

Collection de la Direction
des Études et Recherches
d'Électricité de France

ESE
8

Les cours de
l'École Supérieure
d'Électricité

Françoise Brouaye

La modélisation des incertitudes

Probabilités
Signaux et communications
Statistiques

Préface de
Jean Klein

1621

ISSN 0292-6903

Françoise Brouaye
La modélisation des incertitudes
Probabilités - Signaux et communications - Statistiques

EYROLLES



La modélisation des incertitudes

Probabilités
Signaux et communications
Statistiques

COLLECTION DE LA DIRECTION DES ÉTUDES
ET RECHERCHES D'ÉLECTRICITÉ DE FRANCE

La modélisation des incertitudes

Probabilités
Signaux et communications
Statistiques

Françoise Brouaye

Professeur à l'École Supérieure
d'Électricité

Préface de
Jean Klein

EDITIONS EYROLLES
61, Bd Saint-Germain Paris 5^e
1990

PRÉFACE

La formation des ingénieurs d'une grande Ecole scientifique nécessite une solide culture de base, dont les mathématiques représentent une part essentielle.

Le cours de mathématiques doit réaliser un équilibre harmonieux entre les fondements théoriques et leurs applications que l'ingénieur rencontrera durant sa vie professionnelle.

Seule une bonne maîtrise des bases théoriques donne à l'ingénieur les moyens de construire des modèles qu'il pourra ensuite résoudre à l'aide de méthodes pratiques.

Cet ouvrage présente de manière exemplaire cette méthodologie appliquée aux modèles probabilistes que l'on rencontre dans des domaines de plus en plus nombreux et variés notamment la théorie du signal, le filtrage, le codage, la simulation sur ordinateur et naturellement les méthodes statistiques.

La lecture de cet ouvrage est très enrichissante par la diversité des exemples choisis. Il ne s'agit jamais d'appliquer une collection de techniques systématiques mais toujours de réfléchir au modèle et à son domaine de validité.

L'auteur fait comprendre des concepts difficiles à l'aide, quand il le faut, d'une présentation intuitive suivie d'une formalisation rigoureuse.

Il est heureux qu'un tel ouvrage soit à la disposition des étudiants et des ingénieurs et fait désormais partie des « indispensables ».

Jean KLEIN
Professeur à l'Université PARIS VII
Président du Conseil de la Recherche
de l'Ecole Supérieure d'Electricité

ISSN 0292-6903

© 1990 DIRECTION DES ÉTUDES ET RECHERCHES D'ÉLECTRICITÉ DE FRANCE

La loi du 11 mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'Article 41, d'une part que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause, est illicite » (alinéa 1^{er} de l'Article 40).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les Articles 425 et suivants du Code Pénal.

AVANT-PROPOS

La modélisation d'un problème est une activité très difficile. Un exemple est celui des probabilités. Face à une expérience dont l'issue est incertaine, la démarche probabiliste consiste à introduire un modèle aléatoire et utiliser les résultats donnés par la théorie des probabilités. Le calcul des probabilités est né au XVI^e siècle de l'étude des jeux de hasard (az-zhard signifie dé à jouer en arabe) et s'est développé au XVII^e siècle avec la résolution, par Pascal, de problèmes posés par un joueur invétéré, le Chevalier de Méré, ainsi que par la correspondance entre Fermat et Pascal sur le problème des enjeux. Même s'il ne s'agissait que de probabilités combinatoires, le progrès était grand ; on pouvait ainsi « rationaliser » les incertitudes du jeu. Dans les années et les siècles qui ont suivi, les applications se sont multipliées et la théorie s'est étoffée mais ce n'est qu'en 1930 environ, que les axiomes de base de la théorie ont été posés par Kolmogorov.

L'ouvrage présente cette théorie dans les chapitres 1 à 4. Les chapitres 5 à 7 sont consacrés aux modélisations probabilistes rencontrées en théorie du signal. Les deux derniers chapitres 8 et 9 présentent les méthodes statistiques. J'ai volontairement laissé de côté la statistique mathématique (exhaustivité, complétude...) du fait de son aspect théorique ainsi que l'analyse des données du fait de son aspect méthodologique qui nécessite une grande revue de cas pratiques. Les parties « probabilités » et « statistiques » sont traitées relativement indépendamment de la partie « signal ». En effet, un élève-ingénieur par exemple, peut être amené par ses stages puis par son activité professionnelle, à rencontrer des problèmes extrêmement différents, dans des domaines eux aussi très différents. Il apparaît alors important de dissocier la modélisation en elle-même des diverses méthodes pratiques permettant de résoudre les modèles les plus courants.

Cet ouvrage est issu de cours donnés à l'Ecole Supérieure d'Electricité auxquels ont été ajoutés les développements théoriques qu'un lecteur curieux peut souhaiter rencontrer. Il s'adresse en premier lieu aux élèves des grandes écoles scientifiques et aux étudiants de second cycle universitaire. Le niveau requis est celui du premier cycle (classes préparatoires, Deug), essentiellement la connaissance des calculs de limites, du calcul intégral et de l'algèbre linéaire ; certains approfondissements nécessitent quelques connaissances en théorie de la mesure et en théorie des distributions. En ce qui concerne l'utilisation de modèles probabilistes dans des domaines porteurs d'avenir, je citerai deux exemples qui susciteront, je l'espère, des vocations. En premier lieu, celui de l'intelligence artificielle où la gestion des incertitudes par des méthodes probabilistes est en plein développement. En second lieu le domaine de la cryptographie pour lequel des approches probabilistes permettent l'étude de certains algorithmes assurant théoriquement la confidentialité et l'authenticité des messages transmis.

L'objectif de cet ouvrage est de donner des méthodes illustrées d'exemples ainsi que de bonnes bases théoriques. Il est difficile, en effet, de faire des calculs

de probabilités en restant superficiel et en appliquant des formules mécaniquement ; il faut acquérir la tournure d'esprit probabiliste, le langage et... s'exercer. Les bases de probabilité sont indispensables pour aborder la partie « signal » ou la partie « statistique ». Lorsqu'une démonstration n'est pas développée, une référence bibliographique est précisée. La lecture des démonstrations n'est pas indispensable mais conseillée dans la mesure où elles montrent où se situent les difficultés et au prix de quelles hypothèses on les mène à bien. L'« apprenti modélisateur » aura à réfléchir à la signification de telles hypothèses dans les situations réelles. Chaque chapitre est suivi d'une liste d'exercices dont on trouvera le corrigé à la fin de l'ouvrage.

L'auteur

TABLE DES MATIÈRES

Notations	XVII
Chapitre 1. Espaces probabilisés	1
1.1. <i>Expérience aléatoire</i>	1
1.1.1. Epreuves	1
1.1.2. Événements	1
1.1.3. Exemples	3
1.2. <i>Mesure des événements : probabilité</i>	3
1.2.1. Définition	3
1.2.2. Propriétés de P	4
1.2.3. Formules d'analyse combinatoire	6
1.3. <i>Indépendance et probabilité conditionnelle</i>	6
1.3.1. Définition	6
1.3.2. Probabilité conditionnelle	6
1.3.3. Formule de Bayes	8
1.3.4. Généralisation et exemple	8
1.4. <i>Expérience aléatoire de durée illimitée</i>	10
1.4.1. Jeu de pile ou face de durée illimitée	10
1.4.2. Jeu de pile ou face et codage binaire : construction de la mesure de Lebesgue	10
1.5. <i>Chaînes de Markov homogènes</i>	12
1.5.1. Introduction et définitions	12
1.5.2. Exemple	13
1.5.3. Notions fondamentales dans l'étude des chaînes	15
1.5.4. Principaux résultats	15
<i>Exercices</i>	17
Chapitre 2. Variables aléatoires	21
2.1. <i>Premières définitions</i>	21
2.1.1. Loi de X	21
2.1.2. Variables aléatoires réelles	21
2.1.3. Variables aléatoires discrètes	22
2.1.4. Fonction de répartition	23
2.1.5. Variables à densité	23
2.1.6. Exemple en fiabilité	24
2.2. <i>Variables aléatoires réelles : lois usuelles</i>	25
2.2.1. Variables discrètes	25
2.2.2. Variables à densité	28
2.3. <i>Espérance, moments, fonction caractéristique, transformée de Laplace</i>	29
2.3.1. Espérance	29
2.3.2. Fonction caractéristique (variables réelles)	33

2.3.3. Transformée de Laplace et moments (variables réelles)	33
2.3.4. Tableau récapitulatif	34
2.4. Vecteurs aléatoires	35
2.4.1. Loi	35
2.4.2. Lois marginales	35
2.4.3. Exemple de calcul de loi	36
2.4.4. La loi multinomiale	37
2.4.5. Fonction caractéristique	37
2.4.6. Espérance, matrice de covariance	37
2.4.7. Fonction caractéristique et moments	38
2.5. Indépendance	39
2.5.1. Définition	39
2.5.2. Caractérisations	39
2.6. Vecteurs gaussiens	40
2.6.1. Définition et propriétés	40
2.6.2. Fonction caractéristique	41
2.6.3. Indépendance	41
2.6.4. Densité	42
2.7. Somme de variables indépendantes	42
2.7.1. Loi de la somme	42
2.7.2. Fonction caractéristique, transformée de Laplace	43
2.7.3. Gaussiennes, Qui-deux et loi Gamma	43
2.8. Lois conditionnelles	44
2.8.1. Le cas des densités	44
2.8.2. Espérance conditionnelle	45
Exercices	47
Chapitre 3. Théorèmes de convergence	51
3.1. Convergence en moyenne quadratique	51
3.1.1. Définition	51
3.1.2. Lemme de Loève	51
3.2. Convergence en probabilité et convergence presque sûre	52
3.2.1. Inégalité de Bienaymé-Tchebichev	52
3.2.2. Convergence en probabilité	53
3.2.3. Convergence presque sûre	53
3.2.4. Convergence p.s. pour des moyennes de variables de Bernoulli	54
3.3. Loi du tout ou rien et lemme de Borel-Cantelli	55
3.3.1. Loi du tout ou rien	55
3.3.2. Exemples d'événements asymptotiques	55
3.3.3. Lemme de Borel-Cantelli	56
3.4. Loi des grands nombres	57
3.4.1. Loi faible	57
3.4.2. Enoncé de la loi forte	57
3.4.3. Contre-exemple	57
3.4.4. Extension	58
3.4.5. Convergence de la fonction de répartition empirique	58
3.4.6. Un exemple en théorie du signal	59
3.5. Convergence en loi	60
3.5.1. Définition	60

3.5.2. Caractérisations de la convergence en loi	61
3.5.3. Application : processus de Poisson	61
3.5.4. Théorème de la limite centrale	62
3.5.5. Convergence en loi pour des variables binomiales	63
3.5.6. Convergence en loi pour des variables gaussiennes	64
3.5.7. Convergence en loi pour des variables Qui-deux	64
3.6. Récapitulation et comparaison des différents types de convergence	65
Exercices	66
Chapitre 4. Simulation sur ordinateur de phénomènes aléatoires	69
4.1. Loi uniforme : résultats théoriques (cas discret et cas continu)	69
4.1.1. Rappels	69
4.1.2. Cas discret et cas continu	70
4.1.3. Relations entre variables uniformes continues	70
4.2. Générateurs de nombres pseudo-aléatoires	71
4.2.1. Introduction	71
4.2.2. Propriétés arithmétiques des congruences linéaires	71
4.2.3. Génération de 0, 1 pseudo-aléatoires	72
4.3. Simulation de variables aléatoires discrètes	72
4.3.1. Variable prenant un nombre fini de valeurs	72
4.3.2. Loi de Poisson $\mathcal{P}(\lambda)$	73
4.3.3. Loi géométrique	73
4.4. Simulation de variables aléatoires continues	74
4.4.1. Le théorème fondamental et ses applications	74
4.4.2. Loi gaussienne $\mathcal{N}(m, \sigma^2)$	75
4.5. Simulation de lois plus complexes	76
4.5.1. Lois mélanges	76
4.5.2. Lois de Cox	77
4.6. Exemples de problèmes traités par simulation	77
4.6.1. Minimisation d'un coût	77
4.6.2. Calcul d'une intégrale par la méthode de Monte Carlo	78
4.6.3. Simulation de files d'attente	79
4.7. Le mouvement brownien	79
4.8. Remarques	81
Exercices	81
Chapitre 5. Notions probabilistes utilisées en théorie du signal	85
5.1. Processus aléatoires	85
5.1.1. Définitions	85
5.1.2. Exemples	85
5.1.3. Processus du second ordre	88
5.1.4. Processus du second ordre stationnaires	89
5.1.5. Processus gaussien	89
5.2. Bruit blanc	89
5.2.1. A temps discret	89
5.2.2. A temps continu	90
5.2.3. Tableau récapitulatif	92

5.3. Mesure spectrale de puissance pour un processus stationnaire au second ordre	93
5.3.1. Fonction de covariance et fonction d'auto-corrélation	93
5.3.2. Théorème de Bochner-Herglotz	93
5.3.3. Interprétation	94
5.3.4. Transformée de Fourier (théorème de représentation spectrale)	94
5.4. L'ergodisme	97
5.4.1. Historique	97
5.4.2. Les résultats	98
Exercices	99
Chapitre 6. Le filtrage linéaire	103
6.1. Introduction	103
6.1.1. Filtre RC	103
6.1.2. Dérivation	104
6.1.3. Lissage	104
6.1.4. Modulations	104
6.1.5. A temps discret	104
6.1.6. Dans les communications numériques	104
6.1.7. Le théorème des filtres	104
6.2. Filtrage linéaire à temps continu	105
6.2.1. Intégrales stochastiques	105
6.2.2. Filtre de convolution par la fonction h avec un bruit blanc en entrée	107
6.2.3. Filtre de convolution par la fonction h avec en entrée une fonction aléatoire	109
6.2.4. Dérivation d'une fonction aléatoire	110
6.3. Analyse spectrale (cas stationnaire)	110
6.3.1. Formules dans le domaine spectral	110
6.3.2. Contre-exemple : l'équation de Langevin	111
6.4. Corrélation entre l'entrée et la sortie	112
6.4.1. Cas général	112
6.4.2. Cas stationnaire	112
6.4.3. Exemple : le filtre en quadrature	113
6.5. Formule des interférences	114
6.6. Résumé pour le cas stationnaire	115
6.7. Filtrage linéaire à temps discret	116
6.7.1. Adaptation des formules précédentes	116
6.7.2. Les modèles fondamentaux	116
6.7.3. Le modèle AR(p)	117
6.7.4. Le modèle ARMA(p, q)	120
Exercices	123
Chapitre 7. Codage et communications	127
7.1. Introduction	127
7.2. Codages en blocs	129
7.2.1. Contrôle de parité	129
7.2.2. Codes linéaires (n, k), codes systématiques	130
7.2.3. Codes cycliques (n, k)	133
7.2.4. Les codes polynomiaux	135

7.2.5. Compression avec distorsion	137
7.2.6. Codes de Goppa (codes géométriques)	137
7.3. Codages de longueur variable	139
7.3.1. Introduction	139
7.3.2. Code de Fano-Shannon	140
7.3.3. Code d'Huffman	141
7.3.4. Théorème de Shannon	142
7.3.5. Applications	143
7.4. Détection et correction des erreurs	144
Exercices	145
Chapitre 8. Statistiques	149
8.1. Introduction	149
8.1.1. Les domaines d'utilisation	149
8.1.2. Exemple à temps discret	149
8.1.3. A temps continu	151
8.1.4. Premières remarques	151
8.2. L'estimation	152
8.2.1. Vocabulaire statistique	152
8.2.2. Le critère du maximum de vraisemblance	152
8.2.3. Le critère de Bayes	154
8.2.4. Comparaison de ces deux critères	155
8.2.5. Estimateur de variance minimale	156
8.3. Tests paramétriques	157
8.3.1. Vocabulaire et définitions	157
8.3.2. Le critère de Neyman-Pearson	161
8.3.3. Test Bayésien	161
8.3.4. Hypothèses multiples et critère du maximum de vraisemblance	162
8.4. Tests gaussiens (critère de Neyman-Pearson)	162
8.4.1. Introduction	162
8.4.2. Test sur m, σ^2 étant connu	165
8.4.3. Test sur σ^2	165
8.4.4. Test sur m, σ^2 étant inconnu	167
8.4.5. Comparaison de deux échantillons indépendants	167
8.5. Intervalles de confiance	169
8.5.1. Introduction	169
8.5.2. Estimation de la moyenne	169
8.5.3. Estimation de la variance	170
8.5.4. Relation avec les tests	171
Exercices	171
Chapitre 9. Et un peu plus de statistiques	175
9.1. Loi empirique et tests basés sur la loi empirique	175
9.1.1. Loi empirique	175
9.1.2. Test d'ajustement du χ^2 : méthode pratique	177
9.1.3. Test d'ajustement de Kolmogorov-Smirnov (loi complètement spécifiée)	179
9.1.4. Test d'indépendance du χ^2 (grands échantillons)	181
9.2. Tests non paramétriques (petits échantillons)	182

9.2.1. Introduction	182
9.2.2. Test de la médiane et test du signe	183
9.2.3. Le test signe et rang de Wilcoxon	184
9.2.4. Comparaison de deux échantillons indépendants (test de Mann-Whitney)	185
9.2.5. Test d'indépendance entre deux v.a.r. X et Y (Spearman)	186
9.3. Modèles linéaires	187
9.3.1. La régression	187
9.3.2. Introduction à l'analyse de variance	191
Exercices	194
Corrigé des exercices	197
ANNEXES	219
Annexe 1: Théorème de Kolmogorov	219
Annexe 2: Tables statistiques	220
BIBLIOGRAPHIE	231
INDEX	233

NOTATIONS

\mathbb{N}	ensemble des entiers naturels
\mathbb{Z}	ensemble des entiers relatifs
\mathbb{Q}	ensemble des nombres rationnels
\mathbb{R}	ensemble des nombres réels
\mathbb{C}	ensemble des nombres complexes
cv	converge
v.a.	variable aléatoire (chapitre 2)
v.a.r.	variable aléatoire réelle (chapitre 2)
m.q.	moyenne quadratique (chapitre 3)
p.s.	presque sûrement (chapitre 3)
$\mathcal{P}(\lambda)$	loi de Poisson de paramètre λ (chapitre 2)
$\mathcal{B}(n, p)$	loi binomiale de paramètres n et p (chapitre 2)
$\mathcal{N}(m, \sigma^2)$,	ou aussi $\mathcal{N}(m, \sigma^2)$
	loi normale, ou aussi loi gaussienne, d'espérance m et de variance σ^2 (chapitre 2)
χ_n^2	loi de Qui-deux à n degrés de liberté (chapitre 2) ou Khi-deux
$F_{n,m}$	loi de Fisher (chapitre 8)
$E(X)$	espérance de la variable X (chapitre 2)
$\text{var}(X)$	variance de la variable X (chapitre 2)
$\text{cov}(X, Y)$	covariance de X et Y (chapitre 2)
$E(X/Y)$	espérance conditionnelle de la variable X, conditionnellement à la variable Y (chapitre 2)
$1_A(x)$	fonction indicatrice de l'ensemble A (1 si x appartient à A, 0 sinon)
$L^1(\mu)$	espace des v.a. complexes de module intégrable pour la mesure μ (chapitre 2)
$L^2(\mu)$	espace des v.a. complexes de carré intégrable (chapitre 2).

CHAPITRE 1

ESPACES PROBABILISÉS

1.1. EXPÉRIENCE ALÉATOIRE

1.1.1. Epreuves

On dit qu'une expérience est aléatoire lorsqu'en la répétant dans les mêmes conditions, le résultat n'est pas toujours le même ; par exemple, si on tire une carte au hasard dans un jeu de 52 cartes...

Mathématiquement, une expérience aléatoire est décrite par la donnée de l'ensemble des résultats possibles de l'expérience, appelé *ensemble des épreuves* et noté Ω .

Dans l'exemple du jeu de cartes, on prendrait $\Omega = \{1, 2, \dots, 52\}$ (en numérotant les cartes). Ω peut être un ensemble très complexe et difficile à décrire ; imaginons que l'on choisisse une photo au hasard dans un magasin de développement rapide. Le résultat est alors une image avec des paramètres quantitatifs, dimensions, ..., et d'autres qualitatifs : impression que donne l'image, qualité artistique... L'espace Ω est alors un espace abstrait contenant toutes les informations sur l'expérience.

1.1.2. Événements

Il s'agit d'une notion fondamentale en théorie des probabilités. Un événement A est une partie de Ω mais pas toujours n'importe quelle partie de Ω . Dans le deuxième exemple, même si Ω n'est pas très explicite, on peut regarder l'événement « il y a un objet rouge sur la photo » ou « il s'agit d'une photo sur papier mat ».

On peut faire le lien avec les notations de la théorie des ensembles en disant qu'une épreuve ω appartient à l'ensemble A si l'événement A est réalisé lors de l'épreuve ω . On utilisera alors les opérations ensemblistes : passage au complémentaire, intersection et réunion ; en particulier, ω appartient à $A \cup B$ si et seulement si l'un au moins des deux événements A ou B , est réalisé lors de l'épreuve ω .

Ces opérations sur les ensembles interviennent naturellement dans la pratique ; par exemple, pour un système constitué d'éléments en série, l'événement « le système fonctionne » s'exprime comme une intersection :

$$\{\text{le système fonctionne}\} = \{\text{le 1}^{\text{er}} \text{ élé. fonctionne}\} \cap \\ \cap \{\text{le 2}^{\text{e}} \text{ élé. fonctionne}\} \dots \cap \{\text{le } k\text{-ième élé. fonctionne}\}$$

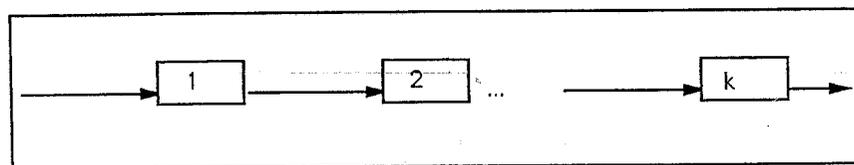


Figure 1.1.

Au contraire, si les éléments sont disposés en parallèle, le système fonctionne dès que l'un des éléments fonctionne et il s'agit d'une réunion.

Les événements peuvent être disjoints ; on dit aussi qu'ils sont incompatibles ou aussi qu'ils s'excluent mutuellement.

Exemple : Considérons une maladie causée par un virus et pour laquelle il y a trois possibilités qui s'excluent mutuellement : être immunisé, être porteur chronique ou n'avoir jamais rencontré le virus. Si on se rend compte que la maladie peut être causée par deux virus, les événements A, B, C qui suivent ne sont pas disjoints :

- A = {le patient est porteur chronique d'au moins un virus }
- B = {le patient est immunisé contre les deux virus }
- C = {il y a au moins un virus que le patient n'a jamais rencontré }

Ce que l'on demande aux événements c'est que, si A et B sont des événements, alors :

- A^c (le complémentaire de A),
- $A \cap B$ (intersection)
- $A \cup B$ (réunion)

soient aussi des événements. Dans la suite on notera $B \setminus A$ l'ensemble $B \cap A^c$ dessiné en pointillé sur la figure 1.2.

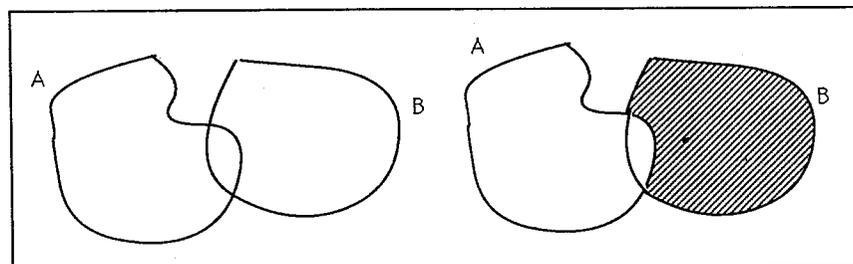


Figure 1.2.

On demande aussi que \emptyset et Ω , événement impossible et événement certain, soient des événements ; et enfin que $\cup A_n$ et $\cap A_n$ soient aussi des événements

lorsque les A_n sont des événements. Cette dernière propriété est importante lorsque Ω n'est plus un ensemble fini. On parle alors de σ -algèbre d'événements :

Définition : On dit que \mathcal{A} est une σ -algèbre si

- 1) $\emptyset \in \mathcal{A}$
- 2) $A \in \mathcal{A}$
 $B \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}, A \cup B \in \mathcal{A}, A \cap B \in \mathcal{A}$
- 3) $\forall n \geq 1 \quad A_n \in \mathcal{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$

1.1.3. Exemples

Exemple 1) La σ -algèbre d'événements, engendrée par une partition.

Si $\Omega = \{1, 2, \dots, 52\}$ et $A_1 = \{3, 6, 9, \dots\}$,
 $A_2 = \{1, 4, 7, \dots\}$, $A_3 = \{2, 5, 8, \dots\}$

En utilisant les opérations de passage au complémentaire, réunion finie, intersection finie, cette partition de Ω engendre la σ -algèbre

$$\mathcal{A} = \{\emptyset, \Omega, A_1, A_2, A_3, A_1 \cup A_2, A_1 \cup A_3, A_2 \cup A_3\}$$

On pourrait faire la même chose avec

$$\Omega = [0, 1] \quad A_1 = [0, 1/3], \quad A_2 =]1/3, 2/3], \quad A_3 =]2/3, 1]$$

Exemple 2) La σ -algèbre engendrée par les ouverts de $[0, 1]$: les boréliens de $[0, 1]$. Il s'agit de la plus petite σ -algèbre contenant tous les ouverts ; c'est aussi la plus petite σ -algèbre contenant tous les intervalles. On remarque qu'un point est un élément de la σ -algèbre,

$$\{x\} = \bigcap_{n=1}^{\infty} \left] x - \frac{1}{n}, x + \frac{1}{n} \right[$$

L'ensemble des rationnels de $[0, 1]$ qui est une réunion dénombrable de points est donc aussi un borélien ; par ailleurs il n'est ni ouvert ni fermé.

1.2. MESURE DES ÉVÉNEMENTS : PROBABILITÉ

Avant d'effectuer l'expérience on ne sait pas si l'événement A sera réalisé ou non ; la donnée d'une probabilité permet de mesurer les chances pour que A se réalise.

1.2.1. Définition

P est une probabilité sur (Ω, \mathcal{A}) si P est une application de \mathcal{A} dans

$[0, 1]$ vérifiant :

- 1) $P(\Omega) = 1$
- 2) $P\left(\bigcup_n A_n\right) = \sum_n P(A_n)$ si les événements A_n sont 2 à 2 disjoints et forment une suite finie ou dénombrable.

1.2.2. Propriétés de P

— En utilisant 2) avec A et A^c , on obtient

$$P(A) + P(A^c) = P(\Omega) = 1$$

relation que l'on écrit

$$P(A^c) = 1 - P(A)$$

En particulier

$$P(\emptyset) = 0$$

— Si A et B sont deux événements, $A \cup B$ se décompose en réunion disjointe :

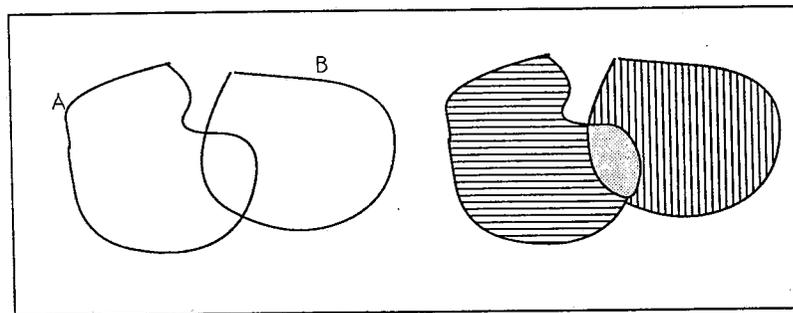


Figure 1.3.

On en déduit la formule utile,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

— Si $A \supset B$, $A = B \cup (A \setminus B)$

et

$$P(A \setminus B) = P(A) - P(B) \text{ si } A \supset B$$

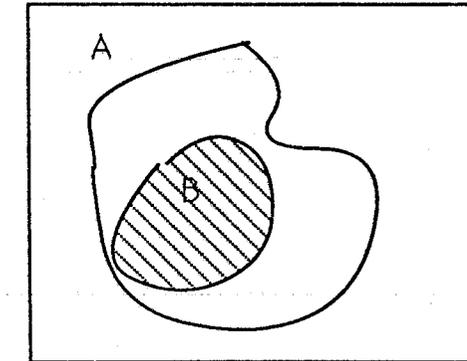


Figure 1.4.

— Si la suite d'événements A_n est croissante

$$\bigcup A_n = A_1 \cup (A_2 \setminus A_1) \cup \dots = A_1 \cup B_2 \cup \dots \cup B_n \cup \dots$$

avec

$$\begin{aligned} B_2 &= A_2 \setminus A_1 \\ B_3 &= A_3 \setminus (A_1 \cup B_2) = A_3 \setminus A_2 \\ &\dots \\ B_n &= A_n \setminus A_{n-1} \\ &\dots \end{aligned}$$

Les événements B_n sont deux à deux disjoints et on obtient

$$P\left(\bigcup_n A_n\right) = \lim_n P(A_n)$$

On a le même résultat dans le cas d'une suite décroissante : la probabilité de l'intersection des A_n est la limite des probabilités.

Exemples :

Exemple 1)

$$\begin{aligned} \Omega &= \{1, 2, \dots, 52\} \\ \mathcal{A} &= \mathcal{F}(\Omega) \text{ et } P(\{k\}) = 1/52 \end{aligned}$$

On a alors

$$P(A) = \text{card}(A)/52$$

Tous les points de Ω ont la même probabilité : on dit que la probabilité est uniforme.

Exemple 2)

$$\Omega = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$$

avec les probabilités respectives $1/9, 2/9, 2/9, 4/9$.

Ω est un ensemble à 4 éléments ; P n'est pas uniforme.

1.2.3. Formules d'analyse combinatoire

Ces formules sont maintenant au programme des classes de Terminales ; en voici un bref rappel :

Le nombre de permutations d'un ensemble à n éléments, est égal à $n!$.

Le nombre de façons de choisir p objets parmi n , noté C_n^p , est

$$C_n^p = \frac{n!}{p! (n-p)!} \quad \text{pour } 0 \leq p \leq n$$

Le nombre de façons de choisir p objets parmi n , en tenant compte de l'ordre est noté A_n^p

$$A_n^p = \frac{n!}{(n-p)!} \quad \text{pour } 0 \leq p \leq n$$

1.3. INDÉPENDANCE ET PROBABILITÉ CONDITIONNELLE

1.3.1. Définition

A et B sont dits indépendants si $P(A \cap B) = P(A)P(B)$, A_1, A_2, \dots, A_n sont dits indépendants si

$$P(B_1 \cap B_2 \cap \dots \cap B_n) = P(B_1)P(B_2) \dots P(B_n)$$

dès que B_i est égal à A_i ou son complémentaire.

On remarquera la différence avec « les événements sont 2 à 2 indépendants » (voir E.1.5).

1.3.2. Probabilité conditionnelle

On va maintenant définir la probabilité conditionnelle notée $P(A/B)$, probabilité pour que A se réalise sachant que B est réalisé. Regardons sur un exemple où tous les chiffres sont fictifs. Si dans la population la proportion de

personnes qui fument est de 60 %, que l'on connaît la proportion de cancers de la gorge parmi les fumeurs et parmi les non-fumeurs, on en déduit le schéma :

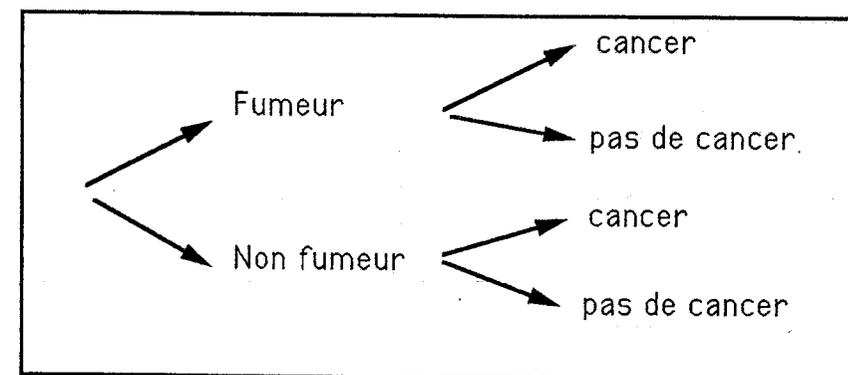


Figure 1.5.

La proportion de cancers dans la population totale est égale à la probabilité pour qu'une personne tirée au hasard soit atteinte ; de même la proportion de cancers parmi la population des fumeurs est égale à la probabilité pour qu'une personne tirée au hasard parmi les fumeurs soit atteinte ;

$$\text{proport. de cancers parmi les fumeurs} = \frac{\text{Nbre de fumeurs atteints/Nbre total}}{\text{Nbre de fumeurs/Nbre total}}$$

Définition : Si $P(B) \neq 0$, on pose $P(A/B) = \frac{P(A \cap B)}{P(B)}$.

$P(A/B)$ est appelée probabilité conditionnelle de A sachant B.

Remarque : On utilisera souvent la relation

$$P(A \cap B) = P(A/B)P(B)$$

Application : Construction d'une probabilité par la donnée de probabilités conditionnelles. Considérons un labyrinthe avec une entrée et 9 sorties ; on ne fait pas de retour en arrière et à chaque bifurcation on tire la direction à pile ou face (Fig. 1.6) :

Dans ces conditions

$$\begin{aligned} P(\text{on sort par 1}) &= 1/2 \\ P(\text{on sort par 2}) &= (1/2)(1/2)(1/2) \\ &\dots \end{aligned}$$

Sur cet exemple, les tirages effectués à chaque bifurcation sont indépendants.

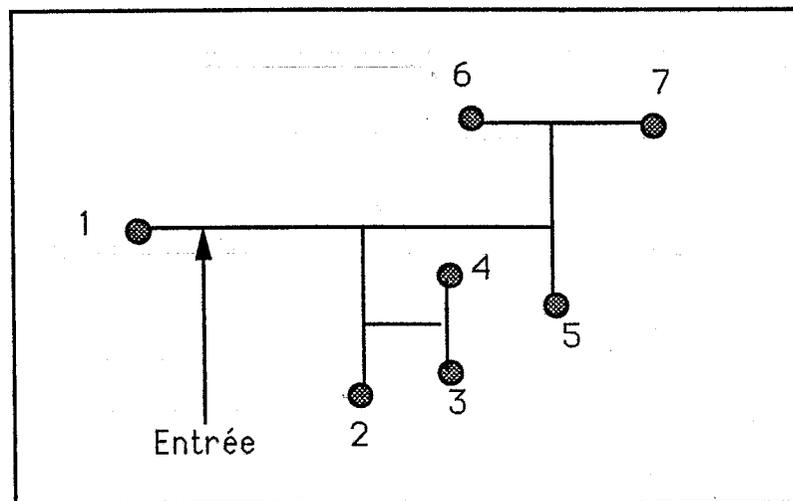


Figure 1.6.

1.3.3. Formule de Bayes

En utilisant la décomposition en événements disjoints

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

on a

$$P(A/B) = \frac{P(B/A)P(A)}{P(B/A)P(A) + P(B/A^c)P(A^c)}$$

On remarque que cette formule renverse le conditionnement.

1.3.4. Généralisation et exemple

Si B_1, B_2, \dots, B_k forment une partition de Ω , comme sur la figure 1.7 où $k = 5$,

alors on peut écrire

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_k)$$

et

$$P(B_1/A) = \frac{P(A/B_1)P(B_1)}{P(A/B_1)P(B_1) + \dots + P(A/B_k)P(B_k)}$$

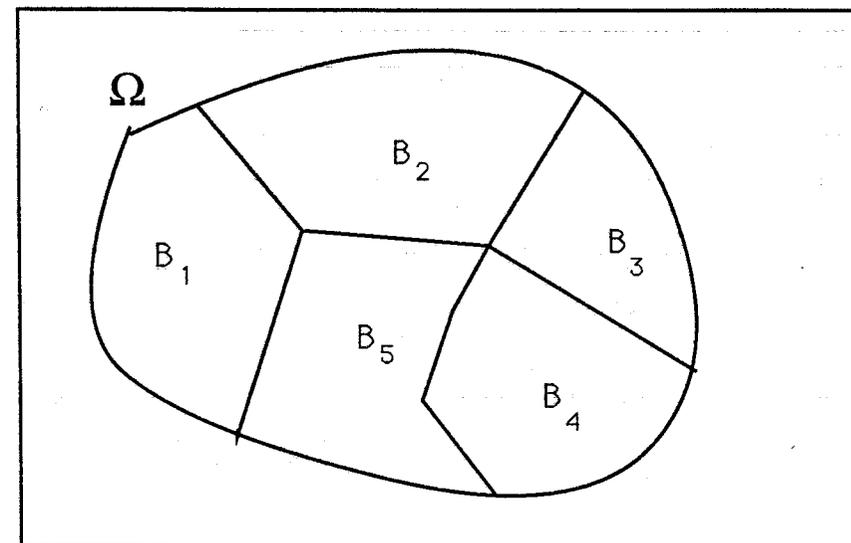


Figure 1.7.

Exemple : Une image peut comporter soit un objet rouge, soit un objet jaune soit un objet vert et ceci avec les probabilités respectives 0,5, 1/3, 1/6. Un système permet de détecter automatiquement, mais cette détection n'est pas certaine et se fait avec la probabilité 0,1 si il s'agit de l'objet rouge 0,2 et 0,9 respectivement pour les objets jaune et vert. On fait un essai et la réponse est positive ; quel objet a le plus de chances d'être présent ? Pour traiter cet exemple on pose

$$B_1 = \{\text{présence de l'objet rouge}\}$$

$$B_2 = \{\text{présence de l'objet jaune}\}$$

$$B_3 = \{\text{présence de l'objet vert}\}$$

L'énoncé nous donne

$$\begin{aligned} P(B_1) &= 0,5 & P(B_2) &= 1/3 & P(B_3) &= 1/6 \\ P(A/B_1) &= 0,1 & P(A/B_2) &= 0,2 & P(A/B_3) &= 0,9 \end{aligned}$$

On en déduit $P(B_1/A) = 3/16$, $P(B_2/A) = 4/16$ et $P(B_3/A) = 9/16$. C'est donc l'objet vert qui a le plus de chances d'être présent.

1.4. EXPÉRIENCE ALÉATOIRE DE DURÉE ILLIMITÉE

1.4.1. Jeu de pile ou face de durée illimitée

En notant 0 si on obtient pile et 1 si on obtient face, l'espace des épreuves au temps n varie avec n :

- au temps 1 $\Omega_1 = \{0,1\}$
- au temps 2 $\Omega_2 = \{00, 01, 10, 11\} = \{0, 1\}^2$
- au temps n $\Omega_n = \{0, 1\}^n$

La probabilité au temps $n + 1$ est compatible avec la probabilité au temps n au sens où c'est une extension. Dans ce cas, un théorème de Kolmogorov nous dit qu'il existe une probabilité sur $\{0, 1\}^{\mathbb{N}}$ qui prolonge toutes ces probabilités, théorème que nous admettrons ; l'énoncé est donné en annexe ; pour la démonstration, voir par exemple, « Bases mathématiques du calcul des probabilités » de J. Neveu Ed. Masson).

1.4.2. Jeu de pile ou face et codage binaire : construction de la mesure de Lebesgue

Si x est un nombre réel de $[0, 1]$, on le code en binaire de la manière suivante :

- si $x \in [0, 1/2[$ le premier chiffre binaire est 0
- si $x \in [1/2, 1]$ le premier chiffre binaire est 1

pour le deuxième chiffre binaire on applique le procédé ci-dessus à $2(x - x_1)$ en appelant x_1 ce premier chiffre binaire :

- $x \in [0, 1/4[\rightarrow 00$
- $x \in [1/4, 1/2[\rightarrow 01$
- $x \in [1/2, 3/4[\rightarrow 10$
- $x \in [3/4, 1[\rightarrow 11$

Cela correspond à la division des intervalles :

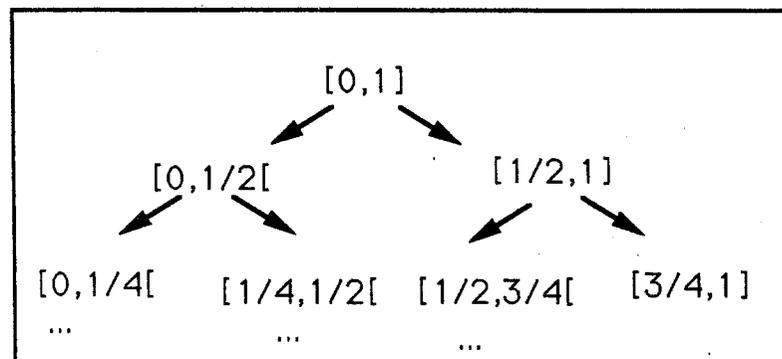


Figure 1.8.

On continue indéfiniment pour obtenir le codage binaire d'un point x quelconque.

Réciproquement, étant donné une suite infinie de 0, 1 notée x_n , on peut lui associer un nombre réel entre 0 et 1 en posant

$$\phi(x_1, x_2, \dots, x_n) = \sum_{n=1}^{\infty} \frac{x_n}{2^n}$$

Ces deux applications sont inverses l'une de l'autre mais un point peut avoir plusieurs codages binaires : $1/2$ s'obtient comme l'image de la suite $(1, 0, \dots)$ et de la suite $(0, 1, 1, \dots)$. De manière générale, les suites

$$\begin{aligned} &(x_1, x_2, \dots, x_n, 0, 1, 1, \dots) \\ \text{et} &(x_1, x_2, \dots, x_n, 1, 0, 0, \dots) \end{aligned}$$

ont même image. Retirons les suites qui n'ont que des 0 à partir d'un certain rang, soit S cet ensemble de suites ; l'application devient une bijection sur $[0, 1]$. Si on met sur $\{0, 1\}^{\mathbb{N}}$, la probabilité du jeu de pile ou face, l'ensemble des suites qui n'ont que des 0 à partir d'un certain rang, est de probabilité nulle :

$$S = \bigcup_n \{ \text{ens. des suites qui n'ont que des 0 à partir du rang } n \}$$

S est donc dénombrable ; la probabilité de n'importe quel point de S est nulle :

$$P(x_1, x_2, \dots, x_n, 0, 0, \dots) = \lim_{p \rightarrow \infty} 1/2^p = 0$$

On construit alors une mesure sur $[0, 1]$, appelée mesure de Lebesgue en posant

$$\lambda(A) = P(\text{ens. des suites } x_1, \dots, x_n \dots \text{ telles que } \phi(x_1, x_2, \dots) \in A)$$

Avec cette définition,

$$\lambda([0, 1/2]) = \lambda([0, 1/2]) = P(\text{ens. des suites qui commencent par 0}) = 1/2$$

et de façon générale

$$\lambda([k/2^n, (k+1)/2^n]) = 1/2^n$$

Tout intervalle peut être approché par des intervalles du type $[k/2^n, (k+1)/2^n[$ et on obtient

$$\lambda([a, b]) = \lambda([a, b]) = \lambda(]a, b]) = b - a$$

La mesure de Lebesgue peut ensuite être prolongée à \mathbb{R} par translation.

D'autre part si on considère les rationnels de $[0, 1]$, on a

$$\lambda(\mathbb{Q} \cap [0, 1]) = \sum \lambda(q_n) = \sum 0 = 0$$

Si on tire un nombre réel au hasard entre 0 et 1, on a 100 % de chances d'obtenir un irrationnel. L'événement « on a tiré un nombre rationnel entre 0 et 1 » est possible mais improbable ; l'expérience est en fait impossible à réaliser puisqu'il faudrait manipuler autre chose que des nombres décimaux.

1.5. CHAÎNES DE MARKOV HOMOGENES

1.5.1. Introduction et définitions

Il s'agit d'un modèle simple qui permet de décrire une évolution aléatoire à temps discret. L'expérience est de durée illimitée comme dans le jeu de Pile ou Face. L'évolution ne dépend du passé que par le dernier instant.

Soit E un ensemble fini ou dénombrable dont les éléments sont notés e_1, e_2, \dots , appelé ensemble des états.

Définitions :

1) Soit π une matrice $k \times k$ si E a k éléments, infinie si E est infini ; π est appelé une probabilité de transition entre les états si ses éléments $\pi_{i,j}$ vérifient

$$\begin{cases} 0 \leq \pi_{i,j} \leq 1 & \forall i, j \\ \sum_j \pi_{i,j} = 1 & \forall i \end{cases}$$

2) On dit que le système évolue suivant une chaîne de Markov homogène si il existe une probabilité de transition π telle que

$$\begin{aligned} P(\text{état } b \text{ à l'instant } n+1 / \text{état } a_0 \text{ à l'instant } 0, \text{ état } a_1 \text{ à l'instant } 1, \dots, \\ \text{état } a_n \text{ à l'instant } n) \\ = P(\text{état } b \text{ à l'instant } 1 / \text{état } a_n \text{ à l'instant } 0) \\ = \pi_{a_n, b}, \text{ si on confond } \pi_{e_i, e_j} \text{ avec } \pi_{i, j} \end{aligned}$$

Remarque : Pour une définition plus générale de système markovien (espace d'états non nécessairement discret, espace de temps non nécessairement discret, chaînes non homogènes dans le temps) on pourra consulter le livre de Dynkin « Théorie des processus markoviens » Ed. Dunod.

Représentation : Lorsque l'espace d'états E est fini, on peut représenter l'évolution du système de façon concise par un graphe où figurent les probabilités de transition entre états ; l'exemple le plus simple est celui d'un canal de transmission qui transmet des symboles binaires 0, 1 avec une proportion ε d'erreurs. Si les symboles binaires passent dans plusieurs canaux successifs,

l'évolution est markovienne suivant le graphe :

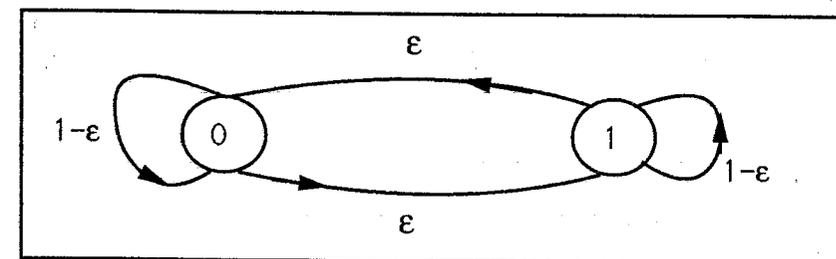


Figure 1.9.

1.5.2. Exemple

Voici un exemple en marketing : on s'intéresse à la part de marché d'un certain produit. Pour une personne prise au hasard, on considère l'expérience associée aux différents achats :

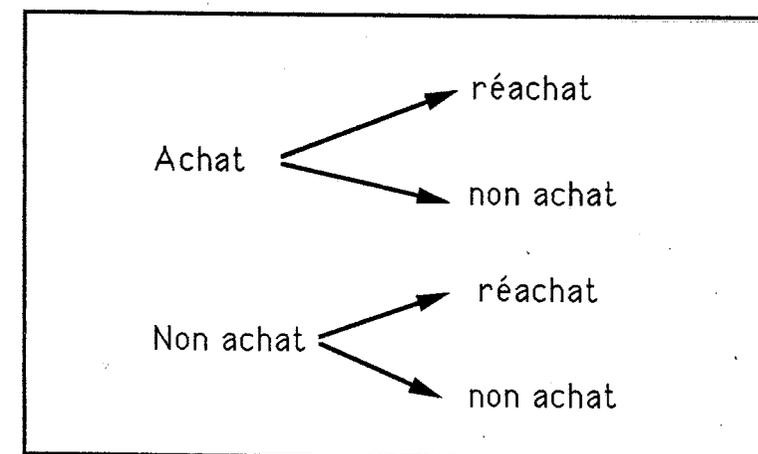


Figure 1.10.

Si l'on note 1 lorsque la personne achète le produit et 0 lorsqu'elle ne l'achète pas, on peut décrire l'expérience par une suite de 0 et 1.

Supposons que la probabilité de passage d'un état, 0 ou 1, à un état, 0 ou 1, ne dépende pas de l'instant où il a lieu ni des instants précédents (on dit que l'on a une chaîne de Markov homogène). On peut alors faire des calculs de probabilité sur l'instant N , regarder l'évolution...

Supposons p et p' non égaux à 0 ou 1 (dans le cas contraire l'évolution est plus simple à étudier). La matrice de transition entre états est la matrice 2×2 suivante :

$$\pi = \begin{pmatrix} p' & 1-p' \\ p & 1-p \end{pmatrix}$$

Appelons μ la loi initiale c'est-à-dire

$${}^t\mu = (\mu_0, \mu_1) = (P(0 \text{ à l'instant } 0), P(1 \text{ à l'instant } 0))$$

μ_1 est la part de marché initiale.

Par récurrence sur n , on montre que la loi à l'instant n vérifie l'équation matricielle

$$(P(0 \text{ à l'instant } n), P(1 \text{ à l'instant } n)) = {}^t\mu \pi^n$$

En utilisant les valeurs propres et vecteurs propres de π on montre alors la convergence de la loi de l'état à l'instant n vers (ν_0, ν_1) :

$$\nu_0 = \frac{1}{1 + \frac{1-p'}{1-p}}, \quad \nu_1 = \frac{1-p'}{1 + \frac{1-p'}{1-p}}$$

${}^t\nu$ vérifie la propriété d'invariance : ${}^t\nu \pi = {}^t\nu$ et la convergence a lieu quelle que soit la loi initiale ce qui veut dire que la part de marché se stabilise sur ν_1 quelle que soit la part de marché initiale. Certaines des propriétés rencontrées sur cet exemple sont toujours vraies ; plus précisément on a toujours la relation matricielle,

$$\begin{aligned} (P(e_1 \text{ à l'instant } n), P(e_2 \text{ à l'instant } n), \dots, P(e_k \text{ à l'instant } n)) = \\ = (P(e_1 \text{ à l'instant } 0), P(e_2 \text{ à l'instant } 0), \dots) \pi^n \end{aligned}$$

et si la loi de l'état à l'instant n converge, alors la limite est une probabilité invariante par π :

$$(\nu_1, \nu_2, \dots, \nu_k) = (\nu_1, \nu_2, \dots, \nu_k) \pi$$

La convergence n'a pas toujours lieu ; en effet si E a 3 éléments et que l'on prend la matrice de transition π :

$$\pi = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

la probabilité invariante par π est $(1/3, 1/3, 1/3)$; si on part du premier état e_1 , $P(e_1 \text{ à l'instant } n)$ est égal à 1 si n est multiple de 3, à 0 sinon.

1.5.3. Notions fondamentales dans l'étude des chaînes

L'une des notions fondamentales est la façon dont on passe d'un état à l'autre, plus précisément on définit ρ_{xy} par

$$\rho_{xy} = P(\text{on passe par } y/\text{état } x \text{ à l'instant } t = 0)$$

En particulier, ρ_{xx} est la probabilité de revenir en x sachant que l'on est parti de x .

On dit que deux états communiquent si ρ_{xy} et ρ_{yx} sont tous deux non nuls.

Un état x est dit de période d si en partant de x les retours à x n'ont lieu qu'à des instants multiples de d au sens où

$$d = \text{pgcd} \{n \text{ tel que } n \geq 1 \text{ et } P(\text{état } x \text{ à l'instant } n/\text{état } x \text{ à l'instant } t = 0)\}$$

L'état x est dit *apériodique* si $d = 1$.

1.5.4. Principaux résultats

1.5.4.1. Etats récurrents et états transients

Proposition : Si ρ_{xx} est égal à 1 alors on passe, avec probabilité 1, une infinité de fois par x en partant de x et x est appelé un état récurrent. Si ρ_{xx} est strictement inférieur à 1 alors, en partant de x , on ne passe, avec probabilité 1, qu'un nombre fini de fois par x ; dans ce cas x est appelé un état transient.

Démonstration succincte : On note $P_y(A)$ la probabilité d'un événement A conditionnellement au fait que l'on part de y . On montre alors l'égalité,

$$P_y(\text{on passe au moins une fois par } x) = \rho_{yx}$$

On considère maintenant l'événement « on passe au moins deux fois par x ».

$$A \text{ y fixé, } A \mapsto P_y(A) \text{ étant une probabilité,}$$

$$P_y(\text{on passe au moins 2 fois par } x) =$$

$$= P_y(\text{on passe au moins 2 fois par } x/\text{on passe au moins 1 fois par } x) \\ \times P_y(\text{on passe au moins 1 fois par } x)$$

On montre alors l'égalité

$$\begin{aligned} P_y(\text{on passe au moins 2 fois par } x/\text{on passe au moins 1 fois par } x) = \\ = P(\text{on passe au moins 1 fois par } x/\text{on part de } x) \\ = \rho_{xx} \end{aligned}$$

et on en déduit

$$P_y(\text{on passe au moins 2 fois par } x) = \rho_{yx} \rho_{xx}$$

On montre de même

$$P_y(\text{on passe au moins } n \text{ fois par } x) = \rho_{yx} \rho_{xx}^{n-1}$$

de sorte que

$$\begin{aligned} P_y(\text{on passe une infinité de fois par } x) &= \\ &= P_y\left(\bigcap_{n=1}^{\infty} \{\text{on passe au moins } n \text{ fois par } x\}\right) \\ &= \lim_{n \rightarrow +\infty} P_y(\text{on passe au moins } n \text{ fois par } x) \\ &= \begin{cases} 0 & \text{si } \rho_{xx} < 1 \\ \rho_{yx} & \text{si } \rho_{xx} = 1 \end{cases} \end{aligned}$$

ce qui donne le résultat avancé en prenant $y = x$.

1.5.4.2. Classification des états

Un état x est soit récurrent soit transient ; il a une période notée d ; si il est récurrent, l'instant de premier retour est fini avec la probabilité 1 mais peut être d'espérance infinie (voir chap. 2 pour la définition de l'espérance).

Proposition : Soient x et y deux états distincts ; si x et y communiquent alors ils sont de même nature, plus précisément :

- ils sont tous deux récurrents ou tous deux transients ;
- ils ont même période ;
- si ils sont récurrents, les temps de retour sont tous deux d'espérance finie ou tous d'espérance infinie.

On trouvera la démonstration dans « Processus stochastiques » de Chrétienne et Faure.

On en déduit la classification des états en états transients regroupés dans l'ensemble E_T et états récurrents qui communiquent :

$$E = E_T \cup (C_1 \cup \dots \cup C_i \dots)$$

Chaque classe C_j est constituée d'états qui communiquent ; deux classes C_i, C_j distinctes ne communiquent pas et lorsqu'on rentre dans une classe C_j on n'en sort plus.

Par exemple, si on considère la matrice de transition suivante :

$$\pi = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/8 & 0 & 7/8 & 0 \\ 1/4 & 1/4 & 0 & 0 & 1/4 & 1/4 \\ 0 & 0 & 3/4 & 0 & 1/4 & 0 \\ 0 & 1/5 & 0 & 1/5 & 1/5 & 2/5 \end{pmatrix}$$

On trouve que les états 4 et 6 sont transients

- 1 et 2 sont récurrents et communiquent
- 3 et 5 sont récurrents et communiquent.

1.5.4.3. Probabilité invariante

On trouvera dans l'ouvrage cité ci-dessus le théorème suivant

Théorème : Si tous les états sont :

- récurrents,
- apériodiques
- avec un temps moyen de retour fini

alors il existe une probabilité invariante unique et quel que soit l'état de départ, lorsque n tend vers l'infini, la loi de l'état à l'instant n se stabilise sur cette probabilité invariante.

EXERCICES

E.1.1. En suivant les notations du paragraphe 1.1.2, on considère le système suivant, constitué de trois branches en parallèle.

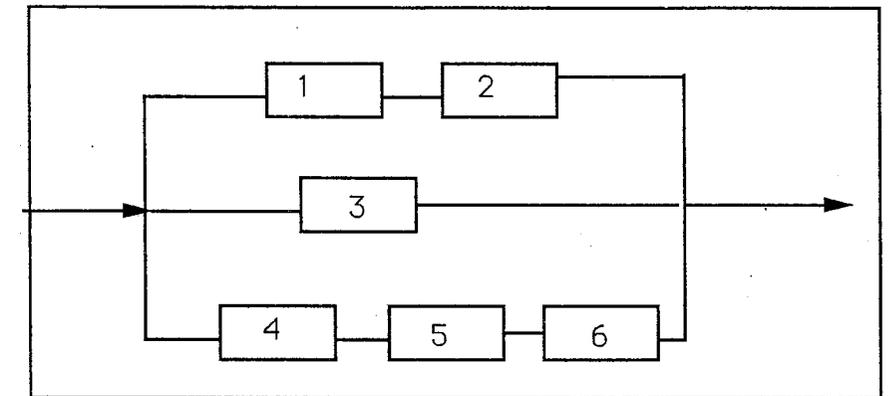


Figure E.1.1.

Exprimer l'événement $A = \{\text{le système fonctionne}\}$ en utilisant les événements $B_i = \{\text{l'élément } i \text{ fonctionne}\}$.

E.1.2. On tire une carte au hasard dans un jeu de 52 cartes. Les événements

$$\begin{aligned} A &= \{ \text{la carte représente un personnage} \} \\ B &= \{ \text{il s'agit d'un carreau} \} \\ C &= \{ \text{la couleur est noire} \} \end{aligned}$$

Sont-ils disjoints ? Calculer leur probabilité respective ainsi que celles des intersections 2 à 2.

E.1.3. On reprend l'exercice E.1.2. Les événements A et B sont-ils indépendants ?

E.1.4. Montrer que si A et B sont indépendants alors A et B^c sont indépendants ainsi que A^c et B, A^c et B^c.

E.1.5. Montrer que si A₁, A₂, ..., A_k sont indépendants alors A₁, A₂, ..., A_{k-1} sont indépendants (ce qui montre la cohérence de la définition).

E.1.6. Soient a, b, c des réels entre 0 et 1. A quelle condition peut-on les interpréter comme des probabilités conditionnelles :

$$a = P(A/A \cup B) \quad b = P(B/B \cup C) \quad c = P(C/A \cup C)$$

où A, B, C sont des événements deux à deux disjoints ?

On rencontre ce problème en Marketing : a est la préférence du client pour un objet lorsqu'il est mis en présence de 2 objets...

E.1.7. Deux boîtes B₁, B₂ contiennent des boules noires et des boules rouges : 10 noires, 5 rouges pour B₁, 3 noires, 7 rouges pour B₂. On choisit une boîte au hasard puis une boule au hasard dans la boîte. Quelle est la probabilité d'obtenir une rouge ?

E.1.8. Un nouveau test est utilisé pour déceler une maladie. On constate que lorsque le patient est malade, le test est positif dans 90 % des cas alors que si le patient n'est pas malade, il est négatif dans 85 % des cas. Un patient se présente. Le test est positif. Quelle est la probabilité pour qu'il soit atteint si on sait qu'il y a 5 % de personnes atteintes dans la population.

E.1.9. On tire deux cartes au hasard dans un jeu de 32 cartes. Calculer la probabilité d'obtenir deux rois conditionnellement au fait qu'il y a une carte rouge et une carte noire.

E.1.10. On considère un système constitué de deux appareils en parallèle (de sorte que le système fonctionne dès que l'un des appareils fonctionne). Chaque appareil est muni d'un système de détection de panne. Le système de détection n'est pas parfait mais il est fiable, c'est-à-dire qu'il ne tombe pas en panne.

Le réparateur quant à lui ne peut réparer qu'un appareil à la fois.

Tracer le graphe de transition entre états, analogue à la figure 1.9 sans indiquer les probabilités, si on appelle état 1 « les 2 appareils fonctionnent », état 2 « un appareil est en panne et la panne est détectée », état 3 « un appareil est en panne

et la panne n'est pas détectée », état 4 « les deux appareils sont en panne ». (Quand on est dans l'état 4, le système ne fonctionne plus et il y a alarme.)

E.1.11. On considère un réseau de stations qui émettent des informations sur un canal partagé. Le temps est discret : $t = 0, 1, \dots$. A chaque instant, les stations essaient ou non d'envoyer un message et le canal ne transmet qu'un message à la fois. Dans le protocole ALOHA, si il y a conflit pour l'accès au canal, les stations réessaient chacune de transmettre leur message avec la probabilité p et ceci jusqu'au succès final (p est un paramètre). On appelle a_j la probabilité de j nouvelles arrivées (indépendantes du temps). Montrer que l'on a une évolution markovienne, si on appelle état k « il y a k messages en attente qui essaient de passer sur le canal », de probabilité de transition π donnée par :

$$\begin{cases} \pi_{0,j} &= a_j ; \quad j = 0, 1, \dots \\ \pi_{i,i-1} &= a_0 i p (1-p)^{i-1} \\ \pi_{i,i} &= a_1 (1-p)^i + a_0 [1 - i p (1-p)^{i-1}] \\ \pi_{i,i+1} &= a_1 [1 - (1-p)^i] \\ \pi_{i,i+j} &= a_j \quad \text{si } j \geq 2 \end{cases}$$

CHAPITRE 2

VARIABLES ALÉATOIRES

2.1. PREMIÈRES DÉFINITIONS

Une variable aléatoire X est une application de Ω dans un ensemble de valeurs, qui est « régulière » vis-à-vis des événements, c'est-à-dire :

$$X: \Omega \rightarrow E$$

E étant muni d'une σ -algèbre, \mathcal{B} , on veut que $X^{-1}(A)$ (événement) soit un événement :

$$\text{si } A \in \mathcal{B}, \quad X^{-1}(A) = \{\omega / X(\omega) \in A\} = \{X \in A\} \in \mathcal{A}$$

si $E = \mathbb{R}$ on choisira la σ -algèbre des boréliens, c'est-à-dire tous les ensembles obtenus à partir des intervalles ouverts par les opérations habituelles de passage au complémentaire, intersection finie ou dénombrable, réunion finie ou dénombrable.

On dit aussi que X est une *application mesurable* de (Ω, \mathcal{A}) dans (E, \mathcal{B}) .

Exemple : On lance un dé et X est le numéro sorti.

2.1.1. Loi de X

C'est la donnée des probabilités $P(X \in A)$ où A varie dans la σ -algèbre de l'espace d'arrivée, \mathcal{B} .

2.1.2. Variables aléatoires réelles

X est une application mesurable de Ω dans \mathbb{R} qui est muni de la σ -algèbre des boréliens. Dans la pratique, toutes les applications que l'on considère sont mesurables et on ne vérifie pas cette propriété ; à ce sujet on pourra consulter le livre de Neveu.

Exemples d'événements :

$$\{\omega / 1/2 < X(\omega) \leq 2,5\} \text{ qui est noté aussi } \{1/2 < X \leq 2,5\}$$

$$\{\omega / X(\omega) \text{ est rationnel}\} \text{ qui est noté aussi } \{X \text{ rationnel}\}.$$

2.1.3. Variables aléatoires discrètes

L'ensemble des valeurs prises par X est discret ; ce sera généralement un ensemble fini, \mathbb{N} ou \mathbb{Z} . Dans le cas fini ou dénombrable, la loi de X est la donnée des probabilités $P(X = x_i)$ que l'on présente dans un tableau :

valeurs	x_1	...	x_n
probabilités	$P(X = x_1)$...	$P(X = x_n)$

Exemples : On lance trois fois un dé et on note X la somme des points ; X est aléatoire entre 3 et 18 ; on montre facilement l'égalité $P(X = k) = P(X = 21 - k)$ en changeant le 6 en 1, le 5 en 2...

x	3	4	5	6...	7	8	9	10	11	...
probabilités	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$	$\frac{27}{216}$...

On joue à pile ou face n fois et on appelle X le nombre de piles :

x	0	1	...	k	...	N
probabilités	$\frac{1}{2^N}$	$\frac{N}{2^N}$...	$C_N^k \frac{1}{2^N}$...	$\frac{1}{2^N}$

On joue à pile ou face jusqu'à ce que l'on obtienne pile et on appelle X le nombre de tirages effectués :

x	0	1	...	N	...
probabilités	$\frac{1}{2}$	$\frac{1}{4}$...	$\frac{1}{2^N}$...

2.1.4. Fonction de répartition

Pour une variable aléatoire réelle, en abrégé v.a.r., c'est la fonction F de \mathbb{R} dans $[0, 1]$ définie par,

$$F(t) = P(X < t)$$

Proposition : F est une fonction croissante, continue à gauche et admettant en tout point une limite à droite égale à $P(X \leq t)$. F(t) tend vers 0 lorsque t tend vers $-\infty$ et vers 1 lorsque t tend vers $+\infty$.

Démonstration laissée au lecteur.

Lorsque la variable est à valeurs dans \mathbb{R}^2 ou plus généralement dans \mathbb{R}^n , on définit la fonction de répartition bien qu'elle soit moins utile que dans le cas réel : c'est une fonction de n variables ; en dimension 2, notons (X, Y) la variable aléatoire à valeurs dans \mathbb{R}^2 ; la fonction de répartition est alors définie par

$$F(t, s) = P(X < t \text{ et } Y < s)$$

Nous admettrons le résultat suivant :

Proposition : Soit X une v.a.r. ; il est équivalent de connaître la loi de X ou de connaître sa fonction de répartition. Le résultat reste vrai pour une v.a. à valeurs dans \mathbb{R}^n .

2.1.5. Variables à densité

On sous-entend densité par rapport à la mesure de Lebesgue sur \mathbb{R} ou sur \mathbb{R}^n . Soit X une variable à valeurs dans \mathbb{R} ou \mathbb{R}^n . La variable est dite à densité si il existe une fonction f(x) à valeurs positives, telle que

$$A \in \mathcal{B} \Rightarrow P(X \in A) = \int_A f(x) dx$$

La figure 2.1 illustre le cas où $A = [a, b]$.

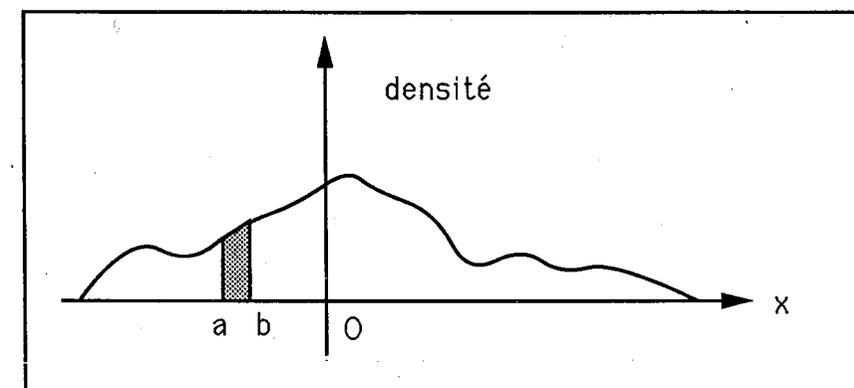


Figure 2.1.

Lorsque la variable est à valeurs dans \mathbb{R}^n il s'agit d'une intégrale multiple. En utilisant les résultats de la théorie de la mesure, de Lebesgue, on a :

Proposition : Soit X une v.a.r. à densité f ; alors F est dérivable et $F'(t) = f(t)$ sauf éventuellement sur un ensemble de points t qui forment un ensemble de mesure de Lebesgue nulle ; si f est continue au point t alors F est dérivable au point t et

$$F'(t) = f(t)$$

Dans la pratique, pour trouver la densité d'une variable réelle on calcule la fonction de répartition puis on dérive.

Remarque : On peut construire une fonction F de \mathbb{R} dans $[0, 1]$ croissante, continue, et une variable aléatoire réelle dont ce soit la fonction de répartition sans que la variable soit à densité (contre exemple construit par Lebesgue) ; néanmoins dans la pratique, les variables seront discrètes, ou à densité, ou un mélange des deux. On rencontre le mélange discret continu dans les problèmes de file d'attente : si un client se présente à un guichet, le temps d'attente avant d'être servi est nul si la file est vide.

2.1.6. Exemple en fiabilité

On s'intéresse à la probabilité de l'événement

$$B_t = \{ \text{le dispositif est non défectueux entre } 0 \text{ et } t \}$$

On appelle fonction de fiabilité, $R(t)$, la probabilité de cet événement :

$$R(t) = P(B_t)$$

On constate que $1 - R(t)$ est une fonction de répartition :

$$1 - R(t) = P(\text{le temps de vie sans défaillance est inférieur à } t)$$

Supposons que la quantité

$$\frac{1}{\Delta t} P(\text{le dispositif tombe en panne entre } t \text{ et } t + \Delta t / \text{il n'est pas tombé en panne entre } 0 \text{ et } t)$$

admet une limite lorsque Δt tend vers 0 ; cette limite est alors appelée taux de défaillance instantanée et notée $\lambda(t)$. Alors,

$$\begin{aligned} \lambda(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \cdot \frac{P(B_{t+\Delta t}^c \cap B_t)}{P(B_t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{R(t) - R(t + \Delta t)}{\Delta t R(t)} \\ &= - \frac{R'(t)}{R(t)} \end{aligned}$$

$R(t)$ est donc solution d'une équation différentielle et vérifie la condition initiale $R(0) = 1$. On en déduit $R(t)$ à l'aide de $\lambda(t)$:

$$R(t) = \exp \left(- \int_0^t \lambda(u) du \right)$$

Un système sans usure serait un système pour lequel $\lambda(t)$ est constant (loi exponentielle) ; on utilise aussi la loi de Weibull :

$$R(t) = \exp \left(- \left(\frac{t}{\alpha} \right)^\beta \right)$$

2.2. VARIABLES ALÉATOIRES RÉELLES : LOIS USUELLES

2.2.1. Variables discrètes

Variable de Bernoulli

p est un paramètre entre 0 et 1

$$X = 0 \text{ ou } 1$$

$$P(X = 0) = p \text{ et } P(X = 1) = 1 - p$$

Loi binomiale $\mathcal{B}(n, p)$

p est un paramètre entre 0 et 1

$$X = 0 \text{ ou } 1, \dots, \text{ ou } n$$

$$P(X = k) = C_n^k p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n$$

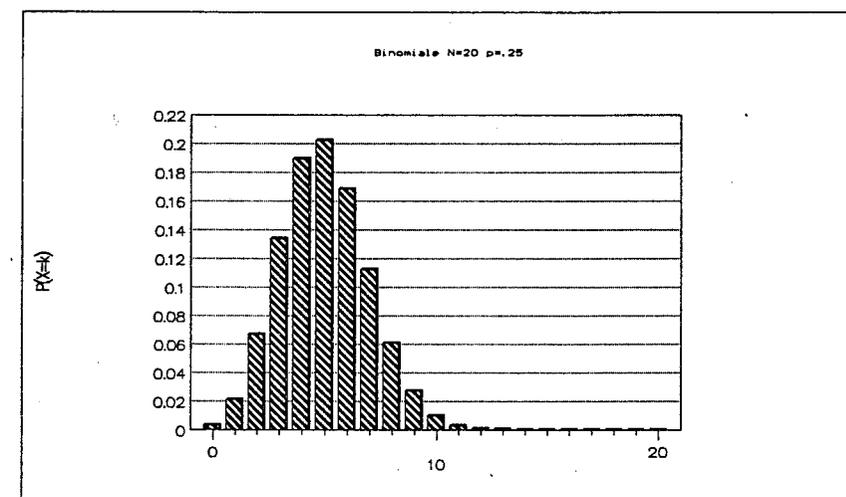


Figure 2.2.

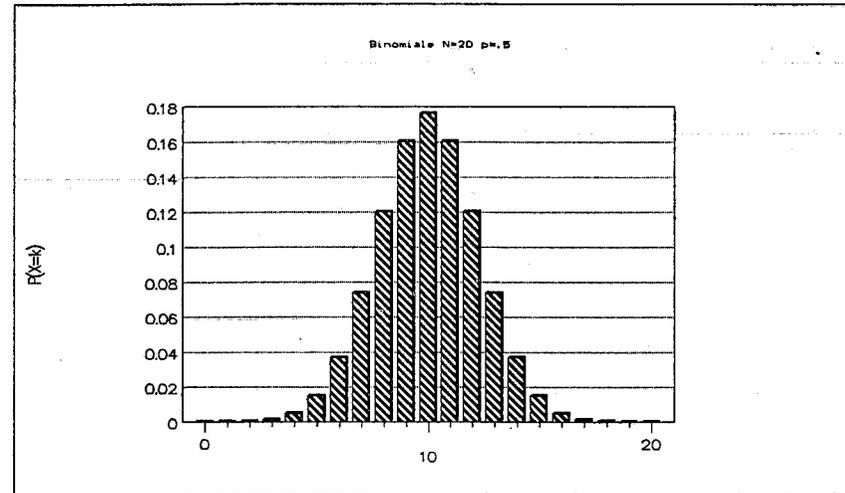


Figure 2.3.

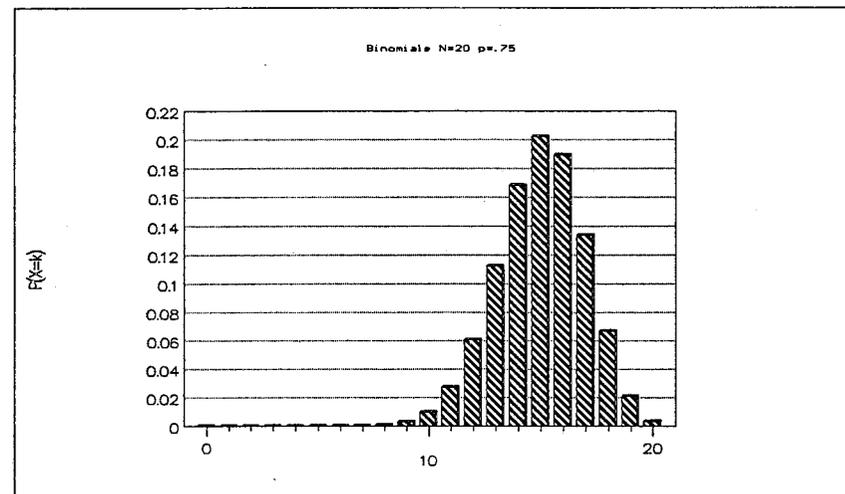


Figure 2.4.

Loi de Poisson $\mathcal{P}(\lambda)$

λ est un paramètre strictement positif

$$X = 0, 1, \dots, n, \dots$$

$$P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

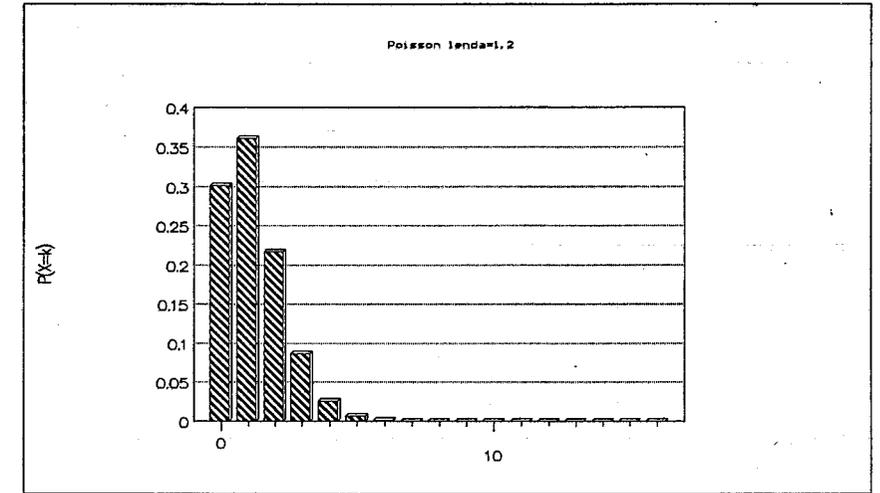


Figure 2.5.

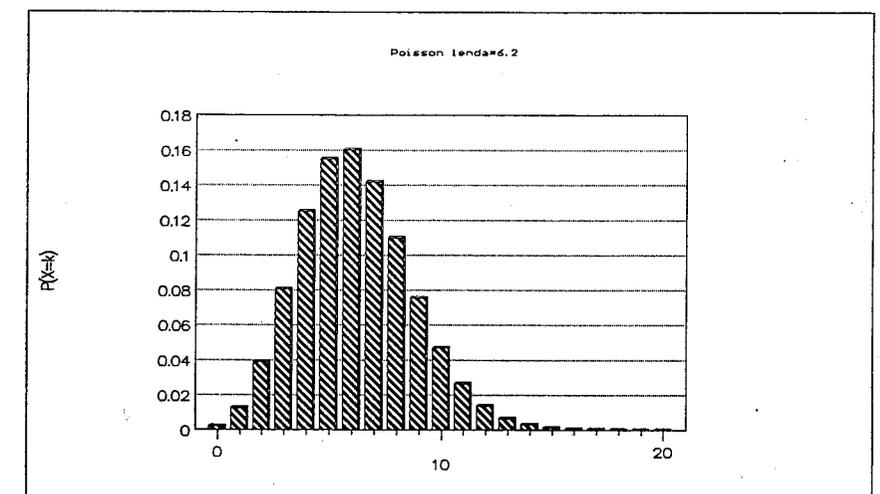


Figure 2.6.

Loi géométrique

p est un paramètre entre 0 et 1

$$X = 0, 1, \dots$$

$$P(X = k) = p(1-p)^{k-1}$$

2.2.2. Variables à densité

Loi exponentielle

λ est un paramètre strictement positif ; il s'agit de la loi ayant pour densité

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Loi uniforme sur un intervalle

La loi uniforme sur l'intervalle borné $[a, b]$ a pour densité

$$f(x) = \begin{cases} 1/(b-a) & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases}$$

Par translation et homothétie la variable reste une variable de loi uniforme ; on s'en sert notamment pour passer d'une variable uniforme sur $[0, 1]$ à une variable uniforme sur n'importe quel intervalle.

Loi gaussienne

C'est une densité qui dépend de deux paramètres réels m et σ , σ étant strictement positif ; la loi est notée $\mathcal{N}(m, \sigma^2)$ et appelée aussi loi normale :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}$$

On admet aussi le cas extrême où $\sigma = 0$: la variable est constante et égale à m .

On montre facilement que si X suit la loi $\mathcal{N}(m, \sigma^2)$ alors $\frac{X-m}{\sigma}$ suit la loi $\mathcal{N}(0, 1)$ appelée loi normale centrée réduite ; de même, par translation et homothétie la variable reste gaussienne :

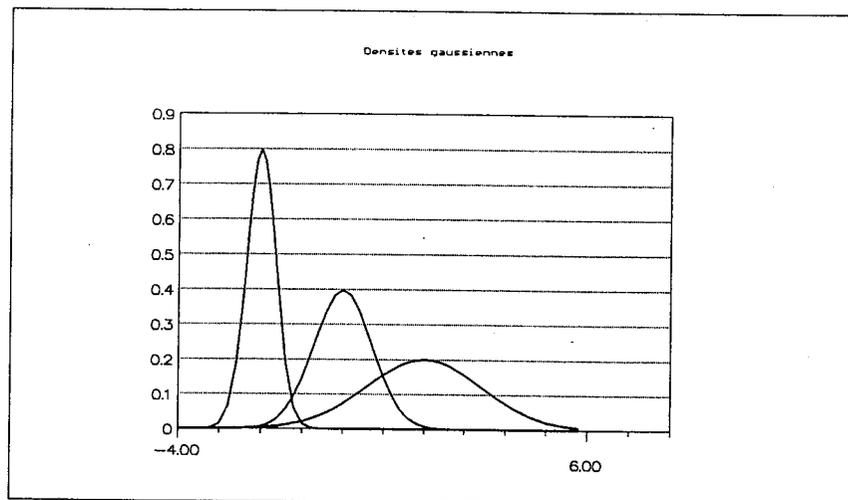


Fig. 2.7. — resp. $m = -2, \sigma = 0,5$; $m = 0, \sigma = 1$; $m = 2, \sigma = 2$.

Loi de Cauchy

C'est la loi de densité

$$f(x) = \frac{1}{\pi(1+x^2)}$$

On rencontre la loi de Cauchy lorsqu'on considère $\tan(\theta)$ où θ est aléatoire de loi uniforme sur $]-\pi/2, \pi/2[$.

Loi Gamma $\gamma(a, \lambda)$ où a et λ sont des paramètres strictement positifs

$$f(x) = \begin{cases} \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

Loi de Erlang (cas particulier de la loi Gamma)

$$f(x) = \begin{cases} \frac{\lambda^{n+1}}{n!} x^n e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

Loi Bêta de paramètres a, b (strictement positifs)

$$f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} & \text{si } x \in]0, 1[\\ 0 & \text{sinon} \end{cases}$$

Loi log-normale

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{1}{x} e^{-\frac{1}{2\sigma^2}[\ln(x)-m]^2} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

Loi de Rayleigh

$$f(x) = \begin{cases} x e^{-\frac{x^2}{2}} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$$

2.3. ESPÉRANCE, MOMENTS, FONCTION CARACTÉRISTIQUE, TRANSFORMÉE DE LAPLACE

2.3.1. Espérance

Soit X une variable aléatoire à valeurs réelles ou complexes.

2.3.1.1. Définition

Si l'intégrale sur Ω , $\int |X(\omega)| dP(\omega)$, est finie, on appelle espérance de X la

valeur de l'intégrale :

$$\int_{\Omega} X(\omega) dP(\omega)$$

L'intégrale est écrite sous une forme « abstraite » mais on comprendra mieux sa signification sur l'exemple ci-dessous.

Exemple fondamental : C'est celui où la variable ne prend qu'un nombre fini de valeurs distinctes a_1, a_2, \dots, a_n

$$\begin{aligned} X &= a_1 & \text{sur } A_1 \\ & a_2 & \text{sur } A_2 \\ & \vdots & \\ & a_n & \text{sur } A_n \end{aligned}$$

$$E(X) = \int_{\Omega} X(\omega) dP(\omega) = a_1 P(A_1) + \dots + a_n P(A_n)$$

Espace $L^1(P)$: C'est l'ensemble des variables aléatoires $X: \Omega \rightarrow \mathbb{C}$ qui admettent une espérance. D'après les résultats de la théorie de la mesure, c'est un espace vectoriel, c'est-à-dire que l'espérance est linéaire,

$$E(aX + bY) = aE(X) + bE(Y) \quad \text{si } X \text{ et } Y \text{ admettent une espérance}$$

Variable constante ou p.s. constante

On dit que X est presque sûrement constante, en abrégé p.s., si il existe une constante a telle que $P(X \neq a) = 0$.

Si X est p.s. égale à a , alors X a pour espérance la constante a .

2.3.1.2. Mesure image et calcul pratique de l'espérance

On appelle mesure image de P par X la probabilité μ sur \mathbb{R} définie par

$$\mu(A) = P(X \in A) = P(\omega / X(\omega) \in A)$$

Si X est une variable à densité f , alors $d\mu(x) = f(x) dx$ (mesure à densité).

Si X est une variable de Bernoulli, alors la mesure μ est une combinaison de deux mesures de Dirac, l'une au point 0, l'autre au point 1 :

$$\mu = p\delta_1 + (1-p)\delta_0$$

Nous admettrons le résultat suivant :

Proposition : X admet une espérance si et seulement si l'intégrale

$$\int_{\mathbb{R}} x d\mu(x)$$

est absolument convergente et dans ce cas l'espérance $E(X)$ est égale à cette intégrale :

$$E(X) = \int_{\mathbb{R}} x d\mu(x)$$

en particulier si X est une variable à densité f , X a pour espérance

$$E(X) = \int_{\mathbb{R}} xf(x) dx$$

lorsque cette intégrale est absolument convergente. D'autre part, si h est une application de \mathbb{R} dans \mathbb{C} , mesurable, la variable $h(X)$ a pour espérance

$$E(h(X)) = \int_{\mathbb{R}} h(x) d\mu(x)$$

lorsque cette intégrale est absolument convergente.

	Variable discrète	Variable à densité
$E(X)$	$\sum_k x_k P(X = x_k)$	$\int_{\mathbb{R}} xf(x) dx$
$E(h(X))$	$\sum_k h(x_k) P(X = x_k)$	$\int_{\mathbb{R}} h(x) f(x) dx$

Exemple de la loi binomiale :

$$P(X = k) = C_N^k p^k (1-p)^{N-k}; \quad 0 \leq k \leq N$$

$$E(X) = \sum_{k=0}^N kP(X = k)$$

Posons $G(t) = \sum_{k=0}^N t^k P(X = k)$; nous avons alors les relations

$$G(t) = [tp + (1-p)]^N$$

$$G'(t) = \sum_{k=0}^N kt^{k-1} P(X = k) = Np[tp + (1-p)]^{N-1}$$

soit

$$E(X) = G'(1) = Np$$

2.3.1.3. Moments (variables réelles)

$E(X)$ est appelé le moment d'ordre 1 de X ; $E(X^2)$ est appelé le moment d'ordre 2 et on peut définir, si il existe, le moment d'ordre n comme $E(X^n)$ lorsque l'intégrale

$$\int_{\Omega} X^n(\omega) dP(\omega)$$

est absolument convergente. Regardons le moment d'ordre 2 :

$$E(X^2) = \int_{\Omega} X^2(\omega) dP(\omega) = \int_{\mathbb{R}} x^2 d\mu(x)$$

Dire que X admet un moment d'ordre 2 équivaut à dire que X appartient à $L^2(P)$. En appliquant l'inégalité de Cauchy-Schwarz sur $L^2(P)$, à X et à la variable constante égale à 1, on obtient :

« si X admet un moment d'ordre 2 alors X admet un moment d'ordre 1 et $|E(X)| \leq \sqrt{E(X^2)}$ ».

En utilisant des inégalités de convexité on montre de même que si X admet un moment d'ordre n alors X admet des moments d'ordre 1, 2, ..., $n-1$.

Variance : Si le moment d'ordre 2 existe, alors on sait que le moment d'ordre 1, c'est-à-dire l'espérance, existe et on peut définir la variance par

$$\text{var}(X) = E[(X - m)^2] \quad \text{où } m = E(X)$$

En particulier une variable X est de variance nulle si et seulement si elle est p.s. constante.

Le lecteur pourra démontrer les relations suivantes :

$$\text{var}(X) = E(X^2) - E(X)^2$$

$$\text{var}(aX + b) = a^2 \text{var}(X)$$

2.3.2. Fonction caractéristique (variables réelles)

Soit X une variable aléatoire réelle. Si t est un paramètre réel, e^{itX} est une variable à valeurs complexes de module 1, donc appartenant à $L^1(P)$; on peut donc calculer son espérance (voir le tableau en fin de chapitre pour les lois usuelles).

Définition : On appelle fonction caractéristique de la v.a.r. X la fonction ϕ de \mathbb{R} dans \mathbb{C} définie par $\phi(t) = E(e^{itX})$.

Le calcul de la fonction caractéristique est facile pour les variables qui ne prennent qu'un nombre fini de valeurs distinctes. Pour les variables réelles à densité le calcul se fait généralement en utilisant le théorème des résidus ou des propriétés des fonctions analytiques car

$$E(e^{itX}) = \int_{\mathbb{R}} e^{itx} f(x) dx$$

Nous admettrons le résultat suivant qui justifie le nom de fonction caractéristique :

Proposition : Il est équivalent de connaître la loi de X ou de connaître sa fonction caractéristique, c'est-à-dire que si deux v.a.r. ont même fonction caractéristique alors elles ont même loi.

Fonction caractéristique et moments : En utilisant les théorèmes de dérivation de la théorie de Lebesgue, si X admet un moment d'ordre 1 alors ϕ est continûment dérivable et on a la relation

$$\phi'(0) = iE(X)$$

Plus généralement si X admet un moment d'ordre n alors ϕ est n fois continûment dérivable et on a

$$\phi^{(n)}(0) = i^n E(X^n)$$

La réciproque est fautive mais si ϕ est infiniment dérivable au point 0 alors X admet des moments à tout ordre.

2.3.3. Transformée de Laplace et moments (variables réelles)

Pour les variables de loi exponentielle et pour un certain nombre de lois sur \mathbb{R}_+ la transformée de Laplace est plus pratique que la fonction caractéristique; par contre, alors que la fonction caractéristique d'une v.a.r. existe toujours, la transformée de Laplace pose des problèmes d'existence.

Définition : Soit X une v.a.r.; lorsque $E(e^{-zX})$ existe on l'appelle transformée de Laplace de X et on la note $\hat{\mu}(z)$.

Le domaine d'existence est une bande verticale de \mathbb{C} contenant l'axe imaginaire.

Exemples :

Binomiale $\mathcal{B}(n, p)$

$$\hat{\mu}(z) = [p e^z + (1 - p)]^n; \quad z \in \mathbb{C}$$

Loi exponentielle de paramètre λ

$$\hat{\mu}(z) = \frac{\lambda}{\lambda + z}; \quad \text{Re}(z) > -\lambda$$

Théorème : Si $\hat{\mu}(z)$ existe sur un domaine du type $a < \text{Re}(z) < b$ avec $a < 0 < b$ alors

$\hat{\mu}(z)$ est holomorphe sur ce domaine et on a

$$\forall n \in \mathbb{N}, \quad E(X^n) = (-1)^n \frac{d^n}{dz^n} \hat{\mu}(z) \Big|_{z=0}$$

Applications : Pour une v.a.r. de loi exponentielle on en déduit

$$E(X^n) = \frac{n!}{\lambda^n}$$

Remarque : Pour les variables X à valeurs entières on définit la fonction génératrice $G(t) = E(t^X)$ (sous réserve de l'existence).

2.3.4. Tableau récapitulatif

	Espérance	Variance	Fonction caractéristique	Transformée de Laplace
Binomiale	np	$np(1-p)$	$[1 - p + p e^{it}]^n$	$[1 - p + p e^{-z}]^n$
Poisson	λ	λ	$e^{\lambda(e^{it}-1)}$	$e^{\lambda(e^{-z}-1)}$
Exponentielle	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - it}$	$\frac{\lambda}{\lambda + z}$
Gaussienne	m	σ^2	$e^{im - t^2 \frac{\sigma^2}{2}}$	$e^{-zm + z^2 \frac{\sigma^2}{2}}$

2.4. VECTEURS ALÉATOIRES

2.4.1. Loi

On considère dans ce paragraphe une variable aléatoire à valeurs dans \mathbb{R}^n que nous appellerons vecteur aléatoire et que nous noterons $X = (X_1, X_2, \dots, X_n)$. Cette notion est essentielle dès que l'on veut faire des calculs de probabilité faisant intervenir plusieurs variables aléatoires. Par exemple, un client a le choix entre deux guichets ; la file d'attente au guichet 1 est vide mais le service est généralement plus long qu'au guichet 2 qui a presque terminé avec son client ; quel guichet a-t-il intérêt à choisir ?

Loi : La loi du vecteur est la donnée de toutes les quantités $P(X \in A)$ où A est un borélien de \mathbb{R}^n (la σ -algèbre des boréliens est la σ -algèbre engendrée par tous les ensembles du type $]a_1, b_1[\times]a_2, b_2[\times \dots \times]a_n, b_n[$).

La loi du vecteur nous renseigne sur le comportement des composantes et sur les « interactions » entre composantes.

Densité : Le vecteur admet une densité si il existe une fonction f de \mathbb{R}^n dans \mathbb{R}_+ telle que pour tout A borélien de \mathbb{R}^n

$$P(X \in A) = \int_A f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

(intégrale multiple)

2.4.2. Loïs marginales

Si on connaît la loi d'un vecteur aléatoire on en déduit la loi de n'importe quelle composante X_i du vecteur en écrivant

$$P(X_1 \in A_1) = P(X \in A_1 \times \mathbb{R} \times \dots \times \mathbb{R})$$

$$P(X_2 \in A_2) = P(X \in \mathbb{R} \times A_2 \times \mathbb{R} \times \dots \times \mathbb{R})$$

...

— Dans le cas d'une variable discrète on obtient les lois des composantes dans les marges en écrivant par exemple (voir le tableau de probabilités ci-dessous)

$$P(X = -1) = P(X = -1 \text{ et } Y = 0) + P(X = -1 \text{ et } Y = 1) + P(X = -1 \text{ et } Y = 2)$$

		Y			
		0	1	2	
X	-1	0,08	0,04	0,125	0,245
	1	0,25	0,125	0,38	0,755
		0,33	0,165	0,505	

— Dans le cas d'une loi à densité $f(x_1, x_2, \dots, x_n)$, la loi de X_1 , par exemple, s'obtient en intégrant sur x_2, \dots, x_n :

$$\begin{aligned} P(X_1 \in A_1) &= \int_{x_1 \in A_1} \int_{x_2, \dots, x_n} f(x_1, x_2, \dots, x_n) dx_2 \dots dx_n \\ &= \int_{A_1} f_1(x_1) dx_1 \end{aligned}$$

avec $f_1(x_1) = \int f(x_1, x_2, \dots, x_n) dx_2 \dots dx_n$

Proposition : Si X a pour densité $f(x_1, x_2, \dots, x_n)$ alors les composantes du vecteur ont chacune une densité : la densité de X_i s'obtient en intégrant f sur les $n - 1$ variables différentes de x_i .

2.4.3. Exemple de calcul de loi

Soit (X, Y) un vecteur aléatoire à valeurs dans \mathbb{R}^2 de densité $f(x, y)$. Cherchons la loi du minimum des deux composantes $\inf(X, Y)$; à cet effet, calculons la fonction de répartition :

$$\begin{aligned} F(t) &= P(\inf(X, Y) < t) \\ &= P(X < Y \text{ et } X < t) + P(Y < X \text{ et } Y < t) \\ &= \iint_D f(x, y) dx dy \end{aligned}$$

où D est la partie non hachurée, sur la figure 2.8.

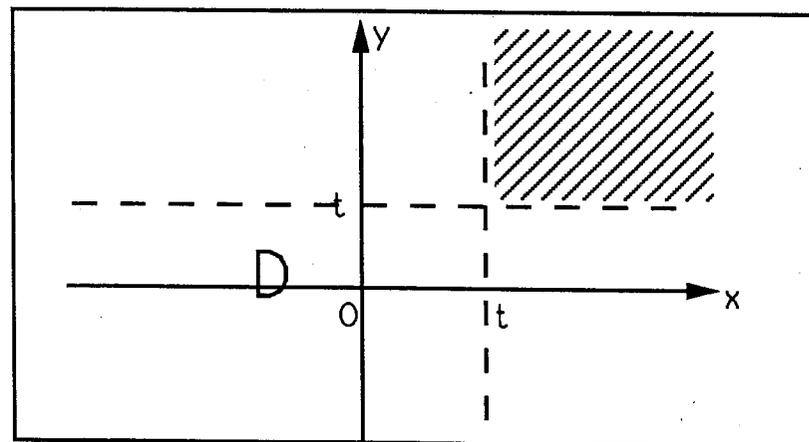


Figure 2.8.

2.4.4. La loi multinomiale

Soit (A_1, A_2, \dots, A_k) une partition de l'espace des épreuves Ω , c'est-à-dire

$$\begin{cases} A_j \neq \emptyset & \text{pour } j = 1, 2, \dots, k \\ A_i \cap A_j = \emptyset & \text{si } i \neq j \\ \bigcup_{j=1}^k A_j = \Omega \end{cases}$$

On note p_j la probabilité de l'événement A_j . Si on réalise l'expérience N fois de façons indépendantes, alors les variables N_1, N_2, \dots, N_k où N_j est le nombre de fois où A_j est réalisé, ne sont pas indépendantes et leur loi conjointe est donnée par

$$\begin{aligned} P(N_1 = n_1 \text{ et } N_2 = n_2 \text{ et } N_k = n_k) &= \\ &= \frac{N!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} \quad \text{si } \begin{cases} n_1 + n_2 + \dots + n_k = N \\ n_1 \geq 0 \dots n_k \geq 0 \end{cases} \\ &= 0 \quad \text{sinon} \end{aligned}$$

En particulier, chacune des variables N_j est de loi binomiale $\mathcal{B}(N, p_j)$.

2.4.5. Fonction caractéristique

Lorsque la variable est à valeurs dans \mathbb{R}^n , la fonction caractéristique est une fonction de n variables.

Définition : La fonction caractéristique du vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ est la fonction de \mathbb{R}^n dans \mathbb{C} définie par

$$\phi(t_1, t_2, \dots, t_n) = E(e^{i(t_1 X_1 + t_2 X_2 + \dots + t_n X_n)})$$

Dans le cas d'une loi à densité, ϕ est pratiquement une transformée de Fourier :

$$\phi(t_1, t_2, \dots, t_n) = \int_{\mathbb{R}^n} f(x_1, x_2, \dots, x_n) e^{i(t_1 x_1 + \dots + t_n x_n)} dx_1 dx_2 \dots dx_n$$

2.4.6. Espérance, matrice de covariance

Si chacune des composantes X_i du vecteur admet une espérance, on appelle espérance du vecteur le vecteur des espérances

$$E(X) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}$$

Si toutes les composantes du vecteur admettent un moment d'ordre 2, c'est-à-dire si les variables aléatoires réelles X_i appartiennent à $L^2(P)$, alors les variables

X_i et les produits $X_i X_j$ appartiennent à $L^1(P)$ (inégalité de Cauchy-Schwarz) et on pose

$$m_i = E(X_i)$$

$$\text{cov}(X_i, X_j) = E\{(X_i - m_i)(X_j - m_j)\}$$

on remarque que

$$\text{cov}(X_i, X_j) = E(X_i X_j) - m_i m_j$$

D'autre part, si deux variables aléatoires réelles, X et Y , admettent un moment d'ordre 2, alors on a la formule utile qui suit (démonstration laissée au lecteur) :

$$\text{var}(aX + bY) = a^2 \text{var}(X) + b^2 \text{var}(Y) + 2ab \text{cov}(X, Y)$$

2.4.7. Fonction caractéristique et moments

En utilisant les résultats de la théorie de la mesure, on obtient :

Si les composantes X_i admettent un moment d'ordre 1, alors la fonction caractéristique admet des dérivées partielles continues et

$$\frac{\partial \phi}{\partial t_k}(0, 0, \dots, 0) = iE(X_k)$$

De même si toutes les composantes du vecteur admettent un moment d'ordre 2 alors la fonction caractéristique admet des dérivées partielles d'ordre 2 continues et

$$\frac{\partial^2 \phi}{\partial t_k \partial t_l}(0, 0, \dots, 0) = -E(X_k X_l)$$

On obtient ainsi les moments à n'importe quel ordre, par dérivation.

2.5. INDÉPENDANCE

2.5.1. Définition

Dans ce qui suit les variables aléatoires sont réelles ou vectorielles :

Deux variables aléatoires X et Y sont indépendantes si quel que soit A et B (éléments de la σ -algèbre appropriée)

$$P(X \in A \text{ et } Y \in B) = P(X \in A) P(Y \in B).$$

n variables X_1, X_2, \dots, X_n sont indépendantes si quels que soient A_1, A_2, \dots, A_n (éléments de la σ -algèbre appropriée) on a

$$P(X_1 \in A_1 \text{ et } X_2 \in A_2 \dots \text{ et } X_n \in A_n) = P(X_1 \in A_1) P(X_2 \in A_2) \dots P(X_n \in A_n)$$

On remarquera la différence avec

« les variables X_1, X_2, \dots, X_n sont 2 à 2 indépendantes »

Par contre, on verra au paragraphe suivant, que pour un vecteur gaussien il suffit que les composantes soient deux à deux indépendantes pour être indépendantes dans leur ensemble.

Si deux v.a.r. X et Y sont indépendantes, alors $\text{cov}(X, Y) = 0$. La réciproque est en général fautive ; elle est vraie dans le cas d'un vecteur gaussien (paragraphe 6). On trouvera un contre-exemple dans l'exercice E.2.18.

2.5.2. Caractérisations

Il existe trois principales caractérisations de l'indépendance de n variables à valeurs réelles X_1, X_2, \dots, X_n .

1) Si le vecteur $X = (X_1, X_2, \dots, X_n)$ admet une densité qui s'écrit sous forme d'un produit du type $f_1(x_1) f_2(x_2) \dots f_n(x_n)$ alors les composantes sont indépendantes ; réciproquement, si les composantes sont indépendantes ont pour densité f_1, f_2, \dots, f_n alors le vecteur X admet pour densité le produit $f_1(x_1) f_2(x_2) \dots f_n(x_n)$. Ce résultat reste vrai si les composantes X_i sont elles-mêmes vectorielles.

2) X_1, X_2, \dots, X_n sont indépendantes si et seulement si la fonction caractéristique s'écrit sous forme du produit des fonctions caractéristiques :

$$\phi(t_1, t_2, \dots, t_n) = E(e^{it_1 X_1}) E(e^{it_2 X_2}) \dots E(e^{it_n X_n})$$

Cette caractérisation peut être adaptée au cas où les composantes X_i sont elles-mêmes vectorielles.

3) X_1, X_2, \dots, X_n sont indépendantes si et seulement si quelles que soient les fonctions h_1, h_2, \dots, h_n de \mathbb{R} dans \mathbb{R}_+ , mesurables

$$E\{h_1(X_1) \cdot h_2(X_2) \dots h_n(X_n)\} = E[h_1(X_1)] \cdot E[h_2(X_2)] \dots E[h_n(X_n)]$$

Pour les démonstrations on pourra se reporter à « Calcul des probabilités » de Renyl.

2.6. Vecteurs gaussiens

On considère une variable aléatoire $X = (X_1, X_2, \dots, X_n)$ à valeurs dans \mathbb{R}^n .

2.6.1. Définition et propriétés

Définition : X est un vecteur gaussien si toute combinaison linéaire de ses composantes est une variable gaussienne. (On rappelle qu'une variable presque sûrement constante est considérée comme une variable gaussienne de variance nulle.)

Lois marginales : Chaque composante du vecteur étant une combinaison linéaire particulière, est donc une variable gaussienne.

Matrice de covariance : On note m_i l'espérance de X_i . Les composantes étant toutes gaussiennes, admettent un moment d'ordre 2 et on peut définir la matrice de covariance que l'on notera Σ : $\Sigma_{ij} = \text{cov}(X_i, X_j)$.

Transformation linéaire : Si A est la matrice d'une transformation linéaire de \mathbb{R}^n dans \mathbb{R}^p , alors AX est un vecteur aléatoire à valeurs dans \mathbb{R}^p ; toute combinaison linéaire de ses composantes étant une combinaison linéaire des X_i est gaussienne : AX est encore un vecteur gaussien.

(AX) a pour espérance $AE(X)$. En ce qui concerne la matrice de covariance de $Y = AX$, on peut remarquer que la matrice d'un vecteur centré U s'obtient en prenant l'espérance de $U'U$:

$$U'U = \begin{pmatrix} U_1 \\ \vdots \\ U_n \end{pmatrix} (U_1, \dots, U_n) = \begin{pmatrix} U_1^2 & U_1 U_2 & \dots & U_1 U_n \\ \vdots & & & \\ U_1 U_n & \dots & \dots & U_n^2 \end{pmatrix}$$

On en déduit

$$\begin{aligned} \Sigma_Y &= E\{(Y - E(Y))'(Y - E(Y))\} \\ &= E\{A(X - m)'(X - m)'A\} \\ &= A\Sigma'A \end{aligned}$$

En particulier, la variance d'une combinaison linéaire $a_1 X_1 + \dots + a_n X_n$ est égale à

$$(a_1, a_2, \dots, a_n) \Sigma \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

Ces calculs sont valables dans les cas non gaussiens ; néanmoins, ils sont particulièrement intéressants dans le cas gaussien puisqu'ils suffisent à déterminer la loi (voir la fonction caractéristique au paragraphe 2.6.2).

Tableau récapitulatif

Vecteur	$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$	$AX = \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix}$
Espérance	$m = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix}$	$Am = \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix}$
Matrice de covariance	Σ	$A\Sigma'A$

2.6.2. Fonction caractéristique

Par définition

$$\phi(t_1, t_2, \dots, t_n) = E(e^{i(t_1 X_1 + t_2 X_2 + \dots + t_n X_n)})$$

$t_1 X_1 + \dots + t_n X_n$ est une variable réelle de loi gaussienne ; ou bien elle est p.s. constante ou bien elle admet une densité gaussienne de paramètres α et β^2 et dans ce cas sa fonction caractéristique est égale (après calcul par les résidus...) à $\exp(i\alpha) \exp\left(-\frac{\beta^2}{2}\right)$ avec

$$\begin{aligned} \alpha &= E(t_1 X_1 + \dots + t_n X_n) = t_1 m_1 + \dots + t_n m_n \\ \beta^2 &= \text{var}(t_1 X_1 + \dots + t_n X_n) \end{aligned}$$

On en déduit donc

$$\phi(t_1, t_2, \dots, t_n) = \exp\left\{i(t_1 m_1 + \dots + t_n m_n) - \frac{1}{2}(t_1, \dots, t_n) \Sigma \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}\right\}$$

2.6.3. Indépendance

En utilisant la caractérisation de l'indépendance par la fonction caractéristique, les composantes X_1, X_2, \dots, X_n sont indépendantes si et seulement si la matrice Σ est diagonale, c'est-à-dire si toutes les covariances sont nulles.

2.6.4. Densité

Σ étant une matrice symétrique réelle peut être diagonalisée. Soit P une matrice de changement de base orthogonale telle que

$$P \Sigma P = D, \quad \text{où } D \text{ est une matrice diagonale}$$

La variable $Y = PX$ a pour matrice de covariance D .

Ses composantes sont donc indépendantes et ont pour variances les éléments diagonaux de D . Ces éléments diagonaux sont donc positifs ou nuls et on obtient le résultat suivant :

Proposition : Si Σ est de rang $r < n$, alors le vecteur aléatoire X appartient presque sûrement à sous-espace affine de dimension r ; si le rang de Σ est égal à n alors X admet pour densité

$$f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(\Sigma)}} \times \exp \left\{ -\frac{1}{2} (x_1 - m_1, x_2 - m_2, \dots, x_n - m_n) \Sigma^{-1} \begin{pmatrix} x_1 - m_1 \\ \vdots \\ x_n - m_n \end{pmatrix} \right\}$$

2.7. SOMME DE VARIABLES INDÉPENDANTES

Dans la pratique, on est souvent amené à faire des calculs de probabilités sur des sommes de variables aléatoires indépendantes ; d'autre part, on retrouve cette situation dans la loi des grands nombres (chap. 3).

2.7.1. Loi de la somme

Lorsque les deux variables indépendantes X et Y sont des variables à densités respectives f et g , la variable $X + Y$ a pour densité le produit de convolution des densités $f * g$; en effet :

$$P(X + Y \in A) = \int_{x+y \in A} f(x) g(y) dx dy = \int_{u \in A} \left[\int_{\mathbb{R}} f(x) g(u-x) dx \right] du$$

Dans le cas de deux variables gaussiennes, X et Y , indépendantes, on sait que le couple est gaussien, et par conséquent la somme est gaussienne ; il suffit alors de calculer l'espérance et la variance pour connaître complètement la loi ce qui évite le calcul du produit de convolution des densités.

Dans le cas de variables à valeurs entières, on obtient une formule de convolution discrète :

$$P(X + Y = n) = \sum_k P(X = k) P(Y = n - k) = \sum_k P(Y = k) P(X = n - k)$$

En particulier, la somme de n variables de Bernoulli, de même paramètre p et indépendantes, suit une loi binomiale $\mathcal{B}(n, p)$.

Variance : Si les deux variables indépendantes X et Y admettent des moments d'ordre 2, $X + Y$ a pour variance la somme des variances (démonstration laissée au lecteur).

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) \quad \text{si } X \text{ et } Y \text{ sont indépendantes}$$

2.7.2. Fonction caractéristique, transformée de Laplace

Sous les mêmes hypothèses que ci-dessus, $X + Y$ a pour fonction caractéristique $\phi_X(t) \phi_Y(t)$.

Transformée de Laplace : Si X et Y sont indépendantes, la transformée de Laplace de la somme est le produit des transformées :

$$E(e^{-z(X+Y)}) = \hat{\mu}_X(z) \hat{\mu}_Y(z)$$

Application aux lois $\gamma(a, \lambda)$:

La loi $\gamma(a, \lambda)$ a pour densité sur \mathbb{R}_+ $x \mapsto \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1}$ où $a, \lambda > 0$.

On trouve que la transformée de Laplace existe pour $\text{Re}(z) > -\lambda$; on a

$$\hat{\mu}(z) = \frac{\lambda^a}{(z + \lambda)^a}$$

Si on effectue le produit de convolution de $\gamma(a, \lambda)$ et $\gamma(b, \lambda)$, la transformée de Laplace est le produit des transformées

$$\frac{\lambda^a}{(z + \lambda)^a} \frac{\lambda^b}{(z + \lambda)^b} = \frac{\lambda^{a+b}}{(z + \lambda)^{a+b}}$$

$$\gamma(a, \lambda) * \gamma(b, \lambda) = \gamma(a + b, \lambda)$$

Cela permet de calculer les moments pour une loi $\gamma(a, \lambda)$:

$$E(X) = \frac{a}{\lambda} \quad \text{et} \quad E(X^2) = \frac{a(a+1)}{\lambda^2}$$

$$E(X^n) = \frac{a(a+1) \dots (a+n-1)}{\lambda^n}$$

2.7.3. Gaussiennes, Qui-deux et lois gamma

Soit X une v.a.r. à densité f et $t > 0$; cherchons la loi de X^2

$$P(X^2 < t) = \int_{0 < x < \sqrt{t}} f(x) dx + \int_{-\sqrt{t} < x < 0} f(x) dx$$

Les intégrales sont ensuite calculées par le changement de variables $u = x^2$.

Si X suit une loi $\mathcal{N}(0, 1)$ alors X^2 suit une loi γ :

$$X^2 \text{ a pour loi } \gamma\left(\frac{1}{2}, \frac{1}{2}\right)$$

et $\sigma^2 X^2$ suit aussi une loi γ :

$$\sigma^2 X^2 \text{ a pour loi } \gamma\left(\frac{1}{2}, \frac{1}{2\sigma^2}\right)$$

Si X_1, X_2, \dots, X_n sont indépendantes de loi $\mathcal{N}(0, 1)$, alors

$$\sigma^2(X_1^2 + X_2^2 + \dots + X_n^2) \text{ a pour loi } \gamma\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right)$$

Lorsque σ est égal à 1, c'est ce que l'on appelle une loi de *Qui-deux* à n degrés de liberté. On rencontre les lois de *Qui-deux* en statistique, lorsqu'on estime la variance d'un échantillon gaussien. La loi de *Qui-deux* à n degrés de liberté est aussi notée χ_n^2 .

2.8. LOIS CONDITIONNELLES

Il s'agit d'une notion subtile, qui s'avère un outil puissant dans les calculs de probabilité. Dans ce paragraphe, on donne uniquement les résultats pour des variables à valeurs dans \mathbb{R} ou \mathbb{R}^n ; pour les résultats plus généraux et les démonstrations on pourra se reporter au chapitre 4 du livre « Probability » de Breiman.

On rappelle la définition de la probabilité conditionnelle de A sachant que B , lorsque l'événement B est de probabilité non nulle :

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Exemple en discret : Si X et Y sont des variables de Poisson, indépendantes de paramètres respectifs λ et μ , conditionnellement à l'événement « $X + Y = n$ » la probabilité pour que $X = k$, est celle d'une loi binomiale $\mathcal{B}(n, p)$ avec

$$p = \frac{\lambda}{\lambda + \mu}$$

(démonstration laissée au lecteur).

2.8.1. Le cas des densités

Soit (X, Y) un vecteur aléatoire de densité $f(x, y)$. On appelle g la densité marginale de Y . On ne peut pas utiliser la formule de la probabilité conditionnelle avec $B = \{Y = y\}$ et $A = \{X = x\}$, ces deux événements étant de probabilité nulle.

On arrive à donner un sens à la fonction

$$x \mapsto f_{X/Y=y_0} = \frac{f(x, y_0)}{\int f(u, y_0) du}$$

que l'on appelle densité conditionnelle de X sachant $Y = y_0$. Elle a la propriété suivante :

$$\begin{aligned} \int_{y_0 \in B} g(y_0) \int_{x \in A} f_{X/Y=y_0}(x) dx \times dy_0 &= \int_{y_0 \in B} g(y_0) \frac{\int_{x \in A} f(x, y_0) dx}{\int f(u, y_0) du} \\ &= \int_A \int_B f(x, y) dx dy = P(X \in A ; Y \in B) \end{aligned}$$

On peut dire grossièrement, que l'on intègre $P(X \in A / Y = y_0)$ en tenant compte des fluctuations de Y .

Remarques importantes : Dans cette définition de la loi de X conditionnellement à $Y = y$, X et Y peuvent être des vecteurs aléatoires ; $X = (X_1, \dots, X_n)$ et $Y = (Y_1, \dots, Y_p)$. Dans la formule qui donne la densité conditionnelle, l'intégrale sur la variable multidimensionnelle x doit alors être écrite avec $dx = dx_1 dx_2 \dots dx_n$.

La loi conditionnelle de X sachant $Y = y$ est une loi de probabilité et, à ce titre, possède une fonction de répartition, éventuellement des moments, une fonction caractéristique... Cette fonction caractéristique conditionnelle peut être très utile dans les calculs de loi.

2.8.2. Espérance conditionnelle

On notera $E(X/Y)$ l'espérance de X conditionnellement à Y : intuitivement, on regarde la moyenne de X sur toutes les épreuves telles que $Y = y_0$. On admettra le résultat suivant qui permet de définir $E(X/Y)$:

Proposition : Si X admet un moment d'ordre 1, alors il existe une variable aléatoire unique (à l'égalité p.s. près), qui est une fonction de Y , que l'on appelle espérance conditionnelle de X sachant Y , qui a la propriété suivante

$$\int_{Y \in A} E(X/Y) dP = \int_{Y \in A} X dP \quad \forall A \in \mathcal{B}$$

Interprétation : Lorsque X admet un moment d'ordre 2, $E(X/Y)$ est la projection orthogonale de X sur l'espace $L^2(Y)$, espace des **variables qui sont fonctions de Y et de carré intégrable** ; en effet, la variable $h(Y)$ qui approxime le mieux X au sens de l'écart quadratique moyen,

$$E[X - h(Y)]^2$$

réalise

$$\inf_h \|X - h(Y)\|^2$$

avec le produit scalaire et la norme définis par :

$$\begin{cases} \|U\|^2 = \langle U, U \rangle = E(U^2) \\ \langle U, V \rangle = E(UV) \end{cases}$$

Or, $E(X/Y)$ est caractérisée par la propriété d'orthogonalité :

$$X - E(X/Y) \perp h(Y)$$

c'est-à-dire

$$E\{[X - E(X/Y)][h(Y)]\} = 0$$

pour toute variable $h(Y)$ qui est une fonction indicatrice :

$$h(\gamma) = 1_{\{\gamma \in A\}}$$

La propriété d'orthogonalité s'étend à l'espace vectoriel engendré par ces fonctions indicatrices et aux limites de suites, c'est dire à toutes les fonctions $h(Y)$ ayant un moment d'ordre 2 fini. Cette propriété d'orthogonalité est très utile dans le cas des variables gaussiennes comme on le verra dans la suite. Énonçons toutes les propriétés de l'espérance conditionnelle :

Théorème : Soit X une variable aléatoire réelle admettant un moment d'ordre 1. Alors,

- 1) $E(X/Y)$ est une variable aléatoire dont l'espérance est égale à celle de X .
- 2) Si X est une fonction de Y , c'est-à-dire $X = h(Y)$, alors $E(X/Y)$ est égale à X .
- 3) Si X et Y sont indépendantes alors $E(X/Y)$ est égale à la constante $E(X)$.
- 4) L'espérance conditionnelle est linéaire par rapport à la variable X , c'est-à-dire

$$E(aX_1 + bX_2/Y) = aE(X_1/Y) + bE(X_2/Y)$$

- 5) Si de plus X admet un moment d'ordre 2, alors $E(X/Y)$ est la projection de X sur $L^2(Y)$; elle est caractérisée par

$$E\{h(Y)[X - E(X/Y)]\} = 0 \quad \forall h \text{ vérifiant } E[h^2(Y)] < +\infty$$

Toutes ces propriétés s'adaptent au cas où X est à valeurs dans \mathbb{R}^n . On pourra consulter à ce sujet le livre de Breiman déjà cité.

Le cas des densités :

Si (X, Y) a pour densité $f(x, y)$, alors $E(X/Y) = h(Y)$ avec

$$h(y) = \int x f_Y(x) dx$$

L'espérance conditionnelle est l'espérance de la loi de probabilité conditionnelle.

Dans le cas d'un couple gaussien (X, Y) , la loi de X conditionnellement à $Y = y$ est une loi gaussienne dont le paramètre de centrage est $E(X/Y)$:

$$f_Y(x) = \frac{1}{\sqrt{2\pi} \sigma_1 \sqrt{1-\rho^2}} \exp \left\{ \frac{1}{2(1-\rho^2)} \left[\frac{(x - E(X/Y))^2}{\sigma_1^2} \right] \right\}$$

$$E(X/Y) = m_1 + \rho \frac{\sigma_1}{\sigma_2} (Y - m_2)$$

Ce résultat se généralise à un vecteur gaussien : Si $(X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_p)$ est un vecteur gaussien, alors la loi conditionnelle du vecteur (X_1, X_2, \dots, X_n) conditionnellement au vecteur (Y_1, \dots, Y_p) est une loi gaussienne centrée sur $E(X/Y)$. Le vecteur $E(X/Y)$ est une fonction affine de Y_1, \dots, Y_p comme dans le cas unidimensionnel ; (voir exercice 2.17).

EXERCICES

E.2.1. Soit X une variable de loi uniforme sur $\{1, 2, \dots, 9\}$. On pose $Y = 2X - 1$ modulo 3. Quelle est la loi de Y ?

E.2.2. Si X suit une loi $\mathcal{B}(18, 1/3)$ quelle est la valeur k la plus probable ? On pourra calculer le rapport $P(X = k + 1)/P(X = k)$.

E.2.3. Mr Chance jette un dé ; si l'as sort, il joue à Pile ou Face pour continuer : si c'est Face il doit s'arrêter en gagnant 10 F ; si c'est Pile il relance le dé et gagne 100 F si c'est l'as qui sort, perd 10 F si ce n'est pas l'as. Si l'as n'est pas sorti au premier envoi du dé, il perd 1 F. On pose

$$\begin{aligned} X &= \text{nombre d'envois du dé} \\ Y &= \text{Gain (positif ou négatif)} \end{aligned}$$

Calculer la loi de X et la loi de Y .

E.2.4. Soit X une v.a.r. de loi exponentielle de paramètre λ . Montrer que λ joue le rôle d'un paramètre d'échelle : on cherchera la loi de aX , a étant un

paramètre. Calculer la médiane, M , de X c'est-à-dire la valeur M qui vérifie

$$P(X < M) = \frac{1}{2} = P(X > M)$$

E.2.5. On considère des particules radioactives, de même nature. On suppose que le temps de désintégration de la particule n° i , est une variable aléatoire X_i de loi exponentielle de paramètre λ . Les variables X_i sont supposées indépendantes. La médiane (voir exercice précédent) est alors appelée temps de demie-vie. Il y a N particules au temps $t = 0$; on appelle Y le nombre de particules désintégrées avant l'instant M . Quelle est la loi de Y ?

E.2.6. Un point est choisi au hasard sur le segment $[0, 1]$ et on appelle X sa distance à $1/3$. Calculer la fonction de répartition et la densité de X .

E.2.7. Recherche de quantiles : Soit X une v.a.r. de loi de Cauchy. Calculer $a, b, c_1, c_2, \dots, c_9$ tels que

$$P(X \leq a) = 1/4; \quad P(X \leq b) = 1/3 \\ P(X \leq c_1) = \frac{1}{10}; \quad P(X \leq c_2) = \frac{2}{10}; \quad \dots; \quad P(X \leq c_9) = \frac{9}{10}$$

a est appelé quantile d'ordre $1/4$, b quantile d'ordre $1/3$, c_i quantile d'ordre $\frac{i}{10}$.

E.2.8. On installe un appareil neuf au temps $t = 0$ et on suppose que l'instant de la première panne, R_1 , est une v.a.r. de loi exponentielle de paramètre λ . On décide de changer l'appareil au temps s même s'il n'est pas tombé en panne. Soit X le temps de vie sans panne de l'appareil. Exprimer X en fonction de s et R_1 . Calculer $P(X = s)$ et $P(X < t)$ pour $0 \leq t < s$.

E.2.9. On reprend les notations du paragraphe 2.1.6 et on suppose

$$R(t) = \exp \left[- \left(\frac{t}{\alpha} \right)^\beta \right]$$

Pour quelles valeurs du paramètre β , la fiabilité $R(t)$ est-elle une fonction décroissante du temps?

E.2.10. Si on sait que pour une v.a. de loi $\mathcal{N}(0, 1)$, il y a 90 % de chances d'obtenir une valeur entre $-1,64$ et $1,64$, calculer A tel que si X est de loi $\mathcal{N}(m, \sigma^2)$, on ait

$$P(m - A \leq X \leq m + A) = 90 \%$$

E.2.11. Montrer que si X suit une loi $\mathcal{N}(m, \sigma^2)$ alors e^X suit une loi log-normale (on calculera la fonction de répartition puis la densité).

E.2.12. On reprend les notations de l'exercice 2.9. Soit X le temps de bon fonctionnement. Quelle est la densité de la v.a. X ? Calculer $E(X)$.

E.2.13. On reprend l'exercice 2.3. Calculer loi du couple (X, Y) . X et Y sont-elles indépendantes? Calculer $E(Y)$.

E.2.14. Deux composants sont disposés en série au sens où le système est défaillant dès que l'un des composants est défaillant. Exprimer la durée de vie aléatoire du système en fonction des durées de vie respectives des éléments, notées X et Y , et calculer la loi si X et Y sont indépendantes, de loi exponentielle. Même question lorsque les éléments sont disposés en parallèle.

E.2.15. Soit X de loi $\mathcal{N}(0, 1)$. En utilisant $(e^{-x^2/2})' = -x e^{-x^2/2}$ calculer $E(X^2)$ et $E(X^4)$. En déduire $E[(X - m)^4]$ lorsque X est de loi $\mathcal{N}(m, \sigma^2)$.

E.2.16. Montrer que pour tout z dans \mathbb{C} , l'intégrale $I(z)$

$$I(z) = \int_{\mathbb{R}} e^{-zx} e^{-x^2/2} dx$$

est absolument convergente. On admettra qu'elle définit une fonction analytique.

En utilisant

$$\int_{-\infty}^{+\infty} e^{-u^2} du = \sqrt{\pi}$$

Calculer $I(z)$ pour z réel. En déduire

$$I(z) = \sqrt{2\pi} e^{z^2/2}$$

et la fonction caractéristique d'une variable $\mathcal{N}(m, \sigma^2)$.

E.2.17. Soit (X, Y) un vecteur gaussien d'espérance $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ et de matrice de covariance $\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$. Donner la loi de $X - 2Y$. A quelles conditions sur a, b , la variable $aX + bY$ est-elle indépendante de $X - 2Y$ et de variance 1. Soit Z une telle variable. Utiliser la décomposition $Y = \alpha(X - 2Y) + \beta Z$ pour calculer la loi conditionnelle de Y sachant $X - 2Y$. En déduire $E(Y/X - 2Y)$.

E.2.18. Soient X, Y, Z trois v.a.r. indépendantes, de loi $\mathcal{N}(0, 1)$. On admettra $E(e^{ix(iY + sZ)})/X = h(X)$ avec $h(x) = E(e^{ix(iY + sZ)})$.

a) En déduire $E(e^{i(xY + sXZ)})$. Montrer que les variables XY et XZ ne sont pas indépendantes.

b) Montrer qu'on a $\text{cov}(XY, XZ) = 0$.

E.2.19. On considère une v.a.r. X , de Bernoulli telle que $P(X = 0) = 1/3$ et $P(X = 1) = 2/3$ (jeu de pile ou face truqué). Suivant le résultat de X , on fait un tirage aléatoire d'une v.a. Y , de loi $\mathcal{N}(0, 1)$ si $X = 0$, de loi $\mathcal{N}(1, 1)$ si $X = 1$. Calculer

$$\lim_{\Delta y \rightarrow 0} P(X = 0/Y \in [y_0, y_0 + \Delta y])$$

Si on observe uniquement Y (la valeur de X a été perdue) quelle valeur peut-on prédire pour X ?

CHAPITRE 3

THÉORÈMES DE CONVERGENCE

3.1. CONVERGENCE EN MOYENNE QUADRATIQUE

3.1.1. Définition

On dit que X_n converge en moyenne quadratique (en abrégé m.q.) vers X si $E(X_n - X)^2$ tend vers 0 lorsque n tend vers l'infini. Si les variables prennent leurs valeurs dans \mathbb{C} , il s'agit de l'espérance du carré du module.

Ainsi, une variable S_n binomiale $\mathcal{B}(n, p)$, S_n/n converge en moyenne quadratique vers p :

$$E\left(\left[\frac{S_n}{n} - p\right]^2\right) = \frac{p(1-p)}{n} \xrightarrow{n \rightarrow +\infty} 0$$

S_n est la somme de n variables de Bernoulli indépendantes et de façon plus générale, si les variables X_i sont indépendantes, d'espérance m et de variance σ^2 , la moyenne empirique

$$\frac{X_1 + X_2 + \dots + X_n}{n}$$

converge en m.q. vers m (démonstration laissée au lecteur).

En théorie du signal (chap. 5), on utilise un critère de convergence en moyenne quadratique dû à Loève :

3.1.2. Lemme de Loève

Soit X_n une suite de v.a. à valeurs dans \mathbb{R} ou \mathbb{C} . La suite X_n converge en moyenne quadratique si et seulement si

$$E(X_n - \bar{X}_m)$$

admet une limite finie c lorsque n et m tendent vers l'infini.

Démonstration : Supposons tout d'abord que la suite converge en m.q. vers X .

On en déduit

$$\begin{aligned} |E[X(\bar{X}_n - \bar{X})]| &\leq \sqrt{E(|X|^2)} \sqrt{E(|X_n - X|^2)} \xrightarrow{n \rightarrow +\infty} 0 \\ |E[\bar{X}(X_n - X)]| &\leq \sqrt{E(|X|^2)} \sqrt{E(|X_n - X|^2)} \xrightarrow{n \rightarrow +\infty} 0 \end{aligned}$$

soit,

$$E(X\bar{X}_n) \xrightarrow{n \rightarrow +\infty} E(|X|^2) \quad \text{et} \quad E(\bar{X}X_n) \xrightarrow{n \rightarrow +\infty} E(|X|^2)$$

En utilisant $E(|X_n - X|^2) \rightarrow 0$, on en déduit aussi

$$E(|X_n|^2) \rightarrow E(|X|^2)$$

et

$$\begin{aligned} E(X_n \bar{X}_m) - E(|X|^2) &= E(X_n \bar{X}_m) - E(X_n \bar{X}) \\ &\quad + E(X_n \bar{X}) - E(|X|^2) \\ &= E[X_n(\bar{X}_m - \bar{X})] + E[\bar{X}(X_n - X)] \end{aligned}$$

et ces deux termes tendent vers 0.

Réciproquement, si $E(X_n \bar{X}_m)$ admet une limite c lorsque n et m tendent vers l'infini alors

$$E(X_n \bar{X}_n) \xrightarrow{n \rightarrow +\infty} c$$

$$\begin{aligned} E(|X_n - X_m|^2) &= E(|X_n|^2) + E(|X_m|^2) - E(X_n \bar{X}_m) - E(X_m \bar{X}_n) \\ &\xrightarrow{n, m \rightarrow +\infty} 0 \end{aligned}$$

la suite X_n qui satisfait le critère de Cauchy dans $L^2(\Omega, \mathcal{Q}, P)$ admet donc une limite dans cet espace, c'est-à-dire converge en m.q.

3.2. CONVERGENCE EN PROBABILITÉ ET CONVERGENCE PRESQUE SÛRE

3.2.1. Inégalité de Bienaymé-Tchebichev

$$P(|X - a| > \varepsilon) \leq \frac{E(|X - a|^2)}{\varepsilon^2}$$

On généralise ce résultat à n'importe quelle puissance positive α :

$$P(|X - a| > \varepsilon) \leq \frac{E(|X - a|^\alpha)}{\varepsilon^\alpha}$$

Démonstration :

$$E|X - a|^2 \geq E[|X - a|^2 1_{\{|X - a| \geq \varepsilon\}}] \geq \varepsilon^2 P(|X - a| \geq \varepsilon)$$

3.2.2. Convergence en probabilité

Définition : On dit que X_n converge en probabilité vers X si pour tout $\varepsilon > 0$, $P(|X_n - X| > \varepsilon)$ tend vers 0 lorsque n tend vers l'infini.

Exemple : Si $P(X_n = 0) = 1 - \frac{1}{n}$ et $P(X_n = 1) = \frac{1}{n}$ alors X_n converge en probabilité vers 0.

Proposition : La convergence en moyenne quadratique entraîne la convergence en probabilité vers la même limite.

Démonstration : On utilise l'inégalité de Bienaymé Tchebichev,

$$P(|X_n - X| > \varepsilon) \leq \frac{E(|X_n - X|^2)}{\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0$$

3.2.3. Convergence presque sûre

Définition : On dit que X_n converge presque sûrement vers X si l'ensemble des épreuves ω pour lesquelles la suite ne converge pas vers $X(\omega)$, est négligeable (inclus dans un ensemble de probabilité nulle).

Par commodité, on remplace « presque sûrement » par son abréviation « p.s. ».

Proposition : Si la suite X_n converge p.s. vers X alors elle converge aussi en probabilité et vers la même limite.

Démonstration : On pose

$$\begin{aligned} N &= \{\omega \in \Omega / X_n(\omega) \text{ ne converge pas vers } X(\omega)\} \\ A_n &= \{|X_n - X| > \varepsilon\} \end{aligned}$$

Si ω appartient à N^c , $X_n(\omega)$ converge vers $X(\omega)$; il appartient alors à tous les A_n^c à partir d'un certain rang c'est-à-dire

$$N^c \subset \bigcup_{n=0}^{+\infty} \left[\bigcap_{p=n}^{+\infty} A_p^c \right]$$

$$0 = P(N) \geq \lim_{n \rightarrow +\infty} P\left(\bigcup_{p=n}^{+\infty} A_p \right)$$

$$P(A_n) \leq P\left(\bigcup_{p=n}^{+\infty} A_p \right) \xrightarrow{n \rightarrow +\infty} 0$$

CQFD

La réciproque n'est pas vraie : voir l'exercice 3.2 pour un contre-exemple.

3.2.4. Convergence p.s. pour des moyennes de variables de Bernoulli

On sait que la somme de n variables de Bernoulli indépendantes et de même paramètre p , suit une loi $\mathcal{B}(n, p)$; plaçons-nous dans cette situation: X_1, X_2, \dots est la suite de variables de Bernoulli et S_n la somme partielle,

$$S_n = X_1 + X_2 + \dots + X_n$$

On a alors

$$E \left(e^{it \left(\frac{S_n}{n} - p \right)} \right) = e^{-ip} [p e^{it/n} + (1-p)]^n$$

on en déduit le moment d'ordre 4 de $\left(\frac{S_n}{n} - p \right)$:

$$\begin{aligned} E \left(\left[\frac{S_n}{n} - p \right]^4 \right) &= \frac{d^4}{dt^4} \{ e^{-ip} [p e^{it/n} + (1-p)]^n \} /_{t=0} \\ &= \frac{p(1-p)}{n^2} \left[3p(1-p) + \frac{1-6p(1-p)}{n} \right] \end{aligned}$$

Il existe N tel que

$$n \geq N \Rightarrow E \left(\left[\frac{S_n}{n} - p \right]^4 \right) \leq \frac{4p^2(1-p)^2}{n^2} = \frac{K}{n^2}$$

On en déduit, pour tout $\varepsilon > 0$

$$n \geq N \Rightarrow P \left(\left| \frac{S_n}{n} - p \right| \geq \varepsilon \right) \leq \frac{K}{n^2 \varepsilon^4}$$

Posons

$$A_n = \left\{ \omega \in \Omega \mid \frac{S_n(\omega)}{n} \notin]-p - \varepsilon, p + \varepsilon[\right\}$$

On a alors

$$\bigcup_{n=k}^{+\infty} A_n = \left\{ \omega \in \Omega \text{ tels que } \exists n \geq k \frac{S_n(\omega)}{n} \notin]-p - \varepsilon, p + \varepsilon[\right\}$$

$$P \left(\bigcup_{n=k}^{+\infty} A_n \right) \leq \sum_{n=k}^{+\infty} P(A_n) \leq \frac{K}{\varepsilon^4} \cdot \sum_{n=k}^{+\infty} \frac{1}{n^2} \rightarrow 0$$

$$\Rightarrow P \left(\bigcap_k \left[\bigcup_{n=k}^{+\infty} A_n \right] \right) = \lim_{k \rightarrow +\infty} P \left(\bigcup_{n=k}^{+\infty} A_n \right) = 0$$

$\frac{S_n(\omega)}{n}$ converge vers p si et seulement si, quel que soit $\varepsilon > 0$, la suite $\frac{S_n(\omega)}{n}$ appartient à l'intervalle $]p - \varepsilon, p + \varepsilon[$ à partir d'un certain rang. Il suffit

en fait de vérifier la propriété sur une suite ε_j qui tend vers 0 en décroissant. Or pour tout j , on a

$$P \left(\frac{S_n}{n}(\omega) \text{ appartient à }]p - \varepsilon_j, p + \varepsilon_j[\text{ à partir d'un certain rang} \right) = 1$$

On en déduit

$$P \left[\bigcap_j \left\{ \frac{S_n}{n} \in]p - \varepsilon_j, p + \varepsilon_j[\text{ à partir d'un certain rang} \right\} \right] = 1 \quad \text{CQFD}$$

Remarque : Cette méthode est applicable dès que $S_n = X_1 + \dots + X_n$ avec les variables X_i indépendantes, de même loi et ayant un moment d'ordre 4 fini. On retrouvera ce résultat dans le paragraphe « loi des grands nombres ».

3.3. LOI DU TOUT OU RIEN ET LEMME DE BOREL-CANTELLI

3.3.1. Loi du tout ou rien

Considérons une suite de v.a. X_n et introduisons pour chaque instant n , la σ -algèbre \mathcal{B}_n des événements du présent et la σ -algèbre \mathcal{F}_n des événements du futur comme suit :

\mathcal{B}_n est la plus petite σ -algèbre qui contient tous les événements du type $\{X_n \in A\}$.

\mathcal{F}_n est la plus petite σ -algèbre qui contient tous les événements du type $\{X_p \in A\}$ avec $p \geq n$.

On définit alors la σ -algèbre des événements asymptotiques, \mathcal{F}_∞ :

$$\mathcal{F}_\infty = \bigcap_n \mathcal{F}_n$$

On a alors le résultat suivant dû à Kolmogorov :

Théorème : Si les variables X_1, X_2, \dots sont indépendantes alors tout événement asymptotique est de probabilité 0 ou 1.

On trouvera la démonstration dans « Calcul des probabilités » de Rényi.

3.3.2. Exemples d'événements asymptotiques

X_n étant une suite de v.a.r., les événements suivants sont asymptotiques :

« La suite X_n passe une infinité de fois dans $[0, 1]$ »

« A partir d'un certain rang, X_n appartient à $[0, 1]$ »

« $\sup_n X_n$ est infini »

« $\inf_n X_n$ est égal à $-\infty$ »

« La série de terme général $\frac{X_n}{n}$ est convergente »

Dans le cas où les variables X_n sont indépendantes, ces événements sont tous de probabilité 0 ou 1 ; à chaque fois, la propriété est vérifiée pour presque tout ω , ou bien, au contraire, n'est vérifiée que sur un ensemble de probabilité nulle.

On montre ainsi que la série $\sum_{n=1}^{+\infty} \frac{X_n(\omega)}{n}$, où les variables X_n sont équidistribuées, indépendantes $X_n = \pm 1$ avec probabilité $\frac{1}{2}$ et $\frac{1}{2}$ respectivement, est presque sûrement convergente : presque tous les choix de signes aléatoires aboutissent à une série semi-convergente (puisque la série de terme général $\frac{1}{n}$ est divergente).

3.3.3. Lemme de Borel-Cantelli

Soit A_n une suite d'événements :

- 1) Si $\sum_n P(A_n) < +\infty$ alors $P\left(\bigcap_{n \geq 1} \bigcup_{k \geq n} A_k\right) = 0$.
- 2) Si $\sum_{n=2}^{+\infty} P\left(A_n \mid \bigcap_{k=1}^{n-1} A_k^c\right) = +\infty$ alors $P\left(\bigcup_{k \geq 1} A_k\right) = 1$.
- 3) Si les A_n sont indépendants dans leur ensemble alors

$$P\left(\bigcap_{n \geq 1} \bigcup_{k \geq n} A_k\right) = 0 \quad \text{si} \quad \sum_n P(A_n) < +\infty$$

$$= 1 \quad \text{si} \quad \sum_n P(A_n) = +\infty$$

Démonstration :

$$1) \text{ On utilise } P\left(\bigcup_{k \geq n} A_k\right) \leq \sum_{k=n}^{+\infty} P(A_k).$$

$$2) \quad P\left(\bigcap_{k \geq 1} A_k^c\right) = \lim_n P\left(\bigcap_{1 \leq k \leq n} A_k^c\right)$$

$$P\left(\bigcap_{1 \leq k \leq n} A_k^c\right) = P\left(\bigcap_{1 \leq k \leq n} A_k^c \mid \bigcap_{1 \leq k \leq n-1} A_k^c\right) P\left(\bigcap_{1 \leq k \leq n-1} A_k^c\right)$$

$$P\left(\bigcap_{k \geq 1} A_k^c\right) = P(A_1^c) \prod_{n=2}^{+\infty} P\left(\bigcap_{1 \leq k \leq n} A_k^c \mid \bigcap_{1 \leq k \leq n-1} A_k^c\right)$$

$$= P(A_1^c) \prod_{n=2}^{+\infty} \left[1 - P\left(A_n \mid \bigcap_{1 \leq k \leq n-1} A_k^c\right)\right]$$

$$= 0 \quad \text{car} \quad \sum_{n \geq 2} P\left(A_n \mid \bigcap_{1 \leq k \leq n-1} A_k^c\right) = +\infty$$

$$\Rightarrow P\left(\bigcup_{k \geq 1} A_k\right) = 1.$$

- 3) Les A_n étant indépendants, $P\left(A_n \mid \bigcap_{k=1}^{n-1} A_k^c\right) = P(A_n)$ et d'après 2)

$$\forall n \quad P\left(\bigcup_{k \geq n} A_k\right) = 1,$$

si la série diverge. On en déduit :

$$P\left(\bigcap_{n \geq 1} \bigcup_{k \geq n} A_k\right) = \lim_n P\left(\bigcup_{k \geq n} A_k\right) = 1$$

Le lemme de Borel-Cantelli est utilisé pour démontrer des convergences presque sûre : à la fin du paragraphe précédent, on a démontré la convergence presque sûre de S_n en utilisant le résultat 1) du lemme de Borel-Cantelli.

3.4. LOI DES GRANDS NOMBRES

3.4.1. Loi faible

Soit X_n une suite de variables indépendantes et équidistribuées, de moyenne m et de variance σ^2 alors la moyenne empirique converge en probabilité vers m .

3.4.2. Énoncé de la loi forte

Soit X_n une suite de variables indépendantes, de même loi et d'espérance m , alors la moyenne empirique converge p.s. vers m :

$$\frac{X_1 + X_2 + \dots + X_n}{n} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} m$$

Nous avons vu en 3.2.4 la démonstration dans un cas particulier ; dans le cas général la démonstration est longue et difficile ; on pourra la trouver dans le livre « Calcul des probabilités » de Neveu ou dans « Probability » de Breimann, ou « Calcul des probabilités » de Rényl.

3.4.3. Contre-exemple

L'hypothèse « le moment d'ordre 1 existe » est fondamentale. En effet, si les variables sont indépendantes et suivent une loi de Cauchy, bien que ces variables n'admettent pas d'espérance (l'intégrale est divergente à l'infini), du fait de la symétrie de la densité on serait tenté de dire que la moyenne empirique converge

vers 0 ; or la moyenne empirique suit encore une loi de Cauchy car

$$E\left(e^{\frac{X_1+X_2+\dots+X_n}{n}}\right) = \left[E\left(e^{\frac{X_1}{n}}\right)\right]^n = \left[e^{-\frac{|t|}{n}}\right]^n = e^{-|t|}$$

Cela veut dire que l'on n'a pas réduit les fluctuations en moyennant les variables.

3.4.4. Extension

Le résultat le plus général que l'on puisse obtenir (grâce à la théorie des martingales) est le suivant :

Théorème : Soit Y_n une suite de v.a.r.

Si $\forall n \geq 1$,

$$E[Y_{n+1}/Y_1, \dots, Y_n] \leq Y_n \quad (\text{resp. } =, \geq)$$

p.s.

$$\text{et } \sup_n E(|Y_n| \mathbb{1}_{\{|Y_n| > A\}}) \rightarrow 0$$

$$A \rightarrow +\infty$$

alors il existe une variable Y intégrable telle que Y_n converge p.s. et en moyenne vers Y .

Remarque : C'est une généralisation de « toute suite de réels, monotone et bornée, converge ».

3.4.5. Convergence de la fonction de répartition empirique

Soit X_1, X_2, \dots une suite de v.a.r. indépendantes, de même loi, de fonction de répartition $F(t)$.

On définit alors la fonction de répartition empirique $F_n(t)$ par :

$$F_n(t) = \frac{\text{Nombre de variables } X_i \text{ qui sont inférieures à } t}{n}$$

A t fixé, on peut appliquer la loi des grands nombres en posant

$$Y_j(\omega) = \begin{cases} 1 & \text{si } X_j(\omega) < t \\ 0 & \text{sinon} \end{cases}$$

de sorte que $F_n(t)$ est la moyenne empirique des variables Y_1, \dots, Y_n :

$$F_n(t) = \frac{Y_1 + Y_2 + \dots + Y_n}{n}$$

et d'après la loi des grands nombres

$$F_n(t) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} E(Y_1) = F(t)$$

Pour tout t , $F_n(t)$ converge presque sûrement vers $F(t)$.

3.4.6. Un exemple en théorie du signal

On observe un signal $s(t)$ sur un intervalle de temps $[0, T]$. On suppose que $s(t)$ est une fonction de \mathbb{R} dans \mathbb{R} , continue périodique, de période τ , et que le temps d'observation T est grand devant la période. On pose

$$G(t) = \frac{\text{mesure}(\{u/0 < u < T \text{ et } s(u) < t\})}{T}$$

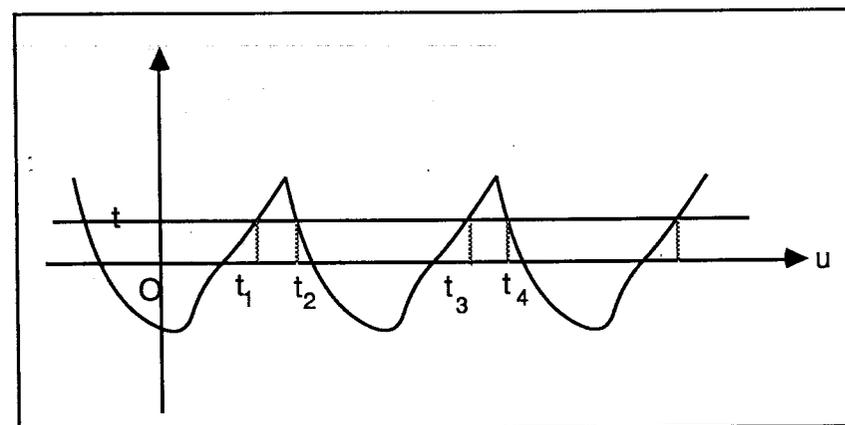


Figure 3.1.

La figure 3.1 permet de voir de calcul de $G(t)$ sur un exemple.

On peut comparer $G(t)$ avec la fonction de répartition de la variable $s(U)$ où U est choisi « au hasard » entre 0 et τ , c'est-à-dire de loi uniforme sur $[0, \tau]$

$$\begin{aligned} F(t) &= P(s(U) < t) = \int_0^\tau \frac{1}{\tau} \cdot \mathbb{1}_{\{s(u) < t\}} du \\ &= \frac{1}{\tau} \cdot \text{mesure}(\{u/0 < u < \tau \text{ et } s(u) < t\}) \end{aligned}$$

Ecrivons la décomposition de T :

$$T = n\tau + r \quad \text{avec } 0 \leq r < \tau$$

$s(t)$ étant un signal périodique, on peut écrire

$$\begin{aligned} \text{mesure} (\{u/0 < u < T \text{ et } s(u) < t\}) &= \\ &= n \text{ mesure} (\{u/0 < u < \tau^- \text{ et } s(u) < t\}) \\ &\quad + \text{mesure} (\{u/0 < u < \tau \text{ et } s(u) < t\}) \end{aligned}$$

ce qui donne,

$$\begin{aligned} G(t) &= \frac{n\tau}{T} F(t) + \frac{\text{mesure} (\{u/0 < u < \tau \text{ et } s(u) < t\})}{T} \\ &= F(t) + a(t) \text{ avec } |a(t)| \leq 2 \frac{\tau}{T} \end{aligned}$$

Lorsque le temps d'observation T augmente, le rapport $\frac{\tau}{T}$ tend vers 0 et $G(t)$ converge vers $F(t)$; la convergence est uniforme par rapport à la variable t .

Si on effectue maintenant n tirages aléatoires, indépendants, de loi uniforme, sur l'intervalle $[0, \tau]$, et que l'on pose

$$F_n(t) = \frac{\text{nombre de v.a.r. } U \text{ telles que } s(U) < t}{n}$$

d'après ce qui précède

$$F_n(t) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} F(t)$$

de sorte que pour n grand et T grand devant τ , on peut affirmer que $F_n(t)$ est peu différent de $G(t)$.

3.5. CONVERGENCE EN LOI

3.5.1. Définition

Dans la notion de convergence en loi on exprime le fait que la loi ressemble de plus en plus à une loi fixée. Il n'y a pas nécessairement convergence ponctuelle des variables.

Si l'on considère la situation de la loi des grands nombres avec X_1, X_2, \dots indépendantes et de même loi $\mathcal{N}(0, 1)$, on voit la loi de S_n/n se stabiliser; en effet,

$$P\left(\frac{S_n}{n} \in [a, b]\right) \rightarrow 0 \text{ si } a < b < 0 \text{ ou } 0 < a < b$$

$$P\left(\frac{S_n}{n} \in [a, b]\right) \rightarrow 1 \text{ si } a < 0 < b$$

Intuitivement la loi de S_n/n converge vers la mesure de Dirac au point 0. En fait, si on regarde $P(S_n/n \in A)$ lorsque n tend vers l'infini

$$P\left(\frac{S_n}{n} \in A\right) \xrightarrow[n \rightarrow +\infty]{} 1 \text{ si } 0 \text{ est à l'intérieur de } A$$

(au sens topologique du terme).

D'où la définition suivante de la convergence en loi :

Définition : Soit X_n une suite de variables aléatoires. On dit que la suite converge en loi si il existe une loi de probabilité μ , telle que $P(X_n \in A) \rightarrow \mu(A)$ pour les événements A dont le bord est μ -négligeable.

La convergence en loi permet de faire des calculs approchés pour n grand, en remplaçant $P(X_n \in A)$ par $\mu(A)$. De telles approximations sont les bienvenues pour les lois binomiales avec n grand, par exemple, car elles évitent de manipuler des factorielles. Tous les résultats se généralisent au cas des vecteurs aléatoires.

3.5.2. Caractérisations de la convergence en loi

La première des caractérisations est basée sur les fonctions de répartition :

Théorème : La suite X_n de v.a.r. converge en loi vers X si et seulement si les fonctions de répartition F_n convergent ponctuellement vers la fonction de répartition F de la variable X en tout point de continuité de F .

On trouvera la démonstration de ce théorème ainsi que ceux qui suivent, dans « Convergence of probability measures » de Billingsley.

La plus importante caractérisation se fait avec les fonctions caractéristiques (transformées de Fourier).

Théorème : La suite X_n converge en loi vers la loi de X si et seulement si la fonction caractéristique de X_n converge ponctuellement vers celle de X :

$$E(e^{itX_n}) \rightarrow E(e^{itX}) \text{ pour tout } t$$

$$n \rightarrow \infty$$

Théorème de Lévy : Si les fonctions caractéristiques convergent vers une fonction qui est continue en 0, alors les variables convergent en loi et la fonction limite est la fonction caractéristique de cette loi.

Autre caractérisation :

$$E(f(X_n)) \rightarrow E(f(X)) \text{ lorsque } f \text{ est continue bornée}$$

3.5.3. Application : Processus de Poisson

Considérons une file d'attente avec un taux d'arrivée par unité de temps α ; on suppose qu'il y a indépendance entre les variables correspondants à ces

intervalles de temps disjoints, que la loi du nombre d'arrivées dans un intervalle de temps ne dépend que de la longueur de cet intervalle et :

$$\begin{aligned} P(\text{aucun client entre } t \text{ et } \Delta t) &= 1 - \alpha \Delta t + \Delta t \varepsilon(\Delta t) \\ P(1 \text{ client exactement entre } t \text{ et } \Delta t) &= \alpha \Delta t + \Delta t \varepsilon(\Delta t) \end{aligned}$$

Appelons N_T le nombre de clients entre 0 et T , T étant un instant quelconque. On découpe l'intervalle de temps $[0, T]$ en n parties égales et on pose

$$\begin{aligned} X_1 &= \text{nombre de clients arrivés entre } 0 \text{ et } T/n \\ X_2 &= \text{nombre de clients arrivés entre } T/n \text{ et } 2T/n \\ &\vdots \\ X_n &= \text{nombre de clients arrivés entre } (n-1)T/n \text{ et } T \end{aligned}$$

Calculons la fonction caractéristique de N_T :

$$\begin{aligned} E(e^{iuN_T}) &= E(e^{i(X_1 + \dots + X_n)}) \\ &= [(1 - \alpha \Delta t) + \alpha \Delta t e^{iu} + \Delta t \varepsilon(\Delta t)]^n \\ &= \left[1 + \alpha \frac{T}{n} (e^{iu} - 1) + \frac{T}{n} \varepsilon\left(\frac{T}{n}\right) \right]^n \end{aligned}$$

Il y a convergence ponctuelle vers $e^{\alpha T(e^{iu} - 1)}$, lorsque n tend vers l'infini, c'est-à-dire lorsque le pas $\Delta t = \frac{T}{n}$ tend vers 0.

On reconnaît la fonction caractéristique d'une loi de Poisson de paramètre αT . On montrerait de même que le nombre d'arrivées entre t_1 et t_2 suit une loi de Poisson de paramètre $\alpha(t_2 - t_1)$. On obtient un processus de Poisson c'est-à-dire un processus qui vérifie :

- Si $0 \leq t_1 < t_2$, $N_{t_2} - N_{t_1}$ suit la loi $\mathcal{P}(\alpha(t_2 - t_1))$
- si $0 < t_1 < t_2 < \dots < t_k$, les variables $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_k} - N_{t_{k-1}}$ sont indépendantes.

On peut alors calculer la loi, du temps d'attente entre deux clients, T ,

$$\begin{aligned} P(T > a) &= P(\text{aucun client entre } 0 \text{ et } a) = P(N_a = 0) = e^{-\alpha a} \\ &= \int_a^{+\infty} \alpha e^{-\alpha u} du \end{aligned}$$

Il s'agit donc d'une loi exponentielle de paramètre α .

3.5.4. Théorème de la limite centrale

Si les variables X_1, X_2, \dots, X_n sont indépendantes, équidistribuées, de moyenne m et de variance σ^2 , alors

$$\sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - m \right) \rightarrow \mathcal{N}(0, \sigma^2)$$

La démonstration se fait en utilisant la fonction caractéristique ϕ des X_i et son développement limité à l'ordre 2 au voisinage de 0 :

$$\begin{aligned} E \left(e^{iu \sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - m \right)} \right) &= e^{-iu \sqrt{n} m} \left[\phi \left(\frac{t}{\sqrt{n}} \right) \right]^n \\ &= e^{-iu \sqrt{n} m} e^{n \log \phi \left(\frac{t}{\sqrt{n}} \right)} \\ &= e^{-\frac{\sigma^2 t^2}{2} + \varepsilon(n)} \rightarrow e^{-\frac{\sigma^2 t^2}{2}} \end{aligned}$$

Ce théorème se généralise au cas où les variables X_i ne sont plus à valeurs dans \mathbb{R} mais dans \mathbb{R}^n , c'est-à-dire sont des vecteurs aléatoires, en supposant l'existence du moment d'ordre 1 et de la matrice de covariance.

3.5.5. Convergence en loi pour des variables binomiales

3.5.5.1. Convergence vers une loi normale

Par application directe du théorème de la limite centrale, si S_n suit une loi $\mathcal{B}(n, p)$, alors

$$\sqrt{n} \left(\frac{S_n}{n} - p \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, p(1-p))$$

Dans la pratique, on remplace la loi de S_n par une loi $\mathcal{N}(np, np(1-p))$ lorsque n est suffisamment grand ; la convergence est d'autant plus rapide que p est proche de 1/2 (voir fig. 3.2).

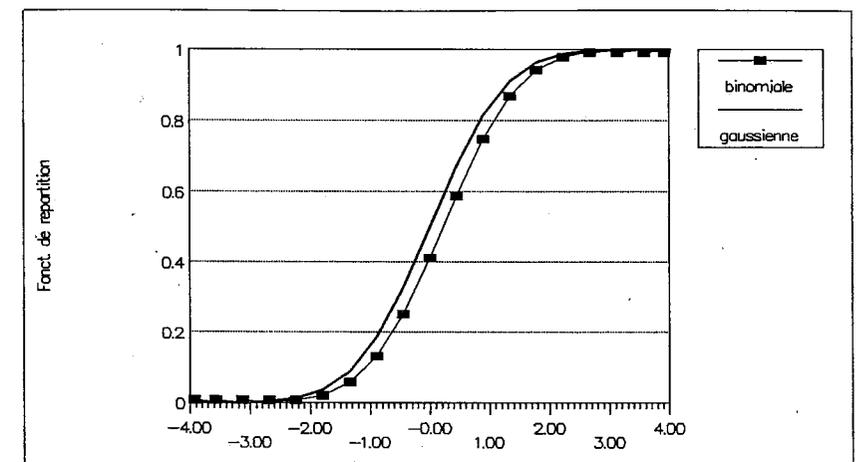


Fig. 3.2. — $n = 20$ et $p = 0,5$.

3.5.5.2. Convergence vers une loi de Poisson

Cette approximation sera utilisée pour n grand et p petit (événements rares). Si n tend vers l'infini et p tend vers 0 avec $np \rightarrow \lambda$, alors S_n converge en loi vers une loi de Poisson de paramètre λ (démonstration à l'aide des fonctions caractéristiques, laissée au lecteur).

Tableau récapitulatif pour $n \geq 30$

	$np \geq 5$ $n(1-p) \geq 5$	$np < 5$	$n(1-p) < 5$
Variable	S_n	S_n	$n - S_n$
Loi approximée par	$\mathcal{N}(np, np(1-p))$	$\mathcal{P}(np)$	$\mathcal{P}(n(1-p))$

Dans le cas $n(1-p) < 5$, on utilisera

$$P(S_n \in A) = P(n - S_n \in n - A)$$

3.5.6. Convergence en loi pour des variables gaussiennes

Si les variables X_n sont de loi $\mathcal{N}(m_n, \sigma_n^2)$

$$\phi_n(t) = e^{im_n t - \frac{\sigma_n^2 t^2}{2}}$$

alors $\phi_n(t)$ converge ponctuellement si et seulement si m_n et σ_n convergent et dans ce cas la limite est encore une fonction caractéristique de gaussienne. Ce résultat se généralise aux vecteurs gaussiens.

3.5.7. Convergence en loi pour les variables Qui-deux

Soit X_1, X_2, \dots une suite de v.a.r. de loi $\mathcal{N}(0, 1)$ et indépendantes. On pose

$$S_n = X_1^2 + X_2^2 + \dots + X_n^2$$

S_n étant la somme de n variables aléatoires indépendantes de même loi, d'espérance 1 et de variance 2, on en déduit par application du théorème de la limite centrale

$$\sqrt{n} \left(\frac{S_n}{n} - 1 \right) \xrightarrow[n \rightarrow +\infty]{\text{en loi}} \mathcal{N}(0, 2)$$

S_n suit une loi de Qui-deux à n degrés de liberté, notée χ_n^2 .

Dans la pratique, pour n supérieur à 30, on remplacera la loi d'un χ_n^2 par la loi $\mathcal{N}(n, 2n)$.

3.6. RÉCAPITULATION ET COMPARAISON DES DIFFÉRENTS TYPES DE CONVERGENCE

La convergence en loi a une importance pratique considérable. D'autre part, tous les types de convergence rencontrés dans ce chapitre entraînent la convergence en loi. En effet si X_n converge en probabilité vers X , alors

$$E(e^{ix_n}) = E(e^{ix_n} 1_{\{|X_n - X| > \eta\}}) + E(e^{ix_n} 1_{\{|X_n - X| \leq \eta\}})$$

Soit $\varepsilon > 0$; choisissons η de façon à avoir $|e^{ix_n} - e^{ix}| \leq \varepsilon/2$ lorsque $|X_n - X| \leq \eta$. Il suffit alors de choisir n assez grand pour avoir $P(|X_n - X| > \eta) \leq \frac{\varepsilon}{4}$ ce qui donne

$$|E(e^{ix_n}) - E(e^{ix})| \leq \varepsilon \text{ pour } n \geq n_0$$

ce qui permet de compléter les résultats obtenus et d'écrire :

$cv \text{ en m.q.} \Rightarrow cv \text{ en proba.} \Rightarrow cv \text{ en loi}$ \uparrow $cv \text{ p.s.}$
--

D'autre part si la suite de variables est bornée p.s. par une constante indépendante de n on a :

$$cv \text{ p.s.} + \left. \begin{array}{l} \text{les variables} \\ \text{sont bornées par une constante} \end{array} \right\} \Rightarrow cv \text{ en m.q.}$$

L'exercice 3.2 permet de mieux comprendre ces différents types de convergence et leur non-équivalence.

Citons des résultats particuliers aux variables indépendantes que l'on peut trouver dans le livre de Breiman :

Théorème : Soit X_1, X_2, \dots une suite de v.a.r. indépendantes ; alors la série de terme général X_n converge en probabilité si et seulement si elle converge p.s.

Corollaire : Soit X_1, X_2, \dots une suite de v.a.r. indépendantes et centrées telles que $\sum E(X_n^2) < +\infty$; alors la série de terme général X_n converge p.s.

Théorème des trois séries (Kolmogorov) : Soit X_1, X_2, \dots une suite de v.a.r. indépendantes ; alors

1) $\sum_n X_n$ cv p.s. $\Rightarrow \forall a > 0$, les 3 séries suivantes sont convergentes :

$$\sum_n P(|X_n| \geq a) \tag{1}$$

$$\sum_n E(X_n 1_{\{|X_n| \leq a\}}) \tag{2}$$

$$\sum_n \text{var}(X_n 1_{\{|X_n| \leq a\}}) \tag{3}$$

2) Si il existe $a > 0$ tel que les séries (1), (2) et (3) sont convergentes, alors $\sum X_n$ converge p.s.

EXERCICES

E.3.1. Montrer les équivalences

$$\begin{array}{ccc} \text{en proba.} & & \text{en proba.} \\ X_n \xrightarrow[n \rightarrow +\infty]{} X \Leftrightarrow X_n - X \xrightarrow[n \rightarrow +\infty]{} 0 & & \\ \text{p.s.} & & \text{p.s.} \\ X_n \xrightarrow[n \rightarrow +\infty]{} X \Leftrightarrow X_n - X \xrightarrow[n \rightarrow +\infty]{} 0 & & \end{array}$$

E.3.2. Soit X_n une suite de v.a. à valeurs dans \mathbb{N} , ainsi que la variable X . Montrer qu'on a l'équivalence

$$\begin{array}{ccc} \text{en loi} & & \\ X_n \xrightarrow[n \rightarrow +\infty]{} X \Leftrightarrow \forall k, P(X_n = k) \xrightarrow[n \rightarrow +\infty]{} P(X = k) & & \end{array}$$

Etudier le cas où $P(X_n = 0) = 1/2 = P(X_n = 1)$ et $X_{n+1} = 1 - X_n$; montrer qu'il n'y a pas convergence en probabilité.

Etudier le cas où les variables X_n sont indépendantes, de loi donnée par $P(X_n = 0) = 1 - \frac{1}{n}$ $P(X_n = 1) = \frac{1}{n}$.

En déduire que la convergence en loi n'est pas équivalente à la convergence en probabilité, qui n'est pas équivalente à la convergence presque sûre.

E.3.3. Soit U_n une suite de v.a. indépendantes, de loi uniforme sur $[a, b]$. On pose

$$S_n = \text{Sup}(U_1, U_2, \dots, U_n) \quad I_n = \text{Inf}(U_1, U_2, \dots, U_n)$$

Montrer que S_n converge presque sûrement vers b et I_n converge p.s. vers a .

E.3.4. Méthode de Monte Carlo.

$$\text{Soit} \quad I = \int_0^1 \sqrt{x^4 + x^2 + x + 1} \, dx$$

$$\text{On pose} \quad f(x) = \sqrt{x^4 + x^2 + x + 1}$$

et on considère les variables $X_i = f(U_i)$ où U_1, U_2, \dots sont indépendantes, de loi uniforme sur $[0, 1]$.

Montrer que $\frac{X_1 + X_2 + \dots + X_n}{n}$ converge presque sûrement vers I .

E.3.5. Le directeur d'un grand magasin désire estimer la proportion d'achats avec un chèque sans provision parmi tous les paiements par chèque. Des études antérieures dans d'autres magasins montrent que cette proportion est aux alentours de 5%. Combien faut-il examiner de dossiers pour estimer cette proportion à 10^{-3} près, avec 95% de chances (on utilisera le théorème de la limite centrale).

E.3.6. Inégalité de Chernoff.

1) Soit X une v.a.r. admettant une transformée de Laplace sur un voisinage de 0. Montrer les inégalités

$$\begin{array}{l} 1_{\{X > a\}} \leq e^{t(X-a)} \quad \forall t > 0 \\ P(X > a) \leq \inf_{t > 0} E(e^{t(X-a)}) \end{array}$$

2) Soit X_1, X_2, \dots des variables de Bernoulli indépendantes et de même loi. On pose $S_n = X_1 + X_2 + \dots + X_n$, de sorte que S_n suit la loi $\mathcal{B}(n, p)$. Montrer

$$\begin{array}{l} P\left(\frac{S_n}{n} - p > a\right) \leq \\ \leq \exp\left\{-n\left[(p+a)\log\left(\frac{1-p}{p} \cdot \frac{p+a}{1-p-a}\right) - \log\frac{1-p}{1-p-a}\right]\right\} \end{array}$$

E.3.7. Expérience de Weldon.

Un dé est lancé 315 672 fois et on constate l'apparition du 5 ou 6, 106 602 fois. Si l'on suppose le dé parfaitement équilibré et que l'on appelle S_n la variable « nombre d'apparitions du 5 ou du 6 », calculer

$$P(S_n/n > 0,337)$$

pour $n = 315 672$. Pensez-vous que les 6 faces du dé sont équiprobables. Expliquer physiquement ce résultat.

E.3.8. On reprend l'étude du protocole ALOHA vu au chapitre 1 (exercice E.1.9). Pour chaque entier k on est dans l'état k à l'instant t si il y a k messages en attente et on pose :

$$I_k(\omega) = \begin{cases} 0 & \text{si } k \text{ n'est jamais atteint} \\ 1 & \text{si le premier départ de l'état } k \text{ se fait vers l'état } k-1 \\ 0 & \text{sinon} \end{cases}$$

Montrer (sans calcul explicite de la somme) que l'on a $\sum_k P(I_k = 1) < +\infty$. En déduire

$$\sum_k I_k \underset{\text{p.s.}}{<} +\infty$$

En déduire que, presque sûrement, à partir d'un certain rang, le nombre de messages en attente ne fait qu'augmenter.

E.3.9. On considère un processus de naissance où X_n est le nombre d'individus à la n -ième génération, c'est-à-dire au temps n . On suppose $X_0 = 1$; chaque individu présent au temps n , laisse place à $Y_{n,i}$ individus de sorte que l'on a

$$X_{n+1} = \sum_{i=1}^{X_n} Y_{n,i}$$

On suppose les variables $Y_{n,i}$ indépendantes, de même loi :

$$\begin{cases} P(Y_{n,i} = k) = p_k & k = 1, 2, \dots \\ P(Y_{n,i} = 0) = 0 & \text{et } P(Y_{n,i} = 1) < 1 \end{cases}$$

a) Montrer que l'on a

$$P(X_{n+1} = X_n / X_{n_0} = x) \leq p_1 \quad \text{pour } n \geq n_0$$

$$P(X_n = X_{n+1} = \dots = X_{n+k} / X_n = x) \leq p_1^k$$

En déduire que la suite X_n tend vers $+\infty$, presque sûrement.

b) On pose $G(t) = E(t^{Y_{1,1}})$ et on suppose que $Y_{1,1}$ admet une espérance m .

Calculer $E(t^{X_{n+1}} / X_n = N)$.

En déduire que $\frac{X_n}{m^n}$ est une martingale, c'est-à-dire,

$$E\left(\frac{X_{n+1}}{m^{n+1}} / X_0, X_1, \dots, X_n\right) = \frac{X_n}{m^n}$$

SIMULATION SUR ORDINATEUR DE PHÉNOMÈNES ALÉATOIRES

4.1. LOI UNIFORME : RÉSULTATS THÉORIQUES (cas discret et cas continu)

4.1.1. Rappels

On rappelle la définition d'une loi uniforme sur un ensemble fini ou sur un intervalle de \mathbb{R} :

Soit X de Ω dans un ensemble fini à N éléments, une v.a. ; X suit une loi uniforme si elle prend les N différentes valeurs avec la même probabilité $1/N$.

Soit X une v.a.r. et $[a, b]$ un intervalle borné ; X suit une loi uniforme sur $[a, b]$ si elle a pour densité

$$f(x) = \begin{cases} 0 & \text{si } x \notin [a, b] \\ \frac{1}{b-a} & \text{si } x \in [a, b] \end{cases}$$

On a alors le résultat suivant : Si X_1, X_2, \dots sont indépendantes, de même loi uniforme sur $\{0, 1, \dots, 9\}$ alors la série

$$\sum_{n=1}^{+\infty} \frac{X_n(\omega)}{10^n}$$

est convergente (pour toute épreuve ω de Ω) et définit une variable aléatoire à valeurs dans $[0, 1]$ qui suit une loi uniforme. Réciproquement, si X suit une loi uniforme sur $[0, 1]$ et que l'on appelle X_n le n -ième chiffre de l'écriture décimale de X , alors les variables X_n sont indépendantes et de loi uniforme sur $\{0, 1, \dots, 9\}$; en effet,

$\forall n \geq 1, P(X_1 = k_1 X_2 = k_2 \dots X_n = k_n)$

$$\begin{aligned} &= P\left(\frac{k_1}{10} + \dots + \frac{k_n}{10^n} \leq X \leq \frac{k_1}{10} + \dots + \frac{k_n}{10^n} + \frac{1}{10^n}\right) \\ &= \frac{1}{10^n} \end{aligned}$$

D'où

$$P(X_i = k_i) = \frac{1}{10} \quad \text{et} \quad P(X_1 = k_1 \dots X_n = k_n) = P(X_1 = k_1) \dots P(X_n = k_n)$$

On a le même résultat si on utilise une décomposition binaire :

$$X = \sum_{n=1}^{+\infty} \frac{X_n}{2^n} \quad \text{avec } X_n = 0 \text{ ou } 1$$

les variables X_n étant uniformes sur $\{0, 1\}$.

Ces deux résultats montrent comment on peut construire théoriquement un nombre réel « au hasard » entre 0 et 1 : on choisit les décimales au hasard et indépendamment les unes des autres. Ce résultat reste théorique puisque dans la pratique on manipule des nombres réels qui sont des décimaux (ayant un nombre fini de décimales non nulles).

4.1.2. Cas discret et cas continu

On peut aussi voir la relation entre variables uniformes discrètes et variables uniformes continues en considérant X_n de loi uniforme sur $\{0, 1, \dots, n-1\}$ et $Y_n = \frac{X_n}{n}$.

Soit $t \in [0, 1]$

$$P(Y_n < t) = P(X_n < nt) = \frac{k}{n} \quad \text{avec } k = \text{partie entière de } nt$$

$$P(Y_n < t) \xrightarrow{n \rightarrow +\infty} t$$

Lorsque n est grand Y_n suit approximativement une loi uniforme sur $[0, 1]$. C'est la méthode employée dans la pratique.

4.1.3. Relations entre variables uniformes continues

Si X suit la loi uniforme sur $[a, b]$ alors $\alpha X + \beta$ suit une loi uniforme sur $[\alpha a + \beta, \alpha b + \beta]$ si $\alpha > 0$; en effet,

$$P(\alpha X + \beta < t) = P\left(X < \frac{t - \beta}{\alpha}\right)$$

$$\frac{d}{dt} P(\alpha X + \beta < t) = \begin{cases} 0 & \text{si } \frac{t - \beta}{\alpha} \notin [0, 1] \\ \frac{1}{\alpha} & \text{sinon} \end{cases}$$

Par exemple, si U suit une loi uniforme sur $[0, 1]$ alors $2\pi U$ suit une loi uniforme sur $[0, 2\pi]$ et $(b-a)U + a$ suit une loi uniforme sur $[a, b]$.

4.2. GÉNÉRATEURS DE NOMBRES PSEUDO-ALÉATOIRES

4.2.1. Introduction

Dans un certain nombre de problèmes scientifiques il est nécessaire d'utiliser une source de nombres aléatoires et lorsqu'on veut comparer plusieurs méthodes, on les compare en faisant passer la même suite de nombres aléatoires ; on peut trouver quelques exemples de tels problèmes au paragraphe 5.

Il existe des banques de nombres au hasard, nombres fabriqués par des moyens mécaniques ou plus généralement, physiques ; ces nombres ont toutes les propriétés requises mais ont l'inconvénient de nécessiter un stockage : il faudra alors prendre de la place mémoire et prendre du temps pour aller lire dans le fichier qui contient ces nombres d'où l'apparition de méthodes de génération de nombres dits pseudo-aléatoires, qui sont calculés à l'intérieur du programme. La méthode la plus fréquemment employée est du type

$$X_{n+1} = aX_n + b \pmod{m}$$

La suite obtenue à partir d'une valeur initiale est censée ressembler à une réalisation d'une suite de variables indépendantes, uniformes sur $\{0, \dots, m-1\}$.

4.2.2. Propriétés arithmétiques des congruences linéaires

Les propriétés arithmétiques des relations de congruences linéaires ont été étudiées de façon intensive dans les années 1960 : on peut trouver l'historique ainsi que les démonstrations dans « The art of computer programming » vol. 2 de Knuth, Editions Addison Wesley.

Une telle suite peut prendre au plus m valeurs différentes et admet une période. On a intérêt à avoir m le plus grand possible ainsi que la période ; cette période dépend de la valeur initiale X_0 , sauf si elle est égale à m car dans ce cas toutes les valeurs de l'ensemble $\{0, \dots, m-1\}$ sont visitées ; il est donc intéressant de choisir les coefficients pour avoir une période égale à m exactement :

Théorème : La période est égale à m si et seulement si

- 1) b est premier avec m ,
- 2) $a-1$ est multiple de tout nombre premier qui divise m , multiple de 4 si m est multiple de 4.

Les relations de congruences linéaires avec $b=0$ sont plus rapides mais ne peuvent donner une période égale à m : la période est égale au plus à $m-1$ si m est premier, à 2^{n-2} si $m=2^n$ et $m \geq 8$, à $\lambda(m)$ dans le cas général avec

$$\begin{aligned} \lambda(2) &= 1, \lambda(4), \lambda(2^n) = 2^{n-2} \quad \text{si } n \geq 3 \\ \lambda(p^n) &= (p-1)p^{n-1} \quad \text{si } p \text{ est premier et } p \geq 3 \\ \lambda(m) &= \text{PPCM}(\lambda(p^{n_1}), \dots, \lambda(p^{n_k})) \quad \text{si } m = p^{n_1} \dots p^{n_k} \end{aligned}$$

La plus grande période est alors atteinte en prenant X_0 premier avec m et a d'ordre maximum modulo m ce qui donne $a = 3$ ou 5 modulo 8 si $m = 2^n$ et $m > 8$.

En ce qui concerne les propriétés statistiques, on pourra consulter le livre de Knuth. Un exemple de générateur donnant de bons résultats est celui de la librairie SANE :

$$X_n = (7^5 X_{n-1}) \bmod (2^{31} - 1)$$

4.2.3. Génération de 0, 1 pseudo-aléatoires

On peut construire une suite X_n de 0, 1 par la relation

$$X_n = a_1 X_{n-1} + a_2 X_{n-2} + \dots + a_k X_{n-k} \bmod 2$$

Si on choisit bien les coefficients, alors la suite des mots binaires de longueur k successifs : $X_1 X_2 \dots X_k, X_{k+1} X_{k+2} \dots X_{2k}, \dots$ a pour période $2^k - 1$; c'est le cas pour $k = 30$, par exemple, si on prend

$$a_1 a_2 \dots a_{30} = 0 \dots 01010011$$

4.3. SIMULATION DE VARIABLES ALÉATOIRES DISCRÈTES

4.3.1. Variable prenant un nombre fini de valeurs

Si la v.a. X prend ses valeurs dans l'ensemble fini $\{x_1, x_2, \dots, x_n\}$ avec les probabilités respectives p_1, p_2, \dots, p_n , on peut simuler X à l'aide d'une variable U uniforme sur $[0, 1]$ en posant :

$$\begin{aligned} X = x_1 & \text{ si } 0 \leq U \leq p_1 \\ & x_2 \text{ si } p_1 < U \leq p_1 + p_2 \\ & \vdots \\ & x_n \text{ si } p_1 + \dots + p_{n-1} < U \leq 1 \end{aligned}$$

Cette méthode est utilisable pour une loi binomiale $\mathcal{B}(n, p)$ si n n'est pas trop grand ; on pourra tenir compte du fait que la somme de deux variables binomiales indépendantes de paramètre p est encore une binomiale de paramètre p : pour construire X de loi $\mathcal{B}(N, p)$ on écrit $N = N_1 + N_2$ et on construit X_1 de loi $\mathcal{B}(N_1, p)$ et X_2 de loi $\mathcal{B}(N_2, p)$ par la méthode précédente de façon à ne faire des simulations que pour des valeurs faibles de n . Lorsque N est supérieur à 30, on utilise les approximations vues au chapitre précédent : approximation par une loi de Poisson ou par une loi gaussienne.

4.3.2. Loi de Poisson $\mathcal{P}(\lambda)$

Soit U_1, U_2, \dots une suite de variables indépendantes et uniformes sur $[0, 1]$. On peut alors montrer par récurrence sur n la relation

$$\begin{aligned} P(U_1 \geq p ; U_1 U_2 \geq p \dots ; U_1 U_2 \dots U_n \geq p ; U_1 U_2 \dots U_{n+1} < p) = \\ = \frac{p}{n!} (-1)^n [\ln(p)]^n \end{aligned}$$

pour $0 < p < 1$.

On pose alors :

$$\begin{aligned} X = 0 & \text{ si } U_1 < e^{-\lambda} \\ & 1 \text{ si } U_1 \geq e^{-\lambda} \text{ et } U_1 U_2 < e^{-\lambda} \\ & \vdots \\ & n \text{ si } U_1 \geq e^{-\lambda} \text{ et } U_1 U_2 \geq e^{-\lambda} \dots ; \text{ et } U_1 U_2 \dots U_n \geq e^{-\lambda} ; \\ & \text{ et } U_1 U_2 \dots U_{n+1} < e^{-\lambda} \end{aligned}$$

et on a

$$P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

c'est-à-dire que X suit la loi $\mathcal{P}(\lambda)$.

4.3.3. Loi géométrique

Soit U de loi uniforme alors la variable définie par

$$X = \text{partie entière de } \frac{\text{Ln}(U)}{\text{Ln}(1-p)}$$

suit une loi géométrique de paramètre p , c'est-à-dire vérifie

$$P(X = k) = (1-p)^{k-1} p$$

Remarque : On trouvera dans l'article de T. Kämpke « Multiple use of random numbers in discrete-event simulation », (Math. and computers in Simulation, n° 31 — 1989 — p. 171, 176) une méthode astucieuse pour obtenir N réalisations indépendantes de la loi de X , lorsque

$$P(X = x_1) = p_1 \dots P(X = x_n) = p_n$$

en utilisant une seule variable uniforme U , de loi uniforme sur $[0, 1]$.

4.4. SIMULATION DE VARIABLES ALÉATOIRES CONTINUES

4.4.1. Théorème fondamental et ses applications

Théorème : Soit X une v.a.r. de fonction de répartition F continue, alors la variable $F(X)$ suit une loi uniforme sur $[0, 1]$.

Démonstration : F est croissante, à valeurs dans $[0, 1]$ et continue. Néanmoins elle n'est pas nécessairement strictement croissante :

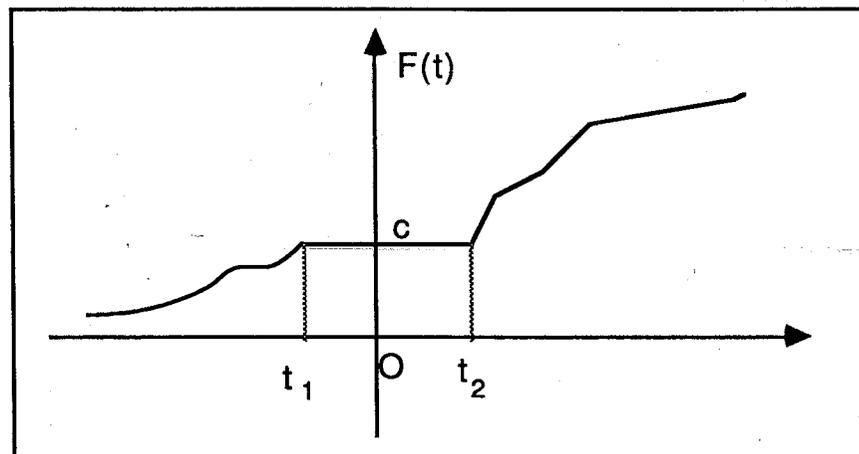


Figure 4.1.

Soit c dans $]0, 1[$ et posons $t(c) = \inf \{t/F(t) = c\}$; on a alors

$$\begin{aligned} F(t) < c &\Leftrightarrow t < t(c) \\ P(F(X) < c) &= P(X < t(c)) = F(t(c)) = c \end{aligned}$$

Ce théorème sera utilisé en prenant $X = F^{-1}(U)$ où U est uniforme sur $[0, 1]$, comme le montrent les exemples ci-dessous.

4.4.1.1. Loi de Cauchy

X suit une loi de Cauchy si sa densité est égale à $\frac{1}{\pi} \cdot \frac{1}{1+X^2}$.

Dans ce cas $F(t) = \frac{1}{2} + \frac{1}{\pi} \text{Arctg}(t)$.

Construction de X : Soit U de loi uniforme sur $[0, 1]$; $\Theta = \pi \left(U - \frac{1}{2} \right)$ est uniforme sur $\left] -\frac{\pi}{2}, \frac{\pi}{2} \right[$ et $\text{tg}(\Theta)$ suit une loi de Cauchy.

$$X = \text{tg} \left[\pi \left(U - \frac{1}{2} \right) \right]$$

4.4.1.2. Loi exponentielle

Comme dans le cas précédent la fonction de répartition est facile à inverser.

$$F(t) = 1 - e^{-\lambda t}$$

Si U est uniforme sur $[0, 1]$ alors $-\frac{1}{\lambda} \ln(1-U)$ suit une loi exponentielle de paramètre 1. $1-U$ est encore de loi uniforme sur $[0, 1]$, d'où la construction :

$$X = -\frac{1}{\lambda} \ln(U)$$

4.4.2. Loi gaussienne $\mathcal{N}(m, \sigma^2)$

On utilise la propriété fondamentale des variables gaussiennes : Si X est de loi $\mathcal{N}(0, 1)$ alors $\sigma X + m$ est de loi $\mathcal{N}(m, \sigma^2)$, pour se ramener au cas $m = 0$ et $\sigma = 1$.

Comme il n'existe pas d'autres méthodes que les méthodes numériques pour calculer F^{-1} , on utilise des techniques différentes pour les gaussiennes : il y en a essentiellement deux ; la première est basée sur le théorème de la limite centrale et consiste à considérer la variable

$$X = U_1 + U_2 + \dots + U_{12} - 6$$

où U_1, \dots, U_{12} sont indépendantes et uniformes sur $[0, 1]$.

X a pour espérance 0, pour variance 1 et une loi proche de celle d'une $\mathcal{N}(0, 1)$.

La seconde méthode consiste à construire deux variables X et Y indépendantes, de loi $\mathcal{N}(0, 1)$ en remarquant :

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

En passant en coordonnées polaires, $X = R \cos \Theta$ et $Y = R \sin \Theta$ et on a

$$\begin{aligned} P(R^2 < t \mid 0 \leq \Theta \leq \theta_0) &= \iint_{\substack{0 < r^2 < t \\ 0 \leq \theta \leq \theta_0}} e^{-\frac{r^2}{2}} r \frac{1}{2\pi} dr d\theta \\ &= \int_0^t \frac{1}{2} e^{-\frac{u}{2}} du \times \int_0^{\theta_0} \frac{1}{2\pi} d\theta \end{aligned}$$

D'où la construction : Soit U et V deux variables indépendantes de loi uniforme sur $[0, 1]$, alors X et Y définies par :

$$\begin{aligned} X &= \sqrt{-2 \ln(U)} \cos(2\pi V) \\ Y &= \sqrt{-2 \ln(U)} \sin(2\pi V) \end{aligned}$$

sont indépendantes et de loi $\mathcal{N}(0, 1)$.

4.5. SIMULATION DE LOIS PLUS COMPLEXES

4.5.1. Lois mélanges

Il s'agit de v.a. dont la densité s'écrit sous la forme :

$$f(x) = \sum_k p_k f_k(x) \quad (1)$$

avec

$$0 \leq p_k \leq 1 \quad \text{et} \quad \sum_k p_k = 1$$

On rencontre ce type de loi par exemple, lorsqu'on a des données de loi $\mathcal{N}(0, 1)$ mélangées avec des données erronées de loi $\mathcal{N}(2, 1)$. Si on appelle p la proportion de données erronées et si on suppose les v.a. indépendantes lorsqu'on choisit une donnée au hasard, elle est la réalisation d'une loi de densité :

$$p\phi(x-2) + (1-p)\phi(x)$$

en appelant ϕ la densité d'une loi $\mathcal{N}(0, 1)$.

La simulation d'une variable de loi donnée par (1) se fait en deux étapes, la première étant la simulation d'une variable discrète prenant les valeurs $1, 2, \dots, n$ avec les probabilités respectives p_1, p_2, \dots, p_n . Si cette variable est égale à k , on simule une variable de densité f_k :

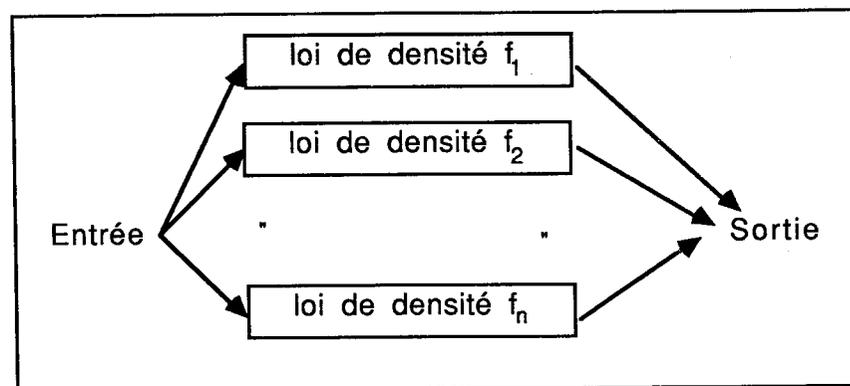


Figure 4.2.

4.5.2. Lois de Cox

Ce sont les lois qui résultent du schéma suivant :

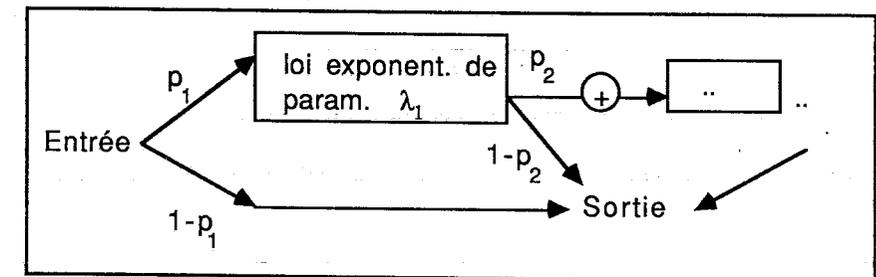


Figure 4.3.

Cox a démontré que toutes les densités de probabilité sur \mathbb{R} ayant une transformée de Laplace rationnelle sont de ce type. On rencontre les lois de Cox dans les problèmes de file d'attente où un « client » rentre dans un réseau de stations ; le temps de service dans une station est supposé de loi exponentielle, et le temps total passé dans le réseau est la somme des temps passés dans les différentes stations.

4.6. EXEMPLES DE PROBLÈMES TRAITÉS PAR SIMULATION

4.6.1. Minimisation d'un coût

Une compagnie d'aviation cherche à rentabiliser ses vols réguliers.

Considérons un avion qui contient N places ; des études antérieures ont montré que lorsqu'un client réserve une place, il a en fait une probabilité p de ne pas annuler sa réservation. La compagnie a donc intérêt à accepter plus de réservations que de places, mais jusqu'à combien ? Elle doit tenir compte du fait que, en cas de conflit (plus de clients présents que de places dans l'avion), le mécontentement des clients non satisfaits va rejallir sur la compagnie sous forme d'un coût. Posons alors

n = nombre maximal de réservations consenties

X = nombre de présents au départ de l'avion

$a(k)$ = coût dû à k places vides dans l'avion

A = profit lorsque l'avion est plein

$b(k)$ = coût dû à k clients mécontents, n'ayant pu avoir de places.

Si n réservations sont prises (n sera choisi supérieur à N), le profit est égal à

$$\text{Profit} = \begin{cases} A - a(k) & \text{si } X = N - k \quad \text{et } k \geq 0 \\ A - b(k) & \text{si } X = N + k \quad \text{et } k \geq 0 \end{cases}$$

On peut essayer, connaissant A , les fonctions a et b , de trouver la valeur de n qui rend maximum l'espérance de la variable aléatoire Profit,

$$E(\text{Profit}) = A - \sum_{k=1}^N a(k) P(X = N - k) - \sum_{k=1}^{n-N} b(k) P(X = N + k)$$

$$P(X = j) = C_j^n (1-p)^j p^{n-j}$$

mais on peut aussi simuler le comportement des clients un grand nombre de fois et calculer le profit moyen obtenu qui, d'après la loi des grands nombres est proche de $E(\text{profit})$. En faisant varier n , on pourra ainsi tracer la courbe du profit moyen en fonction de n et trouver la valeur de n optimale.

4.6.2. Calcul d'une intégrale par la méthode de Monte Carlo

Supposons que nous voulions calculer l'intégrale multiple I :

$$I = \iiint_{\substack{0 \leq x \leq 1 \\ 0 \leq y \leq 1 \\ 0 \leq z \leq 1}} \sqrt{2x^2 + y + 1} \cos(2\pi xz) dx dy dz$$

Cette intégrale peut être interprétée comme une espérance :

$$I = E\{f(U, V, W)\}$$

où

$$f(x, y, z) = \sqrt{2x^2 + y + 1} \cos(2\pi xz)$$

U, V, W sont indépendantes et uniformes sur $[0, 1]$

En simulant $3n$ variables uniformes, on calcule n valeurs de la fonction f : f_1, f_2, \dots dont la moyenne

$$\frac{f_1 + f_2 + \dots + f_n}{n}$$

est proche de l'espérance (loi des grands nombres), c'est-à-dire I .

On peut estimer l'erreur grâce au théorème de la limite centrale ; en effet, si on appelle σ l'écart type de la variable aléatoire $f(U, V, W)$, on a

$$P\left(\frac{\sqrt{n}}{\sigma} \left| \frac{f_1 + \dots + f_n}{n} - I \right| \leq 1,96\right) \rightarrow 95\%$$

Sur l'exemple ci-dessus on ne connaît pas σ mais on peut majorer σ^2 par $E(f(U, V, W)^2)$ qui est elle-même inférieure ou égale à $\frac{13}{6}$ ce qui donne environ (n doit être grand) 95 % de chances de réaliser l'inégalité :

$$\left| \frac{f_1 + \dots + f_n}{n} - I \right| \leq \frac{2,88}{\sqrt{n}}$$

4.6.3. Simulation de files d'attente

Considérons un exemple où n guichets assurent un service à des clients : les temps de service des différents guichets sont supposés aléatoires, indépendants et de loi connue. Les clients arrivent suivant une loi elle aussi connue (en fait, on se donne la loi du temps écoulé entre deux arrivées) et rejoignent une file d'attente unique. Lorsqu'un guichet se libère, le premier client de la file d'attente se présente. On peut alors étudier, par simulation, la relation entre le temps moyen d'attente et le nombre de guichets.

Il existe des langages informatiques qui permettent de décrire et simuler un réseau de files d'attente. Ces langages sont utilisés dans la modélisation des systèmes informatiques pour l'étude du parallélisme de fonctionnement entre les différents composants du système. Le plus couramment utilisé est le langage GPSS (General Purpose Simulation System). On trouvera de nombreux exemples dans le livre « Introduction aux réseaux de files d'attente » de Gelenbe et Pujolle, Editions Eyrolles.

4.7. LE MOUVEMENT BROWNIEN

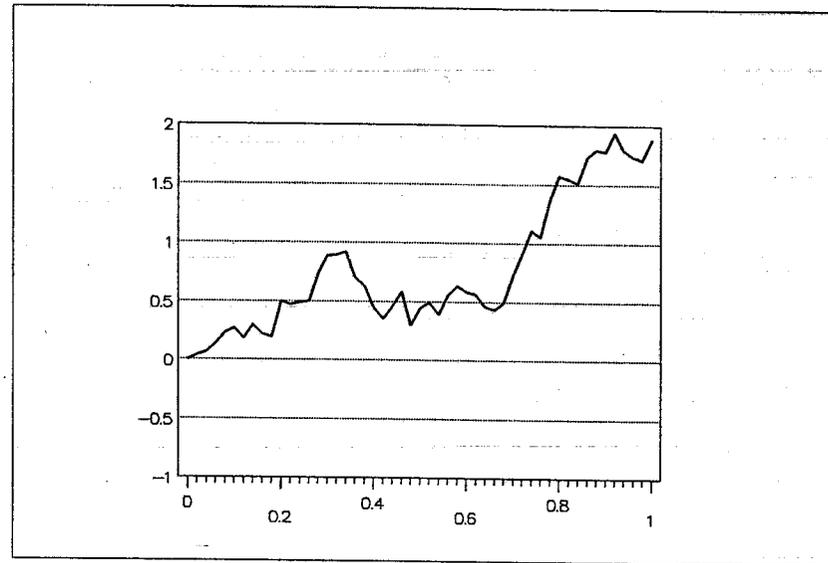
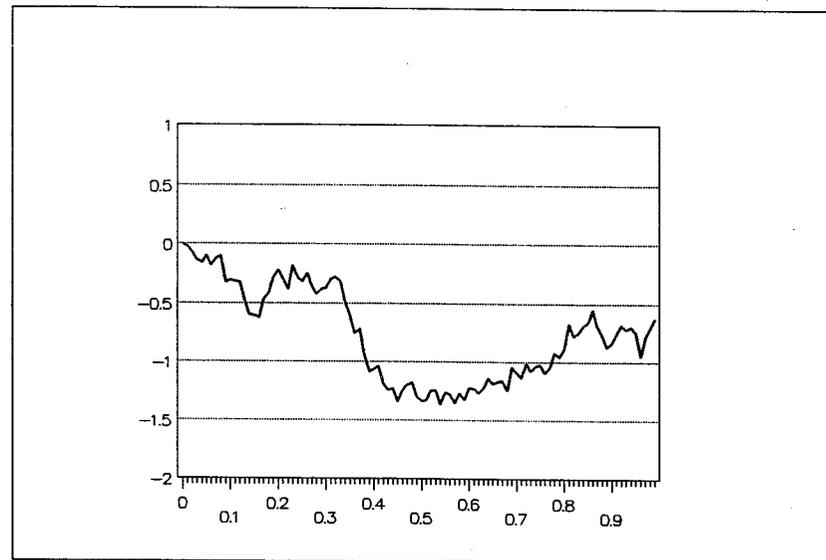
Soit X_1, X_2, \dots, X_n des variables indépendantes de loi $\mathcal{N}(0, 1)$. On construit à partir de ces variables une fonction S_n (aléatoire) définie pour $t \in [0, 1]$, en posant :

$$S_n(t) = \begin{cases} 0 & \text{si } t = 0 \\ \frac{X_1}{\sqrt{n}} & \text{si } t = \frac{1}{n} \\ \vdots \\ \frac{X_1 + \dots + X_k}{\sqrt{n}} & \text{si } t = \frac{k}{n} \\ \vdots \\ \frac{X_1 + \dots + X_n}{\sqrt{n}} & \text{si } t = 1 \end{cases}$$

$S_n(t)$ continue, affine sur chaque intervalle $\left[\frac{k}{n}, \frac{k+1}{n}\right]$

Lorsque n est grand, la fonction obtenue est « agitée » (Figs. 4.4 et 4.5) et proche de ce que l'on appelle un mouvement brownien (voir chap. 5 et le livre de Billingsley). On peut ensuite réaliser une simulation du mouvement d'une particule en suspension dans un liquide et soumise à un bombardement moléculaire, en écrivant l'équation sur la quantité de mouvement :

$$\Delta M = m \Delta V - \beta V \Delta t$$

Fig. 4.4. — Exemple avec $n = 50$.Fig. 4.5. — Exemple avec $n = 100$.

βV est la force de résistance due à la viscosité et M évolue comme un mouvement brownien, de sorte que l'on obtient l'évolution de la vitesse :

$$mV(t + \Delta t) = mV(t) - \beta V(t) \Delta t + \Delta M$$

et on a vu au début de ce paragraphe que l'on pouvait simuler ΔM par $\frac{X}{\sqrt{n}}$.

4.8. REMARQUES

Cette liste n'est pas exhaustive. Il y manque notamment une méthode sophistiquée de simulations qui utilise un changement de probabilité. La méthode est employée pour étudier des événements rares ; le changement de probabilité rend ces événements plus probables donc plus faciles à observer, et améliore la qualité des estimations que l'on peut faire sur le phénomène. On pourra se reporter à l'article « Large deviation and rare events in the study of algorithms » de Cottrel, Fort, Malgouyres. *I.E.E.E., Trans. on Automatic Control*. Vol. AC 28 n° 99 Sept. 83 ainsi que « Stochastic approximation via large deviations : asymptotic properties » de Kushner et Dupuis *SIAM Journal on Control and Optimisation*. Vol. 23 p. 675-695 (1985).

EXERCICES

E.4.1. Si X_n est calculé suivant la relation de récurrence

$$X_{n+1} = aX_n + bX_{n-1} \pmod{m}$$

Quelle est la plus grande période possible ?

E.4.2. 1) Soit X, Y, Z trois v.a.r. indépendantes, de même loi et de densité f . Montrer que l'on a

$$P(X < Y < Z) = \frac{1}{6}$$

2) Montrer que dans la suite définie par la relation (inspirée de Fibonacci) :

$$X_{n+1} = X_n + X_{n-1} \pmod{m}$$

on n'a jamais $X_{n-1} < X_{n+1} < X_n$.

E.4.3. Soit $I = \iint_D f(x, y) dx dy$ où D est un domaine fermé borné de \mathbb{R}^2 tel que $D \subset [a, b] \times [c, d]$. On pose $g(x, y) = f(x, y)$ si $(x, y) \in D$, 0 sinon. Montrer que l'on peut utiliser la méthode de Monte Carlo, pour calculer une valeur approchée de I .

E.4.4. On reprend les notations de l'exercice 4.3. D est maintenant un disque centré sur $(1, 1)$ de rayon 5. Montrer que l'on peut utiliser la méthode de Monte Carlo avec une variable R et une variable θ , de lois uniformes.

E.4.5. Soient X et Y deux variables indépendantes, de loi gaussienne de variance 1 avec $E(X) = m_1$, $E(Y) = m_2$.

1) En utilisant une rotation, montrer que le vecteur $\begin{pmatrix} X \\ Y \end{pmatrix}$ a même matrice de covariance que $\begin{pmatrix} X - m_1 + d \\ Y - m_2 \end{pmatrix}$ avec $d = \sqrt{m_1^2 + m_2^2}$.

2) Soit $\begin{pmatrix} X \\ Y \end{pmatrix}$ un vecteur gaussien de loi $\mathcal{N}\left(\begin{pmatrix} d \\ 0 \end{pmatrix}, I_d\right)$ et $R = \sqrt{X^2 + Y^2}$. Montrer que R a pour densité

$$f(r) = \begin{cases} \frac{r}{\sigma^2} I_0\left(\frac{rd}{\sigma^2}\right) e^{-\frac{r^2+d^2}{2\sigma^2}} & \text{si } r > 0 \\ 0 & \text{sinon} \end{cases}$$

avec

$$I_0(a) = \frac{1}{2\pi} \int_0^{2\pi} e^{a \cos \theta} d\theta$$

(f est la densité d'une loi de Rice).

E.4.6. 1) Montrer qu'en choisissant bien n et les p_i dans la figure 4.3, une loi exponentielle est une loi de Cox.

2) Soit X_1, X_2, \dots des variables aléatoires indépendantes, de même loi exponentielle de paramètre λ . Par récurrence sur n , montrer que $X_1 + X_2 + \dots + X_n$ a pour densité (loi de Erlang) :

$$\frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \quad \text{si } x \geq 0$$

En déduire, en choisissant les p_i dans la figure 4.3, que toute loi de Erlang est une loi de Cox.

E.4.7. Soit X une v.a.r. de densité $f(x)$ telle que

$$\begin{cases} f(x) = 0 & \text{si } x \notin [a, b] \\ 0 \leq f(x) \leq M \end{cases}$$

Soit U_1, U_2 deux v.a.r. indépendantes de loi uniforme sur $[0, 1]$. On pose

$$\begin{cases} V_1 = a + (b-a)U_1 \\ V_2 = MU_2 \end{cases}$$

Si $V_2 < f(V_1)$ on pose $X = V_1$, sinon on recommence depuis le début : on génère deux variables uniformes...

a) Calculer $P(V_2 < f(V_1))$.

b) Montrer que X est une v.a. de densité f . Cette méthode s'appelle la *simulation par rejet* et évite de calculer la fonction réciproque de la fonction de répartition.

CHAPITRE 5

NOTIONS PROBABILISTES UTILISÉES EN THÉORIE DU SIGNAL

5.1. PROCESSUS ALÉATOIRES

5.1.1. Définitions

Les bruits et plus généralement les signaux « incertains » seront représentés par une famille de variables aléatoires, indiquée par le temps : à temps continu, $X_t(\omega)$ et à temps discret $X_n(\omega)$. C'est le cas lorsqu'on effectue un électro-encéphalogramme ; le tracé obtenu est très agité localement et on le modélise par une famille de v.a. X_t ; Si on échantillonne à 100 Hz (100 mesures par seconde), on obtient une famille X_n de v.a., c'est-à-dire à temps discret.

Les variables aléatoires seront à valeurs dans \mathbb{R} , \mathbb{C} ou \mathbb{R}^n . L'application

$$t \mapsto X_t(\omega),$$

s'appelle une *trajectoire*, alors que

$$\omega \mapsto X_t(\omega)$$

décrit l'état à l'instant fixé t . X_t représentera, par exemple, la tension complexe aux bornes d'une résistance.

Le terme « fonction aléatoire » est explicite néanmoins il est de plus en plus remplacé par le terme plus général de « processus aléatoire » où la famille de variables aléatoires n'est plus nécessairement indiquée par le temps mais par une variable d'espace... Lorsque le temps est discret, on parle de série chronologique.

5.1.2. Exemples

5.1.2.1 Le bruit de grenaille

Dans un tube à vide, les électrons sont émis avec un taux α par unité de temps ; le courant au temps t est noté X_t . Il dépend des électrons émis via la formule

$$X_t = \sum_{\substack{\text{électrons} \\ \text{émis entre} \\ 0 \text{ et } t}} g(t - t_e)$$

où g est une fonction qui caractérise le tube et le circuit, et t_e est l'instant d'émission de l'électron e . Dans un modèle déterministe on supposerait que les électrons sont émis de façon régulière au taux α . On montre qu'ils sont émis de façon aléatoire suivant un processus de Poisson : le nombre d'électrons émis durant un intervalle de temps suit une loi de Poisson de paramètre :

$\alpha \cdot$ longueur de l'interv.

D'autre part, si on considère des intervalles de temps disjoints, il y a indépendance.

X_t est donc aléatoire. Si on appelle N_t le nombre d'électrons émis entre 0 et t , N_t saute de +1 à chaque fois qu'un électron est émis et reste constant entre deux instants d'émission. X_t s'exprime à l'aide de dN_t

$$X_t = \int_0^t g(t-u) dN_u$$

N_u a pour espérance αu , ce qui donne,

$$X_t = \alpha \int_0^t g(t-u) du + \int_0^t g(t-u) d(N_u - \alpha u)$$

Le premier terme est déterministe et exprime la tendance ; le deuxième terme est aléatoire, centré et exprime les fluctuations autour de la tendance.

5.1.2.2. Le mouvement brownien

Il s'agit d'un mouvement aléatoire sans sauts (trajectoires continues) mais à comportement extrêmement irrégulier ; il doit son nom au botaniste G. Brown qui observait le pollen en suspension dans l'eau (1827).

Définition :

1) $X_t, t \geq 0$ est un processus brownien normalisé si $X_0 = 0$, à chaque instant t , X_t est de loi $\mathcal{N}(0, t)$ et pour $t > s$, $X_t - X_s$ est indépendant du passé jusqu'à l'instant s .

2) $X_t, t \geq 0$ est un processus brownien de dérive m et de paramètre de diffusion σ^2 si $X_0 = 0$, à chaque instant t , X_t est de loi $\mathcal{N}(mt, \sigma^2 t)$ et pour $t > s$, $X_t - X_s$ est indépendant du passé jusqu'à l'instant s .

On vérifie facilement que $X_t, t \geq 0$ est un processus brownien de dérive m et de paramètre de diffusion σ^2 si et seulement si Y_t

$$Y_t = \frac{1}{\sigma} (X_t - mt)$$

est un brownien normalisé.

Lorsque $m = 0$ et $\sigma^2 = 1$, on dit simplement « mouvement brownien ».

On peut modifier les trajectoires par un phénomène de réflexion sans modifier le caractère brownien :

$$t \mapsto X_{t_0} - [X_t - X_{t_0}]$$

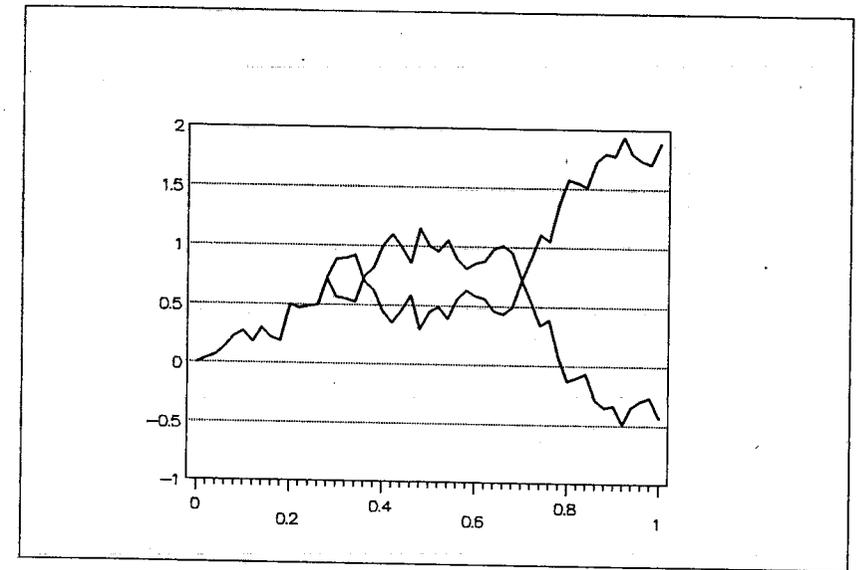


Fig. 5.1. — Modification de trajectoires par réflexion.

Einstein publia en 1909 l'étude du comportement moléculaire d'une particule soumise à un bombardement moléculaire, en supposant que la quantité de mouvement évoluait continûment et suivant la définition 1 ci-dessus.

C'est Wiener, en 1922, qui donna le modèle rigoureux du mouvement et qui montra que *presque toutes les trajectoires* étaient *continues*, ainsi que l'existence d'un tel processus en prenant une suite Y_n de v.a.r. indépendantes, de loi $\mathcal{N}(0, 1)$ et en posant

$$X_t = \frac{t}{\sqrt{\pi}} Y_0 + \sqrt{\frac{2}{\pi}} \sum_{n=1}^{+\infty} \frac{\sin nt}{n} Y_n$$

(voir exercice 4).

Depuis, le mouvement brownien n'a cessé d'intéresser de nombreux mathématiciens (Mandelbrot par exemple) et physiciens (physique des polymères par exemple). En ce qui concerne les trajectoires, citons les résultats suivants :

1) Loi du logarithme itéré (Khintchine en 1924)

$$\lim_{\substack{t \rightarrow 0 \\ t > 0}} \sup \frac{X_t}{\sqrt{2t \ln(\ln 1/t)}} = 1 \text{ p.s.}$$

2) Presque toutes les trajectoires ne sont dérivables en aucun point.

3) Presque toutes les trajectoires ne sont à variation bornée sur aucun intervalle de temps.

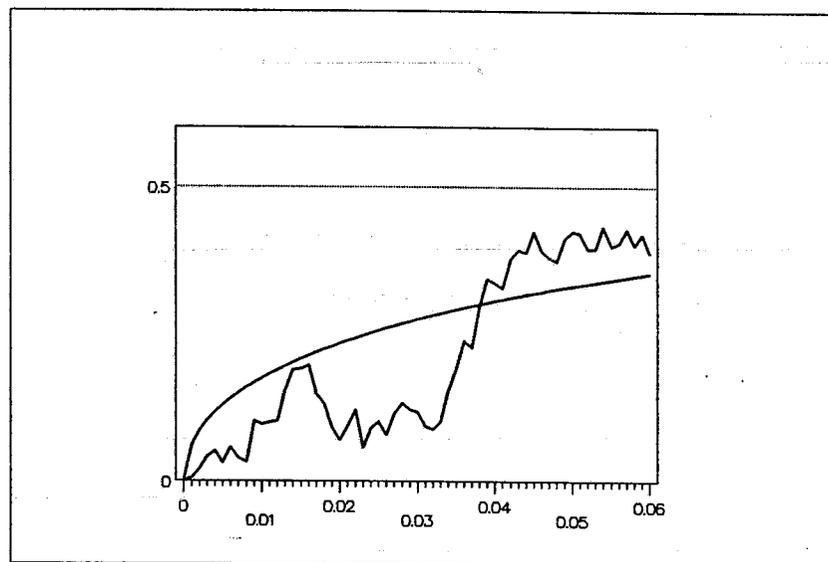


Fig. 5.2. — « Loi du log-itéré ».

4) Presque toutes les trajectoires sont des courbes fractales ayant pour dimension de Hausdorff 1,5 (voir « Dimensions non entières et applications » Cherbit, Ed. Masson).

5.1.3. Processus du second ordre

Il s'agit des processus pour lesquels, à tout instant t , l'état est décrit par une variable ayant un moment d'ordre 2 fini.

Moyenne

La moyenne donne la tendance : c'est la fonction $t \mapsto E(X_t)$.

Fonction de covariance et fonction d'autocorrélation

Fonction de covariance : C'est la fonction de 2 variables t et s (temps discret ou continu) :

$$t, s \mapsto K(t, s) = \text{cov}(X_t, X_s)$$

Dans le cas complexe, on prendra le conjugué de X_s .

La covariance renseigne sur les interactions entre les états à deux instants différents.

Fonction d'autocorrélation du signal : On peut définir de la même manière la fonction $\Gamma(t, s) = E(X_t \bar{X}_s)$.

Les deux fonctions K et Γ sont reliées par

$$K(t, s) = \Gamma(t, s) - E(X_t) E(\bar{X}_s)$$

et on a le résultat suivant :

Proposition : K et Γ sont des fonctions de type positif, c'est-à-dire qu'elles vérifient :

$$\forall n, \forall c_1, \dots, c_n \in \mathbb{C}, \forall t_1, \dots, t_n \quad \sum_{k, \ell} c_k \bar{c}_\ell K(t_k, t_\ell) \geq 0$$

5.1.4. Processus du second ordre stationnaires

Le processus (X_t) est dit stationnaire (on dit aussi *stationnaire fort*) si il a même loi que le processus où on a fait une translation du temps, (Y_t) , $Y_t = X_{t-a}$, et ceci pour tout a de l'espace des temps.

Le processus sera dit *faiblement stationnaire* ou stationnaire au second ordre, si la fonction moyenne et la fonction de covariance sont stationnaire, c'est-à-dire si la moyenne $E(X_t)$ est constante et si $K(t, s)$ ne dépend que de la différence $t - s$ ou, ce qui est équivalent, si $E(X_t)$ est constante et si $\Gamma(t, s)$ ne dépend que de la différence $t - s$. On vérifie immédiatement qu'un processus du second ordre fortement stationnaire est faiblement stationnaire.

Dans le cas où le processus est stationnaire au second ordre, on peut écrire :

$$E(X_t) = m$$

$$K(t, s) = k(t - s) = E(X_t \bar{X}_s) - |m|^2$$

$$\Gamma(t, s) = \gamma(t - s) = E(X_t \bar{X}_s)$$

$$k(\tau) = \gamma(\tau) - |m|^2$$

lorsque la fonction γ est continue (temps continu) on dit que le processus est *continu en moyenne quadratique*.

5.1.5. Processus gaussien

Soit X_t un processus à valeurs réelles. (X_t) est un processus gaussien si, quel que soit n et quels que soient les instants t_1, t_2, \dots, t_n , le vecteur $(X_{t_1}, \dots, X_{t_n})$ est gaussien. Un processus gaussien est un processus du second ordre. La loi ne dépend que de la fonction moyenne et de la fonction de covariance. Si il est stationnaire faiblement, il est donc automatiquement stationnaire fortement.

5.2. BRUIT BLANC

5.2.1. A temps discret

A temps discret un bruit blanc, $B_n, n \in \mathbb{Z}$, est la donnée d'une suite de variables aléatoires indépendantes, centrées et de même variance ; on verra la justification de ce terme « bruit blanc » au paragraphe suivant et comment il peut être étendu.

On aura alors,

$$\text{var} \left(\sum_n a_n B_n \right) = \sum_n \sigma^2 |a_n|^2$$

L'effet du bruit blanc se traduit sur la variance par un apport de $|a_n|^2 \sigma^2$ à chaque instant n .

5.2.2. A temps continu

A temps continu, si on adapte cette formulation, au lieu de considérer une somme, on considère une intégrale en remplaçant les coefficients a_n par une fonction : on regarde l'effet sur une fonction $\phi(t)$; un intervalle $t, t + dt$ aura un apport de

$$\sigma^2 \phi(t) dt$$

sur la variance, la variance totale étant

$$\sigma^2 \int \phi(t)^2 dt$$

Nous allons voir la construction rigoureuse sur \mathbb{R}_+ à partir du processus de Wiener (appelé aussi mouvement Brownien).

Le lecteur peu attiré par ces considérations mathématiques peut se reporter au tableau récapitulatif en fin de paragraphe.

$W_t, t \geq 0$ est une fonction aléatoire qui a les propriétés suivantes :

- 1) A ω fixé, W est une fonction continue de t et $W_0 = 0$.
- 2) W_t est de loi $\mathcal{N}(0, t)$.
- 3) Pour $t > s$, $W_t - W_s$ est indépendant du passé jusqu'à l'instant s .

On en déduit que des accroissements successifs $W_{t_{i+1}} - W_{t_i}$ sont indépendants, centrés et de variances $t_{i+1} - t_i$.

W étant une fonction continue de la variable t peut être considérée comme une distribution. On regarde l'effet sur une fonction test ϕ , c'est-à-dire une fonction indéfiniment dérivable à support compact dans \mathbb{R}_+ :

$$\phi \mapsto \int_{\mathbb{R}_+} \phi(t) W_t dt$$

La dérivée de W_t va nous fournir un bruit blanc. Malheureusement, les trajectoires de W_t sont trop agitées pour que l'on puisse dériver au sens ordinaire des fonctions ; nous allons dériver au sens des distributions, opération qui est toujours possible (voir « Distributions » de Roddier, Ediscience). Par définition, il s'agit de l'application :

$$\phi \mapsto - \langle W_t, \phi'(t) \rangle \stackrel{\text{dét.}}{=} - \int \phi'(t) W_t dt$$

On peut écrire $\langle W_t, \phi'(t) \rangle$ comme limite des sommes de Riemann,

$$\sum_i \phi'(t_i) W_{t_i} (t_{i+1} - t_i) \quad (1)$$

ϕ est suffisamment régulière pour que dans la recherche de cette limite on puisse remplacer $\phi'(t_i) (t_{i+1} - t_i)$ par $\phi(t_{i+1}) - \phi(t_i)$.

On peut alors réarranger les termes de la somme (1), en tenant compte du fait que $\phi(0) = 0$ et $\phi(t) = 0$ pour $t >$ un certain T , qui devient :

$$= \sum_i \phi(t_i) (W_{t_{i+1}} - W_{t_i}) \quad (2)$$

On remarque que (2) est la somme de variables gaussiennes indépendantes et en utilisant les propriétés de W_t , cette somme est une variable gaussienne centrée, de variance

$$\sum_i |\phi(t_i)|^2 (t_{i+1} - t_i)$$

En utilisant les résultats du chapitre 3, (2) converge à ω fixé et aussi en moyenne quadratique, lorsque le pas de la subdivision tend vers 0, vers une variable aléatoire gaussienne centrée, de variance

$$\int_{\mathbb{R}_+} |\phi(t)|^2 dt$$

En faisant la même construction sur \mathbb{R}_- mais de façon symétrique, et en faisant la somme des deux intégrales, on obtient une intégrale stochastique sur \mathbb{R} . On appelle bruit blanc de niveau 1 la distribution aléatoire ainsi obtenue et on la note :

$$\phi \mapsto \int_{\mathbb{R}} \phi(t) dW_t$$

On dit que l'intégrale est stochastique (c'est une variable aléatoire).

On obtient un bruit blanc de niveau $B > 0$ en prenant :

$$\phi \mapsto B \int_{\mathbb{R}} \phi(t) dW_t$$

Si on intègre sur un intervalle $[a, b]$, on a la formule d'intégration par parties

$$\int_a^b \phi(t) dW_t = \phi(a) W_a - \phi(b) W_b - \int_a^b \phi'(t) W_t dt$$

D'autre part on peut étendre l'écriture de l'intégrale stochastique à une fonction f de carré intégrable en l'approchant avec des fonctions ϕ_n indéfiniment dérivables et à support compact ; l'intégrale (qui est une variable aléatoire)

$$\int_{\mathbb{R}} f(t) dW_t$$

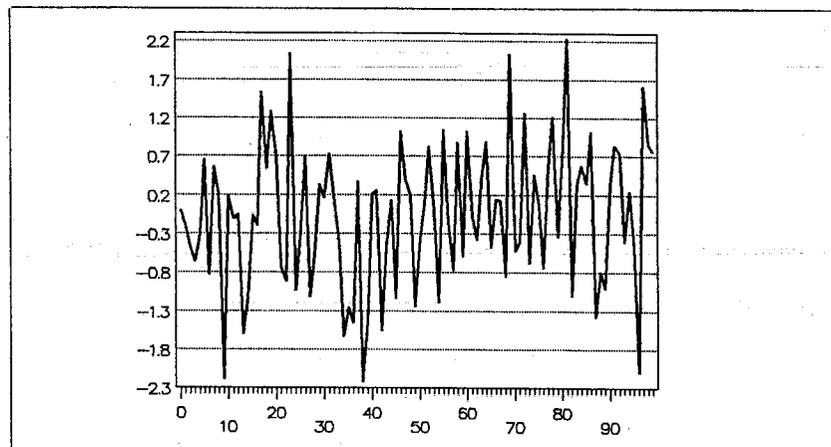


Fig. 5.3. — Bruit blanc.

est la limite en moyenne quadratique des intégrales

$$\int_{\mathbb{R}} \phi_n(t) dW_t$$

Si on considère deux fonctions f et g de carré intégrable, leur covariance est égale à l'intégrale

$$\int_{\mathbb{R}} f(t) \overline{g(t)} dt$$

5.2.3. Tableau récapitulatif

	Variable	Espérance	Variance	Covariance
Temps discret	$\sum a_n B_n$ avec $\sum a_n ^2 < +\infty$	0	$\sum a_n ^2$	$\sum a_n \overline{b_n}$
	$\sum b_n B_n$ avec ...			
Temps continu	$\int_{\mathbb{R}} f(t) dW_t$ avec $\int f(t) ^2 dt < +\infty$	0	$\int_{\mathbb{R}} f(t) ^2 dt$	$\int_{\mathbb{R}} f(t) \overline{g(t)} dt$
	$\int_{\mathbb{R}} g(t) dW_t$ avec ...			

5.3. MESURE SPECTRALE DE PUISSANCE POUR UN PROCESSUS STATIONNAIRE AU SECOND ORDRE

5.3.1. Fonction de covariance et fonction d'autocorrélation

Le processus est supposé stationnaire faiblement. On décompose le processus suivant « une tendance » déterministe et un processus centré :

$$X_t = m + (X_t - m) \text{ avec } m = E(X_t)$$

$$\text{cov}(X_t - m, X_s - m) = E[(X_t - m)(\overline{X_s - m})] = k(t - s)$$

$$= \gamma(t - s) - |m|^2$$

Le processus centré $X_t - m$ a pour fonction d'autocorrélation, la fonction $k(t)$.

Dans le cas où $m = 0$, les deux fonctions k et γ sont égales.

5.3.2. Théorème de Bochner-Herglotz

a) A temps discret : la fonction $\gamma(n)$ est la transformée de Fourier inverse d'une mesure positive bornée sur $[0, 1]$, appelée mesure spectrale de puissance :

$$\gamma(n) = \int_{[0,1]} e^{2i\pi n\lambda} d\mu(\lambda)$$

Si cette mesure admet une densité, on l'appelle *densité spectrale* :

$$\gamma(n) = \int_0^1 e^{2i\pi n\lambda} f(\lambda) d\lambda$$

avec $f(\lambda) \geq 0$.

b) A temps continu, on a le même résultat qu'à temps discret si on suppose de plus γ continue, γ est la transformée d'une mesure positive bornée sur \mathbb{R} :

$$\gamma(t) = \int_{\mathbb{R}} e^{2i\pi t\lambda} d\mu(\lambda)$$

On peut trouver la démonstration de ce théorème dans le livre « Probability theory » de Loeve.

L'existence d'une densité spectrale de puissance est directement liée à la décroissance à l'infini ; plaçons-nous dans le cas centré et à temps discret, les deux fonctions k et γ coïncident et il suffit d'avoir

$$\sum_n |\gamma(n)|^2 < +\infty$$

pour que la mesure spectrale de puissance soit égale à une mesure à densité ; plus précisément :

$$\sum_n |\gamma(n)| < +\infty \Rightarrow \sum_n |\gamma(n)|^2 < +\infty \Rightarrow f(\lambda) = \sum_n e^{-2i\pi n\lambda} \gamma(n)$$

Dans le cas d'un bruit blanc, $k(0) = \sigma^2$ et $k(n) = 0$ si $n \neq 0$, de sorte que la densité spectrale est égale à la constante σ^2 . Toutes les fréquences jouent le même rôle, c'est ce qui explique le terme de « bruit blanc ». Par extension, on a l'habitude d'appeler bruit blanc tout processus B_n , du second ordre, centré ayant une densité spectrale de puissance constante, c'est-à-dire tel que

$$\text{cov}(B_n, B_m) = 0 \quad \text{si } n \neq m \quad \text{et} \quad \sigma^2 \quad \text{si } n = m$$

On a de même à temps continu

$$\int_{\mathbb{R}} |\gamma(t)| dt < +\infty \Rightarrow f(\lambda) = \int_{\mathbb{R}} e^{-2i\pi t \lambda} \gamma(t) dt$$

$$\int_{\mathbb{R}} |\gamma(t)|^2 dt < +\infty \Rightarrow f(\lambda) = \int_{\mathbb{R}} e^{-2i\pi t \lambda} \gamma(t) dt$$

5.3.3. Interprétation

On fera le rapprochement avec le cas déterministe où pour un signal réel de puissance finie, on forme la fonction de corrélation du signal

$$\text{f.c.}(\tau) = \lim_{T \rightarrow +\infty} \frac{1}{2T} \int_{-T}^T s(t) s(t - \tau) dt$$

Cette fonction permet de décrire la répartition de la puissance suivant les différentes fréquences, en regardant sa transformée de Fourier. Pour un processus aléatoire réel stationnaire et ergodique (voir paragraphe 4), l'intégrale (aléatoire)

$$\frac{1}{2T} \int_{-T}^T X_t X_{t-\tau} dt$$

qui est une moyenne temporelle, tend vers la moyenne au sens probabiliste, c'est-à-dire $k(\tau)$. On peut d'ailleurs remarquer que toute mesure expérimentale est une moyenne temporelle car elle ne peut être faite de façon instantanée.

5.3.4. Transformée de Fourier (théorème de représentation spectrale)

A ω fixé on peut considérer la trajectoire

$$t \mapsto X_t(\omega)$$

que le temps soit discret ou continu, et chercher à calculer : la série de Fourier dans le cas discret, la transformée de Fourier dans le cas continu. Ce calcul pose des problèmes comme le montre l'exemple d'un bruit blanc à temps discret : supposons les variables B_n indépendantes de même loi $\mathcal{N}(0, 1)$; la série de Fourier associée :

$$\sum_n B_n e^{2i\pi n \lambda}$$

diverge pour presque toute épreuve ω ; en effet l'ensemble des épreuves ω pour lesquelles la série converge, est un ensemble de probabilité 0 ou 1. Si il était de probabilité 1, d'après le théorème des trois séries, on aboutirait à une contradiction.

Par contre, si X_t est une somme de sinusoides ayant des fréquences constantes et

$$X_t(\omega) = \sum_{n=1}^N A_n(\omega) e^{2i\pi t \lambda_n}$$

avec

$$E(A_n) = 0 \quad E(A_n \bar{A}_m) = 0 \quad \text{si } n \neq m$$

alors on peut représenter X_t :

$$X_t = \int_{\mathbb{R}} e^{2i\pi t \lambda} dZ(\lambda)$$

où $dZ(\lambda)$ est une « mesure aléatoire »

$$dZ(\lambda) = \sum_{n=1}^N A_n \delta_{\lambda_k}(\lambda)$$

$$\gamma(\tau) = \sum_{n=1}^N E(|A_n|^2) e^{2i\pi \tau \lambda_n} = \sum_{n=1}^N \sigma_n^2 e^{2i\pi \tau \lambda_n}$$

$$= \int_{\mathbb{R}} e^{2i\pi \lambda \tau} d\mu(\lambda)$$

$$d\mu(\lambda) = \sum_{n=1}^N \sigma_n^2 \delta_{\lambda_k}(\lambda)$$

Dans le cas général, on a une représentation spectrale abstraite :

Théorème : Soit X_t stationnaire au second ordre, centré, de fonction $\gamma(\tau)$ continue si le temps est continu, de mesure spectrale de puissance $\mu(\lambda)$. Alors il existe un processus aléatoire Z_λ , centré, à accroissements orthogonaux c'est-à-dire

$$E[(Z_b - Z_a)(\bar{Z}_d - \bar{Z}_c)] = 0 \quad \text{si } a < b \leq c < d$$

et tel que

$$\text{var}(Z_b - Z_a) = \mu([a, b])$$

relation que l'on a l'habitude d'écrire

$$E(|dZ_\lambda|^2) = d\mu(\lambda)$$

et on a la représentation spectrale

$$X_t = \int_{\mathbb{R}} e^{2i\pi\lambda t} dZ_\lambda \quad \text{si le temps est continu}$$

$$X_n = \int_{[0,1[} e^{2i\pi\lambda n} dZ_\lambda \quad \text{si le temps est discret}$$

Démonstration : A chaque v.a. X_t on associe, dans le plan spectral la fonction

$$\lambda \mapsto e^{2i\pi\lambda t}$$

Cette application conserve le produit scalaire :

$$\begin{aligned} E(X_t \bar{X}_s) &= \int e^{2i\pi\lambda t} e^{-2i\pi\lambda s} d\mu(\lambda) \\ &= \langle e^{2i\pi\lambda t}, e^{2i\pi\lambda s} \rangle_{L^2(\mu)} \end{aligned}$$

On la prolonge par linéarité et continuité au sous-espace vectoriel fermé, H , de $L^2(\Omega, a, P)$ engendré par toutes les v.a. X_t . L'image de H est un sous-espace vectoriel fermé de $L^2(\mu)$ qui contient toutes les fonctions $e^{2i\pi\lambda t}$, donc toutes les combinaisons de ces fonctions et les limites... On a ainsi, dans l'image, l'application

$$\begin{aligned} 1_{]-\infty, a]}(\lambda) &\quad \text{si le temps est continu} \\ 1_{[0, a]}(\lambda) &\quad \text{si le temps est discret} \end{aligned}$$

Soit Z_a l'image réciproque ; les propriétés annoncées pour Z_a sont faciles à vérifier. Considérons maintenant

$$Y_t = \int e^{2i\pi\lambda t} dZ_\lambda$$

qui est, par définition, la limite en moyenne quadratique des sommes de Riemann

$$\sum e^{2i\pi\lambda_n t} (Z_{\lambda_{n+1}} - Z_{\lambda_n})$$

Y_t a donc pour image

$$\lim \sum_n e^{2i\pi\lambda_n t} 1_{] \lambda_n, \lambda_{n+1}]}(\lambda) = e^{2i\pi\lambda t}$$

Y_t et X_t ont même image et l'application est bijective, donc $Y_t = X_t$.

Exemple : Soit $(B_n)_{n \in \mathbb{Z}}$ un bruit blanc gaussien avec $\sigma = 1$.

On peut calculer Z_a en utilisant

$$\begin{aligned} B_n &\mapsto (\lambda \mapsto e^{2i\pi n\lambda}) \\ 1_{[0, a]}(\lambda) &= a + \sum_{n \neq 0} \frac{1 - e^{-2i\pi n a}}{2i\pi n} e^{2i\pi n\lambda} \\ Z_a &= aB_0 + \sum_{n \neq 0} \frac{1 - e^{-2i\pi n a}}{2i\pi n} \cdot B_n \\ &= aB_0 + \sum_{n=1}^{+\infty} \left[\frac{1 - e^{-2i\pi n a}}{2i\pi n} B_n + \frac{1 - e^{2i\pi n a}}{-2i\pi n} B_{-n} \right] \\ 2 \operatorname{Re}(Z_{a/2}) &= aB_0 + \sum_{n=1}^{+\infty} C_n \end{aligned}$$

avec C_n qui est une variable aléatoire réelle, gaussienne centrée, de variance $\frac{2|1 - e^{-2i\pi n a}|^2}{4\pi^2 n^2} = \frac{2 \sin^2 \pi n a}{\pi^2 n^2}$.

On écrit alors $C_n = \sqrt{2} \frac{\sin \pi n a}{\pi n} \cdot D_n$ avec D_n de loi $\mathcal{N}(0, 1)$ et on a :

$$2 \operatorname{Re}(Z_{a/2}) = aB_0 + \frac{\sqrt{2}}{\pi} \sum_{n=1}^{+\infty} \frac{\sin n\pi a}{n} D_n$$

c'est-à-dire que Z_a est un mouvement brownien sur $[0, 1]$. $\operatorname{Re}(Z_a)$ et $\operatorname{Im}(Z_a)$ sont indépendants (voir E.5.9) et sont des browniens. La variable

$$B_n = \int_{[0, 1]} e^{2i\pi n\lambda} dZ_\lambda$$

est obtenue en intégrant $e^{2i\pi n\lambda}$ par rapport à un bruit blanc complexe, isotrope, dans l'espace des fréquences : cela veut dire que si l'on considère un bruit blanc sur $[0, 1]$ et que l'on cherche à calculer ses coefficients de Fourier, alors ces coefficients sont des v.a. B_n indépendantes, centrées, c'est-à-dire un bruit blanc complexe, isotrope à temps discret.

5.4. L'ERGODISME

5.4.1. Historique

La notion de processus ergodique est au cœur de l'étude des systèmes dynamiques et remonte à Gibbs (1839-1903), précurseur de la physique statistique : suivant le raisonnement de Gibbs, le temps d'observation du phénomène physique peut être considéré comme infini devant l'échelle de temps au niveau moléculaire. Le système passe alors par tous les états possibles en restant plus longtemps, ou en passant plus souvent, dans les états qui sont plus probables, de telle sorte que la moyenne temporelle devient égale à la moyenne sur les états, c'est-à-dire l'espérance mathématique.

C'est Birkhoff (1884-1944) qui formalisa ces notions et démontra les premiers grands théorèmes sur la question dans les années 30. Ce fut alors l'ouverture sur un vaste champ d'investigation.

5.4.2. Les résultats

Les notions d'invariance et d'ergodisme sont définies pour des processus aléatoires non nécessairement stationnaires mais s'énoncent de façon particulière dans le cas stationnaire. Voici des résultats intéressants pour les applications en théorie du signal :

Définitions : Etant donné un processus stationnaire..., $X_{-n}, \dots, X_1, X_2, \dots, X_n, \dots$, un événement est dit invariant si il est invariant par toute translation du temps. Le processus est dit ergodique si tous les événements invariants sont de probabilité 0 ou 1.

Exemple fondamentale : Soit..., X_1, X_2, \dots une suite de v.a. indépendantes et de même loi, alors..., X_1, X_2, \dots est un processus stationnaire ergodique.

On admettra la proposition suivante :

Proposition : Soit X_1, X_2, \dots un processus stationnaire ; alors il existe un processus..., $Y_{-n}, \dots, Y_1, Y_2, \dots, Y_n, \dots$ tel que, pour tout n , positif (Y_1, Y_2, \dots, Y_n) et (X_1, X_2, \dots, X_n) ont même loi, et le processus Y_n est stationnaire.

Théorème : Soit X_1, X_2, \dots un processus stationnaire ergodique. Soit φ une application telle que $E|\varphi(X_1)| < \infty$, alors :

1)

$$\frac{1}{n} [\varphi(X_1) + \dots + \varphi(X_n)] \xrightarrow{n \rightarrow +\infty} E[\varphi(X_1)] \quad (\text{convergence presque sûre})$$

2) De même, si $f(X_1, \dots, X_k)$ admet une espérance alors

$$\frac{1}{n} [f(X_1, \dots, X_k) + f(X_2, \dots, X_{k+1}) + \dots + f(X_{n+1} + \dots + X_{n+k})] \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} E[f(X_1, X_2, \dots, X_k)]$$

Grâce à ce résultat, on peut estimer la densité spectrale de puissance à l'aide du périodogramme (voir exercice E.5.10).

Evidemment, à la lecture de ce qui précède, beaucoup de questions se posent :

- Sur la physique statistique : dans quel sens les états sont-ils aléatoires ?
- Que représente un événement invariant ?
- Quels sont les phénomènes qui sont ergodiques ?

La première et la troisième question représentent des tonnes de connaissances et... de nouvelles questions, mais on peut donner des réponses parcellaires :

En physique statistique, on dit que les états du système sont aléatoires, mais il ne s'agit pas de n'importe quel aléatoire. Le modèle probabiliste est choisi de façon à prendre en compte les lois physiques : types d'interaction, portée de l'interaction... Avec ces modèles, lorsqu'on est dans un état proche d'un état d'équilibre, on peut très bien passer dans un autre état que l'état d'équilibre et même, quitter le bassin d'attraction de cet état d'équilibre. Néanmoins de telles transitions d'états sont extrêmement peu probables bien que possibles. En ce qui concerne les états invariants, pour un processus stationnaire, un événement invariant est un événement terminal, c'est-à-dire qu'il appartient à tous les futurs :

σ -alg. engendrée par X_n, X_{n+1}, \dots

On a de nombreux exemples de phénomènes ergodiques. On peut dire, par exemple, que les systèmes considérés en théorie cinétique des gaz sont ergodiques, qu'un bruit blanc est ergodique.

Un résultat important pour les applications au signal est le suivant :

Théorème : Soit..., $Y_{-n}, \dots, Y_1, Y_2, \dots, Y_n, \dots$ un processus gaussien stationnaire tel que

$$\text{cov}(Y_n, Y_m) \xrightarrow{|n-m| \rightarrow +\infty} 0$$

alors ce processus est ergodique.

Pour les démonstrations on pourra se reporter à « Probability » de Breiman, à Halmos : Lectures on ergodic theory, à « Calcul des probabilités » de Neveu, « Processus aléatoires » de Rozanov.

EXERCICES

E.5.1. Un compteur enregistre le nombre de particules émises au cours du temps. On suppose que le nombre de particules N_t , émises entre 0 et t , obéit aux hypothèses d'un processus de Poisson et que lorsqu'une particule est émise, elle a la probabilité p d'être enregistrée par le compteur. Soit M_t le nombre de particules enregistrées au temps t ; montrer que M_t est encore un processus de Poisson ; que peut-on dire sur le paramètre de ce processus ?

E.5.2. Soit $(W_t)_{t \geq 0}$ un mouvement brownien.

Soit $a > 0$ et $Y_t = \frac{W_{at}}{\sqrt{a}}$. Montrer que Y_t est encore un mouvement brownien.

E.5.3. Soit $(X_t)_{t \geq 0}$ et $(Y_t)_{t \geq 0}$ deux mouvements browniens indépendants. Montrer que

$$Z_t = \begin{cases} X_t & \text{si } t \in [0, 1] \\ X_1 + Y_{t-1} & \text{si } t \in [1, +\infty[\end{cases}$$

est encore un mouvement brownien.

E.5.4. Soit Y_0, Y_1, \dots une suite de v.a. indépendantes, de loi $\mathcal{N}(0, 1)$. En utilisant les résultats du chapitre 3, montrer que la série

$$\frac{t}{\sqrt{\pi}} Y_0 + \sqrt{\frac{2}{\pi}} \sum_{n=1}^{+\infty} \frac{\sin nt}{n} Y_n$$

converge p.s. et en moyenne quadratique. Soit X_t cette limite et $f(t)$ la fonction périodique, de période 2π , paire, et égale à $t - \frac{t^2}{\pi}$ sur $[0, \pi[$. Montrer que l'on a :

$$f(t) = \frac{\pi}{6} - \frac{1}{\pi} \sum_{n=1}^{+\infty} \frac{\cos 2nt}{n^2}$$

En déduire que X_t suit la loi $\mathcal{N}(0, t)$ si t appartient à $[0, \pi[$.

E.5.5. Soient A et B deux v.a. indépendantes de loi $\mathcal{N}(0, 1)$. On pose

$$\begin{aligned} X_t &= A \cos \omega t \\ Y_t &= B \sin \omega t \\ Z_t &= A \cos \omega t + B \sin \omega t \end{aligned}$$

Calculer la fonction moyenne et la fonction de covariance pour chacun des processus. Sont-ils stationnaires ?

E.5.6. Soit A une v.a. de loi $\mathcal{N}(0, 1)$ et θ de loi uniforme sur $[0, 2\pi]$ et indépendante de A. On pose

$$X_t = A \cos(\omega t + \theta)$$

Calculer la fonction moyenne et la fonction de covariance. Le processus est-il stationnaire ?

E.5.7. Calculer la fonction moyenne et la fonction d'autocorrélation d'un processus brownien X_t en utilisant $X_t = (X_t - X_s) + X_s$. On pose

$$\begin{aligned} Y_t &= 0 & \text{si } t = 0 \\ Y_t &= t X_t^1 & \text{si } t > 0 \end{aligned}$$

Calculer $E(Y_t^2)$ et $\text{cov}(Y_t - Y_s, Y_t)$.

En déduire que Y_t est encore brownien.

E.5.8. Soit X_n une chaîne de Markov à espace d'états fini. Soit μ une probabilité invariante. Montrer que si X_0 suit la loi μ , alors la suite X_n est stationnaire.

E.5.9. Soit X_n un processus réel, stationnaire, centré et gaussien. Soit Z_λ le processus à accroissements donné par le théorème de représentation spectrale 5.3.3.

a) Montrer que Z_λ est un processus gaussien, $Z_0 = 0$ et que les accroissements sont indépendants.

b) On pose $Z_\lambda = Z_{1,\lambda} + iZ_{2,\lambda}$ avec Z_1 et Z_2 réels, et $Z(I) = \int_I dZ_\lambda$ que l'on prolonge par périodicité en dehors de $[0, 1]$.

Si I et J sont des intervalles disjoints, calculer $E[Z(I)\overline{Z(J)}]$ et $E[Z(-I)\overline{Z(J)}]$. En déduire

$$E[Z_1(I)Z_2(J)] = 0 = E[Z_1(I)Z_1(J)] = E[Z_2(I)Z_2(J)]$$

Montrer que $E[Z_1(I)Z_2(J)] = 0$ même si les intervalles I et J ne sont pas disjoints.

E.5.10. Soit $(X_n)_{n \in \mathbb{Z}}$ un signal aléatoire centré, stationnaire au second ordre, de densité spectrale de puissance $f(\lambda)$. On suppose

$$\sum_n |\text{cov}(X_n, X_0)| < +\infty$$

a) Montrer que $f(\lambda)$ est continue.

b) On appelle périodogramme à l'instant n , la fonction aléatoire

$$I_n(\lambda) = \frac{1}{n} |X_1 e^{-2i\pi\lambda} + \dots + X_n e^{-2i\pi n\lambda}|^2$$

Calculer $\int_0^1 I_n(\lambda) e^{2i\pi n_0 \lambda} d\lambda$ pour $0 \leq n_0 \leq n-1$. Montrer que si X_n est de plus ergodique, on a

$$\lim_{n \rightarrow +\infty} \int_0^1 I_n(\lambda) e^{2i\pi n_0 \lambda} d\lambda \underset{\text{p.s.}}{=} \text{cov}(X_{n_0}, X_0)$$

CHAPITRE 6

LE FILTRAGE LINÉAIRE

6.1. INTRODUCTION

On parle de filtrage dès qu'on a une opération qui transforme le signal, opération que l'on peut symboliser par une opération entrée-sortie, sur un schéma, en appelant $x(t)$ le signal avant modification et $y(t)$ après modification :

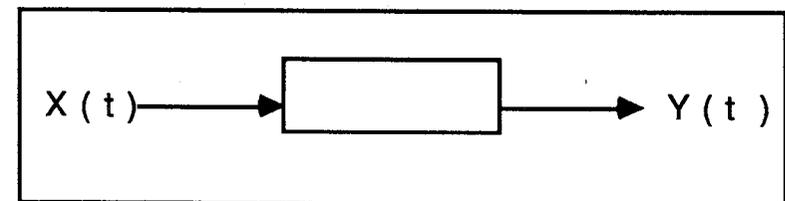


Figure 6.1.

Le temps peut être aussi bien discret que continu.

Voici quelques exemples que l'on peut rencontrer dans l'étude des circuits ou lorsqu'on fait des mesures physiques.

6.1.1. Filtre RC

A $x(t)$ on associe $y(t)$ solution de

$$\begin{cases} y'(t) + \alpha y(t) = x(t) & t > 0 \\ y(0) = 0 \end{cases}$$

soit
$$y(t) = \int_0^t x(u) e^{-\alpha(t-u)} du = x * h(t)$$

en particulier pour le filtre RC

$$h(t) = \frac{e^{-t/RC}}{RC} \cdot u(t) \quad \text{où } u(t) \text{ est la fonction de Heaviside}$$

6.1.2. Dérivation

$$x(t) \mapsto y(t) = x'(t)$$

6.1.3. Lissage

$$\begin{aligned} x(t) \mapsto y(t) &= \frac{1}{2a} \int_{-a}^a x(t+u) du \\ &= x * h(t) \end{aligned}$$

avec

$$h(t) = \frac{1}{2a} \cdot 1_{[-a, a]}(t)$$

6.1.4. Modulations

Dans l'étude des communications, on rencontre l'opération de

— Modulation d'amplitude :

$$\begin{aligned} x(t) \mapsto y(t) &= ax(t) \cos(2\pi\nu_0 t + \phi_0) \\ \text{ou } a[1 + bx(t)] \cos(2\pi\nu_0 t) \end{aligned}$$

— Modulation de phase :

$$x(t) \mapsto y(t) = a \cos(2\pi\nu_0 t + bx(t))$$

— Modulation de fréquences :

$$x(t) \mapsto y(t) = a \cos\left(2\pi\nu_0 t + \int_0^t bx(u) du\right)$$

6.1.5. A temps discret

A une suite x_n on associe une suite y_n .

Par exemple,

$$x_n \mapsto y_n = 2x_n - 3x_{n-1}$$

6.1.6. Dans les communications numériques

$$\text{suite } x_n \mapsto \text{signal analogique } y(t) = \sum_n x_n h(t - nT)$$

6.1.7. Le théorème des filtres

Parmi les exemples cités ci-dessus, deux ne sont pas des opérations linéaires : il s'agit de la modulation de phase et la modulation de fréquence. Dans ce chapitre nous nous intéresserons uniquement aux opérations qui sont :

- linéaires
- stationnaires dans le temps
- continues

Il faudrait évidemment préciser ce que l'on entend par continu : il s'agit d'une notion très large qui permet d'inclure tous les cas classiques rencontrés en physique (voir « Distributions et Signal » de Roubine). On a alors le théorème dit théorème des filtres : tout système linéaire, stationnaire, continu est une opération de convolution par une distribution. Si l'on prend l'exemple de la dérivation,

$$x(t) \mapsto y(t) = x'(t) = (x * \delta')(t)$$

où

$$\langle \delta', \phi \rangle \stackrel{\text{déf.}}{=} -\phi'(0)$$

(Pour ces définitions, on pourra consulter le livre très pédagogique de Roddier : « Distributions » Edisciences.)

Le signal déterministe est généralement bruité de façon additive :

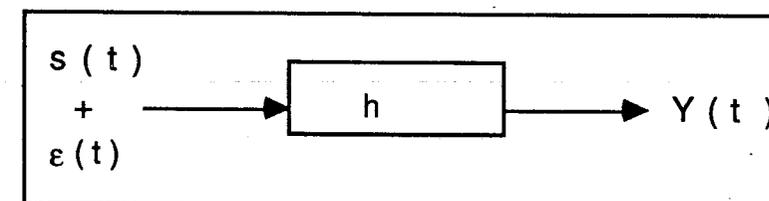


Figure 6.2.

Le bruit passe dans le filtre et modifie le signal de sortie. On va donc étudier les caractéristiques d'un processus aléatoire après passage dans un filtre linéaire.

6.2. FILTRAGE LINÉAIRE À TEMPS CONTINU

6.2.1. Intégrales stochastiques

Lorsqu'on met un bruit à l'entrée d'un filtre du type 6.1.1, on récupère en sortie l'intégrale d'une fonction aléatoire ; nous avons vu au chapitre précédent le sens que l'on pouvait donner à cette intégrale dans le cas d'un bruit blanc ; étudions le cas général d'une fonction aléatoire du second ordre.

Soit X_t une fonction aléatoire centrée, de fonction de covariance $K(s, t)$. Si f est une fonction, on veut donner un sens à l'intégrale

$$\int_{\mathbb{R}} f(t) X_t dt$$

Si les trajectoires $t \mapsto X_t$ sont suffisamment régulières ainsi que la fonction f , par exemple continues, on peut écrire cette intégrale pour chaque épreuve ω . Si on

ne sait rien des trajectoires, on peut donner un sens à cette intégrale en disant que c'est la limite en moyenne quadratique des sommes de Riemann

$$\sum_i f(t_i) X_{t_i} (t_{i+1} - t_i)$$

En effet, d'après le lemme de Loève vu au chapitre 3, les sommes de Riemann convergent en m.q. lorsque le pas de la subdivision tend vers 0 si et seulement si la série

$$\sum_{i,j} f(t_i) \overline{f(t_j)} k(t_i, t_j) (t_{i+1} - t_i) (t_{j+1} - t_j)$$

tend vers une limite finie lorsque le pas de la subdivision tend vers 0 ; d'où le critère pour que l'on puisse définir l'intégrale.

Critère d'intégrabilité en moyenne quadratique.

L'intégrale $\int f(t) X_t dt$ existe en moyenne quadratique si et seulement si

$$\iint f(t) K(t, s) \overline{f(s)} dt ds < +\infty$$

Dans ce cas, $\int f(t) X_t dt$ est une variable centrée, de variance égale à l'intégrale double ci-dessus.

En ce qui concerne les covariances :

$$\text{cov} \left(\int f(t) X_t dt, \int g(t) X_t dt \right) = \iint f(t) K(t, s) \overline{g(s)} dt ds$$

Dans le cas stationnaire,

$$\text{cov} \left(\int f(t) X_t dt, \int g(t) X_t dt \right) = \iint f(t) k(t-s) \overline{g(s)} dt ds$$

$$\text{cov} \left(\int f(t) X_t dt, \int g(t) X_t dt \right) = \int f(t) (k * \overline{g})(t) dt$$

Dans le cas du bruit blanc, la covariance est $\int f(t) \overline{g(t)} dt$, ce qui veut dire que l'on a remplacé la fonction k par la distribution de Dirac δ . La distribution de Dirac δ a pour transformée de Fourier la fonction constante et égale à 1.

On prendra très souvent l'approximation suivante du bruit blanc.

La densité spectrale est égale à la fonction rectangle sur $[-F, F]$, c'est-à-dire

$$\text{cov}(X_t, X_s) = k(t-s) = \frac{\sin(2\pi F(t-s))}{\pi(t-s)}$$

D'autre part, on parlera d'un bruit blanc de niveau B lorsque toutes les variances et covariances ci-dessus sont multipliées par B .

Tableau récapitulatif

Fonction aléatoire stationnaire	Bruit blanc
$\text{cov}(X_t, X_s) = k(t-s)$ Mesure spectrale $d\mu$	k est remplacé par δ Mesure spectrale $d\lambda$
$k(t) = \int e^{2i\pi\lambda t} d\mu(\lambda)$	Densité spectrale égale à 1

6.2.2. Filtre de convolution par la fonction h avec un bruit blanc en entrée

Soit dW_t le bruit blanc, que l'on appellera ε_t pour simplifier. Pour être rigoureux on devrait calculer $h * \varepsilon$ en tant que distribution. Nous ferons le calcul comme s'il s'agissait d'une fonction en remplaçant au moment voulu les intégrales par des intégrales au sens de Riemann-Stieljes. Pour une fonction f on écrirait

$$f * h(x) = \int f(x-u) h(u) du = \int h(x-u) f(u) du$$

Pour le bruit blanc, on écrira

$$\varepsilon * h(x) = \int_{\mathbb{R}} h(x-u) dW_u$$

D'après ce qui précède, $\varepsilon * h(x)$ est une variable gaussienne, centrée de variance égale à

$$\int_{\mathbb{R}} |h(x-u)|^2 du = \int_{\mathbb{R}} |h(u)|^2 du$$

De même les formules de covariance nous donnent :

$$\begin{aligned} \text{cov}(\varepsilon * h(t), \varepsilon * h(s)) &= \int h(t-u) \overline{h(s-u)} du \\ &= \int h(v) \overline{h(v+s-t)} dv \\ &= h * \overline{h}(t-s) \end{aligned}$$

en posant $\check{h}(u) = h(-u)$. Le signal de sortie est encore stationnaire.

Ainsi pour un filtre RC, avec la condition $y(-\infty) = 0$ au lieu de $y(0) = 0$

$$h(t) = \frac{1}{RC} e^{-t/RC} u(t)$$

$$\text{cov}(\varepsilon * h(t), \varepsilon * h(s)) = \frac{1}{2RC} e^{-\frac{|t-s|}{RC}}$$

La sortie a pour densité spectrale

$$\frac{1}{1 + 4\pi^2 R^2 C^2 \lambda^2}$$

Les fréquences élevées sont atténuées, de sorte que l'on retire les variations rapides, et on obtient à la sortie un signal dont les composantes sont corrélées. La corrélation entre deux composantes Y_t et Y_s décroît exponentiellement avec la distance $t - s$.

Application : Bruit thermique.

Une résistance R , à la température T , est source d'un bruit lié à l'agitation des électrons, que l'on appelle bruit thermique. On montre que si on considère le circuit suivant, sans générateur,

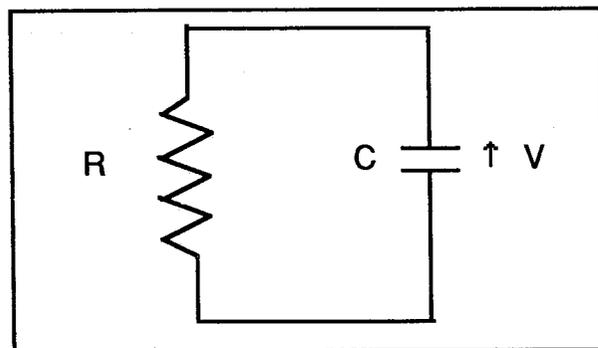


Figure 6.3.

V est en moyenne nul mais V^2 n'est pas nul en moyenne :

$$\frac{1}{2} kT = \frac{1}{2} C \overline{V^2} \text{ (valeur moyenne)}$$

On explique le bruit en disant que l'on a une source aléatoire X_t qui est un bruit blanc de niveau σ^2 . Il s'agit alors d'un circuit RC ayant X_t en entrée et V en sortie. D'après ce qui précède, à la sortie on a une variance égale à

$$E(V^2) = \frac{\sigma^2}{2RC}$$

D'où

$$\frac{kT}{C} = \frac{\sigma^2}{2RC}$$

$$\sigma^2 = 2kRT$$

Une résistance R est la source d'un bruit blanc de niveau $2kRT$, c'est-à-dire proportionnel à la température.

La densité spectrale est constante et égale à $2kRT$. Elle apporte donc $4kRT \Delta\lambda$ à une bande de fréquences de largeur $\Delta\lambda$, si on regroupe les fréquences symétriques. Cela permet de définir une température équivalente même si le bruit n'est pas d'origine thermique. En effet on regarde la puissance utilisable en connectant une résistance R' idéale (ne produisant aucun bruit) :

$$I = V/(R + R')$$

$$R' I^2 = 4 RR' kT \Delta\lambda / (R + R')^2$$

La puissance maximale est obtenue pour $R' = R$ et vaut

$$N = kT \Delta\lambda$$

Pour un bruit quelconque, on définit la température équivalente dans la bande de fréquences $\lambda, \lambda + \Delta\lambda$ par

$$T = \frac{1}{k} \frac{dN}{d\lambda}$$

La comparaison des températures est alors une méthode de comparaison de deux bruits.

6.2.3. Filtre de convolution par la fonction h avec en entrée une fonction aléatoire

Soit ε_t un bruit de fonction de covariance $K(t, s)$, non nécessairement stationnaire. On supposera ε_t centré.

$(\varepsilon * h)$, a pour fonction de covariance

$$\Gamma(t, s) = \text{cov} \left(\int h(t-u) \varepsilon_u du, \int h(s-u) \varepsilon_u du \right)$$

$$= \iint h(t-u) K(u, v) \overline{h(s-v)} du dv$$

Dans le cas stationnaire,

$$K(t, s) = k(t-s)$$

$$\Gamma(t, s) = \int h(t-u) k * \tilde{h}(u-s) du$$

$$= (k * h * \tilde{h})(t-s)$$

$$= \gamma(t-s)$$

On retrouve bien une formule analogue à celle du bruit blanc, pour lequel k est remplacé par δ .

Le filtrage linéaire conserve la stationnarité

6.2.4. Dérivation d'une fonction aléatoire

ε_t est toujours un bruit centré de fonction de covariance $K(s, t)$. La convolution par δ' a pour effet de dériver au sens des distributions lorsque la dérivation au sens classique n'est pas possible ; on est alors obligé de regarder l'effet sur une fonction test ϕ :

$$\langle \delta' * \varepsilon, \phi \rangle = - \langle \varepsilon, \phi' \rangle = - \int \varepsilon_t \phi'(t) dt$$

$$\text{cov} (\langle \delta' * \varepsilon, \phi \rangle, \langle \delta' * \varepsilon, \psi \rangle) = \iint \phi'(t) k(t, s) \overline{\psi'(s)} dt ds$$

Lorsque K est continûment différentiable, on peut intégrer par parties et on obtient,

$$\iint \frac{\partial^2}{\partial t \partial s} K(t, s) \phi(t) \overline{\psi(s)} dt ds$$

c'est-à-dire

$$\Gamma(t, s) = \frac{\partial^2}{\partial t \partial s} K(t, s)$$

et dans le cas stationnaire,

$$\Gamma(t, s) = \gamma(t - s) = -k''(t - s)$$

Dans le cas général il s'agit de dérivée au sens des distributions.

6.3. ANALYSE SPECTRALE (cas stationnaire)

6.3.1. Formules dans le domaine spectral

L'analyse spectrale permet d'exprimer simplement les relations entre l'entrée et la sortie, dans le cas stationnaire, grâce aux transformées de Fourier. On appelle H la transformée de Fourier du filtre (fonction de transfert).

$$\begin{aligned} K(t, s) &= k(t - s) \quad \text{et} \quad \Gamma(t, s) = \gamma(t - s) \\ k(u) &= \int e^{2i\pi u \lambda} d\mu(\lambda) \\ &= \int e^{2i\pi u \lambda} f(\lambda) d\lambda \quad \text{si } k \text{ admet une densité spectrale } f \end{aligned}$$

On sait que l'on a la relation

$$\gamma = k * h * \tilde{h}$$

entre l'entrée et la sortie. Cette relation s'écrit avec les transformées de Fourier,

$$\hat{\gamma} = \hat{k} H \bar{H} = |H|^2 \hat{k}$$

Dans le cas d'une densité spectrale f à l'entrée, on obtient à la sortie une densité spectrale g :

$$g(\lambda) = f(\lambda) |H(\lambda)|^2$$

On pourra calculer les covariances à la sortie en prenant la transformée inverse de g :

$$\gamma(t) = \int_{\mathbb{R}} e^{2i\pi t \lambda} f(\lambda) |H(\lambda)|^2 d\lambda$$

En particulier,

$$\begin{aligned} \text{var}(Y_t) &= \int_{\mathbb{R}} f(\lambda) |H(\lambda)|^2 d\lambda \\ \text{cov}(Y_t, Y_0) &= \int_{\mathbb{R}} e^{2i\pi \lambda t} f(\lambda) |H(\lambda)|^2 d\lambda \end{aligned}$$

6.3.2. Contre-exemple : L'équation de Langevin

On considère l'équation avec condition initiale

$$\begin{cases} y'(t) + \alpha y(t) = X_t, t > 0 \\ y(0) = 0 \end{cases}$$

où X_t est un bruit blanc.

On sait que pour une équation de ce type, la sortie y est le résultat de la convolution de l'entrée par la fonction h ,

$$h(t) = e^{-\alpha t} u(t)$$

qui a pour transformée de Fourier la fonction de transfert H ,

$$H(\lambda) = \frac{1}{2i\pi\lambda + \alpha}$$

Malheureusement, on ne peut pas écrire la formule entre les densités spectrales car l'entrée n'est pas stationnaire (l'espace de temps est \mathbb{R}_+ au lieu de \mathbb{R}). On prend donc la formule du cas non stationnaire.

A la sortie :

$$\begin{aligned} \Gamma(s, t) &= \iint h(s-u) h(t-v) d\mu(u, v) \\ &\quad \uparrow \\ &\quad \text{mesure de Lebesgue sur la diagonale} \\ &\quad \delta(u-v) \\ &= \int h(s-u) h(t-u) du \\ &= e^{-\alpha(s+t)} (e^{2 \inf(s,t)} - 1) \quad \text{pour } s \text{ et } t > 0 \\ &\quad \left\{ \begin{array}{l} E(Y_t) = 0 \\ E(Y_t^2) = \Gamma(t, t) = \frac{1 - e^{-2\alpha t}}{2\alpha} \end{array} \right. \end{aligned}$$

6.4. CORRÉLATION ENTRE L'ENTRÉE ET LA SORTIE

6.4.1. Cas général

$$E(Y_t) = E(X_t) * h \quad \text{et} \quad X_t = E(X_t) + (X_t - E X_t)$$

Pour calculer $\text{cov}(X_t, Y_s)$, on peut supposer l'entrée centrée puisque le filtre est linéaire. Dans ce cas $\text{cov}(X_t, Y_s) = E(X_t \bar{Y}_s)$.

$$\begin{aligned} E(X_t \bar{Y}_s) &= E \left\{ X_t \int \overline{h(s-u)} \bar{X}_u du \right\} \\ &= \int \overline{h(s-u)} E(X_t \bar{X}_u) du \\ &= \int K(t, u) \overline{h(s-u)} du \\ &= K(t, \cdot) * \overline{h(s)} \\ &\quad \uparrow \uparrow \text{la convolution a lieu sur la 2}^\circ \text{ variable de K} \end{aligned}$$

$$\text{De même, } E(Y_t \bar{X}_s) = K(\cdot, s) * h(t) \quad \uparrow \uparrow \text{convolution sur la 1}^\circ \text{ variable de K}$$

6.4.2. Cas stationnaire

Dans le cas d'une entrée *stationnaire*, on obtient une relation stationnaire entre l'entrée et la sortie :

$$E(X_t \bar{Y}_s) = (k * \check{h})(t-s)$$

Cette formule se généralise aux distributions. En particulier lorsque l'entrée est un bruit blanc, k est remplacé par la distribution de Dirac. La sortie est

stationnaire de densité spectrale $|H(\lambda)|^2$ et

$$E(Y_t \bar{X}_s) = h(t-s)$$

Dans le cas où le système est inconnu, cette formule donne une méthode d'estimation de h . On met un bruit blanc à l'entrée et on estime statistiquement $\text{cov}(Y_t, X_0)$, ce qui donne une estimation de $h(t)$.

6.4.3. Exemple : Le filtre en quadrature

Il s'agit d'un filtre linéaire qui à tout signal du type $\cos \omega t$ associe le sinus de même fréquence. Lorsqu'on écrit la convolution au sens des distributions,

$$\sin \omega t = h * \cos \omega t$$

on constate que le gain complexe, $H(\lambda)$, est nécessairement égal à

$$-i \text{sign}(\lambda)$$

On montre dans le cours sur les distributions, que cette fonction de transfert est la transformée de Fourier de la distribution

$$\frac{1}{\pi} \text{pf} \left(\frac{1}{t} \right) \quad (\text{pseudo-fonction})$$

c'est-à-dire,

$$(h * \phi)(t) = \lim_{\eta \rightarrow 0} \int_{|t-u| > \eta} \frac{\phi(u)}{t-u} du$$

La relation entre le signal d'entrée $x(t)$ et le signal de sortie $y(t)$ peut s'écrire dans le domaine spectral :

$$\hat{y}(\lambda) = \hat{x}(\lambda) H(\lambda) = -i \text{sign}(\lambda) \hat{x}(\lambda)$$

$$\text{d'où} \quad \hat{x}(\lambda) = -\hat{y}(\lambda) H(\lambda)$$

Le filtre inverse est donc égal à l'opposé du filtre en quadrature. En particulier, l'image de $\sin \omega t$ est égale à $-\cos \omega t$ et celle de $\exp(i\omega t)$ à $\exp\{i(\omega t - \pi/2)\}$ d'où le nom de filtre en quadrature.

Le filtre analytique associé à tout signal du type $\cos \omega t$, la fonction $\exp(i\omega t)$; Il s'agit de la somme du filtre identité et du filtre en quadrature multiplié par i . Le filtre analytique permet de représenter le signal uniquement avec des fréquences positives :

$$\begin{aligned} x(t) &\rightarrow z(t) = x(t) + iy(t) \\ \hat{x}(\lambda) &\rightarrow \hat{x}(\lambda) + i\hat{x}(\lambda) H(\lambda) \\ z(\lambda) &= 2\hat{x}(\lambda) \quad \text{si } \lambda > 0 \\ &= 0 \quad \text{sinon} \end{aligned}$$

Cette représentation est particulièrement intéressante pour un signal réel quasimonochromatique, c'est-à-dire si $\tilde{x}(\lambda)$ est nul en dehors du domaine de fréquences $|\lambda - \lambda_0| \leq \varepsilon$. On forme le signal analytique z et on pose,

$$z(t) = a(t) \exp(2i\pi\lambda_0 t + i\phi(t))$$

de sorte que,

$$\begin{aligned} a(t) &= |z(t)| \\ \phi(t) &= \arg(z(t) \overline{\exp(-2i\pi\lambda_0 t)}) \end{aligned}$$

$a(t) \exp(i\phi(t))$ n'a aucune fréquence en dehors de l'intervalle $[-\varepsilon, \varepsilon]$, ce qui veut dire qu'il s'agit d'un signal à variations très lentes.

Supposons maintenant que l'entrée soit aléatoire, centrée et stationnaire,

$$\begin{aligned} \gamma &= k * h * \check{h} \\ \hat{\gamma} &= \hat{k} |H|^2 = \hat{k} \end{aligned}$$

Cela veut dire que \check{Y}_t a les mêmes moments d'ordre 2 que X_t . D'autre part, en utilisant le filtrage de X_t ou de Y_t ,

$$\begin{aligned} \text{cov}(X_t, Y_s) &= k * \check{h}(t-s) = -k * h(t-s) \\ \text{cov}(Y_t, X_s) &= k * h(t-s) \end{aligned}$$

la fonction $\text{cov}(Y_t, X_s)$ est donc le résultat du filtrage de la fonction k par le filtre en quadrature : on a la même relation que pour les signaux eux-mêmes.

6.5. FORMULE DES INTERFÉRENCES

L'entrée X_1 passe dans le filtre h_1 , tandis que X_2 passe dans le filtre h_2 .

La formule des interférences donne les relations de covariances entre les sorties Y_1, Y_2 provenant d'entrées corrélées X_1, X_2 .

On note la covariance entre deux processus par :

$$R_{X_1 X_2}(t, s) = \text{cov}(X_{1,t}, X_{2,s})$$

En supposant les deux entrées centrées ce qui ne change pas le calcul des covariances, on a

$$\begin{aligned} R_{Y_1 Y_2}(t, s) &= E\{Y_{1,t} \bar{Y}_{2,s}\} \\ &= E\{(X_1 * h)_t (\bar{X}_2 * \check{h})_s\} \\ &= E\left\{\int \int X_1(t-u) h_1(u) \overline{X_2(s-v)} \overline{h_2(v)} du dv\right\} \\ &= R_{X_1 X_2}(t-u, s-v) h_1(u) \overline{h_2(v)} du dv \end{aligned}$$

$$R_{Y_1 Y_2} = R_{X_1 X_2} * \check{h}_2 * h_1$$

↑ convolution sur la 1^{re} variable
↑ convolution sur la 2^e variable

Dans le cas où $R_{X_1 X_2}$ est stationnaire, il en est de même pour les sorties,

$$\begin{aligned} R_{X_1 X_2}(t, s) &= r_{X_1 X_2}(t-s) \\ R_{Y_1 Y_2}(t, s) &= r_{Y_1 Y_2}(t-s) \\ &= r_{X_1 X_2} * h_1 * \check{h}_2(t-s) \end{aligned}$$

$$r_{Y_1 Y_2}(\tau) = r_{X_1 X_2} * h_1 * \check{h}_2(\tau)$$

En utilisant les transformées de Fourier, on peut écrire

$$\hat{r}_{Y_1 Y_2} = \hat{r}_{X_1 X_2} H_1 \bar{H}_2$$

Toutes les relations des paragraphes précédents sont des cas particuliers de la formule des interférences. On retrouve la formule de corrélation entrée-sortie, par exemple, en faisant

$$\begin{aligned} X_2 &= X_1 \\ h_2 &= \delta \quad \text{c'est-à-dire} \quad H_2(\lambda) = 1 \quad \forall \lambda \end{aligned}$$

6.6. RÉSUMÉ POUR LE CAS STATIONNAIRE

La sortie est aussi stationnaire et on a :

	Entrée X_t	Sortie Y_t
Moyenne au temps t	m_X	$m_Y = h * m_X = H(0)m_X$
Covariance entre les instants $t+s$ et s	$k(t)$	$\gamma(t) = k * h * \check{h}(t)$
Densité spectrale	$f(\lambda)$	$g(\lambda) = f(\lambda) H(\lambda) ^2$
Loi	Processus gaussien	Processus gaussien
Covariance entrée-sortie	$\text{cov}(X_t, Y_s) = k * \check{h}(t-s)$	

6.7. FILTRAGE LINÉAIRE À TEMPS DISCRET

6.7.1. Adaptation des formules précédentes

L'entrée est toujours notée X_n et la sortie Y_n . Nous supposons X_n centré

$$K(n, m) = \text{cov}(X_n, X_m)$$

$$k(n) = \text{cov}(X_n, X_0) \text{ dans le cas stationnaire}$$

Si $k(n)$ a une bonne décroissance vers 0 lorsque n tend vers $\pm \infty$, X_n a pour densité spectrale de puissance $f(\lambda)$:

$$f(\lambda) = \sum_{n=-\infty}^{+\infty} k(n) e^{-2i\pi\lambda n}$$

Toutes les formules précédentes s'adaptent au cas discret en remplaçant les intégrales sur la variable de temps t par des sommes sur \mathbb{Z} et les intégrales sur les fréquences λ par des intégrales sur des fréquences λ variant entre 0 et 1 ou entre $-\frac{1}{2}$ et $\frac{1}{2}$; en particulier, si X_n a pour densité spectrale de puissance $f(\lambda)$, la sortie Y_n a pour densité spectrale de puissance $g(\lambda)$:

$$g(\lambda) = f(\lambda) |H(\lambda)|^2$$

et on a

$$\text{var}(Y_n) = \int_0^1 f(\lambda) |H(\lambda)|^2 d\lambda$$

6.7.2. Les modèles fondamentaux

Dans ce paragraphe, on suppose que l'entrée est un bruit blanc, c'est-à-dire que les variables X_n sont centrées et non corrélées (extension de la notion de bruit blanc vue au § 5.3.2).

Moyennes mobiles d'ordre q : MA(q)

$$Y_n = X_n - b_1 X_{n-1} - \dots - b_q X_{n-q}$$

Auto-régressif d'ordre p : AR(p)

$$Y_n = a_1 Y_{n-1} + a_2 Y_{n-2} + \dots + a_p Y_{n-p} + X_n$$

ARMA(p, q)

$$Y_n + a_1 Y_{n-1} + \dots + a_p Y_{n-p} = b_0 X_n + b_1 X_{n-1} + \dots + b_q X_{n-q}$$

Dans un modèle MA(q), le filtre qui fait passer de X à Y est explicite :

$$Y = h * X$$

$$\begin{cases} h(0) = 1 & h(1) = -b_1 \dots & h(q) = -b_q \\ \text{et} \\ h(k) = 0 & \text{sinon} \end{cases}$$

Au contraire dans un modèle AR(p) ou plus généralement ARMA(p, q), la sortie est donnée sous forme récursive. L'exemple le plus simple de modèle récursif est celui où cherche à estimer l'espérance d'une variable aléatoire en se basant sur la loi des grands nombres c'est-à-dire en prenant la moyenne empirique

$$\frac{X_1 + X_2 + \dots + X_n}{n}$$

Si prend cette relation comme une relation entrée-sortie, l'entrée étant la suite X_1, X_2, \dots , la sortie Y_n

$$Y_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

peut s'écrire sous forme récursive avec des coefficients variables

$$Y_{n+1} = \frac{n}{n+1} Y_n + \frac{1}{n+1} X_{n+1}$$

Le modèle AR(p) est l'équivalent à temps discret d'une équation différentielle d'ordre n , linéaire et à coefficients constants. Il s'écrit

$$(Y * A)_n = X_n$$

$$\text{avec} \quad \begin{cases} A(0) = 1 & A(1) = -a_1 \dots & A(p) = -a_p \\ A(k) = 0 & \text{sinon} \end{cases}$$

6.7.3. Le modèle AR(p)

6.7.3.1. Recherche de filtres

Cherchons les filtres h tels que $Y = h * X$ et leurs propriétés

$$Y = h * X \Rightarrow A * Y = A * (h * X)$$

$$= (A * h) * X$$

↑
sous certaines conditions

L'associativité est vérifiée notamment si les 3 éléments ont leur support borné du même côté : par exemple si h est nul jusqu'à un certain n et X est nul jusqu'à un certain n (A est de support fini).

Dans ce cas on cherche h de support borné à gauche et tel que

$$A * h = \delta$$

On utilise les transformées en z :

$$\left(\sum_n h_n z^{-n} \right) \left(\sum_n A(n) z^{-n} \right) = 1$$

$$\sum_n h_n z^{-n} = \frac{1}{1 - a_1 z^{-1} - \dots - a_p z^{-p}}$$

Pour le modèle le plus simple qui est le modèle AR(1):

$$y_n = a y_{n-1} + x_n$$

$$\frac{1}{1 - a z^{-1}} = 1 + a z^{-1} + a^2 z^{-2} + \dots + a^k z^{-k} + \dots$$

sur $|z| > |a|$

$$\text{et } \frac{1}{1 - a z^{-1}} = \frac{z}{z - a} = \frac{-z/a}{1 - z/a} = \frac{-z}{a} \left\{ 1 + \frac{z}{a} + \frac{z^2}{a^2} + \dots \right\}$$

sur $|z| < |a|$

Dans les deux cas on trouve une suite de coefficients h_n t.q.

$$h * A = \delta$$

La première suite donne une solution causale,

$$\begin{aligned} h(n) &= 0 & \text{si } n < 0 \\ h(n) &= a^n & \text{si } n \geq 0 \end{aligned}$$

et c'est celle-ci que nous choisirons, un filtre non causal n'étant pas réalisable. On a alors

$$y_n = x_n + a x_{n-1} + a^2 x_{n-2} + \dots + a^k x_{n-k} + \dots$$

Il s'agit d'une moyenne mobile d'ordre infini MA(∞).

En ce qui concerne les propriétés du filtre, on rappelle qu'un filtre est dit stable si il associe à toute entrée bornée une sortie bornée. La solution trouvée représente un filtre stable, si et seulement si a est de module strictement inférieur à 1; en effet, en choisissant

$$\begin{aligned} x(n) &= e^{n i \phi} \text{ où } \phi \text{ est l'argument du nombre complexe } a \\ y(0) &= \sum |a|^n \end{aligned}$$

Pour un modèle d'ordre supérieur, la transformée en z s'écrit

$$\begin{aligned} \sum_n h_n z^{-n} &= \frac{1}{1 - a_1 z^{-1} - \dots - a_p z^{-p}} = \frac{z^p}{z^p - a_1 z^{p-1} - \dots - a_p} \\ &= \frac{z^p}{(z - \alpha_1)(z - \alpha_2) \dots (z - \alpha_p)} = \frac{z^p}{P(z)} \\ &= \frac{1}{(1 - \alpha_1 z^{-1}) \dots (1 - \alpha_p z^{-p})} \end{aligned}$$

que l'on développe en éléments simples.

On obtient ainsi un filtre causal, qui est stable si toutes les racines α_i sont de module strictement inférieur à 1. Ce filtre est associé à un modèle MA(∞) (moyenne mobile d'ordre infini).

6.7.3.2. Le cas d'un bruit blanc en entrée

Supposons que l'on ait un filtre h réel et stable, et que l'entrée soit un bruit blanc de variance 1. En posant $e^{2i\pi\lambda} = z$ on a:

$$\gamma(n) = \text{cov}(Y_n, Y_0)$$

$$\hat{\gamma}(\lambda) = \sum_n \gamma(n) e^{-2i\pi n \lambda}$$

$$\hat{\gamma}(\lambda) = |H(\lambda)|^2$$

$$H(\lambda) = \sum_n h_n e^{-2i\pi n \lambda} = \frac{1}{1 - a_1 e^{-2i\pi \lambda} - \dots - a_p e^{2i\pi p \lambda}}$$

$$= \frac{z^p}{P(z)}$$

$$|H(\lambda)|^2 = H(\lambda) \overline{H(\lambda)}$$

$$|H(\lambda)|^2 = \frac{1}{P(z) P(z^{-1})}$$

car

$$\bar{z} = z^{-1} \text{ lorsque } z = e^{2i\pi\lambda}$$

Regardons le cas particulier d'un modèle AR(1); la densité spectrale et les covariances s'obtiennent à partir de la fraction

$$\begin{aligned} \frac{1}{(z - \alpha)(z^{-1} - \alpha)} &= \frac{z}{(z - \alpha)(1 - \alpha z)} = \\ &= \frac{A}{(z - \alpha)} + \frac{B}{(1 - \alpha z)} = \frac{A z^{-1}}{1 - \alpha z^{-1}} + \frac{B}{1 - \alpha z} \end{aligned}$$

On développe le premier terme en série de z^{-1} et le deuxième en série de z . La somme des deux converge sur une couronne contenant strictement le cercle unité

car la racine α est supposée de module strictement inférieur à 1

$$\hat{\gamma}(\lambda) = B \sum_{k \geq 0} \alpha^k e^{2i\pi k \lambda} + A \sum_{k \geq 1} \alpha^{k-1} e^{-2i\pi k \lambda}$$

$$\Rightarrow \begin{cases} \gamma(k) = A\alpha^{k-1} & \text{si } k \geq 1 \\ \gamma(k) = B\alpha^{|k|} & \text{si } k \leq 0 \end{cases}$$

D'autre part, dans le modèle d'ordre p , on a des relations intéressantes sur les $\gamma(n)$ qui sont les suivantes

$$\begin{cases} Y_{n+1} = a_1 Y_n + a_2 Y_{n-1} + \dots + a_p Y_{n+1-p} + X_{n+1} \\ E(Y_{n+1} Y_n) = a_1 E(Y_n^2) + a_2 E(Y_{n-1} Y_n) + \dots + \\ \quad + a_p E(Y_{n+1-p} Y_n) + E(Y_n X_{n+1}) \\ \vdots \\ E(Y_{n+1} Y_{n+1-p}) = a_1 E(Y_{n+1-p} Y_n) + \dots + \\ \quad + a_p E(Y_{n+1-p}^2) + E(Y_{n+1-p} X_{n+1}) \end{cases}$$

Y est obtenu à partir de X par un filtre causal tel que $h(0) = 1$, ce qui veut dire que chaque Y_n est la somme de X_n et d'un élément du passé, donc indépendant de X_n , le système s'écrit donc

$$\begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \gamma(p-2) \\ \vdots & \vdots & \vdots \\ \gamma(p-1) & \dots & \gamma(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = M\vec{a}$$

Cette formule permet d'estimer \vec{a} lorsqu'il est inconnu, à partir d'estimations de $\gamma(0), \gamma(1), \dots, \gamma(p)$. Pour terminer, donnons le niveau de bruit en sortie :

$$\begin{aligned} \text{var}(Y_{n+1}) &= E[Y_{n+1}(a_1 Y_n + \dots + a_p Y_{n+1-p} + X_{n+1})] \\ &= a_1 \gamma(1) + a_2 \gamma(2) + \dots + a_p \gamma(p) + E(Y_{n+1} X_{n+1}) \\ &= a_1 \gamma(1) + a_2 \gamma(2) + \dots + a_p \gamma(p) + E(X_{n+1}^2) \\ &= (a_1, \dots, a_p) M \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} + \sigma^2 \end{aligned}$$

6.7.4. Le modèle ARMA(p, q)

Les modèles ARMA sont intimement liés aux processus à spectre rationnel comme nous allons le montrer.

Définition : Un processus Y_n stationnaire au second ordre, centré est dit à spectre rationnel si il admet une densité spectrale de puissance $f(\lambda)$ qui est une fraction rationnelle de la variable $e^{-2i\pi\lambda}$.

Un processus ARMA, Y_n centré, c'est-à-dire de la forme

$$Y_n + a_1 Y_{n-1} + \dots + a_p Y_{n-p} = b_0 X_n + b_1 X_{n-1} + \dots + b_q X_{n-q}$$

où X_n est un bruit blanc, est à spectre rationnel puisque la fonction de transfert (transformée en z du filtre) est une fraction rationnelle de z^{-1} . La réciproque est vraie comme le montre le théorème ci-dessous.

Théorème : Le processus Y_n stationnaire au second ordre et centré, est un processus ARMA si et seulement si il est à spectre rationnel.

Démonstration : On suppose que Y_n est à spectre rationnel ; soit $f(\lambda)$ sa densité spectrale ;

$$f(\lambda) = \frac{P(e^{-2i\pi\lambda})}{Q(e^{-2i\pi\lambda})}$$

avec

$$\begin{cases} P(z) = a_0 + a_1 z + \dots + a_q z^q \\ Q(z) = b_0 + b_1 z + \dots + b_p z^p \\ P \text{ et } Q \text{ premiers entre eux et } a_q b_p \neq 0 \end{cases}$$

On a

$$P(z^{-1}) = z^{-q} P_1(z) \quad \text{avec} \quad P_1(z) = a_q + a_{q-1} z + \dots + a_0 z^q$$

On écrit de même,

$$Q(z^{-1}) = z^{-p} Q_1(z)$$

La densité spectrale est réelle et positive. En posant $e^{2i\pi\lambda} = z$

$$f(\lambda) = \overline{f(\lambda)} \Rightarrow \frac{z-q}{z-p} \cdot \frac{P_1(z)}{Q_1(z)} = \frac{\overline{P}(z)}{\overline{Q}(z)} \quad \text{où} \quad \overline{P}(z) = \overline{a_0} + \overline{a_1} z + \dots + \overline{a_q} z^q$$

On a donc

$$z^q P_1(z) \overline{Q}(z) = z^p \overline{P}(z) Q_1(z) \quad \text{si } z = e^{2i\pi\lambda}$$

Cette égalité entre polynômes est donc nécessairement vraie pour tout Z de \mathbb{C} . Soit z_0 une racine non nulle de P_1 . P et Q étant premiers entre eux, z_0 annule le polynôme \overline{P} :

$$\overline{P}(z_0) = z_0^q P_1 \left(\frac{1}{z_0} \right)$$

Donc $\frac{1}{z_0}$ est racine de P_1 . On montre de même, que pour toute racine non nulle

de Q_1 , z_0 , nécessairement $\frac{1}{z_0}$ est aussi racine de Q_1 . La relation

$$Z^q P_1 \overline{Q} = Z^p \overline{P} Q_1$$

entraîne des relations intéressantes sur les degrés des polynômes :

$$\begin{cases} p = q & \text{si } 0 \text{ n'est racine ni de P, ni de Q} \\ p = q - n_0 & \text{si } 0 \text{ est racine d'ordre } n_0 \text{ pour P} \\ q = p - n_0 & \text{si } 0 \text{ est racine d'ordre } n_0 \text{ pour Q} \end{cases}$$

$f(\lambda)$ étant la densité spectrale de puissance, est positive et intégrable sur $[0, 1]$. $Q_1(z)$ ne peut donc pas avoir de racine de module 1. Supposons que $e^{2i\pi\theta}$ soit racine de P_1 , d'ordre impair

$$e^{2i\pi\lambda} - e^{2i\pi\theta} = 2i e^{i\pi(\lambda+\theta)} \sin \pi(\lambda - \theta)$$

on aurait donc

$$f(\lambda) = [\sin \pi(\lambda - \theta)]^{2k+1} \varphi(\lambda)$$

avec

$$\varphi(\theta) \neq 0$$

cela contredit le fait que $f(\lambda)$ est positive ou nulle. Comme il y a correspondance entre les racines de module < 1 et celles de module > 1 , P_1 et Q_1 sont de degré pair.

a) Si 0 n'est racine ni de P, ni de Q, P_1 et Q_1 sont de degré p qui est pair et égal à q

$$P_1(z) = \text{Cte} \prod_{|\alpha_k| < 1} (z - \alpha_k) \prod_{|\alpha_k| > 1} (z - \alpha_k) \prod_{|\alpha_k| = 1} (z - \alpha_k)^2$$

$$Q_1(z) = \text{Cte} \prod_{|\beta_k| < 1} (z - \beta_k) \prod_{|\beta_k| > 1} (z - \beta_k)$$

Posons

$$H(z) = \frac{\prod_{|\alpha_k| < 1} (z - \alpha_k) \prod_{|\alpha_k| = 1} (z - \alpha_k)}{\prod_{|\beta_k| < 1} (z - \beta_k)}$$

Alors, si $z = e^{2i\pi\lambda}$

$$H(z) \overline{H(z)} = \text{Cte} \frac{P_1(z)}{Q_1(z)} = \text{Cte} f(\lambda)$$

$$\Rightarrow f(\lambda) = c |H(z)|^2 \quad \text{si } z = e^{2i\pi\lambda}$$

Il suffit d'un filtre h'_n tel que

$$\sum_n h'_n e^{-2i\pi n\lambda} = \frac{1}{H(e^{2i\pi\lambda})}$$

pour obtenir X_n , de densité spectrale de puissance égale à une constante.

b) Si 0 est racine de P ou Q, d'après ce qui a été vu au-dessus,

$$H(z) \overline{H(z)} = \text{Cte} \frac{P_1(z)}{Q_1(z)} = \text{Cte} z^{q-p} f(\lambda) \quad \text{pour } z = e^{2i\pi\lambda}$$

Donc $f(\lambda) = \text{Cte} z^{p-q} |H(z)|^2$ pour $z = e^{2i\pi\lambda}$

$f(\lambda)$ étant positive cela n'est possible que si $p = q$ et on retrouve

$$f(\lambda) = c |H(e^{2i\pi\lambda})|^2$$

Dans les deux cas, on réécrit $H(z)$ pour trouver le filtre inverse :

$$H(z) = \frac{\prod_{|\alpha_k| < 1} z(1 - \alpha_k z^{-1}) \prod_{|\alpha_k| = 1} z(1 - \alpha_k z^{-1})}{\prod_{|\beta_k| < 1} z(1 - \beta_k z^{-1})}$$

$$= \frac{\prod_{|\alpha_k| < 1} (1 - \alpha_k z^{-1}) \prod_{|\alpha_k| = 1} (1 - \alpha_k z^{-1})}{\prod_{|\beta_k| < 1} (1 - \beta_k z^{-1})}$$

et

$$\frac{1}{H(z)} = \sum_{n \geq 0} h'_n z^{-n}$$

EXERCICES

E.6.1. On considère le signal $s_0(t)$, signal rectangle de durée T :

$$s_0(t) = 1 \left[-\frac{T}{2}, \frac{T}{2} \right] (t)$$

$s_0(t)$ passe dans un filtre passe-bas idéal, de fonction de transfert

$$H(\lambda) = 1 \left[-\frac{1}{2}, \frac{1}{2} \right] (\lambda)$$

et a pour sortie le signal $s_1(t)$, $s_0(t)$ est bruité additivement par un bruit blanc de niveau B_0 , de sorte que la sortie est $s_1(t) + Y(t)$.

Calculer le niveau du bruit en sortie : $B_1 = E(Y_t^2)$.

Calculer le rapport signal sur bruit, R ,

$$R = \frac{s_1^2(t)}{\text{var}(Y_t)}$$

- a) lorsque TF est petit
b) lorsque TF est grand.

En déduire qu'il existe une valeur du produit TF qui rend ce rapport maximal.

E.6.2. On considère un filtre de fonction de transfert $H(\lambda)$. Le signal d'entrée $s_0(t)$ est brut additivement par un bruit aléatoire X_t que l'on suppose centré, stationnaire au second ordre et de densité spectrale de puissance $f(\lambda)$. Soit $s_1(t)$ la sortie de $s_0(T)$ et Y_t celle du bruit X_t .

a) Donner l'expression sous forme intégrale du niveau de bruit B_1 en sortie : $B_1 = \text{var}(Y_t)$.

b) Montrer que l'on a

$$|s_1(t_0)|^2 \leq B_1 \int_{\mathbb{R}} \frac{|\hat{s}_0(\lambda)|^2}{f(\lambda)} d\lambda$$

En déduire une condition sur le filtre pour que le rapport signal sur bruit, en sortie, soit maximum.

c) Montrer que si X_t est un bruit blanc, le filtre optimal est un corrélateur c'est-à-dire $h(t) = \text{Cte } s_0(t_0 - t)$. Montrer que le filtre optimal est le même pour tous les signaux $s_0(t) = au(t - \tau)$ où a et τ varient.

E.6.3. Soit X_t un signal aléatoire réel, centré, gaussien, stationnaire dont presque toutes les trajectoires sont continûment dérivables. La fonction de covariance $k(\tau) = \text{cov}(X_t, X_{t+\tau})$ est supposée deux fois continûment dérivable avec $k'(0) = 0$. On pose $Y_t = X'_t$.

a) Montrer que Y_t est stationnaire et calculer sa fonction de covariance.

b) Quelle est la loi du couple gaussien $(X_t, \frac{X_{t+h} - X_t}{h})$? En déduire la loi du couple (X_t, Y_t) .

E.6.4. Soit X_n un signal aléatoire, réel, stationnaire au second ordre, de fonction de covariance $k(n)$. Calculer la prédiction de X_{n+1} , de la forme aX_n qui minimise l'erreur quadratique moyenne :

$$E(|aX_n - X_{n+1}|^2)$$

E.6.5. On reprend les notations de l'exercice 6.3.

Trouver l'interpolation de X_n , du type $aX_{n-1} + bX_{n+1}$ qui minimise l'erreur quadratique moyenne

$$E(|X_n - (aX_{n-1} + bX_{n+1})|^2)$$

E.6.6. Une suite $X_n = \pm 1$ est transmise dans un canal, de réponse impulsionnelle h_n réelle. On suppose X_n aléatoire centré et stationnaire au second ordre ; la

transmission est bruitée de sorte que l'on reçoit Y_n :

$$Y_n = \sum_k h_k X_{n-k} + \varepsilon_n$$

où ε_n est un bruit centré, stationnaire.

On appelle $f(\lambda)$ la densité spectrale de puissance de X_n de $f_2(\lambda)$ celle du bruit ε_n qui est supposé indépendant du signal d'entrée X_n .

Calculer $E[(Y_n - X_n)^2]$ (on donnera l'expression intégrale dans le domaine spectral).

E.6.7. On reprend les notations de l'exercice 6.5. Trouver le paramètre θ qui minimise

$$E[(\theta Y_n - X_n)^2]$$

E.6.8. Soit un système donné par la relation

$$y_{n+1} = \theta y_n + x_{n+1}; \quad n \in \mathbb{Z}$$

a) Déterminer la fonction de transfert du système, c'est-à-dire la transformée en Z . Déterminer le filtre causal qui a cette fonction de transfert : quelle est la réponse impulsionnelle, le filtre est-il stable ?

b) θ étant choisi de façon à ce que le filtre causal soit stable, calculer la loi de Y_n lorsque l'entrée X_n est un bruit blanc.

E.6.9. Soit $X_n = an + b + \varepsilon_n$ où ε_n est un bruit centré, stationnaire au second ordre, de densité spectrale de puissance $f(\lambda)$ et les coefficients a et b sont déterministes. On filtre X_n de la manière suivante :

$$Y_{n+1} = X_{n+1} - X_n$$

Montrer que le processus Y_n est stationnaire au second ordre et calculer sa densité spectrale de puissance.

CHAPITRE 7

CODAGE ET COMMUNICATIONS

7.1. INTRODUCTION

Tout le monde connaît le Morse où chaque lettre de l'alphabet est remplacée par une succession de points et de tirets. Un exemple plus proche des applications actuelles est celui du code ISO (alphabet international n° 5) qui est une version améliorée du code ASCII. Dans ce code, les lettres de l'alphabet, majuscules et minuscules, les chiffres, les symboles servant à construire un texte, ainsi qu'un certain nombre de commandes comme « CR » (carriage return) sont codés en binaires sur 7 bits ; certains mots binaires sont réservés à une utilisation propre à chaque pays comme la lettre « à » en français ou le symbole « ¿ » en espagnol. Dans la transmission de signaux téléphoniques, le signal est numérisé de la façon suivante :

- on échantillonne au pas $\Delta t = 125 \mu s$
- on remplace l'amplitude par quantification : 256 niveaux sont choisis
- chaque niveau est codé sur 8 bits

Pour les modems les symboles 0, 1 sont transmis en envoyant des impulsions aux 2 fréquences 1 300 Hz et 2 100 Hz.

Dans les méthodes dites MIA (modulation d'impulsion en amplitude), la transmission des symboles numériques $a_k \in \{-(M-1), \dots, -1, 1, 2, \dots, M\}$ se fait en générant un signal $s(t)$ analogique,

$$s(t) = \sum_k a_k h(t - kT)$$

où la fonction $h(t)$ est de durée finie, inférieure à T . Ainsi dans le code NRZ binaire (Non Retour à Zéro),

$$a_k = \pm 1 \quad \text{et} \quad h(t) = A 1_{[0, T]}(t)$$

La suite 11-11-11 est remplacée par le signal dessiné sur la figure 7.1, tandis que dans le code RZ,

$$h(t) = A 1_{[0, T/2]}$$

et on obtient le signal dessiné sur la figure 7.2 (en réalité, dans le code RZ, $a_k = 0$ ou 1).

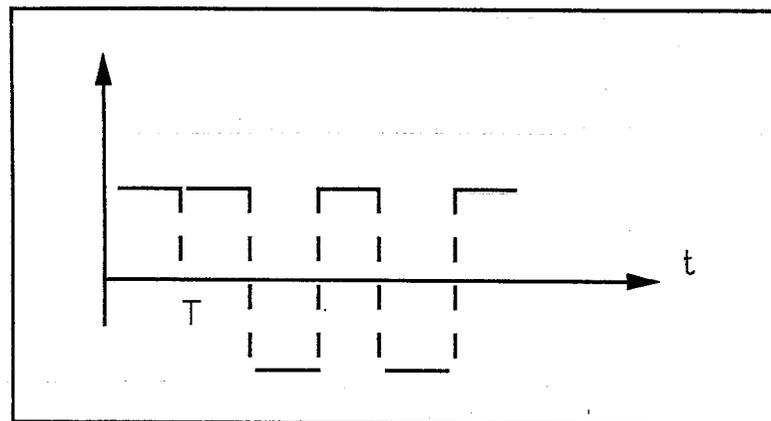


Figure 7.1.

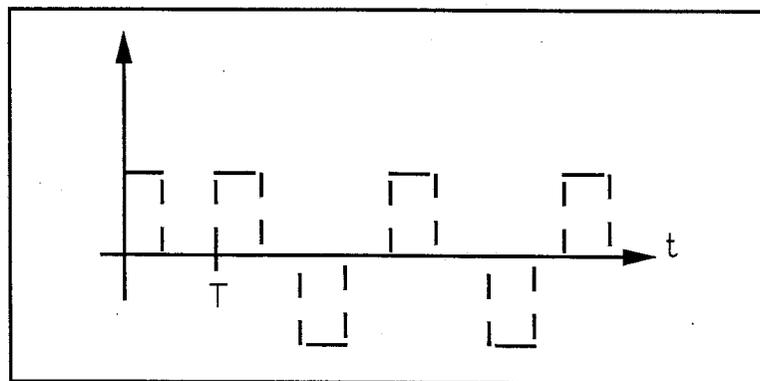


Figure 7.2.

Il existe de nombreux systèmes de modulation ; les codes NRZ et RZ demandent une bonne synchronisation avec le signal d'horloge qui donne le rythme ; d'autre part, le calcul du spectre de puissance montre qu'une grande partie de la puissance est concentrée autour de la fréquence nulle et on leur préfère d'autres codes (bipolaires, biphasé) ; de façon générale un système de modulation est spectralement efficace si le rapport du débit d'information sur la largeur de bande occupée est proche de 1 ; on pourra consulter à ce sujet le livre « Téléinformatique » de Guilbert et Macchi.

En ce qui concerne les techniques de codage elles-mêmes, deux points de vue vont faire jour dans les paragraphes qui suivent : certains codages assurent la

transmission avec une grande sécurité, voire permettent de corriger des erreurs de transmission (§ 2), d'autres permettent au contraire d'éliminer des redondances et d'économiser de l'espace et du temps de transmission (§ 3).

Pour simplifier tous les codages considérés dans ce chapitre sont des codages en binaire ; les éléments à coder seront appelés des messages sans préjuger de leur nature : lettre de l'alphabet, chiffre, mot, mot binaire... A chaque message est associé un mot binaire, c'est-à-dire une succession de 0 et de 1, que l'on appelle un mot code.

7.2. CODAGES EN BLOCS

7.2.1. Contrôle de parité

Soit N le nombre de messages à coder ; on commence par choisir k de façon à ce que le nombre de mots codes soit supérieur ou égal à N , c'est-à-dire tel que $2^k \geq N$; chaque message est codé par un mot binaire de longueur k , $a_1 a_2 \dots a_k$ puis on rajoute un $k + 1$ -ième élément, a_{k+1} , appelé bit de contrôle de parité :

$$a_{k+1} = 0 \quad \text{si} \quad \sum_{p=0}^k a_p \text{ est pair}$$

$$a_{k+1} = 1 \quad \text{sinon}$$

Lorsqu'on transmet un mot code, on vérifie à l'arrivée que la somme des termes est paire (contrôle) ; il y a alors plusieurs possibilités qui s'excluent mutuellement :

- pas d'erreur de transmission
- une erreur exactement
- un nombre pair d'erreurs
- un nombre impair, strictement supérieur à 1, d'erreurs.

Appelons A_1, A_2, A_3, A_4 les événements correspondants et calculons leur probabilité lorsque les erreurs sont supposées indépendantes et que chaque symbole binaire est mal transmis avec la probabilité ε (canal binaire symétrique) :

$$P(A_1) = P(\text{aucune erreur}) = (1 - \varepsilon)^{k+1}$$

$$P(A_2) = P(1 \text{ erreur exactement}) = (k + 1) \varepsilon (1 - \varepsilon)^k$$

$$P(A_3) = P(\text{un nombre pair d'erreurs}) = \sum_{\substack{j \neq 0 \\ j \text{ pair}}} C_{k+1}^j \varepsilon^j (1 - \varepsilon)^{k+1-j}$$

$$P(A_4) = P(\text{un nombre impair d'erreurs} > 1) = \sum_{\substack{j > 1 \\ j \text{ impair}}} C_{k+1}^j \varepsilon^j (1 - \varepsilon)^{k+1-j}$$

Pour $k = 7$ et $\varepsilon = 10^{-3}$ on trouve

$$P(A_1) = 0,99203 \quad P(A_2) = 0,00794 \quad P(A_3) = 2,8 \cdot 10^{-5} \quad P(A_4) = 6 \cdot 10^{-8}$$

En cas d'erreur, c'est l'événement A_2 qui est prépondérant, l'événement A_3 est de probabilité négligeable. Bien que ce code ne permette de détecter qu'un nombre impair d'erreurs, en négligeant le terme $P(A_3)$ on peut considérer qu'il permet de détecter une transmission erronée.

7.2.2. Codes linéaires (n, k) , codes systématiques

Avec les codes linéaires systématiques on généralise le procédé du contrôle de parité en rajoutant au mot binaire de longueur k un certain nombre de symboles binaires qui vont servir au contrôle :

$$^1a = (a_1, a_2, \dots, a_k) \rightarrow (a_1, a_2, \dots, a_k, a_{k+1}, \dots, a_n)$$

Les éléments a_{k+1}, \dots, a_n sont des combinaisons linéaires des k premiers éléments a_1, \dots, a_k . Le code est dit systématique car on rajoute systématiquement $n - k$ symboles binaires. **Tous les calculs sont faits modulo 2** et le code est caractérisée par la matrice dite matrice génératrice du code et notée G :

$$(a_1, \dots, a_k, \dots, a_n) = (a_1, \dots, a_k) G \text{ et } G = \left(\begin{array}{c|c} \begin{array}{ccc} 1 & \dots & 0 \\ 0 & & \\ \vdots & & \\ 0 & & 1 \end{array} & \begin{array}{c} \boxed{P} \\ \vdots \\ \boxed{P} \end{array} \end{array} \right) \begin{array}{l} \uparrow k \\ \downarrow n-k \end{array}$$

En particulier, la première ligne de G est le mot code associé au mot source $(1, 0, \dots, 0), \dots$, la k -ième ligne est le mot code associé au mot source $(0, 0, \dots, 1)$.

Avec cette définition, l'ensemble des mots codes forme un sous-espace vectoriel de $\{0, 1\}^n$, qui est de dimension k . De façon générale, un code linéaire (n, k) est la donnée d'un sous-espace vectoriel U de $\{0, 1\}^n$, de dimension k et tout code linéaire (n, k) peut être réduit à une forme systématique (voir « Geometry and codes » de V. D. Goppa) ; aussi ce paragraphe sera-t-il consacré à l'étude d'un code linéaire **mis sous forme systématique**.

On définit le taux de transmission comme étant le rapport k/n .

Matrice de contrôle

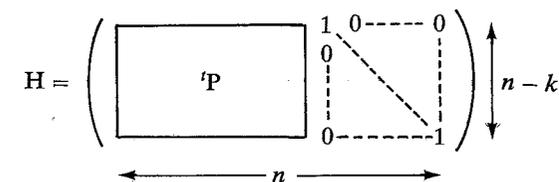
Après transmission du mot code on vérifie qu'il n'y a pas d'erreurs de transmission ; il est nécessaire que le mot reçu (b_1, \dots, b_n) soit de la forme 1aG , c'est-à-dire que l'on ait :

$$\begin{aligned} b_{k+1} &= (b_1, \dots, b_k) \cdot (1^{\text{e}} \text{ colonne de } P) \\ b_{k+2} &= (b_1, \dots, b_k) \cdot (2^{\text{e}} \text{ colonne de } P) \\ &\vdots \\ b_n &= (b_1, \dots, b_k) \cdot (\text{dernière colonne de } P) \end{aligned}$$

Or deux nombres binaires sont égaux si et seulement si leur somme est nulle modulo 2 ; d'où le procédé de contrôle :

$$(b_1, \dots, b_n) = \text{mot de code} \Leftrightarrow \begin{cases} b_{k+1} + (b_1, \dots, b_k) (1^{\text{e}} \text{ col. de } P) = 0 \\ \vdots \\ b_n + (b_1, \dots, b_k) (\text{dernière col. de } P) = 0 \end{cases} \Leftrightarrow Hb = 0$$

avec



Exemple fondamental : Le code de Hamming $(4, 7)$

Ce code a pour matrice génératrice

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

La matrice de contrôle est

$$H = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

On constate que les 16 mots codes, qui sont des mots binaires de longueur 7, différent deux à deux sur au moins 3 symboles binaires, en abrégé, 3 bits (binary digit) :

0000 000	01000 111	1000 101	1100 010
0001 011	0101 100	1001 110	1101 001
0010 110	0110 001	1010 011	1110 100
0011 101	0111 010	1011 000	1111 111

Détection et correction des erreurs dans un code linéaire (n, k) systématique

Le code permet de détecter une transmission erronée lorsque le mot reçu ne vérifie pas la relation de contrôle

$Hb = 0$

En utilisant la linéarité du procédé de contrôle,

$$\begin{aligned} (b_1, \dots, b_n) &= \text{mot code envoyé} + \text{mot erreur} \\ Hb = 0 &\Leftrightarrow H(\text{mot envoyé}) + H(\text{mot erreur}) = 0 \\ &\Leftrightarrow H(\text{mot erreur}) = 0 \\ &\Leftrightarrow \text{le mot erreur est un mot du code} \end{aligned}$$

Une transmission erronée n'est pas détectée par le procédé de contrôle si et seulement si le mot erreur est un mot code.

Il est donc important de calculer la plus petite distance entre deux mots codes au sens suivant :

$$d(m, m') = \underset{\text{(définition)}}{\text{nombre de bits où } m \text{ et } m' \text{ sont différents}}$$

Cette distance s'appelle la distance de Hamming.
On définit la distance minimale entre mots du code

$$d = \underset{\substack{m \neq m' \\ m, m' \text{ mots du code}}}{\text{Inf}} d(m, m')$$

Le code étant linéaire, on a

$$d = \underset{\substack{m \neq 0 \\ m \text{ mot du code}}}{\text{Inf}} d(m, (0, \dots, 0))$$

Pour le code de Hamming $d = 3$.

On remarque que les nombres n, k, d sont reliés ; en effet, les mots sources $(1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, 0, \dots, 0, 1)$ ont pour images des mots codes qui sont à distance deux à deux supérieure ou égale à 2 ; les $n - k$ bits de contrôle sont donc à distance deux à deux supérieure ou égale à $d - 2$, ce qui n'est pas possible si $n - k \leq d - 2$, d'où l'inégalité

$$n - k > d - 2$$

Cette inégalité est appelée inégalité de Singleton.

Le mot erreur, s'il est non nul, aura donc au moins d bits non nuls si il n'est pas détecté et on peut calculer la probabilité d'un tel événement avec les mêmes hypothèses que dans le cas du contrôle de parité (canal binaire symétrique et erreurs indépendantes)

$$P(\text{au moins } d \text{ bits erronés}) = \sum_{j=d}^n C_n^j \varepsilon^j (1 - \varepsilon)^{n-j}$$

En résumé, on a :

$$\begin{aligned} P(\text{transmission erronée non détectée}) &= \sum_{m = \text{mot du code}} \varepsilon^{d(m,0)} (1 - \varepsilon)^{n-d(m,0)} \\ P(\text{erreur détectée/au plus } d - 1 \text{ bits erronés}) &= 1 \end{aligned}$$

Pour le code de Hamming, on a $n = 7, k = 4$ et $d = 3$, ce qui donne

$$P(\text{transmission erronée non détectée}) = \varepsilon^7 + 6 \varepsilon^4 (1 - \varepsilon)^3 + 8 \varepsilon^3 (1 - \varepsilon)^4$$

Pour $\varepsilon = 10^{-3}$, on trouve une probabilité de l'ordre de 10^{-8} .

D'autre part, si d est impair, $d = 2p + 1$, ou si $d = 2p$ et si il y a p bits erronés, le mot erreur vérifie

$$d(\text{mot erreur}, (0, \dots, 0)) = p$$

On est donc sûr de le détecter par le procédé de contrôle ; si $d = 2p + 1$ on peut même corriger les erreurs, en effet, le mot erreur appartient à une boule centrée sur $(0, \dots, 0)$ de rayon p , qui, par définition de d et parce que $d = 2p + 1$, ne contient pas d'autre mot code que $(0, \dots, 0)$ d'où la correction :

$$\text{mot reçu} = \text{mode code} + \text{mot erreur} \rightarrow \text{mot code}$$

ce qui revient à chercher le mot code *le plus proche du mot reçu*.

$$\begin{aligned} \text{Si } d = 2p + 1, \quad &P(\text{détection et correction/il y a au plus } p \text{ bits erronés}) = 1 \\ \text{Si } d = 2p, \quad &P(\text{détection/il y a au plus } p \text{ bits erronés}) = 1 \end{aligned}$$

Le procédé de correction « chercher le mot code le plus proche » est applicable sans être sûr à 100 %, lorsque le nombre d'erreurs est supérieur à p , si le mot reçu admet un mot code unique comme plus proche voisin parmi tous les mots codes ; cela revient à chercher le mot code le plus vraisemblable, c'est-à-dire celui qui réalise

$$\underset{x}{\text{Sup}} P(x \text{ est envoyé}/y \text{ est reçu})$$

Si l'on suppose que les bits sont transmis par un canal binaire symétrique.

7.2.3. Codes cycliques (n, k)

Définition : Un code linéaire (n, k) systématique est dit cyclique si toute permutation circulaire d'un mot code est encore un mot code.

A tout mot code (b_1, \dots, b_n) associons le polynôme

$$P(X) = b_1 X^{n-1} + b_2 X^{n-2} + \dots + b_{n-1} X + b_n$$

Lorsqu'on effectue la permutation circulaire

$$(b_1, b_2, \dots, b_n) \mapsto (b_2, b_3, \dots, b_n, b_1)$$

le polynôme $P(X)$ est transformé en $Q(X)$:

$$\begin{aligned} Q(X) &= b_2 X^{n-1} + \dots + b_n X + b_1 = X[P(X) - b_1 X^{n-1}] + b_1 \\ &= XP(X) - b_1 X^n + b_1 = XP(X) + b_1(X^n + 1) \\ Q(X) &\equiv XP(X) \pmod{X^n + 1} \end{aligned}$$

Au bout de n permutations circulaires on retrouve le mot de départ d'où la caractérisation d'un code cyclique comme suit.

Proposition : Le code est cyclique si et seulement si quel que soit le polynôme code $P(X)$, les polynômes $XP(X) \pmod{X^n + 1}$, $X^2 P(X) \pmod{X^n + 1}$, ..., $X^{n-1} P(X) \pmod{X^n + 1}$, sont aussi des polynômes du code.

Polynôme générateur du code

L'ensemble des polynômes du code n'est pas quelconque ; en effet cet ensemble :

- est inclus dans l'ensemble des polynômes de degré inférieur ou égal à $n - 1$
- est un sous-espace vectoriel sur le corps $F_2 = \{0, 1\}$ car le code est linéaire
- est stable par multiplication par un polynôme quelconque à condition de regarder le résultat modulo le polynôme $(X^n + 1)$.

Suivant un procédé classique en algèbre, soit $g(X)$ un polynôme non nul de plus petit degré parmi les polynômes du code. Montrons que tous les polynômes du code sont multiples de g ; soit $P(X)$ un polynôme quelconque du code ; en faisant la division de P par g on a

$$\begin{aligned} P(X) &= g(X) Q(X) + R(X) \text{ avec } \text{degré}(R) < \text{degré}(g) \\ \text{soit} \\ R(X) &= P(X) - g(X) Q(X) = \text{polynôme du code} \end{aligned}$$

g étant de plus petit degré parmi les polynômes du code on doit donc avoir $R = 0$, c'est-à-dire P multiple de g . Le degré de g est nécessairement égal à $n - k$ si on veut engendrer 2^k éléments par multiplication avec des polynômes.

D'autre part, si on regarde les lignes de la matrice génératrice G , on constate que les polynômes associés sont des polynômes du code qui engendrent par combinaisons linéaires tous les polynômes du code et c'est le polynôme associé à la dernière ligne qui est de plus petit degré ; c'est donc le polynôme associé à la dernière ligne de G qui est le polynôme générateur. Nous allons voir que le polynôme générateur n'est pas quelconque dans la proposition qui suit.

Proposition : Un code linéaire systématique (n, k) est un code cyclique si et seulement si tous les polynômes du code sont multiples du polynôme $g(X)$ associé à la dernière ligne de G et si g est un diviseur de $X^n + 1$.

Démonstration : g est de degré $n - k$. $X^k g(X)$ est donc de degré n :

$$X^k g(X) = X^n + 1 + R(X) \text{ avec } R(X) = X^k g(X) \pmod{X^n + 1}$$

Si le code est cyclique $R(X)$ est un polynôme du code, c'est-à-dire un multiple de $g(X)$ et on en déduit que $X^n + 1$ est multiple de $g(X)$.

Réciproquement, si tous les polynômes du code sont multiples de $g(X)$ et que $g(X)$ divise $X^n + 1$, montrons par récurrence que $X^p g(X)$ est un polynôme du code.

L'avant-dernière ligne de G a pour polynôme $Q(X)$ de degré $n - k + 1$, multiple de $g(X)$

$$Q(X) = Xg(X) + R(X) \text{ avec } d^0(R) \leq n - k$$

Supposons $Xg(X), \dots, X^{k-p-1}g(X)$ polynômes du code.

En considérant la p -ième ligne de G , on obtient un polynôme, P , du code qui est de degré $n - p$ et multiple de $g(X)$:

$$\begin{aligned} P(X) &= X^{k-p}g(X) + R(X) \text{ avec } R(X) \text{ de } d^0 n - p - 1 \\ \Rightarrow R(X) &= \text{multiple de } g(X) \\ &= (X^{k-p-1} + \dots)g(X) \end{aligned}$$

$R(X)$ est donc un polynôme du code d'après l'hypothèse de récurrence. $X^{k-p}g(X)$ est la somme de $P(X)$ et de $R(X)$, et appartient au code par linéarité. C.Q.F.D.

Exemple : Le code de Hamming

Les polynômes associés aux différentes lignes de G sont tous multiples du polynôme générateur $g(X) = X^3 + X + 1$ qui divise $X^7 + 1$.

Les codes cycliques les plus performants sont les codes BCH (de Bose, Ray-Chaudhuri et Hocquenghem) : On pourra consulter à ce sujet « Coding theory » de J. H. van Lint, Ed. Springer Verlag. Ce sont des codes cycliques (n, k) avec

$$n = 2^m - 1; \quad k \geq 2^m - md - 1$$

qui corrigent toute combinaison d'au plus d erreurs. Il existe aussi des codes adaptés à la correction de paquets d'erreurs.

7.2.4. Les codes polynomiaux

7.2.4.1. Construction de codes polynomiaux

La construction d'un code cyclique (n, k) impose de trouver un polynôme de degré $n - k$ qui soit un diviseur de $X^n + 1$, ce qui n'est pas possible si le polynôme $X^n + 1$ est irréductible. On peut trouver des tables donnant les diviseurs de $X^n + 1$ pour différentes valeurs de n .

D'autre part, les avantages de facilité de codage, ne nécessitent pas que le polynôme $g(X)$ divise $X^n + 1$. En effet considérons les codes linéaires, dits

polynomiaux, tels que tous les polynômes du code sont multiples du polynôme $g(X)$ associé à la dernière ligne de G :

Soit a_1, a_2, \dots, a_k un mot source et $A(X)$ son polynôme :

$$A(X) = a_1 X^{k-1} + \dots + a_k$$

$$(a_1, \dots, a_k) \mapsto (a_1, \dots, a_n)$$

$$a_1 X^{n-1} + \dots + a_n = P(X) = (a_1, \dots, a_k) G \begin{pmatrix} X^{n-1} \\ \vdots \\ 1 \end{pmatrix}$$

$$= a_1(\text{Polyn.}) + a_2(\text{Polyn.}) + \dots + a_k(\text{Polyn.})$$

$$= X^{n-k} A(X) + R(X) \text{ avec } d^0(R) \leq n - k - 1$$

P multiple de $g \Rightarrow$

$$\begin{aligned} R(X) &= X^{n-k} A(X) \bmod g(X) \\ P(X) &= X^{n-k} A(X) + [X^{n-k} A(X) \bmod g(X)] \end{aligned}$$

7.2.4.2. Exemple du Minitel

On code des mots binaires de longueur 120 (15 octets) avec le polynôme générateur $X^7 + X^3 + 1$. Ici $n - k = 7$ et $k = 120$. Le codage s'effectue en calculant $X^7 A(X)$ modulo $X^7 + X^3 + 1$. Du point de vue de la protection contre les erreurs le code est plus raffiné qu'il n'y paraît à première vue ; en effet, le mot source, qui est de longueur 120, est constitué de 15 octets, chaque octet contenant en dernière position un bit de parité. Nous venons de voir comment on rajoute 7 bits de contrôle à ces 15 octets ; on rajoute alors un bit de contrôle de parité sur le mot obtenu et on arrive ainsi à 16 octets c'est-à-dire un mot de longueur 128. Au total, il y a 23 bits qui servent à la protection ce qui correspond à 18 % de la longueur totale.

7.2.4.3. Détection et correction des erreurs

Montrons qu'une erreur est automatiquement détectée ; en effet, supposons qu'il y ait une erreur parmi les $n - k$ derniers bits ; si l'erreur avait lieu par exemple sur le $n - k + 2$ -ième bit, et n'était pas détectée, cela voudrait dire que $0 \dots 0010 \dots 0$ est un mot code ce qui est impossible pour un code linéaire (n, k) systématique.

Si l'erreur a lieu sur les k premiers bits et si elle n'est pas détectée, on peut écrire polynôme du mot reçu = polynôme code envoyé + $X^p =$ multiple de $g(X)$.

Si on regarde le reste modulo $g(X)$, on obtient

$$X^p = 0 \bmod g$$

ce qui est impossible.

On pourra même retrouver la position de l'erreur dans le cas où les polynômes $X^n \bmod g, X^{n-1} \bmod g, \dots, X^{n-k} \bmod g$ sont tous différents.

On a vu au paragraphe 7.2.2 le rôle important joué par la distance minimale, pour la correction des erreurs ; le paragraphe 7.2.5 donne quelques précisions sur les relations entre n, k et d . Par ailleurs, on trouvera au paragraphe 7.4, le résultat de Shannon qui donne la relation entre les probabilités d'erreurs, longueur du code et le taux de transmission.

7.2.5. Compression avec distorsion

Un code linéaire (n, k) systématique peut être utilisé pour faire de la compression de mots binaires de longueur n sur des mots binaires de longueur k : on associe à chaque mot de longueur n , qui est élément quelconque de $\{0, 1\}^n$, un mot code (qui est aussi de longueur n) que l'on décode ; on obtient ainsi un mot de longueur k qui servira dans les transmissions car il occupe moins de place...

Le mot code qui remplace le mot de longueur n , doit être le plus proche possible ; cette construction est donc utilisée pour un code dit parfait, c'est-à-dire tel que $d = 2p + 1$ et tout élément de $\{0, 1\}^n$ est dans une boule, de rayon p , centrée sur un mot code ; les éléments d'une telle boule sont tous remplacés par le mot code qui sert de centre qui est ensuite décodé sur un mot de longueur k .

Exemple avec le code de Hamming : Si on considère $\{0, 1\}^7$, ensemble de tous les mots binaires de longueur 7, on peut remplacer chaque mot par le mot code de Hamming qui est le plus proche, puis par le mot source de longueur 4 avant d'effectuer la transmission, ce qui donne pour le mot 1 000 000, s'il n'y a pas d'erreur de transmission le schéma de la figure 7.3.

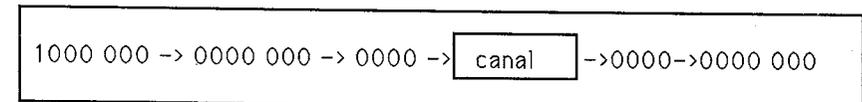


Figure 7.3.

7.2.6. Codes de Goppa (codes géométriques)

Ce sont des codes linéaires au sens où l'on passe d'un mot a à un mot b par une transformation linéaire, l'ensemble des mots codes formant un sous espace vectoriel de $\{0, 1\}^n$.

Ces codes reposent sur les notions de corps fini à q éléments, noté \mathbb{F}_q , q étant une puissance d'un nombre premier, et de courbe algébrique sur un tel corps. Sans rentrer dans la théorie des corps finis (on pourra consulter à ce sujet « Le défi algébrique » de C. Mutafian Ed. Vuibert), on peut considérer l'exemple le plus simple d'un corps fini contenant \mathbb{F}_2 , à savoir le corps \mathbb{F}_4 : sur \mathbb{F}_2 le seul polynôme de degré 2 irréductible, c'est-à-dire non décomposable en

produits de facteurs, est le polynôme $X^2 + X + 1$. Tout élément de $\{0, 1\}^2$ étant identifié à son polynôme, le produit de deux éléments s'obtient en regardant le résultat du produit des deux polynômes, $P(X)Q(X)$, modulo $X^2 + X + 1$; on a alors la table de multiplication sur $\{0, 1\}^2$ qui en fait un corps noté \mathbb{F}_4 :

		00	01	10	11
		0	1	α	α^2
00	0	00	00	00	00
01	1	00	01	10	11
10	α	00	10	11	01
11	α^2	00	11	01	10

En particulier, l'inverse de α est α^2 .

Pour tout m , on peut de la même manière identifier $\{0, 1\}^m$ au corps \mathbb{F}_q à $q = 2^m$ éléments, en choisissant un polynôme de degré m irréductible.

G sera la matrice d'une application linéaire de \mathbb{F}_q^k dans \mathbb{F}_q^n . La distance de Hamming entre deux éléments de \mathbb{F}_q^k , (x_1, x_2, \dots, x_k) et (y_1, \dots, y_k) , est égale, par définition, au nombre d'indices où x_i et y_i diffèrent.

Goppa montre comment construire un code linéaire inclus dans \mathbb{F}_q^n ayant la plus grande taille possible, ce qui correspond à k le plus grand possible, et assurant une distance minimum égale à d fixé (on rappelle que si $d = 2p + 1$ on peut détecter et corriger p erreurs). Ainsi si on reprend l'exemple du corps \mathbb{F}_4 , le code $(6, 3)$ défini par

$$x \in \text{code} \Leftrightarrow Hx = 0$$

avec

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & \alpha & \alpha^2 & 0 & 1 \\ 0 & 1 & \alpha^2 & \alpha & 1 & 0 \end{pmatrix}$$

a pour distance minimale $d = 4$.

On démontre que le taux de transmission, la longueur n et la distance minimale d sont reliés par la contrainte :

$$\sum_{d=0}^D (q-1)^d C_n^d < q^{(1-R)n}$$

Si n est fixé et si on choisit R entre 0 et $\frac{q-1}{q}$ on voit que d est inférieur à une borne, appelée borne de Gilbert.

Or avec le procédé de Goppa, on arrive à construire un code qui approche cette borne à ε près; plus précisément, $\forall \varepsilon > 0 \forall 0 < R < \frac{q-1}{q}$, il existe un code

de Goppa de taux de transmission R , de longueur n et de distance minimale d supérieure ou égale à $d_{\min} - \varepsilon_n$ (on travaille en fait sur la distance relative $\frac{d}{n}$).

Pour la construction de ces codes, on pourra se reporter au livre « Geometry and Codes » de Goppa, Editions Kluwer Academic Publisher group (livre théorique) ou à l'article « A transform approach to Goppa codes » M. Loeloeian, J. Conan — IEEE, *Trans. I.T.* vol. 33 n° 1, Jan. 1987, p. 105-115.

7.3. CODAGES DE LONGUEUR VARIABLE

7.3.1. Introduction

L'idée de codages de longueur variable est basée sur le fait que certains messages à coder sont plus fréquents que d'autres; les messages fréquents ont donc intérêt à être codés sur des mots de longueur plus petite que les messages peu fréquents si on veut améliorer le rendement. La difficulté va résider dans le décodage lorsqu'on envoie les mots codes les uns à la suite des autres: dans le codage en bloc, à la succession de messages $m_1 m_2 \dots$ correspond la succession de blocs de 0 et de 1:

$$a_1 a_2 \dots a_k b_1 b_2 \dots b_k \dots$$

A la réception (après transmission) il suffit de tronçonner par paquets de k puis de décoder chaque mot binaire. Si la longueur du mot code est variable, ce « tronçonnage » n'est plus possible mais une façon de tourner la difficulté est d'imposer au codage d'obéir à ce que l'on appelle la condition du préfixe: « aucun mot code n'est le début d'un autre mot code ». Avec cette condition du préfixe, il y a une seule façon de tronçonner la succession de 0 et de 1 pour retrouver la suite de mots code. Un code qui obéit à la condition du préfixe est dit irréductible.

Lorsqu'on construit le code, on peut vérifier la contrainte du préfixe en utilisant un graphe arborescent: lorsqu'un mot binaire est choisi comme mot code, il est nécessaire de barrer tous ses descendants; la figure 7.4 illustre le cas où on code 4 messages m_1, \dots, m_4 .

La succession des messages est codé par concaténation des mots code; par exemple la suite des 7 messages suivants,

$$m_1 m_1 m_4 m_1 m_2 m_2 m_3$$

sera codée

$$1100010101001$$

Ce codage est sans intérêt par rapport à un codage sur 2 bits si les quatre messages sont équiprobables, c'est-à-dire de probabilité (ou de fréquence d'apparition) 25%. Il devient intéressant si m_1 est plus probable que m_2 qui est lui-même plus probable que m_3 et m_4 ; on voit à cette occasion comment construire

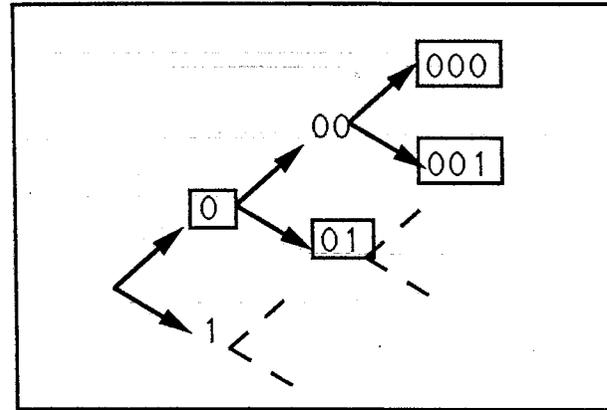


Figure 7.4.

empiriquement un code en utilisant le graphe arborescent et en classant les messages par ordre décroissant des probabilités.

Sur l'exemple ci-dessus la suite codée occupe 13 bits au lieu de 14 en codage de longueur fixe. On verra au paragraphe 7.3.4. que l'on ne peut pas améliorer le rapport

$$\text{longueur moyenne/nombre de messages}$$

au-delà d'une borne inférieure donnée par le théorème de Shannon.

Actuellement les codes en longueur variable sont utilisés pour faire de la compression avant transmission (images envoyées par la sonde américaine Voyager 2 par exemple). Les techniques sont celles employées aux paragraphes 7.3.2. et 7.3.3. ou des raffinements de celles-ci comme l'algorithme FGK (Faller, Gallager et Knuth) qui est un algorithme d'Huffman adaptatif. Pour d'autres méthodes récentes on pourra se reporter à l'article de synthèse « Data compression » de D. A. Lelewer et D. S. Hirschberg-ACM Comp. Surveys 19, n° 3, Sept. 1987, p. 261.

7.3.2. Code de Fano Shannon

Supposons que l'on veuille coder N messages ; si les messages sont numérotés par ordre décroissant de probabilité, c'est-à-dire, m_k est de probabilité p_k avec $p_{k+1} < p_k$, la méthode de Fano Shannon consiste à couper la liste en deux listes de façon à avoir approximativement la même probabilité pour les deux groupes ainsi obtenus. Le premier groupe est affecté du symbole 0, le deuxième du symbole 1 et on recommence à l'intérieur de chacun des groupes jusqu'à épuisement des messages. La chaîne la plus courte correspond à l'élément le plus probable que l'on code avec le plus petit nombre de symboles. Reprenons un exemple avec 4 messages :

$$p_1 = 1/2 \quad p_2 = 1/4 \quad p_3 = 1/8 \quad p_4 = 1/8$$

m_2, m_3, m_4 forment un groupe de messages de probabilité $1/2$. On obtient le code suivant la figure 7.5

m_1	0
m_2	10
m_3	110
m_4	111

La longueur moyenne des mots codes, \bar{n} , est égale à 1,75 (au lieu de 2 en longueur fixe).

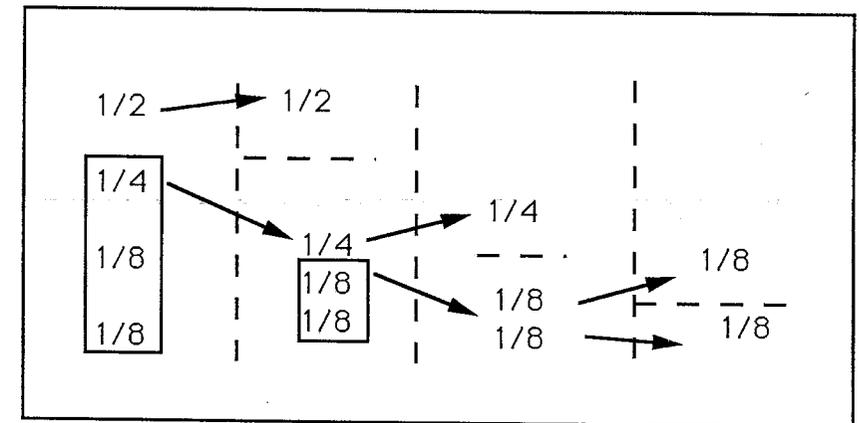


Figure 7.5.

7.3.3. Code d'Huffman

Le code d'Huffman est celui qui rend la longueur moyenne des mots code minimum. Il est basé sur le fait que dans le code optimal, les deux messages les moins probables sont nécessairement codés sur la même longueur (utiliser le graphe arborescent et faire un raisonnement par l'absurde) ; d'autre part le code doit être optimal pour l'ensemble des messages où les deux messages les moins probables ont été « fondus » ensemble et ne représentent plus qu'un seul message de probabilité égale à la somme des deux probabilités, ce qui donne le codage illustré par la figure 7.6.

On suit les flèches et on dispose les chiffres binaires de droite à gauche au fur et à mesure qu'ils se présentent ce qui donne :

m_1	11
m_2	10
m_3	01
m_4	00

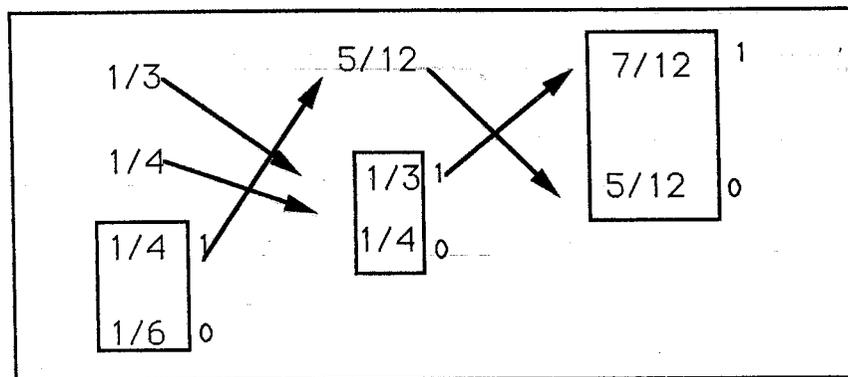


Figure 7.6.

Les messages sont des probabilités suffisamment voisines pour que le codage optimal soit un codage en longueur fixe. Si on change les probabilités en $p_1 = 6/14$ $p_2 = 4/14$ $p_3 = 3/14$ et $p_4 = 1/14$, on trouve le codage

m_1	0
m_2	11
m_3	101
m_4	100

ce qui donne une longueur moyenne \bar{n} égale à $1,86 < 2$.

7.3.4. Théorème de Shannon

La longueur moyenne des mots code est intimement liée à ce que l'on appelle l'entropie de la source de messages ; pour la définir supposons que la source émette les messages successifs de façon indépendante et suivant la même loi de probabilité ; on appelle entropie de la source, l'entropie de la loi de probabilité d'émission des messages, définie ci-dessous, de même pour une variable aléatoire X, on notera H(X) l'entropie de sa loi de probabilité.

Définition : On appelle entropie de la loi de probabilité P : p_1, \dots, p_N , la quantité H(P) définie par

$$H(P) = - \sum_k p_k \log_2 (p_k)$$

On convient de mettre 0 lorsque $p_k = 0$.

L'entropie est une façon de mesurer l'incertitude lors de l'expérience aléatoire ; en effet, posons

$$I(A) = \log_2 (1/P(A))$$

si $P(A) = 1$, il n'y a aucune incertitude sur A et $\log_2 (1/P(A)) = 0$; d'autre part, si B contient A, l'information « B est réalisé » renforce les chances pour que A soit réalisé et on a

$$P(A/B) = P(A \cap B)/P(B) = P(A)/P(B)$$

$$\log_2 (1/P(A/B)) = \log_2 (1/P(A)) - \log_2 (1/P(B))$$

L'incertitude sur A a diminué ; la réalisation de B a amené la quantité d'information $I(B) = \log_2 (1/P(B))$. L'entropie représente alors l'incertitude moyenne lorsqu'on choisit un message suivant la loi de probabilité P. On montre facilement que si X et Y sont deux variables aléatoires et H(X, Y) l'entropie de la loi conjointe, on a

$$H(X, Y) \leq H(X) + H(Y) \text{ (égalité ssi X et Y sont indépendantes)}$$

D'autre part, grâce à l'inégalité $\ln(x) \leq x - 1$, l'entropie d'une loi de probabilité sur un ensemble à N éléments, est toujours inférieure à $\log_2(N)$, la borne supérieure étant atteinte pour la loi uniforme :

$$H(P) \leq \log_2(N)$$

On a alors les résultats suivants sur la longueur moyenne des mots codes d'un code préfixe.

- Théorème :** 1) Pour tout code préfixe, $\bar{n} \geq H(P)$.
 2) Il existe au moins un code préfixe tel que $\bar{n} \leq H(P) + 1$.
 3) (Shannon) la longueur moyenne peut être rendue aussi proche que l'on veut de H(P) à l'aide de codages par blocs de messages.

Le résultat de Shannon sera éclairé par l'application faite au paragraphe 7.3.5. En ce qui concerne les démonstrations on pourra se reporter à « Transmission of information » de Fano, Editions Wiley.

7.3.5. Applications

Nous allons traiter l'exemple d'une source qui débite plus de symboles que ne peut l'accepter un canal, et l'adaptation de la source au canal. En appelant X la variable d'entrée et Y la variable de sortie, on définit $H_X(Y)$ par

$$H_X(Y) = E (H(\text{loi de Y conditionnement à X}))$$

puis on définit la capacité d'un canal par

$$C = \sup_{\text{loi de X}} (H(Y) - H_X(Y))$$

Dans le cas d'un canal binaire symétrique qui échange les symboles avec la probabilité ε ,

$$P(Y = y/X = x) = \varepsilon \text{ si } x \neq y,$$

$$C = 1 + \varepsilon \log_2(\varepsilon) + (1 - \varepsilon) \log_2(1 - \varepsilon).$$

Si la source débite des symboles binaires 0, 1 suivant un débit supérieur à C mais avec un débit d'information inférieur à C, on peut adapter la source au canal ; supposons par exemple que 0 soit de probabilité 80 % et 1 de probabilité 20 % ; si le débit numérique est de 80 symboles par min ; et la capacité du canal de 60 bits/min, on a

$$H(P) = -0,8 \log_2 0,8 - 0,2 \log_2 0,2 = 0,722$$

Taux d'émission d'information de la source = 80 (H(P)) bits/min = 55,76 bits/min < C.

En une minute la source envoie 80 symboles, ce qui correspond à 40 groupes de 2 symboles ; si l'on code les groupes de 2 symboles suivant la méthode d'Huffman, on obtient

$$P(00) = (0,8)^2 = 0,64$$

$$P(01) = P(10) = (0,8)(0,2) = 0,16$$

$$P(11) = (0,2)^2 = 0,04$$

00 est codé 0, 01 est codé 10, 10 est codé 100, 11 est codé 101. La longueur moyenne est égale à 1,56 et $40 \bar{n} = 62,4$; on a donc en moyenne 62,4 bits/min à l'entrée du canal, ce qui n'est pas encore suffisant. On code par groupe de 3 symboles et on trouve $\bar{n} = 2,174$ ce qui donne en moyenne 57,92 bits/min à l'entrée du canal. Avec le codage par groupe de 3, on s'est rapproché de la borne inférieure annoncée par Shannon :

$$\frac{1,56}{2} > \frac{2,17}{3} > H(P) = 0,5...$$

7.4. DÉTECTION ET CORRECTION DES ERREURS

Lorsqu'un mot binaire de longueur n est envoyé dans un canal de transmission, on reçoit à la sortie le mot envoyé auquel s'est ajouté éventuellement un mot erreur e . On peut écrire ce mot e sous la forme

$$e = X_1 X_2 \dots X_n$$

la variable X_i étant égale à 1 s'il y a erreur sur le i -ième bit. Lorsque les chiffres binaires sont envoyés les uns derrière les autres, et lorsque le canal transforme les 1 en 0 et les 0 en 1 dans une proportion p (canal binaire symétrique), on obtient

$$d(e, 0) = X_1 + X_2 + \dots + X_n$$

où les v.a. X_i sont indépendantes, de même loi de Bernoulli. $d(e, 0)$ est donc une v.a. $\mathcal{B}(n, p)$ qui a pour valeur moyenne np . Une inégalité due à Chernoff permet d'évaluer la probabilité pour que $d(e, 0)$ dépasse $n\alpha$ si $\alpha > p$:

$$P(d(e, 0) > n\alpha) < 2^{-n\varphi(\alpha, p)}$$

avec

$$\varphi(\alpha, p) = \alpha \log_2 \alpha + (1 - \alpha) \log_2(1 - \alpha) - \alpha \log_2 p - (1 - \alpha) \log_2(1 - p)$$

On remarque que l'on a $\varphi(\alpha, p) \geq 0$ pour $\frac{1}{2} > \alpha > p > 0$.

Avec cette inégalité, le taux d'erreur tend vers 0 lorsque n tend vers l'infini.

De façon plus générale, un résultat de Shannon montre que l'on peut coder un grand nombre de messages avec un taux d'erreur faible ; plus précisément :

Théorème : Quel que soit R inférieur strictement à la capacité C du canal de transmission (R représente un taux de transmission) ; quel que soit $\varepsilon > 0$, il existe n_0 tel que pour tout $n \geq n_0$, il existe un code de $N =$ partie entière de 2^{nR} mots de longueur n et de probabilité d'erreur inférieure à ε .

EXERCICES

E.7.1. Quantification : Soit θ un angle aléatoire, de loi uniforme sur $\left[0, \frac{\pi}{2}\right]$ et $X = \sin \theta$. Calculer la loi de X. On effectue une quantification sur 4 niveaux puis un codage binaire sur 2 bits. Donner les niveaux à choisir pour avoir 4 mots équiprobables. Calculer $E[(X - \tilde{X})^2]$ (variance du « bruit de quantification ») où \tilde{X} est le milieu de chaque intervalle de quantification.

E.7.2. Soit a_k une suite de symboles égaux à ± 1 , que l'on suppose aléatoires, indépendants, de même loi uniforme (+1 et -1 sont équiprobables). Soit $h(t)$ une fonction intégrable, de durée finie inférieure à T, et

$$X_t = \sum_k a_k h(t - kT)$$

Calculer

$$m(t) = E(X_t)$$

$$K(t, s) = \text{cov}(X_t, X_s)$$

Montrer que l'on a

$$K(t + nT, s + nT) = K(t, s)$$

$$\text{soit } \Gamma(t, s) = \frac{1}{T} \int_0^T K(t + \tau, s + \tau) d\tau$$

Montrer que Γ ne dépend que de $t - s$.

E.7.3. On reprend les notations de l'exercice 7.2.

Soit θ une variable aléatoire de loi uniforme sur $[0, T]$ indépendante des symboles numériques émis et

$$Y_t = \sum_k a_k h(t - kT - \theta)$$

Montrer que Y_t est stationnaire de fonction de covariance $\Gamma(t, s) = \gamma(t - s)$.

Calculer la densité spectrale de puissance dans le cas du code NRZ, la mesure spectrale de puissance dans le cas du code RZ.

E.7.4. On reprend les notations des exercices 7.2 et 7.3. Montrer que si les symboles $a_k = \pm 1$ sont corrélés de façon stationnaire, c'est-à-dire $\text{cov}(a_{n+k}, a_n) = c(k)$ alors

$$E[Y_{t+\tau} \bar{Y}_t] = \sum_k c(k) \frac{1}{T} \int_{\mathbb{R}} h(s + \tau - kT) \bar{h}(s) ds$$

E.7.5. Codeur rotatif de position : on veut coder un angle θ appartenant à $[0, 2\pi]$ sur k bits. Pour simplifier nous étudions le cas où l'angle est codé sur 3 bits (au moins 6 dans la réalité). Dans ce dispositif des photodiodes sont éclairées par une raie lumineuse au travers d'un disque et on déplace la barrette de lecture pour avoir le codage du secteur angulaire comme illustré par la figure E.7.5.

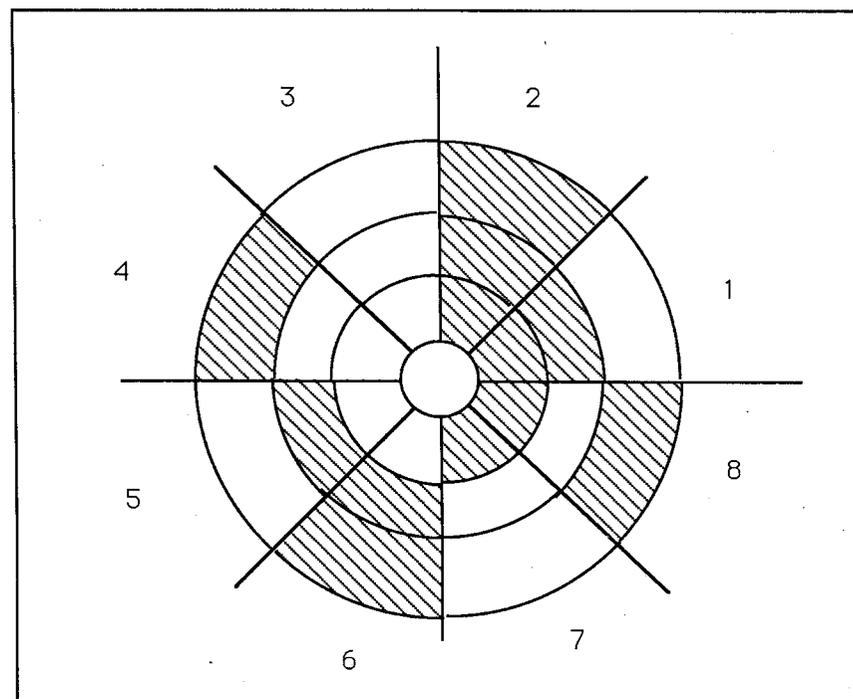


Figure E.7.5.

Le secteur 1 est codé 110, le secteur 2 111, ...

Etudier les erreurs de codages lorsqu'on est proche d'une frontière entre deux secteurs : dans ce cas les diodes ne sont ni franchement éclairées, ni franchement dans l'obscurité. Quel angle peut-on restituer si la position initiale est proche de $\frac{\pi}{4}$? (Dans la réalité on utilise un disque tel qu'à chaque changement de secteur, un seul bit change.)

E.7.6. Soit $g(X) = X^{11} + X^3 + 1$. En utilisant

$$X^{14} = X^3 + 1 \text{ modulo } g(X)$$

Calculer X^5 modulo $g(X)$, X^6 modulo $g(X)$, ..., X^{11} modulo $g(X)$. $g(X)$ est-il un diviseur de $X^{11} + 1$?

E.7.7. Si $n = 7$ et $k = 4$, quels sont les polynômes $g(X)$ à essayer si on cherche tous les codes cycliques $(7, 4)$?

E.7.8. Soit (X, Y) un couple de v.a. à valeurs dans un ensemble fini avec $P_{ij} = P(X = x_i, Y = y_j)$

On pose $H(X, Y) = - \sum_{i,j} p_{ij} \log_2 p_{ij}$

et $H(X/Y) = \sum_j P(Y = y_j) H(X/Y = y_j)$

où $H(X/Y = y_j)$ désigne l'entropie de la loi conditionnelle de X sachant $Y = y_j$.

a) Montrer que l'on a

$$H(X, Y) = H(Y) + H(X/Y)$$

b) Montrer que si X et Y sont indépendantes

$$H(X, Y) = H(X) + H(Y)$$

c) Transmission $Y = f(X)$. Montrer que l'on a

$$H(Y/X) = 0 \text{ et } H(Y) = H(X) = H(X, Y)$$

E.7.9. On considère 8 messages de probabilité respectives

$$\frac{10}{27}, \frac{6}{27}, \frac{4}{27}, \frac{3}{27}, \frac{1}{27}, \frac{1}{27}, \frac{1}{27}, \frac{1}{27}$$

Utiliser la méthode de Fano-Shannon, puis celle d'Huffman pour coder et calculer la longueur moyenne des mots codes.

E.7.10. Calculer la capacité du canal, non symétrique, qui transforme 1 en 0, dans 1% des cas et 0 en 1 dans 2% des cas.

E.7.11. On considère une source markovienne à 2 états qui génère trois symboles A, B, C suivant le schéma de la figure E.7.11.

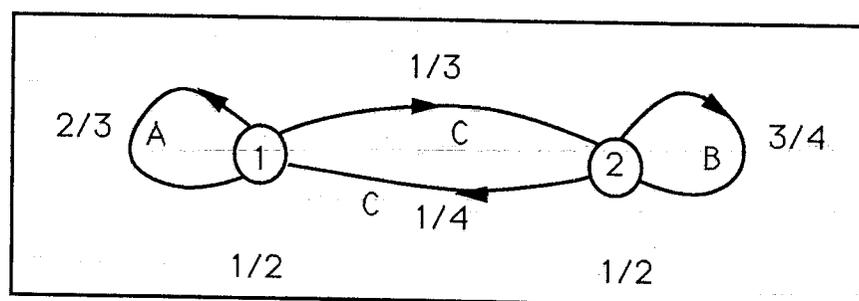


Figure E.7.11.

Calculer l'entropie de l'état 1, l'entropie de l'état 2. On définit l'entropie de la source comme la moyenne des entropies des deux états. Calculer l'entropie de la source. Déterminer les messages de 1 lettre avec leur probabilité, les messages de 2 lettres avec leur probabilité. En déduire l'information moyenne par lettre des messages de 1 lettre, des messages de 2 lettres. Comparer ces valeurs entre elles et avec l'entropie.

Quel est le codage binaire optimal pour les messages de 2 lettres ? Calculer la longueur moyenne ainsi obtenue.

E.7.12. A lance un dé n fois et donne à B l'information suivante : « la moyenne des chiffres obtenus est 4,5 » B voudrait reconstituer l'expérience, c'est-à-dire connaître n_1 , nombre d'apparitions du 1, n_2, \dots, n_6 . On pose $f_i = \frac{n_i}{n}$ et on suppose n, n_1, \dots, n_6 grands. Montrer que l'on a

$$\frac{1}{n} \log_2 P(n_1 \text{ fois } 1, \dots, n_6 \text{ fois } 6) \approx H(f_1, f_2, \dots, f_6) - \log_2(6)$$

En déduire la configuration (n_1, n_2, \dots, n_6) qui rend $P(n_1 \text{ fois } 1, \dots, n_6 \text{ fois } 6)$ maximum sous les contraintes $\sum_i \frac{n_i}{n} = 4,5$ et $\sum n_i = n$.

(On utilisera deux multiplicateurs de Lagrange et on terminera par une résolution graphique).

CHAPITRE 8

STATISTIQUES

8.1. INTRODUCTION

8.1.1. Les domaines d'utilisation

Les méthodes statistiques trouvent des applications dans des domaines aussi variés que l'économie, la biologie, la médecine et la physique ; on peut chercher un plan de contrôle de qualité pour une production, tester la présence d'un signal (détection radar), comparer l'efficacité de plusieurs médicaments, chercher une combinaison de plusieurs engrais qui assure la meilleure productivité, estimer la densité spectrale de puissance d'un signal aléatoire...

Un exemple d'utilisation de méthodes statistiques en théorie du signal est l'estimation de la moyenne d'un processus stationnaire au 2^e ordre ; cet exemple, en apparence simple, montre qu'une étude fine demande déjà quelques développements théoriques ; le lecteur peu soucieux de démonstrations mathématiques pourra se reporter directement au paragraphe 8.1.4 en ayant en tête l'exemple, plus simple, où on cherche à mesurer une grandeur physique et où les mesures sont entachées d'erreurs (voir l'exercice 8.2).

8.1.2. Exemple à temps discret

Soit X_n un signal aléatoire stationnaire.

Supposons X_n de moyenne m , de fonction de covariance k . En décomposant X_n suivant m et le processus centré $X_n - m$, on peut écrire

$$E(X_n X_p) = m^2 + \text{cov}(X_n, X_p)$$

Il suffit d'étudier la partie purement aléatoire de la puissance. On appelle $d\mu(\lambda)$ la mesure spectrale de puissance du processus centré $X_n - m$:

$$\text{cov}(X_{n+m}, X_n) = k(n) = \int_0^1 e^{2i\pi\lambda n} d\mu(\lambda)$$

Si m est inconnu, on cherche à l'estimer grâce à une expérience.

On effectue des observations X_1, X_2, \dots, X_n et on forme la moyenne empirique \bar{X} ,

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$E(\bar{X}) = m$$

\bar{X} est centrée sur le paramètre inconnu que l'on cherche à estimer m . On dit alors que \bar{X} est un estimateur sans biais de m . Cet estimateur est d'autant meilleur qu'il varie peu autour de m ; sa variance est une mesure de ses variations

$$\begin{aligned} \text{var}(\bar{X}) &= E(|\bar{X} - m|^2) \\ &= E\left[\frac{1}{n^2} \sum_{i,j} (X_i - m)(X_j - m)^*\right] \\ &= \frac{1}{n^2} \sum_{i,j} \text{cov}(X_i, X_j) \\ &= \int_0^1 \frac{1}{n^2} \sum_{i,j} e^{2i\pi(\ell-j)\lambda} d\mu(\lambda) \\ &= \int_0^1 \left| \frac{1}{n} \sum_{\ell} e^{2i\pi\ell\lambda} \right|^2 d\mu(\lambda) \\ &= \int_{]0,1[} \left| \frac{1}{n} \sum_{\ell} e^{2i\pi\ell\lambda} \right|^2 d\mu(\lambda) + \mu(\{0\}) \end{aligned}$$

Si la mesure spectrale a une masse en 0, \bar{X} ne converge pas en moyenne quadratique vers m car sa variance est supérieure ou égale à $\mu(\{0\})$. C'est le cas si X_n contient une variable invisible, c'est-à-dire si

$$X_n = Y_n + A$$

avec A une variable centrée indépendante du processus Y_n ; la variable A est impossible à détecter sur l'espérance.

Supposons maintenant que la mesure spectrale n'ait pas de masse en 0, ce qui est le cas lorsqu'on a une densité spectrale, par exemple. On peut intégrer sur l'intervalle $] -0,5, 0,5[$ au lieu de l'intervalle $]0, 1[$. On découpe l'intervalle en deux parties : $[-\eta, \eta]$ et le reste

$$\mu([- \eta, \eta]) \xrightarrow{\eta \rightarrow 0} \mu(\{0\}) = 0$$

L'intégrale sur $[-\eta, \eta]$ est majorée par $\mu([- \eta, \eta])$ car la moyenne des exponentielles est majorée en module par 1. On choisit η de façon que cette masse soit inférieure à un ε que l'on s'est fixé à l'avance.

Pour l'intégrale sur $|\lambda| > \eta$, on utilise le calcul bien connu de somme d'exponentielles qui donne ici,

$$\begin{aligned} \left| \sum_{k=1}^n e^{2i\pi k\lambda} \right| &= \left| \frac{\sin(n\pi\lambda)}{\sin(\pi\lambda)} \right| \leq \frac{1}{\sin(\pi\eta)} \\ \int_0^1 \left| \frac{1}{n} \sum_{k=1}^n e^{2i\pi k\lambda} \right|^2 d\mu(\lambda) &\leq \varepsilon + \frac{1}{n^2 \sin^2(\pi\eta)} \end{aligned}$$

Lorsque n est suffisamment grand cette quantité est inférieure à 2ε et

$$\text{var}(\bar{X}) \leq 2\varepsilon$$

On admettra la proposition suivante qui est plus complète.

Proposition : \bar{X} converge en moyenne quadratique vers m si et seulement si la mesure spectrale de puissance du processus centré ne charge pas l'origine. Si le processus centré admet une densité spectrale continue à l'origine alors \bar{X} converge aussi presque sûrement et dans le cas gaussien,

$$\sqrt{n}(\bar{X} - m) \xrightarrow[n \rightarrow +\infty]{\text{en loi}} N(0, f(0))$$

8.1.3. A temps continu

On observe le processus durant un intervalle de temps fini que l'on suppose égal à $[-T, T]$ par commodité d'écriture. On forme alors la moyenne empirique

$$\bar{X} = \frac{1}{2T} \int_{-T}^T X_t dt$$

\bar{X} est centrée sur m et \bar{X} est obtenue par convolution avec la fenêtre rectangulaire

$$\frac{1}{2T} 1_{[-T, T]}$$

$$\text{var}(\bar{X}) = \int_{-\infty}^{\infty} |H(\lambda)|^2 d\mu(\lambda) = \int_{-\infty}^{\infty} \left[\frac{\sin(2\pi\lambda T)}{2\pi\lambda T} \right]^2 d\mu(\lambda)$$

Par le théorème de convergence dominée, on montre que $\text{var}(\bar{X})$ converge vers $\mu(\{0\})$ lorsque T tend vers l'infini. On admettra que la proposition énoncée à temps discret reste vraie à temps continu.

8.1.4. Premières remarques

Sur cet exemple sont apparus différents points importants qui interviendront dans toute situation statistique : le modèle, le problème posé dans le cadre du modèle choisi et enfin les méthodes statistiques pour traiter le problème.

LE MODÈLE

Choix des lois des variables qui interviennent dans le problème : indépendance ou non, loi de type gaussienne ou non...

Modèle paramétrique : on choisit une loi qui dépend paramétriquement d'un (ou plusieurs) paramètre inconnu.

Modèle non paramétrique : la loi dépend des inconnues de façon non paramétrique.

LES DIFFÉRENTS TYPES DE PROBLÈMES

Estimation d'un ou plusieurs paramètres

Test sur un paramètre

Test d'ajustement sur une loi donnée

Recherche d'un intervalle de confiance

Choix d'un plan d'expérience

Choix d'un plan de contrôle de qualité

Comparaison de plusieurs échantillons provenant de populations différentes

Test de l'indépendance de deux variables

8.2. L'ESTIMATION

8.2.1. Vocabulaire statistique

Chercher un estimateur d'un paramètre réel comme la moyenne d'un processus stationnaire, c'est chercher à associer une valeur à chaque observation. Plus généralement, un estimateur est une fonction :

variables observées \mapsto une valeur possible du paramètre

Il s'agit d'une règle de décision : par exemple, au vu de l'observation de $X_1(\omega)$, $X_2(\omega)$... je décide de la valeur de l'espérance (moyenne théorique des variables). Il y a alors deux points de vue : on peut choisir une règle de décision (connaissance antérieure du phénomène ou considérations intuitives) et évaluer les erreurs possibles, ou bien, au contraire, se donner un critère et chercher une règle de décision qui satisfasse ce critère.

8.2.2. Le critère du maximum de vraisemblance

On observe les variables aléatoires X_1, X_2, \dots, X_n de densité conjointe $f_\theta(x_1, x_2, \dots, x_n)$ où θ est un paramètre à estimer ; ce paramètre peut être unidimensionnel ou non : moyenne inconnue et variance connue, moyenne

connue et variance inconnue, moyenne et variance inconnues en sont quelques exemples ; le n -uple observé (x_1, x_2, \dots, x_n) est d'autant plus probable que la densité est élevée en ce point. Le critère du maximum de vraisemblance consiste donc à chercher la (ou les) valeur du paramètre qui rend la densité maximum :

$$\text{Sup}_\theta f_\theta(x_1, x_2, \dots, x_n)$$

Généralement on travaille avec un n -échantillon, c'est-à-dire n réalisations indépendantes d'une même loi de probabilité.

Dans le cas d'un échantillon gaussien, l'estimateur du maximum de vraisemblance pour l'espérance est \bar{X} . On rencontre cette situation lorsqu'on fait des mesures entachées d'erreurs ; si on répète l'expérience dans les mêmes conditions, les mesures x_i peuvent être considérées comme la somme de la quantité à mesurer θ et d'une variable d'erreur ε_i ; dans un modèle où les erreurs sont accidentelles, on suppose généralement les variables ε_i indépendantes de loi $\mathcal{N}(0, \sigma^2)$, σ^2 représentant la dispersion des mesures. La figure 8.1 illustre le cas où à la suite d'une expérience on a récolté les résultats de mesure 1,2, 2,1, -0,5, 0,8, 1,7, 0,9, 1,4, 1,6, 1,1 ; on cherche le paramètre $\theta =$ moyenne théorique, au vu des valeurs, en supposant les variables de variance 1. La densité $f_\theta(x_1, x_2, \dots, x_n)$ s'écrit

$$f_\theta(x_1, x_2, \dots, x_n) = \text{Cte} \exp\left(-\frac{1}{2} \sum x_i^2\right) \exp(\theta \sum x_i - n\theta^2/2)$$

On regarde cette expression en fixant les valeurs des variables x et on cherche le maximum sur le paramètre θ , maximum qui a lieu pour $\theta =$ moyenne empirique des variables.

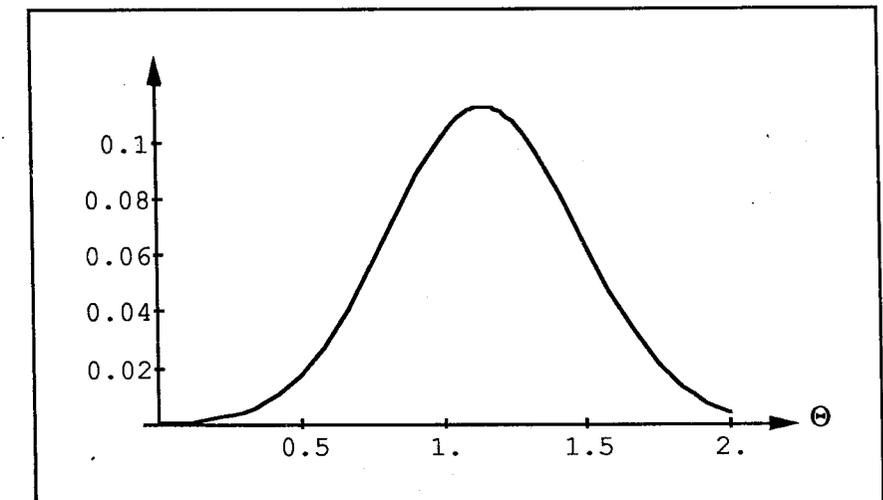


Figure 8.1.

8.2.3. Le critère de Bayes

Il est basé sur le fait que l'on a une idée à priori sur le paramètre : domaine où il se trouve, valeurs plus probables que d'autres...

Concrètement on se donne une probabilité appelée *loi a priori*, pour le paramètre.

Une façon de mesurer l'erreur que l'on fait sur le paramètre (à valeurs réelles) lorsqu'on prend l'estimateur $d(x_1, x_2, \dots, x_n)$ est de calculer l'écart quadratique moyen :

$$E \{d(X_1, X_2, \dots, X_n) - \theta\}^2$$

Cette moyenne doit être comprise comme une moyenne sur les valeurs x_i que peuvent prendre les variables, mais aussi sur les valeurs possibles du paramètre ; un calcul de moyenne étant un calcul de centre de gravité, on introduit une pondération qui tient compte du fait que telle ou telle valeur du paramètre est plus probable que d'autres (ou bien que telle ou telle erreur soit plus grave que d'autres) : supposons qu'il n'y ait que 3 valeurs possibles pour le paramètre ; on se donne une probabilité à priori sur ces 3 valeurs p_1, p_2, p_3 et on calcule l'erreur quadratique moyenne

$$\sum p_i E_{\theta_i} \{d(X_1, X_2, \dots, X_n) - \theta\}^2$$

Le critère de Bayes consiste à choisir l'estimateur qui donne la plus petite erreur moyenne. On montre que cet estimateur est l'espérance conditionnelle du paramètre, conditionnellement à (X_1, X_2, \dots, X_n)

$$d(X_1, X_2, \dots, X_n) = E \{ \theta / X_1, X_2, \dots, X_n \}$$

Exemple : Reprenons le cas de variables gaussiennes de variance 1 ; supposons que la moyenne m soit égale soit à 0 soit à 1 soit à 2 ; la valeur $m = 0$ étant plus probable que les autres et des considérations antérieures ont amené le statisticien à supposer $m = 0$ de proba 1/2, $m = 1$ de probabilité 1/4 ainsi que $m = 2$. En utilisant les résultats sur les lois conditionnelles, on calcule $P(m = 0 / X_1, \dots, X_n)$, $P(m = 1 / X_1, \dots, X_n)$ et $P(m = 2 / X_1, \dots, X_n)$:

$$\begin{aligned} P(m = 0 / X_1 = x_1, \dots, X_n = x_n) &= \\ &= \frac{\frac{1}{2} f_0(x_1, \dots, x_n)}{\frac{1}{2} f_0(x_1, \dots, x_n) + \frac{1}{4} f_1(x_1, \dots, x_n) + \frac{1}{4} f_2(x_1, \dots, x_n)} \end{aligned}$$

$$\begin{aligned} P(m = 1 / X_1 = x_1, \dots, X_n = x_n) &= \\ &= \frac{\frac{1}{4} f_1(x_1, \dots, x_n)}{\frac{1}{2} f_0(x_1, \dots, x_n) + \frac{1}{4} f_1(x_1, \dots, x_n) + \frac{1}{4} f_2(x_1, \dots, x_n)} \end{aligned}$$

$$\begin{aligned} P(m = 2 / X_1 = x_1, \dots, X_n = x_n) &= \\ &= \frac{\frac{1}{4} f_2(x_1, \dots, x_n)}{\frac{1}{2} f_0(x_1, \dots, x_n) + \frac{1}{4} f_1(x_1, \dots, x_n) + \frac{1}{4} f_2(x_1, \dots, x_n)} \end{aligned}$$

$$\begin{aligned} E(m / X_1 = x_1, \dots, X_n = x_n) &= P(m = 1 / X_1 = x_1, \dots, X_n = x_n) + \\ &+ 2 P(m = 2 / X_1 = x_1, \dots, X_n = x_n) = \frac{1 + 2 e^{-\frac{n}{2}} e^{n\bar{x}}}{2 + e^{-\frac{n}{2}} e^{n\bar{x}} + e^{-2n} e^{2n\bar{x}}} \end{aligned}$$

On peut évidemment regarder un autre type d'erreur que l'erreur quadratique ; on se donne une fonction de coût, coût de décider $d(x_1, x_2, \dots)$ lorsque la vraie valeur est θ puis on minimise le coût moyen. De nouveau l'estimateur est basé sur la loi du paramètre conditionnellement à l'observation, loi que l'on appelle *loi a posteriori*.

8.2.4. Comparaison de ces deux critères

Ils ont tous les deux l'avantage de donner une méthode de construction d'un estimateur.

Si l'ensemble des valeurs du paramètre est un ensemble borné, une loi à priori particulière est la loi uniforme ; on constate alors que chercher l'estimateur du maximum de vraisemblance revient à chercher le maximum de la densité à posteriori :

$$P(\theta / X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{f_\theta(x_1, x_2, \dots, x_n) p(\theta)}{\sum_k f_{\alpha_k}(x_1, x_2, \dots, x_n) p(\alpha_k)} \quad (\text{cas discret})$$

$$f(\theta / X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \frac{f_\theta(x_1, x_2, \dots, x_n) p(\theta)}{\int_A^B f_\alpha(x_1, x_2, \dots, x_n) p(\alpha) d\alpha}$$

Si cette densité à posteriori est unimodale et symétrique par rapport à son mode (valeur où a lieu la maximum), comme dans le cas gaussien illustré par la figure 8.1, l'estimateur du maximum de vraisemblance est égal à la moyenne de la loi à posteriori, c'est-à-dire à l'estimateur bayésien.

La méthode du maximum de vraisemblance peut poser quelques problèmes : le maximum n'est pas atteint ou bien n'est pas unique, l'estimateur est biaisé. Prenons l'exemple d'une loi de densité $f(x - \theta)$ avec f densité de Cauchy ; pour un échantillon de taille 2, X et Y , on est amené à chercher le maximum de $f(x - \theta) f(y - \theta)$ où, ce qui est équivalent le minimum de $\text{Ln}[1 + (x - \theta)^2] + \text{Ln}[1 + (y - \theta)^2]$; le graphe de cette fonction est donnée sur la figure 8.2 lorsque $x = -2$ et $y = 2$.

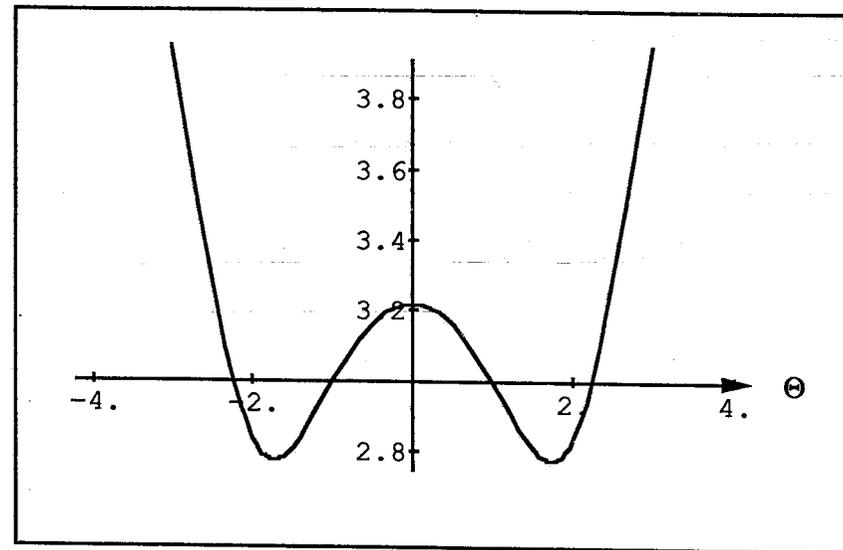


Figure 8.2.

On constate que le minimum n'est pas unique : il y a deux valeurs du paramètre qui répondent au critère.

L'estimateur du maximum de vraisemblance a l'avantage d'avoir de bonnes propriétés de convergence ; sous certaines hypothèses (que l'on trouvera par exemple dans le livre de Dacunha-Castelle et Duflo), il converge presque sûrement vers le paramètre à estimer. C'est le cas dans les exemples courants où on travaille sur un échantillon gaussien...

8.2.5. Estimateur de variance minimale

Etant donné un paramètre θ ou plus généralement une fonction $g(\theta)$ à estimer, on montre que si Y est un estimateur sans biais de $g(\theta)$ alors sa variance est minorée par une borne, appelée *borne de Cramer-Rao* :

$$\text{var}_\theta(Y) \geq \frac{[g'(\theta)]^2}{I(\theta)}$$

où $I(\theta)$ est l'information de Fisher,

$$I(\theta) = E_\theta \left[\frac{\frac{\partial}{\partial \theta} L_\theta}{L_\theta} \right]^2$$

$L_\theta(y)$ étant la densité de Y lorsque le paramètre vaut θ . Le calcul de l'espérance est fait à θ fixé. On remarquera qu'il s'agit d'une dérivée logarithmique. Dans le

cas d'une variable $\mathcal{N}(m, 1)$ avec le paramètre m inconnu, on trouve comme dérivée logarithmique $Y - m$; le calcul de l'information est donc celui de la variance de Y , c'est-à-dire 1 :

$$I(m) = 1 \text{ pour tout } m$$

Si Y avait pour loi $\mathcal{N}(m, \sigma^2)$ avec σ^2 connu, l'information serait encore constante et égale à la variance σ^2 .

On démontre cette inégalité sous un certain nombre d'hypothèses : L_θ est supposée dérivable par rapport à θ , $I(\theta)$ est finie... (voir le livre de Dacunha Castelle et Duflo).

On dit que l'estimateur est *efficace* lorsqu'il atteint la borne de Cramer-Rao.

On montre facilement que si l'on effectue n observations indépendantes de Y , l'information de Fisher est multipliée par n ; la borne de Cramer-Rao est alors divisée par n . Si on augmente le nombre d'observations, on peut s'attendre, sauf cas pathologique, à ce que le paramètre soit mieux estimé.

Ce résultat se généralise au cas où l'on cherche à estimer plusieurs paramètres, c'est-à-dire lorsque θ est un vecteur et $g(\theta)$ à valeurs réelles ; on définit alors la matrice d'information de Fisher :

$$I_{i,j}(\theta) = E_\theta \left[\left(\frac{\partial}{\partial \theta_i} \text{Log } L_\theta \right) \left(\frac{\partial}{\partial \theta_j} \text{Log } L_\theta \right) \right]$$

et on a

$$\text{var}_\theta(Y) \geq \text{'grad } g(\theta) I(\theta)^{-1} \text{grad } g(\theta)$$

8.3. TESTS PARAMÉTRIQUES

8.3.1. Vocabulaire et définitions

Dans un test d'ajustement sur une loi $\mathcal{N}(0, 1)$, c'est-à-dire lorsqu'on se demande si les variables observées suivent une loi $\mathcal{N}(0, 1)$ ou non, l'hypothèse « la loi n'est pas celle d'une $\mathcal{N}(0, 1)$ » n'est pas paramétrisable. Dans ce paragraphe, on s'intéresse aux hypothèses qui peuvent être paramétrées par un ou plusieurs paramètres réels.

On dit que l'on fait un test si l'ensemble des valeurs du paramètre est partagé en deux, chaque partie correspondant à une hypothèse. Les deux hypothèses seront appelées H_0 et H_1 . Nous verrons néanmoins le cas d'hypothèses multiples au paragraphe 8.3.

Exemple de test sur un paramètre :

$$H_0 = \text{« } X \text{ suit une loi } \mathcal{N}(0, 1) \text{ »}$$

$$H_1 = \text{« } X \text{ suit une loi } \mathcal{N}(m, 1) \text{ avec } m > 0 \text{ »}$$

$$(x_1, x_2, \dots, x_n) \mapsto \text{décision} = d(x_1, x_2, \dots, x_n)$$

$$d(x_1, \dots, x_n) = H_0 \text{ si et seulement si } \bar{x} \leq 0,5$$

A chaque observation, on associe une décision : « c'est l'hypothèse H_0 qui est vérifiée » ou bien « c'est l'hypothèse H_1 qui est vérifiée ». Il s'agit alors d'une règle de décision déterministe ; en effet on peut aussi effectuer un tirage au sort (non pas à pile ou face mais qui dépend de l'observation) pour prendre la décision.

Pour une règle de décision déterministe, le test est caractérisé par sa *région de rejet* D :

$$D = \{ \text{valeurs de l'observation pour lesquelles on rejette l'hypothèse } H_0 \}$$

Il y a deux types d'erreurs, l'erreur de première espèce et l'erreur de deuxième espèce :

Décision	Etat	H_0	H_1
H_0			Erreur de décision
H_1		Erreur de décision	

Erreur de 1^{re} espèce (niveau du test).

On dit qu'une hypothèse est simple si, sous cette hypothèse, le paramètre est réduit à un point. On note α l'erreur de 1^{re} espèce, définie par :

$$\alpha = P_{H_0} \text{ (on décide } H_1 \text{)}$$

Dans le cas où l'hypothèse H_0 n'est pas simple, α est définie comme le sup sur les valeurs du paramètre appartenant à H_0

$$\alpha = \text{Sup}_{\theta \in H_0} P_{\theta} \text{ (on décide } H_1 \text{)}$$

Erreur de 2^e espèce :

Si H_1 est vraie et que l'on décide H_0 . La probabilité d'erreur est $\beta(\theta) = P_{\theta} \{D^c\}$ pour chaque θ de H_1 . On regarde généralement la fonction *puissance* du test :

$$\begin{aligned} \alpha(\theta) &= P_{\theta} \{D\} \quad \text{pour } \theta \text{ dans } H_1 \\ &= 1 - \beta(\theta) \end{aligned}$$

Cette fonction est appelée fonction puissance du test, lorsque θ varie dans H_1 .

Exemple : Le service comptabilité d'un grand magasin veut savoir si la proportion de chèques sans provision est supérieure à 5 % ; il s'agit d'une proportion théorique, qui représenterait une moyenne prise sur tous les clients potentiels du magasin. Si 100 dossiers de paiement par chèque sont examinés, l'observation est une variable de loi $\mathcal{B}(100, p)$ et on teste l'hypothèse $H_0 : \langle p = 5 \% \rangle$ contre $H_1 \langle p > 5 \% \rangle$.

Regardons un exemple de taille plus faible ; l'observation est une variable X qui suit une loi $\mathcal{B}(10, p)$ avec p inconnu ; on veut tester $H_0 : \langle p = 0,2 \rangle$ contre $H_1 : \langle p = 0,5 \rangle$. Si on examine la loi pour différentes valeurs de p , on constate que la loi « déplace son poids » vers la droite si on augmente p , comme on peut le voir sur le graphe qui suit.

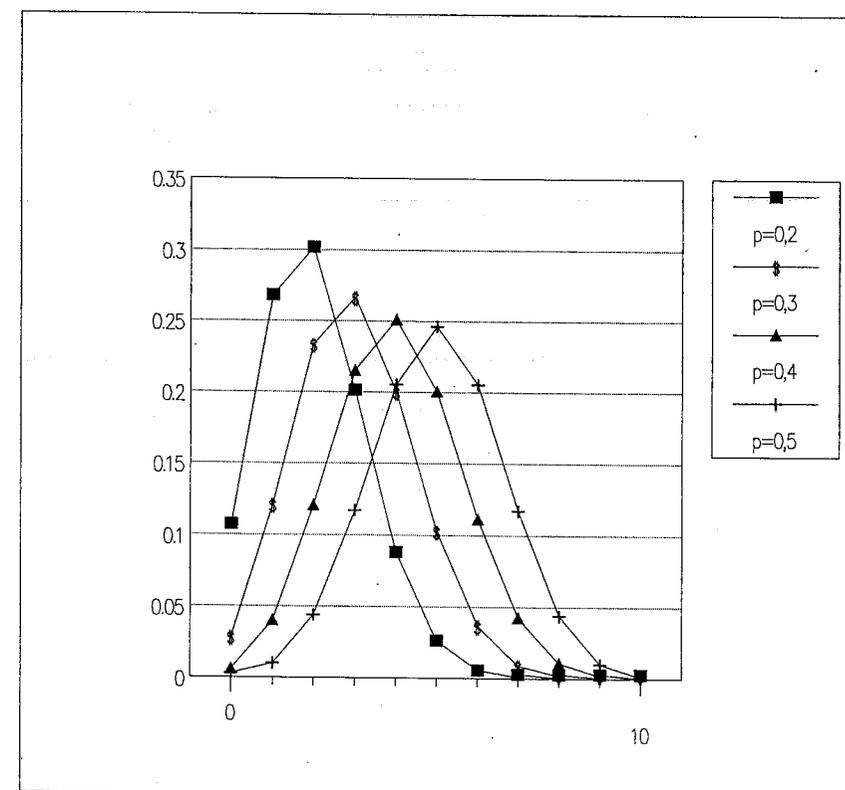


Figure 8.3.

Le tableau des probabilités permet de faire les calculs d'erreurs de première et de deuxième espèce (voir tableau page 160).

Si on choisit comme région de rejet l'ensemble $\{X > 4\}$, on obtient $\alpha = 3 \%$ et $\beta = 37,7 \%$.

p	k	$p(X = k)$	$\sum_{j=0}^k p(X = j)$
0,2	0	0,107374	0,107374
	1	0,268435	0,37581
	2	0,30199	0,6778
	3	0,201327	0,879126
	4	0,08808	0,967207
	5	0,026424	0,993631
	6	0,005505	0,999136
	7	0,000786	0,999922
	8	7,37E-05	0,999996
	9	4,1E-06	1
0,3	10	1,02E-07	1
	0	0,028248	0,028248
	1	0,121061	0,149308
	2	0,233474	0,382783
	3	0,266828	0,649611
	4	0,200121	0,849732
	5	0,102919	0,952651
	6	0,036757	0,989408
	7	0,009002	0,99841
	8	0,001447	0,999856
9	0,000138	0,999994	
0,4	10	5,9E-06	1
	0	0,006047	0,006047
	1	0,040311	0,046357
	2	0,120932	0,16729
	3	0,214991	0,382281
	4	0,250823	0,633103
	5	0,200658	0,833761
	6	0,111477	0,945238
	7	0,042467	0,987705
	8	0,010617	0,998322
9	0,001573	0,999895	
0,5	10	0,000105	1
	0	0,000977	0,000977
	1	0,009766	0,010742
	2	0,043945	0,054688
	3	0,117188	0,171875
	4	0,205078	0,376953
	5	0,246094	0,623047
	6	0,205078	0,828125
	7	0,117188	0,945313
	8	0,043945	0,989258
9	0,009766	0,999023	
10	0,000977	1	

8.3.2. Le critère de Neyman-Pearson

On se fixe un niveau et on cherche, s'il existe, le test le plus puissant parmi les tests de niveau inférieur ou égal à ce niveau donné α .

Lemme de Neyman-Pearson : Si les deux hypothèses sont simples, pour chaque niveau α , il existe un test optimal ; ce test est basé sur le rapport de vraisemblance et il est de niveau α exactement ; dans le cas de v.a. à densité, c'est le test déterministe défini par la région de rejet D (rejet de H_0) :

$$\begin{cases} D = \left\{ \frac{f_1}{f_0} \geq A \right\} \\ P_{H_0} \left(\frac{f_1}{f_0} \geq A \right) = \alpha \end{cases}$$

où f_1 est la densité sous l'hypothèse H_1 , f_0 la densité sous H_0 .

Dans la pratique, connaissant la forme du test, on cherche A de façon à avoir la deuxième relation.

Démonstration dans le cas de densités : Soit D' la région de rejet d'un test de niveau inférieur ou égal à α , c'est-à-dire

$$\int_{D'} f_0(u) du \leq \alpha$$

On a alors l'inégalité

$$[1_D(u) - 1_{D'}(u)] \cdot [f_1(u) - f_0(u)] \geq 0$$

et on en déduit

$$\int [1_D(u) - 1_{D'}(u)] f_1(u) du \geq 0$$

c'est-à-dire

$$\int_{D^c} f_1(u) du \leq \int_{D^c} f_0(u) du \quad \text{CQFD}$$

8.3.3. Test bayésien

Comme dans le problème de l'estimation, on peut pondérer les probabilités d'erreurs ou, suivant le point de vue bayésien, introduire une loi à priori sur le paramètre. On choisit le test qui rend minimum le risque bayésien

$$E[P_\theta \{ \text{on prend la mauvaise décision} \}]$$

Dans le cas d'une hypothèse simple contre une hypothèse simple, on rend minimum

$$\begin{aligned} p_0 P_{\theta_0} \{D\} + p_1 P_{\theta_1} \{D^c\} &= (p_0 P_{\theta_0} - p_1 P_{\theta_1}) \{D\} + p_1 = \\ &= \int_D (p_0 f_0 - p_1 f_1) dx + p_1 \end{aligned}$$

La meilleure décision consiste à prendre $D = \{p_0 f_0 - p_1 f_1 < 0\}$.
C'est un test basé sur le rapport de vraisemblance

$$D = \left\{ \begin{array}{l} f_1 \geq p_0 \\ f_0 \geq p_1 \end{array} \right\}$$

On peut introduire le coût d'une erreur ; la région de rejet devient

$$D = \left\{ \begin{array}{l} f_1 \geq \frac{c_0 p_0}{c_1 p_1} \\ f_0 \geq \frac{c_1 p_1}{c_0 p_0} \end{array} \right\}$$

c_0 : coût de décider H_1 quand H_0 est vrai.
 c_1 : coût de décider H_0 quand H_1 est vrai.

D'autre part, si les coûts sont égaux à 1 et les probabilités a priori égales, on retrouve le critère du maximum de vraisemblance : décider H_1 si $f_1 > f_0$.

8.3.4. Hypothèses multiples et critère du maximum de vraisemblance

Lorsqu'on est en présence de deux hypothèses simples, une méthode de construction de la décision consiste à choisir l'hypothèse la plus vraisemblable ; la méthode est utilisable même si les deux hypothèses ne sont plus simples :

$$\text{On décide } H_1 \text{ si et seulement si } \sup_{\theta \in H_1} f_\theta(x) \geq \sup_{\theta \in H_0} f_\theta(x).$$

Cette méthode est surtout utilisée lorsque les hypothèses sont multiples : au lieu de choisir entre deux hypothèses, on a à choisir entre k hypothèses H_1, H_2, \dots, H_k . En fait ce problème rentre dans le cadre des problèmes d'estimation ; la méthode du maximum de vraisemblance s'exprime simplement par :

On décide H_j si $f_j(x)$ réalise le maximum de f_1, f_2, \dots, f_k (voir exercice 8.11).

8.4. TESTS GAUSSIENS (CRITÈRE DE NEYMAN-PEARSON)

8.4.1. Introduction

Les tests étudiés aux paragraphes suivants sont résumés dans le tableau ci-dessous :

<i>Test sur la moyenne</i>
variance connue
variance inconnue : Test de Student (test T)
<i>Test sur la variance (moyenne inconnue)</i>
Test du χ^2
<i>Comparaison de deux échantillons indépendants</i>
comparaison des variances : Test de Fisher (test F)
comparaison des moyennes (même variance) : test T

Tous ces tests sont basés sur des résultats fondamentaux concernant les variables gaussiennes.

Théorème : Soit X_1, X_2, \dots, X_n n variables indépendantes, de même loi gaussienne $\mathcal{N}(m, \sigma^2)$, alors

- 1) Les variables \bar{X} et $\sum (X_i - \bar{X})^2$ sont indépendantes.
- 2) $\frac{1}{\sigma^2} \sum (X_i - \bar{X})^2$ suit une loi de χ^2_{n-1} (Qui-deux à $n - 1$ degrés de liberté).
- 3) $\frac{\bar{X} - m}{\sqrt{\frac{1}{n(n-1)} \sum (X_i - \bar{X})^2}}$ suit une loi qui ne dépend pas des paramètres m et σ^2 appelée loi T_{n-1} (loi de Student à $n - 1$ degrés de liberté).

Démonstration : Posons $U_i = \frac{X_i - m}{\sigma}$. Les variables U_i sont indépendantes, gaussiennes $\mathcal{N}(0, 1)$ et on a

$$\bar{U} = \frac{\bar{X} - m}{\sigma} \quad \sum_i (U_i - \bar{U})^2 = \frac{1}{\sigma^2} \sum_i (X_i - \bar{X})^2$$

$$\frac{\bar{U}}{\sqrt{\frac{1}{n(n-1)} \sum (U_i - \bar{U})^2}} = \frac{\bar{X} - m}{\sqrt{\frac{1}{n(n-1)} \sum (X_i - \bar{X})^2}}$$

Cela prouve que la loi de $\frac{\bar{X} - m}{\sqrt{\frac{1}{n(n-1)} \sum (X_i - \bar{X})^2}}$ ne dépend ni de m , ni de σ^2 mais uniquement de n , et qu'il suffit de démontrer le théorème dans le cas $m = 0 \quad \sigma = 1$

$$\sum (U_i - \bar{U}) = 0$$

Cette relation montre que les variables $U_i - \bar{U}$ ne sont pas indépendantes.

$$\sum_i (U_i - \bar{U})^2 = (U_1 - \bar{U}, \dots, U_n - \bar{U}) \begin{pmatrix} U_1 - \bar{U} \\ \vdots \\ U_n - \bar{U} \end{pmatrix}$$

$$\begin{pmatrix} U_1 - \bar{U} \\ \vdots \\ U_n - \bar{U} \end{pmatrix} = \begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \dots \\ -\frac{1}{n} & \dots & \dots \\ \vdots & \dots & 1 - \frac{1}{n} \end{pmatrix} \begin{pmatrix} U_1 \\ \vdots \\ U_n \end{pmatrix} = M \begin{pmatrix} U_1 \\ \vdots \\ U_n \end{pmatrix}$$

or

$$M - I = -\frac{1}{n} \begin{pmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{pmatrix}$$

$M - I$ est une matrice de rang 1. M a pour valeur propre $\lambda = 1$ d'ordre $n - 1$ et $\lambda = 0$ d'ordre 1.

On cherche des vecteurs propres orthogonaux et la matrice tP qui fait passer des coordonnées sur la base de vecteurs propres, aux coordonnées sur la base canonique :

$${}^tP = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \dots & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & & \vdots & \vdots \\ 0 & -\frac{2}{\sqrt{6}} & & \vdots & \vdots \\ \vdots & 0 & & \frac{1}{\sqrt{n(n-1)}} & \vdots \\ 0 & \vdots & & \dots & \vdots \\ 0 & 0 & \dots & 0 & -\frac{n-1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n}} \end{pmatrix}$$

$$\sum_i (u_i - \bar{u})^2 = {}^t u {}^t P D P u = {}^t v D v$$

avec

$$v = P u = \begin{pmatrix} \frac{u_1 - u_2}{\sqrt{2}} \\ \frac{u_1 + u_2 - 2u_3}{\sqrt{6}} \\ \vdots \\ \frac{u_1 + \dots + u_{n-1} - (n-1)u_n}{\sqrt{n(n-1)}} \\ \sqrt{n} \bar{u} \end{pmatrix}$$

Si l'on repasse aux variables aléatoires

$$V = P U = \begin{pmatrix} U_1 - U_2 \\ \frac{U_1 + U_2 - 2U_3}{\sqrt{6}} \\ \vdots \\ \frac{U_1 + U_2 + \dots + U_{n-1} - (n-1)U_n}{\sqrt{n(n-1)}} \\ \sqrt{n} \bar{U} \end{pmatrix}$$

V est un vecteur gaussien centré, de matrice de covariance $P'P = I$. Les composantes de V sont donc indépendantes, de loi $N(0, 1)$ et on a

$${}^t V D V = V_1^2 + V_2^2 + \dots + V_{n-1}^2$$

qui est une variable de loi χ_{n-1}^2 . D'autre part $\sqrt{n} \bar{U}$, dernière composante, est indépendante des $n - 1$ premières composantes donc aussi de la somme des carrés de ces composantes, c'est-à-dire $V_1^2 + V_2^2 + \dots + V_{n-1}^2$ qui est aussi égal à $\sum_{i=1}^n (U_i - \bar{U})^2$. CQFD.

8.4.2. Test sur m , σ^2 étant connu

m caractérise la localisation des variables.

Si on teste $H_0 : \langle m = m_0 \rangle$ contre $H_1 : \langle m = m_1 \rangle$, au vu de X_1, X_2, \dots, X_n , on écrit le rapport des densités et on constate que la région de rejet de H_0 est de la forme

$$\{(m_1 - m_0) \bar{X} > \text{Cte}\}$$

Suivant le signe de $m_1 - m_0$, il s'agira d'un domaine $\{\bar{X} > \text{Cte}\}$ ou bien $\{\bar{X} < \text{Cte}\}$. Si on se place dans le premier cas, la constante est choisie de façon à ce que le niveau soit exactement égal au niveau imposé, par exemple 5% :

$$5\% = P_{H_0}(\bar{X} > A)$$

soit, en utilisant une table de $\mathcal{N}(0, 1)$

$$\sqrt{n} \frac{A - m_0}{\sigma} = 1,64$$

Ce test est aussi le test le plus puissant parmi les tests de niveau inférieur ou égal à 5%, de l'hypothèse $H_0 : \langle m \leq m_0 \rangle$ contre l'hypothèse $H_1 : \langle m > m_0 \rangle$.

Voir le tableau 8.1.

8.4.3. Test sur σ^2

σ^2 caractérise la dispersion des variables.

Pour tester $H_0 : \langle \sigma = \sigma_0 \rangle$ contre $H_1 : \langle \sigma = \sigma_1 \rangle$, on choisira une région de rejet de H_0 , de la forme

$$\begin{cases} \left\{ \sum (X_i - \bar{X})^2 > \text{Cte} \right\} & \text{si } \sigma_1 > \sigma_0 \\ \left\{ \sum (X_i - \bar{X})^2 < \text{Cte} \right\} & \text{si } \sigma_1 < \sigma_0 \end{cases}$$

Tableau 8.1.

H_0	H_1	Rejet de H_0 si :	Calcul de A	Table
$m = m_0$	$m > m_0$	$\bar{X} > A$	$P_{m=m_0}(\bar{X} > A) = \alpha$	$\frac{\sqrt{n} \cdot \bar{X} - m_0}{\sigma}$ dans une table de $\mathcal{N}(0, 1)$ en utilisant la variable $\frac{\sqrt{n} \cdot \bar{X} - m_0}{\sigma}$
$m = m_0$	$m < m_0$	$\bar{X} < A$	$P_{m=m_0}(\bar{X} < A) = \alpha$	
$m = m_0$	$m \neq m_0$	$ \bar{X} - m_0 > A$	$P_{m=m_0}(\bar{X} - m_0 > A) = \alpha$	

Estimateur de σ^2 : $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

$m = m_0$	$m > m_0$	$\sqrt{n} \cdot \frac{\bar{X} - m_0}{\sqrt{\hat{\sigma}^2}} > A$	$P_{m=m_0} \left(\sqrt{n} \cdot \frac{\bar{X} - m_0}{\sqrt{\hat{\sigma}^2}} > A \right) = \alpha$	Student à $n-1$ degrés de liberté
$m = m_0$	$m < m_0$	$\sqrt{n} \cdot \frac{\bar{X} - m_0}{\sqrt{\hat{\sigma}^2}} < -A$	$P_{m=m_0} \left(\sqrt{n} \cdot \frac{\bar{X} - m_0}{\sqrt{\hat{\sigma}^2}} < -A \right) = \alpha$	
$m = m_0$	$m \neq m_0$	$\left \sqrt{n} \cdot \frac{\bar{X} - m_0}{\sqrt{\hat{\sigma}^2}} \right > A$	$P_{m=m_0} \left(\left \sqrt{n} \cdot \frac{\bar{X} - m_0}{\sqrt{\hat{\sigma}^2}} \right > A \right) = \alpha$	

X_1, X_2, \dots, X_n sont indép. de loi $\mathcal{N}(m, \sigma^2)$

On trouve la constante en écrivant l'équation qui donne le niveau ; par exemple, si $\sigma_1 > \sigma_0$

$$5\% = P_{H_0}(\sum (X_i - \bar{X})^2 > A)$$

On doit diviser les deux termes de l'inégalité par σ_0^2 pour se ramener à une table de χ^2 (tableau 8.2).

Tableau 8.2.

H_0	H_1	Rejet de H_0 si :	Calcul de A (et B)	Table
$\sigma = \sigma_0$	$\sigma > \sigma_0$	$\hat{\sigma}^2 > A$	$P_{\sigma=\sigma_0}(\hat{\sigma}^2 > A) = \alpha$	χ^2 à $n-1$ degrés de liberté, en utilisant la v.a.r.
$\sigma = \sigma_0$	$\sigma < \sigma_0$	$\hat{\sigma}^2 < A$	$P_{\sigma=\sigma_0}(\hat{\sigma}^2 < A) = \alpha$	
$\sigma = \sigma_0$	$\sigma \neq \sigma_0$	$\hat{\sigma}^2 < A$ ou $\hat{\sigma}^2 > B$	$P_{\sigma=\sigma_0}(\hat{\sigma}^2 < A) = \frac{\alpha}{2}$ et $P_{\sigma=\sigma_0}(\hat{\sigma}^2 > B) = \frac{\alpha}{2}$	$(n-1) \frac{\hat{\sigma}^2}{\sigma_0^2}$

X_1, X_2, \dots, X_n sont indép. de loi $\mathcal{N}(m, \sigma^2)$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

8.4.4. Test sur m, σ^2 étant inconnu

Lorsque σ^2 est connu, on construit la région de rejet de H_0 pour un niveau donné en écrivant par exemple pour le niveau 5%

$$\sqrt{n} \times \frac{A - m_0}{\sigma} = 1,64$$

Dans le cas d'une variance inconnue, on reprend la forme en remplaçant σ^2 par l'estimateur sans biais $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Voir le tableau 8.1.

8.4.5. Comparaison de deux échantillons indépendants

Exemple : Supposons que l'on veuille comparer deux engrais A et B. Deux champs considérés comme identiques (orientation, nature du sol, ...) sont traités l'un avec l'engrais A, l'autre avec l'engrais B ; chaque camp étant découpé en parcelles, on pourra faire une comparaison des productivités obtenues.

Soit X_1, X_2, \dots, X_n et Y_1, Y_2, \dots, Y_m deux échantillons indépendants des lois respectives $\mathcal{N}(m_1, \sigma_1^2)$ et $\mathcal{N}(m_2, \sigma_2^2)$.

On compare tout d'abord les variances en utilisant les estimateurs sans biais habituels :

$$\hat{\sigma}_1^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2$$

$$\hat{\sigma}_2^2 = \frac{1}{m-1} \sum_j (Y_j - \bar{Y})^2$$

La variable $\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$ suit une loi qui ne dépend pas de (m_1, σ_1^2) ni de (m_2, σ_2^2) , mais

uniquement des tailles n et m , appelée loi de Fisher, $F_{n-1, m-1}$.

Voir le tableau 8.3 pour les différents tests.

Dans le cas de deux échantillons de même variance, on estime la variance commune avec

$$\hat{\sigma}^2 = \frac{1}{n+m-2} \left[\sum_i (X_i - \bar{X})^2 + \sum_j (Y_j - \bar{Y})^2 \right]$$

On peut alors comparer les moyennes à l'aide d'un test de Student, en utilisant la variable

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right) \hat{\sigma}^2}}$$

Voir le tableau 8.4 pour les différents tests.

Tableau 8.3.

H ₀	H ₁	Rejet de H ₀ si :	Calcul de A (et B)	Table
$\sigma_1 = \sigma_2$	$\sigma_1 > \sigma_2$	$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} > A$	$P_{\sigma_1 = \sigma_2} \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} > A \right) = \alpha$	Fisher F _{n-1, m-1} en utilisant la variable $\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$
$\sigma_1 = \sigma_2$	$\sigma_1 < \sigma_2$	$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} < A$	$P_{\sigma_1 = \sigma_2} \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} < A \right) = \alpha$	
$\sigma_1 = \sigma_2$	$\sigma_1 \neq \sigma_2$	$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} < A$ ou $\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} > B$	$P_{\sigma_1 = \sigma_2} \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} < A \right) = \frac{\alpha}{2}$ et $P_{\sigma_1 = \sigma_2} \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} > B \right) = \frac{\alpha}{2}$	

$\{X_1, X_2, \dots, X_n$ sont indép. de loi $\mathcal{N}(m_1, \sigma_1^2)$
 $\{Y_1, Y_2, \dots, Y_m$ sont indép. de loi $\mathcal{N}(m_2, \sigma_2^2)$
 les deux échantillons sont indépendants

Tableau 8.4.

H ₀	H ₁	Rejet de H ₀ si :	Calcul de A	Table
$m_1 = m_2$	$m_1 > m_2$	$\frac{1}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma^2}} > A$	$P_{m_1 = m_2} \left(\frac{1}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma^2}} > A \right) = \alpha$	Student à n + m - 2 degrés de liberté
$m_1 = m_2$	$m_1 < m_2$	$\frac{1}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma^2}} < A$	$P_{m_1 = m_2} \left(\frac{1}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma^2}} < A \right) = \alpha$	
$m_1 = m_2$	$m_1 \neq m_2$	$\left \frac{1}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma^2}} \right > A$	$P_{m_1 = m_2} \left(\left \frac{1}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma^2}} \right > A \right) = \alpha$	

$\left\{ \sigma^2 = \frac{1}{n+m-2} \left\{ \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2 \right\} \right.$
 les deux échantillons sont indépendants

8.5. INTERVALLES DE CONFIANCE

8.5.1. Introduction

Supposons que l'on veuille estimer un paramètre réel ; On a vu au paragraphe 8.2. différentes méthodes d'évaluation de ce paramètre à partir des observations ; Il s'agissait d'estimation ponctuelle. Nous allons maintenant chercher des estimations par intervalle. A chaque observation, on associe un intervalle qui a de grandes chances de contenir le paramètre :

$$x_1, \dots, x_n \rightarrow [a(x_1, \dots, x_n), b(x_1, \dots, x_n)]$$

Cet intervalle qui dépend de l'observation, est donc aléatoire. On le choisira de façon à avoir un risque limité par α fixé à l'avance. $1 - \alpha$ est le *coefficient de sécurité*.

$$P \{ \theta \notin [a(X_1, \dots, X_n), b(X_1, \dots, X_n)] \} \leq \alpha$$

soit

$$P \{ a(X_1, \dots, X_n) \leq \theta \leq b(X_1, \dots, X_n) \} \geq 1 - \alpha$$

On choisira, dans la mesure du possible, l'intervalle qui est le plus petit de façon à avoir une meilleure précision sur le paramètre.

C'est la méthode employée le soir des élections pour estimer la proportion de votants pour tel ou tel candidat : on donne une fourchette de valeurs. Ces estimations nécessitent des techniques fines de sondages et de correction de données.

8.5.2. Estimation de la moyenne

Etant donné un n-échantillon d'une loi de moyenne μ et variance σ^2 , si on part de l'estimateur \bar{X} , on peut construire un intervalle en regardant les fluctuations de \bar{X} autour de m ; reprenons l'exemple des mesures

1.2, 2.1, -0.5, 0.8, 1.7, 0.9, 1.4, 1.6, 1.1

$$\bar{X}(\omega) = 1,14\dots$$

$$E(\bar{X}) = m$$

$$\text{var}(\bar{X}) = \sigma^2/n$$

D'autre part, si n est assez grand, on peut utiliser l'approximation par une loi normale

$$\bar{X} - m \text{ est de loi } \mathcal{N}(0, \sigma^2/n)$$

$$\sqrt{n} \frac{\bar{X} - m}{\sigma} \text{ est de loi } \mathcal{N}(0, 1)$$

Les fluctuations d'une loi $\mathcal{N}(0, 1)$ sont connues ; par exemple, on a 95 % de chances d'obtenir une valeur entre $-1,96$ et $1,96$. Si σ est connu, cela nous fournit un intervalle de confiance à 95 % pour m :

$$\bar{X} - 1,96 \sigma / \sqrt{n} \leq m \leq \bar{X} + 1,96 \sigma / \sqrt{n}$$

Si σ est inconnu, on utilise l'estimateur sans biais de σ^2 , $\hat{\sigma}^2$,

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum (X_j - \bar{X})^2$$

ce qui donne sur l'exemple considéré

$$\hat{\sigma}^2 = 0,547...$$

$$0,541 \leq m \leq 1,747$$

en effet, $\sqrt{n}(\bar{X} - m)/\hat{\sigma}$ n'est plus une variable gaussienne, mais suit une loi qui ne dépend que de la taille n de l'échantillon et appelée loi de student à $n - 1$ degrés de libertés. La densité d'une loi de Student a sensiblement la même forme qu'une densité $\mathcal{N}(0,1)$ mais elle est plus étalée. Pour un échantillon de taille 10 et un coefficient de sécurité de 95 %, on trouve l'intervalle,

$$\bar{X} - 2,26 S / \sqrt{n} \leq m \leq \bar{X} + 2,26 S / \sqrt{n}$$

avec

$$S = \sqrt{\hat{\sigma}^2}$$

8.5.3. Estimation de la variance

Si on part de l'estimateur $\hat{\sigma}^2$,

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2$$

on connaît les fluctuations d'une loi de χ^2 : par exemple, pour un échantillon de taille 10 et un coefficient de sécurité de 98 %, une variable χ^2 à 9 degrés de liberté ayant 98 % de chances d'être entre 2,088 et 21,666, on obtient l'intervalle,

$$\frac{\sum (X_i - \bar{X})^2}{21,666} \leq \sigma^2 \leq \frac{\sum (X_i - \bar{X})^2}{2,088}$$

On remarquera l'arbitraire dans le choix de l'intervalle $[2,088, 21,666]$ par la variable χ^2 ; en effet, une loi de χ^2 n'étant pas symétrique par rapport à une valeur particulière, on n'a pas de raison de faire un découpage symétrique. Sur l'exemple ci-dessus, on a pris le découpage symétrique par commodité :

$$P(\text{variable } \chi^2 \leq 2,088) = 1 \% = P(\text{var } \chi^2 \geq 21,666)$$

8.5.4. Relation avec les tests

L'estimation par intervalle est liée à la construction d'un test bilatère sur le paramètre. Si l'on sait construire le test de l'hypothèse H_0 « la moyenne est m » contre l'hypothèse H_1 « la moyenne est différente de m » au niveau α pour chaque valeur m , on en déduit un intervalle de confiance au niveau α :

observations \rightarrow ensembles des m tels que l'hypothèse H_0
est acceptée avec ces observations

EXERCICES

E.8.1. Un grand magasin désire estimer la proposition p de chèques sans provision ; 150 paiements par chèques sont examinés et 13 chèques sont dans ce cas. Donner une estimation de p , que l'on notera $\hat{p}(\omega)$. Faire un calcul approché de

$$P(|p - \hat{p}| \leq 0,5 \%)$$

en utilisant une approximation gaussienne.

E.8.2. Dans une expérience pour l'étude d'une antenne, on utilise un réseau de sondes, c'est-à-dire, de capteurs de champ électromagnétique. On étudie tout d'abord la sensibilité des sondes ; à cet effet, on envoie une onde à la fréquence 10 GHz qui arrive en incidence normale sur le réseau de sondes et on mesure le signal reçu sur chaque sonde. Les mesures obtenues pour 17 sondes sont données sous la forme partie réelle et partie imaginaire :

(6,953, 0,422)	(7,295, -0,881)	(8,086, -2,260)
(7,021, 0,325)	(6,279, -1,762)	
(6,259, -1,795)	(7,021, -2,488)	(5,419, -1,215)
(5,312, -0,981)	(6,152, -1,096)	(4,736, -1,794)
(6,474, -0,784)	(6,035, -2,509)	(5,498, -0,070)
(6,796, -0,853)	(6,074, 0,311)	(4,941, 0,163)

En appelant X la partie réelle et Y la partie imaginaire et en supposant (X, Y) gaussien, estimer $E(X)$ et $E(Y)$ par la méthode du maximum de vraisemblance (on suppose que les mesures faites sur les différentes sondes sont des variables aléatoires indépendantes et de même loi).

E.8.3. Soit X une variable aléatoire réelle de densité $f_\theta(x)$

$$f_\theta(x) = \frac{1}{\pi} \times \frac{1}{1 + (x - \theta)^2}$$

avec θ inconnu. Montrer que $X - \theta$ suit une loi de Cauchy. En déduire la loi de $\frac{X_1 + \dots + X_n}{n}$ si les X_i sont indépendantes et de densité f_θ . On estime le

paramètre de localisation θ par $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$. Que vaut

$$P\left(|\bar{X} - \theta| > \frac{1}{100}\right) ?$$

E.8.4. Estimation d'un instant initial θ .

Soit X_1, X_2, \dots, X_n des v.a.r. indépendantes, de même densité $f_\theta(x)$:

$$f_\theta(x) = e^{-(x-\theta)} 1_{[\theta, +\infty[}(x)$$

a) Montrer que $X - \theta$ suit une loi exponentielle.

b) Calculer $E(\bar{X})$ et $E\left(\inf_i X_i\right)$. En déduire deux estimateurs sans biais de θ .

Quel est l'estimateur qui a la plus petite variance ?

c) Calculer l'estimateur du maximum de vraisemblance.

E.8.5. Le nombre d'appels téléphoniques en une minute au standard d'une école, est supposé aléatoire et noté X . Si X suit une loi de Poisson de paramètre λ avec λ inconnu, quel est l'estimateur du maximum de vraisemblance si on dispose de n observations X_1, X_2, \dots, X_n ?

E.8.6. On reprend les notations de l'exercice précédent.

On ne dispose que d'une seule observation et on suppose le paramètre λ lui-même aléatoire, de loi exponentielle de paramètre 10^{-1} . On cherche un estimateur proportionnel à X , cX , qui minimise

$$E\left[\frac{(cX - \lambda)^2}{\lambda}\right]$$

de façon à tenir compte de la variance de X . Quel est le meilleur choix de c ?

E.8.7. On reprend les notations de l'exercice précédent en supposant λ aléatoire, de loi exponentielle de paramètre 10^{-1} . On dispose de n observations X_1, X_2, \dots, X_n . Quel est l'estimateur bayésien de λ , c'est-à-dire l'estimateur $d(X_1, \dots, X_n)$ qui minimise

$$E\{[d(X_1, X_2, \dots, X_n) - \lambda]^2\}$$

λ ayant pour densité $\frac{1}{10} e^{-\lambda/10}$ si $\lambda > 0$.

E.8.8. Un producteur contrôle la qualité d'un lot de 1 500 objets avant de le mettre sur le marché. Il veut un plan de contrôle sur au plus 100 objets, de façon que si p , pourcentage théorique de défectueux, est inférieur à 1 %, on ait au plus 20 % de chances de rejeter le lot et, pour satisfaire les clients, si $p \geq 5$ % on ait au moins 10 % de rejeter le lot.

Si il choisit de prélever 40 objets au hasard et de rejeter le lot dès qu'il y a un objet non conforme, ce plan satisfait-il les contraintes imposées ?

E.8.9. Afin de contrôler le réglage d'une machine, on prélève 10 objets. On suppose que la mesure faite sur un objet au hasard est une variable $\mathcal{N}(\mu, \sigma^2)$. On obtient les valeurs suivantes :

99,450	98,654	99,335	102,020	101,503
100,610	100,898	99,897	101,707	102,677

a) Donner une estimation de σ (qui caractérise la précision de la mesure).

b) Peut-on accepter l'hypothèse H_0 « la machine est réglée sur $\mu = 100$ » contre l'hypothèse H_1 : « $\mu > 100$ » avec un risque $\alpha = 5$ % ?

E.8.10. Soit $f_\theta(x)$ la densité $\frac{1}{\pi} \times \frac{1}{1 + (x - \theta)^2}$.

Montrer que l'information de Fisher $I(\theta)$ est finie et ne dépend pas de θ .

E.8.11. Problème de démodulation.

On observe un signal Y_t sur l'intervalle $[0, 1]$. On sait que $Y_t = s(t) + B_t$ où B_t est un bruit blanc gaussien de niveau σ^2 et $s(t)$ est l'un des 4 signaux $s_i(t)$:

$$s_1(t) = \sin \omega t$$

$$s_2(t) = \sin\left(\omega t + \frac{\pi}{2}\right)$$

$$s_3(t) = \sin(\omega t + \pi)$$

$$s_4(t) = \sin\left(\omega t + \frac{3\pi}{2}\right)$$

où ω est connu. On choisit T de la forme $\frac{k\pi}{2\omega}$ et on calcule les quantités $A_i(\omega)$:

$$A_i(\omega) = \int_0^T Y_t(\omega) s_i(t) dt$$

Calculer la loi du vecteur $\begin{pmatrix} A_1 \\ \vdots \\ A_4 \end{pmatrix}$ sous les 4 hypothèses H_i : « $s(t) = s_i(t)$ ». On

choisit l'hypothèse la plus vraisemblable au vu du vecteur $\begin{pmatrix} A_1 \\ \vdots \\ A_4 \end{pmatrix}$. Exprimer cette

règle de décision.

E.8.12. Problème de détection.

On observe un signal aléatoire Y_t sur l'intervalle $[0, T]$. On veut tester la présence d'un signal connu $s(t)$. On suppose que Y_t s'écrit $s(t) + B_t$ ou

B_t suivant que le signal $s(t)$ est envoyé ou non, B_t étant un bruit blanc gaussien de niveau σ^2 . On pose

$$\begin{aligned} H_0 : Y_t &= B_t \\ H_1 : Y_t &= s(t) + B_t \end{aligned}$$

et on calcule $A(\omega) = \int_0^T Y_t(\omega) s(t) dt$.

Calculer la loi de A sous les deux hypothèses H_0 et H_1 . Construire le test bayésien, basé sur l'observation de A , en supposant l'hypothèse H_1 de probabilité 10^{-1} . Calculer l'erreur de 1^{re} espèce, l'erreur de 2^e espèce et l'erreur bayésienne.

Montrer que la règle de décision bayésienne se réalise grâce à un intégrateur et une comparaison à un seuil.

E.8.13. Démodulation par échantillonnage.

On considère un signal $Y_t = s(t) + B_t$ où $s(t) = s_0(t)$ sous l'hypothèse H_0 , $s_1(t)$ pour l'hypothèse H_1 ; par exemple, en modulation par saut de fréquence, on utilise $s_0(\omega) = \sin \omega_0 t$ et $s_1(t) = \sin \omega_1 t$. B_t est un bruit gaussien de densité spectrale de puissance $B_0 1_{[-F, F]}(\lambda)$. On échantillonne Y_t au pas $\frac{1}{2F}$ et on pose $Y_n = Y_{n\Delta t}$. Construire le test bayésien basé sur l'observation de Y_1, Y_2, \dots, Y_N , en supposant les deux hypothèses H_0 et H_1 équiprobables et $T = \frac{N}{2F}$ avec F suffisamment grand pour remplacer les intégrales sur $[0, T]$ par des sommes de Riemann.

Montrer que la règle de décision s'obtient en calculant $\int_0^T [s_1(t) - s_0(t)] Y_t dt$ et en comparant cette variable à un seuil. Calculer les erreurs de 1^{re} et 2^e espèce.

CHAPITRE 9

ET UN PEU PLUS DE STATISTIQUES

9.1. LOI EMPIRIQUE ET TESTS BASÉS SUR LA LOI EMPIRIQUE

9.1.1. Loi empirique

Supposons que l'on fasse des mesures de résistances sur un échantillon de taille 100 provenant d'un lot de résistances de valeur nominale 200 k Ω et que l'on obtienne : 193, 196, ... on peut représenter graphiquement la distribution des valeurs prises par l'échantillon en dessinant l'histogramme, avec en ordonnée le nombre d'éléments ayant une résistance r (fig. 9.1) :

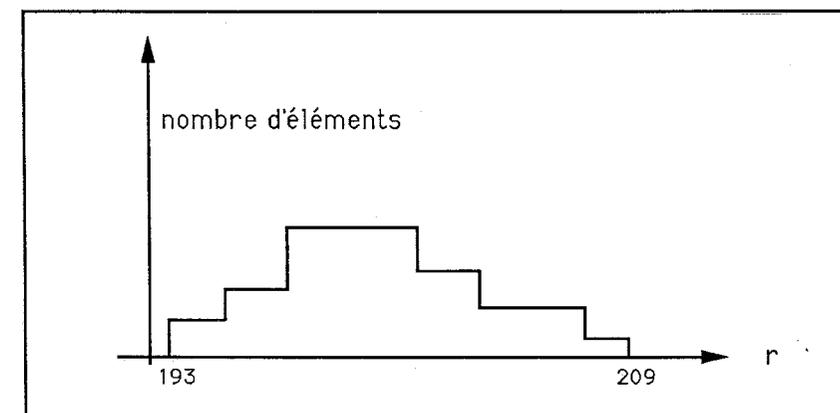


Figure 9.1.

Supposons que les 100 valeurs proviennent de variables aléatoires indépendantes et de même loi R_1, R_2, \dots, R_{100} , de fonction de répartition $F(t)$. Alors en appelant N_1 la variable égale à 1 si $R_1 < t$, 0 sinon :

$$\begin{aligned} N_1 &= 1 \quad \text{si } R_1 < t, \quad 0 \text{ sinon} \\ N_2 &= 1 \quad \text{si } R_2 < t, \quad 0 \text{ sinon} \\ &\vdots \\ N_n &= 1 \quad \text{si } R_n < t, \quad 0 \text{ sinon} \end{aligned}$$

on peut appliquer la loi des grands nombres et on obtient :

$$\text{pourcentage de var. inf. à } t = \frac{N_1 + N_2 + \dots + N_n}{n} \xrightarrow{n \rightarrow +\infty} F(t)$$

On appellera $G_n(t)$ la fonction de répartition empirique définie par

$$G_n(t) = \frac{\text{nombre de valeurs obtenues qui sont inférieures à } t}{n}$$

C'est la fonction de répartition associée à la distribution de probabilité discrète

$$P_n(X = r) = \frac{\text{nombre de valeurs obtenues qui sont égales à } r}{n}$$

D'autre part, on remarque que la moyenne empirique est égale à l'espérance de la loi empirique.

D'après ce qui a été dit plus haut, la fonction de répartition empirique converge ponctuellement vers la fonction de répartition sous-jacente.

La distribution empirique nous permet donc « d'estimer » la loi théorique. On admettra le résultat qui suit.

Théorème : Soit A_1, A_2, \dots, A_k une partition de l'espace des épreuves Ω ; on répète l'expérience, de façon indépendante, n fois et on appelle n_i le nombre de fois où A_i est réalisé, alors

$$\sum_i \frac{(n_i - np_i)^2}{np_i} \xrightarrow{n \rightarrow +\infty} \text{en loi } \chi_{k-1}^2$$

où

$$p_i = P(A_i)$$

C'est ce théorème que l'on utilise dans le test d'ajustement du χ^2 (§ 9.1.2).

On peut, de même, construire la loi empirique associée à un échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$ d'une loi de probabilité sur \mathbb{R}^2 :

$$P_n((x, y)) = \frac{\text{nombre de } (X_i, Y_i) \text{ égaux à } (x, y)}{n}$$

On peut, en particulier, calculer les variances, covariances de cette loi empirique :

$$\text{variance empirique de } X = \frac{1}{n} \sum (X_i - \bar{X})^2$$

$$\text{covariance empirique du couple } (X, Y) = \frac{1}{n} \sum (X_i - \bar{X})(Y_i - \bar{Y})$$

$$\text{coefficient de corrélation empirique} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2} \sqrt{\sum (Y_i - \bar{Y})^2}} = r_n$$

Si le couple (X, Y) admet des variances et covariances, alors le coefficient de corrélation empirique converge vers le coefficient de corrélation théorique ρ :

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)} \sqrt{\text{var}(Y)}}$$

ρ est compris entre -1 et 1 ; $|\rho| = 1$ si et seulement si il existe une relation de dépendance affine entre X et Y . D'autre part si X et Y sont indépendantes, ρ est nul; la réciproque n'est vraie que si le vecteur (X, Y) est gaussien. Pour tester l'indépendance entre les deux composantes X et Y , on utilisera le test du χ^2 (§ 9.1.4) pour les grands échantillons ou le test de Spearman pour les petits échantillons.

9.1.2. Test d'ajustement du χ^2 : méthode pratique

9.1.2.1. Ajustement sur une loi complètement spécifiée

Soit $F_0(t)$ la fonction de répartition de la loi sur laquelle on ajuste. Par exemple, on mesure la durée de vie d'une ampoule électrique et on se demande si la loi est exponentielle d'espérance 750 h.

On teste

$$H_0 : \text{« la loi a pour fonction de répartition } F_0 \text{ »}$$

contre

$$\text{« la fonction de répartition n'est pas égale à } F_0 \text{ »}$$

avec

$$F_0(t) = 1 - e^{-t/750}$$

Les N données sont réparties dans k intervalles $]-\infty, a_1], [a_1, a_2], \dots, [a_{k-1}, +\infty[$.

On appelle n_i le nombre de variables qui appartiennent au i -ième intervalle. n_i est une variable aléatoire d'espérance Np_i où p_i est la probabilité théorique d'appartenir au i -ième intervalle. On a par exemple, $p_2 = F(a_2) - F(a_1)$.

On calcule la variable D ,

$$D = \sum_i \frac{(n_i - Np_i)^2}{Np_i}$$

La loi de D est difficile à calculer (voir ex. 9.1) mais on sait que D converge en loi vers un χ_{k-1}^2 lorsque N tend vers l'infini. La convergence est d'autant meilleure que les intervalles sont de probabilité sensiblement égale et pas trop faible. Dans la pratique on utilise la loi limite dès que

$$N \geq 30 \text{ et } Np_i \geq 5 \text{ pour tout } i$$

Lorsque la condition $Np_i \geq 5$ n'est pas vérifiée on opère un regroupement d'intervalles jusqu'à ce que la condition soit vérifiée.

Exemple : On mesure la durée de vie d'une ampoule électrique et on se demande si la loi est exponentielle d'espérance 750 h. On réalise l'expérience 100 fois et en divisant par 750, on se ramène à une loi exponentielle de paramètre 1 : la question devient « les données suivantes proviennent-elles d'une loi exponentielle de paramètre 1 ? »

Durée	0 à 0,3	0,3 à 0,5	0,5 à 0,7	0,7 à 1	1 à 2	2 à 3	> 3
n_i	23	10	12	16	25	9	5

L'effectif théorique pour la dernière case est $100 P(X > 3)$

$$P(X > 3) = \int_3^{+\infty} e^{-t} dt = e^{-3} = 0,0498$$

On regroupe donc les deux dernières cases de façon à avoir un effectif théorique supérieur à 5

Durée	0 à 0,3	0,3 à 0,5	0,5 à 0,7	0,7 à 1	1 à 2	≥ 2
n_i	23	10	12	16	25	14
np_i	25,9	13,5	11	12,9	23,3	13,5
$\frac{(n_i - np_i)^2}{np_i}$	0,32	0,91	0,09	0,74	0,12	0,02

On trouve $D(\omega) = 2,2$. On utilise une table de χ^2_5 ce qui donne pour $\alpha = 5\%$ la région de rejet $\{D \geq 11\}$. La différence entre les effectifs observés et les effectifs théoriques n'est pas significative et on garde l'hypothèse H_0 « la loi est exponentielle de paramètre 1 ».

9.1.2.2. Ajustement sur une loi avec paramètres inconnus

On utilise la même méthode après avoir estimé le ou les paramètres inconnus par la méthode du maximum de vraisemblance. Par exemple, on observe une v.a. à valeurs entières et on se demande si la loi est une loi de Poisson.

L'expérience a été réalisée 200 fois et on a obtenu

Valeur	0	1	2	3	4
n_i	109	65	22	3	1

On estime le paramètre λ pour la méthode au maximum de vraisemblance :

$$\hat{\lambda} = \frac{1(65) + 2(22) + 3(3) + 4(1)}{200} = 0,61$$

Les effectifs théoriques sont calculés pour une loi de Poisson de paramètre 0,61 et on regroupe les cases de façon à avoir un effectif théorique au moins égal à 5

Valeur	0	1	2	3	4
n_i	109	65	22	3	1
np_i	108,7	66,3	20,2	4,1	0,7
$\frac{(n_i - np_i)^2}{np_i}$	$8 \cdot 10^{-4}$	0,0255		0,04	
					$D(\omega) = 0,066$

On admettra le calcul du nombre de degrés de liberté :

Nombre de d° de liberté = = nombre de cases - 1 - nombre de paramètres estimés

Sur l'exemple traité ci-dessus, on a donc 1 seul degré de liberté ; pour le niveau $\alpha = 5\%$, la valeur critique est 3,841 :

$$\text{région de rejet} = \{D > 3,841\}$$

On accepte donc l'hypothèse H_0 « les variables suivent une loi de Poisson ».

9.1.3. Test d'ajustement de Kolmogorov Smirnov (loi complètement spécifiée)

Le test de Kolmogorov Smirnov ne s'applique qu'aux v.a.r. ayant une fonction de répartition continue ; d'autre part on ne l'utilise que pour des tailles faibles car il nécessite de calculer la fonction de répartition à tous les points de mesure obtenus par l'expérience.

Soit F la fonction de répartition (inconnue) et F_0 une fonction de répartition complètement spécifiée ; le test s'écrit

$$H_0 : \langle F = F_0 \rangle \quad \text{contre} \quad H_1 : \langle F \neq F_0 \rangle$$

La fonction de répartition empirique construite sur un n -échantillon est une fonction en escalier, qui saute à chaque valeur prise par l'une des variables. On regarde si la fonction de répartition empirique est proche de F_0 . La figure 9.2

illustre l'exemple de l'ajustement sur une loi exponentielle de paramètre 1 avec les 10 données :

3,22	2,26
0,95	0,04
1,74	0,3
1,12	0,64
0,36	0,37

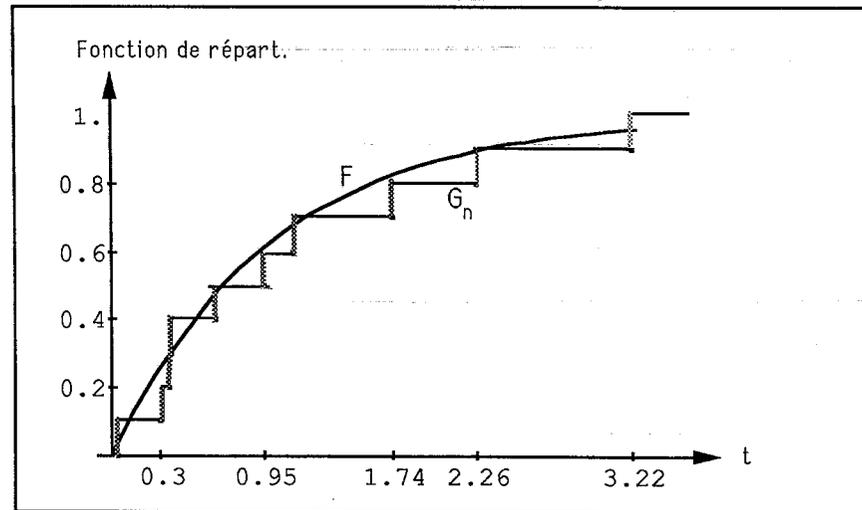


Figure 9.2.

On admettra le résultat suivant qui permet de construire le test :

Théorème : Soit X_1, \dots, X_n un n -échantillon de fonction de répartition F continue et G_n la fonction de répartition empirique construite à l'aide du n -échantillon ; alors si $F = F_0$, la v.a. D définie par

$$D = \text{Sup}_t |G_n(t) - F_0(t)|$$

suit une loi qui ne dépend pas de la loi F_0 continue sur laquelle on ajuste, mais uniquement de la taille de l'échantillon.

La loi de la variable est tabulée pour les différentes valeurs de n . La région de rejet de l'hypothèse H_0 est de la forme $\{D > a\}$; a est choisi de façon à avoir un niveau que l'on s'est fixé à l'avance. En ce qui concerne le calcul de $D(\omega)$ on remarque qu'il suffit de regarder l'écart entre la fonction de répartition empirique et la fonction de répartition théorique aux points de discontinuité puis de prendre le sup de tous ces écarts ce qui donne sur l'exemple ci-dessus

t	$F_0(t)$	$G_n(t)$	$G_n(t+0)$	Plus grand écart
0,04	0,039	0	10^{-1}	0,039
0,3	0,259	10^{-1}	$2,10^{-1}$	0,159
0,36	0,302	$2,10^{-1}$	$3,10^{-1}$	0,102
0,37	0,309	$3,10^{-1}$	$4,10^{-1}$	0,091
0,64	0,473	$4,10^{-1}$	$5,10^{-1}$	0,073
0,95	0,613	$5,10^{-1}$	$6,10^{-1}$	0,113
1,12	0,674	$6,10^{-1}$	$7,10^{-1}$	0,074
1,74	0,824	$7,10^{-1}$	$8,10^{-1}$	0,124
2,26	0,896	$8,10^{-1}$	$9,10^{-1}$	0,096
3,22	0,960	$9,10^{-1}$	1	0,060

$$D(\omega) = 0,159$$

On trouve la valeur critique dans une table. Pour $\alpha = 5\%$ $P(D_{10} > 0,41) = 5\%$. On garde donc l'hypothèse H_0 : « La loi est exponentielle de paramètre 1 ». Contrairement au test du χ^2 , on ne peut pas utiliser le test de Kolmogorov Smirnov lorsque un ou plusieurs paramètres de la loi, sont inconnus.

9.1.4. Test d'indépendance du χ^2 (grands échantillons)

Il s'agit d'un test d'indépendance entre deux caractères ou deux variables aléatoires. Par exemple, peut-on au vu des observations ci-dessous, affirmer qu'il y a indépendance entre la couleur des yeux et la couleur des cheveux ?

	yeux	bleus	verts, gris	bruns
cheveux				
Blonds		1 768	946	115
Roux		47	53	16
Châtains		807	1 387	438
Noirs		189	746	288

Si il y a indépendance entre les deux caractères, alors pour chaque case, la probabilité doit être le produit des probabilités correspondantes :

$$P(\text{yeux bleus et cheveux blonds}) = P(\text{yeux bleus}) P(\text{cheveux blonds})$$

...

Ces probabilités sont inconnues et on les estime

$$P(\text{yeux bleus}) = \frac{1768 + 47 + 807 + 189}{\text{nombre total}}, \dots$$

Il suffit d'estimer $P(\text{yeux bleus})$, $P(\text{yeux verts ou gris})$ car $P(\text{yeux bruns}) = 1 - [P(\text{yeux bleus}) + P(\text{yeux verts ou gris})]$.

De même, pour les cheveux, on a trois probabilités à estimer. Le nombre de degrés de liberté est encore égal au nombre de cases $- 1 - \text{nombre de paramètres à estimer}$, soit 6. De façon générale, le calcul du nombre de degrés de liberté se fait suivant la règle suivante :

<p>Nombre de degrés de liberté = $(a - 1)(b - 1)$ a = nombre de cases pour la 1^{re} variable b = nombre de cases pour la 2^e variable</p>

Quant à la variable D , elle s'écrit

$$D = \sum_{i,j} \frac{(n_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}}$$

où

$$\begin{cases} p_{ij} = \hat{p}_i \hat{q}_j \\ \hat{p}_i = \text{estimation de } p_i, & \text{probabilité de la } i\text{-ième case} \\ & \text{pour la 1^{re} variable} \\ \hat{q}_j = \text{estimation de } q_j, & \text{probabilité de la } j\text{-ième case} \\ & \text{pour la 2^e variable} \end{cases}$$

$P(\text{yeux bleus et cheveux blonds})$ est estimé par $\frac{2811}{6800} \times \frac{2829}{6800}, \dots$

On trouve ici $D(\omega) > 306$ en calculant le premier terme de la somme. On rejette l'hypothèse d'indépendance, même au niveau $\alpha = 1\%$ (la valeur critique est égale à 16,81).

9.2. TESTS NON PARAMÉTRIQUES (petits échantillons)

9.2.1. Introduction

Considérons l'exemple suivant : un nouveau médicament semble avoir un effet sur le rythme cardiaque ; on mesure alors le rythme cardiaque avant et après absorption du médicament sur 30 personnes ; on appelle X et Y les variables avant et après. Pour tester l'hypothèse H_0 « il n'y a aucun effet » contre l'hypothèse H_1 « il y a un effet du médicament », on considère la variable $X - Y$; On supposera que l'on a un 30-échantillon de cette variable (ce qui veut dire qu'il n'y a pas de liens entre les 30 mesures effectuées sur les 30 malades).

On peut interpréter ce problème comme un problème de test sur la moyenne de $X - Y$, supposer $X - Y$ gaussienne...

On peut aussi interpréter ce problème comme un problème de test sur la médiane de $X - Y$, médiane qui est définie, lorsque la fonction de répartition est continue, par :

$$\text{médiane de la loi de } Z = \text{Inf } \{m / P(Z < m) = P(Z > m) = 50\%\}$$

L'hypothèse H_0 s'écrit, H_0 « la médiane de $X - Y$ est nulle ».

Le deuxième modèle est plus général que le modèle gaussien et on peut tester l'hypothèse H_0 avec un autre test (voir §§ 9.2.2 et 9.2.3). De façon générale, les modèles paramétriques sont souvent trop restrictifs et supportent mal les écarts au modèle comme dans le cas où les données sont contaminées, c'est-à-dire lorsqu'une partie d'entre elles provient d'une autre loi. Les méthodes non paramétriques visent la construction de procédures robustes pour les problèmes de test ou d'estimation. En ce qui concerne l'estimation, on pourra se reporter à « Non parametric statistical methods » Hollander-Wolf, Editions Wiley.

9.2.2. Test de la médiane et test du signe

Soit X_1, \dots, X_n n v.a.r. indépendantes, de même loi, de fonction de répartition continue F . La médiane m , de la loi est alors définie par

$$F(m) = 50\%$$

lorsque F est strictement croissante.

Supposons F inconnue. On peut tester « $m = m_0$ » où m_0 est une valeur fixée à l'avance, en considérant la variable $Z = \text{nombre de v.a.r. qui sont supérieurs à } m_0$; en effet cette variable suit une loi binomiale dont le paramètre est égal à $1 - F(m_0)$. Considérons l'exemple d'un test bilatère effectué sur 10 observations :

$$H_0 : \langle m = 1 \rangle \text{ contre } H_1 : \langle m \neq 1 \rangle$$

Z suit une loi $\mathcal{B}(10, p)$ avec $p = P(X > 1)$; sous H_0 , p est égal à 0,5 tandis que sous H_1 p est différent de 0,5. On aboutit donc à un test bilatère sur p . La région de rejet de l'hypothèse H_0 , D , est égale à $\{Z \leq k\} \cup \{Z \geq 10 - k\}$ avec k le plus grand entier tel que

$$P_{H_0}(Z \leq k) \leq \alpha/2$$

si α est le niveau du test.

Utilisons la même méthode pour comparer deux échantillons appariés $(X_1, \dots, X_n), (Y_1, \dots, Y_n)$; on travaille sur les différences $X - Y$ et le test s'appelle alors test du signe pour une raison qui va apparaître immédiatement. Reprenons l'exemple du médicament en appelant m la médiane de $X - Y$:

$$H_0 = \langle m = 0 \rangle, \quad H_1 = \langle m < 0 \rangle \text{ et } Z = \text{nombre de } X - Y \text{ positifs}$$

Il s'agit d'un test unilatère sur m .

Sous l'hypothèse H_0 , Z suit une loi $\mathcal{B}(30, 0,5)$ tandis que sous l'hypothèse H_1 , il s'agit d'une loi $\mathcal{B}(30, p)$ avec $p = 1 - F(0)$ qui est strictement supérieur à 0,5.

$D = \{Z \geq k\}$ avec k le plus petit entier tel que $P\{Z \geq k\} \leq \alpha$, α étant le niveau que l'on s'est choisi pour le test.

Pour $\alpha = 5\%$, on trouve $k = 20$, c'est-à-dire $k/n = 66,6\%$.

On rejette l'hypothèse H_0 « le nouveau médicament n'a pas d'effet » si on a un taux de réussite supérieur ou égal à 66,6%.

L'erreur de première espèce représenté la probabilité de décider que le nouveau médicament est efficace à tort (dans ce cas, c'est le hasard qui aurait donné un taux de réussite supérieur à 66,6%).

9.2.3. Le test signe et rang de Wilcoxon

Il s'agit encore d'un test sur la médiane, m , de la loi, test qui est généralement employé pour comparer deux échantillons appariés (exemple des médicaments).

L'hypothèse H_0 est du type « $m = m_0$ », l'hypothèse H_2 peut être unilatère, $m > m_0$ ou $m < m_0$, ou bilatère.

On suppose la fonction de répartition continue et la loi symétrique par rapport à sa valeur médiane (voir exercice 9.5). Soit (X_1, X_2, \dots, X_n) le n -échantillon ; on ordonne $|X_1 - m_0|, \dots, |X_n - m_0|$ dans l'ordre croissant, soit R_i le rang de $|X_i - m_0|$, et on forme le tableau des signes et des rangs :

Variable	$X_1 \dots$
Rang de $ X_i - m_0 $	$R_1 \dots$
Signe de $X_i - m_0$	$+ \dots$

La variable W définie par

$$W = \sum_{i=1}^n R_i 1_{\{X_i - m_0 > 0\}}$$

construite par Wilcoxon, prend en compte à la fois les signes et les grandeurs. Sa loi, sous l'hypothèse H_0 , ne dépend pas de la fonction de répartition choisie mais uniquement de la taille de l'échantillon.

La démarche est ensuite la même que pour le test du signe. On choisit un niveau α .

$H_1 : \langle m > m_0 \rangle$: On rejette H_0 pour $W \geq k$
 k est le plus petit entier tel que $P_{H_0}(W \geq k) \leq \alpha$.

$H_1 : \langle m < m_0 \rangle$: On rejette H_0 pour $W \leq k$
 k est le plus grand entier tel que $P_{H_0}(W \leq k) \leq \alpha$.

$H_1 : \langle m \neq m_0 \rangle$: On rejette H_0 pour $W \leq k_1$ ou $W \geq k_2$ avec

$$k_2 = \frac{n(n+1)}{2} - k_1$$

et k_1 plus grand entier tel que $P_{H_0}(W \leq k_1) \leq \frac{\alpha}{2}$.

Pour comparer deux échantillons appariés (X_1, \dots, X_n) (Y_1, \dots, Y_n) on forme les différences $X_i - Y_i$ et on teste

H_0 : « la médiane de $X - Y$ est nulle »

De nouveau le test peut être unilatère ou bilatère.

Le test est utilisé pour de petits échantillons lorsque l'hypothèse d'une loi gaussienne, par exemple, paraît abusive.

9.2.4. Comparaison de deux échantillons indépendants (Test de Mann Whitney)

Soit (X_1, X_2, \dots, X_n) un n -échantillon d'une loi de fonction de répartition F et (Y_1, Y_2, \dots, Y_m) un m -échantillon, indépendant du premier, d'une loi de fonction de répartition G , F et G sont inconnues.

Nous voulons tester l'hypothèse H_0 : « les deux échantillons proviennent de la même loi », c'est-à-dire,

$H_0 : \langle F = G \rangle$ contre $H_1 : \langle F \neq G \rangle$

Supposons $m \geq n$ et appelons U la variable égale au nombre de fois où une variable Y dépasse une variable X :

$$U = \sum_{i,j} 1_{\{Y_j > X_i\}}$$

Sous l'hypothèse H_0 , la variable U , introduite par Mann et Whitney, a une loi qui ne dépend que de n et m , et pas de la fonction de répartition F . La région de rejet de H_0 est de la forme $\{U \leq k_1\} \cup \{U \geq k_2\}$ avec

$$P_{H_0}(U \leq k_1) + P_{H_0}(U \geq k_2) \leq \alpha$$

On remarque que si l'on pose

$$V = \sum_{i,j} 1_{\{X_i > Y_j\}}$$

alors

$$\begin{cases} U + V = nm \\ P(U \geq k) = P(V \leq nm - k) \end{cases}$$

Ce test peut aussi être utilisé de façon unilatère pour tester une hypothèse de translation sur la loi, sans préjuger de la forme de la loi :

$$H_0 : \langle F = G \rangle \text{ et } H_1 : \langle G(t) < F(t) \forall t \rangle$$

F est inconnue et on teste un déplacement de la loi vers la droite (fig. 9.3). Le test devient unilatère :

$$\text{rejet de } H_0 \Leftrightarrow U \geq k$$

k est le plus petit entier tel que $P_{H_0}(U \geq k) \leq \alpha$.

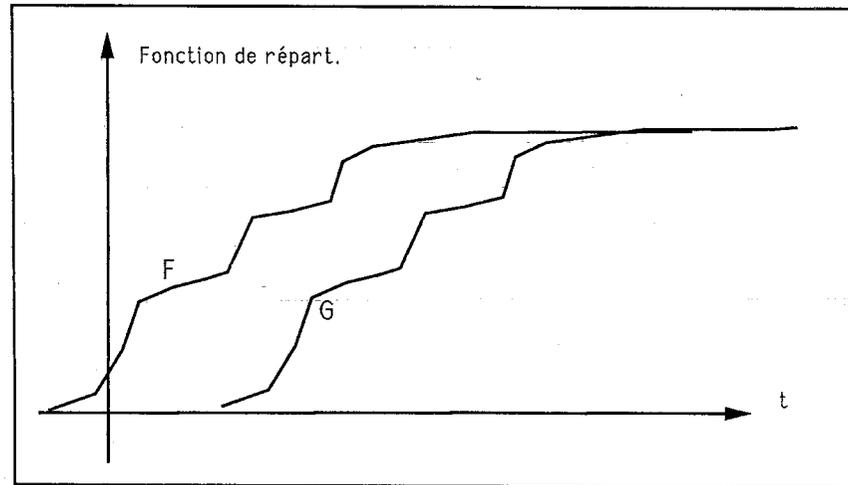


Figure 9.3.

De nouveau, il s'agit d'un test que l'on utilise pour de petits échantillons lorsque l'hypothèse d'une loi gaussienne n'est pas adaptée au problème ou abusive.

Remarque : Il existe un autre test non paramétrique pour comparer deux échantillons indépendants, construit par Kolmogorov et Smirnov. Dans ce test on compare les fonctions de répartition empiriques associées aux deux échantillons, F_n et G_m , en prenant

$$\sup_t |F_n(t) - G_m(t)|$$

9.2.5. Test d'indépendance entre deux v.a.r. X et Y (Spearman)

Soit (X, Y) un couple de v.a.r. continues.

Pour tester l'hypothèse H_0 « X et Y sont indépendantes » à l'aide d'un n -échantillon $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, Spearman a construit une variable aléatoire basée sur les rangs ; on ordonne X_1, \dots, X_n dans le sens croissant, et on appelle R_i le rang de X_i ; de même, on ordonne Y_1, \dots, Y_n dans le sens croissant et on appelle S_i le rang de Y_i .

La variable r_s , de Spearman, est définie par

$$r_s = 1 - \frac{6 \sum_i (R_i - S_i)^2}{n(n^2 - 1)}$$

r_s est comprise entre -1 et 1 . Sous l'hypothèse H_0 , r_s doit être proche de 0 d'où la région de rejet :

$$\text{rejet de } H_0 \Leftrightarrow |r_s| \geq a$$

a est calculé de façon à avoir un niveau inférieur ou égal à α , le plus proche possible de α :

$$P_{H_0}(|r_s| \geq a) \leq \alpha$$

La loi r_s est tabulée, mais pour $n \geq 11$ on utilise une approximation : $r_s \sqrt{\frac{n-2}{1-r_s^2}}$ suit approximativement une loi de Student à $n-2$ degrés de liberté, sous l'hypothèse H_0 .

9.3. MODÈLES LINÉAIRES

9.3.1. La régression

Dans la méthode de régression, on cherche à mettre en évidence une relation entre la variable à étudier, Y , et une, ou plusieurs, variables explicatives : relation entre le chiffre d'affaires et le nombre de clients, l'indice des prix et le temps...

La variable Y , qui est la variable à expliquer, s'appelle variable exogène ou dépendante. On cherche une relation

$$y = f(x)$$

où f dépend linéairement de un ou plusieurs paramètres inconnus. L'exemple le plus simple est celui de la régression linéaire simple où on cherche les coefficients a et b (équation d'une droite),

$$y = ax + b \quad (1)$$

au vu des observations (x_i, y_i) .

On utilise la méthode des moindres carrés :

$$\inf_{a,b} \sum_{i=1}^n [y_i - (ax_i + b)]^2$$

la méthode des moindres carrés peut tout aussi bien être utilisée pour ajuster les données sur un polynôme (fig. 9.4) :

$$y = a + bx + cx^2$$

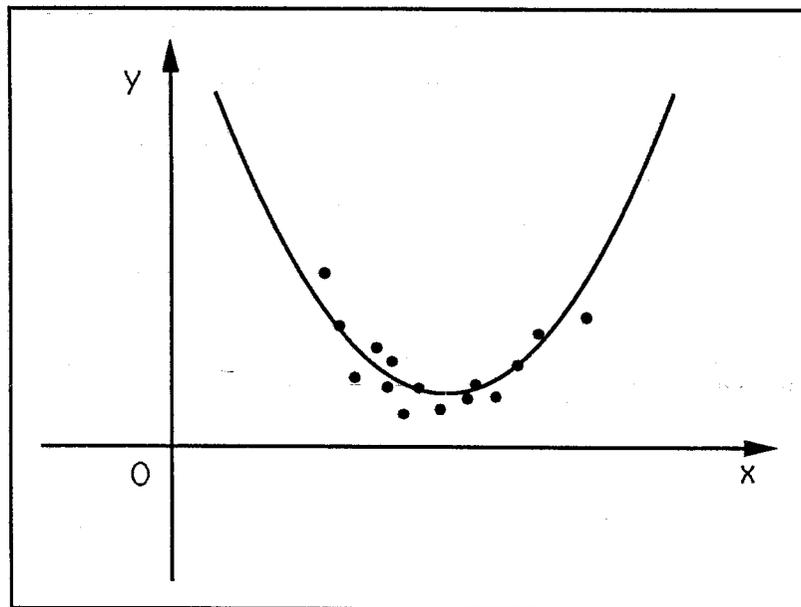


Figure 9.4.

ou sur des fonctions trigonométriques

$$y = a + b \cos x + c \sin x$$

ou toute combinaison de fonctions connues, par exemple

$$y = af_1(x) + bf_2(x) \tag{2}$$

Pour la régression linéaire, on obtient

$$\begin{cases} \hat{b} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \\ \hat{a} = \bar{y} - \hat{b}\bar{x} \end{cases}$$

avec $\bar{x} = \frac{x_1 + \dots + x_n}{n}$ et $\bar{y} = \frac{y_1 + \dots + y_n}{n}$

De façon plus générale, dans le modèle (2), les coefficients \hat{a} , \hat{b} obtenus par les moindres carrés sont solutions du système linéaire

$$\begin{cases} \frac{\partial}{\partial a} \sum_i [y_i - af_1(x_i) - bf_2(x_i)]^2 = 0 \\ \frac{\partial}{\partial b} \sum_i [y_i - af_1(x_i) - bf_2(x_i)]^2 = 0 \end{cases}$$

c'est-à-dire

$$\begin{pmatrix} \sum_i y_i f_1(x_i) \\ \sum_i y_i f_2(x_i) \end{pmatrix} = \begin{pmatrix} \sum_i [f_1(x_i)]^2 & \sum_i f_1(x_i) f_2(x_i) \\ \sum_i f_1(x_i) f_2(x_i) & \sum_i [f_2(x_i)]^2 \end{pmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$$

Pour tout modèle $y = f(x)$ où f dépend linéairement de plusieurs paramètres inconnus, la méthode des moindres carrés aboutit à la résolution d'un système linéaire dont la matrice est symétrique.

L'introduction d'un modèle probabiliste permet de construire des intervalles de confiance, faire des prédictions, faire des tests d'hypothèses...

Etudions le cas d'un phénomène qui évolue au cours du temps : Y_i est la mesure de Y à l'instant t_i .

L'expérience peut être répétée, de sorte que les instants t_i ne sont pas nécessairement tous distincts.

Modèle de la régression linéaire simple :

$$\begin{aligned} Y_i &\text{ est de loi } \mathcal{N}(at_i + b, \sigma^2) \\ Y_1, \dots, Y_n &\text{ sont indépendantes} \end{aligned}$$

les coefficients a et b sont inconnus. On remarque que dans la méthode des moindres carrés, les coefficients \hat{a} et \hat{b} sont obtenus en cherchant la projection orthogonale du vecteur $\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$ sur le sous-espace vectoriel de \mathbb{R}^n , V , engendré

par $\begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}$ et $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$.

On peut alors écrire

$$\begin{aligned} Y_i &= at_i + b + \varepsilon_i \\ \left\| \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \right\|^2 &= \left\| \text{proj}_V^\perp \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \right\|^2 + \left\| \text{proj}_V \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \right\|^2 \\ \hat{a} \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} + \hat{b} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} &= a \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} + b \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \text{proj}_V^\perp \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \end{aligned}$$

Il en résulte que $\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$ est un vecteur gaussien, centré sur $\begin{pmatrix} a \\ b \end{pmatrix}$, indépendant de $\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} - \hat{a} \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} - \hat{b} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ qui s'écrit comme la somme d'un vecteur déterministe et du vecteur aléatoire $\text{proj}_{V^\perp}(\varepsilon)$; on a de plus

$$\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} - \hat{a} \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} - \hat{b} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + (\hat{a} - a) \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} + (\hat{b} - b) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n [Y_i - (\hat{a}t_i + \hat{b})]^2 + (a - \hat{a})^2 \sum (t_i - \bar{t})^2 + n[(a - \hat{a})\bar{t} + (b - \hat{b})]^2$$

$\frac{1}{\sigma^2} \sum_{i=1}^n \varepsilon_i^2$ suit une loi χ_n^2 . On admettra que la variable $\sum [Y_i - \hat{a}t_i - \hat{b}]^2$ appelée variable résiduelle, suit une loi $\sigma^2 \cdot \chi_{n-2}^2$.

On en déduit une estimation de σ^2 :

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n [Y_i - \hat{a}t_i - \hat{b}]^2$$

On peut tester l'hypothèse $H_0: a = 0$ et $b = 0$ en considérant

$$Z = \frac{(a - \hat{a})^2 \sum (t_i - \bar{t})^2 + n[(a - \hat{a})\bar{t} + (b - \hat{b})]^2}{\sum_i [Y_i - \hat{a}t_i - \hat{b}]^2} \times \frac{n-2}{2}$$

qui suit une loi $F_{2, n-2}$ sous l'hypothèse H_0 .

La région de rejet de H_0 est du type

$$\{Z > A\}$$

En fait, le test que l'on fait généralement est celui de l'hypothèse « $a = 0$ » contre « $a \neq 0$ »; sous l'hypothèse H_0 , il n'y a pas d'effet du temps. A cet effet, on calcule la variance de \hat{a} :

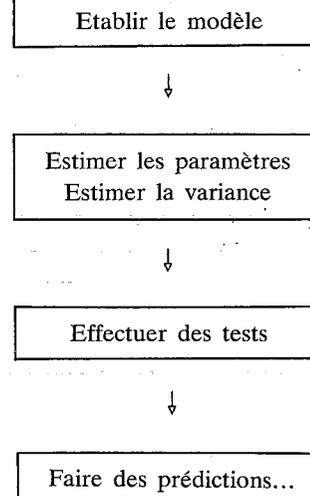
$$\text{var}(\hat{a}) = \sigma^2 \times \frac{1}{\sum_i (t_i - \bar{t})^2}$$

On utilise un test de Student, en remplaçant σ^2 par l'estimateur trouvé ci-dessus:

$$\text{rejet de l'hypothèse } a = 0 \Leftrightarrow \sqrt{\sum (t_i - \bar{t})^2} \cdot \frac{|\hat{a}|}{\sqrt{\hat{\sigma}^2}} \geq B$$

B est choisi de façon à avoir un niveau égal à α fixé à l'avance (dans une table de loi de Student à $n-2$ degrés de liberté).

Ce qui précède n'est qu'une introduction à la méthode de régression. On retiendra que la méthode s'effectue en plusieurs étapes:



La méthode est particulièrement intéressante lorsqu'il y a plusieurs variables explicatives; on parle alors de régression linéaire multiple: on ajuste les données sur la surface d'équation

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_k x_k$$

On peut alors chercher un sous-ensemble de variables qui explique bien la variable y , c'est-à-dire tester la nullité d'un sous-ensemble de coefficients.

D'autre part, on pourra chercher à ajuster sur un modèle qui ne dépend pas linéairement des coefficients inconnus. Pour tous ces développements sur la régression on pourra consulter le livre « La régression » de R. Tomassone, E. Lesquoy et C. Millier, Ed. Masson.

9.3.2. Introduction à l'analyse de variance

Comme dans le paragraphe précédent, il s'agit de modèles gaussiens. L'analyse de la variance permet d'expliquer une variable Y par une, ou des, variables qui ne sont pas traitées quantitativement: par exemple, le rendement d'une parcelle de terrain en fonction du type d'engrais utilisé, du type de semence...

Nous ne traitons ici que le cas de l'analyse de variance à *un facteur*, c'est-à-dire d'une seule variable explicative; le nombre de modalités pour la variable explicative est appelé le *nombre de niveaux*. Soit k le nombre de niveaux.

Notations : $Y_{i,j}$ est la valeur de Y pour le i -ième niveau du facteur ; j est un indice de répétition
 n_i est le nombre de répétitions de l'expérience pour le i -ième niveau

$$Y_{i..} = \frac{1}{n_i} (Y_{i,1} + \dots + Y_{i,n_i})$$

$$Y_{..} = \frac{1}{n} \sum_{i,j} Y_{i,j} \text{ avec } n = \sum n_i$$

Modèle : $Y_{i,j}$ est de loi $\mathcal{N}(\mu_i, \sigma^2)$. Les variables $Y_{i,j}$ sont indépendantes.

D'un niveau à l'autre, l'espérance de la variable change si les μ_i sont distincts. On testera l'effet global du facteur en testant l'hypothèse H_0 : « $\mu_1 = \mu_2 = \dots = \mu_k$ » contre son contraire. Comme dans le cas de la régression linéaire, les coefficients inconnus sont estimés par la méthode des moindres

carrés ce qui revient à trouver la projection orthogonale du vecteur $\begin{pmatrix} Y_{1,1} \\ Y_{1,n_1} \\ Y_{2,1} \\ \vdots \\ Y_{k,n_k} \end{pmatrix}$

sur le sous-espace vectoriel engendré par les k vecteurs

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

qui ont respectivement n_1 fois 1, ..., n_k fois 1. On obtient

$$\hat{\mu}_i = Y_{i..}$$

$$\frac{1}{\sigma^2} \sum_{i,j} (Y_{ij} - \hat{\mu}_i)^2 \text{ suit une loi } \chi^2_{n-k}$$

Exemple : On veut comparer trois engrais A, B, C. Pour chaque engrais, cinq parcelles sont traitées dans les mêmes conditions et on obtient

Engrais A : 41, 38, 39, 40, 40
 Engrais B : 36, 37, 38, 37, 39
 Engrais C : 38, 40, 37, 41, 41

$$\hat{\mu}_1 = 39,6 \quad \hat{\mu}_2 = 37,4 \quad \hat{\mu}_3 = 39,4 \quad \hat{\sigma}^2 = 1,96$$

Pour tester l'hypothèse H_0 : « Les trois engrais ont le même effet sur le rendement » on compare les sommes résiduelles provenant du modèle général

d'une part, et du modèle $\mu_1 = \mu_2 \dots$ d'autre part ; si les espérances sont égales, l'espérance commune est estimée par la moyenne empirique de toutes les variables, $Y_{..}$, et la somme résiduelle est alors $\sum_{i,j} (Y_{ij} - Y_{..})^2$. On admettra que

sous l'hypothèse H_0 , la variable

$$\frac{n-k}{k-1} \cdot \frac{\sum_i n_i (Y_{i..} - Y_{..})^2}{\sum_{i,j} (Y_{i,j} - Y_{i..})^2}$$

suit une loi $F_{k-1, n-k}$ (Loi de Fisher) d'où la règle de décision :

$$\text{Rejet de } H_0 \Leftrightarrow \frac{n-k}{k-1} \cdot \frac{\sum_i n_i (Y_{i..} - Y_{..})^2}{\sum_{i,j} (Y_{i,j} - Y_{i..})^2} \geq A$$

A étant choisi de façon à avoir un niveau égal à α que l'on s'est fixé à l'avance. Sur l'exemple précédent, on obtient les valeurs suivantes

$$\sum_{i,j} (Y_{i,j} - Y_{i..})^2 = 23,6$$

$$\sum_i n_i (Y_{i..} - Y_{..})^2 = 14,8$$

$$Y_{..} = 38,8$$

$$\frac{n-k}{k-1} \cdot \frac{\sum_i n_i (Y_{i..} - Y_{..})^2}{\sum_{i,j} (Y_{i,j} - Y_{i..})^2} = 3,76$$

Pour un niveau α égal à 5 %, on trouve dans une table de Loi de Fischer, qu'une variable $F_{2,12}$ à 95 % de chances d'être inférieure à 3,88, 5 % de chances d'être supérieure. On accepte donc l'hypothèse H_0 , « il n'y a pas effet de l'engrais utilisé ».

Si on effectue la comparaison des engrais A et B, sans tenir compte de l'engrais C, on utilise un test de Student ; le nombre de degré de liberté est égal à 8 car σ^2 est alors estimé par

$$\frac{1}{8} \left[\sum_j (Y_{1j} - Y_{1.})^2 + \sum_j (Y_{2j} - Y_{2.})^2 \right]$$

La variable de Student est :

$$\frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{2}{5} \sqrt{\hat{\sigma}^2}}} = 2,48$$

Pour un test unilatère au niveau 5 %, la valeur critique est égale à 2,3. La valeur obtenue, 2,48, est à l'extérieur de l'intervalle $]-\infty, 2.3]$ où on accepte l'hypothèse $H_0: \mu_1 = \mu_2$. On décide donc « $\mu_1 > \mu_2$ » ; $\hat{\mu}_1$ est significativement supérieur à $\hat{\mu}_2$.

On trouvera ces méthodes dans « Théorie et méthodes statistiques » de P. Dagnelie, Editions Presses Agronomiques de Gembloux.

EXERCICES

E.9.1. M. Dupont dépose 6 lettres à la poste. En supposant les boîtes à lettres vides avant son passage, combien il y a-t-il de possibilités pour le triplet (N_1, N_2, N_3) en appelant N_1 le nombre de lettres dans la boîte « PARIS », N_2 le nombre de lettres dans la boîte « banlieue » et N_3 le nombre de lettres dans la boîte « PROVINCE » ? En admettant

$$P(N_1 = k_1 \text{ et } N_2 = k_2 \text{ et } N_3 = k_3) = \frac{6!}{k_1! k_2! k_3!} \left(\frac{1}{3}\right)^6$$

si $k_i \geq 0$ et $\sum_i k_i = 6$, calculer $P(N_1 = N_2 = N_3)$.

On pose

$$D = \frac{1}{2} [(N_1 - 2)^2 + (N_2 - 2)^2 + (N_3 - 2)^2]$$

En utilisant des valeurs $k_1 \leq k_2 \leq k_3$, calculer la loi de D.

E.9.2. 56 personnes sont interrogées sur leur couleur préférée. Peut-on considérer que les trois couleurs Rouge, Bleu, Vert sont aussi attractives si on obtient les résultats suivants :

Rouge : 17 Bleu : 22 Vert : 17

E.9.3. Au cours d'une élection présidentielle, 3 candidats notés A, B, C se présentent. Un sondage est effectué dans trois villes :

Ville :	1	2	3
Candidat			
A	48	41	36
B	32	45	26
C	23	40	22

Peut-on considérer qu'il y a indépendance entre l'appartenance à une ville et la façon de voter ?

E.9.4. Une particule radioactive à une durée de vie qui est considérée comme une v.a. X de loi exponentielle de paramètre λ . Calculer l'espérance et la médiane (la médiane s'appelle alors le temps de $\frac{1}{2}$ vie).

E.9.5. Soit X une v.a.r. dont la loi est symétrique par rapport à l'origine, c'est-à-dire

$$P(X \in A) = P(X \in -A)$$

On suppose la fonction de répartition, F, continue. Montrer que l'on a $F(-t) = 1 - F(t)$.

Calculer la fonction de répartition de $|X|$. Montrer que $\text{sign}(X)$ et $|X|$ sont indépendantes. Que peut-on en déduire pour les variables $\text{sign}(X_1), \dots, \text{sign}(X_n), |X_1|, \dots, |X_n|$ lorsque X_1, \dots, X_n est un n -échantillon ?

E.9.6. Soit (X_1, \dots, X_n) et (Y_1, \dots, Y_m) deux échantillons indépendants d'une même loi et U la variable du test de Mann-Withney

$$U = \sum_{i,j} 1_{\{X_i > Y_j\}}$$

Quel est l'ensemble des valeurs possibles pour U ? Montrer qu'on a la relation

$$U = \sum_{i=1}^n \text{rang}(X_i) - \frac{n(n+1)}{2}$$

où le rang de X_i est calculé en considérant les $n+m$ variables dans leur ensemble, et en ordonnant dans le sens croissant.

E.9.7. On reprend les mêmes notations qu'à l'exercice précédent en supposant la fonction de répartition, F, continue. On pose

$$A_i = F(X_i) \quad \text{et} \quad B_j = F(Y_j)$$

Montrer que les variables $A_1, \dots, A_n, B_1, \dots, B_m$ sont indépendantes, de loi uniforme sur $[0, 1]$ et

$$U = \sum_{i,j} 1_{\{A_i > B_j\}}$$

En déduire que la loi de U ne dépend pas de F.

E.9.8. On considère un échantillon de taille 11, de densité $f_\theta(x) = \frac{1}{\pi} \times \frac{1}{1 + (x - \theta)^2}$ et le test

$$H_0: \theta = 0 \quad \text{contre} \quad H_1: \theta > 0$$

Construire le test de niveau 5 % ayant pour région de rejet $\{\bar{X} \geq A\}$.

Calculer la fonction puissance $\beta(\theta)$ et comparer avec celle obtenue avec le test du signe de même niveau (on pourra comparer les pentes en $\theta = 0$).

E.9.9. Une société publie les résultats des exercices de 1984 à 1988 ; parmi ces résultats, figurent le nombre de salariés et le montant des sommes versées au titre des avantages sociaux (sécurité sociale, ...).

Estimer les coefficients a et b , de la régression linéaire en appelant Y le montant des sommes versées

	1984	1985	1986	1987	1988
Montant	254 698	345 450	504 261	547 994	574 213
Nombre de salariés	233	255	279	297	319

Estimer σ^2 et tester l'hypothèse $a = 0$.

CORRIGÉ DES EXERCICES

Chapitre 1

$$\text{E.1.1. } A = [B_1 \cap B_2] \cup B_3 \cup [B_4 \cap B_5 \cap B_6]$$

$$\text{E.1.2. } A \cap B = \{\text{personnage du carreau}\} \neq \emptyset$$

$$A \cap C = \{\text{personnage à trèfle ou à pique}\} \neq \emptyset$$

$$B \cap C = \emptyset$$

$$P(A) = 3/13 \quad P(B) = 1/4 \quad P(C) = 1/2$$

$$P(A \cap B) = 3/52 \quad P(A \cap C) = 3/26 \quad P(B \cap C) = 0$$

E.1.3. $P(A \cap B) = P(A)P(B)$; A et B sont donc indépendants ce qui confirme l'intuition.

$$\text{E.1.4. } P(A \cap B^c) + P(A \cap B) = P(A)$$

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A)[1 - P(B)] = P(A)P(B^c)$$

A et B^c sont donc indépendants ; il en sera donc de même pour A^c et B , A^c et B^c .

E.1.5. Soit B_1, \dots, B_{k-1} une suite d'événements tels que $B_i = A_i$ ou A_i^c

$$P(B_1 \cap B_2 \dots \cap B_{k-1}) =$$

$$= P(B_1 \cap B_2 \dots \cap B_{k-1} \cap A_k) + P(B_1 \cap B_2 \dots \cap B_{k-1} \cap A_k^c)$$

B_1, \dots, B_{k-1} sont donc indépendants et toute sous famille finie sera aussi constituée d'événements indépendants.

E.1.6. On doit avoir $a = \frac{P(A)}{P(A) + P(B)}$... Il est donc nécessaire d'avoir $(1-a)(1-b)(1-c) = abc$. Réciproquement si cette condition est remplie, en prenant

$$P(A) = \frac{1}{1 + \frac{1-a}{a} + \frac{1-b}{b} \times \frac{1-a}{a}}$$

on a bien $a = \frac{P(A)}{P(A) + P(B)}$... et $P(A) + P(B) + P(C) = 1$.

E.1.7.

$$P(\text{rouge}) = P(\text{rouge/boîte 1})P(\text{boîte 1}) + P(\text{rouge/boîte 2})P(\text{boîte 2}) = 31/60$$

E.1.8.

$$P(\text{atteint/test positif}) = (0,05 \times 0,9) / [0,05 \times 0,9 + 0,95 \times 0,15] = 31,6 \%$$

E.1.9.

$$P(1 \text{ carte rouge et } 1 \text{ carte noire}) = 2 \left[\frac{1}{2} \times \frac{26}{51} \right] = \frac{26}{51} = 50,98 \%$$

$$P(\text{roi rouge et roi noir}) = 2 \left[\frac{2}{52} \times \frac{2}{51} \right] = \frac{2}{13 \times 51} = 0,3 \%$$

$$P(2 \text{ rois}/1 \text{ rouge et } 1 \text{ noir}) = 1/169 = 0,6 \%$$

E.1.10.

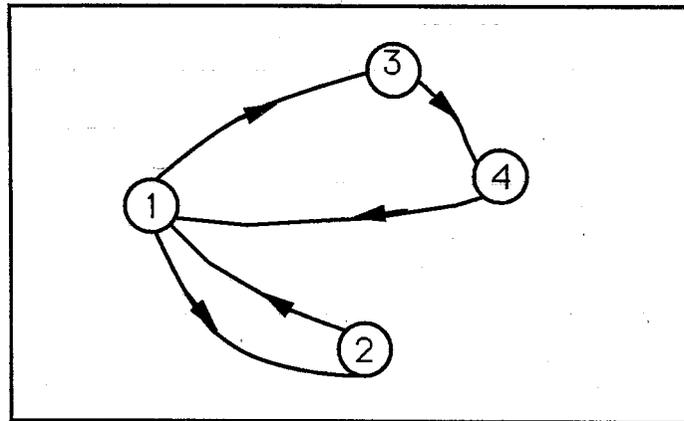


Figure E.1.10.

E.1.11. Si au temps t , il y a n messages en attente (en conflit), chacun essaie de repasser au temps $t + 1$. Le nombre de messages en attente diminue de un, à condition qu'il n'y ait pas d'arrivée et qu'un seul émette.

Si un seul nouveau message se présente, il passe sur le canal de transmission à condition que les messages en attente n'émettent pas.

Chapitre 2

E.2.1. Y suit une loi uniforme sur $\{0, 1, 2\}$ (proba. $1/3$ pour chaque valeur).E.2.2. $\frac{P(X = k + 1)}{P(X = k)} > 1 \Leftrightarrow k \leq 5$. Le maximum a lieu pour $k = 6$.E.2.3. $P(X = 1) = 11/12$; $P(X = 2) = 1/12$;
 $P(Y = -1) = 5/6$; $P(Y = 10) = 1/12$; $P(Y = 100) = 1/72$;
 $P(Y = -10) = \frac{5}{72}$ E.2.4. $P(aX < t) = P\left(X < \frac{t}{a}\right) = 1 - e^{-\lambda \frac{t}{a}}$ si $a > 0$.

aX suit une loi exponentielle de paramètre $\frac{\lambda}{a}$ et λX suit une loi exponentielle de paramètre λ .

On trouve $M = \text{Ln}(2)/\lambda$ E.2.5. Y suit une loi $\mathcal{B}\left(N, \frac{1}{2}\right)$ qui est d'espérance $\frac{N}{2}$.E.2.6. $0 \leq t \leq \frac{1}{3}$ $P(X < t) = 2t$

$$\frac{1}{3} < t \leq \frac{2}{3} \quad P(X > t) = \frac{2}{3} - t \Rightarrow P(X \leq t) = t + \frac{1}{3}$$

$$\text{Densité : } f(t) = 2 \quad \text{si } 0 \leq t \leq \frac{1}{3}$$

$$= 1 \quad \text{si } \frac{1}{3} \leq t \leq \frac{2}{3}$$

E.2.7. $P(X \leq t) = \frac{1}{2} + \frac{1}{\pi} \text{Arctg } t$

$$a = -1 \quad b = \sqrt{3}/3$$

$$c_1 = -3,08 \quad c_2 = -1,38 \quad c_3 = -0,73 \quad c_4 = -0,32 \quad c_5 = 0$$

les autres valeurs sont obtenues par symétrie par rapport à l'origine.

E.2.8. $X = \text{Inf}(s, R_1)$
 $P(X = s) = P(R_1 > s) = e^{-\lambda s}$. Si $0 \leq t < s$, on a

$$P(X < t) = P(R_1 < t) = 1 - e^{-\lambda t}$$

E.2.9. $R(t)$ est strictement décroissante lorsque $\beta > 0$.

Pour $\beta = 1$, on retrouve une loi exponentielle (le paramètre α est strictement positif).

E.2.10. $\frac{X - m}{\sigma}$ suit une loi $\mathcal{N}(0, 1)$. On en déduit

$$A = \sigma \times 1,64$$

E.2.11. $P(e^X < t) = 0$ si $t \leq 0$ et $P(e^X < t) = P(X < \text{Log } t)$ sinon ; la densité est donc nulle si $t \leq 0$ et vaut pour $t > 0$

$$\frac{d}{dt} (P(X < \log t)) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(\log t - m)^2}{\sigma^2}\right] \times \frac{1}{t}$$

E.2.12. Soit $F(t)$ la fonction de répartition. Pour $t \geq 0$ $F(t) = 1 - R(t)$. La densité est donc égale à $\beta \frac{t^{\beta-1}}{\alpha^\beta} e^{-\frac{t^\beta}{\alpha^\beta}}$.

$$E(X) = \int_0^{+\infty} \alpha u^{1/\beta} e^{-u} du = \alpha \Gamma\left(\frac{1}{\beta} + 1\right)$$

E.2.13. $P(X = 1 \text{ et } Y = -1) = 5/6$

		Y			
		-10	-1	10	100
X	+1	0	5/6	1/12	0
	2	5/72	0	0	1/72

X et Y ne sont pas indépendantes en raison des zéros dans le tableau de probabilité.

$$E(Y) = 50/72$$

E.2.14. Soit T la durée de vie. $T = \inf(X, Y)$ si les éléments sont disposés en série, $T = \sup(X, Y)$ si les éléments sont disposés en parallèle :

$$P(\inf(X, Y) > t) = P(X > t \text{ et } Y > t) = e^{-\lambda t} e^{-\mu t}$$

$\inf(X, Y)$ suit encore une loi exponentielle ; le paramètre est la somme des deux paramètres.

$$P(\sup(X, Y) < t) = P(X < t) e^{-\mu t} = (1 - e^{-\lambda t})(1 - e^{-\mu t}) \\ = 1 - e^{-\lambda t} - e^{-\mu t} + e^{-(\lambda + \mu)t}$$

Il ne s'agit plus d'une loi exponentielle ; la densité est

$$\lambda e^{-\lambda t} + \mu e^{-\mu t} - (\lambda + \mu) e^{-(\lambda + \mu)t}$$

E.2.15.

$$E(X) = \int_{-\infty}^{+\infty} x \times \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} [-x e^{-x^2/2}]_{-\infty}^{+\infty} = 0$$

Pour calculer $E(X^2)$ on intègre par parties en posant $u = x$ $V' = x e^{-x^2/2}$. On trouve $E(X^2) = 1$.

De même, pour le calcul de $E(X^4)$ on pose $u = x^3$ et $V' = x e^{-x^2/2}$ et on trouve $E(X^4) = 3$; d'où $E[(X - m)^4] = 3 \sigma^4$ lorsque X est de loi $\mathcal{N}(m, \sigma^2)$.

E.2.16. Si $z = a + ib$

$$|e^{-zx} e^{-x^2/2}| = e^{-ax} e^{-x^2/2} = e^{-(x+a)^2/2} \cdot e^{a^2/2}$$

Pour $b = 0$, on a $I(z) = \sqrt{2\pi} e^{a^2/2} = \sqrt{2\pi} e^{z^2/2}$. La fonction $g(z) = \sqrt{2\pi} e^{z^2/2}$ est analytique sur \mathbb{C} tout entier ; $I(z)$ et $g(z)$ qui sont égales sur l'axe réel sont donc égales sur \mathbb{C} tout entier (principe des zéros isolés pour une fonction analytique).

E.2.17. $X - 2Y$ est de loi $\mathcal{N}(-2, 4)$. Le vecteur $\begin{pmatrix} X - 2Y \\ aX + bY \end{pmatrix}$ est gaussien car il est obtenu par une application linéaire appliquée à un vecteur gaussien ; on cherche a et b de façon à avoir $\text{cov}(X - 2Y, aX + bY) = 0$. On obtient $b = 0$ c'est-à-dire que X est indépendant de $X - 2Y$. On a

$$Y = -\frac{1}{2}(X - 2Y) + \frac{1}{2}X$$

Conditionnellement à $X - 2Y = z$, Y suit une loi

$$\mathcal{N}\left(-\frac{1}{2}z, \frac{1}{4}\right) \text{ et on a } E(Y/X - 2Y) = -\frac{1}{2}(X - 2Y)$$

E.2.18. a) $E(e^{i(tXY + sYZ)}) = E(h(X))$ et

$$h(X) = e^{-1/2 X^2(t^2 + s^2)} \\ E(h(X)) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2(t^2 + s^2)} dx = \frac{1}{1 + t^2 + s^2}$$

$\frac{1}{1 + t^2 + s^2}$ ne peut s'écrire comme le produit d'une fonction de s par une fonction de t , $f(s) g(t)$, sinon on aurait $f(s) g(0) = \frac{1}{1 + s^2}$ et $f(0) g(t) = \frac{1}{1 + t^2}$. Les deux variables XY et XZ ne sont donc pas indépendantes.

b) $\text{cov}(XY, XZ) = E(X^2 YZ) = E(X^2) E(Y) E(Z) = 0$

La covariance est nulle bien que les deux variables ne soient pas indépendantes !

$$E.2.19. P(X = 0/Y \in [y_0, y_0 + \Delta y]) = \frac{P(X = 0 \text{ et } Y \in [y_0, y_0 + \Delta y])}{P(Y \in [y_0, y_0 + \Delta y])}$$

$$P(X = 0 \text{ et } Y \in [y_0, y_0 + \Delta y]) = P(X = 0) P(Y \in [y_0, y_0 + \Delta y] | X = 0) \\ = \frac{1}{3} \int_{y_0}^{y_0 + \Delta y} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

On obtient finalement

$$\lim_{\Delta y \rightarrow 0} P(x = 0/Y \in [y_0, y_0 + \Delta y]) = \frac{e^{-y_0^2/2}}{e^{-y_0^2/2} + 2e^{-(y_0 - 1)^2/2}}$$

Si on observe $Y(\omega) = y_0$, il est raisonnable de prédire pour X la valeur la plus probable : on choisit 0 si $P(X = 0/Y = y_0) \geq \frac{1}{2}$ c'est-à-dire si $y_0 \leq \frac{1}{2} - \text{Ln}(2)$.

Chapitre 3

E.3.1. Soit $\varepsilon > 0$ et $Y_n = X_n - X$

$$P(|Y_n| > \varepsilon) \rightarrow 0 \Leftrightarrow X_n \rightarrow X$$

en probabilité. On traite de même la convergence presque sûre.

E.3.2. Si toutes les v.a. X_n ainsi que X sont à valeurs entières, leurs fonctions de répartition sont constantes entre 2 entiers et

$$P(X_n = k) = F_n\left(k + \frac{1}{2}\right) - F_n\left(k - \frac{1}{2}\right)$$

d'où l'équivalence demandée en utilisant les fonctions de répartition.

Si $X_{n+1} = 1 - X_n$ et $P(X_n = 0) = P(X_n = 1) = 1/2$, X_n converge bien en loi mais ne peut converger en probabilité ; en effet si X_n converge vers X alors X prend ses valeurs dans $\{0, 1\}$ et $P(X_n = X \text{ et } X_{n+1} = X) = 0$ ce qui contredit le fait que $P(X_n = X) \xrightarrow[n \rightarrow +\infty]{} 1$.

De même, dans l'exemple suivant

$$P(X_n = 0 = X_{n+1} = \dots = X_{n+p}) = \prod_{k=n}^{n+p} \left(1 - \frac{1}{k}\right) \xrightarrow[p \rightarrow +\infty]{} 0$$

$$\Rightarrow P(X_n = 0 \text{ à partir d'un certain rang}) = 0$$

E.3.3. S_n est une suite croissante majorée par b et I_n une suite décroissante minorée par a .

Posons

$$S_\infty(\omega) = \lim_n S_n(\omega) \quad \text{et} \quad I_\infty(\omega) = \lim_n I_n(\omega)$$

Si il existe $c \in]a, b[$ tel que

$$P(a \leq S_\infty \leq c) > 0$$

alors

$$P(\forall n \quad U_n \leq c) > 0$$

or

$$P(U_n \leq c) = \frac{c-a}{b-a} < 1 \quad \text{et} \quad P(\forall n \quad U_n \leq c) = 0$$

D'où

$$P(S_\infty \leq c) = 0 \quad \text{et} \quad S_\infty = b \quad \text{p.s.}$$

De même

$$I_\infty = a \quad \text{p.s.}$$

E.3.4. Les variables X_i ont pour espérance l'intégrale ; on peut donc appliquer la loi des grands nombres.

E.3.5. Si X suit une loi $\mathcal{B}(n, p)$ avec p proche de 0,05 et si n est grand, alors $\sqrt{n} \left(\frac{X}{n} - p\right)$ suit approximativement une loi $\mathcal{N}(0, 0,05 \times 0,95)$. On obtient

une v.a. $\mathcal{N}(0, 1)$ en prenant $Y = \frac{\sqrt{n}}{\sqrt{0,05 \times 0,95}} \left(\frac{X}{n} - p\right)$. On veut

$$P\left(\left|\frac{X}{n} - p\right| \leq 10^{-3}\right) \geq 95\%$$

On obtient alors $n \geq 183$.

E.3.6. 1) On regarde les deux cas $X > a$ et $X \leq a$ puis on prend l'espérance de chaque membre de l'inégalité.

2) On applique 1) à $X = \frac{S_n}{n} - p$

$$E\{\exp[t(X - a)]\} = \exp\{n \log[1 - p(e^{t/n} - 1)] - t(p + a)\}$$

Inf $[\log[1 - p + pu] - (\log u)(p + a)]$ est atteint pour $u > 1$

$$u = \frac{(p + a)(1 - p)}{p(1 - p - a)} \quad \text{soit} \quad t = n \log \frac{(p + a)(1 - p)}{p(1 - p - a)} > 0$$

On trouve $P\left(\frac{S_n}{n} - p > a\right) \leq \exp(-nh(a, p)) \xrightarrow[n \rightarrow +\infty]{} 0$ car $h(a, p) > 0$.

E.3.7. En supposant le dé parfaitement équilibré, S_n suit une loi $\mathcal{B}\left(n, \frac{1}{3}\right)$.

Si on utilise le résultat de l'exercice 3.6. (on peut aussi utiliser le théorème de la limite centrale) avec $p = 1/3$ et $p + a = 0,337$, on a

$$P\left(\frac{S_n}{n} \geq 0,337\right) \leq e^{-4n \cdot 10^{-5}} \leq 3,3 \cdot 10^{-6} \quad \text{pour } n = 315\,672$$

Or l'expérience a donné $\frac{S_n(\omega)}{n} = \frac{106\,602}{315\,672} = 0,3376\dots$ l'interprétation de Weldon : la face du 6 est creusée de 6 trous, celle du 5 de 5 trous, ce qui a pour effet de déplacer le centre de gravité qui n'est plus à égale distance des 6 faces.

E.3.8. On part de l'état 0 : $X_0 = 0$. Soit τ le premier instant où on passe par l'état k , alors

$$P_0(I_k = 1) = P_0(\tau < +\infty) \pi_{k, k-1} \times \frac{1}{1 - \pi_{k, k}}$$

$\sum_k P_0(I_k = 1) \leq \sum_k \frac{\pi_{k, k-1}}{1 - \pi_{k, k}}$ qui est une série convergente. Si l'espérance d'une

v.a. positive est finie alors la v.a. est presque sûrement finie. On montre facilement que tous les états communiquent. Ils sont alors tous transients ou tous

récurrents. Supposons tous les états récurrents, alors on montre que l'on aurait pour tous i, k

$$P(\text{on atteint } k/X_0 = i) = 1$$

Presque sûrement, le nombre d'états visités en partant de l'état i , est infini ce qui contredit le résultat obtenu sur $\sum_k I_n$.

E.3.9. a)

$$\begin{aligned} P(X_{n+1} = X_n/X_{n_0} = x) &= P(X_{n+1} = k/X_n = k) P(X_n = k/X_{n_0} = x) \\ &= p_1^k P(X_n = k/X_{n_0} = x) \leq p_1 \end{aligned}$$

$$P(X_n = X_{n+1} \dots = X_{n+k}/X_n = x) = p_1^{kx} \leq p_1^k$$

$$\left\{ \lim_n X_n < +\infty \right\} = \bigcup_k \{X_k = X_{k+1} = \dots\}$$

car la suite est croissante à valeurs entières

$$P(X_k = X_{k+1} = \dots) = \sum_j P(X_k = X_{k+1} \dots / X_k = j) P(X_k = j) = 0$$

$$b) E(t^{X_{n+1}}/X_n = N) = [p_1 t + p_2 t^2 + \dots]^N = [G(t)]^N.$$

En dérivant par rapport à t et en faisant $t = 1$, on obtient

$$E(X_{n+1}/X_n = N) = Nm \text{ soit } E(X_{n+1}/X_n) = mX_n$$

ou encore

$$E\left(\frac{X_{n+1}}{m^{n+1}} \middle/ \frac{X_n}{m^n}\right) = \frac{X_n}{m^n}$$

(X_{n+1} ne dépend du passé que pour le dernier instant).

Chapitre 4

E.4.1. Il y a m^2 façons de choisir (X_n, X_{n-1}) dans $\{0, 1, \dots, m-1\}$. La période est donc au plus égale à m^2 .

E.4.2. 1) On a $P(X = Y = Z) = 0$ car les variables sont à densité :

$$P(X < Y < Z) = P(Y < X < Z) = P(Z < X < Y) \dots$$

(les 6 permutations sont équiprobables) et

$$P(X < Y < Z) + P(Y < X < Z) + \dots = 1$$

d'où le résultat.

2) Supposons $X_{n-1} < X_{n+1} < X_n$.

On aboutit à une contradiction si $X_n + X_{n-1} < m$.

Supposons $X_n + X_{n-1} \geq m$; on a alors

$$X_{n+1} = X_n + X_{n-1} - m$$

on aboutit à $X_n > m$ ce qui est impossible.

E.4.3. Soit $U_1, U_2, \dots, V_1, V_2, \dots$ des variables aléatoires indépendantes, de loi uniforme sur $[0, 1]$. On pose

$$X_i = a + (b-a)U_i \text{ et } Y_i = c + (d-c)V_i \quad Z_i = g(X_i, Y_i)$$

et on applique la loi des grands nombres à \bar{Z} .

E.4.4. On écrit l'intégrale en coordonnées polaires

$$x = 1 + r \cos \theta \quad y = 1 + r \sin \theta \quad g(r, \theta) = f(1 + r \cos \theta, 1 + r \sin \theta) r$$

E.4.5. 1) On utilise la rotation qui envoie $\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$ sur $\begin{pmatrix} d \\ 0 \end{pmatrix}$.

$R\left(\frac{X}{Y}\right)$ a encore comme matrice de covariance, la matrice Id .

2) $P(R < t) = \iint_{\sqrt{x^2+y^2} < t} \frac{1}{2\pi} \exp\left\{-\frac{1}{2}[x^2 + d^2 + y^2 - 2dx]\right\} dx dy$ et on passe en coordonnées polaires.

E.4.6. 1) On prend $p_1 = 1, p_2 = p_3 = \dots = 0$.

2) On fait la démonstration par récurrence sur n en utilisant la formule de convolution pour la densité d'une somme de deux v.a. indépendantes.

On prend $p_1 = p_2 = \dots = p_n = 1$.

$$E.4.7. a) P(V_2 < f(V_1)) = \iint_{\substack{v_2 < f(v_1) \\ a \leq v_1 \leq b \\ 0 \leq v_2 \leq M}} \frac{1}{M} \times \frac{1}{b-a} dv_1 dv_2 = \frac{1}{M(b-a)} = P$$

$$\begin{aligned} b) P(X < t) &= \sum_{n=1}^{+\infty} P(V_{2n} < f(V_{2n-1}) \text{ et } V_{2n-1} < t) \times (1-p)^{n-1} = \\ &= \int_a^t f(v_1) dv_1. \end{aligned}$$

Chapitre 5

E.5.1. Soit M_t le nombre de particules enregistrées au temps t :

$$\begin{aligned} P(M_t = k) &= \sum_{n \geq k} P(N_t = n \text{ et } p_t = k) = \sum_{n \geq k} C_n^k p^k (1-p)^{n-k} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\ &= e^{-\lambda p t} \frac{(\lambda p t)^k}{k!} \end{aligned}$$

On démontre ensuite l'indépendance pour des intervalles de temps disjoints (l'enregistrement est supposé indépendant du phénomène d'émission de particules).

E.5.2. $Y_t - Y_s$ est de loi $\mathcal{N}\left(0, \frac{1}{a}(at - as)\right)$. Les propriétés d'indépendance sont conservées.

E.5.3. Les propriétés d'indépendance sont vérifiées

$$\begin{aligned} \text{var}(Z_t) &= t \quad \text{si } t \in [0, 1] \\ &= \text{var}(X_1) + \text{var}(Y_{t-1}) \quad \text{si } t > 1 \end{aligned}$$

E.5.4. D'après un résultat du paragraphe 3.6, il suffit de regarder la cv de la série de terme général $E(X_n^2)$ avec $X_n = \frac{\sin nt}{n} Y_n$.

E.5.5. Pour $t > s$ $E(X_t) = E(X_s) = 0$ $E(X_t X_s) = \cos \omega t \cos \omega s$. De même pour Y_t : $E(Y_t Y_s) = \sin \omega t \sin \omega s$. X_t et Y_t ne sont pas stationnaires alors que Z_t est (gaussien) stationnaire : $E(Z_t Z_s) = \cos \omega(t-s)$.

E.5.6. $E(X_t) = E(A) E(\cos(\omega t + \theta)) = 0$ car A et θ sont indépendantes

$$\begin{aligned} E(X_t X_s) &= E(A^2) E\left[\frac{1}{2} \cos(\omega t + \omega s + 2\theta) + \frac{1}{2} \cos \omega(t-s)\right] \\ &= \frac{1}{2} \cos \omega(t-s) \end{aligned}$$

X_t est donc stationnaire au second ordre. Il est aussi fortement stationnaire.

E.5.7. Si $s < t$ $X_t - X_s$ est indépendant de X_s , d'où $E(X_t X_s) = \text{Inf}(t, s)$

$$E(Y_t^2) = t^2 \times \frac{1}{t} = t \quad \text{et} \quad E[(Y_t - Y_s) Y_s] = t - s \quad \text{si } s < t$$

Y_t est gaussien et vérifie les hypothèses d'un mouvement brownien (on regarde les variances et les covariances).

E.5.8. Si on appelle μ_n la loi qui régit l'état à l'instant n alors $\mu_n = \pi^n \mu = \mu$. De même γ_n , loi du vecteur (X_n, X_{n+1}) ne dépend pas de n , la loi de (X_n, X_{n+1}, X_{n+2}) ne dépend pas de n .

E.5.9. a) Z_λ est obtenu par limite (au sens $L^2(P)$) de combinaisons linéaires des variables X_n . Donc toute combinaison linéaire de Z_λ est encore une v.a. gaussienne ; il s'agit d'un processus gaussien.

b) D'après l'isométrie on a

$$E[Z(I) \overline{Z(J)}] = \int_{[0,1]} 1_I(\lambda) 1_J(\lambda) d\mu(\lambda) = \text{réelle}$$

D'autre part X_{-t} ayant pour image, le conjugué de l'image de X_t , on a : $Z(-I) = \overline{Z(I)}$.

En décomposant en parties réelles et imaginaires, on obtient donc $\forall I, J$ $E(Z_2(I) Z_1(J)) = 0$. Si de plus I et J sont disjoints alors $E(Z(I) \overline{Z(J)}) = 0$ et on obtient $E(Z_1(I) Z_1(J)) = 0 = E(Z_2(I) Z_2(J))$.

E.5.10. $\text{cov}(X_n, X_0) = \int_0^1 e^{2i\pi n\lambda} f(\lambda) d\lambda$ par définition.

La série de Fourier de $f(\lambda)$ est uniformément convergente, f est donc continue et égale à la somme de sa série de Fourier.

$$\int_0^1 I_n(\lambda) e^{2i\pi n_0\lambda} d\lambda = \frac{1}{n} \sum_{p=0}^{n-n_0} \bar{X}_p X_{p+n_0} = \frac{n-n_0+1}{n} \times \frac{1}{n-n_0+1} \sum_{p=0}^{n-n_0} \bar{X}_p X_{p+n_0}$$

Si le processus est ergodique cette somme converge presque sûrement vers $E(X_{p+n_0} \bar{X}_p)$.

Chapitre 6

E.6.1. On fait un changement d'échelle sur le temps pour se ramener au cas où $T = 1$, en posant $x_0(t) = s_0(tT)$, $x_1(t) = s_1(tT)$. La fonction de transfert est alors $1 \left[-\frac{TF}{2}, \frac{TF}{2} \right] (\lambda)$. B_0 est remplacé par $\frac{B_0}{T}$. La variance en sortie est égale à

$$\int \frac{B_0}{T} \left| 1 \left[-\frac{TF}{2}, \frac{TF}{2} \right] (\lambda) \right|^2 d\lambda = B_0 F$$

$$x_1(t) = \int_{t-1/2}^{t+1/2} \frac{\sin \pi TF u}{\pi u} du$$

$x_1^2(t)$ passe par un maximum en $t = 0$ (voir la figure E.6.1. pour l'exemple $TF = 1$). On regarde R pour $t = 0$.

a) Si TF est petit, R est peu différent de $\frac{T^2 F}{B_0}$.

b) Si TF est grand, R est de l'ordre de $\frac{1}{B_0 F}$.

On peut donc dire qu'à T fixé, R passe par un maximum lorsque F varie de 0 à $+\infty$.

E.6.1. ($TF = 1$).

E.6.2. a) $B_1 = \int_{\mathbb{R}} f(\lambda) |H(\lambda)|^2 d\lambda$ et $s_1(t) = \int \hat{s}_0(\lambda) H(\lambda) e^{2i\pi\lambda t} d\lambda$.

b) On applique l'inégalité de Cauchy-Schwartz au produit $\left[\frac{\hat{s}_0(\lambda)}{\sqrt{f(\lambda)}} \right] \cdot [\sqrt{f(\lambda)} H(\lambda)]$. Le maximum est atteint en cas de colinéarité, c'est-à-dire pour

$$H(\lambda) = \frac{\hat{s}_0(\lambda)}{f(\lambda)} \times \text{Cte}$$

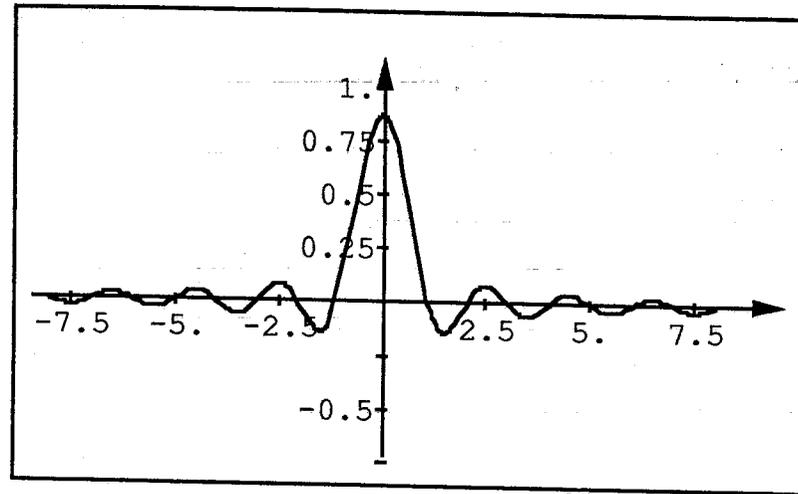


Figure E.6.1. - TF = 1.

Lorsque $s_0(t) = au(t - \tau)$, un maximum en $t = t_0$ pour $|u|$ se traduit par un maximum en $t = t_0 + \tau$ pour s_1 et c'est en ce point que l'on calcule R.

E.6.3. a) En utilisant le lemme de Loeve vu au paragraphe 3.1.2, on montre que $E\left(\frac{X_{t+h_n} - X_t}{h_n} \cdot \frac{X_{t+h_m} - X_t}{h_m}\right)$ converge et que $\frac{X_{t+h} - X_t}{h}$ converge en m.q. Par ailleurs il y a convergence p.s. vers X'_t ; on a donc aussi la convergence en moyenne :

$$E\left(\left|\frac{X_{t+h} - X_t}{h} - X'_t\right|\right) \xrightarrow{h \rightarrow 0} 0$$

On obtient par passage à la limite $E(Y_t) = 0$ et

$$E(Y_t^2) = k''(0); \quad E(Y_{t+\tau} Y_t) = k''(\tau)$$

b) Par passage à la limite le couple (X_t, Y_t) est gaussien centré, de matrice de covariance

$$\begin{pmatrix} k(0) & k'(0) \\ k'(0) & k''(0) \end{pmatrix}$$

E.6.4. On développe l'espérance et on trouve que le minimum est atteint pour $a = k(1)/k(0)$.

E.6.5. a et b doivent être solutions du système linéaire (obtenu par dérivation)

$$\begin{cases} k(0)a + bk(2) = k(1) \\ k(2)a + k(0)b = k(1) \end{cases}$$

soit

$$a = b = \frac{k(1)}{k(0) + k(2)}$$

$$\text{E.6.6. } Y_n - X_n = [(h - \delta) * X]_n + \varepsilon_n$$

$$E(Y_n - X_n)^2 = \int_0^1 [|H(\lambda) - 1|^2 f_1(\lambda) + f_2(\lambda)] d\lambda$$

$$\text{E.6.7. } \theta = E(Y_n^2)/E(Y_n X_n)$$

$$\text{E.6.8. } H(z) = \frac{1}{1 - \theta z^{-1}} = \sum_{n=0}^{\infty} \theta^n z^{-n}$$

Le filtre est stable si et seulement si $|\theta| < 1$.

2) Si X_n est un bruit blanc de niveau B_0 , Y_n a pour densité spectrale de puissance

$$\frac{1}{1 + \theta^2 - 2\theta \cos 2\pi\lambda} = \frac{1}{1 - \theta} \sum_n \theta^{|n|} e^{-2i\pi n\lambda} \text{ cov}(Y_n, Y_0) = \frac{\theta^{|n|}}{1 - \theta}$$

$$\text{E.6.9. } Y_{n+1} = a + (\varepsilon_{n+1} - \varepsilon_n)$$

Y_n est stationnaire au second ordre et a pour d.s.p. $2f(\lambda) (1 - \cos 2\pi\lambda)$.

Chapitre 7

E.7.1. Pour $t \in [0, 1]$ $P(X < t) = P(\theta < \text{Arcsin } t)$.

La densité est égale à $\frac{2}{\pi} \times \frac{1}{\sqrt{1-t^2}}$.

On veut

$$P(0 < X < t_1) = 1/4 = P(t_1 < X < t_2) = P(t_2 < X < t_3) = P(t_3 < X < 1)$$

avec $0 < t_1 < t_2 < t_3 < 1$; ce qui donne

$$t_1 = \sin \frac{\pi}{8} = 0,3827 \quad t_2 = \frac{\sqrt{2}}{2} = 0,7071 \quad t_3 = 0,9239$$

si $X \in [0, t_1]$, $\tilde{X} = \frac{t_1}{2}$ et $(X - \tilde{X})^2 = \left(X - \frac{t_1}{2}\right)^2$

d'où

$$E(X - \tilde{X})^2 = \int_0^{t_1} \frac{2}{\pi} \left(\sin \theta - \frac{t_1}{2}\right)^2 d\theta + \frac{2}{\pi} \int_{t_1}^{t_2} \left(\sin \theta - \frac{t_1+t_2}{2}\right)^2 d\theta + \frac{2}{\pi} \int_{t_2}^{t_3} \left(\sin \theta - \frac{t_2+t_3}{2}\right)^2 d\theta + \frac{2}{\pi} \int_{t_3}^{1/2} \left(\sin \theta - \frac{1+t_3}{2}\right)^2 d\theta = 6,5 \times 10^{-3}$$

E.7.2. $m(t) = 0$

$$K(t, s) = E(X_t X_s) = \sum_k h(t - kT) h(s - kT)$$

$$\Gamma(t + \alpha, s + \alpha) = \frac{1}{T} \int_{\alpha}^{\alpha+T} K(t + u, s + u) du = \Gamma(t, s)$$

$$\begin{aligned} \text{Donc } T(t, s) = \gamma(t - s) &= \frac{1}{T} \sum_k \int_{k - kT - T}^{s - kT} h(t - s + u) h(u) du \\ &= \frac{1}{T} \int_{-\infty}^{+\infty} h(t - s + u) h(u) du \end{aligned}$$

E.7.3. $E(Y_t) = 0$ et $\gamma(t, s) = \frac{1}{T} \int_0^T K(t - \theta, s - \theta) d\theta$.

Pour le code NRZ, $h(t) = 1$ sur $[0, T]$, on trouve que $\gamma(t - s)$ est la fonction triangulaire

$$\gamma(t - s) = 1 - \frac{|t - s|}{T} \quad \text{pour } |t - s| \leq T$$

Dans le code RZ, $h(t) = 1$ sur $[0, T/2]$ mais $a_k = 0$ ou 1 de sorte que

$$E(a_k^2) - [E(a_k)]^2 = \frac{1}{4}$$

$$E(Y_t) = \frac{1}{2} \quad E(Y_t Y_s) = \text{cov}(Y_t, Y_s) + \frac{1}{4}$$

$\text{cov}(Y_t, Y_s)$ se calcule comme précédemment et c'est une fonction triangulaire sur $[-\frac{T}{2}, \frac{T}{2}]$ qui vaut $\frac{1}{2}$ en 0. Pour le code NRZ, la mesure spectrale de puissance est une mesure à densité tandis que pour le code RZ, on a en plus $\frac{1}{4} \delta$, c'est-à-dire une raie en 0 due au fait que le processus n'est pas centré.

E.7.4. On développe l'espérance en tenant compte des corrélations entre symboles.

E.7.5. Pour θ proche de $\frac{\pi}{4}$, θ peut être codé 110 ou 111. Si au décodage, on associe $\frac{\pi}{8}$ au mot 110, $\frac{3\pi}{8}$ au mot 111, ... alors l'erreur est proche de $\frac{\pi}{8}$. Si θ est proche de $\frac{\pi}{2}$ l'erreur peut être plus importante : on peut avoir 100 ou 010 ou 001 ou ... c'est-à-dire n'importe quel autre mot code. Si on code 100 cela correspond au secteur 7 c'est-à-dire une erreur de $\pi + \frac{\pi}{8}$.

E.7.6. $X^5 = X^4 + X \text{ mod } g = X^3 + X + 1 \text{ mod } g$ en recommençant on trouve $X^{11} = X^3 + X^2 + 1 \text{ mod } g$. Donc g ne divise pas $X^{11} + 1$.

E.7.7. On cherche les diviseurs de $X^7 + 1$ de degré 3 exactement ; le coefficient constant doit être égal à 1 ; il reste 4 possibilités pour les coefficients.

E.7.8. a) On écrit

$$\log p_{ij} = \log [P(X = i / Y = y_j)] + \log (P(Y = y_j))$$

b) Si X et Y sont indépendantes $P(X = x_i / Y = y_j) = P(X = x_i)$.

c) La loi conditionnelle de Y sachant $X = x_i$ est concentrée sur la valeur $f(x_i)$ et l'entropie conditionnelle est nulle.

E.7.9. Les 8 messages sont codés dans l'ordre

$$11, 10, 011, 010, 0011, 0010, 0001, 0000$$

par la méthode de Fano Shannon, ce qui donne une longueur moyenne $\bar{n} = 2,55$ au lieu de 3 en longueur fixe. Par la méthode d'Huffman on obtient

$$0, 111, 110, 100, 10101, 10100, 10111, 10110$$

ce qui donne $\bar{n} = 2,55...$ la même valeur que pour la méthode Fano Shannon.

E.7.10. On calcule $H(Y) - H_X(Y)$ et on cherche le maximum sur $p = P(X = 0)$. On doit résoudre

$$(1 - 0,03) \text{Ln} \left[\frac{p(1 - 0,02) + (1 - p) \times 0,01}{p \times 0,02 + (1 - p)(1 - 0,01)} \right] = 0,104$$

On trouve $p = 0,494$ et $C = 0,889$.

E.7.11. L'état 1 a pour entropie 0,918, l'état 2 : 0,811, la source : 0,865.

A a pour probabilité $\frac{1}{3}$, B $\frac{7}{24}$ et C $\frac{3}{8}$.

Pour les messages de 2 lettres :

AA	AC	CB	CC	CA	BB	BC
$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{9}{32}$	$\frac{3}{32}$

Les messages de 1 lettre ont une information moyenne égale à 1,577. Pour les messages de 2 lettres, l'entropie est égale à 2,527 soit 1,263 pour lettre. On remarque que l'entropie de la source est inférieure à 1,263 qui est inférieur à 1,577 (on montre qu'il y a convergence). Par le codage d'Huffman les messages sont codés par :

BB → 10	AA → 01	CB → 110	AC → 001
BC → 000	CC → 1111	CA → 1110	

La longueur moyenne est $\bar{n} = 2,66$.

E.7.12. En utilisant

$$\log(n_i!) = \log \left[\left(\frac{n_i}{e} \right)^{n_i} \sqrt{2\pi n_i} \right] + \frac{1}{12 n_i} + \frac{1}{n_i} \varepsilon(n_i)$$

On cherche le maximum de $\frac{n!}{n_1! \dots n_6!} \left(\frac{1}{6}\right)^n$ sous les contraintes :

$$n_1 + \dots + n_6 = n \quad \text{et} \quad \sum \frac{n_i}{n} = 4,5$$

$$\frac{\partial}{\partial f_i} \left[H(f_1, \dots, f_6) + \lambda (\sum i f_i - 4,5) + \mu (\sum f_i - 1) \right] = 0$$

On trouve $f_i = ab^i$ et on détermine les constantes a et b avec les contraintes, soit à résoudre $a = f(b) = \frac{b}{1-b^6}$ et $4,5 = a(a + 6a^2b^4)$ c'est-à-dire $b = g(a)$. La figure E.7.12 représente le graphe de $f(g(a)) - a$; on trouve $a = 1,168$ et $b = 0,818$.

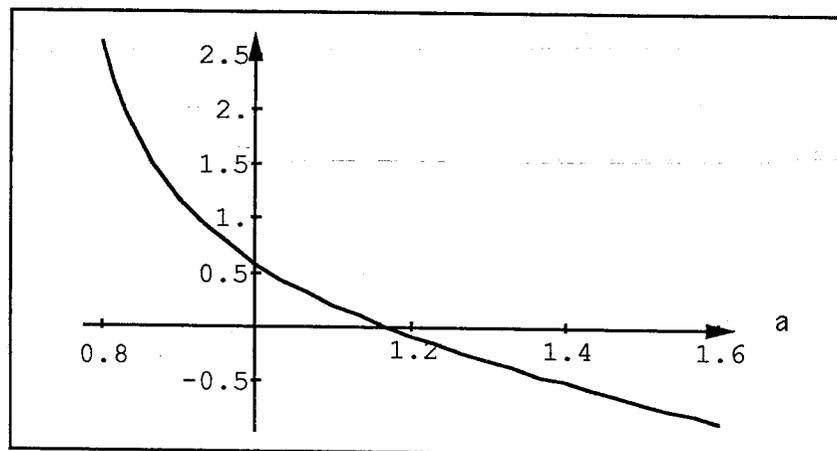


Figure E.7.12.

Chapitre 8

E.8.1. $\hat{p}(\omega) = \frac{13}{150} = 8,6\%$, $\hat{p}(\omega)$ est la réalisation d'une v.a. de loi $\mathcal{B}(150, p)$; en faisant l'approximation gaussienne on a

$$P(|p - \hat{p}| \leq 0,5\%) = P(\sqrt{150}|p - \hat{p}| \leq 0,5\% \times \sqrt{150})$$

$p(1-p)$ est proche de $7,9 \cdot 10^{-2}$. Donc, en utilisant une table de la loi $\mathcal{N}(0,1)$, la probabilité est égale à 79,2%.

E.8.2. Pour un couple gaussien $\begin{pmatrix} X \\ Y \end{pmatrix}$, l'estimateur du maximum de vraisemblance de $m_X = E(X)$ et $m_Y = E(Y)$ s'obtient en prenant la moyenne empirique de X et Y respectivement (même si les deux composantes sont corrélées)

$$\hat{m}_X = \frac{X_1 + \dots + X_n}{n} \quad \hat{m}_Y = \frac{Y_1 + \dots + Y_n}{n}$$

E.8.3. On a vu au chapitre 3 que la moyenne de n variables de Cauchy indépendantes, était encore une variable de Cauchy. On en déduit que \bar{X} a encore pour densité $f_\theta(x)$

$$P\left(|\bar{X} - \theta| > \frac{1}{100}\right) = 1 - \int_{-\frac{1}{100}}^{\frac{1}{100}} \frac{1}{\pi} \times \frac{1}{1+x^2} dx = 99,3\%$$

E.8.4. a) $X - \theta$ suit une loi exponentielle de paramètre 1.

b) $E(\bar{X}) = \theta + 1$ $E\left(\text{Inf}_i(X_i)\right) = \theta + \frac{1}{n}$ en calculant la densité de la variable $\text{Inf}_i(X_i - \theta)$, qui est celle d'une loi exponentielle de paramètre n .

$\bar{X} - 1$ et $\text{Inf}_i(X_i) - \frac{1}{n}$ sont deux estimateurs sans biais de θ ; le premier a pour variance $\frac{1}{n}$, le second $\frac{1}{n^2}$.

c) L'estimateur du maximum de vraisemblance est égal à $\text{Inf}_i(X_i)$. Pour la même technique que dans l'exercice E.3.3, on montre que $\text{Inf}_i(X_i)$

converge presque sûrement vers θ (estimateur consistant).

E.8.5. $\text{Sup}_\lambda \left[e^{-n\lambda} \frac{\lambda^{x_1}}{x_1!} \dots \frac{\lambda^{x_n}}{x_n!} \right]$ est atteint pour $\hat{\lambda} = \frac{x_1 + \dots + x_n}{n}$.

E.8.6. $\frac{1}{\lambda} E_\lambda[(tX - \lambda)^2] = c^2 + \lambda(c-1)^2$ expression dont on prend l'espérance en utilisant la loi de λ . Le minimum de $c^2 + 10(c-1)^2$ est atteint pour $c = \frac{10}{11}$.

E.8.7. On sait que l'estimateur bayésien est égal dans ce cas (erreur quadratique) à $E(\lambda/X_1, \dots, X_n)$, la loi conditionnelle de λ sachant $X_1 = x_1 \dots X_n = x_n$ est donnée par une densité du type

$$c e^{-\left(n + \frac{1}{10}\right)\lambda} \lambda^{x_1 + \dots + x_n}$$

Il s'agit donc d'une loi Gamma $\gamma\left(x_1 + x_2 + \dots + x_n + 1, n + \frac{1}{10}\right)$ avec

$$c = \frac{\left(n + \frac{1}{10}\right)^{x_1 + x_2 + \dots + x_n + 1}}{\Gamma(x_1 + x_2 + \dots + x_n + 1)}$$

D'où $E(\lambda/X_1 = x_1 \dots X_n = x_n) = \frac{x_1 + x_2 + \dots + x_n + 1}{n + \frac{1}{10}}$

$$E(\lambda/X_1, \dots, X_n) = \frac{X_1 + \dots + X_n}{n} \times \frac{n}{n + \frac{1}{10}} + \frac{1}{n + \frac{1}{10}}$$

E.8.8. Si $p \leq 10^{-2}$

$$P(\text{rejet du lot}) = 1 - (1-p)^{40} \approx 1 - (0,99)^{40} = 33,1 \%$$

La première contrainte n'est pas satisfaite.

La deuxième contrainte est satisfaite.

E.8.9. a) σ^2 est estimé par $\frac{1}{9} \sum (x_i - \bar{x})^2 = 1,734$ et $\hat{\sigma} = 1,317$.

b) On utilise un test de Student unilatère ; en considérant $\frac{\bar{X} - 100}{\frac{1}{\sqrt{n}} \hat{\sigma}}$ et on

rejette l'hypothèse H_0 si $\sqrt{10} \frac{\bar{X} - 100}{\hat{\sigma}} > 1,833$. Ce rapport est égal à 1,62 sur les valeurs mesurées ; on garde l'hypothèse H_0 .

E.8.10. On remarque $I(\theta)$ s'exprime à l'aide de la dérivée logarithmique de la densité :

$$\frac{\partial}{\partial \theta} \log f_{\theta}(x) = \frac{2(x - \theta)}{1 + (x - \theta)^2} \quad \text{et} \quad I(\theta) = \frac{4}{\pi} \int_{-\infty}^{+\infty} \frac{u^2}{[1 + u^2]^3} du$$

soit $I(\theta) = \frac{1}{2}$

$$\text{E.8.11. } A_i = \int_0^T s_i(t) s(t) dt + \int_0^T s_i(t) B_i dt.$$

Le vecteur $\begin{pmatrix} A_1 \\ \vdots \\ A_4 \end{pmatrix}$ est gaussien, centré sur $\begin{pmatrix} \int_0^T s_1(t) s(t) dt \\ \vdots \\ \int_0^T s_4(t) s(t) dt \end{pmatrix}$ de matrice de

variance-covariance égale à M avec $M_{ij} = \sigma^2 \int_0^T s_i(t) s_j(t) dt$ soit

$$M_{ii} = \sigma^2 \frac{T}{2}; \quad M_{1,2} = M_{1,4} = 0; \quad M_{1,3} = -\frac{\sigma^2 T}{2} = M_{2,4} = M_{3,4}$$

La matrice de variance-covariance est la même sous les quatre hypothèses ; par contre l'espérance change.

Il n'est pas nécessaire de calculer M^{-1} pour avoir l'estimateur. En effet, si m_i est l'espérance sous l'hypothèse H_i , on vérifie

$$'(a - m_i) M^{-1} (a - m_i) = 'a M^{-1} a + 'm_i M^{-1} m_i - 2 'a M^{-1} m_i$$

$$M e_i = m_i \sigma^2 \Rightarrow 'm_i M^{-1} m_i = \frac{2}{T} \sigma^2 \quad \text{et} \quad 'a M^{-1} m_i = 'a e_i \sigma^2$$

D'où la règle de décision : on décide H_1 si

$$A_1 = \text{Max} (A_1, A_2, A_3, A_4)$$

on décide H_2 si $A_2 = \text{Max} (A_1, \dots, A_4) \dots$

E.8.12. A est une v.a. gaussienne, de variance $\sigma^2 T$, centrée sur 0 sous l'hypothèse H_0 , centrée sur $\int_0^T s^2(t) dt$ sous H_1 . On rejette l'hypothèse H_0 lorsque le rapport des densités $\frac{f_1}{f_0}$ est supérieur à 9, c'est-à-dire si

$$A \geq \frac{1}{2} \int_0^T s^2(t) dt + \frac{\text{Log } 9}{\int_0^T s^2(t) dt}$$

E.8.13. Les variables Y_1, \dots, Y_N forment un vecteur gaussien à composantes indépendantes grâce au pas choisi, de variance $B_0 F$. On choisit H_0 si et seulement si

$$\sum_{i=1}^N (y_i - s_0(t_i))^2 \leq \sum_{i=1}^N (y_i - s_1(t_i))^2$$

c'est-à-dire si et seulement si

$$\sum_{i=1}^N y_i [s_1(t_i) - s_0(t_i)] \leq \frac{1}{2} \left[\sum_{i=1}^N s_1^2(t_i) - \sum_{i=1}^N s_0^2(t_i) \right]$$

la variable $A = \int_0^T [s_1(t) - s_0(t)] Y_t dt$ est gaussienne, de variance

$$\sigma^2 = B_0 \int_{-F}^F |\hat{s}_1(\lambda) - \hat{s}_0(\lambda)|^2 d\lambda$$

L'espérance est égale à m_0 sous H_0 , m_1 sous H_1 :

$$m_0 = \int_0^T s_0(t) [s_1(t) - s_0(t)] dt \quad \text{et} \quad m_1 = \int_0^T s_1(t) [s_1(t) - s_0(t)] dt$$

En appelant ϕ la fonction de répartition d'une loi $\mathcal{N}(0, 1)$, l'erreur de 1^{re} espèce est égale à $1 - \phi\left(\frac{c - m_0}{\sigma}\right)$ et l'erreur de 2^e espèce à $\phi\left(\frac{c - m_1}{\sigma}\right)$ en posant

$$c = \frac{1}{2} \int_0^T [s_1^2(t) - s_0^2(t)] dt$$

On obtient

$$c - m_0 = \frac{1}{2} \int_0^T (s_1^2(t) - s_0^2(t)) dt = - (c - m_1)$$

de sorte que les deux erreurs sont égales.

Chapitre 9

E.9.1. En décomposant suivant les différentes valeurs de k_1 on a 28 possibilités pour le triplet (N_1, N_2, N_3)

$$P(N_1 = N_2 = N_3) = P(N_1 = N_2 = N_3 = 2) = \frac{6!}{(2!)^3} \left(\frac{1}{3}\right)^6 = \frac{10}{81}$$

D	0	1	3	4	7	12
Pr.	$\frac{10}{3^4}$	$\frac{40}{3^4}$	$\frac{10}{3^4} + \frac{20}{3^5}$	$\frac{10}{3^4}$	$\frac{4}{3^4}$	$\frac{1}{3^5}$

La probabilité est maximum pour la valeur 1. D représente la distance (du χ^2) entre la loi empirique et la loi uniforme (voir § 9.1).

E.9.2. On teste l'hypothèse H_0 : « la loi est uniforme »

$$P(\text{rouge}) = P(\text{bleu}) = P(\text{vert}) = \frac{1}{3}$$

$$\sum_{i=1}^3 \frac{(n_i - np_i)^2}{np_i} = 0,892 \quad (\chi^2 \text{ à } 2 \text{ degrés de liberté})$$

Au niveau 5 %, on décide de ne pas rejeter l'hypothèse H_0 : les valeurs ne sont pas significativement différentes. On ne sait pas avec quelle erreur on prend cette décision. (On ne connaît l'erreur que dans le cas où on rejette l'hypothèse H_0 .)

E.9.3. $\sum_{i,j} \frac{(n_{ij} - n\hat{p}_i \hat{q}_j)^2}{n\hat{p}_i \hat{q}_j} = 5,44$, valeur non significative pour un χ^2 à 4 degrés de liberté et un niveau de 5 %. On accepte l'hypothèse d'indépendance.

$$E.9.4. \quad E(X) = \frac{1}{\lambda} \text{ soit } \lambda = \frac{1}{E(X)}$$

$$P(X < t_0) = P(X > t_0) = 0,5 \quad \text{pour } t_0 = -\frac{1}{\lambda} \text{Ln}(0,5) = -E(X) \cdot \text{Ln}(0,5)$$

Statistiquement, pour un grand nombre de particules on en aura la moitié désintégrées avant l'instant t_0 (aux aléas près).

$$E.9.5. \quad P(X < -t) = P(X > t) \Rightarrow F(-t) = 1 - F(t) \text{ si } t \geq 0$$

$$P(|X| < t) = P(-t < X < t) = F(t) - F(-t) = 2F(t) - 1$$

$$P(X > 0 \text{ et } |X| < t) = P(0 < X < t) = F(t) - F(0) = F(t) - \frac{1}{2} = \\ = P(|X| < t) \times \frac{1}{2} = P(|X| < t) P(X > 0)$$

Pour un n -échantillon, les variables $(\text{sign}(X_1), \dots, \text{sign}(X_n), |X_1|, \dots, |X_n|)$ sont indépendantes. Ce résultat sert à calculer la loi de statistique de Wilcoxon sous l'hypothèse H_0 .

E.9.6. Les deux cas extrêmes sont

$$X_1(\omega) < X_2(\omega) \dots < X_n(\omega) < Y_1(\omega) < \dots < Y_m(\omega)$$

et

$$Y_1(\omega) < Y_2(\omega) \dots < Y_m(\omega) < X_1(\omega) < X_2(\omega) \dots < X_n(\omega)$$

Dans le premier cas $U = 0$, dans le second $U = nm$.

Toutes les valeurs intermédiaires peuvent être obtenues. Ordonnons X_1, X_2, \dots, X_n : $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ avec $X_{(1)} = \text{Inf}(X_1, X_2, \dots, X_n) \dots$ et ordonnons les $n + m$ variables. On a alors

$$U = [\text{rg}(X_{(1)} - 1)] + [\text{rg}(X_{(2)} - 2)] + \dots + [\text{rg}(X_{(n)} - n)] \\ = \sum \text{rg}(X_i) - \frac{n(n+1)}{2}$$

E.9.7. $P(A_i < t) = P(F(X_i) < t) = t$ si $t \in [0, 1]$ en distinguant le cas où F est strictement croissante au voisinage de t , du cas où F n'est pas strictement croissante.

F étant croissante, conserve l'ordre et U est inchangé, on remarque

$$P(X_i = X_j) = 0 = P(X_i = Y_j)$$

D'autre part, on ne peut avoir $X_i < X_j$ et $F(X_i) = F(X_j)$ avec une probabilité positive, car les ensembles où F est constante sont de probabilité nulle.

E.9.8. On sait que \bar{X} a même loi que X_1 (chap. 3)

$$P_{H_0}(\bar{X} \geq A) = 5 \% \quad \text{pour } A = \text{tg}(0,45 \pi) = 6,313$$

$$\beta(\theta) = P_\theta(\bar{X} \geq A) = P_0(\bar{X} - \theta \geq A - \theta) = \frac{1}{2} - \frac{1}{\pi} \text{Arctg}(A - \theta)$$

$$\beta'(0) = 8 \cdot 10^{-3}$$

Pour le test du signe

$$P_{H_0}(S \geq 9) = 3,3 \% \quad \text{et} \quad P_{H_0}(S \geq 8) = 11,3 \%$$

On prend comme région de rejet de H_0 l'ensemble {nombre de signes positifs ≥ 9 }

$$P_\theta(X > 0) = P_0(X - \theta > -\theta) = \frac{1}{2} + \frac{1}{\pi} \text{Arctg} \theta$$

et

$$\frac{d}{d\theta} P_\theta(S \geq 9) |_{\theta=0} = \frac{1}{2\pi} \left[\frac{990}{2^{11}} \right] = 0,769$$

La puissance croît plus vite avec le test du signe.

Pour être rigoureux, on aurait dû prendre $\alpha = 3,3 \%$ pour le premier test, ce qui aurait donné $A = 9,61$ et $\beta'(0) = 3 \cdot 10^{-3}$. Le test du signe sépare mieux les deux hypothèses.

E.9.9. On trouve $\hat{a} = 3,96 \cdot 10^3$, $\hat{b} = -650,26 \cdot 10^3$, $\hat{\sigma}^2 = 1\,657,8 \cdot 10^6$

$$\text{var}(\hat{a}) \text{ est estimée par } \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2} = 0,3614 \cdot 10^6$$

$\frac{\hat{a}}{\sqrt{\text{est de la variance}}} = 6,6$ est à comparer à la valeur critique pour une variable de Student à 3 degrés de liberté. Si $\alpha = 5 \%$ cette valeur critique est égale à 3,182 et on rejette l'hypothèse $H_0 = \langle a = 0 \rangle$.

ANNEXE 1

THÉORÈME DE KOLMOGOROV

a) **Définition :** Une classe C de parties d'un ensemble d'épreuves Ω , est dite compacte si quelle que soit la suite C_n d'éléments de C dont l'intersection est vide, il existe un entier N pour lequel $\bigcap_{n=1}^N C_n = \emptyset$.

b) **Enoncé du théorème :** Soit $\{P_{t_1, \dots, t_n}\}$ une famille de probabilités définies sur les espaces produits finis de la famille $\{(\Omega_t, \mathcal{A}_t); t \in T\}$ et telle que si $T_1 \subset T_2$ sont deux sous-ensembles finis de T , la restriction de P_{T_2} à $\bigotimes_{T_1} \mathcal{A}_t$ est égale

à P_{T_1} . On suppose que pour tout $t \in T$, il existe une sous-classe compacte, C_t , de \mathcal{A}_t telle que, pour tout A de \mathcal{A}_t , on a

$$P(A) = \sup_{\substack{C \in \mathcal{C}_t \\ C \subset A}} P(C)$$

alors il existe une probabilité P unique sur l'espace produit $\left(\prod_T \Omega_t, \bigotimes_T \mathcal{A}_t \right)$ qui prolonge chacune des probabilités $P_{(t_1, t_2, \dots, t_n)}$.

c) L'hypothèse d'existence d'une classe compacte C_t est automatiquement vérifiée lorsque Ω_t est égal à \mathbb{R} ou \mathbb{R}^n et que \mathcal{A}_t est la σ -algèbre des boréliens.

ANNEXE 2

TABLES STATISTIQUES

Loi binomiale

La table donne les probabilités $P(X = k) = C_n^k p^k (1-p)^{n-k}$.

n	k	p=0.1	p=0.2	p=0.3	p=0.4	p=0.5
5	0	0.5905	0.3277	0.1681	0.0778	0.0313
	1	0.3280	0.4096	0.3601	0.2592	0.1563
	2	0.0729	0.2048	0.3087	0.3456	0.3125
	3	0.0081	0.0512	0.1323	0.2304	0.3125
	4	0.0004	0.0064	0.0283	0.0768	0.1563
	5	0.0000	0.0003	0.0024	0.0102	0.0313
6	0	0.5314	0.2621	0.1176	0.0467	0.0156
	1	0.3543	0.3932	0.3025	0.1866	0.0938
	2	0.0984	0.2458	0.3241	0.3110	0.2344
	3	0.0146	0.0819	0.1852	0.2765	0.3125
	4	0.0012	0.0154	0.0595	0.1382	0.2344
	5	0.0001	0.0015	0.0102	0.0369	0.0938
	6	0.0000	0.0001	0.0007	0.0041	0.0156
7	0	0.4783	0.2097	0.0824	0.0280	0.0078
	1	0.3720	0.3670	0.2471	0.1306	0.0547
	2	0.1240	0.2753	0.3177	0.2613	0.1641
	3	0.0230	0.1147	0.2269	0.2903	0.2734
	4	0.0026	0.0287	0.0972	0.1935	0.2734
	5	0.0002	0.0043	0.0250	0.0774	0.1641
	6	0.0000	0.0004	0.0036	0.0172	0.0547
	7	0.0000	0.0000	0.0002	0.0016	0.0078
8	0	0.4305	0.1678	0.0576	0.0168	0.0039
	1	0.3826	0.3355	0.1977	0.0896	0.0313
	2	0.1488	0.2936	0.2965	0.2090	0.1094
	3	0.0331	0.1468	0.2541	0.2787	0.2188
	4	0.0046	0.0459	0.1361	0.2322	0.2734
	5	0.0004	0.0092	0.0467	0.1239	0.2188
	6	0.0000	0.0011	0.0100	0.0413	0.1094
	7	0.0000	0.0001	0.0012	0.0079	0.0313
	8	0.0000	0.0000	0.0001	0.0007	0.0039

n	k	p=0.1	p=0.2	p=0.3	p=0.4	p=0.5
9	0	0.3874	0.1342	0.0404	0.0101	0.0020
	1	0.3874	0.3020	0.1556	0.0605	0.0176
	2	0.1722	0.3020	0.2668	0.1612	0.0703
	3	0.0446	0.1762	0.2668	0.2508	0.1641
	4	0.0074	0.0661	0.1715	0.2508	0.2461
	5	0.0008	0.0165	0.0735	0.1672	0.2461
	6	0.0001	0.0028	0.0210	0.0743	0.1641
	7	0.0000	0.0003	0.0039	0.0212	0.0703
	8	0.0000	0.0000	0.0004	0.0035	0.0176
	9	0.0000	0.0000	0.0000	0.0003	0.0020
10	0	0.3487	0.1074	0.0282	0.0060	0.0010
	1	0.3874	0.2684	0.1211	0.0403	0.0098
	2	0.1937	0.3020	0.2335	0.1209	0.0439
	3	0.0574	0.2013	0.2668	0.2150	0.1172
	4	0.0112	0.0881	0.2001	0.2508	0.2051
	5	0.0015	0.0264	0.1029	0.2007	0.2461
	6	0.0001	0.0055	0.0368	0.1115	0.2051
	7	0.0000	0.0008	0.0090	0.0425	0.1172
	8	0.0000	0.0001	0.0014	0.0106	0.0439
	9	0.0000	0.0000	0.0001	0.0016	0.0098
	10	0.0000	0.0000	0.0000	0.0001	0.0010
11	0	0.3138	0.0859	0.0198	0.0036	0.0005
	1	0.3835	0.2362	0.0932	0.0266	0.0054
	2	0.2131	0.2953	0.1998	0.0887	0.0269
	3	0.0710	0.2215	0.2568	0.1774	0.0806
	4	0.0158	0.1107	0.2201	0.2365	0.1611
	5	0.0025	0.0388	0.1321	0.2207	0.2256
	6	0.0003	0.0097	0.0566	0.1471	0.2256
	7	0.0000	0.0017	0.0173	0.0701	0.1611
	8	0.0000	0.0002	0.0037	0.0234	0.0806
	9	0.0000	0.0000	0.0005	0.0052	0.0269
	10	0.0000	0.0000	0.0000	0.0007	0.0054
	11	0.0000	0.0000	0.0000	0.0000	0.0005

n	k	p=0.1	p=0.2	p=0.3	p=0.4	p=0.5
12	0	0.2824	0.0687	0.0138	0.0022	0.0002
	1	0.3766	0.2062	0.0712	0.0174	0.0029
	2	0.2301	0.2835	0.1678	0.0639	0.0161
	3	0.0852	0.2362	0.2397	0.1419	0.0537
	4	0.0213	0.1329	0.2311	0.2128	0.1208
	5	0.0038	0.0532	0.1585	0.2270	0.1934
	6	0.0005	0.0155	0.0792	0.1766	0.2256
	7	0.0000	0.0033	0.0291	0.1009	0.1934
	8	0.0000	0.0005	0.0078	0.0420	0.1208
	9	0.0000	0.0001	0.0015	0.0125	0.0537
	10	0.0000	0.0000	0.0002	0.0025	0.0161
	11	0.0000	0.0000	0.0000	0.0003	0.0029
12	0.0000	0.0000	0.0000	0.0000	0.0002	
13	0	0.2542	0.0550	0.0097	0.0013	0.0001
	1	0.3672	0.1787	0.0540	0.0113	0.0016
	2	0.2448	0.2680	0.1388	0.0453	0.0095
	3	0.0997	0.2457	0.2181	0.1107	0.0349
	4	0.0277	0.1535	0.2337	0.1845	0.0873
	5	0.0055	0.0691	0.1803	0.2214	0.1571
	6	0.0008	0.0230	0.1030	0.1968	0.2095
	7	0.0001	0.0058	0.0442	0.1312	0.2095
	8	0.0000	0.0011	0.0142	0.0656	0.1571
	9	0.0000	0.0001	0.0034	0.0243	0.0873
	10	0.0000	0.0000	0.0006	0.0065	0.0349
	11	0.0000	0.0000	0.0001	0.0012	0.0095
	12	0.0000	0.0000	0.0000	0.0001	0.0016
13	0.0000	0.0000	0.0000	0.0000	0.0001	
14	0	0.2288	0.0440	0.0068	0.0008	0.0001
	1	0.3559	0.1539	0.0407	0.0073	0.0009
	2	0.2570	0.2501	0.1134	0.0317	0.0056
	3	0.1142	0.2501	0.1943	0.0845	0.0222
	4	0.0349	0.1720	0.2290	0.1549	0.0611
	5	0.0078	0.0860	0.1963	0.2066	0.1222
	6	0.0013	0.0322	0.1262	0.2066	0.1833
	7	0.0002	0.0092	0.0618	0.1574	0.2095
	8	0.0000	0.0020	0.0232	0.0918	0.1833
	9	0.0000	0.0003	0.0066	0.0408	0.1222
	10	0.0000	0.0000	0.0014	0.0136	0.0611
	11	0.0000	0.0000	0.0002	0.0033	0.0222
	12	0.0000	0.0000	0.0000	0.0005	0.0056
	13	0.0000	0.0000	0.0000	0.0001	0.0009
14	0.0000	0.0000	0.0000	0.0000	0.0001	

n	k	p=0.1	p=0.2	p=0.3	p=0.4	p=0.5
15	0	0.2059	0.0352	0.0047	0.0005	0.0000
	1	0.3432	0.1319	0.0305	0.0047	0.0005
	2	0.2669	0.2309	0.0916	0.0219	0.0032
	3	0.1285	0.2501	0.1700	0.0634	0.0139
	4	0.0428	0.1876	0.2186	0.1268	0.0417
	5	0.0105	0.1032	0.2061	0.1859	0.0916
	6	0.0019	0.0430	0.1472	0.2066	0.1527
	7	0.0003	0.0138	0.0811	0.1771	0.1964
	8	0.0000	0.0035	0.0348	0.1181	0.1964
	9	0.0000	0.0007	0.0116	0.0612	0.1527
	10	0.0000	0.0001	0.0030	0.0245	0.0916
	11	0.0000	0.0000	0.0006	0.0074	0.0417
	12	0.0000	0.0000	0.0001	0.0016	0.0139
	13	0.0000	0.0000	0.0000	0.0003	0.0032
	14	0.0000	0.0000	0.0000	0.0000	0.0005
15	0.0000	0.0000	0.0000	0.0000	0.0000	

Loi de Poisson

La table donne : $P(X = k) = e^{-\lambda} \lambda^k / k!$

λ k	1	2	3	4	5	6
0	0.3679	0.1353	0.0498	0.0183	0.0067	0.0025
1	0.3679	0.2707	0.1494	0.0733	0.0337	0.0149
2	0.1839	0.2707	0.2240	0.1465	0.0842	0.0446
3	0.0613	0.1804	0.2240	0.1954	0.1404	0.0892
4	0.0153	0.0902	0.1680	0.1954	0.1755	0.1339
5	0.0031	0.0361	0.1008	0.1563	0.1755	0.1606
6	0.0005	0.0120	0.0504	0.1042	0.1462	0.1606
7	0.0001	0.0034	0.0216	0.0595	0.1044	0.1377
8	0.0000	0.0009	0.0081	0.0298	0.0653	0.1033
9	0.0000	0.0002	0.0027	0.0132	0.0363	0.0688
10	0.0000	0.0000	0.0008	0.0053	0.0181	0.0413
11	0.0000	0.0000	0.0002	0.0019	0.0082	0.0225
12	0.0000	0.0000	0.0001	0.0006	0.0034	0.0113
13	0.0000	0.0000	0.0000	0.0002	0.0013	0.0052
14	0.0000	0.0000	0.0000	0.0001	0.0005	0.0022
15	0.0000	0.0000	0.0000	0.0000	0.0002	0.0009
16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003
17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
...						

Loi normale centrée réduite $\mathcal{N}(0,1)$

La table donne la valeur de $P(X < t)$ où t est écrit avec 2 décimales :
 $t = a 0,1 + b 0,01$ par exemple $P(X < 0,12) = 0,5478$.

t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Loi de khi-deux (ou Qui-deux)

La table donne la valeur t pour laquelle $P(X < t) = p$ lorsque X suit une loi de khi-deux à n degrés de liberté.

n \ p	p					
	1%	5%	10%	90%	95%	99%
1	0.0002	0.0039	0.0158	2.7055	3.8414	6.6348
2	0.0201	0.1026	0.2107	4.6052	5.9915	9.2104
3	0.1148	0.3518	0.5844	6.2514	7.8148	11.3448
4	0.2971	0.7107	1.0636	7.7795	9.4878	13.2769
5	0.5543	1.1455	1.6103	9.2364	11.0705	15.0862
6	0.8721	1.6354	2.2041	10.6446	12.5916	16.8119
7	1.2390	2.1673	2.8331	12.0170	14.0672	18.4756
8	1.6465	2.7326	3.4895	13.3616	15.5073	20.0900
9	2.0879	3.3251	4.1681	14.6837	16.9190	21.6662
10	2.5582	3.9403	4.8652	15.9872	18.3071	23.2089
11	3.0534	4.5748	5.5778	17.2750	19.6752	24.7251
12	3.5705	5.2260	6.3038	18.5493	21.0261	26.2172
13	4.1069	5.8918	7.0415	19.8119	22.3621	27.6880
14	4.6603	6.5706	7.7895	21.0641	23.6849	29.1412
15	5.2293	7.2609	8.5467	22.3072	24.9958	30.5778
16	5.8121	7.9616	9.3122	23.5418	26.2963	32.0000
17	6.4077	8.6717	10.0852	24.7690	27.5871	33.4086
18	7.0148	9.3904	10.8649	25.9895	28.8694	34.8053
19	7.6327	10.1170	11.6509	27.2036	30.1436	36.1911
20	8.2603	10.8508	12.4426	28.4120	31.4105	37.5662
21	8.8971	11.5913	13.2396	29.6151	32.6706	38.9320
22	9.5424	12.3380	14.0415	30.8133	33.9245	40.2892
23	10.1956	13.0905	14.8479	32.0069	35.1725	41.6383
24	10.8562	13.8484	15.6587	33.1963	36.4151	42.9799
25	11.5239	14.6114	16.4734	34.3816	37.6526	44.3138
26	12.1980	15.3791	17.2919	35.5632	38.8852	45.6418
27	12.8783	16.1514	18.1139	36.7412	40.1133	46.9628
28	13.5646	16.9278	18.9392	37.9160	41.3372	48.2786
29	14.2563	17.7083	19.7677	39.0875	42.5570	49.5881
30	14.9533	18.4926	20.5992	40.2561	43.7731	50.8924

Pour $n > 30$, on pourra utiliser l'approximation gaussienne avec la même espérance et la même variance.

Loi de Student

La table donne la valeur de t pour laquelle $P(X > t) = \alpha$ lorsque X est une variable de Student à n degrés de liberté.

n \ α	10%	5%	2.5%	1%
4	1.5332	2.1318	2.7764	3.7470
5	1.4759	2.0150	2.5706	3.3649
6	1.4398	1.9432	2.4469	3.1427
7	1.4149	1.8946	2.3646	2.9980
8	1.3968	1.8595	2.3060	2.8965
9	1.3830	1.8331	2.2622	2.8214
10	1.3722	1.8125	2.2281	2.7638
11	1.3634	1.7959	2.2010	2.7181
12	1.3562	1.7823	2.1788	2.6810
13	1.3502	1.7709	2.1604	2.6503
14	1.3450	1.7613	2.1448	2.6245
15	1.3406	1.7531	2.1315	2.6025
16	1.3368	1.7459	2.1199	2.5835
17	1.3334	1.7396	2.1098	2.5669
18	1.3304	1.7341	2.1009	2.5524
19	1.3277	1.7291	2.0930	2.5395
20	1.3253	1.7247	2.0860	2.5280
21	1.3232	1.7207	2.0796	2.5176
22	1.3212	1.7171	2.0739	2.5083
23	1.3195	1.7139	2.0687	2.4999
24	1.3178	1.7109	2.0639	2.4922
25	1.3163	1.7081	2.0595	2.4851
26	1.3150	1.7056	2.0555	2.4786
27	1.3137	1.7033	2.0518	2.4727
28	1.3125	1.7011	2.0484	2.4671
29	1.3114	1.6991	2.0452	2.4620
30	1.3104	1.6973	2.0423	2.4573
31	1.3095	1.6955	2.0395	2.4528
32	1.3086	1.6939	2.0369	2.4487
33	1.3077	1.6924	2.0345	2.4448
34	1.3070	1.6909	2.0322	2.4412
35	1.3062	1.6896	2.0301	2.4377
36	1.3055	1.6883	2.0281	2.4345
37	1.3049	1.6871	2.0262	2.4314
38	1.3042	1.6860	2.0244	2.4286
39	1.3036	1.6849	2.0227	2.4258
40	1.3031	1.6839	2.0211	2.4233

Pour $n > 120$, on utilisera l'approximation par une loi $\mathcal{N}(0, 1)$.

Loi de Fisher

La table donne t tel que $P(X > t) = 5\%$ lorsque X suit une loi de Fisher de degrés respectifs n et m (n est le nombre de degrés de liberté du numérateur).

n \ m	1	2	3	4	5	6	7	8	9	10
1	161.447	199.499	215.7073	224.58	230.161	233.985	236.768	238.882	240.543	241.882
2	18.5128	19.000	19.1643	19.246	19.2964	19.329	19.3532	19.3710	19.3848	19.3959
3	10.128	9.5521	9.2761	9.1169	9.0129	8.9419	8.8856	8.8450	8.8127	8.7875
4	7.7086	6.9443	6.5913	6.3881	6.2559	6.1627	6.0939	6.0412	5.9989	5.9650
5	6.6079	5.7861	5.4097	5.1920	5.0505	4.9504	4.8761	4.8186	4.7724	4.7349
6	5.9874	5.1433	4.7572	4.5337	4.3872	4.2837	4.2065	4.1468	4.0992	4.0598
7	5.5914	4.7374	4.3468	4.1204	3.9716	3.8660	3.7870	3.7256	3.6765	3.6366
8	5.3177	4.4590	4.0662	3.8377	3.6874	3.5806	3.5004	3.4382	3.3881	3.3470
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3737	3.2928	3.2295	3.1789	3.1373
10	4.9646	4.1028	3.7082	3.4781	3.3259	3.2173	3.1354	3.0717	3.0203	2.9783
11	4.8443	3.9823	3.5875	3.3567	3.2039	3.0946	3.0123	2.9481	2.8963	2.8537
12	4.7472	3.8853	3.4904	3.2591	3.1058	2.9962	2.9134	2.8485	2.7963	2.7534
13	4.6672	3.8056	3.4106	3.1791	3.0255	2.9152	2.8320	2.7669	2.7143	2.6710
14	4.6001	3.7389	3.3440	3.1123	2.9582	2.8478	2.7643	2.6986	2.6458	2.6021
15	4.5431	3.6823	3.2874	3.0556	2.9012	2.7905	2.7066	2.6407	2.5876	2.5437
20	4.3512	3.4928	3.0985	2.8661	2.7109	2.5989	2.5140	2.4471	2.3929	2.3478
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3344	2.2662	2.2107	2.1646

n \ m	11	12	13	14	15	20	30
1	242.982	243.9056	244.6901	245.362	245.950	248.012	250.0935
2	19.4049	19.4125	19.4189	19.4244	19.429	19.4458	19.4624
3	8.7611	8.7448	8.7261	8.7164	8.7013	8.6625	8.6166
4	5.9351	5.9109	5.8902	5.8727	5.8578	5.8028	5.7450
5	4.7042	4.6774	4.6551	4.6361	4.6186	4.5582	4.4962
6	4.0272	4.0002	3.9765	3.9562	3.9383	3.8738	3.8077
7	3.6030	3.5745	3.5504	3.5293	3.5106	3.4448	3.3762
8	3.3128	3.2838	3.2589	3.2373	3.2183	3.1504	3.0797
9	3.1024	3.0728	3.0476	3.0256	3.0060	2.9365	2.8637
10	2.9430	2.9131	2.8871	2.8647	2.8450	2.7740	2.6997
11	2.8179	2.7876	2.7614	2.7386	2.7186	2.6465	2.5704
12	2.7173	2.6867	2.6602	2.6372	2.6168	2.5436	2.4664
13	2.6346	2.6036	2.5770	2.5535	2.5331	2.4588	2.3803
14	2.5655	2.5342	2.5072	2.4836	2.4630	2.3879	2.3082
15	2.5068	2.4752	2.4481	2.4243	2.4035	2.3275	2.2468
20	2.3099	2.2776	2.2494	2.2249	2.2033	2.1241	2.0391
30	2.1256	2.0921	2.0630	2.0374	2.0148	1.9316	1.8409

Test de Kolmogorov-Smirnov (ajustement sur la fonction de répartition F)

$$X = \sup_t |F_n(t) - F(t)|$$

n	t tel que $P(X > t) = 0.05$	t tel que $P(X > t) = 0.01$
5	0.56	0.67
6	0.52	0.62
7	0.47	0.56
8	0.45	0.54
9	0.42	0.50
10	0.41	0.49
15	0.34	0.40
20	0.29	0.36
25	0.27	0.32

Pour n grand, t est équivalent à $1,36/\sqrt{n}$ pour 5% et à $1,63/\sqrt{n}$ pour 1%.

Test signes et rangs (Wilcoxon)

La table donne la valeur critique k telle que :

n (taille de l'échantillon)	$P(X \leq k) \leq 5\%$ et $P(X \leq k+1) > 5\%$	$P(X \leq k) \leq 1\%$ et $P(X \leq k+1) > 1\%$
	k	k
5	0	
6	2	
7	3	0
8	5	1
9	8	3
10	10	5
11	13	7
12	17	9
13	21	12
14	25	15
15	30	19
16	35	23
17	41	27
18	47	32
19	53	37
20	60	43

$n \geq 21$: Approximation gaussienne avec $m = E(X) = \frac{n(n+1)}{4}$ et
 $\text{var}(X) = \frac{n(n+1)(2n+1)}{24}$.

Test de Mann-Whitney

L'échantillon des X est de taille n , celui des Y de taille m

$$V = \sum 1_{\{X_i > Y_j\}} = \sum \text{rg}(X_i) - n(n+1)/2$$

On pourra utiliser la relation $\sum \text{rg}(X_i) + \sum \text{rg}(Y_j) = (n+m)(n+m+1)/2$ pour calculer $P(V > k) = P(U < nm - k)$ avec

$$U = \sum \text{rg}(Y_j) - m(m+1)/2 = nm - V.$$

U et V ont même loi, symétrique par rapport à $nm/2$.

La table donne la valeur critique k telle que $P(V \leq k)$ soit le plus proche possible de 0,05.

$n \backslash m$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1														
2				0	0	0	1	1	1	1	2	2	2	3
3		0	0	1	2	2	3	3	4	5	5	6	7	7
4		0	1	2	3	4	5	6	7	8	9	10	11	12
5	0	1	2	4	5	6	8	9	11	12	13	15	16	18
6	0	2	3	5	7	8	10	12	14	16	17	19	21	23
7	0	2	4	6	8	11	13	15	17	19	21	24	26	28
8	1	3	5	8	10	13	15	18	20	23	26	28	31	33
9	1	3	6	9	12	15	18	21	24	27	30	33	36	39
10	1	4	7	11	14	17	20	24	27	31	34	37	41	44
11	1	5	8	12	16	19	23	27	31	34	38	42	46	50
12	2	5	9	13	17	21	26	30	34	38	42	47	51	55
13	2	6	10	15	19	24	28	33	37	42	47	51	56	61
14	2	7	11	16	21	26	31	36	41	46	51	56	61	66
15	3	7	12	18	23	28	33	39	44	50	55	61	66	72

BIBLIOGRAPHIE**Probabilités**

RENYL. — Calcul des probabilités. Ed. *Dunod*.

BREIMAN. — Probability. Ed. *Addison Wesley* (contient quelques beaux développements théoriques).

Exercices : JAFFARD. — Collection comprendre et appliquer. Ed. *Masson*.

Signal, codage

PICINBONO. — Théorie des signaux et systèmes. Ed. *Dunod, Université*.

MACCHI, GUILBERT. — Téléinformatique. Ed. *Masson*.

HOËL, PORT, STONE. — Introduction to stochastic processes. Ed. *Houghton Mifflin*.

Statistiques

SANDERS, MURPH, Eng. — « Les statistiques, une approche nouvelle ». Ed. *McGraw-Hill*.

Exercices : JAFFARD. — Collection comprendre et appliquer. Ed. *Masson*.

Pour approfondir

COMTET. — Analyse combinatoire. Collection Sup. Ed. *PUF*.

COTTELL, DUHAMEL, GENON-CATALOT. — Exercices de probabilités. Diffusion *Belin*.

NEVEU. — Bases mathématiques du calcul des probabilités. Ed. *Masson*.

DACUNHA-CASTELLE, DUFLO. — Probabilités et Statistiques, Ed. *Masson* (2 volumes + 1 volume d'exercices).

DYNKIN. — Théorie des processus markoviens. Ed. *Dunod*.

FAURE, CHRETIENNE. — Processus stochastiques, leurs graphes, leurs usages. Ed. *Gauthier-Villars*.

ROUBINE. — Introduction à la théorie de la communication. Ed. *Masson* (en 3 volumes).

AZENCOTT, DACUNHA-CASTELLE. — Séries d'observations irrégulières. Ed. *Masson*.

CRAMER, LEADBETTER. — Stationary and related stochastic processes. Ed. *Wiley*.

VAN LINT. — Coding theory. Ed. *Springer Verlag*.

GOPPA. — Geometric codes. Ed. *Kluwer, Academic Publishers*.

KNUTH. — The art of computer programming. Vol. 2. Ed. *Addison Wesley*.

HOLLANDER, WOLFE. — Non parametric statistical methods. Ed. *Wiley*.

TOMASSONE, LESQUOY, MILLIER. — La régression. Ed. *Masson*.

SCHEFFE. — Analysis of variance. Ed. *Wiley*.

Pour d'autres tables statistiques

S. KOKOSKA, C. NEVISON. — Statistical tables and formulae. Ed. *Springer-Verlag*.

INDEX

- Adaptation (d'une source à un canal), 143
ALOHA, 19, 67
- Bayes
 Formule de —, 8
 Estimateur de —, 154
Test bayésien —, 161
- Biais (estimateur sans), 150
BCH (code), 135
Bienayme-Tchebichev (inégalité de), 52
Bochner-Herglotz (théorème de), 93
- Bruit
 — de grenaille, 85
 — de quantification, 145
 — thermique, 108
- Capacité (d'un canal), 143, 147
Causal (filtre), 118, 125
Chernoff (inégalité de), 67, 145
Compression, 137
Contrôle de qualité, 172, 173
Cramer-Rao (inégalité de), 156
Cyclique (code), 133
- Débit d'information, 144
Démodulation, 173
Détection, 131, 144, 173
E.E.G. (électro-encéphalogramme), 85
Entropie, 142, 147, 148
- Facteurs, 191
Fano-Shannon (code de), 140
Fiabilité, E.1.1, 17, E.1.10, 18, 24, E.2.8 et E.2.9, 48
Fibonacci, E.4.2, 82
Files d'attente, 79
Filtre, 113
Fisher (information de) 156
Fractal, 88
- Gain complexe, 113
- Hamming
 code de —, 131, 135, 137
 distance de —, 132, 138
Huffmann (code d'), 141, 144, 147
- Lévy (théorème de), 61
Loève (Lemme de), 51
- Loi
 — a priori, 154
 — a posteriori, 155
 — de Bernoulli, 25
 — binomiale, 25
 — Cauchy, 29
 — Erlang, 29
 — exponentielle, 28
 — de Fisher, 167
 — Gamma, 29
 — de Gauss, 28
 — géométrique, 27
 — log-normale, 29, E.2.11, 48
 — multinomiale, 37, 194
 — de Poisson, 26, 64
 — de Rayleigh, 29
 — de Student, 163
 — de Weibull, 25
- Mann-Withney, 185, 195
Martingale, 58, 68
Médiane, 47, 48, 183, 195
Minitel, 136
Modulation, 104
Monte-Carlo (méthode de), 66
Mouvement brownien, 79, 86, 90
- Négligeable, 53, 61
Niveau (d'un test), 158, 161
Niveaux (d'un facteur), 191
- Parfait (code), 137
Préfixe (code), 139
Processus
 — de Poisson, 61
 — de Wiener, 90
Périodogramme, 101
Polynomial (code), 135
Pseudo-aléatoires (nombres)
- Quantiles, E.2.7, 48
Qui-deux
 Loi de —, 43, 64
 test d'ajustement du —, 177

test d'indépendance du —, 181	Simulation par rejet, 83
test sur la variance —, 165	Spearman (coefficient de corrélation de), 186
Rapport signal-bruit, 123, 124	Taux de transmission, 130
Résiduelle (variable), 190	Weldon (expérience de), 67
Série chronologique, 85	Wilcoxon (test de), 184, E.9.5, 195
Stable (filtre), 118, 119, 125	
Shannon (théorème de), 142	

Dans la même collection :

1. **Chabert (J.)**. — *Calcul des courbes de remous*, 1956, épuisé.
2. **Larras (J.)**. — *Plages et côtes de sable*, 1957, épuisé.
3. **Valembois (J.)**. — *Memento d'hydraulique pratique*, 1958, corrigé et mis à jour, 1963, nouveau tirage 1977, épuisé.
4. **Bouvard (M.)**. — *Barrages et prises d'eau en rivières*, 1958, 2^e édition revue et augmentée, 1960, épuisé.
5. **Ginocchio (R.)**. — *Aménagements hydroélectriques*, 1959, épuisé.
6. **Réméniéras (G.)**. — *L'hydrologie de l'Ingénieur*, 1960, 2^e éd. revue et augmentée, 1965, nouveaux tirages 1976, 1980, 1986.
7. **Lencastre (A.)**. — *Manuel d'hydraulique générale*, 1961, nouveau tirage 1976.
8. *La Mécanique des fluides et la magnétohydrodynamique*, 1962, épuisé.
9. **Ailleret (P.)**. — *Energétique. Les besoins d'énergie*, 1963, nouveau tirage 1974, épuisé.
10. **Condoyannis (G.)**. — *Le Russe scientifique*, 1964 (en vente chez Ophrys).
11. **Larras (J.)**. — *Embouchures, estuaires, lagunes et deltas*, 1964, épuisé.
12. **Schneebeli (G.)**. — *Hydraulique souterraine*, 1966, nouveaux tirages 1978, 1987.
13. **Idel'cik (I.E.)**. — *Memento des pertes de charge*, 1969, nouveaux tirages 1978, 1986.
14. **Carlier (M.)**. — *Hydraulique générale et appliquée*, 1972, nouveaux tirages 1980, 1986.
15. **Larras (J.)**. — *Hydraulique et granulats*, 1972.
16. **Théron (R.)**. — *Recherche sur l'étanchéité des lacs de barrage en pays karstique*, 1973.
17. *Aéro-Hydro-Elasticité*. (Cycle de conférences CEA-EDF, septembre 1972), 1973.
18. **Lebreton (J.C.)**. — *Dynamique fluviale*, 1974.
19. **Jamme (G.)**. — *Travaux fluviaux*, 1974.
20. **Chambadal (P.)**. — *La thermodynamique*, 1974.
21. *Techniques de mesure dans les écoulements*. (Cycle de conférences CEA-EDF, septembre 1973), 1974.
22. **Ishii (M.)**. — *Thermo-fluid dynamic theory of two-phase flow*, 1975.
23. **André (H.)**, **Audinet (M.)**, **Mazéran (G.)**, **Richer (C.)**. — *Hydrométrie pratique des cours d'eau*, 1976.
24. **Gary (C.)**, **Moreau (M.)**. — *L'effet de couronne en tension alternative*, 1976.
25. **Bonnin (J.)**. — *Hydraulique urbaine appliquée aux agglomérations de petite et moyenne importance*, 1977, nouveau tirage 1986.
26. **Chambadal (P.)**. — *La turbine à gaz*, 1976.
27. *Nouveaux systèmes énergétiques : cycles thermodynamiques associés* (Cycle de conférences CEA-EDF, septembre-octobre 1974), 1976.
28. *Les sources froides des centrales électriques*. (Cycle de conférences CEA-EDF, octobre 1975), 1977.
29. **Maurel (E.)**, **Roux (D.)**, **Dupont (D.)**. — *Techniques opérationnelles d'ordonnement*, 1977.
30. **Chouard (Ph.)**, **Michel (H.)**, **Simon (M.F.)**. — *Bilan thermique d'une maison solaire*, 1977.

31. **Valid (R.)**. — *La mécanique des milieux continus et le calcul des structures*, (Ecole d'été d'analyse numérique, C.E.A.-INRIA-E.D.F., 1976), 1977, épuisé.
32. *Phénomènes thermiques et hydrauliques non stationnaires*. (Cycle de conférences CEA-EDF, octobre 1976), 1978.
33. **Ginocchio (R.)**. — *L'énergie hydraulique*, 1978.
34. **Meyer (B.), Baudoin (C.)**. — *Méthodes de programmation*, 1978, nouveau tirage 1980, 3^e édition revue et corrigée 1984, nouveau tirage 1986.
35. **Bidard (R.), Bonnin (J.)**. — *Energétique et turbomachines*, 1979.
36. **Larras (J.)**. — *Physique de la houle et des lames*, 1979.
37. **Gondran (M.), Minoux (M.)**. — *Graphes et algorithmes*, 1979, 2^e édition, revue et complétée, 1985.
38. *Energie électrique et environnement*, 1980. (Publié sous la direction de R. Ginocchio).
39. **Pagès (A.), Gondran (M.)**. — *Fiabilité des systèmes*, 1980.
40. **Raviart (P.A.)**. — *Les méthodes d'éléments finis en mécanique des fluides* (Ecole d'été d'analyse numérique, CEA-EDF-INRIA, juillet 1979), 1981.
41. *La réalisation des logiciels graphiques interactifs* (Ecole d'été d'informatique, C.E.A.-E.D.F.-INRIA, juillet 1979, ouvrage réalisé sous la direction de Lucas (M.)), 1981, réimpression en 1985.
42. **Bonnin (J.)**. — *Aide-mémoire d'hydraulique urbaine*, 1982.
43. **Ailleret (P.)**. — *Essai de théorie de la normalisation*, 1982.
44. **Arrighi (R.)**. — *Evolution des disjoncteurs à haute et moyenne tensions*, 1983.
45. **Reynaud (J.F.)**. — *Concentration par évaporation et recompression mécanique de vapeur*, 1984.
46. **Liénart (P.), François (P.)**. — *Acoustique industrielle et environnement. I. Eléments fondamentaux d'acoustique physique et perceptive*. (Monographie publiée sous les auspices du Groupement des acousticiens de langue française), 1983.
47. **Bonnin (J.)**. — *L'eau dans l'antiquité : l'hydraulique avant notre ère*, 1984.
48. **Duderstadt (J.J.), Lewis (E.E.), Bardos (C.)**. — *Neutron transport equation* (Ecole d'été d'analyse numérique, CEA-EDF-INRIA, 1981), en anglais, 1983.
49. **Golub (G.H.), Meurant (G.A.)**. — *Résolution numérique des grands systèmes linéaires* (Ecole d'été d'analyse numérique, CEA-EDF-INRIA, 1980), 1983.
50. **Persoz (H.), Santucci (G.), Lemoine (J.C.), Sapet (P.)**. — *La planification des réseaux électriques*, 1984.
51. **Le Roy (G.), Gary (C.), Hutzler (B.), Lalot (J.), Dubanton (C.)**. — *Les propriétés diélectriques de l'air et les très hautes tensions*, 1984.
52. **SOGREAH, Laboratoire National d'Hydraulique**. — *Catalogue sédimentologique des côtes françaises, côtes de la Méditerranée* (ouvrage publié sous les auspices du Secrétariat d'Etat auprès du Ministre des Transports, chargé de la mer), 1984.
53. **Miquel (J.)**. — *Guide pratique d'estimation des probabilités de crues*, 1984.
54. **Bouvard (M.)**. — *Barrages mobiles et ouvrages de dérivation, à partir de rivières transportant des matériaux solides* (En collaboration avec la Section d'Hydraulique Fluviale de l'A.I.R.H.), 1984.

55. **Mathieu (J.), Jeandel (D.)**. — *La simulation des modèles de turbulence et leurs applications, Volume 1* (Ecole d'été d'analyse numérique, C.E.A.-E.D.F.-INRIA, juillet 1982), 1984.
56. **Launder (B.E.), Reynolds (W.C.), Rodi (W.)**. — *Turbulence models and their applications, Volume 2* (Ecole d'été d'analyse numérique, C.E.A.-E.D.F.-INRIA, juillet 1982), en anglais, 1984.
57. **Bergman (D.), Lions (J.L.), Papanicolaou (G.), Murat (F.), Sanchez-Palencia (E.), Tartar (L.)**. — *Les méthodes de l'homogénéisation : théorie et applications en physique* (Ecole d'été d'analyse numérique, C.E.A.-INRIA-E.D.F., juillet 1983), 1985.
58. **Cormary (Y.), Nicolas (C.)**. — *La thermique des serres*, 1985.
59. **Borghi (R.), Clavin (P.), Liñán (A.), Pelcé (P.), Sivashinsky (G.I.)**. — *Modélisation des phénomènes de combustion* (Ecole d'été sur les problèmes non linéaires appliqués, C.E.A.-INRIA-E.D.F., octobre 1983), 1985.
60. *Fortran 77. Guide pour l'écriture de programmes portables*, ouvrage réalisé sous la direction de Ficheux-Vapné (F.), 1985.
61. **Laboratoire Central d'Hydraulique de France, Laboratoire National d'Hydraulique**. — *Catalogue sédimentologique des côtes françaises, de la frontière belge au Mont-Saint-Michel* (ouvrage publié sous les auspices du Secrétariat d'Etat auprès du Ministre de l'Urbanisme, du Logement et des Transports, chargé de la mer), 1986.
62. **Biau (D.)**. — *Les applications industrielles du chauffage par rayonnement infrarouge*, 1986.
63. **Fournié (R.)**. — *Les isolants en électrotechnique. Concepts et théories*, 1986.
64. **Bernier (J.), Chaloin (B.), Coeffé (Y.), Feuillet (J.)**. — *Le dimensionnement des digues à talus*, 1987, épuisé.
65. **Laboratoire Central d'Hydraulique de France**. — *Catalogue sédimentologique des côtes françaises, de la baie du Mont-Saint-Michel à la frontière espagnole* (ouvrage publié sous les auspices du Secrétariat d'Etat à la mer), 1987.
66. **Lesueur (C.)**. — *Rayonnement acoustique des structures ; Vibroacoustique, Interactions fluide-structure*, 1988.
67. **Villemeur (A.)**. — *La sûreté de fonctionnement des systèmes industriels*, 1988.
68. *Méthodes spectrales* (Ecole d'été sur les problèmes non linéaires appliqués C.E.A.-INRIA-E.D.F., janvier 1986), 1988.
69. **Gibert (R.J.)**. — *Vibrations des structures — Interactions avec les fluides — Sources d'excitation aléatoires* (Ecole d'été d'analyse numérique, C.E.A.-E.D.F.-INRIA, juillet 1986), 1988.
70. **Bamberger (Y.), Donéa (J.), Laval (H.), Planchard (J.), Shaw (R.P.)**. — *Aspects théoriques et numériques de la dynamique des structures* (Ecole d'été d'analyse numérique, C.E.A.-E.D.F.-INRIA, juillet 1986), 1988.
71. **Cladé (J.)**. — *Electrotechnique*, 1989.
72. **Frêne (J.), Nicolas (D.), Degueurce (B.), Berthe (D.), Godet (M.)**. — *Lubrification hydrodynamique. Paliers et Butées*, 1990.

ESE 1. **Roubine (E.)**. — *Distributions-Signal*, 1982, épuisé.

ESE 2. **Barret (P.)**. — *Régimes transitoires des machines tournantes électriques*, 1982, 2^e édition revue et corrigée, 1987.

- ESE 3. **Dattée (F.)**. — *Electronique. Concepts de base* (avec la collaboration de D. Venot), 1983, 2^e édition revue et corrigée, 1987.
- ESE 4. **Mésa (F.)**. — *Méthodes d'études des circuits électriques*, 1985, 2^e édition revue et corrigée, 1987, nouveau tirage 1989.
- ESE 5. **Escané (J.M.)**. — *Synthèse des circuits passifs et actifs - Filtres*, 1987.
- ESE 6. **Laborne (H.)**. — *Convertisseurs assistés par un réseau alternatif. Volume 1. Concepts, critères de choix, dimensionnement*, 1989.
- ESE 7. **Laborne (H.)**. — *Convertisseurs assistés par un réseau alternatif. Volume 2. Environnement*, en cours d'édition.
- ESE 8. **Brouaye (F.)**. — *La modélisation des incertitudes. Probabilités, Signaux et communications, Statistiques*, 1990.

Dans les publications de cette collection, la Direction des Etudes et recherches d'Electricité de France laisse aux auteurs la liberté d'exprimer leur point de vue personnel sur les sujets traités.