $\mathbf{P}$ sur $\mathfrak{T}$. ∎

On peut se demander en quoi ce théorème est utile pour la vie de tous les jours ; mais le lecteur attentif aura d'ores et déjà compris qu'il permet de justifier l'existence de suites de variables aléatoires dont on se donne les fonctions de répartition conjointes :

■ **Corollaire C.4 (Existence de suites de variables aléatoires)**
Soit $(F_n)_{n\geqslant 1}$ une suite compatible de fonctions de distribution. Alors il existe un espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ et une suite $(X_n)_{n\geqslant 1}$ de variables aléatoires tels que, pour tout $k \geqslant 1$, la fonction de répartition du vecteur $(X_1, X_2, \ldots, X_k)$ soit F_k.

DÉMONSTRATION : On munit l'espace $(\Omega, \mathfrak{T}) = (\mathbb{R}^{\mathbb{N}^*}, \mathfrak{B}^\infty)$ de la probabilité $\mathbf{P}$ définie dans le théorème précédent. On pose alors, pour tout entier k et pour tout élément $\omega = (\omega_n)_{n\geqslant 1}$ de Ω :

$$X_k(\omega) := \omega_k.$$

Il est alors immédiat de vérifier que

$$\begin{aligned}
\mathbf{P}\{X_1 \leqslant a_1, \ldots, X_k \leqslant a_k\} &= \mathbf{P}_0\{X_1 \leqslant a_1, \ldots, X_k \leqslant a_k\} && \textit{élément de } \mathfrak{A}\\
&= \mathbf{P}\{\omega_1 \leqslant a_1, \ldots, \omega_k \leqslant a_k\} && \textit{définition } (*)\\
&= F_k(a_1, a_2, \ldots, a_k) && \textit{par construction de } \mathbf{P}
\end{aligned}$$

quels que soient l'entier k et les réels $a_1, a_2, \ldots, a_k$ choisis. ∎

■ **Corollaire C.5 (Existence de suites de variables aléatoires indépendantes)**
Soit $(F_n)_{n\geqslant 1}$ une suite de fonctions de répartition. Alors il existe un espace probabilisé $(\Omega, \mathfrak{T}, \mathbf{P})$ et des variables $(X_n)_n$ indépendantes tels que, pour tout $k \geqslant 1$, F_k soit la fonction de répartition de X_k.

DÉMONSTRATION : Il suffit de poser $G_k = F_1 \otimes F_2 \otimes \cdots \otimes F_k$, c'est-à-dire

$$G_k(x_1, x_2, \ldots, x_k) = F_1(x_1)\, F_2(x_2) \cdots F_k(x_k)$$

pour tout $k \geqslant 1$, et d'appliquer le théorème précédent avec la famille des G_k, qui vérifient bien les conditions de compatibilité :

$$G_{n+p}(x_1, x_2, \ldots, x_n, +\infty, \ldots, +\infty) = F_1(x_1) \cdots F_n(x_n) \cdot \underbrace{F_{n+1}(+\infty) \cdots F_{n+p}(+\infty)}_{1}$$

$$= G_n(x_1, x_2, \ldots, x_n).$$

La relation $G_k = F_1 \otimes F_2 \otimes \cdots \otimes F_k$ assure que les variables construites sont bien indépendantes. ∎

(4). C'est donc également la plus petite tribu rendant mesurables les projections π_n.

(5). Une démonstration élégante peut être trouvé par exemple dans l'ouvrage d'Alfred Rényi, *Probability theory* [87] ; elle généralise la démonstration du paragraphe A71.

Un point évidemment capital de la théorie axiomatique de Kolmogorov est de montrer qu'il existe des suites $(X_n)_{n\geqslant 1}$ de variables aléatoires mutuellement indépendantes identiquement distribuées, c'est-à-dire des « répliques » indépendantes d'une variable aléatoire X, dont on a fait usage dès que se sont posées les questions de lois limites, et notamment à l'époque des Bernoulli, de Pascal, de Laplace et de Gauss. Encore une fois, le formalisme mathématique précis et rigoureux a suivi, de plusieurs siècles, l'utilisation d'un objet mathématique : l'action prime avant tout. De même, les physiciens manipulent des intégrales de Feynman depuis plus d'un demi siècle, intégrale dont les fondements rigoureux ne sont pas encore achevés.

Cependant, l'étape d'achèvement rigoureux, fût-elle tardive, est primordiale, sous peine de jeter le discrédit à l'ensemble de la construction. Le lecteur angoissé à l'idée de manipuler des objets qui, peut-être, n'existent pas — comme c'est arrivé dans l'histoire des mathématiques — aura ressenti un intense soulagement à la lecture de ces théorèmes.

■ **Corollaire C.6 (Existence d'une suite de répliques de X)**
Soit X *une variable aléatoire. Alors il existe une suite* $(X_n)_{n\geqslant 1}$ *de variables aléatoires indépendantes, et de même loi que* X.

Terminons en signalant que le théorème fondamental de Kolmogorov admet une généralisation à une infinité non dénombrable de contraintes, permettant notamment d'établir l'existence de processus stochastiques indicés par un temps continu. Nous renvoyons le lecteur intéressé à Billingsley [9].

Annexe D

Tables

D.1 LANGAGE ENSEMBLISTE ET LANGAGE PROBABILISTE

Notation	Langage ensembliste	Langage probabiliste
ω ($\omega \in \Omega$)	élément	événement atomique
A (A $\subset \Omega$)	partie de Ω	événement
$\omega \in$ A	ω est élément de A	A est réalisé
A $\subset$ B	A inclus dans B	l'événement A implique l'événement B
A $\cup$ B	réunion	réalisation de A *ou* de B
A $\cap$ B	intersection	réalisation simultanée de A *et* de B
A^c ou $\Omega \setminus$ A	complémentaire	non-A (événement contraire)
A $\setminus$ B	A privé de B	réalisation de A mais pas de B
$\varnothing$	partie vide	événement impossible
Ω		événement certain
A $\cap$ B = $\varnothing$	parties disjointes	A et B sont des événements incompatibles



A $\cap$ B A $\setminus$ B A $-$ B A $\cup$ B A $+$ B

$\bigcup_{n=0}^{\infty} A_n$	union de $(A_n)_{n\geqslant 0}$	réalisation d'au moins un des événements A_n
$\bigcap_{n=0}^{\infty} A_n$	intersection de $(A_n)_{n\geqslant 0}$	réalisation de tous les événements A_n
$A_n \uparrow A$	la suite $(A_n)_{n\geqslant 0}$ est croissante et $A = \bigcup_{n=0}^{\infty} A_n = \lim_n \uparrow A_n$	
$A_n \downarrow A$	la suite $(A_n)_{n\geqslant 0}$ est décroissante et $A = \bigcap_{n=0}^{\infty} A_n = \lim_n \downarrow A_n$	
$\limsup A_n$	$\bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{A_n \text{ ; i.s.}\}$	réalisation d'une infinité d'événements parmi $A_0, A_1, \ldots$
$\liminf A_n$	$\bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k$	réalisation de tous les événements $A_0, A_1, \ldots$, sauf peut-être un nombre fini d'entre eux (= réalisation de tous les événements à partir d'un certain rang)

D.2 PROBABILITÉS

A83 **Axiomes**

Un **espace probabilisé** est un triplet $(\Omega, \mathfrak{T}, \mathbf{P})$ où Ω est un ensemble quelconque, $\mathfrak{T}$ une tribu sur Ω et $\mathbf{P}$ une probabilité sur $(\Omega, \mathfrak{T})$, c'est-à-dire une application $\mathbf{P} : \mathfrak{T} \to \mathbb{R}$ vérifiant les propriétés suivantes :

- ⋆ **Axiome 1 :** $\mathbf{P}$ est à valeurs positives ;
- ⋆ **Axiome 2 :** $\mathbf{P}(\Omega) = 1$;
- ⋆ **Axiome 3 :** $\mathbf{P}$ est **σ-additive** : pour toute suite $(A_n)_{n\geqslant 0}$ d'éléments deux à deux disjoints de $\mathfrak{T}$,

$$\mathbf{P}\Big(\sum_{n=0}^{\infty} A_n\Big) = \sum_{n=0}^{\infty} \mathbf{P}(A_n).$$

Sous l'hypothèse des axiomes 1 et 2 et de la propriété d'additivité ($\mathbf{P}(A+B) = \mathbf{P}(A) + \mathbf{P}(B)$), l'axiome 3 est équivalent à chacun des axiomes 3′, 3″ et 3‴ résumés ci-dessous :

Axiome 3	σ-additivité	A_n disjoints	alors	$\mathbf{P}\Big(\sum_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbf{P}(A_n)$
Axiome 3′	continuité monotone $\uparrow$	si $B_n \uparrow B$,	alors	$\mathbf{P}(B_n) \uparrow \mathbf{P}(B)$
Axiome 3″	continuité monotone $\downarrow$	si $C_n \downarrow C$,	alors	$\mathbf{P}(C_n) \downarrow \mathbf{P}(C)$
Axiome 3‴	continuité monotone en $\varnothing$	si $D_n \downarrow \varnothing$,	alors	$\mathbf{P}(D_n) \downarrow 0$

A84 **Théorèmes d'intégration**

Nom	Hypothèses	Résultat
Convergence monotone	$0 \leqslant X_n \uparrow X$	$\mathbf{E}(X_n) \uparrow \mathbf{E}(X)$
Séries positives	$X_n \geqslant 0$	$\mathbf{E}\Big(\sum_{n=1}^{\infty} X_n\Big) = \sum_{n=1}^{\infty} \mathbf{E}(X_n)$
Lemme de Fatou	$0 \leqslant X_n \to X$	$\mathbf{E}(X) \leqslant \liminf \mathbf{E}(X_n)$
Convergence dominée	$X_n \to X$, $\lvert X_n\rvert \leqslant Y$, avec $Y \in L^1$	$\mathbf{E}\lvert X_n - X\rvert \to 0$ et $\mathbf{E}(X_n) \to \mathbf{E}(X)$
Intégration terme à terme	$\sum \mathbf{E}\lvert X_n\rvert$ converge	$\mathbf{E}\Big(\sum_{n=1}^{\infty} X_n\Big) = \sum_{n=1}^{\infty} \mathbf{E}(X_n)$
Cvce monotone conditionnelle	$0 \leqslant Y_n \uparrow Y$	$\mathbf{E}(Y_n \,\|\, \mathfrak{G}) \to \mathbf{E}(Y \,\|\, \mathfrak{G})$
Cvce dominée conditionnelle	$Y_n \to Y$, $\lvert Y_n\rvert \leqslant T \in L^1$	$\mathbf{E}(Y_n \,\|\, \mathfrak{G}) \to \mathbf{E}(Y \,\|\, \mathfrak{G})$

$$L^1 \subset L^2 \subset \cdots \subset L^p \subset L^{p+1} \subset \cdots \qquad \|X\|_1 \leqslant \|X\|_2 \leqslant \cdots \leqslant \|X\|_p \leqslant \|X\|_{p+1} \leqslant \cdots$$

A85 **Formules utiles**

$$\sum_{n=0}^{\infty} n\,x^n = \frac{x}{(1-x)^2} \qquad \sum_{n=0}^{\infty} n^2\,x^n = \frac{x^2+x}{(1-x)^3} \tag{D.1}$$

$$\int_0^{+\infty} \mathrm{e}^{-\alpha t}\,\mathrm{d}t = \frac{1}{\alpha} \qquad \int_0^{+\infty} t^n\,\mathrm{e}^{-\alpha t}\,\mathrm{d}t = \frac{n!}{\alpha^{n+1}}$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{-t^2/2}\,\mathrm{d}t = 1 \qquad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^{2n}\,\mathrm{e}^{-t^2/2}\,\mathrm{d}t = \frac{(2n)!}{2^n\,n!}$$

$$\mathbf{E}(X) = \sum_{n=0}^{\infty} \mathbf{P}\{X > n\} \qquad \mathbf{E}(X) = \int_0^{+\infty} \big(1 - F(x)\big)\,\mathrm{d}x$$

si $X(\Omega) \subset \mathbb{N}$ — si $X \geqslant 0$ à densité

Lois discrètes

Nom	$X(\Omega)$	$\mathbf{P}\{X=k\}$	$\mathbf{E}(X)$	$\mathbf{V}(X)$	$G(t)$	$\varphi(t)$
Uniforme	$[1\,;n]$	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\frac{t-t^{n+1}}{n(1-t)}$	$\frac{e^{it}-e^{i(n+1)t}}{n(1-e^{it})}$
Bernoulli $\mathscr{B}(p)$	$\{0,1\}$	$\begin{cases}p_1=p\\p_0=1-p\end{cases}$	p	$p(1-p)$	$1-p+pt$	$1-p+pe^{it}$
Binomiale $\mathscr{B}(n\,;\,p)$	$[0\,;n]$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$	$(1-p+pt)^n$	$(1-p+pe^{it})^n$
Géométrique $\mathscr{G}(p)$	$\mathbb{N}^*$	$(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pt}{1-(1-p)t}$	$\frac{pe^{it}}{1-(1-p)e^{it}}$
Poisson $\mathscr{P}(\lambda)$	$\mathbb{N}$	$e^{-\lambda}\frac{\lambda^k}{k!}$	λ	λ	$e^{\lambda(t-1)}$	$e^{\lambda(e^{it}-1)}$
Hypergéométrique $\mathscr{H}(N,n,p)$		$\frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}}$	np	$npq\frac{N-n}{N-1}$		

Lois continues

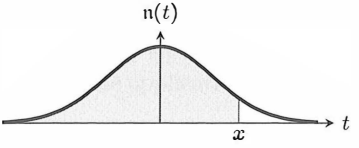
Nom	$X(\Omega)$	Densité $f(x)$	$\mathbf{E}(X)$	$\mathbf{V}(X)$	$\varphi(t)$
Uniforme $\mathscr{U}([a\,;b])$	$[a\,;b]$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{1}{b-a}\frac{e^{itb}-e^{ita}}{it}$
Normale $\mathscr{N}(m\,;\,\sigma^2)$	$\mathbb{R}$	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}$	m	σ^2	$e^{imt}e^{-\sigma^2t^2/2}$
Exponentielle $\mathscr{E}(\alpha)$	$\mathbb{R}^+$	$\alpha e^{-\alpha x}$	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$	$\frac{1}{1-it/\alpha}$
Cauchy $\mathscr{C}(m\,;\,a)$	$\mathbb{R}$	$\frac{a}{\pi}\frac{1}{a^2+(x-m)^2}$	$\nearrow$	$\nearrow$	$e^{imt-a\lvert t\rvert}$
χ^2_n	$\mathbb{R}^+$	$\frac{1}{2\Gamma(n/2)}\left(\frac{x}{2}\right)^{\frac{n}{2}-1}e^{-x/2}$	n	$2n$	$(1-2it)^{-n/2}$
Beta $B(a,b)$	$]0\,;1[$	$\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	
Gamma $\Gamma(\alpha,\beta)$	$\mathbb{R}_+^*$	$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$\left(1-\frac{it}{\beta}\right)^{-\alpha}$

Lois approchées

Loi et paramètres	Loi approchée	Conditions pratiques
Hypergéométrique $\mathscr{H}(N,n,p)$	$\mathscr{B}(n\,;\,p)$	$N\geqslant 10\,n$
Binomiale $\mathscr{B}(n\,;\,p)$	$\mathscr{P}(np)$	$p\leqslant 0,1\quad n\geqslant 30$
Poisson $\mathscr{P}(\mu)$	$\mathscr{N}(\mu\,;\,\mu)$	$\mu\geqslant 15$
Binomiale $\mathscr{B}(n\,;\,p)$	$\mathscr{N}(np\,;\,np(1-p))$	$np\geqslant 15\quad n(1-p)\geqslant 15$

A87 Loi normale centrée réduite

$$\mathfrak{N}(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\, \mathrm{d}t.$$



Les décimales se lisent sur les lignes, et on ajoute les centièmes rangés en colonnes. Par exemple, la valeur de $\mathfrak{N}(1,93)$ est donnée à l'intersection de la ligne 1,9 et de la colonne 0,03, et l'on peut lire $\mathfrak{N}(1,93) = 0,9732$, à 10^{-4} près. Pour les valeurs négatives de x, on utilise la relation $\mathfrak{N}(-x) = 1 - \mathfrak{N}(x)$.

x	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7793	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8906	.8925	.8943	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9986	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

x	**3.7**	**3.8**	**3.9**	**4.0**	**4.5**	**5**
$1 - \mathfrak{N}(x)$	$1,08 \cdot 10^{-4}$	$7,23 \cdot 10^{-5}$	$4,81 \cdot 10^{-5}$	$3,17 \cdot 10^{-5}$	$3,40 \cdot 10^{-6}$	$2,87 \cdot 10^{-7}$

Annexe E

Lexique français-anglais

français	*anglais*
absolument continue	*absolutely continuous*
aléatoire	*random*
algèbre	*algebra, field*
atome	*atom*
biais	*bias*
borélien	*Borel set*
branchement (processus de)	*branching process*
classe monotone	*monotone class, d-system, λ-system, Dynkin system*
coefficient d'aplatissement	*kurtosis*
coefficient de corrélation	*(linear) correlation coefficient*
convergence en loi / étroite	*convergence in law/distribution*
convergence étroite	*weak convergence*
convergence en probabilité	*convergence in probability*
convergence presque sûre	*almost sure convergence*
covariance	*covariance*
dé équilibré	*fair dice*
degré de liberté (d.d.l.)	*degree of freedom (d.o.f.)*
dénombrable	*countable, enumerable*
densité de probabilité	*density function*
— conjointe, marginale	*joint, marginal —*
écart type	*standard deviation*
échantillon	*sample*
épreuve ω	*sample point*
équiprobables	*equally likely*
espace des épreuves	*sample space*
espérance	*expectation, expected value, mean value*
estimateur sans biais	*unbiased estimator*
estimateur convergent	*consistent estimator*
événement	*event*
fonction de répartition	*(cumulative) distribution function c.d.f.*
— conjointe, marginale	*joint, marginal —*
hypothèse nulle	*null hypothesis*
incompatibles	*mutually exclusive*

indépendantes (v.a.)	*independant random variables*
inégalité de Bienaymé-Tchebychev	*Chebychev inequality*
khi-deux (loi du χ^2)	*chi-square distribution*
lemme de classes monotones	*Dynkin's π-λ theorem*
	monotone class theorem
loi	*law, distribution*
— de l'arcsinus	*arcsine law*
— conditionnelle	*conditional distribution*
— conjointe	*joint distribution*
— des grands nombres	*law of large numbers*
— marginale	*marginal distribution*
— normale	*gaussian distribution*
— de Fisher	*F-distribution*
— de Student	*t-distribution*
— du tout ou rien	*zero-one law*
pile/face	*heads/tail*
marche aléatoire	*random walk*
maximum de vraisemblance	*maximum likelihood*
mesure	*measure*
mouvement brownien	*brownian motion*
pièce équilibrée	*unbiased coin*
presque partout	*almost everywhere*
presque sûr, sûrement (p.s.)	*almost sure, surely (a.s.)*
probabilité	*probability*
probabilité *a priori*	*prior probability*
probabilité conditionnelle	*conditional probability*
probabilité de B sachant A	*probability of B given A*
sous- / sur-martingale	*sub- / super-martingale*
stable par passage au complémentaire	*closed under complementation*
test uni-, bilatéral	*one-tailed, two-tailed test*
théorème central limite	*central limit theorem*
tirage de Bernoulli	*Bernoulli trial*
tirage à *pile* ou *face*	*coin toss*
tirage	*trial, drawing*
tribu	*σ-algebra, σ-field*
univers Ω	*sample space*
variable aléatoire	*random variable*
variance	*variance*
chaîne de Markov	*Markov chain*
état transitoire	*transient state*
état récurrent	*recurrent state*
états communicants	*communicating states*
loi stationnaire	*stationnary distribution*
chaîne de Markov fortement irréductible	*regular Markov chain*
classe, chaîne irréductible	*indecomposable class, chain*

Références

[1] Milton ABRAMOWITZ et Irene A. STEGUN. *Handbook of mathematical functions.* Dover, 1972.

[2] Walter APPEL. *Mathématiques pour la physique et les physiciens.* H&K, 4e édition, 2009.

[3] Nicolas BACAËR. *Histoires de mathématiques et de populations.* Cassini, 2008.

[4] Thomas BAYES. An essay towards solving a problem in the doctrine of chances. *Phil. Trans.*, **53**, p. 370–418, 1764.

[5] John BELL. *Speakable and unspeakable in Quantum Mechanics.* Cambridge University Pres, 1988.

[6] Michel BENAÏM et Nicole EL KAROUI. *Promenade aléatoire.* Les éditions de l'École polytechnique, 2004.

[7] Bernard BERCU et Djèlil CHAFAÏ. *Modélisation stochastique et simulation.* Dunod, 2007.

[8] Sergueï Natanovitch BERNSTEIN. Démonstration du théorème de Weierstrass fondé sur le calcul des probabilités. *Comm. Soc. Math. Kharkhov., série 2*, **13**, p. 1–2, 1913.

[9] Patrick BILLINGSLEY. *Probability and measure.* John Wiley & Sons, 3e édition, 1995.

[10] Ludwig BOLTZMANN. *Lectures on gas theory.* Dover, 1995. Traduction de *Vorlesungen über Gastheorie*, paru en deux volumes en 1896 et 1898.

[11] Émile BOREL. *Probabilité et certitude.* PUF — Que sais-je ? n° 445, 1969.

[12] BOULET. *Notes, tome 7 : Formicapunk.* Delcourt, 2012. (Ne pas hésiter à voir également le blog `www.bouletcorp.com`, si vous cliquez sur « Aléatoire », vous aurez — peut-être — une Note parlant de mathématiques.).

[13] Léon BRILLOUIN. *Les statistiques quantiques et leurs applications.* Les Presses Universitaires de France, 1930.

[14] Henri BROCH. *Au cœur de l'extraordinaire.* Collection « Zététique », 2002.

[15] Frank BURK. *A garden of integrals.* The mathematical association of America, 2007.

[16] Éric CHARPENTIER, Annick LESNE et Nikolaï NIKOLSKI. *L'héritage de Kolmogorov en mathématiques.* Belin, 2004.

[17] Thomas H. CORMEN, Charles E. LEISERSON, Ronald L. RIVEST et Clifford STEIN. *Introduction aux algorithmes.* Dunod, 2e édition, 2001.

[18] RAND CORPORATION, editor. *A Million Random Digits with 100,000 Normal Deviates.* Free Press, 1955.

[19] Michel CRITON. Les jeux de dés. *Gazette de la Société Mathématique de France*, **128**, p. 46–61, avril 2011.

[20] Paul DEHEUVELS. *La probabilité, le hasard et la certitude.* PUF, 4e édition, 2008. Collection *Que sais-je ?*

[21] Jean-Paul DELAHAYE. *Information, complexité et hasard.* Hermès Science Publications, 2e édition, 1999.

[22] Jean-François DELMAS et Benjamin JOURDAIN. *Modèles aléatoires.* Mathématiques & Applications, 57. Springer, 2006.

[23] Michel DEMAZURE. *Cours d'algèbre.* Cassini, 2008.

[24] Persi DIACONIS, Susan HOLMES et Richard MONTGOMERY. Dynamical bias in the coin toss. *SIAM Rev.*, **49**(2), p. 211–235, 2007.

[25] Bernard DIU, Claudine GUTHMANN, Danielle LEDERER et Bernard ROULET. *Éléments de physique statistique.* Hermann, 1996.

[26] François DRESS et Michel MENDÈS FRANCE. La suite des puissances de 3/2 : Grandes et petites énigmes mathématiques. *La Recherche*, **346**, p. 34–37, 2001.

[27] Bertrand DUPLANTIER. Two-dimensional copolymers and exact conformal multifractality. *Phys. Rev. Letter*, **82**, p. 880–883, 1999.

[28] Bertrand DUPLANTIER. Le mouvement brownien, "divers et ondoyant". In *Séminaire Poincaré*, volume 1, p. 155–212, 2005.

[29] Evgueni B. DYNKIN et V. A. USPENSKII. *Mathematical conversations.* Dover, 2006.

[30] Arthur ENGEL. *Processus aléatoires pour les débutants.* Cassini, 2011.

[31] Paul ERDŐS et Alfred RÉNYI. On a classical problem of probability theory. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, **6**, p. 215–220, 1961.

[32] K. Bruce ERICKSON. The strong law of large numbers when the mean is undefined. *Transactions of the American Mathematical Society*, **185**, 1973.

[33] William FELLER. *An introduction to probability theory and its applications, Vol I.* Wiley, 1968.

[34] Dominique FOATA et Aimé FUCHS. *Processus stochastiques.* Dunod, 2002.

[35] Patrick X. GALLAGHER. On the distribution of primes in short intervals. *Mathematika*, **23**, p. 4–9, 1976.

[36] Claude GASQUET et Patrick WITOMSKI. *Analyse de Fourier et applications.* Dunod, 2000.

[37] Nicolas GAUVRIT et Jean-Paul DELAHAYE. Pourquoi la loi de Benford n'est pas mystérieuse. *Mathématiques et sciences humaines*, **182-2**, p. 7–15, 2008. (http://msh.revues.org/10363).

[38] Josiah Willard GIBBS. *Principes élémentaires de mécanique statistique.* Gauthier-Villars, 1925.

[39] Xavier GOURDON. http://numbers.computation.free.fr/Constants/constants.html.

[40] Carl GRAHAM et Denis TALAY. *Simulation stochastique et méthodes de Monte-Carlo.* Les éditions de l'École polytechnique, 2011.

[41] Andrew GRANVILLE. Harald Cramér and the distribution of prime numbers. *Scandinavian Actuarial Journal*, **1**, p. 12–28, 1995.

[42] Andrew GRANVILLE et K. SOUNDARARAJAN. Sieving and the Erdős-Kac Theorem. In Andrew GRANVILLE et Zeév RUDNICK, editors, *Proc. NATO Adv. Study Inst. on Equidistribution in Number Theory*, volume Sci Ser. II, Math. Phys. Chem. 237, p. 15–27. Springer, 2006.

[43] Robert E. GREENE et Steven G. KRANTZ. *Function theory of one complex variable.* Wiley Interscience, 1997.

[44] Geoffrey GRIMMETT et David STIRZAKER. *Probability and Random Processes.* Oxford, 2009.

[45] Godfrey H. HARDY et Edward M. WRIGHT. *An introduction to the theory of numbers.* Oxford University Press, 2008. Réédition de l'ouvrage de 1938.

[46] Adam J. HARPER. Two new proofs of the Erdős-Kac theorem, with bound on the rate of convergence, by Stein's method for distributional approximations. *Math. Proc. Cam. Phil. Soc.*, **147**, p. 95–114, 2009.

[47] Edwin HEWITT et Leonard J. SAVAGE. Symmetric measures on cartesian products. *Transactions of the American Mathematical Society*, **80**, p. 470–501, 1955.

[48] Theodore P. HILL. Base-invariance implies benford's law. *Proceedings of the American Mathematical Society*, **123**(3), p. 887–895, 1995.

[49] Marc HINDRY. *Arithmétique.* Calvage & Mounet, 2008.

[50] Kerson HUANG. *Statistical Mechanics.* John Wiley & sons, 1963.

[51] Claude ITZYKSON et Jean-Michel DROUFFE. *Théorie statistique des champs.* EDP Sciences, 1999.

[52] Claude ITZYKSON et Jean-Bernard ZUBER. *Quantum field theory.* Mc-Graw Hill, 1980.

[53] Jean JACOD et Philip PROTTER. *L'essentiel en théorie des probabilités.* Cassini, 2003.

[54] Benjamin JOURDAIN. *Probabilités et statistique.* Ellipses, 2009.

[55] Ioannis KARATZAS et Steven E. SHREVE. *Brownian motion and stochastic calculus.* Graduate texts in Mathematics, 113. Springer-Verlag, 2000.

[56] Nicholas KATZ et Peter SARNAK. Zeroes of zeta functions and symmetry. *Bull. A. M. S.*, **36**(1), p. 1–26, January 1999.

[57] Alexander S. KECHRIS. *Classical descriptive set theory.* Springer-Verlag, 1995.

[58] Daniel KEHLMANN. *La Nuit de l'Illusioniste.* Actes Sud, 2010. Traduction Juliette Aubert.

[59] Alexandre Iakovlevitch KHINTCHINE. Sur la loi des grands nombres. *Comptes rendus de l'Académie des Sciences*, **189**, p. 477–479, 1929.

[60] Donald E. KNUTH. *The Art of Computer Programming, vol. II : Seminumerical Algorithms.* Addison-Wesley, troisième édition, 1998.

[61] Andreï Nikolaiévitch KOLMOGOROV. Über die analytischen Methoden in der Wahrscheinlichkeitrechnung. *Math. Ann.*, **104**, p. 415–458, 1931.

[62] Andreï Nikolaiévitch KOLMOGOROV. *Grundbegriffe der Wahrscheinlichkeitrechnung.* Springer, 1933. Traduit en anglais sous le titre *Foundations of the theory of probability*, Chelsea Publishing Compagny, 1950 et 1956.

[63] Leonid B. KORALOV et Yakov G. SINAI. *Theory of probability and random processes.* Universitext. Springer, deuxième édition, 2007.

[64] Emmanuel KOWALSKI. Bernstein polynomials and Brownian motion. *The American Mathematical Monthly*, **113**(10), p. 865–886, décembre 2006.

[65] Emmanuel KOWALSKI. Averages of Euler products, distribution of singular series and the ubiquity of Poisson distribution. *Acta Arithmetica*, **148**(2), p. 153–187, 2011.

[66] Emmanuel KOWALSKI et Ashkan NIKEGHBALI. Mod-Poisson convergence in probability and number theory. *International Mathematics Research Notices*, 2010. doi: 10.1093/imrn/rnq019.

[67] Serge LANG. *Complex analysis.* Springer Verlag, 1995.

[68] Pierre-Simon de LAPLACE. *Théorie analytique des Probabilités.* Mme Veuve Courcier, Imprimeur Libraire pour les mathématiques, 1812.

[69] Gregory F. LAWLER, Oded SCHRAMM et Wendelin WERNER. The dimension of the planar Brownian frontier is 4/3. *Math. Res. Lett.*, **8**, p. 401–411, 2001.

[70] Gregory F. LAWLER et Wendelin WERNER. Intersection exponents for planar Brownian motion. *Ann. Prob.*, **27**(4), p. 1601–1642, 1999.

[71] Emmanuel LESIGNE. *Pile ou Face, une introduction aux théorèmes limites du Calcul des Probabilités.* Opuscules. Ellipses, 2001.

[72] Benoît MANDELBROT. *The Fractal Geometry of Nature.* Freeman, 1982.

[73] Roger MANSUY et Rached MNEIMNÉ. *Algèbre linéaire, réduction des endomorphismes.* Vuibert, 2012.

[74] Philippe MARTIN. *Une initiation à l'intégrale fonctionnelle en physique quantique et statistique.* Cahiers de physique. Presses polytechniques et universitaires romandes, 1996.

[75] Jean MEEUS. *Calculs astronomiques à l'usage des amateurs.* Société astronomique de France, 1986.

[76] Nicholas METROPOLIS et Stanisław ULAM. The Monte-Carlo method. *Journal of American Statistical Association*, **44**, p. 335–341, 1949.

[77] Jean-Yves OUVRARD. *Probabilités 2.* Cassini, 2000.

[78] Jean PERRIN. *Les atomes.* PUF, 1948.

[79] George PÓLYA et Gabor SZEGŐ. *Problems and theorems in analysis.* Springer, 1978.

[80] Luca PRATELLI. La loi des grands nombres pour une suite échangeable. In *Séminaire de probabilité (Strasbourg)*, volume 23, p. 527–530, 1989.

[81] Daniel REVUZ et Marc YOR. *Continuous martingales and Brownian motion.* Springer Verlag, 3e édition, 2005.

[82] Christiane ROUSSEAU et Yvan SAINT-AUBAIN. *Mathématiques et technologie.* Springer, 2008.

[83] Walter RUDIN. *Analyse réelle et complexe.* Masson, 1991.

[84] David RUELLE. *Hasard et chaos.* Odile Jacob, 1991.

[85] David RUELLE. *Chance and chaos.* Princeton University Press, 1993.

[86] Alfred RÉNYI. *Foundations of probability.* Dover, 2007. Réédition de l'ouvrage de 1970 originellement publié par Holden-Day.

[87] Alfred RÉNYI. *Probability Theory.* Dover, 2007. Réimpression de l'ouvrage publié chez North-Holland en 1970, version augmentée du *Calcul des probabilités* paru chez Dunod en 1966.

[88] Daniel SAADA. *Tribus et probabilités sur les univers infinis.* `www.daniel-saada.eu`, 2009.

[89] Erwin SCHRÖDINGER. *Qu'est-ce que la vie ?* Point Sciences. Christian Bourgois, 1986. (*What's Life?*, 1967).

[90] Jean-Pierre SERRE. *Cours d'arithmétique.* PUF, 1994.

[91] Albert Nicolaievitch SHIRYAEV. *Probability.* Springer-Verlag, 1996.

[92] Robert SOLOVAY et Volker STRASSEN. A fast Monte-Carlo test for primality. *SIAM Journal on Computing*, **6**(1), p. 84–85, 1977.

[93] Elias STEIN et Rami SHAKARCHI. *Functional analysis : an introduction to further topics in Analysis.* Princeton University Press, 2008.

[94] Ian STEWART. *Dieu joue-t-il aux dés ? Les mathématiques du chaos.* Collection Champs. Flammarion, 1998.

[95] Tom STOPPARD. Rosencrantz and Guildenstern are dead. Un film de Tom Stoppard, d'après la pièce éponyme, avec Gary Oldman et Richard Dreyfuss.

[96] Tom STOPPARD. *Rosencrantz & Guildenstern are dead.* Faber, 1991.

[97] Gérald TENENBAUM et Michel MENDÈS FRANCE. *Les nombres premiers.* Collection « Que sais-je ? ». PUF, 2000.

[98] Colin J. THOMPSON. *Classical equilibrium statistical mechanics.* Clarendon Press Oxford, 1988.

[99] Hermann WEYL. Über die Gleichverteilung von Zahlen mod. Eins. *Mathematische Annalen*, **77**(3), p. 313–352, 1916.

[100] David WILLIAMS. *Probability with martingales.* Cambridge University Press, 1991.

[101] Thomas A. WITTEN et Leonard M. SANDER. Diffusion-limited aggregation, a kinetic critical phenomenon. *Physical Review Letters*, **47**, p. 1400–1403, 1981.

[102] Claude ZUILY et Hervé QUEFFÉLEC. *Analyse pour l'agrégation.* Dunod, 2006.

Index

Les numéros de page en **gras** renvoient aux définitions principales, ceux en *italique* à des exercices. Les noms en PETITES CAPITALES renvoient aux indications biographiques. Les numéros de page précédés par la lettre « H » renvoient à des définitions données dans la partie « Heuristique », qui peuvent être plus informelles que les autres.

Symboles

0–1 (loi du — de Kolmogorov) 419
$\mathbf{1}_A$ 133
$\mathfrak{B}(\mathbb{C})$ 357
$\mathfrak{B}^\infty$ 699
$\mathfrak{B}^n$ 699
Γ (fonction d'Euler) 594, 675
Γ (loi —) 235, *279*
Ω ..... [H34]
χ^2 (loi du —) 235
δ_a (mesure de Dirac) 84, **92**
γ (loi —) 235
simulation *530*
ω ..... [H34]
π–λ (théorème —) 704
π-système 703
— et indépendance 705
$\preccurlyeq$ 701
$\sigma(X_0, X_1, \ldots, X_n)$ 418
$\sigma(X)$ 341
$\sigma(\mathfrak{F})$ 81
σ-additivité ..... [H44], **84**
σ-algèbre **79**
σ-finie (mesure —) 707
σ-sous-additivité 87, *97*
$\sim$ $(u_n \sim v_n)$ 669
$\otimes$ (produit direct/tensoriel) 263
$\otimes$ (tribu produit)
$\mathfrak{F} \otimes \mathfrak{G}$ 699
$\bigotimes \mathfrak{F}_i$ 699
ζ (fonction — de Riemann) *507*, 586

A

Absolument continue
densité — 93
fonction — 675
mesure — 707
variable aléatoire — 224
Absorbant
classe —e 562
état — 562
Accroissements
indépendants 605
stationnaires 605
Adaptée (suite — à une filtration) 483
Additive 85
σ-additive ..... [H44], **84**
sous-additive 85
Adjoint d'un endomorphisme 693
Aiguille de Buffon 534
Aléa ..... [H34]
Aléatoire
matrice — *339*, 585
variable — ..... *voir* Variable aléatoire
Alembert (erreur de d'—) 39
Algèbre **78**
engendrée 79
sans atomes *96*
Anniversaires (paradoxe des —) *50*
Aplatissement 252
Application mesurable ..... [H58], 291, 700
Approximation
binomiale par normale 449
binomiale par Poisson 448
Poisson par normale 451
Arc sinus (loi de l'—) *241*, 647
Arrêt (temps d'—) **350**, 567
Asymétrie (coefficient d'—) 252
Asymptotique (tribu —) 417
Atome 79
algèbre sans —s *96*
d'une partition 470
tribu sans —s *96*
Atomique (événement) ..... [H34]
Atteinte (temps d'—) 567
Autoadjoint (endomorphisme —) 694
Avogadro (nombre d'—) 630
Axiomes
1 à 3 84
3′, 3″ et 3‴ 86

B

B(a, b) (loi bêta) 235
$\mathscr{B}(n\,;\,p)$ ..... *voir* Binomiale (loi)
$\mathscr{B}(p)$ 133
Bêta (loi —) 235
Baguette brisée *278*
Banach (espace de —) 693
Base hilbertienne 688
Bayes (formule de —) 104, 105
version continue 273
Benford (loi de —) 546
Bernoulli
épreuve de — 133
loi de — 133
Bernstein
formule de — 495
inégalité de — 328
théorème de — 502
Bertrand (série de —) 669
Biais d'un estimateur 462
Bienaymé (égalité de —) 186, 270, 350
Bienaymé-Tchebychev (inég. de —) . 158, 253, **327**
d'ordre k 327
Bilatérale (loi exponentielle —) *280*
Binomial (coefficient —) 690
Binomiale (loi —) **134**
approximation par Poisson 448
approximation par une loi normale . 449
calcul de l'espérance 149
Boltzmann
facteur de — 64, **593**
formule de — 596
Boole
algèbre de — 78
inégalité de — 87, *97*
Borel
lemme de Borel-Cantelli 117
théorème des nombres normaux de — 218

Borélienne (fonction —) 700
Boréliens
de $\mathbb{C}$ 357
de $\mathbb{R}$ 81, **697**
de $\mathbb{R}^n$ 332, **699**
Borne inférieure/supérieure 79
Bortkiewicz Władisław 137
Boson ..... [H41], 598
Boule ouverte 693
Box-Muller (méthode de —) 521
Branchement (processus de —) 195
Brownien ..... *voir* Mouvement
Bruit de grenaille 607
Buffon, comte de 534
Buffon (aiguille de —) 534
Bywater John 612

C

Cantelli
lemme de Borel- — 117
loi forte 432, *434*
théorème de Glivenko- — 429
Cantor
diagonale de — 667
fonction de — 678
Carathéodory (théorème de —) 94, **707**
Cauchy
loi de — 234
somme de variables de — *363*
produit de — 684
suite de — 668
Cauchy-Schwarz (inégalité de —) 156, 249, **321**, 334
dans un espace préhilbertien 687
Causalité 459
Cavalerie Légère 317
Central (théorème-limite —) **438**, 443
Centré
moment —
d'une variable à densité 247
d'une variable discrète 152
variable —e 156, 252
Chaîne de Markov *199*, **552**
Chaleur (équation de la —) 616
Champernowne (nombre de —) 219
Changement de variable 686
aléatoire 237
Chapeaux (problème des —) 149
Chapman-Kolmogorov (équation de —) 557, 616
Chenille *50*, *576*
Classe
absorbante 562
fermée 562
indécomposable 561
récurrente 567
transitoire 567
Classe monotone 703
engendrée 704
lemme de — 704
Coalitions (lemme des —)
cas à densité 265
cas discret 179
(démonstration) 706
pour des tribus 343, **705**
pour des variables aléatoires .. **344**, **705**
Cochran (théorème de —) 379
Coefficient
d'asymétrie 252
binomial 690
de corrélation 187, 270, 333
Collectionneur (problème du —) 150, **163**
Communicants (états —) 559
Commutativement convergente (série —) . 671
Comparaison de séries positives 670
Complet
espace probabilisé — 87
espace vectoriel normé — 693
système — 88
système — induit par X 128
système quasi- — 88
tribu complète 87
Complexion 591
Composées (th. des probabilités —) [H48], 103
Conditionnelle
densité — 272
espérance ..... *voir* Espérance
probabilité
sachant une partition 487
sachant une tribu 488
sachant une v.a. 488
probabilité — ..... [H47], 102
Confiance (intervalle de —) 463
Conjointe
densité — 258
fonction de répartition — 257
loi —
cas à densité 257
cas discret 174
Continue
fonction — à droite 674
variable aléatoire — 224
Continuité
absolue — 675
correction de — 449
monotone 86
radiale (lemme d'Abel) 684
théorème de — 362, 398, **401**
uniforme 674
Convergence
commutative 671
dominée (th. de — discret) 165, 674
dominée (th. de —) 316
étroite 393, **395**
dans L^p 392
en loi **393**, 395
monotone 315
(démonstration) 329
en moyenne 392
presque sûre 385
en probabilité **388**
simple 682
d'une suite de parties 116
uniforme 682
Convergent (estimateur —) 462
Convolution
des densités 349
de distributions de Dirac 349
des fonctions de répartition 348
formule de — 348
formule de — discrète 182
des mesures de probabilité 348
Correction de continuité 449
Corrélation
et causalité 459
coefficient de — 187, 270, 333
Coupon collector's problem 150, **163**
Covariance **322**
d'un couple à densité 269
d'un couple discret 183
matrice de — 333
changement de base 334
Cramér (modèle de —) 544
Crible (formule du —) 89
Critère de Weyl 405

Croissance (propriété de —)
pour les probabilités ... 85
pour les tribus ... 82
Cylindre ... 204, 624, **699**, **715**

D

d'Alembert (erreur de —) ... 39
Dés non transitif ... *662*
De Moivre (théorème de —-Laplace) ... 449
de Morgan (règles de —) ... 38
Décomposition de Lebesgue ... 298, 678
Décorrélation ... 325
Décorrélées (variables aléatoires —) ... 184
Demi-vie ... 251
Dénombrable ... 667
Dénombrablement additive ... [H44]
Dénombrement ... 667
Densité
au sens des distributions ... 299
conditionnelle ... 272
conjointe ... 258
marginale ... 261, 335
d'une mesure ... 707
d'une probabilité ... 93
sur $\mathbb{R}^n$ ... 332
d'une variable aléatoire ... [H62], 224
d'un vecteur gaussien ... 377
Dés non transitifs ... 662
Destin ... [H34]
Déterministe
système — ... 33
variable — ... **33**, 352
Diagonale de Cantor ... 667
Diffuse (loi —) ... 224
Dini (théorème de —) ... 682
Dirac
convolution de distributions de — ... 349
distribution de — ... 299, **690**
mesure de — ... 84, **92**
Dirichlet
problème de — ... 571, 630
théorème de — ... 671
Discontinuité
de 1re/ de 2e espèce ... 676
saut de — ... 677
Discrète
tribu — ... 79
variable aléatoire — ... [H61], **127**
Distribution ... 689
de Dirac ... 299, **690**
convolution ... 349
de Maxwell-Boltzmann ... 64, 595
Dominée (th. de convergence —) ... 316
cas discret ... 165, 674
Doob
inégalité des remontées ... 486
lemme de — -Dynkin ... 478, 701
Droite de régression ... 372, **456**
Durée de vie ... 251
Dynkin
lemme de Doob- — ... 478, 701
théorème de — ... 704

E

$\mathscr{E}(\alpha)$ ... 230
Écart-type ... **314**
d'une variable à densité ... 248
d'une variable discete ... 155
Échec ... 133
Égalité
de Bienaymé ... 186, 270, 350
de Parseval ... 688
Ehrenfest (modèle des urnes d'—) ... 574
Einstein Albert ... 630
Empirique
fonction de répartition — ... 428
moyenne — ... 460
Endomorphisme autoadjoint ... 694
Endomorphisme orthogonal ... 694
Engendrée
algèbre — ... 79
classe monotone — ... 704
tribu — ... 81, 341, 418
Entropie ... 596
Épreuves (espace des —) ... [H34], 82
Équation
de la chaleur ... 616
de Chapman-Kolmogorov ... 557, 616
Équidistribuée (suite —) ... 405
Erdős Paul ... 209
Erdős-Kac (théorème d'—) ... 581
Ergodique
matrice / chaîne — ... 564, 695
théorème — ... 565, 695
Erreur
de d'Alembert ... 39
fonction d'— « erf » ... 232
Espace
de Banach ... 693
des épreuves ... [H34], 82
euclidien ... 693
mesurable ... **82**
probabilisé ... **84**
probabilisable ... **82**
vectoriel normé ... 693
complet ... 693
Espérance
calcul pratique ... 149
conditionnelle
sachant une partition ... 470
sachant une tribu ... 477
sachant une v.a. ... **478**
sachant une v.a. à densité ... 277
sachant une v.a. discrète ... 190, 191, **474**
définition alternative ... 300, 301
définition générale ... **311**
infinie ... 309, 311
(cas discret) ... 147
linéarité ... 312
cas discret ... 148
totale (formule de l'—) ... 190
d'une v.a. complexe ... 311, **357**
d'une v.a. à densité ... [H72], **245**
d'une v.a. discrète ... [H71], **143**
d'une v.a. simple ... 308
d'une v.a. positive ... 309
Estimateur ... 462
Étagée ... *voir* Simple
État
—s communicants ... 559
absorbant ... 562
final ... 562
microscopique ... 590
périodique ... 563
récurrent ... 567
transitoire (transient) ... 567
Étroite (convergence —) ... 393, **395**
Euclidien (espace —) ... 693
Euler
constante d'—-Mascheroni ... 669
fonction Γ d'— ... 594, 675
identité d'— ... *507*
indicatrice d'— ... *507*

Événement ..... [H35], **82**
atomique ..... [H34]
—s incompatibles 82
—s indépendants 112
—s mutuellement indépendants 113
négligeable ..... [H38], 87
presque sûr ..... [H38], 87
de queue 417
récurrent 208
régénératif 207
symétrique 420
transient 208
Expérience (résultat d'une —) ..... [H34]
Exponentielle
bilatérale (loi —) *280*
loi — ..... [H62], 230
lien avec la loi de Poisson ... 241, 514
simulation 515
Extension (th. d'— de Kolmogorov) 714

F

Face (trois — d'affilée) *50*
Facteur de Boltzmann 64, **593**
Factoriel (moment —) 166
Factorisation (lemme de —) 478, 701
Faible (loi — des grands nombres) ... 211, **424**
de Khintchine 424
Famille sommable 672
Fatou
inégalités de — 124
lemme de — 315
Feller William 209
Fermé 693
classe —e 562
Fermion 598
Feu tricolore (temps d'attente à un —) *296*, *355*
Filtration 483
Final (état —) 562
Fluctuation (intervalle de —) 463
Fonction
Γ d'Euler 594, 675
borélienne 700
de Cantor 678
d'erreur 232
génératrice 160
de Haar 628
harmonique 571
indicatrice d'un événement 133
mesurable ..... [H58], 291, 700
de Rademacher 216
de répartition ..... [H58], 130
conjointe 257
d'une probabilité sur $\mathbb{R}$ 90
empirique 428
de Schauder 628, *639*
singulière 297, 677
-test 689
uniformément continue 674
à variation bornée 675
Formule
de Bayes 104, 105
version continue 273
de Bernstein 495
de Boltzmann 596
de convolution 182, 348, **349**
du crible 89
de l'espérance totale 190
de König-Huygens 154, 248, **314**
de Poincaré 89
des probabilités composées .... [H48], 103
des probabilités totales ..... [H47], 103
de Stirling 669
preuve probabiliste 497
de transfert **318**
pour un couple à densité 263
pour un couple discret 176
pour une variable à densité .. 246, 320
pour une variable discrète 152
de Vandermonde 140, 691
Forte (loi — des grands nombres) ... 212, **425**, 426
(démonstration de Cantelli) 434
(démonstration de Pratelli) 432
Forte (propriété de Markov —) 567
Fortement irréductible (matrice —) .. 564, 695
Fortune
comment gagner des —s 350
Dame — ..... [H34]
Frobenius (théorème de Perron- —) .. 565, 695
Fubini (théorème de —)
cas général 335, 685
pour les séries doubles 672, 673, *696*

G

$\mathcal{G}(p)$ 134
Galton
paradoxe de — 662
processus de — -Watson 195
Gamma
fonction Γ d'Euler 594, 675
loi $\Gamma(\alpha, \beta)$ 235, *279*
gamma
loi $\gamma(\nu)$ 235
simulation 519, *530*
Gaussien
loi —ne ..... *voir* Loi normale
variable —ne 366
vecteur — 374
Génératrice (fonction —) 160
Géométrique (loi —) 134
simulation 513
Gibbs Willard 590
Gibbs (paradoxe de —) 596
Glivenko-Cantelli (théorème de —) 429
Gouy Louis Georges 612
Grands nombres
loi faible des — 211, **424**
de Khintchine 424
loi forte des — 212, **425**, 426
(démonstration de Cantelli) 434
(démonstration de Pratelli) 432
Grenaille (bruit de —) 607
Grenouille (paradoxe des —) *279*
Grossière
divergence — 669
tribu — 79
Guildenstern & Rosencrantz 27
Gumbel (loi de —) 166

H

Haar (fonctions de —) 628
Harmonique
nombre — 163
série — 669
fonction — 571
Heine (théorème de —) 674
Hewitt-Savage (loi du tout ou rien de —) . 420
Hilbertienne (base —) 688
Hölder (inégalités de —) 167
Huygens (théorème de König- —) ... 154, 248, **314**
Hypergéométrique (loi —) 139
Hypothèse de Riemann 586

I

Identité
de Wald 193, 351
d'Euler *507*
Image
probabilité — 294
réciproque d'une partie 58
Indépendance
accroissements indépendants 605
de deux événements 112
mutuelle
de classes d'événements 343
d'événements 113
de tribus 343
de variables aléatoires **344**
de variables discrètes 178
de variables à densité 264
représentation graphique 342, 354
de deux v.a. **341, 342**
à densité 263
discrètes 177
Indicatrice
d'Euler *507*
fonction — d'un événement 133
Inégalité
de Bernstein 328
de Bienaymé-Tchebychev .. 158, 253, **327**
d'ordre k 327
de Boole 87, *97*
de Cauchy-Schwarz 156, 249, 321, 334, 687
de Doob 486
de Fatou 124
de Hölder 167
de Jensen **328**, 477
de Kolmogorov 650
de Lyapounov 329
de Markov **326**
pour une v.a. à densité 253
pour une v.a. discrète 158
pour une v.a. discrète *167*
des remontées de Doob 486
triangulaire 670
inf 79
Initiale (loi —) 552
Inspection (paradoxe de l'—) 72, 609
Instantanée (loi —) 556
Intégrable (v.a. —) 311
Intégrale
gaussienne 718
de Paley-Wiener-Zygmund *639*
selon une mesure de probabilité 308–314
Intégration terme à terme 151, 316
Intervalle
de confiance 463
de fluctuation 463
Invariante (loi, probabilité, mesure —) ... 557
Irréductible
matrice fortement — 564, 695
matrice, chaîne — 562, 695
Itéré (loi du logarithme —) 444

J

Jacobi (symbole de —) 540
Jacobien 686
Jensen (inégalités de —) **328**, 477
Jeu non transitif 661
Joker ..... [H43]
Jumeaux (entiers « premiers »—) *549*

K

Kac (théorème d'Erdős- —) 581
KAKUTANI Shizuo 632
Khintchine (théorème de —) 424
Kolmogorov
axiomes de — 84
équation de Chapman- — 557, 616
inégalité de — 650
loi des grands nombres de — .. **425**, 426
loi du tout ou rien de — 419
test de — -Smirnov 430
théorème d'extension 714
König-Huygens (théorème de —) 154, 248, **314**
Kurtosis 252

L

$\mathcal{L}^0$ ou $\mathcal{L}^0(\Omega)$ 291
$\mathcal{L}^0_b(\Omega)$ 291
$\mathcal{L}^0_+(\Omega)$ 291
L^1 313
$\mathcal{L}^1$ ..... [H71], 311
L^2 321
$\mathcal{L}^2$ 321
L^p 324
Lachesis ..... [H34]
Lagrange (multiplicateurs de —) 687
Laplace
— -Gauss (loi de —) 231
théorème de de Moivre- — 449
$\mathcal{L}^p$ 314
Lebesgue
décomposition de — 298, 678
mesure de — sur [0 ; 1] 92
tribu de — 88
LEEUWENHOEK Antoni 612
Lemme
d'Abel 684
de Borel-Cantelli 117
de classe monotone 703, **704**
des coalitions
cas à densité 265
cas discret 179
(démonstration) 706
pour des tribus 343, **705**
pour des variables aléatoires **344**, **705**
de Doob-Dynkin 478, 701
de factorisation 478, 701
de Fatou 315
des moments 378
Lévy (théorème de —) 362, **401**
lim inf et lim sup 114
lim ↑, lim ↓ 80
Limite
inférieure/supérieure 114
d'une suite de parties 116
d'une suite monotone de parties 80
théorème — central **438**, 443
Logarithme (loi du — itéré) 444
Loi ..... [H56], **294**
a priori/*a posteriori* 273
de l'Arc sinus 647
bêta B(a, b) 235
de Benford 546
de Bernoulli 133
binomiale 134
approximation normale 449
approximation par Poisson 448
calcul de l'espérance 149
de Cauchy 234
somme de variables de — *363*
conditionnelle
cas à densité 271
cas discret 175, 189

conjointe
cas à densité 257
cas discret 174
convergence en — 393, 395
diffuse 224
du tout ou rien
de Hewitt-Savage 420
de Kolmogorov 419
exponentielle ..... [H62], 230
lien avec la loi de Poisson ... 241, 514
simulation 515
somme de deux v.a. *279*
faible des grands nombres 211, **424**
de Khintchine 424
forte des grands nombres .. 212, **425**, 426
(démonstration de Cantelli) 434
(démonstration de Pratelli) 432
Gamma $\Gamma(\alpha, \beta)$ 235, *279*
gamma $\gamma(\nu)$ 235
simulation 519, *530*
gaussienne ..... *voir* Loi normale
géométrique 134
simulation 513
de Gumbel 166
hypergéométrique 139
initiale 552
instantanée 556
invariante 557
du χ^2 235
du logarithme itéré 444
marginale
cas à densité 258, 335
cas discret 174, 175
normale **231**
quotient de deux v.a. 266
simulation 520
somme de deux v.a. 266
des petits nombres 137
de Poisson 136
approximation normale 451
lien avec l'exponentielle 241, 514
mode d'une — *141*
simulation 514
du premier succès 134
de Rayleigh 521
sans mémoire 136, 230
uniforme
à densité 229
discrète 133
d'une variable aléatoire **294**
Lyapounov (inégalités de —) 329

M

Marche aléatoire
discrète 612
limite continue 613, 614
Marginale
— s d'un couple 175, 258
densité — 261, 335
loi —
cas à densité 258, 335
cas discret 174, 175
Markov
chaîne de — stationnaire *199*, 552
inégalité de — **326**
pour une v.a. à densité 253
pour une v.a. discrète 158, *167*
processus de — 627
propriété de — 552
propriété de — forte 567
Martingale 482, **483**
Masse de Dirac 84, **92**
Matching problem 149
Maternité (nuits agitées à la —) *464*
Matrice
aléatoire *339*, 585
de covariance 333
changement de base 334
ergodique 564, 695
fortement irréductible 564, 695
irréductible 562, 695
jacobienne 686
positive 555, 694
régulière 564, 695
stochastique **552**, 555, **694**
symétrique définie positive 694
symétrique positive 694
de transition 552
Maximum
de variables aléatoires *267*, *279*, *280*, *329*
de vraisemblance 645
MAXWELL James Clerk 590
Maxwell-Boltzmann (distribution de —) 64, 595
Médiane d'une v.a. à densité 228
Mémoire (loi sans —) 136, 230
Menteurs (la chaîne des —) *50*
Mesurable
espace — 82
fonction — ..... [H58], 291, 700
selon une partition 475
selon une tribu 478, 701
un ensemble non — 95
Mesure
σ-finie 707
absolument continue 707
de Dirac 84, **92**
de Lebesgue sur [0 ; 1] 92
produit 332
de Wiener 623
Méthode
de Box-Muller 521
de Box-Muller avec rejet 522
de Monte-Carlo ..... *voir* Monte-Carlo
du rejet 427, **516**, 538
Métro (temps d'attente du —) 274
Middle-square (algorithme —) 526, *530*
Mixte (variable aléatoire —) 296
Mode *141*
Modèle de Cramér 544
Moment **314**
centré
d'une variable à densité 247
d'une variable discrète 152
d'une variable à densité 247
factoriel *166*
lemme des —s 378
récupération des —s 162
Monotone
classe — 703
engendrée 704
lemme de classe — 704
théorème de convergence — 315
(démonstration) 329
Monte-Carlo (méthode de —) **533**
pour calculer une intégrale 536
pour estimer une moyenne 534
pour le problème de Dirichlet 630
pour la primalité 540
pour effectuer un tri 545
Morgan (règles de de —) 38

Mouvement brownien
  construction canonique 624
  covariance 625
  définition 621
  existence 622
  loi (mesure de Wiener) 623
Moyenne
  empirique 460
  règles de la valeur — 570, 571
Multiplicateurs de Lagrange 687
Multivariée (distribution normale —) 374
Mutuellement indépendants (évts. —) 113

N

$\mathcal{N}(m\ ;\ \sigma^2)$ 231
Naissances 72
Needham John Turberville 612
Négligeable ..... [H38], **87**
Nikodým (théorème de Radon- —) 707
Nombre
  de Champernowne 219
  normal 218, 498
  —s de Stirling 162, **691**
Normal (nombre —) 218, 498
Normale (loi —) **231**
  quotient de deux v.a. 266
  simulation 520
  somme de deux v.a. 266
Norme 693
Nornes ..... [H34]

O

O, o 669
Observable 82
Ordre (statistique d'—) *339*
Orthogonal (endomorphisme —) 694
Orthonormée (famille — de L^2) 325
Ouvert 693

P

$\mathcal{P}(\lambda)$ 136
Paley-Wiener-Zygmund (intégrale de —) . *639*
Paquets (th. de sommation par —) 674
Paradoxe
  des anniversaires *50*
  de Galton 662
  de Gibbs 596
  des grenouilles *279*
  de l'inspection *72*, *609*
  de Walter Penney *575*, 661
  de Saint-Pétersbourg 658
  téléphonique *339*
Parseval (égalité de —) 688
Partie
  fermée 693
  ouverte 693
  positive/négative 147, **310**
  singulière 297, 677
Penney (paradoxe de W. —) *575*, 661
Périodique (état —) 563
Perrin Jean 630
Perron (théorème de — -Frobenius) .. 565, 695
Petits nombres (loi des —) 137
π-système 703
  — et indépendance 705
Poincaré (formule de —) 89
Points d'un processus 604
Poisson
  loi de — 136
    approximation normale 451
    lien avec l'exponentielle 241, 514
    modes d'une — *141*
  simulation 514
  processus de — 604
Pólya (théorème de récurrence) 627
Polynômes de Bernstein 502
Positive (matrice —) 555, 694
Pratelli (Luca —) 432
Presque sûr ..... [H38], 87
  convergence presque sûre 385
Presque sûrement ..... [H38]
  $(X_n)_{n\geqslant 1} \to X$ p.s. 385
Primalité (tests probabilistes de —) 540
Probabilisable (espace —) 82
Probabilité **84**
  —s composées (th. des —) .... [H48], 103
  conditionnelle ..... [H47], **102**
    sachant une partition 487
    sachant une v.a. discrète 488
  convergence en — 388
  image ..... [H57], 294
  invariante 557
  produit 332, 700
  —s totales (formule des —) ... [H47], 103
  de transition 552
Problème
  des chapeaux (des conjoints) 149
  du collectionneur 150, **163**
  de Dirichlet 571, 630
Processus
  de branchement 195
  de comptage 604
  de Galton-Watson 195
  markovien 627
  de Poisson 604
  stochastique 604
Produit
  de Cauchy 684
  direct/tensoriel 263
  probabilité — 332, 700
  tribu — 332, 699
Programme (en environnement instable) . *352*
Projection orthogonale 687
Propriété
  de Markov 552
  de Markov forte 567
Pseudo-inverse d'une fonction croissante .. 679

Q

Quadratique (risque —) 462
Quasi-complet (système —) 88
Queue (événement de —) 417
Quotient de deux v.a. normales 266

R

Rademacher
  fonctions de — 216
  suite de — 205
Radial (lemme de continuité —e d'Abel) . 684
Radon (théorème de — -Nikodým) 707
Rang d'un cylindre 204
Rayleigh (loi de —) 521
Rayon d'une série entière 683
  aléatoire 651
Réalisation (théorème de —) 711
Réarrangement *339*
Réciproque (image — d'une partie) . [H58], 292
Recuit simulé 545
Récupération des moments 162
Récurrent
  (mouvement brownien) 627
  événement — 208
  classe —e 567
  état — 567
Réduite (variable —) 156, 252

Régénératif (événement —) 207
Règles
de de Morgan 38
de la valeur moyenne 570, 571
Régression (droite de —) 372, **456**
Régulière (matrice / chaîne) 564, 695
Rejet (méthode du —) 427, 516
Remontées (inégalité des —) 486
Répartition (fonction de —) ... [H58], 130, **294**
d'un couple à densité 257
empirique 428
d'une probabilité sur $\mathbb{R}$ 90
Résultat d'une expérience ..... [H34]
Retour
temps de — 567
temps moyen de — 569
Riemann
fonction ζ de — *507*, 586
hypothèse de — 586
Riesz-Fisher (théorème de —) 323
Risque quadratique 462
Rosencrantz & Guildenstern 27

S

Saint-Pétersbourg (paradoxe de —) 658
Saut de discontinuité 677
Schauder (fonctions de —) 628
Séparable (tribu —) 88
Séquence
homogène 205
de trois *face* *50*
Série
aléatoire 421, 649
de Bertrand 669
entière 683
aléatoire 499, 651
harmonique 669
de variables positives 315
Simple (convergence —) 682
Simple (variable aléatoire —) 127, 302
espérance 308
Simulation
d'une loi continue 515
d'une loi exponentielle 515
d'une loi $\gamma(a)$ 519
d'un vecteur gaussien 523
d'une loi géométrique 513
d'une loi normale 520
d'une loi de Poisson 514
Singe 4, 117
Singulière
fonction — 297, 677
partie — 297, 677
Skorokhod (th. de représentation de —) .. 410
Slutsky (théorème de —) **402**, 463
Smirnov (test de Kolmogorov- —) 430
Smoluchowski Marian 630
Sommable (famille —) 672
Somme de deux variables
à valeurs dans $\mathbb{Z}$ 176
binomiales 183
à densité 261
à densité indépendantes 265
discrètes indépendantes 182
exponentielles *279*
normales 266
de Poisson 183
Sondage *51*, *465*
Sous-additivité 85, *97*
dénombrable 87
Sous-martingale 484
Sous-tribu 79
Spectral (théorème —) 694
Stabilité
de la loi binomiale 183
de la loi normale 266
de la loi de Poisson 183
Stationnaire
accroissements —s 605
chaîne de Markov — 552
Statistique d'ordre *339*
Steinhaus (théorème de —) 499
Stieltjes (intégrale de —) 301
Stirling
formule de — 669
preuve probabiliste 497
nombres de — 162, **691**
Stochastique (matrice) 552, 555, 694
Stratégie
pour gagner à *pile-face* 350
pour gagner un trésor *662*
Succès 133
loi du premier — 134
Suite
de Cauchy 668
équidistribuée 405
de Rademacher 205
uniformément distribuée 405
sup 79
Sur-martingale 484
Symétrique
événement — 420
matrice — (définie) positive 694
variable aléatoire — 229, 247
Système
complet d'événements 88
induit par X 128
quasi-complet d'événements 88
total 688

T

Tchebychev (inégalité de —) ... 158, 253, **327**
d'ordre k 327
Temps
d'arrêt **350**, 567
d'attente
au feu tricolore *296*, *355*
au rendez-vous *279*
de retour 567
moyen 569
Test
fonction-test 689
de Kolmogorov-Smirnov 430
de primalité 540
Théorème
de Bayes 104, 105
version continue 273
de Bernstein 502
de Carathéodory 94, **707**
central limit **438**, 443
de Chapman-Kolmogorov 557, 616
de Cochran 379
de comparaison de séries positives .. 670
de continuité
(fonctions caractéristiques) . 362, **401**
(fonctions génératrices) 398
de convergence dominée 316
de convergence dominée discret 165, 674
de convergence monotone 315
(démonstration) 329
de convolution 265, **349**
de de Moivre-Laplace 449
de décomposition de Lebesgue .. 298, 678
de Dini 682
de Dirichlet 671
d'Erdős-Kac 581

ergodique 565, 695
d'extension de Kolmogorov 714
de Fubini
cas général 335, 685
pour les séries doubles . 672, 673, *696*
de Glivenko-Cantelli 429
de Heine 674
d'intégration terme à terme 151, 316
de Khintchine 424
de Kolmogorov 714
de König-Huygens 154, 248, **314**
de Lévy 362, **401**
limite central **438**, 443
de Perron-Frobenius 565, 695
des probabilités composées .... [H48], 103
de Radon-Nikodým 707
de réalisation 711
de récupération des moments 162
de récurrence de Pólya 627
de Riesz-Fisher 323
de représentation de Skorokhod 410
de Slutsky **402**, 463
de sommation par paquets 674
spectral 694
de Steinhaus 499
de transfert **318**
pour un couple à densité 263
pour un couple discret 176
pour une variable discrète 152
pour une variable à densité .. 246, 320
de Weierstrass 501, 683
de Wick 378
Total
formule de l'espérance —e 190
formule des probabilités —es .. [H47], 103
système — 688
Tour (*tower property*) 475
Tout ou rien (loi du — de Kolmogorov) 419, 420
Transfert (théorème de —) **318**
pour un couple à densité 263
pour un couple discret 176
pour une variable à densité 246, 320
pour une variable discrète 152
Transient
(mouvement brownien) 627
état — 567
événement — 208
Transitif
dés non —s *662*
jeu non — 661
Transition (matrice, probabilité de —) ... 552
Transitoire
classe —e 567
état — 567
Tri (méthode de — rapide randomisé) 545
Triangulaire (inégalité —) 670
Tribu **79**
asymptotique 417
complète 87
discrète 79
engendrée 81
engendrée par plusieurs v.a. 418
engendrée par une v.a. 341, 701
grossière 79
de Lebesgue 88
produit 332, 699
de queue 417
sans atomes *96*
séparable 88
sous- — 79
sur $\mathbb{C}$ 357
triviale 79
Trois *face* d'affilée *50*
Tronquée (variable —) 296

U

$\mathscr{U}([a\,;b])$ 229
$\mathscr{U}([a\,;b])$ 133
Uniforme
convergence — 682
loi — à densité 229
loi — discrète 133
Uniformément
continue (fonction —) 674
distribuées (suites —) 405
Univers ..... [H34], **82**
— -image 127
Urnes d'Ehrenfest 574

V

Valeur moyenne (règles de la —) 570, 571
Vandermonde (formule de —) 140, 691
Variable aléatoire ..... [H56], 291
absolument continue 224
centrée 156, 252
complexe 311, **357**
continue 224
—s décorrélées 184
à densité ..... [H62], 224
espérance ..... [H72], **245**
déterministe **33**, 352
discrète ..... [H61], **127**
espérance ..... [H71], **143**
espérance
définition alternative 300, 301
définition générale **311**
gaussienne 366
—s indépendantes 341
intégrable 309
mixte 296
réduite 156, 252
simple 127, 302
espérance 308
symétrique 229, 247
tronquée 296
Variance **314**
d'une v.a. à densité 248
d'une v.a. discrète 153
Variation bornée (fonction à —) 675
Vecteur
aléatoire 331
gaussien 374
simulation 523
Version
de l'espérance conditionnelle 472
de la probabilité conditionnelle . 488, 489
Vieillissement (propriété de non —) .. 136, 230
Vitali (contre-exemple de —) 95
Voisinage 693
von Mises (paradoxe des anniversaires) *50*
Vraisemblance (maximum de —) 645

W

Wald (identités de —) 193, 351
Watson (processus de Galton- —) 195
Weierstrass (théorème de —) 501, 683
Weyl (critère de —) 405
Wick (théorème de —) 378
WIENER Norbert 623
Wiener (mesure de —) 623

Z

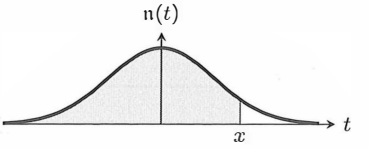
Zéro-un (loi du — de Kolmogorov) 419
Zeta (fonction — de Riemann) *507*, 586

Mis en page avec LaTeX 2_ε sous environnement Linux

Figures réalisées avec TikZ, Python, Scilab et Xfig

LOI NORMALE CENTRÉE RÉDUITE

$$\mathfrak{N}(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\, \mathrm{d}t.$$



Les décimales se lisent sur les lignes, et on ajoute les centièmes rangés en colonnes. Par exemple, la valeur de $\mathfrak{N}(1,93)$ est donnée à l'intersection de la ligne $1,9$ et de la colonne $0,03$, et l'on peut lire $\mathfrak{N}(1,93) = 0,9732$, à 10^{-4} près. Pour les valeurs négatives de x, on utilise la relation $\mathfrak{N}(-x) = 1 - \mathfrak{N}(x)$.

x	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7793	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8906	.8925	.8943	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9986	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

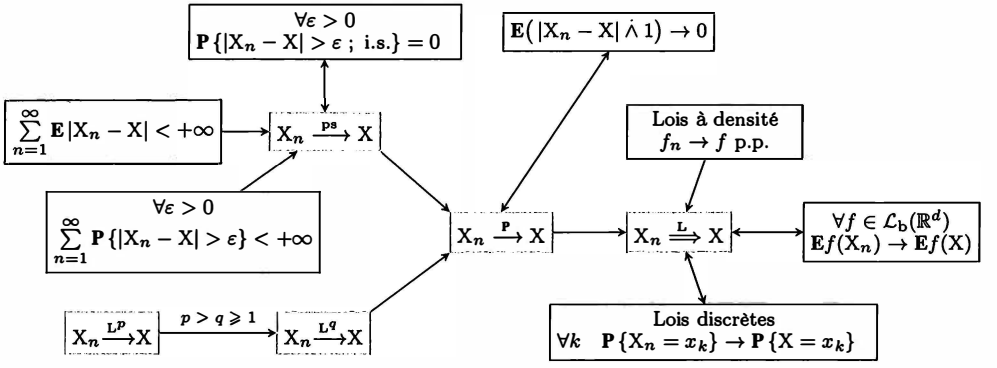
x	**3.7**	**3.8**	**3.9**	**4.0**	**4.5**	**5**
$1 - \mathfrak{N}(x)$	$1,08 \cdot 10^{-4}$	$7,23 \cdot 10^{-5}$	$4,81 \cdot 10^{-5}$	$3,17 \cdot 10^{-5}$	$3,40 \cdot 10^{-6}$	$2,87 \cdot 10^{-7}$

THÉORÈMES D'INTÉGRATION

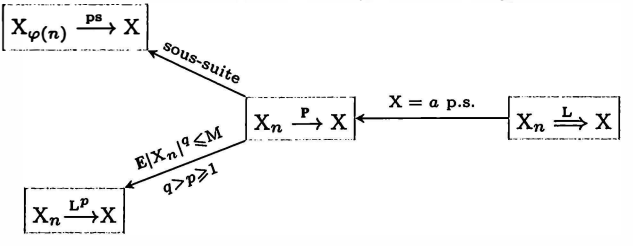
Nom	Hypothèses	Résultat
Convergence monotone	$0 \leqslant X_n \uparrow X$	$\mathbf{E}(X_n) \uparrow \mathbf{E}(X)$
Séries positives	$X_n \geqslant 0$	$\mathbf{E}\Big(\sum_{n=1}^{\infty} X_n\Big) = \sum_{n=1}^{\infty} \mathbf{E}(X_n)$
Lemme de Fatou	$0 \leqslant X_n \to X$	$\mathbf{E}(X) \leqslant \liminf \mathbf{E}(X_n)$
Convergence dominée	$X_n \to X$, $\lvert X_n\rvert \leqslant Y$, avec $Y \in L^1$	$\mathbf{E}\lvert X_n - X\rvert \to 0$ et $\mathbf{E}(X_n) \to \mathbf{E}(X)$
Intégration terme à terme	$\sum \mathbf{E}\lvert X_n\rvert$ converge	$\mathbf{E}\Big(\sum_{n=1}^{\infty} X_n\Big) = \sum_{n=1}^{\infty} \mathbf{E}(X_n)$
Cvce monotone conditionnelle	$0 \leqslant Y_n \uparrow Y$	$\mathbf{E}(Y_n \,\|\, \mathfrak{G}) \to \mathbf{E}(Y \,\|\, \mathfrak{G})$
Cvce dominée conditionnelle	$Y_n \to Y$, $\lvert Y_n\rvert \leqslant T \in L^1$	$\mathbf{E}(Y_n \,\|\, \mathfrak{G}) \to \mathbf{E}(Y \,\|\, \mathfrak{G})$

X discrète	X à densité
$\mathbf{E}(X) = \sum_{k\in K} x_k \mathbf{P}\{X = x_k\}$	$\mathbf{E}(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx$
$\mathbf{E}\varphi(X) = \sum_{k\in K} \varphi(x_k) \mathbf{P}\{X = x_k\}$	$\mathbf{E}\varphi(X) = \int_{-\infty}^{+\infty} \varphi(x)\, f(x)\, dx$
$\mathbf{E}(X) = \sum_{n=0}^{\infty} \mathbf{P}\{X > n\}$	$\mathbf{E}(X) = \int_0^{+\infty} [1 - F(x)]\, dx$
si $X(\Omega) \subset \mathbb{N}$	si $X \geqslant 0$

MODES DE CONVERGENCE



Relations entre les principaux types de convergence.



Réciproques conditionnelles entre les principaux types de convergence.

LANGAGE ENSEMBLISTE ET LANGAGE PROBABILISTE

Notation	Langage ensembliste	Langage probabiliste
ω	élément	événement atomique
A	partie de Ω	événement
$\omega \in A$	ω est élément de A	A est réalisé
$A \subset B$	A inclus dans B	l'événement A implique l'événement B
$A \cup B$	réunion	réalisation de A ou de B
$A \cap B$	intersection	réalisation simultanée de A *et* de B
A^c	complémentaire $\Omega \setminus A$	non-A (événement contraire)
$A \setminus B$	A privé de B	réalisation de A mais pas de B
$\varnothing$	partie vide	événement impossible
Ω		événement certain
$A \cap B = \varnothing$	parties disjointes	A et B sont des événements incompatibles



$\bigcup_{n=0}^{\infty} A_n$	union de $(A_n)_{n\geqslant 0}$	réalisation d'au moins un des événements A_n
$\bigcap_{n=0}^{\infty} A_n$	intersection de $(A_n)_{n\geqslant 0}$	réalisation de tous les événements A_n
$A_n \uparrow A$	la suite $(A_n)_{n\geqslant 0}$ est croissante et $A = \bigcup_{n=0}^{\infty} A_n = \lim_n \uparrow A_n$	
$A_n \downarrow A$	la suite $(A_n)_{n\geqslant 0}$ est décroissante et $A = \bigcap_{n=0}^{\infty} A_n = \lim_n \downarrow A_n$	
$\limsup A_n$	$\bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{A_n \text{ ; i.s.}\}$	réalisation d'une infinité d'événements parmi $A_0, A_1, \ldots$ (i.s. = « infiniment souvent »)
$\liminf A_n$	$\bigcup_{n=0}^{\infty} \bigcap_{k=n}^{\infty} A_k$	réalisation de tous les événements $A_0, A_1, \ldots$ sauf peut-être un nombre fini d'entre eux (= réalisation de tous les événements à partir d'un certain rang)

AXIOMES PROBABILISTES

Un **espace probabilisé** est un triplet $(\Omega, \mathfrak{T}, \mathbf{P})$ où Ω est un ensemble quelconque, $\mathfrak{T}$ une tribu sur Ω et $\mathbf{P}$ une probabilité sur $(\Omega, \mathfrak{T})$, c'est-à-dire une application $\mathbf{P} : \mathfrak{T} \to \mathbb{R}$ vérifiant les propriétés suivantes :

- ⋆ **Axiome 1 :** $\mathbf{P}$ est à valeurs positives ;
- ⋆ **Axiome 2 :** $\mathbf{P}(\Omega) = 1$;
- ⋆ **Axiome 3 :** $\mathbf{P}$ est **σ-additive** : pour toute suite $(A_n)_{n\geqslant 0}$ d'éléments deux à deux disjoints de $\mathfrak{T}$,

$$\mathbf{P}\Big(\sum_{n=0}^{\infty} A_n\Big) = \sum_{n=0}^{\infty} \mathbf{P}(A_n).$$

Sous l'hypothèse des axiomes 1 et 2, et de la propriété d'additivité ($\mathbf{P}(A + B) = \mathbf{P}(A) + \mathbf{P}(B)$), l'axiome 3 ($\sigma$-additivité) est équivalent à chacun des axiomes 3′, 3″ et 3‴ résumés ci-dessous :

Axiome 3	σ-additivité	si les A_n sont disjoints	alors	$\mathbf{P}\Big(\sum_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathbf{P}(A_n)$
Axiome 3′	continuité monotone $\uparrow$	si $B_n \uparrow B$,	alors	$\mathbf{P}(B_n) \uparrow \mathbf{P}(B)$
Axiome 3″	continuité monotone $\downarrow$	si $C_n \downarrow C$,	alors	$\mathbf{P}(C_n) \downarrow \mathbf{P}(C)$
Axiome 3‴	continuité monotone en $\varnothing$	si $D_n \downarrow \varnothing$,	alors	$\mathbf{P}(D_n) \downarrow 0$

TABLES DES PRINCIPALES LOIS

Lois discrètes

Nom	$X(\Omega)$	$\mathbf{P}\{X=k\}$	$\mathbf{E}(X)$	$\mathbf{V}(X)$	$G(t)$	$\varphi(t)$
Uniforme	$[\![1\,;\,n]\!]$	$\dfrac{1}{n}$	$\dfrac{n+1}{2}$	$\dfrac{n^2-1}{12}$	$\dfrac{t-t^{n+1}}{n(1-t)}$	$\dfrac{e^{it}-e^{i(n+1)t}}{n(1-e^{it})}$
Bernoulli $\mathscr{B}(p)$	$\{0,1\}$	$\begin{cases} p_1 = p \\ p_0 = 1-p \end{cases}$	p	$p(1-p)$	$1-p+pt$	$1-p+pe^{it}$
Binomiale $\mathscr{B}(n\,;\,p)$	$[\![0\,;\,n]\!]$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	$(1-p+pt)^n$	$(1-p+pe^{it})^n$
Géométrique $\mathscr{G}(p)$	$\mathbb{N}^*$	$(1-p)^{k-1}\,p$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$	$\dfrac{pt}{1-(1-p)t}$	$\dfrac{pe^{it}}{1-(1-p)e^{it}}$
Poisson $\mathscr{P}(\lambda)$	$\mathbb{N}$	$e^{-\lambda}\,\dfrac{\lambda^k}{k!}$	λ	λ	$e^{\lambda(t-1)}$	$e^{\lambda(e^{it}-1)}$
Hypergéométrique $\mathscr{H}(N,n,p)$		$\dfrac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}}$	np	$npq\,\dfrac{N-n}{N-1}$		

Lois continues

Nom	$X(\Omega)$	Densité $f(x)$	$\mathbf{E}(X)$	$\mathbf{V}(X)$	$\varphi(t)$
Uniforme $\mathscr{U}([a\,;\,b])$	$[a\,;\,b]$	$\dfrac{1}{b-a}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{1}{b-a}\,\dfrac{e^{itb}-e^{ita}}{it}$
Normale $\mathscr{N}(m\,;\,\sigma^2)$	$\mathbb{R}$	$\dfrac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\dfrac{(x-m)^2}{2\sigma^2}\right\}$	m	σ^2	$e^{imt}\,e^{-\sigma^2 t^2/2}$
Exponentielle $\mathscr{E}(\alpha)$	$\mathbb{R}^+$	$\alpha\,e^{-\alpha x}$	$\dfrac{1}{\alpha}$	$\dfrac{1}{\alpha^2}$	$\dfrac{1}{1-it/\alpha}$
Cauchy $\mathscr{C}(m\,;\,a)$	$\mathbb{R}$	$\dfrac{a}{\pi}\,\dfrac{1}{a^2+(x-m)^2}$	$\diagdown$	$\diagdown$	$e^{imt-a\lvert t\rvert}$
χ^2_n	$\mathbb{R}^+$	$\dfrac{1}{2\Gamma(n/2)}\left(\dfrac{x}{2}\right)^{\frac{n}{2}-1} e^{-x/2}$	n	$2n$	$(1-2it)^{-n/2}$
Beta $B(a,b)$	$]0\,;\,1[$	$\dfrac{x^{a-1}\,(1-x)^{b-1}}{B(a,b)}$	$\dfrac{a}{a+b}$	$\dfrac{ab}{(a+b)^2\,(a+b+1)}$	
Gamma $\Gamma(\alpha,\beta)$	$\mathbb{R}_+^*$	$\dfrac{\beta^\alpha}{\Gamma(\alpha)}\,x^{\alpha-1}\,e^{-\beta x}$	$\dfrac{\alpha}{\beta}$	$\dfrac{\alpha}{\beta^2}$	$\left(1-\dfrac{it}{\beta}\right)^{-\alpha}$

Lois approchées

Loi	Paramètres	Loi approchée	Conditions pratiques
Hypergéométrique	$\mathscr{H}(N,n,p)$	$\mathscr{B}(n\,;\,p)$	$N \geqslant 10\,n$
Binomiale	$\mathscr{B}(n\,;\,p)$	$\mathscr{P}(np)$	$p \leqslant 0,1 \quad n \geqslant 30$
Poisson	$\mathscr{P}(\lambda)$	$\mathscr{N}(\lambda\,;\,\lambda)$	$\lambda \geqslant 15$
Binomiale	$\mathscr{B}(n\,;\,p)$	$\mathscr{N}(np\,;\,np(1-p))$	$np \geqslant 15 \quad n(1-p) \geqslant 15$

Les questions les plus importantes de la vie ne sont, pour la plupart, que des problèmes de probabilité.

Pierre Simon de Laplace

Ce livre s'adresse à tous ceux que les probabilités intéressent, d'un point de vue pratique ou théorique, et qui ne sont pas encore spécialistes. Il intéressera tout particulièrement les étudiants, enseignants et praticiens des sciences et techniques.

Ni traité théorique, ni, à l'autre extrême, collection d'exemples épars, il pioche dans ces deux approches pour montrer pas à pas les applications de la théorie. Celle-ci est exposée en trois temps :

- d'abord des chapitres présentant les idées clefs, avec un formalisme minimal, pour construire l'intuition ;
- ensuite, une approche pratique du calcul des probabilités via les variables aléatoires discrètes puis les variables à densité ;
- enfin, la théorie générale de l'intégration selon une mesure de probabilité est expliquée.

L'ouvrage est complété par des chapitres présentant des applications des probabilités à d'autres domaines comme la physique, l'arithmétique ou le calcul numérique. Des méthodes informatiques permettant de simuler les lois usuelles du hasard sont présentées en détail.

Aucune connaissance préalable en probabilités ou en théorie de la mesure n'est requise.

Avec son cheminement progressif, qui combine pédagogie et rigueur, ses très nombreux exemples, ses 165 exercices corrigés et ses chapitres originaux, *Probabilités pour les non-probabilistes* vous accompagnera pendant plusieurs années.



Walter Appel enseigne les mathématiques en classes préparatoires au Lycée du Parc (Lyon). Normalien, agrégé, docteur, il est également l'auteur d'un dictionnaire de mathématiques et d'un ouvrage de référence en mathématiques pour la physique.

38,90 €

ISBN: 978-2-35141-326-5

« J'ai lu avec beaucoup de plaisir cet excellent livre. Walter Appel y réussit un véritable tour de force : présenter une quantité encyclopédique de connaissances, avec pédagogie et précision mathématique. Les notions de la théorie des probabilités y sont introduites de façon graduelle en prenant appui sur des cas simples (variables aléatoires discrètes) pour ensuite présenter des concepts avancés (théorie de la mesure et intégrale de Lebesgue) et finir sur une introduction à des sujets actuels (mouvement brownien, estimation statistique, etc.). Sa présentation moderne et le nombre phénoménal d'exemples et d'applications concrètes placent ce livre au-dessus du lot des références classiques sur la théorie des probabilités. Il est parfait pour des étudiants en Licence souhaitant acquérir les bases sur les probabilités discrètes, ainsi que pour les étudiants de Master qui abordent la théorie de la mesure. Ce sera également le compagnon de route idéal pour les personnes préparant les concours du CAPES et de l'agrégation de mathématiques. »

Gabriel Peyré, chercheur au CNRS, Université Paris-Dauphine

« Le livre de Walter Appel est une excellente introduction aux probabilités. Il pourra notamment servir d'ouvrage de référence aux étudiants en sciences économiques et gestion. Les deux premières parties correspondent bien aux programmes et aux besoins des étudiants jusqu'au M1, tandis que les troisième et quatrième parties, très didactiques, permettront aux étudiants de M2 de se familiariser avec la théorie de l'intégration et ses applications en probabilités. Tous les chapitres de l'ouvrage comportent des exercices corrigés ; la lecture est rendue agréable grâce à une typographie particulièrement soignée. »

Bertrand Crettez, Professeur de sciences économiques à l'Université Panthéon-Assas Paris II

« Ce livre présente une collection impressionnante de thèmes classiques de la théorie des probabilités, ainsi que des thèmes plus originaux qui sont autant d'interactions fascinantes des probabilités avec la science, notamment les liens avec la physique et la théorie c
pédagogiqu
(ou l'amate
là un manu
cours de se
dans ses recherches. »

Ashkan Nikeghbali, Professeur en probabilités à l'Université de Zurich