

Enseignement des mathématiques

Initiation aux probabilités

**Traduction de la quatrième
édition américaine**

Sheldon M. Ross

Traduit de l'américain par Christian Hofer et Frédéric Dorsaz

DANS LA COLLECTION «ENSEIGNEMENT DES MATHÉMATIQUES»
DIRIGÉE PAR LE PROFESSEUR ROBERT C. DALANG

Calcul différentiel et intégral

Jacques Douchet et Bruno Zwahlen

1 *Fonctions réelles d'une variable réelle*

2 *Fonctions réelles de plusieurs variables réelles – Exercices résolus*

3 *Fonctions réelles d'une variable réelle – Exercices résolus*

4 *Fonctions réelles de plusieurs variables réelles – Exercices résolus*

Algèbre linéaire

Renzo Cairoli

Cours d'analyse, 3 volumes

Srishti-D. Chatterji

Algèbre linéaire

Robert C. Dalang, Amel Chaabouni

Recherche opérationnelle pour ingénieurs I

Dominique de Werra, Thomas M. Liebling, Jean-François Hêche

Recherche opérationnelle pour ingénieurs II

Jean-François Hêche, Thomas M. Liebling, Dominique de Werra

Analyse, Recueil d'exercices et aide-mémoire vol. 1 et 2

Jacques Douchet

Analyse avancée pour ingénieur

Bernard Dacorogna, Chiara Tanteri

Introduction à l'analyse numérique

Jacques Rappaz, Marco Picasso

Les Presses polytechniques et universitaires romandes sont une fondation scientifique dont le but est principalement la diffusion des travaux de l'École polytechnique fédérale de Lausanne, d'autres universités francophones ainsi que des écoles techniques supérieures. Le catalogue de leurs publications peut être obtenu par courrier aux Presses polytechniques et universitaires romandes, EPFL – Centre Midi, CH-1015 Lausanne, par E-Mail à ppur@epfl.ch, par téléphone au (0)21 693 41 40, ou par fax au (0)21 693 40 27.

www.ppur.org

Version originale: **A first course in probability**

Copyright © 1994 (1976, 1984, 1988) Macmillan Publishing Company

Traduction de la quatrième édition américaine revue et augmentée

ISBN 2-88074-327-3

© 1987, 1990, 1995, 1996, 1999, 2002, 2004, Presses polytechniques et universitaires romandes, CH – 1015 Lausanne

Tous droits réservés.

Reproduction, même partielle, sous quelque forme ou sur quelque support que ce soit, interdite sans l'accord écrit de l'éditeur.

Imprimé en Suisse

Avant-propos à l'édition française

Publié par Macmillan, New York, sous le titre «A first course in probability», cet ouvrage a été écrit en anglais en 1976 par Sheldon M. Ross. Il fait partie de la nouvelle génération de textes d'introduction au calcul des probabilités. Nouveau dans le sens qu'il s'éloigne de l'ouvrage phare de Feller «An introduction to probability theory and its applications», en mettant l'accent sur les notions de probabilité conditionnelle, plutôt que sur les aspects combinatoires de la probabilité.

Dès sa parution, ce livre a connu un succès énorme auprès des universités et des collèges américains. Le succès de la traduction française, publiée par les Presses polytechniques romandes en 1987, n'en fut pas moins grand car l'ouvrage comblait une lacune de la littérature traitant du calcul des probabilités en langue française. Il a très vite été reconnu comme un texte excellemment adapté aux besoins des étudiants qui cherchent une introduction au sujet à la fois directe et rigoureuse sans un appareil mathématique trop lourd.

Cette troisième édition française, basée sur la quatrième version américaine de 1994, retient le même nombre de chapitres que la précédente, mais se distingue par une nouvelle organisation de la matière. En particulier, les notions de l'espérance mathématique et de la variance d'une variable aléatoire sont avancées au chapitre 4, immédiatement après l'introduction des variables aléatoires discrètes. Cela entraîne des changements dans les chapitres suivants et donne lieu à toute une classe de problèmes nouveaux. Par conséquent, le nombre total des exercices théoriques et des problèmes s'élève dans cette édition à presque 600. Le succès du texte de S. Ross réside en premier lieu dans le choix excellent des exemples, exercices et problèmes et cet ouvrage restera très attractif dans les années à venir, pour des cours d'initiation aux probabilités au niveau universitaire.

Cette nouvelle version a été préparée par Monsieur F. Dorsaz. Je le remercie du soin et de l'engagement manifestés pour cette tâche.

Peter Nüesch

Préface

«... On réalise en fin de compte que la théorie des probabilités n'est tout simplement que le bon sens réduit à du calcul. Elle nous fait apprécier avec exactitude ce que l'esprit bien fait sent déjà par une sorte d'instinct, souvent sans être capable d'en rendre compte... Il est remarquable que cette science, qui a pris son origine dans l'étude des jeux de chance, soit devenue l'objet le plus important de la connaissance humaine. Les questions les plus importantes de la vie ne sont en réalité, pour l'essentiel, que des problèmes de probabilité».

Ainsi pensait le «Newton» des Français, le célèbre mathématicien et astronome Pierre Simon, marquis de Laplace. On est en droit de penser que l'illustre marquis – qui fut d'ailleurs l'un des grands contributeurs à l'essor des probabilités – a un peu exagéré. Il n'en est pas moins certain que la théorie des probabilités est devenue un outil d'importance fondamentale pour un nombre considérable de scientifiques, d'ingénieurs, de médecins, de juristes et d'industriels. En fait l'homme éclairé a appris à ne plus demander «est-ce ainsi?» mais plutôt «quelle est la probabilité qu'il en soit ainsi?».

Ce livre se veut une introduction élémentaire à la théorie mathématique des probabilités pour les étudiants qui possèdent assez de connaissances préalables en calcul différentiel et intégral, qu'ils travaillent en mathématiques, dans les sciences de l'ingénieur et même dans n'importe quelle science en général (y compris les sciences sociales et du management). Il essaie de présenter non seulement la partie mathématique de la théorie des probabilités mais aussi, et à travers une foule d'exemples, les nombreuses applications possibles de cette connaissance.

Dans le chapitre 1 sont présentés les principes de base de l'analyse combinatoire, qui sont extrêmement utiles pour le calcul des probabilités.

Dans le chapitre 2 on considère les axiomes de la théorie des probabilités et on montre comment ils peuvent être utilisés pour calculer les probabilités auxquelles on s'intéresse. Ce chapitre inclut une preuve de l'importante (et malheureusement souvent négligée) propriété de continuité des probabilités, qui est alors utilisée pour la résolution d'un paradoxe.

Le chapitre 3 traite des très importantes notions de probabilité conditionnelle et d'indépendance d'événements. Par une série d'exemples, nous illustrerons comment les probabilités conditionnelles interviennent non seulement quand des informations partielles sont disponibles mais aussi comme outils pour nous permettre de calculer des probabilités plus facilement, même si aucune information partielle n'est présente. Cette technique qui permet efficacement d'obtenir des probabilités en conditionnant réapparaît au chapitre 7, où nous l'utilisons avec la notion d'espérance conditionnelle.

Dans les chapitres 4, 5 et 6 est discuté le concept de variable aléatoire. Les variables aléatoires discrètes sont traitées au chapitre 4, les variables continues au chapitre 5 et les variables conjointes au chapitre 6. Les importants concepts d'espérance et de variance d'une variable aléatoire sont introduits dans les chapitres 4 et 5. Ces quantités sont alors déterminées pour plusieurs types courants de variables aléatoires.

Des propriétés supplémentaires de l'espérance sont présentées dans le chapitre 7. De nombreux exemples illustrant l'utilité du résultat «l'espérance d'une somme de variables aléatoires est égale à la somme de leurs espérances» sont également donnés. Ce chapitre comprend d'autre part une section sur l'espérance conditionnelle, incluant son utilisation en vue de la prédiction, et une autre sur les fonctions génératrices des moments. Enfin, la dernière section présente la distribution normale multivariée ainsi qu'une preuve simple concernant la distribution conjointe de la moyenne et de la variance d'un échantillon provenant d'une distribution normale.

Au chapitre 8 sont présentés les principaux résultats théoriques de la théorie des probabilités. Nous démontrerons en particulier la loi forte des grands nombres et le théorème central limite. Notre démonstration de la loi forte est relativement simple en admettant que les variables aléatoires ont un quatrième moment fini, et celle du théorème central limite repose sur le théorème de continuité de Levy. Des inégalités sur les probabilités sont aussi présentées dans ce chapitre, telles que l'inégalité de Markov, celle de Chebyshev et les bornes de Chernoff. La dernière section du chapitre 8 donne une borne pour l'erreur induite par l'approximation d'une probabilité concernant la somme de variables aléatoires indépendantes de Bernoulli par la probabilité correspondante d'une variable aléatoire de Poisson de même espérance.

Le chapitre 9 présente quelques thèmes choisis tels que les chaînes de Markov, le processus de Poisson ainsi qu'une introduction à la théorie de l'information et du codage.

Le chapitre 10 traite des aspects de la simulation de façon plus étoffée que dans l'édition précédente.

De nombreux exemples sont traités tout au long du texte et le lecteur trouvera aussi quantité d'exercices – où l'on a distingué des exercices théoriques et des problèmes – proposés pour approfondissement. Un grand soin a été porté à la formulation de ces exemples et problèmes. Une solution à la plupart des problèmes est indiquée à la fin de l'ouvrage tandis que pour les enseignants un recueil de solutions est disponible.¹

Nous aimerions remercier les correcteurs suivants: Thomas R. Fischer, Texas A & M University; Jay Devore, California Polytechnic University, San Luis Obispo; Robb J. Muirhead, University of Michigan; David Heath, Cornell University; M. Samuels, Purdue University; I.R. Savage, Yale University; R. Müller, Stanford University; K. B. Athreya, Iowa State University; Phillip Beckwith, Michigan Tech; Howard Bird, St. Cloud State University; Steven Chiappari, Santa Clara University; James Clay, University of Arizona at Tucson; Francis Conlan, University of Santa Clara; Fred Leysieffer, Florida State University; Ian McKeague, Florida State University; Helmut Mayer, University of Georgia; N. U. Prabhu, Cornell University; Art Schwartz, University of Michigan at Ann Arbor; Therese Shelton, Southwestern University; and Allen Webster, Bradley University.

¹ Seulement dans la version anglaise. Pour l'obtenir, s'adresser directement à Macmillan Publishing Company 866 Third Avenue, New York, New York 10 022.

Table des matières

PRÉFACE	VII
CHAPITRE 1 ANALYSE COMBINATOIRE	
1.1 Introduction	1
1.2 Principe fondamental de dénombrement	2
1.3 Permutations	3
1.4 Combinaisons	6
1.5 Coefficients multinomiaux	10
1.6 Répartition de boules dans des urnes	12
1.7 Exercices théoriques	14
1.8 Problèmes	17
CHAPITRE 2 AXIOMES DES PROBABILITÉS	
2.1 Introduction	23
2.2 Ensemble fondamental et événement	23
2.3 Axiomes des probabilités	28
2.4 Quelques théorèmes élémentaires	31
2.5 Ensembles fondamentaux à événements élémentaires équiprobables	35
2.6 Théorème de passage à la limite	44
2.7 Probabilité en tant que mesure du crédit accordé à un fait	48
2.8 Exercices théoriques	49
2.9 Problèmes	53
CHAPITRE 3 PROBABILITÉ CONDITIONNELLE ET INDÉPENDANCE	
3.1 Introduction	61
3.2 Probabilités conditionnelles	61
3.3 Formule de Bayes	66

	3.4 Événements indépendants	75
	3.5 Fonction de probabilité conditionnelle	90
	3.6 Exercices théoriques	97
	3.7 Problèmes	102
CHAPITRE 4	VARIABLES ALÉATOIRES	
	4.1 Variables aléatoires	115
	4.2 Fonctions de répartition	120
	4.3 Variables aléatoires discrètes	123
	4.4 Espérance	126
	4.5 Espérance d'une fonction d'une variable aléatoire	128
	4.6 Variance	132
	4.7 Variable de Bernoulli et variable binomiale	134
	4.8 Variable aléatoire de Poisson	144
	4.9 Autres lois discrètes	154
	4.10 Exercices théoriques	163
	4.11 Problèmes	168
CHAPITRE 5	VARIABLES ALÉATOIRES CONTINUES	
	5.1 Introduction	183
	5.2 Espérance et variance de variables aléatoires continus ...	186
	5.3 Variable aléatoire uniforme	191
	5.4 Variables aléatoires normales	195
	5.5 Variables aléatoires exponentielles	208
	5.6 Autres distributions continues	216
	5.7 Distribution d'une fonction de variable aléatoire	221
	5.8 Exercices théoriques	223
	5.9 Problèmes	228
CHAPITRE 6	VARIABLES ALÉATOIRES SIMULTANÉES	
	6.1 Définition des distributions simultanées	235
	6.2 Variables aléatoires indépendantes	245
	6.3 Sommes de variables aléatoires indépendantes	255
	6.4 Distributions conditionnelles	260
	6.5 Statistiques d'ordre	265
	6.6 Changement de variables multidimensionnelles	270
	6.7 Exercices théoriques	277
	6.8 Problèmes	281
CHAPITRE 7	PROPRIÉTÉS DE L'ESPÉRANCE	
	7.1 Introduction	289
	7.2 Espérance d'une somme de variables aléatoires	290
	7.3 Covariance, variance de sommes, corrélation	305
	7.4 Espérance conditionnelle	316
	7.5 Espérance conditionnelle et prédiction	328
	7.6 Fonctions génératrices des moments	333
	7.7 Autres propriétés des variables aléatoires normales	334
	7.8 Définition générale de l'espérance mathématique	347

	7.9 Exercices théoriques	349
	7.10 Problèmes	358
CHAPITRE 8	THÉORÈMES LIMITES	
	8.1 Introduction	371
	8.2 Loi faible des grands nombres	371
	8.3 Théorème central limite	375
	8.4 Loi forte des grands nombres	382
	8.5 Autres inégalités	385
	8.6 Bornes pour l'erreur de probabilité commise en approximant une loi binomiale par une loi de Poisson	391
	8.7 Exercices théoriques	394
	8.8 Problèmes	396
CHAPITRE 9	THÈMES CHOISIS DE PROBABILITÉ	
	9.1 Processus de Poisson	401
	9.2 Chaînes de Markov	404
	9.3 Surprise, incertitude, entropie	410
	9.4 Théorie du codage et entropie	414
	9.5 Exercices théoriques et problèmes	420
	9.6 Références	423
CHAPITRE 10	SIMULATION	
	10.1 Introduction	425
	10.2 Techniques générales pour la simulation de variables aléatoires continues	428
	10.3 Simulation de variables aléatoires discrètes	436
	10.4 Techniques de la réduction de la variance	438
	10.5 Problèmes	442
	SOLUTIONS À QUELQUES PROBLÈMES CHOISIS	447
	INDEX	453

CHAPITRE I

Analyse combinatoire

1.1 INTRODUCTION

Examinons d'emblée un problème typique de ceux mettant en jeu la notion de probabilité. Un système de communication est composé de n antennes identiques alignées. Ce système ne pourra alors capter de signal incident – il sera alors qualifié de fonctionnel – qu'aussi longtemps que deux antennes consécutives ne seront pas défectueuses. Si on découvre qu'exactly m des n antennes sont défectueuses, quelle est la probabilité que ce système reste fonctionnel?

Etudions par exemple le cas particulier où $n = 4$ et $m = 2$. Le système peut alors se trouver dans l'une des 6 configurations suivantes:

0	1	1	0
0	1	0	1
1	0	1	0
0	0	1	1
1	0	0	1
1	1	0	0

où 1 signifie que l'antenne fonctionne et 0 qu'elle est défectueuse. Comme notre système sera fonctionnel dans les trois premières configurations mais pas dans les trois dernières, il semble raisonnable d'attribuer à la probabilité cherchée la valeur $\frac{3}{6} = \frac{1}{2}$. On pourrait de manière similaire calculer la probabilité que le système fonctionne pour des valeurs quelconques de m et de n . Plus précisément il faudrait calculer le nombre de configurations qui maintiennent le système fonctionnel et le diviser par le nombre de toutes les configurations possibles.

Cet exemple permet de réaliser qu'il est souhaitable de disposer d'une méthode efficace pour dénombrer les différentes situations pouvant se présenter lors d'une expérience. En fait, bien des problèmes en théorie des probabilités peuvent être résolus simplement en comptant le nombre de manières différentes selon lesquelles un certain événement peut se réaliser. Par convention on appelle *analyse combinatoire* la théorie mathématique du dénombrement.

1.2 PRINCIPE FONDAMENTAL DE DÉNOMBREMENT

1.2.1 Version restreinte

Ce *principe de dénombrement* (ci-dessous théorème 1.1) sera essentiel par la suite. Il établit en gros que si une expérience peut produire m résultats et une autre n , alors il y a mn résultats possibles lorsqu'on considère ces deux expériences ensemble.

Théorème 1.1

Supposons qu'il faille réaliser deux expériences. Si l'expérience 1 peut produire l'un quelconque de m résultats et si, pour chacun d'entre eux, il y a n résultats possibles pour l'expérience 2, alors il existe mn résultats pour les deux expériences prises ensemble.

DÉMONSTRATION. On peut obtenir la démonstration en énumérant tous les résultats des deux expériences comme suit:

$$\begin{array}{l} (1, 1), (1, 2), \dots, (1, n) \\ (2, 1), (2, 2), \dots, (2, n) \\ \vdots \\ (m, 1), (m, 2), \dots, (m, n) \end{array}$$

Dans ce tableau un résultat a été noté (i, j) si l'expérience 1 a produit le i -ème de ses résultats et si l'expérience 2 a produit le j -ème des siens. On voit que l'ensemble des résultats possibles est composé de m lignes de n éléments chacune, ce qui démontre le résultat annoncé. ■

Exemple 1.1 Une petite communauté se compose de dix hommes et de leurs fils, chaque homme ayant trois fils. Si un homme et l'un de ses fils doivent être désignés «père et fils exemplaires», combien y a-t-il de choix différents possibles?

SOLUTION. En considérant le choix du père comme la première expérience et ensuite le choix de l'un de ses fils comme la seconde, nous concluons d'après le principe fondamental qu'il y a $10 \cdot 3 = 30$ choix possibles. ■

1.2.2 Principe fondamental généralisé

Lorsqu'il y a plus de deux expériences à réaliser, le principe fondamental peut être généralisé comme suit:

Théorème 1.2

Si r expériences doivent être réalisées et sont telles que la première peut produire l'un quelconque de n_1 résultats, et si pour chacun d'entre eux il y a n_2 résultats possibles pour la 2^e expérience, et si pour chaque résultat des deux premières expériences il y en a n_3 pour la 3^e expérience, et ainsi de suite, il y aura alors au total $n_1 \cdot n_2 \cdot \dots \cdot n_r$ résultats pour les r expériences prises ensemble.

1.2.3 Exemples d'applications du principe fondamental

Exemple 1.2 Le comité de planification d'un collège est constitué de 3 étudiants de première année, 4 de deuxième, 5 de troisième et 2 de dernière année. Un sous-comité de 4 étudiants comportant un représentant de chaque classe doit être choisi. Combien peut-on former de sous-comités?

SOLUTION. Nous pouvons considérer le choix d'un sous-comité comme le résultat combiné de 4 expériences distinctes, chacune consistant à choisir un unique représentant dans l'une des classes. Par conséquent, en application de la version généralisée du principe fondamental, il y a $3 \cdot 4 \cdot 5 \cdot 2 = 120$ sous-comités possibles. ■

Exemple 1.3 Combien de plaques minéralogiques portant un matricule de 7 caractères peut-on former si les 3 premiers caractères sont des lettres et les 4 derniers des chiffres?

SOLUTION. En application de la version généralisée du principe de base, la réponse est $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 175\,760\,000$. ■

Exemple 1.4 Combien de fonctions définies sur n points peut-on construire si ces fonctions ne peuvent prendre pour valeur que 0 ou 1?

SOLUTION. Numérotons de 1 à n les points. Puisque $f(i)$ ne peut prendre pour chaque $i = 1, 2, \dots, n$ que deux valeurs, il y a 2^n de ces fonctions. ■

Exemple 1.5 Dans l'exemple 1.3, combien de plaques minéralogiques pourrait-on avoir si l'on excluait que les lettres ou les chiffres se répètent?

SOLUTION. Dans ce cas, il y aurait $26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 \cdot 7 = 78\,624\,000$ plaques possibles. ■

1.3 PERMUTATIONS

1.3.1 Permutations d'objets distinguables

Combien existe-t-il d'arrangements ordonnés des lettres a , b et c ? Par énumération directe nous en trouvons 6, à savoir: abc , acb , bac , bca , cab et cba . Chaque arrangement est par convention appelé *permutation*. Il y a ainsi 6 permutations possibles des éléments d'un ensemble de 3 objets. Ce résultat aurait également pu être construit à partir du principe fondamental; la première lettre de la permutation peut être n'importe laquelle des 3, la deuxième lettre peut ensuite être choisie parmi les 2 restantes tandis que la troisième ne peut plus faire l'objet d'aucun choix. Ainsi, il y a $3 \cdot 2 \cdot 1 = 6$ permutations possibles.

L'expression $n!$, dite *n factorielle* est définie par l'équation

$$n! = n(n-1)(n-2) \dots 3 \cdot 2 \cdot 1 \quad (1.1)$$

Supposons maintenant que nous ayons n objets. Un raisonnement analogue à celui que nous venons d'utiliser ci-dessus établit le théorème suivant:

Théorème 1.3

Le nombre de permutations de n objets est $n!$.

1.3.2 Exemples de permutations

Exemple 1.6 Combien d'ordres à la batte peut-on avoir pour une équipe de baseball de 9 joueurs¹?

SOLUTION. Il existe $9! = 362\,880$ ordres selon lesquels les joueurs peuvent se succéder à la batte. ■

Exemple 1.7 Un cours de théorie des probabilités est suivi par 6 hommes et 4 femmes. Un examen a lieu, puis les étudiants sont classés selon leur note. On suppose exclu que deux étudiants obtiennent la même note.

- Combien de classements peut-on avoir?
- Si les hommes sont classés entre eux uniquement et les femmes entre elles, combien de classements globaux peut-on avoir?

SOLUTION.

- Comme chaque classement correspond à un certain arrangement ordonné de 10 personnes, on voit que la réponse à cette partie du problème est $10! = 3\,628\,800$.
- Comme il y a $6!$ classements des hommes entre eux et $4!$ classements des femmes entre elles, il résulte par application du principe fondamental qu'il y aura dans ce cas $(6!)(4!) = (720)(24) = 17\,280$ classements possibles. ■

Exemple 1.8 M. Jones va disposer 10 livres sur un rayon de sa bibliothèque. Quatre d'entre eux sont des livres de mathématiques, trois de chimie, deux d'histoire et un de langue. Jones aimerait ranger ses livres de façon que tous les livres traitant du même sujet restent groupés. Combien y a-t-il de dispositions possibles?

SOLUTION. Il y a $4! 3! 2! 1!$ dispositions telles que les livres de mathématiques se présentent devant, derrière eux les livres de chimie, puis ceux d'histoire, enfin celui de langue. Pour chaque autre ordre de présentation des sujets il y a de même $4! 3! 2! 1!$ dispositions des livres. Par conséquent, comme ces ordres de présentation des sujets sont au nombre de $4!$, la réponse cherchée est $4! 4! 3! 2! 1! = 6\,912$. ■

1.3.3 Permutations d'objets partiellement indistinguables

Nous allons maintenant nous attacher à déterminer le nombre de permutations dans un ensemble de n objets quand certains de ces objets sont indistinguables les uns des autres. Pour mieux saisir de quoi il s'agit, considérons l'exemple suivant:

¹ Ndt: A tour de rôle, tous les joueurs doivent servir à la batte lors du jeu, sans répétition.

Exemple 1.9 Combien d'arrangements différents peut-on former avec les lettres $P E P P E R$?

SOLUTION. On remarquera d'abord qu'il existe $6!$ permutations des lettres $P_1 E_1 P_2 P_3 E_2 R$ lorsque les trois P et les deux E sont distingués les uns des autres. Cependant, considérons l'une quelconque de ces permutations – $P_1 P_2 E_1 P_3 E_2 R$ par exemple –. Si nous permutoons les P entre eux et les E entre eux, l'arrangement résultant sera encore de la même forme $P P E P E R$. En fait, chacune des $3! 2!$ permutations

$$\begin{array}{ll} P_1 P_2 E_1 P_3 E_2 R & P_1 P_2 E_2 P_3 E_1 R \\ P_1 P_3 E_1 P_2 E_2 R & P_1 P_3 E_2 P_2 E_1 R \\ P_2 P_1 E_1 P_3 E_2 R & P_2 P_1 E_2 P_3 E_1 R \\ P_2 P_3 E_1 P_1 E_2 R & P_2 P_3 E_2 P_1 E_1 R \\ P_3 P_1 E_1 P_2 E_2 R & P_3 P_1 E_2 P_2 E_1 R \\ P_3 P_2 E_1 P_1 E_2 R & P_3 P_2 E_2 P_1 E_1 R \end{array}$$

est de la forme $P P E P E R$. Par conséquent il y aura $6!/(3! 2!) = 60$ arrangements possibles des lettres $P E P P E R$. ■

Plus généralement, grâce au même raisonnement que celui utilisé dans l'exemple 1.9, on établit le théorème suivant:

Théorème 1.4

Il y a

$$\frac{n!}{n_1! n_2! \cdots n_r!} \quad (1.2)$$

permutations différentes de n objets parmi lesquels n_1 sont indistinguables entre eux, n_2 autres entre eux également, ..., n_r entre eux.

1.3.4 Exemples de permutations d'objets partiellement indistinguables

Exemple 1.10 Parmi les 10 participants à un tournoi d'échec, on compte 4 russes, 3 américains, 2 anglais et un brésilien. Si dans le classement du tournoi on ne peut lire que la liste des nationalités des joueurs mais pas leur identité, à combien de classements individuels différents une telle liste correspond-elle?

SOLUTION. Il y a

$$\frac{10!}{4! 3! 2! 1!} = 12,600$$

classements possibles. ■

Exemple 1.11 On compose des signaux en alignant des drapeaux suspendus. Combien de ces signaux peut-on former si parmi les drapeaux à disposition 4 sont blancs, 3 sont rouges, 2 sont bleus et si tous les drapeaux d'une même couleur sont indistinguables?

SOLUTION. Il y a

$$\frac{9!}{4!3!2!} = 1260$$

signaux différents. ■

1.4 COMBINAISONS

1.4.1 Définitions

Nous serons souvent intéressés à déterminer le nombre de groupes de r objets qu'il est possible de former sans répétition à partir d'un total de n objets. Par exemple, combien de groupes de 3 objets peut-on construire en tirant parmi les 5 objets A , B , C , D et E ? Pour y répondre, on peut raisonner comme suit: puisqu'il y a 5 façons de choisir le premier objet, 4 de choisir ensuite le deuxième et 3 de choisir le dernier, il y a donc $5 \cdot 4 \cdot 3$ façons de composer des groupes de 3 objets en tenant compte de l'ordre dans lequel ces objets sont choisis. Cependant, un triplet donné, par exemple le triplet constitué des objets A , B et C , apparaîtra 6 fois. En effet, chacune des permutations ABC , ACB , BAC , BCA , CAB et CBA sera distinguée lorsqu'on tient compte de l'ordre. Il en résulte que le nombre total de groupes pouvant être formés est

$$\frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10$$

Plus généralement, $n(n-1) \dots (n-r+1)$ représente le nombre de manières de choisir un groupe de r objets parmi n lorsqu'on tient compte de l'ordre. Comme chaque groupe de r objets sera distingué $r!$ fois dans ce dénombrement, le nombre de groupes de r objets pris dans un ensemble de n sera

$$\frac{n(n-1) \cdots (n-r+1)}{r!} = \frac{n!}{(n-r)! r!}$$

L'expression $\binom{n}{r}$, pour $r \leq n$, est définie par l'équation²

$$\binom{n}{r} = \frac{n!}{(n-r)! r!} \quad (1.3)$$

Une *combinaison* de r objets pris parmi n est tout sous-ensemble de r objets choisis sans répétition dans un ensemble en contenant n .

Théorème 1.5

$\binom{n}{r}$ est le nombre de combinaisons de r objets pris parmi n , ou encore le nombre de groupes de taille r si, dans le choix, l'ordre n'est pas considéré comme significatif.

² Par convention $0!$ a pour valeur 1. Donc $\binom{n}{0} = \binom{n}{n} = 1$. De plus $\binom{n}{i} = 0$ lorsque $i < 0$ ou $i > n$.

1.4.2 Exemples de calcul de combinaisons

Exemple 1.12 On veut former un comité comprenant 3 des 20 personnes d'un groupe. Combien y a-t-il de ces comités?

SOLUTION. Il y a $\binom{20}{3} = \frac{20 \cdot 19 \cdot 18}{3 \cdot 2 \cdot 1} = 1\,140$ comités possibles. ■

Exemple 1.13 A partir d'un groupe de 5 hommes et de 7 femmes, combien de comités différents composés de 2 hommes et de 3 femmes peut-on former? Qu'en est-il si 2 des femmes s'entendent mal et refusent de siéger simultanément au comité?

SOLUTION. Comme il y a $\binom{5}{2}$ groupes possibles de 2 hommes et $\binom{7}{3}$ groupes possibles de 3 femmes, il y a selon le principe fondamental $\binom{5}{2} \cdot \binom{7}{3} = \frac{5 \cdot 4}{3 \cdot 2} \cdot \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 350$ comités de 2 hommes et 3 femmes.

Considérons maintenant le cas où deux des femmes refusent de siéger ensemble au comité. Comme il y aura $\binom{2}{0}\binom{5}{3}$ groupes possibles de trois femmes ne contenant aucune des deux ennemies en question et $\binom{2}{1}\binom{5}{2}$ groupes contenant exactement l'une des deux, il y aura par conséquent $\binom{2}{0}\binom{5}{3} + \binom{2}{1}\binom{5}{2} = 30$ groupes de 3 femmes ne contenant pas les deux ennemies à la fois. Puisqu'il y a $\binom{5}{2}$ façons de choisir les 2 hommes, il sera possible au total de composer $30 \cdot \binom{5}{2} = 300$ comités différents. ■

Exemple 1.14 Considérons un ensemble de n antennes alignées dont m sont défectueuses et $n-m$ en état de marche. Supposons que les antennes défectueuses soient indiscernables entre elles et que celles qui marchent le soient également entre elles. Combien de configurations peut-on trouver pour lesquelles deux antennes défectueuses ne sont jamais voisines?

SOLUTION. Imaginons d'abord un alignement composé des seules $n-m$ antennes fonctionnelles. Si maintenant deux antennes défectueuses ne doivent jamais être voisines, les espaces entre les antennes fonctionnelles ne peuvent contenir chacun qu'au plus une antenne défectueuse. Considérons le schéma suivant:

o F o F o F o ... o F o F o

où F désigne un emplacement d'antenne fonctionnelle et o un emplacement pour au plus une antenne en panne. Parmi les $n-m+1$ positions du type o il faut en choisir m où mettre effectivement les antennes défectueuses. Il y a par conséquent $\binom{n-m+1}{m}$ dispositions pour lesquelles on trouve toujours une antenne fonctionnelle au moins entre deux antennes défectueuses. ■

1.4.3 Identité remarquable

L'identité suivante entre grandeurs combinatoires est très utile:

Théorème 1.6

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r} \quad 1 \leq r \leq n \quad (1.4)$$

DÉMONSTRATION. L'équation (1.4) peut être démontrée analytiquement mais aussi grâce à l'argument combinatoire suivant: considérons un groupe de n objets et fixons notre attention sur l'un d'entre eux en particulier, appelons-le objet 1. Il y a alors $\binom{n-1}{r-1}$ combinaisons de taille r qui contiennent l'objet 1 (puisque chaque combinaison de ce genre est formée en choisissant $r-1$ objets parmi les $n-1$ restants). Il y a également $\binom{n-1}{r}$ combinaisons de taille r ne contenant pas l'objet 1. Comme il y a au total $\binom{n}{r}$ combinaisons de taille r , (1.4) se trouve vérifiée. ■

1.4.4 Théorème du binôme

Les nombres $\binom{n}{r}$ sont souvent appelés *coefficients binomiaux* en raison de leur rôle dans le théorème du binôme.

Théorème 1.7

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} \quad (1.5)$$

Nous allons exposer deux démonstrations du théorème du binôme. La première est obtenue au moyen d'un raisonnement par induction, tandis que la seconde est basée sur des considérations d'analyse combinatoire.

DÉMONSTRATION PAR INDUCTION. Pour $n = 1$, (1.5) se réduit à

$$x + y = \binom{1}{0} x^0 y^1 + \binom{1}{1} x^1 y^0 = x + y$$

Admettons que (1.5) soit vérifiée pour $n-1$. Alors

$$\begin{aligned} (x + y)^n &= (x + y)(x + y)^{n-1} \\ &= (x + y) \sum_{k=0}^{n-1} \binom{n-1}{k} x^k y^{n-1-k} \\ &= \sum_{k=0}^{n-1} \binom{n-1}{k} x^{k+1} y^{n-1-k} + \sum_{k=0}^{n-1} \binom{n-1}{k} x^k y^{n-k} \end{aligned}$$

En posant $i = k+1$ dans la première somme et $i = k$ dans la seconde on obtient

$$\begin{aligned} (x + y)^n &= \sum_{i=1}^n \binom{n-1}{i-1} x^i y^{n-i} + \sum_{i=0}^{n-1} \binom{n-1}{i} x^i y^{n-i} \\ &= x^n + \sum_{i=1}^{n-1} \left[\binom{n-1}{i-1} + \binom{n-1}{i} \right] x^i y^{n-i} + y^n \\ &= x^n + \sum_{i=1}^{n-1} \binom{n}{i} x^i y^{n-i} + y^n \\ &= \sum_{i=0}^n \binom{n}{i} x^i y^{n-i} \end{aligned}$$

où l'avant-dernière transformation est obtenue grâce à (1.4). Ce théorème se trouve donc démontré par induction.

DÉMONSTRATION PAR UN ARGUMENT D'ANALYSE COMBINATOIRE. Considérons le produit suivant:

$$(x_1 + y_1)(x_2 + y_2) \cdots (x_n + y_n)$$

En le développant on obtient une somme de 2^n termes, chaque terme étant un produit de n facteurs. Chacun des 2^n termes de la somme contiendra à son tour soit le facteur x_i , soit y_i et ceci pour tout $i = 1, 2, \dots, n$. Par exemple:

$$(x_1 + y_1)(x_2 + y_2) = x_1x_2 + x_1y_2 + y_1x_2 + y_1y_2$$

Combien maintenant de ces 2^n termes de la somme auront-ils k facteurs en x et $(n-k)$ en y ? Comme chaque terme constitué par k des x_i et $(n-k)$ des y_i correspond au choix d'un groupe de k des n valeurs x_1, x_2, \dots, x_n , il y aura $\binom{n}{k}$ de ces termes. Par conséquent, en posant $x_i = x, y_i = y$ pour $i = 1, 2, \dots, n$ nous voyons que

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} \quad \blacksquare$$

1.4.5 Exemples d'application du théorème de binôme

Exemple 1.15 Développer $(x+y)^3$.

SOLUTION.

$$\begin{aligned} (x + y)^3 &= \binom{3}{0} x^0 y^3 + \binom{3}{1} x^1 y^2 + \binom{3}{2} x^2 y + \binom{3}{3} x^3 y^0 \\ &= y^3 + 3xy^2 + 3x^2y + x^3 \end{aligned} \quad \blacksquare$$

Exemple 1.16 Combien y a-t-il de sous-ensembles d'un ensemble à n éléments?

SOLUTION. Puisqu'il y a $\binom{n}{k}$ sous-ensembles de taille k , la réponse est

$$\sum_{k=0}^n \binom{n}{k} = (1 + 1)^n = 2^n$$

On pourrait aussi obtenir ce résultat en assignant à chaque élément de l'ensemble soit le nombre 0 soit le nombre 1. A chaque assignation complète correspond de manière biunivoque un sous-ensemble: celui constitué de tous les éléments auxquels a été attribuée la valeur 1. Comme il y a 2^n jeux d'assignations possibles, on obtient bien le résultat précédent. Notons que nous avons admis comme sous-ensemble celui ne contenant aucun élément (c'est-à-dire l'ensemble vide). Par conséquent, le nombre de sous-ensembles non-vides est $2^n - 1$. \blacksquare

1.5 COEFFICIENTS MULTINOMIAUX

1.5.1 Introduction

Nous traiterons dans cette section du problème suivant: un ensemble de n objets distincts doit être divisé en r groupes de tailles respectives n_1, n_2, \dots, n_r , avec $\sum_{i=1}^r n_i = n$. De combien de manières peut-on le faire?

Pour le savoir remarquons qu'il y a $\binom{n}{n_1}$ possibilités de choix pour le premier groupe; pour chacun de ces choix il y a $\binom{n-n_1}{n_2}$ possibilités de choix pour le deuxième groupe; pour chaque choix des deux premiers groupes il y a $\binom{n-n_1-n_2}{n_3}$ possibilités pour le troisième groupe et ainsi de suite. En utilisant alors la version généralisée du principe fondamental de dénombrement il y aura

$$\begin{aligned} & \binom{n}{n_1} \binom{n-n_1}{n_2} \cdots \binom{n-n_1-n_2-\cdots-n_{r-1}}{n_r} \\ &= \frac{n!}{(n-n_1)! n_1!} \frac{(n-n_1)!}{(n-n_1-n_2)! n_2!} \cdots \frac{(n-n_1-n_2-\cdots-n_{r-1})!}{0! n_r!} \\ &= \frac{n!}{n_1! n_2! \cdots n_r!} \end{aligned}$$

divisions possibles.

Soit r nombres n_1, n_2, \dots, n_r tels que $n_1 + n_2 + \dots + n_r = n$. Le terme $\binom{n}{n_1, n_2, \dots, n_r}$ est défini par l'équation

$$\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!} \quad (1.6)$$

Théorème 1.8

Le coefficient $\binom{n}{n_1, n_2, \dots, n_r}$ représente le nombre de répartitions possibles de n objets en r groupes distincts de tailles respectives n_1, n_2, \dots, n_r .

1.5.2 Exemples d'application du théorème 1.8

Exemple 1.17 Le poste de police d'une petite ville compte 10 agents. Si l'organisation de ce poste est d'avoir 5 agents en patrouille, 2 au poste travaillant activement et les 3 autres au poste également mais de réserve, à combien de répartitions de ces agents en trois groupes ainsi définis peut-on procéder?

SOLUTION. Il y a $\frac{10!}{5!2!3!} = 2520$ répartitions. ■

Exemple 1.18 Il faut répartir 10 filles en deux équipes A et B de 5 personnes chacune. L'équipe A sera placée dans une ligue et l'équipe B dans une autre. Combien y a-t-il de répartitions possibles?

SOLUTION. Il y en a $\frac{10!}{5!5!} = 252$. ■

Exemple 1.19 Pour disputer un match de basketball, 10 garçons se répartissent en deux équipes de 5. De combien de manières peuvent-ils procéder?

SOLUTION. Il faut remarquer que cet exemple est différent du précédent car ici l'ordre des deux équipes n'a pas d'importance: plus précisément il n'y a pas d'équipe A se distinguant d'une équipe B , mais seulement 2 groupes de 5 garçons. Par conséquent, la solution est

$$\frac{10!/5!5!}{2!} = 126 \quad \blacksquare$$

1.5.3 Théorème multinomial

Ce théorème généralise le théorème binomial. Sa démonstration fera l'objet d'un exercice.

Théorème 1.9

$$(x_1 + x_2 + \dots + x_r)^n = \sum_{\substack{(n_1, \dots, n_r): \\ n_1 + \dots + n_r = n}} \binom{n}{n_1, n_2, \dots, n_r} x_1^{n_1} x_2^{n_2} \dots x_r^{n_r} \quad (1.7)$$

La somme est ici faite sur tous les vecteurs à composantes entières non négatives (n_1, n_2, \dots, n_r) tels que $n_1 + n_2 + \dots + n_r = n$.

Les coefficients $\binom{n}{n_1, n_2, \dots, n_r}$ sont appelés *coefficients multinomiaux*.

Exemple 1.20

$$\begin{aligned} (x_1 + x_2 + x_3)^2 &= \binom{2}{2, 0, 0} x_1^2 x_2^0 x_3^0 + \binom{2}{0, 2, 0} x_1^0 x_2^2 x_3^0 \\ &\quad + \binom{2}{0, 0, 2} x_1^0 x_2^0 x_3^2 + \binom{2}{1, 1, 0} x_1^1 x_2^1 x_3^0 \\ &\quad + \binom{2}{1, 0, 1} x_1^1 x_2^0 x_3^1 + \binom{2}{0, 1, 1} x_1^0 x_2^1 x_3^1 \\ &= x_1^2 + x_2^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 2x_2x_3 \quad \blacksquare \end{aligned}$$

1.6 RÉPARTITION DE BOULES DANS DES URNES

Il y a r^n possibilités de répartir n boules discernables dans r urnes discernables également. Cela provient du fait que chaque boule peut être mise dans l'une quelconque des r urnes. Supposons maintenant que les n boules deviennent indiscernables. Combien peut-on alors obtenir de répartitions?

Comme les boules sont indiscernables, le résultat de l'expérience qui consiste à répartir les n boules dans nos r urnes est décrit par un vecteur (x_1, x_2, \dots, x_r) où x_i représente le nombre de boules contenues dans la i -ème urne. Le problème revient alors à trouver le nombre de vecteurs (x_1, x_2, \dots, x_r) à composantes entières non négatives tels que

$$x_1 + x_2 + \dots + x_r = n$$

Pour le calculer, commençons par considérer le nombre de solutions entières positives. Pour cela, imaginons qu'il y a n objets indiscernables alignés et que nous voulons les diviser en r groupes non vides. Ces objets peuvent être représentés comme suit:

$$0 \cdot 0 \cdot 0 \cdot 0 \cdot \dots \cdot 0 \cdot 0 \cdot 0$$

où les 0 représentent les n objets, les points de séparation symbolisant les $n-1$ espaces entre ces objets. Pour notre calcul, il suffit de désigner $r-1$ des $n-1$ espaces comme points de division. Si par exemple $n = 8$ et $r = 3$ on peut choisir les deux séparations comme suit:

$$ooo|ooo|oo$$

Le vecteur correspondant sera $x_1 = 3, x_2 = 3, x_3 = 2$. Comme il y a $\binom{n-1}{r-1}$ choix possibles nous venons de démontrer la proposition suivante:

Théorème 1.10

Il y a $\binom{n-1}{r-1}$ vecteurs distincts à composantes entières et positives satisfaisant à la relation

$$x_1 + x_2 + \dots + x_r = n, \quad x_i > 0, \quad i = 1, \dots, r$$

Pour obtenir le nombre des solutions non négatives (et non plus positives) il suffit de remarquer que le nombre de solutions non négatives de $x_1 + x_2 + \dots + x_r = n$ est le même que celui des solutions positives de $y_1 + y_2 + \dots + y_r = n + r$ (on le voit en posant $y_i = x_i + 1, i = 1, \dots, r$). Ceci permet de démontrer la proposition suivante, en utilisant la précédente:

Théorème 1.11

Il y a $\binom{n+r-1}{n}$ vecteurs distincts à composantes entières et non négatives satisfaisant à la relation

$$x_1 + x_2 + \dots + x_r = n \tag{1.8}$$

Exemple 1.21 Combien l'équation $x_1 + x_2 = 3$ a-t-elle de solutions entières et non négatives?

SOLUTION. Il y en a $\binom{3+2-1}{3} = 4$. Nommément, (0,3), (1,2), (2,1), (3,0). ■

Exemple 1.22 Une personne dispose de 20 000 dollars à investir sur quatre placements potentiels. Chaque mise doit se monter à un nombre entier de milliers de dollars. Entre combien de stratégies d'investissement cette personne a-t-elle le choix si elle décide de risquer la totalité des 20 000 dollars? Qu'en est-il si on admet qu'elle puisse n'investir qu'une partie seulement de la somme?

SOLUTION. Soit x_i le nombre de milliers de dollars placés dans l'affaire i , $i = 1, 2, 3, 4$. Si la totalité de l'argent doit être investie, on aura

$$x_1 + x_2 + x_3 + x_4 = 20 \quad x_i \geq 0$$

Par suite il y a, en vertu du théorème 1.11, $\binom{23}{3} = 1771$ stratégies d'investissement possibles. Si par contre on ne doit pas nécessairement investir tout l'argent, désignons par x_5 le montant gardé en réserve. Une stratégie d'investissement pourra alors être représentée par un vecteur à composantes non négatives $(x_1, x_2, x_3, x_4, x_5)$ tel que

$$x_1 + x_2 + x_3 + x_4 + x_5 = 20$$

En vertu du théorème 1.11 il y a donc ici $\binom{24}{4} = 10\,626$ stratégies possibles. ■

Exemple 1.23 Combien le développement de $(x_1 + x_2 + \dots + x_r)^n$ compte-t-il de termes?

SOLUTION.

$$(x_1 + x_2 + \dots + x_r)^n = \sum \binom{n}{n_1, \dots, n_r} x_1^{n_1} \dots x_r^{n_r}$$

où la somme est prise sur tous les vecteurs (n_1, \dots, n_r) à composantes non négatives entières tels que $n_1 + n_2 + \dots + n_r = n$. Selon le théorème 1.11 le nombre de ces vecteurs, et par suite de termes dans la somme, est $\binom{n+r-1}{r}$. ■

Exemple 1.24 Reprenons l'exemple 1.14 dans lequel nous avons un ensemble de n objets dont m étaient défectueux et indiscernables entre eux tandis que les $n - m$ autres étaient en bon état (et également indiscernables entre eux). Notre but est toujours de déterminer le nombre de séquences dans lesquelles deux objets défectueux ne sont jamais voisins. Pour cela, imaginons qu'on aligne d'abord les seuls objets défectueux et qu'il va ensuite falloir placer les objets en état de marche. Désignons par x_1 le nombre d'objets en bon état à gauche du premier objet endommagé, x_2 le nombre d'objets en bon état entre les deux premiers objets endommagés et ainsi de suite. On peut construire un schéma:

$$x_1 \ 0 \ x_2 \ 0 \ \dots \ x_m \ 0 \ x_{m+1}$$

Il y aura ainsi au moins un objet en bon état entre une paire d'objets défectueux si $x_i > 0$ pour $i = 2, \dots, m$. Le nombre de configurations acceptables sera donc égal au nombre de vecteurs (x_1, \dots, x_{m+1}) qui satisfont à $x_1 + \dots + x_{m+1} = n - m$ et à $x_1 \geq 0, x_{m+1} \geq 0, x_i > 0$ pour $i = 2, \dots, m$.

Mais en posant $y_1 = x_1 + 1$, $y_i = x_i$ pour $i = 2, \dots, m$ et $y_{m+1} = x_{m+1} + 1$ nous voyons que ce nombre est aussi le nombre de vecteurs y_1, \dots, y_{m+1} à composantes positives satisfaisant l'équation

$$y_1 + \dots + y_{m+1} = n - m + 2$$

En vertu du théorème 1.10 il existe $\binom{n-m+1}{m}$ de ces configurations, ce qui corrobore la solution de l'exemple 1.14.

Supposons maintenant que nous nous intéressions au nombre de configurations dans lesquelles toute paire d'objets défectueux est coupée par au moins deux objets en bon état. Selon le même raisonnement que celui mené ci-dessus, ce nombre est égal au nombre de vecteurs satisfaisant

$$x_1 + \dots + x_{m+1} = n - m \quad x_1 \geq 0, x_{m+1} \geq 0, x_i \geq 2, i = 2, \dots, m$$

En posant $y_1 = x_1 + 1$, $y_i = x_i - 1$ pour $i = 2, \dots, m$ et $y_{m+1} = x_{m+1} + 1$ on constate que ce nombre est encore égal à celui des solutions à valeurs positives de l'équation

$$y_1 + \dots + y_{m+1} = n - 2m + 3$$

Ce nombre de configurations est donc, selon le théorème 1.10, $\binom{n-2m+2}{m}$ ■

1.7 EXERCICES THÉORIQUES

1.7.1 Donner une preuve de la version généralisée du principe fondamental de dénombrement.

1.7.2 On réalise deux expériences consécutivement. La première peut présenter m résultats différents. Pour le i -ème de ces résultats la seconde peut présenter n_i résultats, $i = 1, \dots, m$. Quel est le nombre d'issues possibles lorsqu'on considère les deux expériences comme faisant un tout?

1.7.3 De combien de manières peut-on choisir r objets parmi n si l'ordre de tirage est significatif?

1.7.4 Donner un argument d'analyse combinatoire pour justifier l'égalité $\binom{n}{r} = \binom{n}{n-r}$.

1.7.5 Il y a $\binom{n}{r}$ permutations de n boules parmi lesquelles r sont noires et $n-r$ blanches. Expliquer ce résultat grâce à un argument d'analyse combinatoire.

1.7.6 Donner une démonstration analytique de l'équation (1.4).

1.7.7 Démontrer que

$$\binom{n+m}{r} = \binom{n}{0} \binom{m}{r} + \binom{n}{1} \binom{m}{r-1} + \dots + \binom{n}{r} \binom{m}{0}$$

lorsque $r \leq n$, $r \leq m$.

On pourra considérer un groupe comprenant n hommes et m femmes puis déterminer le nombre de sous-groupes de taille r .

1.7.8 Vérifier que pour $n \geq 4$

$$\binom{n+1}{4} = \frac{\binom{n}{2}}{3}$$

Présenter ensuite un argument d'analyse combinatoire en faveur de cette équation. Auparavant on considérera un groupe de $n+1$ objets desquels un est jugé spécial. Montrer que les deux membres de l'identité ci-dessus représentent le nombre de sous-groupes de taille 4. Pour l'expression de droite, supposer qu'on mette initialement de côté l'objet spécial et choisir alors deux des $\binom{n}{2}$ sous-groupes de taille 2 parmi les n objets restants. Si les deux sous-groupes choisis n'ont pas d'objet commun, utiliser alors leur union comme sous-groupe de taille 4; s'ils ont un objet commun, utiliser alors les 3 objets distincts de leur union plus l'objet spécial comme sous-groupe de taille 4. En utilisant cette approche pour obtenir tous les sous-groupes de taille 4, combien de fois apparaît chaque sous-groupe?

1.7.9 Fournir un argument d'analyse combinatoire pour expliquer que $\binom{n}{r}$ est égal à $\binom{n}{r, n-r}$.

1.7.10 Montrer que

$$\begin{aligned} \binom{n}{n_1, n_2, \dots, n_r} &= \binom{n-1}{n_1-1, n_2, \dots, n_r} \\ &+ \binom{n-1}{n_1, n_2-1, \dots, n_r} + \dots \\ &+ \binom{n-1}{n_1, n_2, \dots, n_r-1} \end{aligned}$$

1.7.11 Démontrer le théorème multinomial.

1.7.12 Montrer que pour $n > 0$

$$\sum_{i=0}^n (-1)^{n-i} \binom{n}{i} = 0$$

1.7.13

a) Démontrer l'identité suivante par induction:

$$\sum_{k=1}^n k \binom{n}{k} = n2^{n-1}$$

b) Fournir un argument d'analyse combinatoire pour l'identité précédente en considérant un ensemble de n personnes et en déterminant de deux manières le nombre de compositions possibles pour un comité assorti d'un président. On peut considérer les étapes intermédiaires suivantes :

- Combien de comités de taille k avec son président peut-on composer?
- Combien y a-t-il de compositions possibles pour un président et les autres membres du comité?

c) Vérifier l'identité suivante pour $n = 1, 2, 3, 4$ et 5 :

$$\sum_{k=1}^n \binom{n}{k} k^2 = 2^{n-2} n(n+1)$$

Pour une démonstration de cette égalité par l'analyse combinatoire, considérer un ensemble de n personnes et soutenir que les deux membres en présence représentent le nombre de comités assortis d'un président et d'un secrétaire, le cumul étant possible. On peut considérer les quelques étapes intermédiaires suivantes:

- combien y a-t-il de comités comprenant k personnes exactement?
- Combien y a-t-il de choix pour lesquels on observe un cumul des fonctions? (réponse : $n 2^{n-1}$)
- Combien de choix évitent-ils le cumul?

d) Prouver maintenant que

$$\sum_{k=1}^n \binom{n}{k} k^3 = 2^{n-3} n^2(n+3)$$

1.7.14 De combien de manières peut-on répartir m boules indiscernables dans r urnes de telle façon que la i -ème urne contienne au moins m_i boules? On admet que

$$n \geq \sum_{i=1}^r m_i$$

RÉPONSE. $\binom{n - \sum m_i + r - 1}{n - \sum m_i}$.

1.7.15 Montrer que

$$\binom{n+r-1}{n} = \sum_{i=0}^n \binom{n-i+r-2}{n-i}$$

On pourra appliquer le théorème 1.11.

1.7.16 Montrer qu'il y a $\binom{n-1}{k} \binom{n-1}{n-r-k}$ solutions à l'équation $x_1 + \dots + x_n = n$ pour lesquelles k exactement des termes de la somme sont nuls.

1.7.17 On considère une fonction $f(x_1, \dots, x_n)$ de n variables. Combien de dérivées partielles d'ordre r y a-t-il?

1.7.18 Utiliser l'exercice 1.7.7 pour démontrer que

$$\binom{2n}{n} = \sum_{k=0}^n \binom{n}{k}^2$$

1.7.19

a) En utilisant un raisonnement par induction et l'identité suivante

$$\binom{m}{k} = \binom{m-1}{k-1} + \binom{m-1}{k}$$

montrer que

$$\binom{n+r}{n} = \sum_{j=0}^n \binom{j+r-1}{j}$$

b) Donner une seconde démonstration en montrant que chacun des deux membres de l'égalité précédente est égal au nombre de solutions distinctes entières et non négatives de l'inégalité

$$x_1 + x_2 + \dots + x_r \leq n$$

Pour voir que $\binom{n+r}{n}$ est égal au nombre de solutions non négatives, voir que le nombre de solutions non négatives de $x_1 + \dots + x_r \leq n$ est le même que le nombre de solutions non négatives de $x_1 + \dots + x_r + x_{r+1} = n$.

1.7.20 On veut choisir un comité de j personnes dans un ensemble en comptant n . De ce comité on veut aussi choisir un sous-comité de taille i , $i \leq j$.

a) Ecrire une identité d'analyse combinatoire en calculant de deux manières le nombre de choix pour le comité et son sous-comité. On supposera dans un cas que le comité est tiré d'abord puis son sous-comité et dans le second cas que c'est l'inverse.

b) Utiliser le résultat de a) pour la démonstration de l'identité suivante:

$$\sum_{j=i}^n \binom{n}{j} \binom{j}{i} = \binom{n}{i} 2^{n-i}, \quad i \leq n$$

c) Utiliser le résultat de a) et l'exercice 1.7.12 pour prouver que

$$\sum_{j=i}^n \binom{n}{j} \binom{j}{i} (-1)^{n-j} = 0, \quad i \leq n$$

1.8 PROBLÈMES

1.8.1 Combien existe-t-il de plaques minéralogiques à 7 caractères

- si les deux premiers sont des lettres et les 5 autres des chiffres?
- Même question en supposant que les répétitions de lettres ou de chiffres sur la même plaque sont exclues.

1.8.2 John, Jim, Jay et Jack ont formé un orchestre à 4 instruments. Si chacun des garçons peut jouer des 4 instruments, combien d'arrangements peut-on concevoir? Que se passe-t-il si John et Jim peuvent jouer des 4 instruments mais si Jay et Jack ne savent jouer qu'au piano ou à la batterie?

1.8.3 Les indicatifs téléphoniques des Etats-Unis et du Canada sont composés de trois chiffres. Le premier chiffre est un entier compris entre 2 et 9; le deuxième est soit 0 soit 1; le troisième est un entier compris entre 1 et 9. Combien y a-t-il d'indicatifs possibles? Combien y a-t-il d'indicatifs commençant par 4?

1.8.4 Une célèbre chanson enfantine commence de la façon suivante :

En allant à St Ives
J'ai rencontré un homme avec 7 femmes.
Chaque femme avait 7 sacs.
Chaque sac contenait 7 chats.
Chaque chat avait 7 chatons.

Combien de chatons le voyageur a-t-il rencontrés?

1.8.5 On doit asseoir sur un rang 4 Américains, 3 Français et 3 Anglais. Les gens de même nationalité doivent rester ensemble. Combien de dispositions peut-on imaginer?

1.8.6 De combien de manières peut-on asseoir en rang 3 garçons et 3 filles?

- a) Même question si les garçons doivent rester ensemble et les filles aussi.
- b) Même question si seuls les garçons doivent rester ensemble.
- c) Même question si deux personnes du même sexe ne doivent jamais voisiner.

1.8.7 Combien d'arrangements différents peut-on faire avec les lettres des mots suivants:

- a) PINTE
- b) PROPOSE
- c) MISSISSIPPI
- d) ARRANGE?

1.8.8 Un enfant possède 12 cahiers: 6 noirs, 4 rouges, 1 blanc et 1 bleu. S'il tient à placer les noirs les uns derrière les autres, de combien de manières peut-il les ranger?

1.8.9 De combien de manières peut-on asseoir 8 personnes en rang si:

- a) aucune restriction n'est mise;
- b) les personnes *A* et *B* veulent être ensemble;
- c) les hommes ne doivent avoir que des voisines et inversement, en supposant qu'il y a 4 hommes et 4 femmes;
- d) les hommes, qui sont au nombre de 5, doivent rester ensemble;
- e) les personnes forment 4 couples de gens mariés et si chaque couple doit rester réuni?

1.8.10 De combien de manières peut-on placer 3 romans, 2 livres de mathématiques et 1 de chimie sur une étagère si:

- a) aucune restriction n'est mise;
- b) les livres de mathématiques doivent être rangés ensemble et les romans aussi;
- c) seuls les romans doivent être rangés ensemble?

1.8.11 On veut choisir dans un club comptant 10 membres un président, un secrétaire et un trésorier; le cumul est exclu. De combien de manières peut-on attribuer ces charges si:

- a) aucune restriction n'est imposée;
- b) *A* et *B* refusent d'officier ensemble;
- c) *C* et *D* officieront ensemble ou pas du tout;
- d) *E* doit avoir une charge;
- e) *F* n'accepte que la charge de président?

1.8.12 Cinq prix doivent être décernés à des étudiants méritants choisis dans une classe de 30 personnes (par exemple «meilleure performance académique», «meilleur leadership», etc). Combien de résultats peut-on avoir si:

- a) le cumul des prix est admis;
- b) le cumul n'est pas possible?

1.8.13 On considère un groupe de 20 personnes. Si chaque personne serre la main de toutes les autres, combien y a-t-il de poignées de main?

1.8.14 Combien de mains de poker existe-t-il? Le jeu comprend 52 cartes, une main en contient 5.

1.8.15 On veut former un comité de 7 personnes, dont 2 républicains, 2 démocrates et 3 indépendants. On a le choix parmi 5 républicains, 6 démocrates et 4 indépendants. De combien de manières peut-on procéder?

1.8.16 Un étudiant doit répondre à 7 des 10 questions d'un examen;

- a) de combien de manières peut-il les choisir?
- b) Même question s'il est obligé de choisir au moins 3 des 5 premières questions.

1.8.17 Dans un groupe de 8 femmes et 6 hommes, on doit former un comité de 3 hommes et 3 femmes. Combien de comités différents peut-on former si :

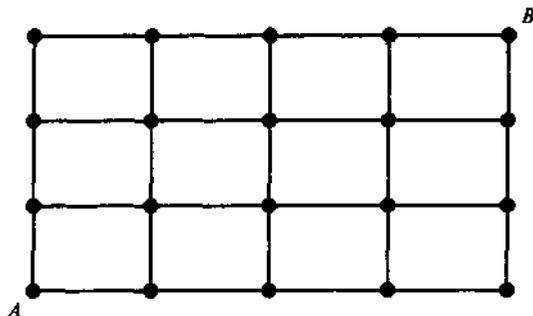
- a) 2 des hommes refusent d'être ensemble dans le comité?
- b) 2 des femmes refusent d'être ensemble dans le comité?
- c) 1 homme et 1 femme refusent d'être ensemble dans le comité?

1.8.18 Une femme a 8 amies et décide d'en inviter 5 à prendre le thé.

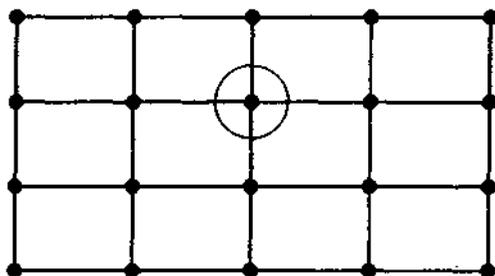
- a) De combien de manières peut-elle s'y prendre si deux d'entre elles sont en mauvais termes et ne viendront en aucun cas ensemble?
- b) Et si au contraire deux d'entre elles ne viendront que si l'autre est aussi invitée?

1.8.19 On considère le treillis de points ci-dessous. On suppose qu'en partant du point A , on peut à chaque pas soit monter d'un cran, soit aller à droite. On continue à avancer ainsi jusqu'à ce que le point B soit atteint. Combien de chemins différents peut-on prendre pour aller de A à B ?

Noter que pour atteindre B à partir de A , il faut faire 4 pas à droite et 3 vers le haut.



1.8.20 Dans le problème 1.8.19, combien de chemins passant par le point encerclé ci-dessous peut-on prendre pour aller de A à B ?



1.8.21 Un laboratoire de recherches en psychologie du rêve dispose de 3 chambres à deux lits. Trois paires de vrais jumeaux sont étudiées. On veut placer chaque paire dans une chambre et assigner à chacun un lit bien déterminé. De combien de manières peut-on organiser l'expérience?

1.8.22 Développer $(3x^2 + y)^5$.

1.8.23 Pour une partie de bridge chacun des 4 joueurs reçoit 13 cartes. Le jeu en compte 52. Combien y a-t-il de donnes possibles?

1.8.24 Développer $(x_1 + 2x_2 + 3x_3)^4$.

1.8.25 Si 12 personnes doivent être réparties en 3 comités comptant respectivement 3, 4 et 5 individus, de combien de manières peut-on s'y prendre?

1.8.26 Un homme veut offrir un total de 7 cadeaux à ses 3 enfants. L'aîné en recevra 3 et les autres 2. De combien de manières peut-il procéder?

1.8.27 Si 8 tableaux noirs doivent être affectés à 4 écoles, de combien de manières peut-on les répartir? Qu'en est-il si chaque école doit recevoir au moins un tableau?

1.8.28 Huit nouveaux professeurs vont être envoyés dans 4 écoles.

- Combien y a-t-il d'affectations possibles?
- Qu'en est-il si l'on impose que chaque école recevra deux professeurs?

1.8.29 Un ascenseur quitte le rez-de-chaussée avec 8 personnes (groom non compris). Lorsqu'il parvient au 6^e étage, il est vide.

- De combien de manières le groom a-t-il pu percevoir le départ des 8 personnes si pour lui elles se ressemblent toutes?
- Qu'en est-il s'il peut faire la différence entre un homme et une femme, l'ascenseur contenant 5 hommes et 3 femmes au départ?

1.8.30 Lors d'une vente aux enchères, une collection de 4 Dali, 5 Van Gogh et 6 Picasso fait face à 5 collectionneurs. Toutes les oeuvres partent. La journaliste en charge de couvrir l'événement n'a à noter que le nombre des Dali, Van Gogh et Picasso acquis par chaque collectionneur. Combien de résultats sont-ils possibles dans ces conditions?

1.8.31 Dix haltérophiles sont engagés dans une compétition par équipe. L'équipe américaine compte 3 champions, l'équipe soviétique 4, l'équipe de Chine populaire 2 et le dernier homme est canadien. Le score publié n'indique que la nationalité des haltérophiles, sans leur nom.

- Dans ce cas, combien y a-t-il de listes de scores possibles?
- Combien y en a-t-il si les Etats-Unis ont un concurrent placé dans les trois meilleurs et deux dans les trois derniers?

1.8.32 Dix délégués de 10 pays – dont l'URSS, la France, la Grande-Bretagne et les Etats-Unis – s'assoient sur un rang. De combien de manières est-ce possible si le français et l'anglais tiennent à être voisins tandis que l'américain et le soviétique ne veulent pas l'être?

1.8.33 Une personne a 20 000 dollars à placer sur 4 affaires potentielles. Chaque investissement doit être un nombre entier de milliers de dollars et il existe un engagement minimum pour chaque affaire que l'on retiendra. Ces minima sont respectivement 2, 2, 3 et 4 milliers de dollars. Combien de stratégies d'investissement y a-t-il si:

- un investissement doit être fait sur chaque affaire;
- au moins 3 des 4 affaires doivent être couvertes?

1.8.34 Montrer que

$$i \binom{n}{i} = n \binom{n-1}{i-1}$$

CHAPITRE 2

Axiomes des probabilités

2.1 INTRODUCTION

Dans ce chapitre nous commencerons par une introduction au concept de probabilité d'un événement puis nous montrerons comment ces probabilités peuvent être calculées dans certaines situations. Nous aurons préalablement besoin, cependant, des concepts d'ensemble fondamental et d'événement d'une expérience.

2.2 ENSEMBLE FONDAMENTAL ET ÉVÉNEMENT

2.2.1 Définitions

Considérons une expérience dont l'issue n'est pas prévisible. Bien que l'issue de l'expérience ne soit pas connue d'avance, admettons cependant que l'ensemble des issues possibles est connu, lui. Cet ensemble des issues possibles à l'expérience est désigné comme l'*ensemble fondamental* de l'expérience et est noté S . Quelques exemples suivent.

- Si le résultat de l'expérience équivaut à la détermination du sexe d'un nouveau-né, alors

$$S = \{g, f\}$$

où le résultat g signifie que l'enfant est un garçon tandis que f désigne une fille.

- Si l'issue de l'expérience est l'ordre d'arrivée à une course entre 7 chevaux ayant les positions de départ 1, 2, 3, ..., 7, alors

$$S = \{\text{toutes les permutations de } (1, 2, \dots, 7)\}$$

soit $7!$ au total.

- Si l'expérience consiste à jeter deux pièces, alors l'ensemble fondamental est constitué des 4 points suivants:

$$S = \{(P,P), (P,F), (F,P), (F,F)\}.$$

On note le résultat (P,P) si les deux pièces montrent pile,
 (P,F) si la première pièce montre pile et la seconde face,
 (F,P) si la première pièce montre face et l'autre pile,
 (F,F) si les deux pièces montrent face.

- Si l'expérience consiste à jeter deux dés, alors l'ensemble fondamental comprend les 36 points suivants:

$$S = \{(i, j) \mid i, j = 1, 2, 3, 4, 5, 6\}$$

où l'événement (i, j) est réputé survenir si le dé le plus à gauche montre i et l'autre j .

- Si l'expérience consiste à mesurer en heures la durée de vie d'un transistor, alors l'ensemble fondamental est égal à l'ensemble des nombres réels non négatifs, c'est-à-dire

$$S = \{x: 0 \leq x < \infty\}$$

Tout sous-ensemble E de l'ensemble fondamental est appelé *événement*. Un événement est donc un ensemble correspondant à divers résultats possibles de l'expérience. Si un résultat de l'expérience est compris dans E , on dit que E est réalisé. Voici quelques exemples d'événements.

Dans le premier exemple ci-dessus, si $E = \{g\}$, alors E est l'événement que l'enfant est un garçon. De même, si $F = \{f\}$, alors F est l'événement que l'enfant est une fille.

Dans le deuxième exemple, si

$$E = \{\text{tous les résultats dans } S \text{ commençant par } 3\}$$

alors E est l'événement que le cheval N° 3 gagne la course.

Dans le troisième exemple, si $E = \{(P,P)(P,F)\}$, alors E est l'événement «la première pièce montre pile».

Dans le quatrième exemple, si $E = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$, E est l'événement «la somme des dés est 7».

Dans le cinquième exemple, si $E = \{x: 0 \leq x \leq 5\}$, E est l'événement «le transistor dure moins de 5 heures».

2.2.2 Opérations sur les événements

Première opération: union

Pour toute paire d'événements E et F d'un ensemble fondamental S , nous imposons au nouvel événement $E \cup F$ de contenir chaque point se trouvant dans E , dans F ou dans les deux à la fois. En clair, l'événement $E \cup F$ sera réalisé si soit E soit F l'est. Prenons le cas du premier exemple: si l'événement E est $\{g\}$ et si F est $\{f\}$, alors

$$E \cup F = \{g, f\}$$

ce qui revient d'ailleurs à dire que $E \cup F$ est l'ensemble fondamental S tout entier. Dans le cas du troisième exemple, on pourrait poser $E = \{(P,P), (P,F)\}$ et $F = \{(P,F)\}$. On aurait alors $E \cup F = \{(P,P), (P,F), (F,P)\}$. Ainsi $E \cup F$ sera réalisé si l'une des

pièces au moins montre pile. L'événement $E \cup F$ est appelé *l'union de l'événement E et de l'événement F*.

Deuxième opération: intersection

De même pour toute paire d'événements E et F on peut aussi définir le nouvel événement EF , appelé *intersection* de E et F , comme l'ensemble des réalisations qui sont à la fois dans E et dans F . Cela veut dire que l'événement EF ne sera réalisé que si E et F le sont à la fois. On peut illustrer ceci grâce au troisième exemple: si $E = \{(P,P), (P,F), (F,P)\}$ est l'événement où au moins une pièce donne pile et si $F = \{(P,F), (F,P), (F,F)\}$ est celui où au moins une pièce donne face, alors

$$EF = \{(P,F), (F,P)\}$$

est l'événement «une pièce montre pile et l'autre face».

Événement vide

Dans le quatrième exemple, si $E = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$ est l'événement «la somme des dés est 7» et $F = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$ est «la somme des dés est 6», alors l'événement EF ne contient aucune réalisation et par conséquent ne peut survenir. Puisqu'il faut donner un nom à un tel événement, on l'appellera *l'événement vide* et on le notera \emptyset . (\emptyset désigne donc l'événement ne contenant aucun point). Si $EF = \emptyset$, alors E et F sont dits *mutuellement exclusifs*.

Extension des définitions

On définit l'union et l'intersection de plus de deux événements de la même manière: si E_1, E_2, \dots sont des événements, leur union, notée $\bigcup_{n=1}^{\infty} E_n$, est par définition l'événement qui contient chaque point qui se trouve dans E_n pour au moins une valeur de $n = 1, 2, \dots$. De même l'intersection des événements E_n , notée $\bigcap_{n=1}^{\infty} E_n$, est par définition l'événement comprenant tous les points qui sont dans tous les événements E_n à la fois, $n = 1, 2, \dots$.

Troisième opération: complémentation

Finalement, pour chaque événement E le nouvel événement E^c devra par définition contenir tous les points de l'ensemble fondamental S qui ne sont pas dans E . Dans le quatrième exemple, si l'événement $E = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$, alors E^c sera réalisé lorsque la somme des dés n'est pas égale à 7. On notera par ailleurs que $S^c = \emptyset$ puisqu'il faut bien que l'expérience débouche sur une réalisation.

2.2.3 Représentation graphique d'événements

Pour toute paire d'événements E et F , si tous les points de E sont aussi dans F alors on dit que E est contenu dans F et on écrit $E \subset F$ (ou, ce qui est équivalent, $F \supset E$). Ainsi, si $E \subset F$ la réalisation de E entraîne automatiquement celle de F . Si $E \subset F$ et $F \subset E$, nous dirons que E et F sont égaux et écrirons $E = F$.

Une représentation graphique très utile pour l'illustration des relations logiques entre les événements est le diagramme de Venn. L'ensemble S est représenté par tous

les points d'un grand rectangle et les événements $E, F, G \dots$ sont représentés par tous les points situés à l'intérieur de cercles inclus dans le rectangle. Des événements d'intérêt particulier peuvent ensuite être mis en évidence en ombrant les aires appropriées du diagramme. Par exemple, dans les trois diagrammes de Venn montrés sur la figure 2.1, les zones ombrées représentent respectivement les événements $E \cup F$, EF et E^c . Le diagramme de Venn de la figure 2.2 indique que $E \subset F$.

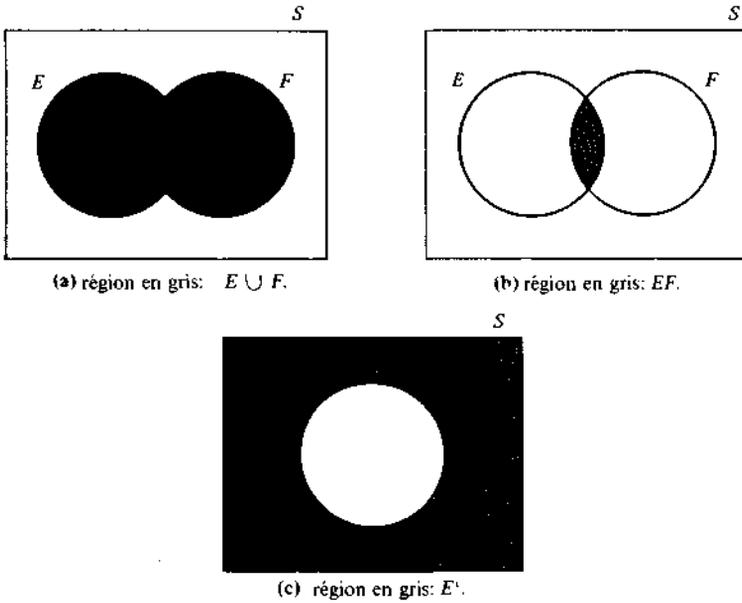


Fig. 2.1

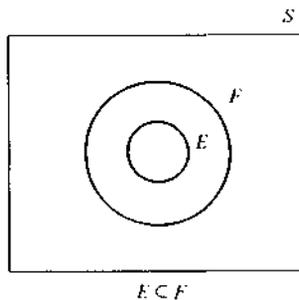


Fig. 2.2

2.2.4 Propriétés des opérations sur les événements

Les opérations d'union, d'intersection et de complément d'événements obéissent à certaines règles rappelant celles de l'algèbre. En voici quelques-unes:

Commutativité $E \cup F = F \cup E$ $EF = FE$
 Associativité $(E \cup F) \cup G = E \cup (F \cup G)$ $(EF)G = E(FG)$
 Distributivité $(E \cup F)G = EG \cup FG$ $EF \cup G = (E \cup G)(F \cup G)$

Ces relations sont démontrables en établissant que chaque résultat d'expérience contenu dans l'événement situé à gauche du signe d'égalité est aussi contenu dans l'événement de droite et inversement. Un moyen de le faire est d'utiliser les diagrammes de Venn. La distributivité par exemple peut être établie grâce à la séquence des diagrammes de la figure 2.3.

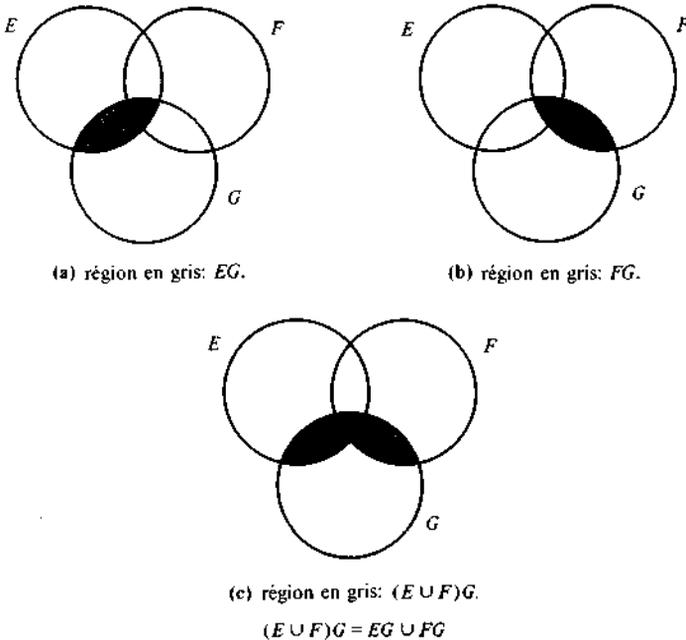


Fig. 2.3

Les relations suivantes entre les trois opérations de base consistant à former des unions, des intersections ou des complémentations, sont connues sous le nom de *lois de DeMorgan* et sont très utilisées.

$$\left(\bigcup_{i=1}^n E_i \right)^c = \bigcap_{i=1}^n E_i^c$$

$$\left(\bigcap_{i=1}^n E_i \right)^c = \bigcup_{i=1}^n E_i^c$$

Pour démontrer la première de ces lois de DeMorgan, supposons d'abord que x soit un point de $\left(\bigcup_{i=1}^n E_i \right)^c$. Alors x n'est pas dans $\bigcup_{i=1}^n E_i$, ce qui signifie que x n'est contenu dans aucun des événements E_i , $i = 1, 2, \dots, n$. Ceci implique que x est contenu à son tour dans E_i^c pour chaque $i = 1, \dots, n$ et donc contenu dans $\bigcap_{i=1}^n E_i^c$. Pour la réciproque,

supposons que x soit un point de $\bigcap_{i=1}^n E_i^c$. Alors x appartient à chaque E_i^c , $i = 1, 2, \dots, n$.

Cela signifie que x n'est pas contenu dans $\bigcup_{i=1}^n E_i$, ce qui entraîne enfin que x est contenu dans $(\bigcup_{i=1}^n E_i)^c$. Ceci prouve la première des lois de DeMorgan. Pour prouver la deuxième loi de DeMorgan, nous utilisons la première pour obtenir

$$\left(\bigcup_{i=1}^n E_i^c \right)^c = \bigcap_{i=1}^n (E_i^c)^c$$

ce qui, du fait que $(E^c)^c = E$, est équivalent à

$$\left(\bigcup_{i=1}^n E_i^c \right)^c = \bigcap_{i=1}^n E_i$$

En prenant le complément des deux membres de l'équation ci-dessus, on obtient précisément le résultat voulu:

$$\bigcup_{i=1}^n E_i^c = \left(\bigcap_{i=1}^n E_i \right)^c$$

2.3 AXIOMES DES PROBABILITÉS

2.3.1 Diverses approches

Un moyen de définir la probabilité d'un événement est de le faire en termes de fréquence relative. Une telle définition est habituellement formulée ainsi: on suppose qu'une expérience d'ensemble fondamental S est exécutée plusieurs fois sous les mêmes conditions. Pour chaque événement E de S on définit $n(E)$ comme le nombre de fois où l'événement E survient lors des n premières répétitions de l'expérience. Alors $P(E)$, la probabilité de l'événement E , est définie par

$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}$$

Cela veut dire que $P(E)$ est définie comme la limite du pourcentage du nombre de fois où E survient par rapport au nombre total des répétitions. C'est donc la fréquence limite de E .

Bien que la définition précédente soit intuitivement commode, et qu'elle doive toujours rester à l'esprit du lecteur, elle possède un sérieux inconvénient. Nous ne savons en fait pas si $n(E)$ va converger vers une limite constante qui sera la même pour chaque séquence de répétitions de l'expérience. Dans le cas du jet d'une pièce par exemple, peut-on être sûr que la proportion de piles sur les n premiers jets va tendre vers une limite donnée lorsque n grandit à l'infini? En plus, même si elle converge vers une certaine valeur, peut-on être sûr que nous obtiendrons de nouveau la même proportion limite de piles si l'expérience est entièrement répétée une deuxième fois?

Les partisans de la définition d'une probabilité en termes de fréquence relative répondent d'habitude à cette objection en faisant remarquer que la convergence de $n(E)$ est une hypothèse, ou un *axiome*, du système. Cependant, admettre que $n(E)/n$ va nécessairement converger vers une certaine valeur fixe semble être une hypothèse très complexe. Car, bien que nous puissions effectivement espérer qu'une telle fréquence limite constante existe, il ne semble pas évident du tout à priori que ce soit nécessairement le cas.

2.3.2 Trois axiomes classiques

En fait, il semble plus raisonnable d'admettre pour les probabilités un ensemble d'axiomes plus simples et intuitivement acceptables, pour ensuite essayer de démontrer qu'une telle fréquence limite existe dans un certain sens. Cette dernière approche est celle de l'axiomatique moderne de la théorie des probabilités et nous l'adopterons. En particulier, nous admettrons que pour chaque événement E de l'ensemble fondamental S il existe une valeur $P(E)$ appelée probabilité de E . Nous admettrons alors que ces probabilités satisfont à un certain groupe d'axiomes qui sont, espérons que le lecteur en conviendra, en accord avec notre notion intuitive des probabilités.

Considérons une expérience dont l'ensemble fondamental est S . Pour chaque événement E de l'espace S nous admettons qu'un nombre $P(E)$ existe et satisfait aux trois axiomes suivants:

Axiome 2.1

$$0 \leq P(E) \leq 1$$

Axiome 2.2

$$P(S) = 1$$

Axiome 2.3

Pour chaque séquence d'événements mutuellement exclusifs E_1, E_2, \dots (c'est-à-dire d'événements pour lesquels $E_i E_j = \emptyset$ si $i \neq j$).

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

$P(E)$ est appelé la probabilité de l'événement E .

L'axiome 2.1 énonce ainsi «la probabilité que le résultat de l'expérience soit un point de E est un certain nombre compris entre 0 et 1». L'axiome 2.2 énonce que le résultat sera un point de S avec une probabilité de 1. L'axiome 2.3 énonce que pour chaque séquence d'événements mutuellement exclusifs la probabilité qu'au moins l'un de ces événements survienne est simplement la somme de leurs probabilités respectives.

2.3.3 Quelques conséquences immédiates

Considérons une séquence d'événements E_1, E_2, \dots où $E_1 = S, E_i = \emptyset$ pour $i > 1$; comme ces événements sont mutuellement exclusifs et comme $S = \bigcup_{i=1}^{\infty} E_i$, nous aurons grâce à l'axiome 2.3:

$$P(S) = \sum_{i=1}^{\infty} P(E_i) = P(S) + \sum_{i=2}^{\infty} P(\emptyset)$$

ce qui implique

$$P(\emptyset) = 0$$

Cela veut dire que l'événement vide ou toujours faux a pour probabilité 0.

Autre conséquence remarquable, il en découle également que pour toute suite finie d'événements mutuellement exclusifs E_1, E_2, \dots, E_n

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i) \quad (2.1)$$

Ceci résulte de l'axiome 2.3 en posant $E_i = \emptyset$ pour toutes les valeurs de i supérieures à n . L'axiome 2.3 équivaut à l'équation (2.1) quand l'ensemble fondamental est fini (expliquer pourquoi). Cependant, lorsque l'ensemble fondamental contient un nombre infini de points, la formulation plus générale que donne l'axiome 2.3 devient nécessaire.

Exemple 2.1 Notre expérience consiste à jeter une pièce. En admettant que pile a autant de chances d'apparaître que face, les axiomes nous donnent

$$P(\{\text{pile}\}) = P(\{\text{face}\}) = \frac{1}{2}$$

Si par contre la pièce est biaisée et si nous estimons que pile a deux fois plus de chances d'apparaître que face, on aura

$$P(\{\text{pile}\}) = \frac{2}{3} \quad P(\{\text{face}\}) = \frac{1}{3} \quad \blacksquare$$

Exemple 2.2 En jetant un dé et en supposant que les six faces ont les mêmes chances d'apparaître, on aura $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$. De l'axiome 2.3 il résulte que la probabilité de tirer un nombre pair est

$$P(\{2, 4, 6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2} \quad \blacksquare$$

L'admission de l'existence d'une fonction P d'ensembles (ndt: fonction dont les arguments sont des ensembles et sous-ensembles), définie sur les événements d'un ensemble fondamental S et satisfaisant les axiomes 2.1, 2.2 et 2.3 constitue l'approche mathématique moderne de la théorie des probabilités. On peut espérer que le lecteur accordera aux axiomes un caractère naturel et en accord avec le concept intuitif de probabilité lié à la chance et au hasard. De plus, en utilisant ces axiomes nous serons capables de prouver que si une expérience est répétée plusieurs fois, alors avec une

probabilité de 1 la proportion du nombre de fois où un événement spécifique E survient sera égale à $P(E)$. Ce résultat, connu sous le nom de loi forte des grands nombres, sera présenté au chapitre 8. En addition nous présenterons, dans la section 2.7, une autre interprétation possible des probabilités: celle consistant à considérer la probabilité comme une mesure du crédit apporté par une personne à une assertion.

Nous avons supposé que $P(E)$ est définie pour tous les événements E de l'ensemble fondamental S . En fait, lorsque S est infini non dénombrable, on ne définit $P(E)$ que sur des événements appelés mesurables. Cependant, nous ne nous soucions pas de cette restriction, tous les événements d'intérêt pratique étant mesurables.

2.4 QUELQUES THÉORÈMES ÉLÉMENTAIRES

Dans cette section nous allons démontrer quelques théorèmes simples concernant les probabilités. Nous remarquons d'abord que E et E^c sont toujours mutuellement exclusifs et puisque $E \cup E^c = S$, nous avons grâce aux axiomes 2.2 et 2.3:

$$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c)$$

Ceci équivaut encore à l'énoncé du théorème 2.4 suivant:

Théorème 2.4

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

On peut commenter ce théorème comme suit: la probabilité qu'un événement n'arrive pas est 1 moins la probabilité qu'il survienne. Par exemple, si la probabilité d'obtenir pile lors du lancer d'une pièce est $\frac{3}{8}$, la probabilité d'obtenir face doit être $\frac{5}{8}$. ■

Notre second théorème affirme que si l'événement E est contenu dans l'événement F , alors la probabilité de E n'est pas plus grande que celle de F .

Théorème 2.5

Si $E \subset F$, alors $P(E) \leq P(F)$.

PREUVE. Du fait que $E \subset F$, on peut exprimer F ainsi:

$$F = E \cup E^c F$$

E et $E^c F$ étant mutuellement exclusifs, on tire de l'axiome 2.3 que

$$P(F) = P(E) + P(E^c F)$$

ce qui prouve le résultat puisque $P(E^c F) \geq 0$. ■

Le théorème 2.5 indique par exemple que la probabilité d'obtenir le nombre 1 avec un dé est inférieure ou égale à celle de tirer une valeur impaire avec ce dé.

Le théorème suivant donne la relation entre la probabilité de la réunion de deux événements d'une part, les probabilités individuelles et la probabilité de l'intersection d'autre part.

Théorème 2.6

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

PREUVE. Pour obtenir une formule donnant $P(E \cup F)$, remarquons d'abord que $E \cup F$ peut être écrit comme l'union de deux éléments disjoints E et $E^c F$. Nous tirons alors de l'axiome 2.3 que

$$\begin{aligned} P(E \cup F) &= P(E \cup E^c F) \\ &= P(E) + P(E^c F) \end{aligned}$$

De plus, comme $F = EF \cup E^c F$, nous tirons de nouveau de cet axiome

$$P(F) = P(EF) + P(E^c F)$$

ou encore

$$P(E^c F) = P(F) - P(EF)$$

ce qui achève la démonstration. ■

On aurait aussi pu démontrer le théorème 2.6 en faisant usage du diagramme de Venn comme le démontre la figure 2.4.

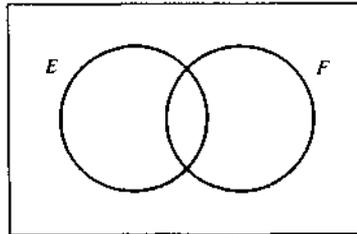


Fig. 2.4

Divisons le diagramme en trois parties disjointes comme celles représentées sur la figure 2.5.

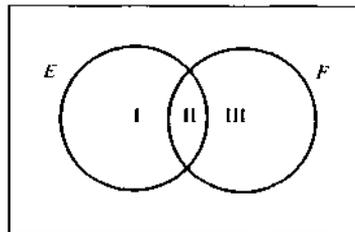


Fig. 2.5

La section I représente tous les points de E qui ne sont pas dans F (c'est-à-dire $E F^c$); la section II représente tous ceux qui sont dans E et dans F (c'est-à-dire EF); la section III représente tous ceux de F qui ne sont pas dans E (c'est-à-dire $E^c F$).

Sur la figure 2.5 nous voyons que

$$\begin{aligned} E \cup F &= I \cup II \cup III \\ E &= I \cup II \\ F &= II \cup III \end{aligned}$$

Comme I, II et III sont disjoints, il résulte de l'axiome 2.3:

$$\begin{aligned} P(E \cup F) &= P(I) + P(II) + P(III) \\ P(E) &= P(I) + P(II) \\ P(F) &= P(II) + P(III) \end{aligned}$$

ce qui montre que

$$P(E \cup F) = P(E) + P(F) - P(II)$$

et le théorème 2.6 est ainsi démontré, puisque $II = EF$. ■

Exemple 2.3 Supposons que l'on jette deux pièces et que chacun des quatre points de l'ensemble fondamental $S = \{(P,P), (P,F), (F,P), (F,F)\}$ soit de même probabilité $\frac{1}{4}$. Soient $E = \{(P,P), (P,F)\}$ et $F = \{(P,P), (F,P)\}$, c'est-à-dire que E est l'événement «la première pièce tombe sur pile» et F l'événement «la deuxième pièce tombe sur pile». Le théorème 2.6 nous donne la probabilité $P(E \cup F)$ que soit la première soit la deuxième pièce tombe sur pile

$$\begin{aligned} P(E \cup F) &= P(E) + P(F) - P(EF) \\ &= \frac{1}{2} + \frac{1}{2} - P(\{(P,P)\}) \\ &= 1 - \frac{1}{4} \\ &= \frac{3}{4} \end{aligned}$$

Cette probabilité aurait bien évidemment pu être calculée directement puisque

$$P(E \cup F) = P(\{(P,P), (P,F), (F,P)\}) = \frac{3}{4} \quad \blacksquare$$

Nous pouvons aussi calculer la probabilité que l'un quelconque des trois événements E , F ou G survienne:

$$P(E \cup F \cup G) = P[(E \cup F) \cup G]$$

qui vaut par le théorème 2.6:

$$P(E \cup F) + P(G) - P[(E \cup F)G]$$

A ce point, l'équivalence des événements $(E \cup F)G$ et $EG \cup FG$ résulte de la distributivité des opérations, ce qui permet d'écrire

$$\begin{aligned} P(E \cup F \cup G) &= P(E) + P(F) - P(EF) + P(G) - P(EG \cup FG) \\ &= P(E) + P(F) - P(EF) + P(G) - P(EG) - P(FG) + P(EGFG) \\ &= P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG) \end{aligned}$$

En fait, on peut démontrer par induction la généralisation suivante:

Théorème 2.7

$$\begin{aligned}
 & P(E_1 \cup E_2 \cup \dots \cup E_n) \\
 &= \sum_{i=1}^n P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \dots \\
 &\quad + (-1)^{r+1} \sum_{i_1 < i_2 < \dots < i_r} P(E_{i_1} E_{i_2} \dots E_{i_r}) \\
 &\quad + \dots + (-1)^{n+1} P(E_1 E_2 \dots E_n)
 \end{aligned}$$

La somme $\sum_{i_1 < i_2 < \dots < i_r} P(E_{i_1} E_{i_2} \dots E_{i_r})$ est prise sur les $\binom{n}{r}$ sous-ensembles possibles de taille r de l'ensemble $\{1, 2, \dots, n\}$.

En clair, le théorème 2.7 signifie que la probabilité de l'union de n événements est égale à la somme des probabilités de ces événements pris un à un moins la somme des probabilités de ces événements pris deux à deux plus la somme des probabilités de ces événements pris trois à trois, et ainsi de suite.

REMARQUE. Pour donner un argument non inductif pour le théorème 2.7, notons d'abord que si un point de l'ensemble fondamental S n'est membre d'aucun ensemble E_i , alors sa probabilité ne contribue en rien aux deux membres de l'égalité. Au contraire, supposons qu'un point appartienne à exactement m ensembles E_i , où $m > 0$. Alors, puisqu'il se trouve dans $\bigcup_i E_i$, sa probabilité est comptée une fois dans $P(\bigcup_i E_i)$; mais comme ce point est contenu dans $\binom{m}{k}$ sous-groupes du type $E_{i_1} E_{i_2} \dots E_{i_k}$, sa probabilité est comptée

$$\binom{m}{1} - \binom{m}{2} + \binom{m}{3} - \dots \pm \binom{m}{m}$$

fois dans le terme de droite du théorème 2.7.

Donc, pour $m > 0$, nous devons montrer que

$$1 = \binom{m}{1} - \binom{m}{2} + \binom{m}{3} - \dots \pm \binom{m}{m}$$

Comme $1 = \binom{m}{0}$, l'équation précédente est équivalente à

$$\sum_{i=0}^m \binom{m}{i} (-1)^i = 0$$

et la dernière équation provient du théorème du binôme :

$$0 = (-1 + 1)^m = \sum_{i=0}^m \binom{m}{i} (-1)^i (1)^{m-i}$$

2.5 ENSEMBLES FONDAMENTAUX À ÉVÉNEMENTS ÉLÉMENTAIRES ÉQUIPROBABLES

2.5.1 Méthode de calcul des probabilités

Pour de nombreuses expériences il est naturel d'admettre que chaque élément, ou *événement élémentaire*, de l'ensemble fondamental a la même probabilité d'apparaître. Plus précisément, considérons une expérience dont l'ensemble fondamental S est fini, disons $S = \{1, 2, \dots, N\}$. Il est alors souvent naturel de supposer que

$$P(\{1\}) = P(\{2\}) = \dots = P(\{N\})$$

ce qui implique du fait des axiomes 2.2 et 2.3 (dire pourquoi) que

$$P(\{i\}) = \frac{1}{N} \quad i = 1, 2, \dots, N$$

De ceci et de l'axiome 2.3 il résulte que pour tout événement E

$$P(E) = \frac{\text{nombre de points dans } E}{\text{nombre de points dans } S}$$

En clair, si nous admettons que toutes les issues d'une expérience ont la même probabilité de survenir, la probabilité d'un événement E quelconque est égale à la proportion dans l'ensemble de définition de points qui sont contenus dans E .

2.5.2 Exemples

Exemple 2.4 Si deux dés sont jetés, quelle est la probabilité que la somme des faces soit 7?

SOLUTION. Nous résoudrons ce problème en faisant l'hypothèse que les 36 issues possibles sont équiprobables. Puisqu'il y a 6 issues, à savoir (1,6), (2,5), (3,4), (4,3), (5,2) et (6,1), qui donnent une somme de 7 pour les deux dés, la probabilité est $\frac{6}{36} = \frac{1}{6}$. ■

Exemple 2.5 Si deux boules sont tirées au hasard d'un bol en contenant 6 blanches et 5 noires, quelle est la probabilité qu'une des boules tirées soit blanche et l'autre noire?

SOLUTION. Si nous considérons l'ordre dans lequel les boules sont choisies comme significatif, l'ensemble fondamental comprend $11 \cdot 10 = 110$ points. De plus, il y a $6 \cdot 5 = 30$ manières de tirer pour lesquelles la première boule est blanche et la seconde noire. On compte de même $5 \cdot 6 = 30$ manières de tirer pour lesquelles la première boule est noire et la seconde blanche. De ce fait, si tirer au hasard signifie que chacun des 110 points de l'ensemble fondamental a la même probabilité de survenir, nous voyons que la probabilité cherchée est

$$\frac{30 + 30}{110} = \frac{6}{11}$$

Ce problème aurait aussi pu être résolu en considérant que chaque résultat de l'expérience est caractérisé par l'ensemble (non ordonné) des boules tirées. De ce point de vue, il y aurait $\binom{11}{2} = 55$ éléments dans l'ensemble fondamental. Il est facile de voir que l'hypothèse «tous les événements sont équiprobables» quand l'ordre est jugé significatif implique celle que les événements sont aussi équiprobables quand l'ordre de tirage n'est plus considéré significatif (à prouver). En utilisant cette seconde représentation de l'expérience, nous voyons donc que la probabilité cherchée est:

$$\frac{\binom{6}{1} \binom{5}{1}}{\binom{11}{2}} = \frac{6}{11}$$

ce qui évidemment est en accord avec la réponse précédente. ■

Exemple 2.6 Un comité de 5 personnes doit être choisi parmi les 6 hommes et 9 femmes d'un groupe. Si le choix est le résultat du hasard, quelle est la probabilité que le comité soit composé de 3 hommes et 2 femmes?

SOLUTION. Admettons que «choix dû au hasard» signifie que chacune des $\binom{15}{5}$ combinaisons possibles a les mêmes chances d'apparaître. La probabilité cherchée sera donc égale à:

$$\frac{\binom{6}{3} \binom{9}{2}}{\binom{15}{5}} = \frac{240}{1001}$$

Exemple 2.7 Une main de poker comprend 5 cartes¹. Si celles-ci ont des valeurs consécutives et ne sont pas de la même couleur, nous dirons que la main est une suite. Par exemple, une main comprenant le cinq, le six, le sept, le huit de pique et le neuf de cœur est une suite. Quelle est la probabilité de se voir distribuer une suite?

SOLUTION. On commence par admettre que les $\binom{52}{5}$ mains possibles au poker sont toutes équiprobables. Pour déterminer le nombre de tirages qui sont des suites, on va déterminer en premier lieu le nombre de tirages pour lesquels la main comprendra un as, un deux, un trois, un quatre et un cinq (sans s'intéresser à la question de savoir si l'on a une suite). Du fait que l'as peut être l'un des quatre as du jeu, qu'il en est de même pour le deux, le trois, le quatre et le cinq, il découle qu'il y a 4^5 tirages livrant exactement un as, un deux, un trois, un quatre et un cinq.

Donc, puisque dans 4 de ces tirages toutes les cartes seront de la même couleur (une telle main est appelée suite royale), il résulte qu'il y a $4^5 - 4$ mains qui sont des suites

¹ Ndt: il s'agit ici d'un jeu de 52 cartes où l'on distingue 4 couleurs.

commençant à l'as. De même il y en a $4^5 - 4$ qui contiennent un dix, un valet, une dame, un roi et l'as. Il y a donc $10(4^5 - 4)$ mains qui sont des suites. La probabilité désirée est donc

$$\frac{10(4^5 - 4)}{\binom{52}{5}} \approx .0039$$

■

Exemple 2.8 Une main de poker de 5 cartes est appelée main pleine si elle comprend 3 cartes de la même valeur et 2 autres, mais de même valeur entre elles également. Une main pleine comprend donc trois cartes d'une sorte plus une paire. Quelle est la probabilité de se voir distribuer une main pleine?

SOLUTION. De nouveau nous admettons que chacune des $\binom{52}{5}$ mains possibles est de même probabilité. Pour déterminer le nombre de mains pleines possibles, nous noterons d'abord qu'il y a $\binom{4}{3} \cdot \binom{4}{2}$ combinaisons différentes de, disons, deux 10 et trois valets. Comme il y a 13 choix différents pour le choix de la paire et après ce choix 12 autres possibilités pour la valeur des 3 cartes restantes, il résulte que la probabilité d'une main pleine est

$$\frac{13 \cdot 12 \cdot \binom{4}{2} \binom{4}{3}}{\binom{52}{5}} \approx .0014$$

■

Exemple 2.9 Lors d'une partie de bridge, les 52 cartes du paquet sont réparties entre les 4 joueurs.

- Quelle est la probabilité qu'un joueur reçoive les 13 piques?
- Quelle est la probabilité que chaque joueur reçoive un as?

SOLUTION. a) Il y a $\binom{52}{13, 13, 13, 13}$ répartitions possibles des cartes entre les 4 joueurs. Comme il y a $\binom{39}{13, 13, 13}$ répartitions possibles des cartes pour lesquelles un joueur donné détient les 13 piques, il en résulte que la probabilité désirée est donnée par

$$\frac{4 \binom{39}{13, 13, 13}}{\binom{52}{13, 13, 13, 13}} \approx 6.3 \times 10^{-12}$$

b) Pour déterminer le nombre de tirages dans lesquels chaque joueur reçoit exactement un as, mettons les as de côté et notons qu'il y a $\binom{48}{12, 12, 12, 12}$ répartitions possibles des 48 cartes restantes lorsque chaque joueur en reçoit 12. Comme il y a 4!

manières de répartir les 4 as pour que chaque joueur en reçoive 1, nous voyons que le nombre de tirages possibles où chaque joueur reçoit exactement 1 as est $4! \binom{48}{12, 12, 12, 12}$. Donc la probabilité désirée est

$$\frac{4! \binom{48}{12, 12, 12, 12}}{\binom{52}{13, 13, 13, 13}} \approx .105$$

■

L'exemple suivant illustre le fait que les résultats en probabilité peuvent être tout à fait surprenants au premier abord.

Exemple 2.10 Si n personnes sont présentes dans une pièce, quelle est la probabilité que leurs anniversaires tombent sur des jours tous différents? Quelle valeur faut-il donner à n pour que cette probabilité descende en dessous de $\frac{1}{2}$?

SOLUTION. Comme chaque personne peut célébrer son anniversaire lors de n'importe lequel des 365 jours de l'an, il y a au total $(365)^n$ situations possibles (on exclut le cas des gens nés un 29 février). En admettant que chaque situation est équiprobable, on voit que la probabilité cherchée est $(365)(364)(363)\dots(365-n+1)/(365)^n$. On sera surpris d'apprendre que lorsque n vaut 23, cette probabilité est inférieure à $\frac{1}{2}$. Cela veut dire que si 23 personnes se trouvent dans une pièce, la probabilité qu'au moins deux d'entre elles aient leur anniversaire le même jour dépasse $\frac{1}{2}$. Beaucoup de gens sont surpris par un tel résultat. Peut-être encore plus surprenant cependant est que cette probabilité augmente à 0,97 quand il y a 50 personnes dans la pièce. Et avec 100 personnes dans la pièce, les chances sont à plus de 3 000 000 contre 1 (ou encore: la probabilité est supérieure à $(3 \times 10^6)/(3 \times 10^6 + 1)$) pour qu'au moins 2 personnes aient leur anniversaire le même jour. ■

Exemple 2.11 Une équipe de football est composée de 20 joueurs attaquants et 20 joueurs défensifs. Les joueurs doivent être regroupés par paires pour qu'on puisse composer des chambrées de deux. Si le regroupement est fait au hasard, quelle est la probabilité qu'il n'y ait pas de paires mixtes de camarades de chambre? Quelle est la probabilité qu'il y ait $2i$ paires mixtes, $i = 1, 2, \dots, 10$?

SOLUTION. Il y a

$$\binom{40}{2, 2, \dots, 2} = \frac{(40)!}{(2!)^{20}}$$

manières de répartir les 40 joueurs en 20 paires ordonnées. Cela veut dire qu'il y a $(40)!/2^{20}$ manières de répartir les joueurs en une paire numéro 1, une paire numéro 2 et ainsi de suite. De ce fait, il y a $40!/2^{20}(20)!$ manières de répartir les joueurs en paires non ordonnées. De plus, puisqu'une répartition ne livrera pas de paire mixte si les attaquants (resp. les défenseurs) sont appariés entre eux, il s'ensuit qu'il y a

$[20!/2^{10}(10)!]^2$ répartitions de ce genre. De ce fait, la probabilité p_0 de n'avoir aucune paire mixte de camarades de chambre est donnée par

$$P_0 = \frac{\left(\frac{(20)!}{2^{10}(10)!}\right)^2}{\frac{(40)!}{2^{20}(20)!}} = \frac{[(20)!]^3}{[(10)!]^2(40)!}$$

Pour déterminer P_{2i} , la probabilité qu'il y ait $2i$ paires mixtes, remarquons d'abord qu'il y a $\binom{20}{2i}^2$ manières de choisir les $2i$ défenseurs et les $2i$ attaquants qui composeront les paires mixtes. Les $4i$ joueurs peuvent être appariés en $(2i)!$ paires mixtes. Ceci du fait que le premier attaquant peut être apparié avec n'importe lequel des $2i$ défenseurs, le second attaquant avec n'importe lequel des $2i - 1$ défenseurs restants, et ainsi de suite. Comme les $20 - 2i$ défenseurs (resp. attaquants) restants doivent être appariés entre eux, il s'ensuit qu'il y a

$$\binom{20}{2i}^2 (2i)! \left[\frac{(20 - 2i)!}{2^{10-i}(10 - i)!} \right]^2$$

répartitions qui mènent à $2i$ paires mixtes. D'où

$$P_{2i} = \frac{\binom{20}{2i}^2 (2i)! \left[\frac{(20 - 2i)!}{2^{10-i}(10 - i)!} \right]^2}{\frac{(40)!}{2^{20}(20)!}} \quad i = 0, 1, \dots, 10$$

Les P_{2i} , $i = 0, 1, \dots, 10$ peuvent maintenant être calculées ou approximées en faisant usage d'un résultat dû à Stirling, montrant que $n^{n+1/2} e^{-n} \sqrt{2\pi}$ approche $n!$. On obtient par exemple

$$\begin{aligned} P_0 &= 1.3403 \times 10^{-6} \\ P_{10} &= .345861 \\ P_{20} &= 7.6068 \times 10^{-6} \end{aligned} \quad \blacksquare$$

Les trois prochains exemples illustrent l'utilité du théorème 2.7. Dans l'exemple 2.12, l'introduction des probabilités nous permet d'obtenir une solution rapide à un problème de dénombrement.

Exemple 2.12 36 membres d'un club jouent au tennis, 28 jouent au squash et 18 jouent au badminton. En outre, 22 membres jouent au tennis et au squash, 12 pratiquent le tennis et le badminton, 9 jouent au squash et au badminton et 4 pratiquent les trois sports. Combien de membres de ce club pratiquent au moins un des trois sports?

SOLUTION. Soit N le nombre de membres du club, et introduisons les probabilités en admettant qu'un membre du club est sélectionné au hasard. Si pour chaque sous-ensemble C de membres du club, on appelle $P(C)$ la probabilité que le membre sélectionné appartienne à C , alors

$$P(C) = \frac{\text{nombre de membres dans } C}{N}$$

Maintenant, T étant l'ensemble des membres qui jouent au tennis, S celui de ceux qui pratiquent le squash et B celui de ceux qui jouent au badminton, le théorème 2.7 entraîne que

$$\begin{aligned} P(T \cup S \cup B) &= P(T) + P(S) + P(B) - P(TS) - P(TB) - P(SB) + P(TSB) \\ &= \frac{36 + 28 + 18 - 22 - 12 - 9 + 4}{N} \\ &= \frac{43}{N} \end{aligned}$$

D'où nous pouvons conclure que 43 membres pratiquent au moins un des sports.

L'exemple qui suit possède deux avantages: non seulement il donne lieu à un résultat quelque peu étonnant, mais il est aussi d'intérêt théorique.

Exemple 2.13 Problème de rencontre

Une réception réunit N invités, tous des hommes. Chacun jette son chapeau au milieu de la pièce. On mélange les chapeaux puis chacun en choisit un au hasard.

- Quelle est la probabilité qu'aucun des hommes ne choisisse son propre chapeau?
- Quelle est la probabilité que k des hommes exactement sélectionnent leur propre chapeau?

SOLUTION. a) Nous répondrons à la première partie en calculant d'abord la probabilité complémentaire qu'au moins un homme choisisse son propre chapeau. Désignons par E_i , $i = 1, 2, \dots, N$ l'événement «le i -ème homme choisit son propre chapeau». Alors en vertu du théorème 2.7, la probabilité $P\left(\bigcup_{i=1}^N E_i\right)$ qu'au moins un homme ait choisi son propre chapeau est donnée par

$$\begin{aligned} P\left(\bigcup_{i=1}^N E_i\right) &= \sum_{i=1}^N P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \dots \\ &\quad + (-1)^{n+1} \sum_{i_1 < i_2 < \dots < i_n} P(E_{i_1} E_{i_2} \dots E_{i_n}) \\ &\quad + \dots + (-1)^{N+1} P(E_1 E_2 \dots E_N) \end{aligned}$$

Considérons le résultat de cette expérience comme un vecteur de N nombres où le i -ème élément est le numéro du chapeau choisi par le i -ème homme. Il y a alors $N!$ tirages possibles [le résultat $(1, 2, \dots, N)$ signifie, par exemple, que chaque homme a choisi son propre chapeau]. De plus, E_1, E_2, \dots, E_n , l'événement que

chacun des n hommes i_1, i_2, \dots, i_n choisisse son propre chapeau, peut survenir de $(N-n)[N-(n+1)] \dots 3 \cdot 2 \cdot 1 = (N-n)!$ manières possibles; car, pour les $N-n$ hommes restants, le premier peut choisir n'importe lequel parmi $N-n$ chapeaux, le second peut choisir parmi $N-(n+1)$ chapeaux et ainsi de suite. Ainsi, en admettant que les $N!$ tirages possibles soient équiprobables, nous voyons que

$$P(E_{i_1} E_{i_2} \dots E_{i_n}) = \frac{(N-n)!}{N!}$$

Aussi, nous pouvons écrire, puisqu'il y a $\binom{N}{n}$ termes dans $\sum_{i_1 < i_2 < \dots < i_n} P(E_{i_1} E_{i_2} \dots E_{i_n})$:

$$\sum_{i_1 < i_2 < \dots < i_n} P(E_{i_1} E_{i_2} \dots E_{i_n}) = \frac{N!(N-n)!}{(N-n)!n!N!} = \frac{1}{n!}$$

et par conséquent

$$P\left(\bigcup_{i=1}^N E_i\right) = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{N+1} \frac{1}{N!}$$

Ainsi la probabilité qu'aucun des hommes ne choisisse son chapeau est

$$1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^N}{N!}$$

Cette probabilité est, pour N grand, approximativement égale à $e^{-1} \approx 0,36788$. En d'autres termes, pour des grandes valeurs de N , la probabilité qu'aucun des hommes ne sélectionne son propre chapeau est d'environ 0,37 (bien des lecteurs auront sans doute plutôt pensé à tort que cette probabilité tendrait vers 1 lorsque N devient infini).

b) Pour obtenir la probabilité qu'exactly k des N hommes choisissent leur propre chapeau, fixons dans un premier temps notre attention sur un groupe particulier de k hommes. Le nombre de manières pour que ces k hommes et eux seulement choisissent leur propre chapeau est égal au nombre de manières pour que les $N-k$ autres hommes choisissent parmi leurs chapeaux sans qu'aucun d'entre eux ne tombe sur le sien. Mais comme

$$1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^{N-k}}{(N-k)!}$$

est la probabilité qu'aucun de ces $N-k$ hommes tirant dans le groupe de leurs chapeaux ne tombe sur le sien, il en résulte que le nombre de cas où nos k hommes et eux seuls ont tiré leur propre chapeau est

$$(N-k)! \left[1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^{N-k}}{(N-k)!} \right]$$

De ce fait, puisqu'il y a $\binom{N}{k}$ compositions possibles pour le groupe des k hommes, il y aura

$$\binom{N}{k} (N-k)! \left[1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{N-k}}{(N-k)!} \right]$$

situations où exactement k hommes ont tiré leur propre chapeau. La probabilité cherchée est donc

$$\frac{\binom{N}{k} (N-k)! \left[1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{N-k}}{(N-k)!} \right]}{N!} \\ = \frac{1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{N-k}}{(N-k)!}}{k!}$$

qui, pour N grand, devient approximativement $e^{-1}/k!$. Ces valeurs $e^{-1}/k!$, $k = 0, 1, \dots$, ont une importance théorique. Elles représentent en effet les valeurs associées à la distribution de Poisson. Ce point sera développé au chapitre 4.¹ ■

Pour illustrer autrement encore l'utilité du théorème 2.7, on peut citer l'exemple suivant.

Exemple 2.14 Si 10 couples mariés sont assis au hasard autour d'une table, calculer la probabilité qu'aucune femme ne soit assise à côté de son mari.

SOLUTION. Si nous désignons par E_i , $i = 1, 2, \dots, 10$ l'événement que le couple i est réuni, il en résulte que la probabilité cherchée est $1 - P\left(\bigcup_{i=1}^{10} E_i\right)$. Mais en vertu du théorème 2.7

$$P\left(\bigcup_{i=1}^{10} E_i\right) = \sum_1^{10} P(E_i) - \cdots + (-1)^{n+1} \sum_{i_1 < i_2 < \cdots < i_n} P(E_{i_1} E_{i_2} \cdots E_{i_n}) \\ + \cdots - P(E_1 E_2 \cdots E_{10})$$

Pour calculer $P(E_{i_1}, E_{i_2}, \dots, E_{i_n})$, nous remarquerons d'abord qu'il y a $19!$ manières d'asseoir 20 personnes autour d'une table ronde (pourquoi?). Le nombre de configurations qui aboutissent à ce qu'un ensemble déterminé de n hommes soient assis à côté de leur femme peut être calculé de manière très simple. On se représente d'abord chacun des n couples de gens mariés comme étant une entité en soi. Nous aurions dans ce cas à disposer $20 - 2n + n = 20 - n$ entités autour de la table et il y a clairement $(20 - n)!$ de ces dispositions. Finalement, comme chacun des n couples de gens mariés peut s'asseoir de deux manières différentes, il en résulte qu'il y a $2^n(20 - n)!$

¹ Voir l'exemple 3.30 pour une autre approche de ce problème.

dispositions qui aboutissent à ce qu'un groupe fixé de n hommes soient assis à côté de leur femme. Donc

$$P(E_{i_1} E_{i_2} \cdots E_{i_n}) = \frac{2^n (19-n)!}{(19)!}$$

De ce fait, le théorème 2.7 livre que la probabilité de trouver au moins un couple réuni est

$$\binom{10}{1} 2^1 \frac{(18)!}{(19)!} - \binom{10}{2} 2^2 \frac{(17)!}{(19)!} + \binom{10}{3} 2^3 \frac{(16)!}{(19)!} - \cdots - \binom{10}{10} 2^{10} \frac{9!}{(19)!} \approx .6605$$

et la probabilité cherchée est 0,3395. ■

Exemple 2.15 Considérons une équipe d'athlétisme qui vient de terminer sa saison avec un palmarès final de n victoires et m défaites. En examinant la séquence des victoires et défaites nous espérons déterminer si l'équipe a eu des suites d'épreuves pendant lesquelles elle avait une chance plus grande de gagner que pendant les autres. Un moyen d'éclairer un peu cette question est de compter le nombre de chaînes de victoires et de voir ensuite quelle probabilité aurait ce résultat en admettant que les $(n+m)!/(n!m!)$ séquences possibles comprenant n victoires et m défaites sont équiprobables. Par «chaîne de victoires» nous entendons une séquence ininterrompue de victoires. Par exemple, si $n = 10$, $m = 6$ et si la séquence des résultats est V V D D V V V D V D D D V V V V, alors il y a eu 4 chaînes de victoires – la première de longueur 2, la seconde de 3, la troisième de 1 et la quatrième de 4.

Supposons maintenant qu'une équipe enregistre n victoires et m défaites. En admettant que les $(n+m)!/(n!m!) = \binom{n+m}{n}$ séquences sont équiprobables, déterminons la probabilité qu'il y ait exactement r chaînes de victoires. Pour l'obtenir, considérons d'abord n'importe quel vecteur d'entiers positifs x_1, x_2, \dots, x_r avec $x_1 + x_2 + \dots + x_r = n$, et voyons combien de séquences comprennent r suites de victoires dans lesquelles la i -ème chaîne est de taille x_i , $i = 1, \dots, r$. Pour toute telle séquence, si nous désignons par y_1 le nombre de défaites avant la première chaîne de victoires, y_2 celui avant la deuxième chaîne de victoires, y_{r+1} celui après la dernière chaîne de victoires, alors les y_i satisfont

$$y_1 + y_2 + \cdots + y_{r+1} = m \quad y_1 \geq 0, y_{r+1} \geq 0, y_i > 0, i = 2, \dots, r$$

et la séquence peut être schématiquement représentée par

$$\underbrace{DD\dots D}_{y_1} \underbrace{VV\dots V}_{x_1} \underbrace{DD\dots D}_{y_2} \underbrace{VV\dots V}_{x_2} \dots \underbrace{VV\dots V}_{x_r} \underbrace{DD\dots D}_{y_{r+1}}$$

De ce fait, le nombre de séquences qui donnent lieu à r chaînes de victoires – la i -ème de longueur x_i , $i = 1, \dots, r$ – est égal au nombre d'entiers y_1, \dots, y_{r+1} qui satisfont les conditions ci-dessus ou, de manière équivalente, au nombre d'entiers positifs

$$\bar{y}_1 = y_1 + 1, \quad \bar{y}_i = y_i, i = 2, \dots, r, \bar{y}_{r+1} = y_{r+1} + 1$$

qui satisfont

$$\bar{y}_1 + \bar{y}_2 + \cdots + \bar{y}_{r+1} = m + 2$$

En vertu du théorème 1.10, il y a $\binom{m+1}{r}$ séquences de ce genre. De ce fait le nombre total de séquences donnant suite à r chaînes de victoires est $\binom{m+1}{r}$ multiplié par le nombre de solutions entières positives à l'équation $x_1 + \dots + x_r = n$. Et donc, de nouveau en vertu du théorème 1.10, il y a $\binom{m+1}{r} \binom{n-1}{r-1}$ séquences livrant r chaînes de victoires. Comme il y a $\binom{m+n}{n}$ séquences équiprobables, nous pouvons déduire que

$$P(\{r \text{ chaînes de victoires}\}) = \frac{\binom{m+1}{r} \binom{n-1}{r-1}}{\binom{m+n}{n}}, \quad r \geq 1$$

Par exemple, si $n = 8$, $m = 6$, alors la probabilité d'avoir 7 chaînes est $\binom{7}{7} \binom{7}{6} / \binom{14}{8} = \frac{1}{429}$ si les $\binom{14}{8}$ résultats sont tous équiprobables. De ce fait, si la séquence était V D V D V D V D V V D V D V, alors nous pourrions suspecter que la probabilité pour l'équipe de gagner a été variable avec le temps. En particulier, la probabilité que l'équipe gagne semble être très haute quand elle a perdu sa dernière épreuve et particulièrement basse quand elle l'a gagnée. A l'extrême inverse, si la séquence avait été V V V V V V V D D D D D D, alors il n'y aurait eu qu'une chaîne de victoires. Comme $P(\{1 \text{ chaîne}\}) = \binom{7}{1} \binom{7}{0} / \binom{14}{8} = \frac{1}{429}$, il semble de nouveau improbable que la probabilité de victoire de l'équipe soit restée invariable tout au long des 14 épreuves. ■

2.6 THÉORÈME DE PASSAGE À LA LIMITE

Une suite d'événements $\{E_n, n \geq 1\}$ est dite suite croissante, si

$$E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots$$

alors qu'elle est dite décroissante si

$$E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots$$

Si $\{E_n, n \geq 1\}$ est une suite croissante d'événements, alors nous définissons un nouvel événement noté $\lim_{n \rightarrow \infty} E_n$:

$$\lim_{n \rightarrow \infty} E_n = \bigcup_{i=1}^{\infty} E_i$$

De même, si $\{E_n, n \geq 1\}$ est une suite décroissante d'événements, nous définissons $\lim_{n \rightarrow \infty} E_n$ par:

$$\lim_{n \rightarrow \infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

Démontrons maintenant le théorème 2.8.

Théorème 2.8

Si $\{E_n, n \geq 1\}$ est une suite soit croissante, soit décroissante d'événements, alors

$$\lim_{n \rightarrow \infty} P(E_n) = P(\lim_{n \rightarrow \infty} E_n)$$

PREUVE. Supposons d'abord que $\{E_n, n \geq 1\}$ est une suite croissante et définissons les événements $F_n, n \geq 1$ par

$$F_1 = E_1$$

$$F_n = E_n \left(\bigcup_{i=1}^{n-1} E_i \right)^c = E_n E_{n-1}^c \quad n > 1$$

où nous avons utilisé le fait que $\bigcup_{i=1}^{n-1} E_i = E_{n-1}$, puisque les événements sont emboîtés.

En d'autres termes, F_n comprend les points de E_n qui ne sont dans aucun des E_i qui le précèdent. Il est facile de voir que les F_n sont des événements s'excluant mutuellement et tels que

$$\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i \quad \text{et} \quad \bigcup_{i=1}^n F_i = \bigcup_{i=1}^n E_i \quad \text{pour tout } n \geq 1$$

Ainsi

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} E_i\right) &= P\left(\bigcup_{i=1}^{\infty} F_i\right) \\ &= \sum_1^{\infty} P(F_i) \quad (\text{Axiome 2.3}) \\ &= \lim_{n \rightarrow \infty} \sum_1^n P(F_i) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n F_i\right) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n E_i\right) \\ &= \lim_{n \rightarrow \infty} P(E_n) \end{aligned}$$

ce qui prouve le résultat lorsque $\{E_n, n \geq 1\}$ est croissante.

Si $\{E_n, n \geq 1\}$ est une suite décroissante, alors $\{E_n^c, n \geq 1\}$ est croissante; de ce fait, d'après les équations précédentes,

$$P\left(\bigcup_{i=1}^{\infty} E_i^c\right) = \lim_{n \rightarrow \infty} P(E_n^c)$$

Mais, comme $\bigcup_{i=1}^{\infty} E_i^c = \left(\bigcap_{i=1}^{\infty} E_i\right)^c$, nous voyons que

$$P\left(\left(\bigcap_{i=1}^{\infty} E_i\right)^c\right) = \lim_{n \rightarrow \infty} P(E_n^c)$$

ou, de manière équivalente,

$$1 - P\left(\bigcap_1^{\infty} E_i\right) = \lim_{n \rightarrow \infty} [1 - P(E_n)] = 1 - \lim_{n \rightarrow \infty} P(E_n)$$

$$P\left(\bigcap_1^{\infty} E_i\right) = \lim_{n \rightarrow \infty} P(E_n)$$

ce qui prouve le résultat. ■

Exemple 2.16 Probabilité et paradoxe

Supposons que nous ayons une urne infiniment grande et une collection infinie de boules numérotées 1, 2, ..., n, Considérons l'expérience réalisée comme il suit: à minuit moins une, les boules 1 à 10 sont placées dans l'urne et la boule 10 est retirée (on admet que le retrait est instantané). A minuit moins 30 secondes, les boules 11 à 20 sont placées dans l'urne et la boule 20 retirée. A minuit moins 15 secondes, les boules 21 à 30 sont introduites et la boule 30 est retirée. A minuit moins 7 secondes et demie, etc. La question intéressante est: combien y a-t-il de boules dans l'urne à minuit?

La réponse à cette question est clairement qu'il y a une infinité de boules dans l'urne à minuit puisque toute boule dont le numéro n'est pas de la forme $10n$, $n \geq 1$, aura été placée dans l'urne et n'aura pas été retirée avant minuit. Ainsi le problème est résolu lorsque l'expérience est réalisée comme nous l'avons décrite.

Cependant, modifions l'expérience et supposons qu'à minuit moins une les boules 1 à 10 sont placées dans l'urne et la boule 1 est retirée. A minuit moins 30 secondes, les boules 11 à 20 sont introduites et la boule 2 est retirée. A minuit moins 15 secondes, les boules 21 à 30 sont introduites et la boule 3 retirée. A minuit moins 7 secondes et demie, les boules 31 à 40 sont introduites et la boule 4 est retirée, et ainsi de suite. Dans cette nouvelle expérience, combien y a-t-il de boules dans l'urne à minuit?

De manière assez surprenante, la réponse est maintenant que l'urne est vide à minuit. En effet, dans le premier cas seules les boules numérotées $10n$, $n \geq 1$, sont retirées; alors que dans le second cas toutes les boules sont finalement retirées.

Supposons maintenant que dès qu'il faut retirer une boule, celle-ci est prise au hasard parmi les boules déjà présentes dans l'urne. Cela veut dire qu'à minuit moins 1, les boules 1 à 10 sont placées dans l'urne et une boule est retirée au hasard, et ainsi de suite. Dans ce cas, combien de boules y a-t-il dans l'urne à minuit?

SOLUTION. Nous montrerons qu'avec une probabilité 1 l'urne est vide à minuit. Considérons d'abord la boule 1. Définissons par E_n l'événement «la boule 1 est encore dans l'urne après que les n premiers retraits ont été effectués». Clairement,

$$P(E_n) = \frac{9 \cdot 18 \cdot 27 \cdots (9n)}{10 \cdot 19 \cdot 28 \cdots (9n + 1)}$$

Pour comprendre cette équation, il suffit de voir que si la boule 1 est encore dans l'urne après les n premiers retraits, la première boule retirée peut être choisie parmi

9, la seconde parmi 18 (il y a 19 boules dans l'urne au moment du deuxième retrait, mais l'une d'elles est la boule 1) et ainsi de suite. Le dénominateur est obtenu de la même manière.

Or, l'événement «la boule 1 est dans l'urne à minuit» est précisément $\bigcap_{n=1}^{\infty} E_n$. Comme les événements E_n , $n \geq 1$, forment une suite décroissante, il résulte du théorème 2.8 que:

$$P \{ \text{la boule numéro 1 se trouve dans l'urne à minuit} \}$$

$$\begin{aligned} &= P \left(\bigcap_{n=1}^{\infty} E_n \right) \\ &= \lim_{n \rightarrow \infty} P(E_n) \\ &= \prod_{n=1}^{\infty} \left(\frac{9n}{9n+1} \right) \end{aligned}$$

Nous allons maintenant montrer que

$$\prod_{n=1}^{\infty} \frac{9n}{9n+1} = 0$$

Comme

$$\prod_{n=1}^{\infty} \left(\frac{9n}{9n+1} \right) = \left[\prod_{n=1}^{\infty} \left(\frac{9n+1}{9n} \right) \right]^{-1}$$

cela revient à montrer

$$\prod_{n=1}^{\infty} \left(1 + \frac{1}{9n} \right) = \infty$$

Or, pour tout $m \geq 1$

$$\begin{aligned} \prod_{n=1}^{\infty} \left(1 + \frac{1}{9n} \right) &\geq \prod_{n=1}^m \left(1 + \frac{1}{9n} \right) \\ &= \left(1 + \frac{1}{9} \right) \left(1 + \frac{1}{18} \right) \left(1 + \frac{1}{27} \right) \cdots \left(1 + \frac{1}{9m} \right) \\ &> \frac{1}{9} + \frac{1}{18} + \frac{1}{27} + \cdots + \frac{1}{9m} \\ &= \frac{1}{9} \sum_{i=1}^m \frac{1}{i} \end{aligned}$$

Par conséquent, en faisant tendre m vers l'infini et en utilisant le fait que $\sum_{i=1}^{\infty} 1/i = \infty$, on obtient

$$\prod_{n=1}^{\infty} \left(1 + \frac{1}{9n} \right) = \infty$$

Donc, en notant F_i l'événement «la boule i est dans l'urne à minuit», nous avons montré que $P(F_i) = 0$. On peut alors montrer que $P(F_i) = 0$ pour tout i (le même raisonnement établi par exemple que $P(F_i) = \prod_{n=2}^{\infty} [9n/(9n+1)]$ pour $i = 11, 12, \dots, 20$). Ainsi, la probabilité $P(\bigcup_{i=1}^{\infty} F_i)$ que l'urne ne soit pas vide à minuit satisfait

$$P\left(\bigcup_{i=1}^{\infty} F_i\right) \leq \sum_{i=1}^{\infty} P(F_i) = 0$$

en vertu de l'inégalité de Boole (voir exercices 2.8.8 et 2.8.20). Aussi l'urne sera-t-elle vide à minuit, avec une probabilité de 1. ■

2.7 PROBABILITÉ EN TANT QUE MESURE DU CRÉDIT ACCORDÉ À UN FAIT

Jusqu'à présent nous avons interprété la probabilité d'un événement d'une expérience donnée comme étant une mesure de la fréquence d'apparition de l'événement lorsque l'expérience est répétée sans fin. Cependant, il existe d'autres usages du terme *probabilité*. Par exemple, nous avons tous entendu des déclarations du genre «il est probable à 90% que Shakespeare ait écrit *Hamlet*», ou «la probabilité qu'Oswald ait agi seul lors de l'assassinat de Kennedy est 0,8». Comment devons-nous interpréter ces affirmations?

L'interprétation la plus simple et naturelle est que les probabilités citées sont des mesures du crédit qu'un individu porte à la déclaration qu'il fait. En d'autres termes, un individu prononçant les déclarations ci-dessus est assez certain qu'Oswald a agi seul et plus certain encore que Shakespeare a écrit *Hamlet*. Cette interprétation des probabilités comme mesure d'une croyance est qualifiée d'approche *personnelle ou subjective* des probabilités.

Il semble logique de supposer qu'une telle mesure du crédit porté aux choses doit satisfaire tous les axiomes des probabilités. Par exemple, si nous sommes certains à 70% que Shakespeare ait écrit *Jules César* et certains à 10% que l'auteur ait en fait été Marlowe, alors il est logique de supposer que nous sommes certains à 80% que l'auteur ait été soit Shakespeare soit Marlowe. Aussi, que nous interprétions les probabilités comme mesure de croyance ou comme fréquence d'apparition à long terme, leurs propriétés mathématiques sont inchangées.

Exemple 2.17 Supposons que dans une course disputée par 7 chevaux vous sentiez que chacun des 2 premiers a 20% de chances de gagner, que les chevaux 3 et 4 ont chacun 15% de chance et que les 3 derniers ont 10% de chance chacun. Avez-vous avantage à parier à 1 contre 1 que le gagnant sera l'un des 3 premiers chevaux ou de parier, à 1 contre 1 toujours, que le vainqueur sera l'un des chevaux 1, 5, 6, 7?

SOLUTION. Calculée d'après vos probabilités personnelles sur l'issue de la course, votre probabilité de gagner le premier pari est $0,2 + 0,2 + 0,15 = 0,55$ tandis qu'elle est de $0,2 + 0,1 + 0,1 + 0,1 = 0,5$ pour le second. La première mise est donc plus intéressante. ■

Il faut remarquer qu'en supposant que les probabilités subjectives relatives à un individu sont cohérentes avec les axiomes de probabilité, nous parlons d'une personne idéale plus que réelle. Par exemple, si nous devons demander à quelqu'un ce qu'il ou elle pense des chances qu'il

- pleuve aujourd'hui
- pleuve demain
- pleuve aujourd'hui et demain
- pleuve aujourd'hui ou demain

il est bien possible qu'après quelques réflexions cette personne puisse donner respectivement 30%, 40%, 20% et 60% comme réponses. Mais malheureusement de telles réponses (ou de telles probabilités subjectives) ne sont pas cohérentes avec les axiomes de probabilité (pourquoi ne le sont-elles pas?). Nous espérons bien naturellement qu'après le lui avoir fait remarquer le répondeur finira par modifier ses réponses (une possibilité acceptable est respectivement 30%, 40%, 10% et 60%).

2.8 EXERCICES THÉORIQUES

2.8.1 Démontrer les relations suivantes

- $EF \subset E \subset E \cup F$
- Si $E \subset F$, alors $F^c \subset E^c$
- $F = FE \cup FE^c$ et $E \cup F = E \cup E^c F$
- $(\bigcup_1^x E_i)F = \bigcap_1^x E_i F$ et $(\bigcap_1^x E_i) \cup F = \bigcup_1^x (E_i \cup F)$.

2.8.2 Pour toute suite d'événements E_1, E_2, \dots , définir une nouvelle suite F_1, F_2, \dots d'événements s'excluant mutuellement (c'est-à-dire tels que $F_i F_j = \emptyset$ dès que $i \neq j$) et tels que pour tout $n \geq 1$

$$\bigcup_1^n F_i = \bigcup_1^n E_i$$

2.8.3 Soient E, F et G trois événements. Trouver des expressions pour les événements suivants que l'on dira réalisés lorsque, de E, F et G ,

- E seul l'est
- E et G le sont mais pas F
- au moins l'un des trois l'est
- au moins deux d'entre eux le sont
- les trois le sont
- aucun ne l'est
- au plus l'un des trois l'est
- au plus deux d'entre eux le sont
- exactement deux le sont
- au plus trois le sont.

2.8.4 Trouver une expression simple pour les événements suivants:

- a) $(E \cup F)(E \cup F^c)$;
 b) $(E \cup F)(E^c \cup F)(E \cup F^c)$;
 c) $(E \cup F)(F \cup G)$.

2.8.5 Soit S un ensemble donné. Si pour un certain $k > 0$, S_1, S_2, \dots, S_k sont des sous-ensembles disjoints non vides de S tels que $\bigcup_{i=1}^k S_i = S$, alors nous appelons l'ensemble $\{S_1, \dots, S_k\}$ une *partition* de S . Désignons par T_n le nombre de partitions différentes de $\{1, 2, \dots, n\}$. On aura $T_1 = 1$ (puisque la seule partition est $S_1 = \{1\}$) et $T_2 = 2$ (puisque les deux partitions possibles sont $\{\{1, 2\}, \{\{1\}, \{2\}\}$). Montrer que $T_3 = 5$, $T_4 = 15$ en exhibant toutes les partitions

$$T_{n+1} = 1 + \sum_{k=1}^n \binom{n}{k} T_k$$

et appliquer ce résultat au calcul de T_{10} .

Une façon de choisir une partition de $n + 1$ objets est de nommer un des objets «spécial». Nous obtenons alors différentes partitions en choisissant tout d'abord k , $k = 0, 1, \dots, n$, puis un sous-ensemble de taille $n - k$ parmi les objets non spéciaux et en choisissant ensuite n'importe quelle partition T_k des k objets non spéciaux restants. En ajoutant l'objet spécial au sous-ensemble de taille $n - k$, nous obtenons une partition des $n + 1$ objets.

2.8.6 On suppose qu'une expérience est répétée n fois. Pour chaque événement E de l'ensemble fondamental, soit $n(E)$ le nombre de fois où l'événement E survient; on définit $f(E)$ par $f(E) = n(E)/n$. Montrer que $f(\cdot)$ satisfait aux axiomes 2.1, 2.2 et 2.3.

2.8.7 Prouver que

$$P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E^c F G) - P(E F^c G) - P(E F G^c) - 2P(E F G).$$

2.8.8 Prouver l'inégalité de Boole

$$P\left(\bigcup_1^n E_i\right) \leq \sum_1^n P(E_i)$$

2.8.9 Si $P(E) = 0,9$ et $P(F) = 0,8$, montrer que $P(EF) \geq 0,7$. De manière plus générale, démontrer l'inégalité de Bonferroni, à savoir

$$P(EF) \geq P(E) + P(F) - 1$$

2.8.10 Montrer que la probabilité qu'un seul exactement des événements E et F se réalise est $P(E) + P(F) - 2P(EF)$.

2.8.11 Montrer que

$$P(EF^c) = P(E) - P(EF).$$

2.8.12 Montrer que

$$P(E^cF^c) = 1 - P(E) - P(F) + P(EF).$$

2.8.13 Démontrer le théorème 2.7 par induction.

2.8.14 Une urne contient M boules blanches et N noires. Si un échantillon de taille r est tiré au hasard, quelle est la probabilité qu'il contienne exactement k boules blanches? Que dire du cas $M = k = 1$?

2.8.15 Généraliser l'inégalité de Bonferroni à n événements ce qui, explicitement, revient à montrer que

$$P(E_1E_2 \cdots E_n) \geq P(E_1) + \cdots + P(E_n) - (n-1)$$

2.8.16 On considère le problème de rencontre (exemple 2.13) et on désigne par A_N le nombre de manières telles que chacun des N hommes n'ait pas tiré son propre chapeau. Prouver que

$$A_N = (N-1)(A_{N-1} + A_{N-2})$$

Cette formule, conjointement avec les conditions aux limites $A_1 = 0$ et $A_2 = 1$, peut être utilisée pour le calcul de A_N . Elle livre la probabilité voulue $A_N/N!$, probabilité de n'avoir aucune rencontre. On sait qu'après que le premier homme ait choisi un chapeau qui n'est pas le sien, il reste $N-1$ hommes qui doivent choisir dans un groupe de $N-1$ chapeaux. De ce groupe manque d'ailleurs le chapeau de l'un de ces hommes, précédemment tiré. Il y a ainsi un homme qui ne peut plus trouver son propre chapeau et un chapeau qui n'a plus de propriétaire. Montrer qu'on peut n'avoir aucune rencontre aussi bien si l'homme de trop sélectionne le chapeau de trop que s'il en sélectionne un autre.

2.8.17 Désignons par f_n le nombre de manières de jeter une pièce n fois sans que deux piles successifs n'apparaissent. Montrer que

$$f_n = f_{n-1} + f_{n-2} \quad n \geq 2, \quad \text{où } f_0 \equiv 1, f_1 \equiv 2$$

Si P_n désigne la probabilité que des piles successifs n'apparaissent jamais lors de n jets, trouver P_n (en fonction de f_n) lorsqu'on admet que toutes les séquences de n jets sont équiprobables. Calculer P_{10} .

RÉPONSE. $P_{10} = 144/2^{10} = 0,141$.

2.8.18 On considère une expérience dont l'ensemble fondamental comprend une infinité dénombrable de points. Montrer qu'il est impossible que chaque point ait la

même probabilité. Est-ce que tous les points peuvent avoir une probabilité strictement positive?

2.8.19 On considère l'exemple 2.15 qui traite du nombre de chaînes de victoires obtenues lors d'une combinaison au hasard de n victoires et m défaites. On considère maintenant le nombre total de chaînes, donc chaînes de victoires et chaînes de défaites. Montrer que

$$P\{2k \text{ chaînes}\} = 2 \frac{\binom{m-1}{k-1} \binom{n-1}{k-1}}{\binom{m+n}{n}}$$

$$P\{2k+1 \text{ chaînes}\} = \frac{\binom{m-1}{k-1} \binom{n-1}{k} + \binom{m-1}{k} \binom{n-1}{k-1}}{\binom{m+n}{n}}$$

2.8.20 A partir de l'inégalité de Boole pour un nombre fini d'événements, montrer que pour toute suite infinie d'événements $E_i, i \geq 1$,

$$P\left(\bigcup_1^{\infty} E_i\right) \leq \sum_1^{\infty} P(E_i)$$

2.8.21 Montrer que si $P(E_i) = 1$ pour tout $i \geq 1$, alors $P(\bigcap_1^{\infty} E_i) = 1$.

2.8.22 Pour une suite d'événements $E_i, i \geq 1$, on définit un nouvel événement, appelé $\limsup E_i$, comprenant tous les événements contenus dans un nombre infini de $E_i, i \geq 1$. Montrer que

$$\limsup_i E_i = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} E_i$$

2.8.23 Montrer que si $\sum_{i=1}^{\infty} P(E_i) < \infty$, alors $P(\limsup E_i) = 0$.

Ceci est un résultat important qui énonce que si $\sum_{i=1}^{\infty} P(E_i) < \infty$, alors la probabilité qu'un nombre infini de E_i survienne est 0.

Pour ce calcul, utiliser l'inclusion

$$\limsup_i E_i \equiv \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} E_i \subset \bigcup_{i=1}^{\infty} E_i$$

2.9 PROBLÈMES

Les problèmes 2.9.1 à 2.9.4 portent sur les sections 2.1 et 2.2.

2.9.1 Une boîte contient 3 jetons, un rouge, un vert et un bleu. On considère l'expérience consistant à tirer au hasard un jeton dans la boîte, à l'y remettre puis à en tirer un second. Décrire l'ensemble fondamental. Même question si le second jeton est tiré sans qu'on ait remis le premier.

2.9.2 Un dé est jeté jusqu'à ce qu'un 6 sorte, ce qui marque la fin de l'expérience. Quel est l'ensemble fondamental pour cette expérience? Notons par E_n l'événement « n jets sont nécessaires pour obtenir le premier 6». Quels points de l'espace fondamental sont contenus dans E_n ? Décrire $(\bigcup_1^{\infty} E_n)^c$.

2.9.3 On jette deux dés. On note par E l'événement «la somme des dés est impaire», par F l'événement «au moins l'un des dés montre 1», et par G «la somme des dés est 5». Décrire EF , $E \cup F$, FG , EF^c et EFG .

2.9.4 Trois joueurs, A , B et C , jettent une pièce à tour de rôle. Le premier qui obtient pile a gagné. L'ensemble fondamental S de cette expérience peut être décrit comme suit:

$$S = \left\{ \begin{array}{l} 1, 01, 001, 0001, \dots, \\ 0000 \dots \end{array} \right.$$

- Donner une interprétation des points de S .
- Décrire les événements suivants en termes de ces points:
premier événement: $A = \text{«}A \text{ gagne}\text{»}$;
deuxième événement: $B = \text{«}B \text{ gagne}\text{»}$;
troisième événement: $(A \cup B)^c$.

On admettra que A joue d'abord, puis B et enfin C .

2.9.5 Une cafétéria propose un menu composé de trois plats. On choisit un plat principal, un féculent et un dessert. Les choix possibles sont donnés ci-dessous.

	Choix
Plat principal	poulet ou rosbif
Féculent	pâtes ou riz ou pommes de terre
Dessert	glace ou gelée ou tarte aux pommes ou pêches

Une personne choisit un plat de chaque catégorie.

- a) Combien de menus possibles y a-t-il dans l'ensemble fondamental?
- b) Soit A l'événement: «on choisit la glace». Combien y a-t-il de menus possibles dans A ?
- c) Soit B l'événement: «on choisit le poulet». Combien y a-t-il de menus possibles dans B ?

- d) Donner tous les menus possibles de l'événement AB .
- e) Soit C l'événement: «on choisit le riz». Combien y a-t-il de menus possibles dans C ?
- f) Donner tous les menus possibles de l'événement ABC .

Les problèmes 2.9.5 à 2.9.36 portent sur les sections 2.3 à 2.6.

2.9.6 Un magasin accepte les cartes de crédit American Express ou VISA. 24% de ses clients possèdent une carte American Express, 61% une carte VISA et 11% possèdent les deux. Quel est le pourcentage de clients possédant une carte de crédit acceptée par le magasin?

2.9.7 60% des élèves d'une école ne portent ni bague ni collier. 20% portent une bague et 30% ont un collier. Si un des élèves est choisi au hasard, quelle est la probabilité qu'il porte

- a) une bague ou un collier?
 b) une bague et un collier?

2.9.8 Un client du rayon costumes d'un magasin achètera un costume avec une probabilité .22, une chemise avec une probabilité .30 et une cravate avec une probabilité .28. Le client achètera un costume et une chemise avec une probabilité .11, un costume et une cravate avec une probabilité .14 et une chemise et une cravate avec une probabilité .10. Un client achètera les trois vêtements avec une probabilité .06. Quelle est la probabilité qu'un client achète

- a) aucun vêtement;
 b) exactement un des vêtements?

2.9.9 Une école propose trois cours de langue : un en espagnol, un en français et un en allemand. Ces cours sont ouverts aux 100 élèves de l'école. Il y a 28 étudiants en espagnol, 26 en français et 16 en allemand. Il y a 12 étudiants qui suivent l'espagnol et le français, 4 qui suivent l'espagnol et l'allemand et 6 qui étudient le français et l'allemand. De plus, 2 élèves suivent les trois cours.

- a) Si un élève est choisi au hasard, quelle est la probabilité qu'il ou elle ne fasse partie d'aucun de ces cours?
 b) Si un élève est choisi au hasard, quelle est la probabilité qu'il ou elle suive exactement un cours de langue?
 c) Si 2 élèves sont choisis au hasard, quelle est la probabilité qu'au moins un des deux suive un cours de langue?

2.9.10 Une ville de 100 000 habitants compte trois journaux locaux: I, II et III. Les proportions de lecteurs pour ces journaux sont:

I : 10% I et II : 8% I et II et III : 1%
 II : 30% I et III : 2%
 III : 5% II et III : 4%.

Ces proportions nous indiquent par exemple que 8 000 personnes lisent à la fois les journaux I et II.

- a) Trouver le nombre de personnes ne lisant qu'un journal.

- b) Combien de personnes lisent au moins deux journaux?
 c) Il est un quotidien du soir, tandis que I et III sortent le matin. Combien de personnes lisent-elles au moins un journal du matin plus celui du soir?
 d) Combien de personnes lisent-elles un journal du matin seulement et le journal du soir?
 e) Combien de personnes ne lisent aucun journal?

2.9.11 Les données suivantes ont été fournies par l'étude d'un groupe de 1 000 abonnés d'un certain magazine. Concernant leur emploi, état civil et niveau d'éducation les réponses furent: 312 actifs, 470 personnes mariées, 525 bacheliers dont 42 actifs, 147 bacheliers mariés, 86 actifs mariés dont 25 bacheliers. Montrer que les effectifs compilés lors de cette étude sont inexacts. Pour cela, désigner par A , M et B respectivement l'ensemble des gens actifs, celui des gens mariés et celui des bacheliers. Supposer qu'une des 1 000 personnes est tirée au hasard et utiliser le théorème 2.7 pour montrer que si ces nombres sont corrects, alors $P(A \cup M \cup B) > 1$.

2.9.12 On distribue les cartes d'un paquet en comptant 52. Quelle est la probabilité que la 14ème carte distribuée soit un as? Quelle est la probabilité que le premier as survienne à la 14ème carte?

2.9.13 On admet que les $\binom{52}{5}$ mains possibles au poker sont équiprobables. Quelle est la probabilité de recevoir:

- a) une couleur? (Une main est appelée couleur lorsque les 5 cartes sont des piques seulement, ou des trèfles, ou des cœurs, ou des carreaux)
 b) Une paire? (C'est le cas lorsqu'on reçoit a, a, b, c, d où a, b, c et d sont de différentes valeurs)
 c) Deux paires (correspondant à a, a, b, b, c)?
 d) Un brelan (a, a, a, b, c)?
 e) Un carré (a, a, a, a, b)?

2.9.14 On peut jouer au poker en jetant simultanément 5 dés¹. Montrer que:

- a) $P\{5 \text{ cartes différentes}\} = 0,0926$
 b) $P\{1 \text{ paire}\} = 0,4630$
 c) $P\{2 \text{ paires}\} = 0,2315$
 d) $P\{\text{brelan}\} = 0,1543$
 e) $P\{\text{main pleine: } 3 + 2\} = 0,0386$
 f) $P\{\text{carré}\} = 0,0193$
 g) $P\{\text{poker de } 5\} = 0,0008$.

2.9.15 Huit tours sont disposées au hasard sur un jeu d'échec. Calculer la probabilité qu'aucune ne puisse en prendre une autre, donc qu'aucune ligne ni colonne ne contienne plus qu'une tour.

2.9.16 On tire d'un paquet de cartes normal (52 cartes) deux cartes au hasard. Quelle est la probabilité qu'elles forment un black jack, ou autrement dit, que l'une soit un as et l'autre un dix, un valet, une dame ou un roi?

¹ Ndt : ces dés sont identiques entre eux et leurs six faces sont toutes différentes.

2.9.17 On jette deux dés. Quelle est la probabilité que la somme des points soit i ? Faire le calcul pour $i = 2, 3, \dots, 11, 12$.

2.9.18 On jette deux dés jusqu'à ce qu'une somme de 5 ou 7 apparaisse. Trouver la probabilité qu'on s'arrête sur une somme de 5. Pour cela, désigner par E_n l'événement «une somme de 5 apparaît au n -ième double jet et sur les $n-1$ premiers jets ni la somme de 5 ni celle de 7 n'apparaît». Calculer $P(E_n)$ et montrer que $\sum_{n=1}^{\infty} P(E_n)$ est la probabilité cherchée.

2.9.19 On joue au «craps» comme suit: un joueur lance deux dés. Si la somme résultante est 2, 3 ou 12, le joueur a perdu. Si la somme est 7 ou 11, il gagne. Dans les autres cas, le joueur continue à lancer les dés jusqu'à ce qu'il sorte soit le premier résultat qu'il a tiré soit 7. Si c'est 7, il perd. Si c'est son résultat initial, il gagne. Calculer la probabilité de gagner sur un jeu.

On pourra pour cela poser $E_i =$ «le résultat initial est i et le joueur finit par gagner».

La probabilité cherchée est $\sum_{i=2}^{12} P(E_i)$. Pour calculer $P(E_i)$, poser $E_{i,n} =$ «la somme initiale est i et le joueur gagne au n -ième coup». Montrer que $P(E_i) = \sum_{n=1}^{\infty} P(E_{i,n})$.

2.9.20 Une urne contient trois boules rouges et sept noires. Les joueurs A et B tirent une boule à tour de rôle jusqu'à ce qu'une rouge sorte, A commençant. Trouver la probabilité que A tire la première boule rouge. On ne remet pas les boules tirées.

2.9.21 Une urne contient cinq boules rouges, six bleues et huit vertes. Si un groupe de trois boules est tiré au hasard, quelle est la probabilité que celles-ci soient toutes de la même couleur? Ou de couleurs différentes? Même question si chaque boule tirée est remise après qu'on ait noté sa couleur (cette méthode s'appelle *échantillonnage avec remise*).

2.9.22 Une urne A contient trois boules noires et trois rouges, alors que l'urne B en contient six et quatre respectivement. On tire une boule dans chaque urne. Quelle est la probabilité que les boules soient de la même couleur?

2.9.23 Une équipe de basket-ball réduite à trois joueurs comprend un arrière, un avant et un centre. On choisit trois hommes dans autant d'équipes de cette composition à raison d'un homme par équipe. Quelle est la probabilité d'obtenir une nouvelle équipe complète? Et celle de tirer trois joueurs de la même spécialisation?

2.9.24 Un groupe est formé de g garçons et f filles. Tous sont alignés au hasard, c'est-à-dire que chacune des $(g + f)!$ permutations possibles est de même probabilité. Quelle est la probabilité que la personne occupant la i -ième position soit une fille, $1 \leq i \leq g + f$?

2.9.25 Une forêt abrite vingt cerfs. Cinq sont capturés, marqués et relâchés. Un peu plus tard, quatre sont de nouveau capturés. Quelle est la probabilité que deux d'entre eux soient marqués? Quelles hypothèses faites-vous?

2.9.26 Dans une loterie, un joueur doit choisir 8 nombres entre 1 et 40. Le tirage sélectionne 8 numéros parmi ces 40 nombres. En admettant que le tirage est équiprobable pour les $\binom{40}{8}$ combinaisons, quelle est la probabilité que le joueur ait

- a) les 8 bons numéros;
- b) 7 numéros parmi les 8 bons;
- c) au moins 6 numéros parmi les 8 bons?

2.9.27 Le second Comte de Yarborough paria à 1000 contre 1 qu'une main de 13 cartes au bridge contiendrait au moins un 10 ou une carte de valeur supérieure (c'est-à-dire un dix, un valet, une reine, un roi ou un as). Aujourd'hui, on appelle une main qui n'a pas de carte supérieure à 9 une Yarborough. Quelle est la probabilité qu'au bridge, une main sélectionnée au hasard soit une Yarborough?

2.9.28 30 psychiatres et 24 psychologues participent à une conférence. Trois personnes parmi ces 54 sont choisies pour présenter un exposé. Quelle est la probabilité qu'au moins un psychologue soit choisi?

2.9.29 Deux cartes sont choisies aléatoirement parmi un jeu de 52 cartes. Quelle est la probabilité

- a) que ce soient 2 as;
- b) qu'elles aient la même valeur?

2.9.30 Un professeur donne à sa classe 10 problèmes en expliquant que l'examen final consistera à résoudre 5 de ces 10 problèmes, choisis aléatoirement. Si un étudiant sait résoudre 7 des 10 problèmes, quelle est la probabilité qu'il ou elle réponde correctement

- a) aux 5 problèmes;
- b) à au moins 4 des 5 problèmes?

2.9.31 Un tiroir contient n chaussettes dont 3 rouges. Quelle doit être la valeur de n pour que, si on choisit 2 chaussettes aléatoirement, la probabilité qu'elles soient les deux rouges soit $1/2$?

2.9.32 Une ville compte cinq hôtels. Si lors d'une journée trois personnes louent une chambre, quelle est la probabilité qu'elles le fassent dans trois hôtels différents? Quelles hypothèses faites-vous?

2.9.33 Il y a quatre réparateurs de télévision dans une ville. Quatre appareils tombent en panne. Quelle est la probabilité que i exactement des réparateurs soient appelés? Résoudre le problème pour $i = 1, 2, 3, 4$. Quelles hypothèses faites-vous?

2.9.34 Quelle est la probabilité de tirer au moins un 6 lorsqu'on jette un dé quatre fois?

2.9.35 On répète n fois le lancer de deux dés. Calculer la probabilité que le six apparaisse au moins une fois. Quelle valeur donner à n pour que cette probabilité atteigne $\frac{1}{2}$?

2.9.36 a) On aligne N personnes, dont A et B . Quelle est la probabilité que A et B soient voisins? b) Qu'en est-il si toutes les personnes sont disposées en cercle?

2.9.37 On tire un comité de quatre personnes au hasard d'une assemblée comprenant trois étudiants de première année, quatre de deuxième, quatre de troisième et trois de dernière. Trouver la probabilité que le comité tiré se compose:

- a) d'une personne de chaque année;
- b) de deux étudiants de deuxième et de troisième année;
- c) seulement d'étudiants de deuxième et de troisième années.

2.9.38 Une personne possède n clés dont une seule ouvre sa porte. a) Si elle les essaie au hasard en éliminant celles qui ne marchent pas, quelle est la probabilité que la porte s'ouvre au k -ième essai? b) Qu'en est-il si elle n'élimine pas les clés essayées?

2.9.39 Combien de personnes faut-il pour que la probabilité qu'au moins deux d'entre elles aient leur anniversaire le même mois soit au moins $1/2$? Admettre que tous les mois sont équiprobables.

2.9.40 Si 12 personnes sont dans une même pièce, quelle est la probabilité qu'aucune d'entre elles ne soit née le même mois?

2.9.41 On considère un groupe de 20 personnes. Quelle est la probabilité que, parmi les 12 mois de l'année, il y ait 4 mois contenant chacun exactement 2 anniversaires et 4 mois contenant chacun exactement 3 anniversaires?

2.9.42 Un groupe de six hommes et six femmes est divisé au hasard en deux sous-groupes de même taille. Quelle est la probabilité que chaque sous-groupe ait la même composition?

2.9.43 Lors d'une donnée de bridge¹, quelle est la probabilité que vous ayez cinq piques et votre partenaire les huit autres?

2.9.44 Supposons que l'on distribue au hasard n boules dans N compartiments. Trouver la probabilité que m boules tombent dans le premier compartiment. On admettra que les N^n répartitions sont équiprobables.

2.9.45 Un cabinet contient 10 paires de chaussures et on en tire 8 chaussures au hasard. Quelle est la probabilité

- a) qu'il n'y ait aucune paire;
- b) qu'il y ait une paire exactement?

2.9.46 Une équipe de basket-ball compte 6 joueurs noirs et 4 blancs. Si les joueurs sont répartis en chambrées de deux personnes, quelle est la probabilité qu'on trouve deux chambrées mixtes?

Ndt: 52 cartes réparties entre quatre joueurs

2.9.47 On dispose sur un rang 4 couples mariés au hasard. Quelle est la probabilité qu'aucun mari ne soit situé à côté de sa femme?

2.9.48 Calculer, en utilisant le théorème 2.7, la probabilité qu'une main de bridge soit dépourvue d'au moins une des quatre couleurs. On remarquera que la probabilité n'est pas

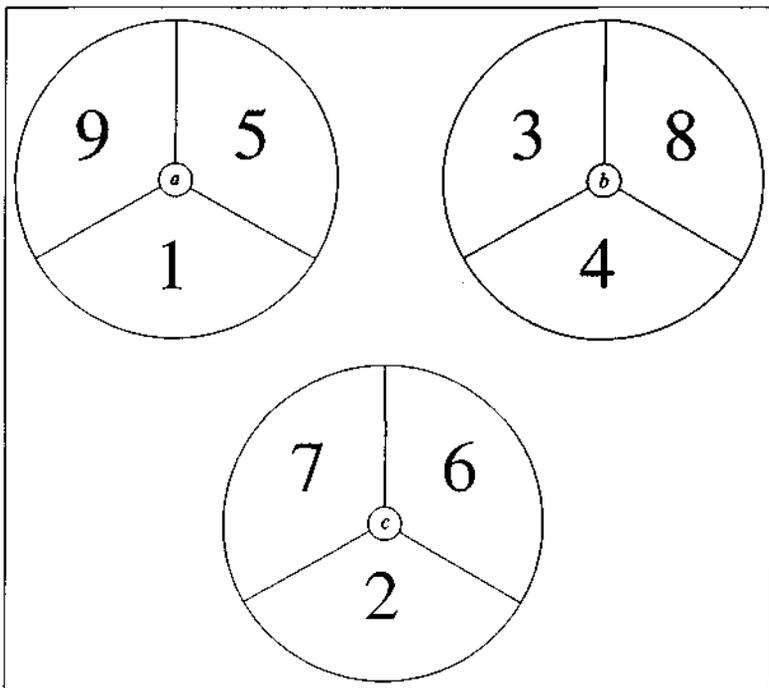
$$\frac{\binom{4}{1} \binom{39}{13}}{\binom{52}{13}}$$

Pourquoi pas?

2.9.49 Calculer la probabilité qu'une main de 13 cartes (tirées parmi 52) contienne

- l'as et le roi de l'une des quatre couleurs;
- les quatre cartes de l'une des treize valeurs.

2.9.50 Deux personnes jouent au jeu suivant. Le joueur *A* choisit une des trois roues ci-dessous et le joueur *B* choisit une des deux restantes. Les deux joueurs font tourner leur roue et celui qui tombe sur le numéro le plus grand gagne. En admettant que les trois régions de chaque roue sont équiprobables, préféreriez-vous être le joueur *A* ou le joueur *B*? Expliquer votre réponse!



Probabilité conditionnelle et indépendance

3.1 INTRODUCTION

Nous allons présenter dans ce chapitre l'un des plus importants concepts de la théorie des probabilités, celui de *probabilité conditionnelle*. L'importance de ce concept est de deux ordres. En premier lieu on s'intéresse souvent à calculer des probabilités lorsqu'une partie de l'information concernant le résultat de l'expérience est disponible; dans une telle situation les probabilités cherchées sont justement des probabilités conditionnelles. Deuxièmement, même lorsqu'aucune information partielle n'est disponible, il est souvent avantageux d'utiliser un détour par certaines probabilités conditionnelles pour réussir le calcul des probabilités cherchées.

3.2 PROBABILITÉS CONDITIONNELLES

3.2.1 Présentation intuitive

Supposons que nous jetions deux dés et que chacun des 36 événements élémentaires ait la même probabilité de survenir, soit $\frac{1}{36}$. Supposons encore que nous puissions observer le premier dé, qui donne un 3. Sur la base de cette information, quelle est dès lors la probabilité que la somme des deux dés donne 8? Pour calculer cette probabilité on peut procéder comme suit: le dé initial étant un 3, il ne peut plus y avoir que 6 événements dans notre expérience, à savoir: (3,1), (3,2), (3,3), (3,4), (3,5) et (3,6). Puisque chacun de ces événements a originellement la même probabilité d'apparaître, ils auront encore des probabilités égales. Autrement dit, étant donné que le premier dé est un 3, la probabilité (conditionnelle) de chacun des événements (3,1), (3,2), (3,3), (3,4), (3,5) et (3,6) devient $\frac{1}{6}$, tandis que la probabilité (conditionnelle) des 30 autres événements de l'ensemble fondamental devient 0. Aussi la probabilité cherchée est-elle ici $\frac{1}{6}$.

Si nous désignons respectivement par E et F les événements «la somme des dés est 8» et «le premier dé donne 3», une probabilité comme celle calculée dans l'exemple ci-dessus est appelée probabilité conditionnelle que E apparaisse sachant que F est réalisé et est notée $P(E|F)$.

3.2.2 Généralisation

On s'inspire de la même démarche pour dériver une formule générale donnant $P(E|F)$ pour tout événement E et F : si F est réalisé, alors E apparaîtra chaque fois qu'on aura affaire à un événement de E et de F à la fois; en d'autres termes, ce sera un événement de EF . Par ailleurs, comme nous savons que F est réalisé, cet ensemble devient notre nouvel ensemble fondamental, d'ailleurs réduit; par conséquent la probabilité conditionnelle de l'événement E sera donnée par comparaison de la probabilité non conditionnelle de EF avec la probabilité non conditionnelle de F . On débouche ainsi sur la définition suivante. Si $P(F) > 0$, la **probabilité conditionnelle** de E sera

$$P(E|F) = \frac{P(EF)}{P(F)} \quad (3.1)$$

Exemple 3.1 Une pièce de monnaie est lancée deux fois. Si nous supposons que les quatre points de l'ensemble fondamental $S = \{(F,F), (F,P), (P,F), (P,P)\}$ sont équiprobables, quelle est la probabilité conditionnelle que les deux jets amènent «face» sachant que le premier est déjà un «face»?

SOLUTION. En désignant par $E = \{(F,F)\}$ l'événement «les 2 jets amènent face» et par $F = \{(F,F), (F,P)\}$ «le premier jet donne face», la probabilité voulue est donnée par

$$\begin{aligned} P(E|F) &= \frac{P(EF)}{P(F)} \\ &= \frac{P(\{F, F\})}{P(\{(F, F), (F, P)\})} \\ &= \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2} \end{aligned}$$

■

Exemple 3.2 Une urne contient 10 billes blanches, 5 jaunes et 10 noires. Une bille est tirée au hasard de l'urne et l'on constate qu'elle n'est pas noire. Quelle est la probabilité qu'elle soit jaune?

SOLUTION. Soit J l'événement «la bille tirée est jaune» et soit N^c l'évènement «elle n'est pas noire». De (3.1) on tire

$$P(J|N^c) = \frac{P(JN^c)}{P(N^c)}$$

Cependant, $JN^c = J$ puisque la bille sera à la fois jaune et non noire, si et seulement si elle est jaune. Nous obtenons ainsi, en supposant que chacune des 25 billes a la même chance d'être choisie:

$$P(J|N^c) = \frac{\frac{5}{25}}{\frac{15}{25}} = \frac{1}{3}$$

Il faut noter qu'on aurait aussi pu déduire cette probabilité en travaillant directement avec l'ensemble fondamental réduit. Comme nous savons en effet que la bille choisie n'est pas noire, le problème se réduit à calculer la probabilité qu'une bille soit jaune lorsqu'elle est choisie au hasard dans une urne en contenant 10 blanches et 5 jaunes. Cette probabilité est évidemment $\frac{5}{15} = \frac{1}{3}$. ■

Si on sait les événements équiprobables, il est souvent plus facile de calculer une probabilité conditionnelle en considérant l'ensemble fondamental réduit qu'en invoquant (3.1).

Exemple 3.3 Dans un jeu de bridge chacun des quatre joueurs – appelés Est, Ouest, Nord et Sud – reçoit 13 des 52 cartes. Si Nord et Sud ont un total de 8 piques entre eux, quelle est la probabilité qu'Est ait 3 des 5 piques restants? ■

SOLUTION. La méthode de calcul la plus rapide est probablement ici de travailler avec l'ensemble fondamental réduit. Plus précisément, Nord et Sud ont un total de 8 piques parmi leurs 26 cartes. Il reste donc 26 cartes dont 5 piques exactement à répartir entre les mains d'Est et d'Ouest. Toutes les répartitions étant équiprobables, la probabilité conditionnelle qu'Est ait exactement 3 piques parmi ses 13 cartes sera donc

$$\frac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}} = .339$$

Exemple 3.4 L'entreprise pour laquelle travaille M. Jones organise un dîner pour ceux de ses employés ayant au moins un fils. Chacun de ces employés est invité à se présenter avec son aîné. On sait que Jones a deux enfants et il est invité au dîner. Quelle est alors la probabilité que ses enfants soient tous deux des garçons? On suppose que l'ensemble fondamental est $S = \{(g,g), (g,f), (f,g), (f,f)\}$ et que tous ces événements sont équiprobables. [(g,f) par exemple signifie que l'enfant le plus âgé est un garçon et que l'autre est une fille.]

SOLUTION. Le fait de savoir que Jones a été invité au dîner est équivalent à savoir qu'il a au moins un fils. Ainsi, en désignant par E l'événement «des deux enfants sont des garçons» et par F l'événement «au moins l'un des deux enfants est un garçon», la probabilité $P(E|F)$ cherchée est

$$\begin{aligned} P(E|F) &= \frac{P(EF)}{P(F)} \\ &= \frac{P(\{(g, g)\})}{P(\{(g, g), (g, f), (f, g)\})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \end{aligned}$$

Bien des gens se trompent en évaluant cette probabilité à $\frac{1}{2}$; ils admettent dans leur raisonnement que l'enfant non présent au dîner a autant de chances d'être un garçon qu'une fille. L'hypothèse que ces deux probabilités sont identiques est fautive: initialement en effet, il y avait quatre événements d'égale probabilité. Dès l'information «au moins l'un des enfants est un garçon» connue, on sait que l'événement final n'est pas (f,f) . Il nous reste ainsi trois événements équiprobables (g,g) , (f,g) , (g,f) . Ceci montre que l'événement «l'enfant de Jones non présent au dîner est une fille» est deux fois plus probable que son contraire. ■

3.2.3 Applications

En multipliant les deux membres de (3.1) par $P(F)$, nous obtenons

$$P(EF) = P(F)P(E|F) \quad (3.2)$$

Cette équation signifie en clair: la probabilité que E et F apparaissent à la fois est égale à la probabilité que F apparaisse multipliée par la probabilité conditionnelle de E si on sait que F est survenu. L'équation (3.2) est ainsi assez souvent utilisée pour calculer des probabilités d'intersections.

Exemple 3.5 Céline hésite entre suivre un cours de français et en suivre un de chimie. Bien qu'à vrai dire, elle préfère la chimie, elle estime à $\frac{1}{2}$ la probabilité d'obtenir la note A au cours de français contre $\frac{2}{3}$ seulement pour la chimie. Céline décide de baser sa décision sur le jet d'une pièce de monnaie équilibrée; quelle est la probabilité qu'elle obtienne la note A en chimie?

SOLUTION. En désignant par C l'événement «Céline suit le cours de chimie» et par A «elle obtient la note A dans le cours qu'elle choisit», la probabilité cherchée peut s'écrire $P(CA)$. On calcule cette dernière probabilité en utilisant (3.2) comme suit:

$$\begin{aligned} P(CA) &= P(C)P(A|C) \\ &= \left(\frac{1}{2}\right)\left(\frac{2}{3}\right) = \frac{1}{3} \end{aligned} \quad \blacksquare$$

Exemple 3.6 Une urne contient 8 boules rouges et 4 blanches. On tire sans remise deux boules de l'urne et admet qu'à chaque étape tous les tirages possibles sont équiprobables. Quelle est la probabilité que les deux boules tirées soient rouges?

SOLUTION. Appelons R_1 et R_2 respectivement les événements «la première boule tirée est rouge» et «la seconde est rouge». Si la première boule sélectionnée est rouge, il reste dès lors 7 boules rouges et 4 boules blanches. Donc $P(R_2|R_1) = \frac{7}{11}$; comme $P(R_1)$ vaut évidemment $\frac{8}{12}$, la probabilité demandée est

$$\begin{aligned} P(R_1R_2) &= P(R_1)P(R_2|R_1) \\ &= \left(\frac{8}{12}\right)\left(\frac{7}{11}\right) = \frac{14}{33} \end{aligned}$$

On pourrait évidemment calculer cette probabilité ainsi:

$$P(R_1 R_2) = \frac{\binom{8}{2}}{\binom{12}{2}}$$

■

3.2.4 La règle de multiplication

Une généralisation de l'équation (3.2), qui donne une expression pour la probabilité de l'intersection d'un nombre arbitraire d'événements, est parfois appelée la règle de multiplication.

La règle de multiplication :

$$P(E_1 E_2 E_3 \dots E_n) = P(E_1) P(E_2 | E_1) P(E_3 | E_1 E_2) \dots P(E_n | E_1 \dots E_{n-1})$$

Pour démontrer la règle de multiplication, on applique la définition de la probabilité conditionnelle au membre de droite. Cela donne

$$P(E_1) \frac{P(E_1 E_2)}{P(E_1)} \frac{P(E_1 E_2 E_3)}{P(E_1 E_2)} \dots \frac{P(E_1 E_2 \dots E_n)}{P(E_1 E_2 \dots E_{n-1})} = P(E_1 E_2 \dots E_n)$$

Nous allons à présent employer la règle de multiplication pour résoudre selon une seconde approche l'exemple 2.9 (question b) du chapitre 2.

Exemple 3.7 Un jeu ordinaire de 52 cartes est divisé aléatoirement en 4 piles de 13 cartes chacune. Calculer la probabilité que chaque pile contienne exactement 1 as.

SOLUTION. Définissons les événements E_i , $i = 1, 2, 3, 4$, de la manière suivante:

$E_1 = \{\text{L'as de pique est dans une des piles}\}$

$E_2 = \{\text{L'as de pique et l'as de cœur sont dans des piles différentes}\}$

$E_3 = \{\text{Les as de pique, cœur et carreau sont dans des piles différentes}\}$

$E_4 = \{\text{Les 4 as sont dans des piles différentes}\}$

La probabilité cherchée est $P(E_1 E_2 E_3 E_4)$ et, par la règle de multiplication,

$$P(E_1 E_2 E_3 E_4) = P(E_1) P(E_2 | E_1) P(E_3 | E_1 E_2) P(E_4 | E_1 E_2 E_3)$$

Or $P(E_1) = 1$ puisque E_1 est l'ensemble fondamental S .

$$P(E_2 | E_1) = \frac{39}{51}$$

puisque la pile contenant l'as de pique recevra 12 des 51 cartes restantes.

$$P(E_3 | E_1 E_2) = \frac{26}{50}$$

puisque les piles contenant les as de pique et cœur recevront 24 des 50 cartes restantes; et finalement

$$P(E_4 | E_1 E_2 E_3) = \frac{13}{49}$$

De là, nous obtenons la probabilité que chaque pile possède exactement un as:

$$P(E_1 E_2 E_3 E_4) = \frac{39 \cdot 26 \cdot 13}{51 \cdot 50 \cdot 49} \approx .105$$

Ceci signifie qu'il y a environ 10.5% de chances que chaque pile contienne un as (le problème 20 utilise la règle de multiplication d'une autre façon pour résoudre ce problème).

REMARQUE. Notre définition de $P(E | F)$ correspond à l'interprétation des probabilités comme étant la limite des fréquences relatives. Pour le voir, supposons qu'on réalise n répétitions de l'expérience, avec n grand. Nous prétendons que si on considère uniquement les expériences où F se produit, alors $P(E | F)$ sera égale à la limite de la proportion de celles où E se produit également. Pour vérifier ceci, on remarque que puisque $P(F)$ est la limite de la proportion des expériences où F se produit, cela entraîne que lors de n répétitions de l'expérience, F se produira environ $nP(F)$ fois. De la même manière, E et F se produiront toutes les deux approximativement lors de $nP(EF)$ expériences. Donc, dans les $nP(F)$ expériences environ où F se produit, la proportion de celles où E se produit aussi est à peu près égale à

$$\frac{nP(EF)}{nP(F)} = \frac{P(EF)}{P(F)}$$

Comme cette approximation s'améliore lorsque n croît, nous voyons que notre définition de $P(E | F)$ est appropriée.

3.3 FORMULE DE BAYES

3.3.1 Présentation intuitive de la formule des probabilités totales

Soient E et F deux événements quelconques. Nous pouvons écrire E sous la forme $E = EF \cup EF^c$, car tout élément de E doit se trouver soit dans E et F à la fois, soit dans E mais pas dans F (voir fig. 3.1). Comme évidemment EF et EF^c s'excluent mutuellement, on peut écrire en vertu de l'axiome 2.3:

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)[1 - P(F)]. \end{aligned} \quad (3.3)$$

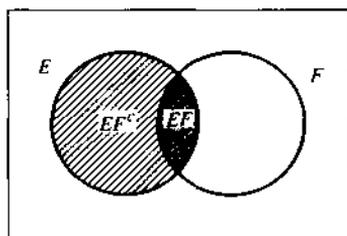


Fig. 3.1

EF : zone en gris

EF^c : zone hachurée

$$E = EF \cup EF^c$$

L'équation (3.3), appelée *formule des probabilités totales*, peut être interprétée de la façon suivante: la probabilité de l'événement E est une moyenne pondérée de la probabilité conditionnelle de E lorsque F est apparu et de la probabilité conditionnelle du même E lorsque F n'est pas apparu, les poids étant les probabilités des événements conditionnants.

Cette formule est extrêmement utile puisqu'elle nous permet dans bien des cas de déterminer la probabilité d'un événement en commençant par le conditionner selon l'apparition ou non d'un autre événement. En d'autres mots, il existe de nombreuses situations où il est difficile de calculer directement la probabilité d'un événement mais où il est par contre possible de la calculer connaissant ses probabilités conditionnelles si certains événements sont réalisés. Quelques exemples illustrent cette démarche.

Exemple 3.8 (1^{re} partie) Une compagnie d'assurance estime que les gens peuvent être répartis en deux classes: ceux qui sont enclins aux accidents et ceux qui ne le sont pas. Ses statistiques montrent qu'un individu enclin aux accidents a une probabilité de 0,4 d'en avoir un dans l'espace d'un an; cette probabilité tombe à 0,2 pour les gens à risque modéré. On suppose que 30% de la population appartient à la classe à haut risque. Quelle est alors la probabilité qu'un nouvel assuré soit victime d'un accident durant l'année qui suit la signature de son contrat?

SOLUTION. Nous obtiendrons la probabilité de l'événement cité en le conditionnant selon que le signataire de la police est ou n'est pas enclin aux accidents. On note par A_1 l'événement «le signataire aura un accident dans l'année qui suit l'établissement du contrat» et par A «le signataire est enclin aux accidents». La probabilité $P(A_1)$ voulue est alors donnée par

$$\begin{aligned} P(A_1) &= P(A_1|A)P(A) + P(A_1|A^c)P(A^c) \\ &= (.4)(.3) + (.2)(.7) = .26 \end{aligned}$$

■

Exemple 3.8 (2^e partie) Un nouveau signataire a un accident dans l'année qui suit la signature de son contrat. Quelle est la probabilité qu'il fasse partie de la classe à haut risque?

SOLUTION. Cette probabilité est $P(A|A_1)$, donnée par

$$\begin{aligned} P(A|A_1) &= \frac{P(AA_1)}{P(A_1)} \\ &= \frac{P(A)P(A_1|A)}{P(A_1)} \\ &= \frac{(.3)(.4)}{.26} = \frac{6}{13} \end{aligned}$$

■

3.3.2 Introduction à la formule de Bayes

Formellement, la *formule de Bayes* est simplement dérivée de (3.1) et (3.3). L'expression générale de cette formule est donnée plus loin (formule (3.6)). Les exemples suivants vont néanmoins servir à donner une idée intuitive du champ d'application de cette formule importante.

Exemple 3.9 Un étudiant répond à une question à choix multiple. De deux choses l'une: soit il connaît la réponse, soit il la devine. Soit p la probabilité que l'étudiant connaisse la réponse et donc $1 - p$ celle qu'il la devine. On admet que l'étudiant qui devine répondra correctement avec probabilité $1/m$ où m est le nombre de réponses possibles. Quelle est la probabilité conditionnelle qu'un étudiant connaisse la réponse à une question s'il y a répondu correctement?

SOLUTION. Soient C et K respectivement les événements «l'étudiant répond correctement à la question» et «il connaît vraiment la réponse». Alors

$$\begin{aligned} P(K|C) &= \frac{P(KC)}{P(C)} \\ &= \frac{P(C|K)P(K)}{P(C|K)P(K) + P(C|K^c)P(K^c)} \\ &= \frac{p}{p + (1/m)(1 - p)} \\ &= \frac{mp}{1 + (m - 1)p} \end{aligned}$$

En prenant par exemple $m = 5$ et $p = 1/2$, la probabilité qu'un(e) étudiant(e) connaisse la réponse à une question sachant qu'il ou elle a répondu correctement sera ainsi $\frac{5}{6}$.

■

Exemple 3.10 Un laboratoire d'analyses du sang assure avec une fiabilité de 95 % la détection d'une certaine maladie lorsqu'elle est effectivement présente. Cependant, le test indique aussi un résultat faussement «positif» pour 1 % des personnes réellement saines à qui on l'applique (c'est-à-dire qu'une personne saine testée sera déclarée

malade une fois sur cent). Si 0,5% de la population porte effectivement la maladie, quelle est la probabilité qu'une personne soit vraiment malade lorsqu'on la déclare telle sur la base du test?

SOLUTION. Soit D l'événement «la personne soumise au test est porteuse de la maladie» et E l'événement «le résultat du test est positif». La probabilité $P(D|E)$ voulue est donnée par

$$\begin{aligned} P(D|E) &= \frac{P(DE)}{P(E)} \\ &= \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)P(D^c)} \\ &= \frac{(.95)(.005)}{(.95)(.005) + (.01)(.995)} \\ &= \frac{95}{294} \approx .323 \end{aligned}$$

Ainsi 32% seulement des personnes dont les résultats au test sont positifs ont vraiment la maladie. Comme beaucoup d'étudiants sont surpris de ce résultat (ils s'attendent souvent à une valeur beaucoup plus élevée puisque le test sanguin semble être bon), il n'est pas inutile de donner un autre argument moins rigoureux que le précédent mais plus parlant.

Puisque 0,5% de la population est réellement affectée par cette maladie, sur 200 personnes testées l en moyenne l'aura. Le test décèlera ce cas avec une probabilité de 0,95. En moyenne donc, sur 200 personnes testées, on détectera correctement 0,95 cas. D'autre part, parmi les 199 personnes saines le test va à tort détecter (199)(0,01) cas de maladie. Si l'on résume, à 0,95 cas de maladie correctement détectés s'ajoutent en moyenne 1,99 cas faussement positifs (cas de personnes saines en réalité). Dès lors, la proportion de résultats corrects quand le test est positif n'est que de

$$\frac{.95}{.95 + (199)(.01)} = \frac{95}{294} \approx .323 \quad \blacksquare$$

L'équation (3.3) est également utile lorsqu'on cherche à réévaluer des probabilités à la lumière d'informations supplémentaires. On peut illustrer cela au moyen des exemples suivants.

Exemple 3.11 Considérons un médecin placé devant le dilemme suivant: «Lorsque je suis certain, à au moins 80%, que mon patient est affecté d'une maladie bien précise, je recommande toujours une intervention chirurgicale. Tandis que si j'en suis moins certain, je prescris des tests complémentaires qui peuvent être chers et parfois pénibles. Dans le cas de Jones, j'étais initialement certain à 60% seulement qu'il souffrait de cette maladie, aussi ai-je prescrit un test A qui donne toujours un résultat positif lorsque le patient est vraiment malade et presque jamais dans le cas contraire. Le résultat de ce test étant positif, j'étais prêt à recommander une opération quand Jones m'informa qu'il était diabétique, ce qu'il n'avait pas dit jusque-là. Cette indication complique le diagnostic: bien que cela ne change en rien mon estimation originale de

maladie avec 60% de risques, cela affecte par contre l'interprétation du résultat du test A. Ce test en effet, alors qu'il ne donne jamais de résultat positif si le patient est sain, conduit malheureusement à un tel résultat – erroné – chez 30% des *diabétiques* ne souffrant pas de la maladie. A partir de là, que faire? Encore des tests ou une opération immédiate?»

SOLUTION. En vue de décider si oui ou non, il faut recourir à une opération chirurgicale, le médecin doit premièrement calculer la nouvelle probabilité que Jones soit malade dès lors qu'on sait le test A positif. Soit D l'événement «Jones a cette maladie» et E «le résultat du test est positif». La probabilité conditionnelle $P(D|E)$ cherchée est calculable ainsi:

$$\begin{aligned} P(D|E) &= \frac{P(DE)}{P(E)} \\ &= \frac{P(D)P(E|D)}{P(E|D)P(D) + P(E|D^c)P(D^c)} \\ &= \frac{(.6)1}{1(.6) + (.3)(.4)} \\ &= .833 \end{aligned}$$

Notons que nous avons calculé la probabilité d'avoir un résultat de test positif en conditionnant par les événements que Jones a ou n'a pas la maladie et en utilisant alors l'information que, Jones étant diabétique, sa probabilité conditionnelle $P(E|D^c)$ de donner un résultat positif s'il n'est pas malade est 0,3. Ainsi, comme le médecin estime à présent à plus de 80% les risques pour Jones d'être atteint, il recommandera d'opérer. ■

Exemple 3.12 L'inspecteur chargé d'une enquête criminelle est à un certain stade convaincu à 60% de la culpabilité d'un suspect donné. On découvre alors une *nouvelle* pièce à conviction permettant d'affirmer que le criminel cherché possédait un certain attribut (il était par exemple gaucher, ou alors chauve, ou aux cheveux bruns). Or 20% de la population possède ce même attribut. Comment l'inspecteur doit-il réapprécier la culpabilité du suspect s'il se trouve que celui-ci a ce fameux attribut?

SOLUTION. Désignons par G l'événement «le suspect est coupable» et par C «il possède le même attribut que le criminel». Nous aurons

$$\begin{aligned} P(G|C) &= \frac{P(GC)}{P(C)} \\ &= \frac{P(C|G)P(G)}{P(C|G)P(G) + P(C|G^c)P(G^c)} \\ &= \frac{1(.6)}{1(.6) + (.2)(.4)} \\ &= .882 \end{aligned}$$

où nous avons supposé que la probabilité pour le suspect d'avoir l'attribut s'il est en fait innocent est 0,2, la proportion normale dans la population. ■

Exemple 3.13 Lors du championnat du monde de bridge qui se déroula en mai 1965 à Buenos Aires, la fameuse équipe britannique Terrence Reese-Boris Schapiro fut accusée de tricher au moyen d'un système de signaux de doigts qui pouvait indiquer le nombre de cœurs détenus par les joueurs. Reese et Schapiro nièrent et finalement une expertise fut menée par la British Bridge League. L'expertise fut organisée sous la forme d'un procès avec partie plaignante et défense, chacune d'elles ayant le pouvoir de faire comparaître tout témoin et de le soumettre à un interrogatoire contradictoire. Au cours de ces débats, le représentant de la partie plaignante examina certaines mains jouées par Reese et Schapiro puis affirma que leur manière de mener le jeu pour ces mains corroborait l'hypothèse qu'ils étaient coupables d'avoir une information malhonnête sur la répartition des cœurs. Au même moment, le représentant de la défense releva que leur manière de jouer était en parfait accord avec leur tactique habituelle. La partie plaignante soutint alors qu'aussi longtemps que leur jeu resterait en ligne avec l'hypothèse de fraude, il faudrait retenir ce fait à l'appui de celle-ci. Que pensez-vous de l'attitude de la partie plaignante?

SOLUTION. Le problème est fondamentalement ici de déterminer l'influence d'une donnée supplémentaire (dans l'exemple traité, la tactique de jeu) sur la probabilité d'une hypothèse donnée. Si nous désignons par H cette hypothèse (telle que la fraude dans le cas de Reese et Schapiro) et par E la donnée supplémentaire, alors

$$\begin{aligned} P(H|E) &= \frac{P(HE)}{P(E)} \\ &= \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|H^c)[1 - P(H)]} \end{aligned} \quad (3.4)$$

où $P(H)$ est notre évaluation de la vraisemblance de l'hypothèse avant que la nouvelle donnée ne soit connue. Celle-ci sera à l'appui de l'hypothèse si elle la rend plus vraisemblable, c'est-à-dire si $P(H|E) \geq P(H)$. D'après l'équation (3.4) ce sera le cas si

$$P(E|H) \geq P(E|H)P(H) + P(E|H^c)[1 - P(H)]$$

ou, de façon équivalente, si

$$P(E|H) \geq P(E|H^c)$$

En d'autres mots, une donnée nouvelle ne peut être retenue en faveur d'une hypothèse donnée que si elle est plus vraisemblable en supposant l'hypothèse vraie qu'en la supposant fautive. En fait, la nouvelle probabilité dépend de l'ancienne et du rapport de ces deux probabilités conditionnelles, puisque (3.4) donne

$$P(H|E) = \frac{P(H)}{P(H) + [1 - P(H)] \frac{P(E|H^c)}{P(E|H)}}$$

De ce fait, dans le problème considéré, la tactique de jeu ne peut être retenue en faveur de l'hypothèse de fraude que lorsqu'une telle tactique est plus probable si l'équipe anglaise triche que dans le cas contraire. Comme la partie plaignante n'a jamais cherché à le prétendre, son opinion que les faits sont de nature à établir la fraude n'est pas valable. ■

3.3.3 Formule des probabilités totales généralisée

L'équation (3.3) peut être généralisée de la manière suivante: supposons que F_1, F_2, \dots, F_n soient des événements s'excluant mutuellement et tels que

$$\bigcup_{i=1}^n F_i = S$$

Cela revient à dire en d'autres termes qu'exactly un des événements F_1, F_2, \dots, F_n se produira. En écrivant

$$E = \bigcup_{i=1}^n EF_i$$

et en utilisant le fait que les événements $EF_i, i = 1, \dots, n$ s'excluent mutuellement, on obtient:

$$\begin{aligned} P(E) &= \sum_{i=1}^n P(EF_i) \\ &= \sum_{i=1}^n P(E|F_i)P(F_i) \end{aligned} \quad (3.5)$$

L'équation (3.5) montre ainsi qu'étant donné un jeu d'événements F_1, F_2, \dots, F_n desquels un et un seul surviendra, on peut calculer $P(E)$ en commençant par conditionner selon les F_i . Ou encore, l'équation (3.5) établit que $P(E)$ est une moyenne pondérée des $P(E|F_i)$, les poids valant la probabilité des événements sur lesquels on conditionne.

3.3.4 Formule de Bayes généralisée

Supposons maintenant que E s'est réalisé et que nous cherchions à déterminer la probabilité que l'un des F_j se soit aussi réalisé. On déduit de l'équation (3.5) le théorème suivant:

Théorème 3.1

$$\begin{aligned} P(F_j|E) &= \frac{P(EF_j)}{P(E)} \\ &= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)} \end{aligned} \quad (3.6)$$

L'équation (3.6) est appelée formule de Bayes, du nom du philosophe anglais Thomas Bayes. Si nous traitons les événements F_j comme les *hypothèses* possibles sur une question donnée, la formule de Bayes joue un rôle utile en nous montrant comment les opinions a priori sur ces hypothèses [à savoir, $P(F_j)$] doivent être modifiées à la lumière du résultat de l'expérience.

Exemple 3.14 Un avion est porté disparu. On pense que l'accident a pu arriver aussi bien dans n'importe laquelle de trois régions données. Notons par $1 - \alpha_i$ la probabilité qu'on découvre l'avion dans la région i s'il y est effectivement. Les valeurs α_i représentent donc la probabilité de manquer l'avion lors des recherches. On peut l'attribuer à diverses causes d'ordre géographique ou à la végétation propre à la région. Quelle est la probabilité que l'avion se trouve dans la i -ème région si les recherches dans la région 1 n'ont rien donné, $i = 1, 2, 3$?

SOLUTION. Soient R_i , $i = 1, 2, 3$ les événements «l'avion est tombé dans la région i ». Soit aussi E l'événement «les recherches dans la région 1 sont restées infructueuses». On tire de la formule de Bayes, pour $i = 1$:

$$\begin{aligned} P(R_1|E) &= \frac{P(ER_1)}{P(E)} \\ &= \frac{P(E|R_1)P(R_1)}{\sum_{i=1}^3 P(E|R_i)P(R_i)} \\ &= \frac{(\alpha_1)^{\frac{1}{3}}}{(\alpha_1)^{\frac{1}{3}} + (1)^{\frac{1}{3}} + (1)^{\frac{1}{3}}} \\ &= \frac{\alpha_1}{\alpha_1 + 2} \end{aligned}$$

Pour $j = 2, 3$

$$\begin{aligned} P(R_j|E) &= \frac{P(E|R_j)P(R_j)}{P(E)} \\ &= \frac{(1)^{\frac{1}{3}}}{(\alpha_1)^{\frac{1}{3}} + \frac{1}{3} + \frac{1}{3}} \\ &= \frac{1}{\alpha_1 + 2}, \quad j = 2, 3 \end{aligned}$$

On remarquera que la probabilité a posteriori (c'est-à-dire conditionnelle) que l'avion se trouve dans la région j une fois que l'on sait que la fouille de la région 1 n'a rien donné est plus grande pour $j = 2$ ou 3 que la probabilité a priori, tandis qu'elle l'est moins pour la région 1. On s'attendait à ce résultat puisque la fouille infructueuse dans la région 1 tend à faire diminuer la probabilité que l'avion s'y trouve, augmentant par-là même les chances qu'il soit ailleurs. On remarque aussi que la probabilité conditionnelle qu'il soit dans la région 1 si la recherche n'a rien donné est une fonction

croissante de la probabilité α_j de l'y manquer, ce qui est généralement intuitivement prévisible: plus α_j est grande, plus il est raisonnable d'attribuer l'échec des recherches à la «malchance» plutôt qu'au fait que l'avion n'y est pas. Inversement, $P(R_j|E)$, $j \neq 1$ est une fonction décroissante de α_j . ■

L'exemple suivant a souvent été utilisé par de peu scrupuleux étudiants pour abuser des camarades plus naïfs.

Exemple 3.15 On considère 3 cartes à jouer de même forme. Cependant, les deux faces de la première carte ont été colorées en noir, les deux faces de la deuxième carte en rouge tandis que la troisième porte une face noire et l'autre rouge. On mélange les trois cartes au fond d'un chapeau puis une carte tirée au hasard en est extraite et placée au sol. Si la face apparente est rouge, quelle est la probabilité que l'autre soit noire?

SOLUTION. Soient RR , NN et RN respectivement les événements, «la carte choisie est entièrement rouge», «entièrement noire» et «bicolore». Soit encore R l'événement, «la face apparente de la carte tirée est rouge». On aura

$$\begin{aligned} P(RN|R) &= \frac{P(RN \cap R)}{P(R)} \\ &= \frac{P(R|RN)P(RN)}{P(R|RR)P(RR) + P(R|RN)P(RN) + P(R|NN)P(NN)} \\ &= \frac{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)}{\left(1\right)\left(\frac{1}{3}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) + 0\left(\frac{1}{3}\right)} = \frac{1}{3} \end{aligned}$$

Ainsi, la réponse est $\frac{1}{3}$. Certaines personnes pourtant estiment la réponse à $\frac{1}{2}$, pensant qu'à partir du moment où le côté rouge apparaît il reste 2 situations équiprobables: soit la carte tirée est entièrement rouge, soit elle est bicolore. Leur erreur est ici d'admettre cette hypothèse d'équiprobabilité. Il faut se souvenir en effet que chaque carte possède deux faces que l'on distinguera pour la commodité. Il y a alors 6 événements élémentaires équiprobables pour cette expérience, que l'on appellera R_1 , R_2 , N_1 , N_2 , R_3 , N_3 . L'événement R_1 sera réalisé si c'est la première face de la carte unicolore rouge qui est apparente. R_2 le sera si c'est la seconde face de la même carte qui est visible. R_3 si c'est le côté rouge de la carte bicolore et ainsi de suite. La face cachée de la carte tirée sera noire seulement si c'est R_3 qui a lieu. La probabilité cherchée est donc la probabilité conditionnelle de R_3 sachant que R_1 , R_2 ou R_3 a eu lieu, laquelle est manifestement $\frac{1}{3}$. ■

Exemple 3.16 Les assistants sociaux travaillant pour une clinique psychiatrique sont si occupés qu'en moyenne seuls 60% des patients prospectifs téléphonant pour la première fois obtiendront une communication avec l'un de ces assistants. On demande aux autres de laisser leur numéro de téléphone. Trois fois sur quatre un assistant trouve le temps de rappeler encore le jour même, autrement le rappel a lieu le

lendemain. L'expérience a montré que dans cette clinique, la probabilité que le patient prospectif demande une consultation est 0,8 s'il a pu parler immédiatement à un assistant, tandis qu'elle tombe à 0,6 et 0,4 respectivement s'il y a eu rappel du patient le jour même ou le lendemain. a) Quel pourcentage parmi les gens qui appellent demanderont-ils une consultation ? b) Quel pourcentage des gens en consultation n'ont pas eu à attendre qu'on les rappelle ?

SOLUTION. On définit les événements suivants:

C : «le patient demande une consultation»;

I : «le patient obtient immédiatement un entretien téléphonique avec un assistant social»;

M : «le patient est rappelé plus tard, le jour même»;

L : «le patient est rappelé plus tard, le lendemain».

Alors,

$$\begin{aligned} P(C) &= P(C|I)P(I) + P(C|M)P(M) + P(C|L)P(L) \\ &= (.8)(.6) + (.6)(.4)(.75) + (.4)(.4)(.25) \\ &= .70 \end{aligned}$$

où nous avons utilisé le fait que $P(M) = (0,4)(0,75)$ et que $P(L) = (0,4)(0,25)$. Ceci répond à la première question. Quant à la seconde, on remarquera que

$$\begin{aligned} P(I|C) &= \frac{P(C|I)P(I)}{P(C)} \\ &= \frac{(.8)(.6)}{.7} \\ &\approx .686 \end{aligned}$$

Ainsi, 69% environ des patients en consultation ont obtenu immédiatement un entretien avec un assistant social. ■

3.4 ÉVÉNEMENTS INDÉPENDANTS

3.4.1 Indépendance de deux événements

Les exemples vus dans ce chapitre jusqu'à présent ont montré que la probabilité conditionnelle de E sachant que F est réalisé n'est en général pas égale à $P(E)$, la probabilité non conditionnelle de E . En d'autres termes, le fait de savoir que F est survenu influence la probabilité de E . Dans les cas où $P(E|F)$ est bien égal à $P(E)$, l'événement E est dit indépendant de F . Plus précisément, E est indépendant de F si le fait de savoir que F est survenu ne change pas la probabilité de E .

Du fait que $P(E|F) = P(EF)/P(F)$, on voit que l'indépendance de E et F équivaut à

$$P(EF) = P(E)P(F) \quad (3.7)$$

Comme cette équation est symétrique en E et F , il en résulte que lorsque E est indépendant de F , F l'est aussi de E . On débouche ainsi sur la définition suivante:

Deux événements E et F sont dits *indépendants* si l'équation (3.7) est vérifiée. Deux événements sont *dépendants* s'ils ne sont pas indépendants.

3.4.2 Exemples d'indépendance de deux événements

Exemple 3.17 On tire au hasard une carte d'un paquet de 52 cartes à jouer ordinaires. Désignons par E l'événement «la carte tirée est un as» et par F «elle est un pique». Alors E et F sont indépendants. En effet, $P(EF) = \frac{1}{52}$ d'une part, alors que d'autre part $P(E) = \frac{4}{52}$ et $P(F) = \frac{13}{52}$. ■

Exemple 3.18 On jette deux pièces et suppose que les 4 résultats possibles sont équiprobables. On désigne par A «la première pièce montre pile» et par B «la seconde montre face». A et B sont indépendants puisque $P(AB) = P(\{(P,F)\}) = \frac{1}{4}$ d'une part, et $P(A) = P(\{(P,P),(P,F)\}) = \frac{1}{2}$, $P(B) = P(\{(P,F),(F,F)\}) = \frac{1}{2}$ d'autre part. ■

Exemple 3.19 On jette deux dés équilibrés. E_1 est l'événement «la somme des dés est 6» et F désigne «le premier dé donne 4». Dans ce cas

$$P(E_1F) = P(\{(4, 2)\}) = \frac{1}{36}$$

alors que

$$P(E_1)P(F) = \left(\frac{5}{36}\right)\left(\frac{1}{6}\right) = \frac{5}{216}$$

E_1 et F ne sont donc pas indépendants. Intuitivement la raison en est claire: si l'on espère obtenir une somme de 6 sur les deux dés, l'apparition d'un 4 sur le premier dé, ou d'un 1, d'un 2, 3 ou d'un 5, laisse espérer d'atteindre ce résultat. Par contre si le premier dé donne déjà 6, il n'y a plus aucune chance d'obtenir 6 au total. En d'autres termes, la probabilité d'obtenir 6 sur deux dés dépend clairement du résultat apparu sur le premier dé. E_1 et F ne peuvent donc être indépendants.

Désignons maintenant par E_2 l'événement «la somme des dés est 7». E_2 est-il indépendant de F ? La réponse est oui, cette fois-ci, car

$$P(E_2F) = P(\{(4, 3)\}) = \frac{1}{36}$$

alors que

$$P(E_2)P(F) = \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{1}{36}$$

Nous laissons au lecteur le soin de découvrir un argument intuitif justifiant l'indépendance entre «la somme des dés est 7» et le résultat donné par le premier dé. ■

Exemple 3.20 Soit E l'événement «le prochain président des U.S.A. sera un Républicain» et soit F «il y aura un tremblement de terre important d'ici un an». La plupart des personnes accepteraient d'admettre qu'ici E et F sont indépendants. Par contre il n'en serait pas nécessairement de même concernant E et G , où G est «un conflit majeur éclatera dans les deux ans suivant l'élection de ce président». ■

3.4.3 Indépendance par rapport au complémentaire

Nous allons maintenant montrer que si E est indépendant de F , il l'est aussi de F^c .

Théorème 3.2

Si E et F sont indépendants, E et F^c le sont aussi.

DÉMONSTRATION. D'une part $E = EF \cup EF^c$, d'autre part EF et EF^c sont mutuellement exclusifs. On peut donc écrire

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E)P(F) + P(EF^c) \end{aligned}$$

ou, de façon équivalente,

$$\begin{aligned} P(EF^c) &= P(E)[1 - P(F)] \\ &= P(E)P(F^c) \end{aligned}$$

ce qui établit le résultat. ■

Ainsi, lorsque E est indépendant de F , la probabilité que E survienne n'est influencée ni par l'information que F est réalisé ni par celle que F ne l'est pas.

3.4.4 Indépendance entre plusieurs événements

Supposons maintenant que E soit indépendant de F et aussi de G . Est-ce que E le sera de FG ? Aussi surprenant que cela puisse paraître, la réponse est non, comme le montre l'exemple suivant.

Exemple 3.21 On jette deux dés équilibrés. Soient E l'événement «la somme est 7», F l'événement «le premier dé montre 4» et G «le second dé donne 3». On a vu dans l'exemple 3.18 que E et F sont indépendants, le même raisonnement permettant d'affirmer que E et G le sont. Cependant E n'est manifestement pas indépendant de FG puisque $P(E|FG) = 1$. ■

3.4.5 Indépendance totale de trois événements

On comprend grâce à cet exemple qu'une bonne définition de l'indépendance de trois événements ne peut pas se limiter à exiger que les événements soient indépendants par paires dans les $\binom{3}{2}$ combinaisons possibles. On est ainsi mené à la définition suivante.

Trois événements E , F et G sont dits totalement indépendants si

$$\begin{aligned} P(EFG) &= P(E)P(F)P(G) \\ P(EF) &= P(E)P(F) \\ P(EG) &= P(E)P(G) \\ P(FG) &= P(F)P(G) \end{aligned}$$

3.4.6 Conséquence

Il faut noter que lorsque E , F et G sont totalement indépendants, E sera indépendant de tout événement formé à partir de F et G . On peut le montrer pour $F \cup G$ par exemple:

$$\begin{aligned} P[E(F \cup G)] &= P(EF \cup EG) \\ &= P(EF) + P(EG) - P(EFG) \\ &= P(E)P(F) + P(E)P(G) - P(E)P(FG) \\ &= P(E)[P(F) + P(G) - P(FG)] \\ &= P(E)P(F \cup G) \end{aligned}$$

3.4.7 Indépendance totale de n événements

Il est évidemment possible d'étendre la définition d'indépendance totale à plus de trois événements: un ensemble d'événements E_1, E_2, \dots, E_n est dit totalement indépendant si pour tout sous-ensemble $E_{1'}, E_{2'}, \dots, E_{r'}$, $r \leq n$.

$$P(E_{1'} E_{2'} \cdots E_{r'}) = P(E_{1'}) P(E_{2'}) \cdots P(E_{r'})$$

3.4.8 Indépendance totale d'une infinité d'événements

Enfin nous dirons que par définition un ensemble infini d'événements est totalement indépendant si tout sous-ensemble fini d'entre eux est totalement indépendant.

3.4.9 Epreuves indépendantes

Il arrive parfois que l'expérience étudiée consiste à effectuer une suite d'expériences partielles. Si par exemple l'expérience de base consiste à répéter le jet d'une pièce, on peut considérer chaque jet comme l'une de ces expériences partielles. Comme dans bien des cas il est raisonnable d'admettre que l'issue de tout groupe d'expériences partielles n'a aucun effet sur celle des autres, on considérera que ces expériences partielles sont totalement indépendantes. De ceci on peut donner une formulation plus rigoureuse: considérons tout jeu d'événements E_1, E_2, \dots, E_n tels que E_i soit complètement déterminé quant à sa réalisation par le résultat de la i -ème expérience partielle. Si un tel ensemble est nécessairement totalement indépendant, alors les expériences partielles sont dites ensemble totalement indépendant d'événements.

Si toutes ces expériences partielles sont identiques – c'est-à-dire si elles ont toutes le même (sous-)ensemble fondamental et sont toutes affectées de la même fonction de probabilité –, alors ces expériences partielles sont appelées *épreuves*.

3.4.10 Exemples d'épreuves indépendantes

Exemple 3.22 On réalise une séquence infinie d'épreuves indépendantes. Chaque épreuve donne soit un succès, soit un échec avec probabilités p et $1 - p$ respectivement.

Quelle est la probabilité pour:

- qu'il survienne au moins un succès parmi les n premières épreuves;
- qu'il survienne exactement k succès parmi les n premières épreuves;
- que toutes les épreuves donnent des succès.

SOLUTION. Dans le but de déterminer plus facilement la probabilité d'avoir au moins 1 succès parmi les n premières épreuves, on préférera calculer la probabilité de l'événement complémentaire (aucun succès lors des n premières épreuves). Notons E_i l'événement «la i -ème épreuve donne un échec». En utilisant la propriété d'indépendance totale, la probabilité de n'obtenir aucun succès est

$$P(E_1 E_2 \cdots E_n) = P(E_1)P(E_2) \cdots P(E_n) = (1 - p)^n$$

Ainsi, la réponse au a) est $1 - (1 - p)^n$.

Pour obtenir b), considérons une séquence de n événements qui comprenne k succès et $n - k$ échecs dans un ordre bien précis. Cette séquence apparaîtra – en supposant l'indépendance totale des épreuves – avec une probabilité $p^k(1 - p)^{n-k}$. Comme il y a $\binom{n}{k}$ de ces séquences (il y a $n!/k!(n - k)!$ combinaisons de k succès et $n - k$ échecs), le deuxième de nos résultats est

$$P\{\text{exactement } k \text{ succès}\} = \binom{n}{k} p^k (1 - p)^{n-k}$$

Quant à c), par analogie avec a), on peut dans un premier temps écrire que la probabilité de n'avoir que des succès lors des n premières épreuves sera

$$P(E_1^c E_2^c \cdots E_n^c) = p^n$$

Aussi peut-on écrire en utilisant la propriété de continuité des probabilités (section 2.6) que la probabilité $P(\bigcap_1^\infty E_i^c)$ cherchée est

$$\begin{aligned} P\left(\bigcap_{i=1}^{\infty} E_i^c\right) &= P\left(\lim_{n \rightarrow \infty} \bigcap_{i=1}^n E_i^c\right) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcap_{i=1}^n E_i^c\right) \\ &= \lim_{n \rightarrow \infty} p^n \\ &= \begin{cases} 0 & \text{si } p < 1 \\ 1 & \text{si } p = 1 \end{cases} \quad \blacksquare \end{aligned}$$

Exemple 3.23 Un système comprenant n composants est appelé système en parallèle s'il fonctionne dès qu'au moins l'un de ces composants fonctionne (voir fig. 3.2). Dans le cas d'un tel système, si son i -ème composant fonctionne indépendamment de tous les autres et avec une probabilité p_i , $i = 1, 2, \dots, n$, quelle est la probabilité de son fonctionnement?

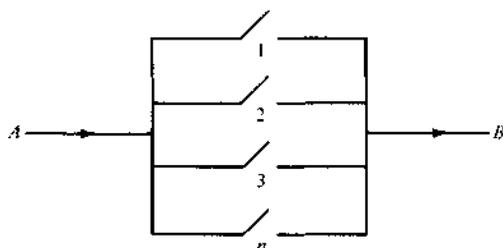


Figure 3.2 Système en parallèle fonctionnant dès que le courant peut passer de A à B

SOLUTION. Soit A_i l'événement «le composant i fonctionne». Alors

$$\begin{aligned}
 P(\text{«le système fonctionne»}) &= 1 - P(\text{«le système ne fonctionne pas»}) \\
 &= 1 - P(\text{«aucun composant ne fonctionne»}) \\
 &= 1 - P\left(\bigcap_i A_i^c\right) \\
 &= 1 - \prod_{i=1}^n (1 - p_i) \quad \text{grâce à l'indépendance}
 \end{aligned}$$

■

3.4.11 Un exemple important résolu

Exemple 3.24 Une séquence d'épreuves indépendantes consiste à jeter plusieurs fois une paire de dés réguliers. On appelle résultat la somme des chiffres apparents. Quelle est la probabilité qu'on voie sortir un résultat valant 5 avant qu'un 7 n'apparaisse?

SOLUTION. Désignons par E_n l'événement que durant les $n - 1$ premières épreuves il n'apparaisse ni 5 ni 7, et qu'à la n -ème épreuve un 5 sorte. On cherche la probabilité

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$$

Par ailleurs, $P(5 \text{ sort lors d'une épreuve quelconque}) = \frac{4}{36}$ et $P(7 \text{ sort lors d'une épreuve quelconque}) = \frac{6}{36}$. Du fait de l'indépendance des épreuves, on obtient donc

$$P(E_n) = \left(1 - \frac{10}{36}\right)^{n-1} \frac{4}{36}$$

et donc

$$\begin{aligned}
 P\left(\bigcup_{n=1}^{\infty} E_n\right) &= \left(\frac{1}{9}\right) \sum_{n=1}^{\infty} \left(\frac{13}{18}\right)^{n-1} \\
 &= \frac{1}{9} \frac{1}{1 - \frac{13}{18}} \\
 &= \frac{2}{5}
 \end{aligned}$$

On aurait aussi pu obtenir ce résultat en passant par des probabilités conditionnelles. Si E désigne l'événement étudié («5 apparaît avant 7»), $P(E)$ peut être calculée en conditionnant suivant le résultat de la première épreuve comme suit: notons par F l'événement «la première épreuve donne 5», par G «elle donne 7» et par H «elle ne donne ni 5 ni 7». En conditionnant sur l'événement qui se produit, on obtient

$$P(E) = P(E|F)P(F) + P(E|G)P(G) + P(E|H)P(H)$$

Cependant

$$\begin{aligned} P(E|F) &= 1 \\ P(E|G) &= 0 \\ P(E|H) &= P(E) \end{aligned}$$

Ces deux premières probabilités sont évidentes. La troisième égalité résulte du fait que si la première épreuve ne donne ni 5 ni 7, on se retrouve exactement dans la situation de départ: l'expérimentateur va répéter le jet des deux dés jusqu'à ce qu'un 5 ou un 7 apparaisse. On peut également remarquer que du fait de l'indépendance des épreuves le résultat de la première d'entre elles n'affecte pas celui des autres. Pour conclure, comme $P(F) = \frac{4}{36}$, $P(G) = \frac{6}{36}$ et $P(H) = \frac{26}{36}$, il vient

$$P(E) = \frac{1}{9} + P(E)\frac{13}{18}$$

ou

$$P(E) = \frac{2}{5}$$

Le lecteur remarquera que cette réponse est conforme à l'intuition basée sur les probabilités $\frac{4}{36}$ pour un 5 et $\frac{6}{36}$ pour un 7: on peut en effet penser que les chances sont à 4 contre 6 pour le premier, ce qui donne bien une probabilité de $\frac{4}{10}$.

Le même argument montre que si E et F sont des événements s'excluant mutuellement lors d'une expérience donnée, en répétant cette expérience pour réaliser une séquence d'épreuves indépendantes, on aura pour probabilité que E survienne avant F

$$\frac{P(E)}{P(E) + P(F)}$$

■

3.4.12 Problème des points

L'exemple suivant illustre un problème qui a pris une place d'honneur dans la théorie des probabilités, le célèbre *problème des points*. En termes généraux, voici de quoi il s'agit: deux joueurs engagent des mises et participent à un jeu quelconque; le gagnant empochera les mises. Mais ils sont interrompus avant la fin du jeu, alors qu'ils n'ont pour l'instant que des scores «intermédiaires» ou partiels. Comment doit-on partager les mises?

Le problème a été soumis pour la première fois au mathématicien français Blaise Pascal en 1654 par le Chevalier de Méré, alors joueur professionnel. Pour attaquer le problème, Pascal introduisit une idée importante: celle que la proportion des mises

méritées par chaque concurrent doit dépendre de leurs probabilités respectives de gagner à partir de là si le jeu était poursuivi. Pascal étudia quelques situations spéciales, mais surtout établit alors une correspondance épistolaire avec le célèbre Français Fermat dont la réputation de mathématicien brillant était immense. L'échange de lettres qui en résulta mena non seulement à une solution complète du problème des points, mais en même temps à celle de bien d'autres questions liées aux jeux de chance. Cette correspondance dont on parla beaucoup et que certains considèrent comme la naissance de la théorie des probabilités eut également l'avantage de stimuler l'intérêt porté aux probabilités par les mathématiciens européens car Pascal et Fermat étaient considérés comme deux des meilleurs mathématiciens de l'époque. Par exemple, le jeune génie néerlandais Huygens se déplaça à Paris peu après la parution de leur correspondance pour discuter ces problèmes et leurs solutions. L'intérêt et l'activité dans ce nouveau domaine s'étendit rapidement.

Exemple 3.25 Problème des points

On réalise des épreuves indépendantes, le succès ayant p pour probabilité et l'échec $1 - p$. Quelle est la probabilité que n succès apparaissent avant qu'il n'en soit de même pour m échecs? Nous admettons que lors d'un succès, c'est un joueur A qui marque un point, tandis qu'en cas d'échec, c'est B . La probabilité demandée nous ramène au problème des points si A et B en sont au stade où A doit marquer n points de plus pour gagner, contre m pour B .

Nous donnerons deux solutions. La première est due à Fermat, la seconde à Pascal.

SOLUTION (de Fermat). Désignons par $P_{n,m}$ la probabilité que n succès apparaissent avant que m échecs ne le fassent. En conditionnant sur le résultat de la première épreuve, on obtient (expliquer pourquoi):

$$P_{n,m} = pP_{n-1,m} + (1-p)P_{n,m-1} \quad n \geq 1, m \geq 1$$

Pour $P_{n,m}$ on peut résoudre cette équation grâce aux conditions limites évidentes $P_{n,0} = 0$ et $P_{0,m} = 1$. Mais plutôt que de s'enfoncer dans les détails ennuyeux de cette solution, voyons celle proposée par Pascal.

SOLUTION (de Pascal). Celui-ci donna l'argument que pour obtenir n succès sans que m échecs aient eu lieu, il est nécessaire et suffisant qu'il y ait eu au moins n succès parmi les $m + n - 1$ premières épreuves. (On supposera que même si le jeu est terminé avant le dernier de ces essais, on continue jusqu'à compléter la séquence.) Cet argument est fondé. En effet, s'il y a eu au moins n succès lors des $m + n - 1$ premiers essais, il y a également au plus $m - 1$ échecs et l'on obtient bien n succès avant m échecs. Inversement, s'il y a moins de n succès lors des $m + n - 1$ premiers essais, il y a nécessairement au moins m échecs. Dans ce cas on n'observera pas les n succès voulus avant le m -ième échec.

La probabilité d'avoir k succès sur $m + n - 1$ essais est, d'après l'exemple 3.21,

$$\binom{m+n-1}{k} p^k (1-p)^{m+n-1-k}$$

Par conséquent, la probabilité voulue (n succès avant le m -ème échec) sera

$$P_{n,m} = \sum_{k=n}^{m+n-1} \binom{m+n-1}{k} p^k (1-p)^{m+n-1-k}$$

Une autre solution au problème des points est présentée dans l'exercice théorique 3.6.10.

Pour illustrer le problème des points, on supposera que lors d'un jeu chacun de deux joueurs misera A francs et que chacun a la même probabilité d'emporter un essai donné ($p = \frac{1}{2}$). On dira qu'un joueur a gagné s'il a n points. A un moment donné le premier joueur a 1 point et l'autre 0. Celui-là a donc droit à

$$2AP_{n-1,n} = 2A \sum_{k=n-1}^{2n-2} \binom{2n-2}{k} \left(\frac{1}{2}\right)^{2n-2}$$

Maintenant,

$$\begin{aligned} \sum_{k=n-1}^{2n-2} \binom{2n-2}{k} &= \sum_{k=n-1}^{2n-2} \binom{2n-2}{2n-2-k} \\ &= \sum_{i=n-1}^0 \binom{2n-2}{i} \end{aligned}$$

où la dernière identité résulte de la substitution de i à $2n-2-k$. Ainsi,

$$\begin{aligned} 2 \sum_{k=n-1}^{2n-2} \binom{2n-2}{k} &= \sum_{k=0}^{2n-2} \binom{2n-2}{k} + \binom{2n-2}{n-1} \\ &= (1+1)^{2n-2} + \binom{2n-2}{n-1} \end{aligned}$$

ce qui donne droit, pour le premier joueur, à

$$A \left[1 + \left(\frac{1}{2}\right)^{2n-2} \binom{2n-2}{n-1} \right] \quad \blacksquare$$

3.4.13 Problème de la ruine du joueur

L'exemple suivant traite du célèbre problème dit de la ruine du joueur.

Exemple 3.26 Deux joueurs A et B misent sur les résultats successifs du jet répété d'une pièce. A chaque jet A reçoit une unité de la part de B si pile est sorti tandis qu'il paie une unité à B dans le cas contraire. Ils poursuivent le jeu tant qu'aucun des deux n'est ruiné. On suppose que les jets sont indépendants et que le côté pile de la pièce apparaît avec une probabilité p . Soient encore i et $N-i$ les fortunes initiales de A et B respectivement. Quelle est la probabilité que A gagne?

SOLUTION. Désignons par E l'événement « A finit par tout gagner, étant parti avec i unités alors que B en avait $N - i$ ». Pour marquer clairement l'influence de la fortune initiale de A , on notera par P_i la probabilité de E . Nous allons obtenir une expression de $P(E)$ en conditionnant suivant le résultat du premier jet comme suit: soit H l'événement «le premier jet donne pile». Alors,

$$\begin{aligned} P_i &= P(E) = P(E|H)P(H) + P(E|H^c)P(H^c) \\ &= pP(E|H) + (1-p)P(E|H^c) \end{aligned}$$

Si le premier jet donne pile, la situation à l'issue du premier pari est: A possède $i + 1$ unités et B en a $N - (i + 1)$. Du fait de l'indépendance de jets ayant tous la même probabilité p de donner pile, la situation est du point de vue de A exactement la même que si le jeu allait commencer avec comme conditions initiales $i + 1$ unités pour A et $N - (i + 1)$ pour B . Donc

$$P(E|H) = P_{i+1}$$

et similairement

$$P(E|H^c) = P_{i-1}$$

Aussi obtient-on, en posant $q = 1 - p$

$$P_i = pP_{i+1} + qP_{i-1} \quad i = 1, 2, \dots, N-1 \quad (3.8)$$

Utilisons les conditions limites évidentes $P_0 = 0$ et $P_N = 1$ pour résoudre les équations (3.8). Comme $p + q = 1$, ces équations équivalent à

$$pP_i + qP_i = pP_{i+1} + qP_{i-1}$$

ou

$$P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1}) \quad i = 1, 2, \dots, N-1 \quad (3.9)$$

Comme $P_0 = 0$, on peut tirer de l'équation (3.9)

$$\begin{aligned} P_2 - P_1 &= \frac{q}{p}(P_1 - P_0) = \frac{q}{p}P_1 \\ P_3 - P_2 &= \frac{q}{p}(P_2 - P_1) = \left(\frac{q}{p}\right)^2 P_1 \\ &\vdots \\ P_i - P_{i-1} &= \frac{q}{p}(P_{i-1} - P_{i-2}) = \left(\frac{q}{p}\right)^{i-1} P_1 \\ &\vdots \\ P_N - P_{N-1} &= \frac{q}{p}(P_{N-1} - P_{N-2}) = \left(\frac{q}{p}\right)^{N-1} P_1 \end{aligned} \quad (3.10)$$

L'addition des $i - 1$ premières équations de (3.10) donne

$$P_i - P_1 = P_1 \left[\left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + \dots + \left(\frac{q}{p}\right)^{i-1} \right]$$

ou

$$P_i = \begin{cases} \frac{1 - (q/p)^i}{1 - (q/p)} P_1 & \text{si } \frac{q}{p} \neq 1 \\ iP_1 & \text{si } \frac{q}{p} = 1 \end{cases}$$

Utilisons alors le fait que $P_N = 1$. On obtient

$$P_1 = \begin{cases} \frac{1 - (q/p)}{1 - (q/p)^N} & \text{si } p \neq \frac{1}{2} \\ \frac{1}{N} & \text{si } p = \frac{1}{2} \end{cases}$$

et ainsi

$$P_i = \begin{cases} \frac{1 - (q/p)^i}{1 - (q/p)^N} & \text{si } p \neq \frac{1}{2} \\ \frac{i}{N} & \text{si } p = \frac{1}{2} \end{cases} \quad (3.11)$$

Soit Q_i la probabilité que B finisse par tout gagner, étant parti avec $N - i$ unités tandis que A en possédait i . Par symétrie avec la situation traitée ci-dessus, mais en remplaçant p par q et i par $N - i$ on obtient

$$Q_i = \begin{cases} \frac{1 - (p/q)^{N-i}}{1 - (p/q)^N} & \text{si } q \neq \frac{1}{2} \\ \frac{N - i}{N} & \text{si } q = \frac{1}{2} \end{cases}$$

Maintenant, étant donné que $q = \frac{1}{2}$ est équivalent à $p = \frac{1}{2}$, dans le cas où $q \neq \frac{1}{2}$ on trouve

$$\begin{aligned} P_i + Q_i &= \frac{1 - (q/p)^i}{1 - (q/p)^N} + \frac{1 - (p/q)^{N-i}}{1 - (p/q)^N} \\ &= \frac{p^N - p^N (q/p)^i}{p^N - q^N} + \frac{q^N - q^N (p/q)^{N-i}}{q^N - p^N} \\ &= \frac{p^N - p^{N-i} q^i - q^N + q^i p^{N-i}}{p^N - q^N} \\ &= 1 \end{aligned}$$

Comme ce résultat reste valable lorsque $p = q = \frac{1}{2}$, on aura toujours

$$P_i + Q_i = 1$$

ce qui, en d'autres mots, établit que la probabilité qu'il y ait un gagnant est 1. Ou encore que la probabilité que le jeu se poursuive indéfiniment, la fortune de A oscillant constamment entre 1 et $N - 1$, est 0. (Le lecteur doit être attentif au fait qu'il existe a priori trois issues à ce jeu: A gagne, B gagne, ni l'un ni l'autre ne l'emporte et le jeu se poursuit indéfiniment. Nous venons de montrer que ce dernier événement est de probabilité nulle).

A titre d'illustration numérique des résultats qui précèdent, si A partait avec 5 unités et B avec 10, il aurait une chance sur trois de l'emporter lorsque $p = \frac{1}{2}$. La probabilité qu'il gagne sauterait à

$$\frac{1 - (\frac{2}{3})^5}{1 - (\frac{2}{3})^{15}} \approx .87$$

si p valait 0,6. ■

3.4.14 Problème de la durée du jeu

Un cas spécial du problème de la ruine du joueur, aussi connu sous le nom de *problème de la durée de jeu*, fut proposé par le Français Fermat en 1657. Dans la version qu'il proposait – et qui fut d'ailleurs résolue par Huygens –, A et B disposaient chacun de 12 pièces. Pour gagner une pièce, ils devaient jouer avec trois dés comme suit: chaque fois que 11 sort (en faisant la somme des trois dés, l'identité du lanceur étant indifférente), A donne une pièce à B . Chaque fois que 14 sort, c'est B qui donne une pièce à A . La première personne qui gagne toutes les pièces gagne le jeu. Comme $P(\text{tirer } 11) = \frac{27}{216}$ et $P(\text{tirer } 14) = \frac{15}{216}$, on réalise grâce à l'exemple 3.23 que du point de vue de A il s'agit ici précisément du problème de la ruine du joueur avec $p = \frac{15}{42}$, $i = 12$ et $N = 24$. La solution à la formulation générale du problème de la ruine du joueur fut donnée par le mathématicien Jacques Bernoulli dans une publication parue huit ans après sa mort, en 1713.

3.4.15 Application à un test d'efficacité

Voici une application du même problème à un test de médicaments: admettons que l'on vienne de développer deux nouveaux produits pour le traitement d'une maladie donnée. Le médicament i a un taux d'efficacité P_i , $i = 1, 2$. On entend par là que tout patient traité avec ce médicament a une probabilité P_i de guérir. Ces taux ne sont pas connus cependant et nous cherchons une méthode pour décider si $P_1 > P_2$ ou $P_1 < P_2$. Pour permettre le choix, on considère le test suivant: on traite paires de patients les uns après les autres. Un membre de la paire reçoit le produit 1 et l'autre le produit 2. Après le traitement d'une paire on détermine le résultat puis on passe au traitement de la paire suivante jusqu'à ce que le total des guérisons attribuées à l'une des drogues dépasse le total de celles attribuées à l'autre d'un nombre déterminé

d'avance. On peut formaliser ceci ainsi:

$$X_j = \begin{cases} 1 & \text{si le patient de la } j\text{-ème paire ayant reçu le médicament 1 est guéri} \\ 0 & \text{sinon} \end{cases}$$

$$Y_j = \begin{cases} 1 & \text{si le patient de la } j\text{-ème paire ayant reçu le médicament 2 est guéri} \\ 0 & \text{sinon.} \end{cases}$$

Pour un entier positif M fixé d'avance, le test s'arrête à la paire N où N est la première valeur de n telle que

$$X_1 + \cdots + X_n - (Y_1 + \cdots + Y_n) = M$$

ou

$$X_1 + \cdots + X_n - (Y_1 + \cdots + Y_n) = -M$$

Dans le premier cas, on décide que $P_1 > P_2$ et dans le second, c'est l'inverse.

Pour déterminer si ce test est bon, nous aimerions connaître la probabilité qu'il amène à une décision fautive. Plus précisément, P_1 et P_2 étant fixés et P_1 étant supérieur à P_2 , quelle est la probabilité que le test fasse déclarer à tort que $P_2 > P_1$? Pour déterminer cette probabilité, il faut d'abord remarquer qu'après chaque test portant sur une paire la différence des succès cumulés de chaque médicament peut augmenter de 1 avec probabilité $P_1(1 - P_2)$, puisque telle est la probabilité que le médicament 1 amène une guérison sans que le médicament 2 en fasse autant. Cette différence peut diminuer de 1 avec probabilité $(1 - P_1)P_2$ ou encore rester inchangée avec probabilité $P_1P_2 + (1 - P_1)(1 - P_2)$. Négligeons ces dernières paires. Avec les autres la différence augmentera de 1 avec une probabilité

$$P = P(\text{elle augmente de 1} | \text{elle augmente ou diminue de 1})$$

$$= \frac{P_1(1 - P_2)}{P_1(1 - P_2) + (1 - P_1)P_2}$$

et diminuera de 1 avec une probabilité

$$1 - P = \frac{P_2(1 - P_1)}{P_1(1 - P_2) + (1 - P_1)P_2}$$

Ainsi, la probabilité que le test fasse déclarer que $P_2 > P_1$ est égale à celle qu'un joueur pouvant gagner une unité avec probabilité P perde M de ces unités avant qu'il ait pu en accumuler M de plus qu'au départ. Mais l'équation (3.11) donne cette probabilité, avec ici $i = M$, $N = 2M$:

$P(\text{le test décide que } P_2 > P_1)$

$$\begin{aligned}
 &= 1 - \frac{1 - \left(\frac{1-P}{P}\right)^M}{1 - \left(\frac{1-P}{P}\right)^{2M}} \\
 &= 1 - \frac{1}{1 + \left(\frac{1-P}{P}\right)^M} \\
 &= \frac{1}{1 + \gamma^M}
 \end{aligned}$$

où

$$\begin{aligned}
 \gamma &= \frac{P}{1-P} \\
 &= \frac{P_1(1-P_2)}{P_2(1-P_1)}
 \end{aligned}$$

Par exemple, si $P_1 = 0,6$ et $P_2 = 0,4$, la probabilité de décision erronée est de 0,017 lorsqu'on fixe M à 5, diminuant à 0,0003 si l'on prend $M = 10$.

Supposons que nous ayons un ensemble d'éléments et que nous voulions déterminer si au moins un des éléments a une certaine propriété. Nous pouvons démarrer ce problème de manière probabiliste en choisissant aléatoirement un élément de l'ensemble de telle sorte que chaque élément ait une probabilité positive d'être choisi. On peut répondre à la question en calculant la probabilité que l'élément sélectionné au hasard ait la propriété voulue. Si cette probabilité est positive, nous avons alors établi qu'au moins un des éléments de l'ensemble possède la propriété et si elle vaut zéro, alors aucun des éléments ne l'a.

Le dernier exemple de la section illustre cette technique.

Exemple 3.27 Un graphe complet à n sommets est défini comme étant un ensemble de n points (appelés sommets) dans le plan et de $\binom{n}{2}$ segments (appelés arêtes) reliant chaque paire de sommets. La figure 3.3 montre un graphe complet à 3 sommets. Supposons que chaque arête d'un graphe complet à n sommets doit être colorée soit en rouge soit en bleu. Pour un entier k fixé, la question est de savoir s'il existe une façon de colorier les arêtes de telle sorte qu'aucun ensemble de k sommets n'ait ses $\binom{k}{2}$ arêtes de la même couleur. On peut montrer, de manière probabiliste, que si n n'est pas trop grand la réponse est oui.

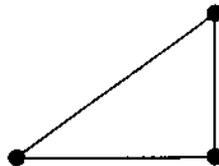


Figure 3.3

SOLUTION. Supposons que chaque arête peut, de manière indépendante, être coloriée soit en rouge soit en bleu de manière équiprobable. Ainsi, chaque arête a une probabilité $1/2$ d'être rouge. On numérote les $\binom{n}{k}$ ensembles de k sommets et on définit les événements E_i , $i = 1, \dots, \binom{n}{k}$ de la façon suivante:

$E_i = \{\text{Toutes les arêtes du } i\text{-ième ensemble de } k \text{ sommets ont la même couleur}\}$

Maintenant, puisque chacune des $\binom{k}{2}$ arêtes d'un ensemble de k sommets a autant de chance d'être bleue ou rouge, la probabilité-*qu'elles soient toutes de la même couleur* est

$$P(E_i) = 2 \left(\frac{1}{2} \right)^{\binom{k}{2}}$$

Par conséquent, puisque

$$P(\cup_i E_i) \leq \sum_i P(E_i) \quad (\text{inégalité de Boole})$$

on obtient que $P(\cup_i E_i)$, la probabilité qu'il y ait un ensemble de k sommets dont toutes les arêtes sont de la même couleur, satisfait :

$$P(\cup_i E_i) \leq \binom{n}{k} \left(\frac{1}{2} \right)^{\binom{k}{2}}$$

Si $\binom{n}{k} \left(\frac{1}{2} \right)^{\binom{k}{2}} < 1$ ou, de manière équivalente, si

$$\binom{n}{k} < 2^{\binom{k}{2}}$$

alors la probabilité qu'au moins un des $\binom{n}{k}$ ensembles de k sommets ait toutes ses arêtes de la même couleur est inférieure à 1. Par conséquent, sous la condition précédente sur n et k , il y a une probabilité positive qu'aucun ensemble de k sommets n'ait toutes ses arêtes de la même couleur. Mais ceci implique qu'il y a au moins une façon de colorier les arêtes de telle sorte qu'aucun ensemble de k sommets n'ait toutes ses arêtes de la même couleur. ■

REMARQUES. a) Bien que la discussion ci-dessus établisse une condition sur n et k qui garantit l'existence d'une combinaison de couleurs satisfaisant la propriété désirée, elle ne donne aucune information sur la façon d'obtenir une telle combinaison. (Une possibilité serait tout simplement de choisir les couleurs au hasard, de voir si le résultat satisfait la propriété et de recommencer jusqu'à ce que la propriété soit satisfaite.)

b) La méthode utilisant les probabilités pour résoudre un problème à la base purement déterministe est appelée la méthode probabiliste¹.

D'autres exemples de cette méthode sont donnés dans l'exercice théorique 19 et dans l'exemple 7.18 du chapitre 7.

¹ Voir N. Alon, J. Spencer, and P. Erdos, *The Probabilistic Method* (New York: John Wiley & Sons, Inc., 1992).

3.5 FONCTION DE PROBABILITÉ CONDITIONNELLE

3.5.1 $P(\cdot | F)$ est une fonction de probabilité

Les probabilités conditionnelles satisfont à toutes les propriétés des probabilités ordinaires. Le théorème 3.3 le démontre, puisqu'il établit que $P(E|F)$ satisfait aux trois axiomes d'une probabilité.

Théorème 3.3

(a) $0 \leq P(E|F) \leq 1$.

(b) $P(S|F) = 1$.

(c) Si $E_i, i = 1, 2, \dots, n$ sont des événements qui s'excluent mutuellement, alors

$$P\left(\bigcup_{i=1}^{\infty} E_i | F\right) = \sum_{i=1}^{\infty} P(E_i | F)$$

DÉMONSTRATION. Pour démontrer la partie (a), nous devons établir que $0 \leq P(EF)/P(F) \leq 1$. L'inégalité de gauche est évidente; celle de droite résulte du fait que $EF \subset F$, ce qui implique $P(EF) \leq P(F)$.

La partie (b) est prouvée par

$$P(S|F) = \frac{P(SF)}{P(F)} = \frac{P(F)}{P(F)} = 1$$

La partie (c) l'est également car

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} E_i | F\right) &= \frac{P\left(\left(\bigcup_{i=1}^{\infty} E_i\right)F\right)}{P(F)} \\ &= \frac{P\left(\bigcup_{i=1}^{\infty} E_i F\right)}{P(F)} \text{ puisque } \left(\bigcup_{i=1}^{\infty} E_i\right)F = \bigcup_{i=1}^{\infty} E_i F \\ &= \frac{\sum_{i=1}^{\infty} P(E_i F)}{P(F)} \\ &= \sum_{i=1}^{\infty} P(E_i | F) \end{aligned}$$

où l'avant-dernière égalité est justifiée par le fait que $E_i E_j = \emptyset$ entraîne que $E_i F E_j F = \emptyset$. ■

Si nous posons $Q(E) = P(E|F)$, $Q(E)$ peut être considérée grâce au théorème 3.3 comme une fonction de probabilité sur les événements de S . Aussi toutes les propositions établies jusque-là pour des fonctions de probabilité s'appliquent à $Q(E)$. Par exemple,

$$Q(E_1 \cup E_2) = Q(E_1) + Q(E_2) - Q(E_1 E_2)$$

ou, de manière équivalente,

$$P(E_1 \cup E_2 | F) = P(E_1 | F) + P(E_2 | F) - P(E_1 E_2 | F)$$

Par ailleurs, on peut définir la probabilité conditionnelle $Q(E_1 | E_2) = Q(E_1 E_2) / Q(E_2)$. D'une application de l'équation (3.3) il résulte que

$$Q(E_1) = Q(E_1 | E_2)Q(E_2) + Q(E_1 | E_2^c)Q(E_2^c) \quad (3.12)$$

or,

$$\begin{aligned} Q(E_1 | E_2) &= \frac{Q(E_1 E_2)}{Q(E_2)} \\ &= \frac{P(E_1 E_2 | F)}{P(E_2 | F)} \\ &= \frac{\frac{P(E_1 E_2 F)}{P(F)}}{\frac{P(E_2 F)}{P(F)}} \\ &= P(E_1 | E_2 F) \end{aligned}$$

et dès lors, (3.12) permet d'écrire:

$$P(E_1 | F) = P(E_1 | E_2 F)P(E_2 | F) + P(E_1 | E_2^c F)P(E_2^c | F)$$

3.5.2 Exemples

Exemple 3.8 Revenons à l'exemple 3.8 qui s'intéresse à une compagnie d'assurances qui pense qu'on peut diviser les gens en deux classes distinctes: ceux qui sont enclins aux accidents et les gens à faible risque. Sur une période d'un an une personne à haut risque sera victime d'un accident avec probabilité 0,4 contre 0,2 pour une personne à faible risque. Quelle est la probabilité conditionnelle pour un nouveau client d'avoir un accident dans sa deuxième année de contrat s'il a eu un accident durant la première année?

SOLUTION. Soit A l'événement «le client est à haut risque» et soient A_i , $i = 1, 2$ «il a eu un accident durant la i -ème année». On peut calculer la probabilité $P(A_2 | A_1)$ demandée en conditionnant sur le fait que le client est ou n'est pas à haut risque de la manière suivante:

$$P(A_2 | A_1) = P(A_2 | A A_1)P(A | A_1) + P(A_2 | A^c A_1)P(A^c | A_1)$$

Maintenant

$$P(A | A_1) = \frac{P(A_1 A)}{P(A_1)} = \frac{P(A_1 | A)P(A)}{P(A_1)}$$

On avait supposé $P(A)$ égal à $\frac{3}{10}$ et montré que $P(A_1) = 0,26$ (dans l'exemple 3.8).
Donc

$$P(A|A_1) = \frac{(.4)(.3)}{.26} = \frac{6}{13}$$

et donc

$$P(A^c|A_1) = 1 - P(A|A_1) = \frac{7}{13}$$

Puisque $P(A_2|AA_1) = .4$ et $P(A_2|A^cA_1) = .2$, on voit que

$$P(A_2|A_1) = (.4)\frac{6}{13} + (.2)\frac{7}{13} \approx .29$$

■

L'exemple qui suit traite de la théorie des chaînes de résultats.

Exemple 3.29 On réalise une expérience composée d'épreuves indépendantes. La probabilité de succès est p , d'échec $q = 1 - p$. Nous nous demandons la probabilité qu'une chaîne de n succès consécutifs ait lieu avant que n'apparaisse une chaîne de m échecs.

SOLUTION. Soit E l'événement étudié. Pour calculer $P(E)$, nous commençons par conditionner suivant l'issue de la première épreuve. On obtient alors, en posant que H désigne l'événement «la première épreuve livre un succès»

$$P(E) = pP(E|H) + qP(E|H^c) \quad (3.13)$$

Admettons que la première épreuve ait été un succès. A partir de là l'une des possibilités d'obtenir n succès avant que n'apparaissent m échecs serait de n'avoir que des succès sur les $n - 1$ épreuves suivantes. Aussi conditionnerons-nous sur le fait que cela arrive ou n'arrive pas. Notons par F l'événement «des épreuves 2 à n sont toutes des succès». On obtient

$$P(E|H) = P(E|FH)P(F|H) + P(E|F^cH)P(F^c|H) \quad (3.14)$$

Or, $P(E|FH)$ est clairement égal à 1, et d'autre part si F^cH survient, c'est que la première épreuve fut un succès mais l'une au moins des $n - 1$ suivantes fut un échec. Mais lorsque cet échec se réalise, on se trouve dans la même situation que si on commençait l'expérience avec un échec car la chaîne de succès est brisée. Donc

$$P(E|F^cH) = P(E|H^c)$$

Comme l'indépendance des épreuves entraîne celle de F et H et comme $P(F) = p^{n-1}$, on tire de (3.14)

$$P(E|H) = p^{n-1} + (1 - p^{n-1})P(E|H^c) \quad (3.15)$$

On peut obtenir d'une manière très similaire une expression de $P(E|H^c)$. Plus précisément soit G l'événement «les épreuves 2 à m sont toutes des échecs». Dans ce cas

$$P(E|H^c) = P(E|GH^c)P(G|H^c) + P(E|G^cH^c)P(G^c|H^c) \quad (3.16)$$

Or GH^c est l'événement «les m premières épreuves donnent toutes des échecs» et par conséquent $P(E|GH^c) = 0$. De plus si $G^c H^c$ se réalise, la première épreuve est un échec, mais il y a au moins un succès parmi les $m - 1$ épreuves suivantes. Comme ce succès brise la chaîne des échecs, on peut dire que

$$P(E|G^cH^c) = P(E|H)$$

Ainsi obtient-on, en utilisant $P(G^c|H^c) = P(G^c) = 1 - q^{m-1}$ et aussi (3.16),

$$P(E|H^c) = (1 - q^{m-1})P(E|H) \quad (3.17)$$

Les solutions de (3.15) et (3.17) sont

$$P(E|H) = \frac{p^{n-1}}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}}$$

et

$$P(E|H^c) = \frac{(1 - q^{m-1})p^{n-1}}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}}$$

et donc

$$\begin{aligned} P(E) &= pP(E|H) + qP(E|H^c) \\ &= \frac{p^n + qp^{n-1}(1 - q^{m-1})}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}} \\ &= \frac{p^{n-1}(1 - q^m)}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}} \end{aligned} \quad (3.18)$$

On peut d'ailleurs noter que, du fait de la symétrie du problème, cette formule livre la probabilité d'obtenir une chaîne de m échecs avant l'apparition d'une chaîne de n succès. Il suffit d'intervertir dans (3.18) p et q d'une part, n et m de l'autre. Cette probabilité serait donc

$P(\text{chaîne de } m \text{ échecs avant une chaîne de } n \text{ succès})$

$$= \frac{q^{m-1}(1 - p^n)}{q^{m-1} + p^{n-1} - q^{m-1}p^{n-1}} \quad (3.19)$$

De plus, comme on voit que la somme des probabilités données par (3.18) et (3.19) est 1, il est certain qu'il finira par se produire soit une chaîne de n succès soit une autre de m échecs.

A titre d'illustration de (3.18), on peut dire que lors du jet répété d'une pièce équilibrée la probabilité de voir sortir une chaîne de 2 piles avant que n'apparaisse une chaîne de 3 faces est $\frac{7}{10}$; cette probabilité monte à $\frac{5}{6}$ pour des chaînes de 2 piles contre 4 faces. ■

Dans l'exemple suivant nous allons reprendre le problème de rencontre de Montmort (exemple 2.13). Nous lui donnerons cette fois-ci une solution en passant par des probabilités conditionnelles.

Exemple 3.30 Lors d'une réunion de n hommes chacun enlève son chapeau. Les chapeaux sont mélangés et chacun en tire un au hasard. On dira qu'il y a rencontre lorsque quelqu'un a tiré son propre chapeau.

- a) Quelle est la probabilité qu'il n'y ait pas de rencontre ?
 b) Quelle est celle qu'il y ait exactement k rencontres ?

SOLUTION. a) Désignons par E l'événement «il n'y a aucune rencontre»; pour faire clairement apparaître l'influence de n nous écrivons $P_n = P(E)$. L'idée de départ est de conditionner sur le fait que le premier homme a ou n'a pas tiré son propre chapeau, événements que nous noterons R et R^c . Alors:

$$P_n = P(E) = P(E|R)P(R) + P(E|R^c)P(R^c)$$

Il est évident que $P(E|R) = 0$ et donc

$$P_n = P(E|R^c) \frac{n-1}{n} \quad (3.20)$$

Ceci dit, $P(E|R^c)$ est la probabilité qu'il n'y ait pas de rencontre lorsque $n-1$ hommes tirent chacun un chapeau d'un tas en comptant $n-1$ mais ne comprenant pas le chapeau de l'un de ces hommes (le chapeau tiré par le premier individu). Lorsqu'il n'y a pas de rencontre, deux cas de figure peuvent se présenter: soit l'homme «en trop» (son chapeau n'est pas dans le tas) ne tire pas le chapeau «en trop» (celui du premier homme), soit il le tire. La probabilité d'absence de rencontre dans le premier cas est simplement P_{n-1} ; on peut s'en convaincre en considérant que le chapeau «appartient» à l'homme en trop. Comme la probabilité de l'alternative est $[1/(n-1)]P_{n-2}$, nous avons maintenant

$$P(E|R^c) = P_{n-1} + \frac{1}{n-1} P_{n-2}$$

et donc en remplaçant dans (3.20)

$$P_n = \frac{n-1}{n} P_{n-1} + \frac{1}{n} P_{n-2}$$

ou, de manière équivalente,

$$P_n - P_{n-1} = -\frac{1}{n} (P_{n-1} - P_{n-2}) \quad (3.21)$$

Par ailleurs, P_n est la probabilité de n'avoir aucune rencontre, donc

$$P_1 = 0 \quad P_2 = \frac{1}{2}$$

et par conséquent, l'équation (3.21) donne

$$P_3 - P_2 = -\frac{(P_2 - P_1)}{3} = -\frac{1}{3!} \quad \text{ou} \quad P_3 = \frac{1}{2!} - \frac{1}{3!}$$

$$P_4 - P_3 = -\frac{(P_3 - P_2)}{4} = \frac{1}{4!} \quad \text{ou} \quad P_4 = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}$$

On peut voir que la formule générale est

$$P_n = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots + \frac{(-1)^n}{n!}$$

b) Pour obtenir la réponse à la question, à savoir la probabilité d'observer exactement k rencontres, considérons un groupe quelconque de k hommes. La probabilité qu'eux et eux seulement choisissent leur propre chapeau est

$$\frac{1}{n} \frac{1}{n-1} \dots \frac{1}{n-(k-1)} P_{n-k} = \frac{(n-k)!}{n!} P_{n-k}$$

où P_{n-k} est la probabilité conditionnelle qu'aucun des $n-k$ autres hommes ne tire son propre chapeau. Mais il y a $\binom{n}{k}$ manières de déterminer le groupe initial des k hommes qui tireront leur chapeau; par conséquent la probabilité demandée est

$$\frac{P_{n-k}}{k!} = \frac{\frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^{n-k}}{(n-k)!}}{k!} \quad \blacksquare$$

3.5.3 Indépendance conditionnelle

Un concept important en théorie des probabilités est celui d'indépendance conditionnelle entre événements. Deux événements E_1 et E_2 seront dits *conditionnellement indépendants* selon F si la probabilité conditionnelle de E_1 , F étant réalisé, n'est pas affectée par l'information que E_2 est ou n'est pas survenu. On peut écrire plus formellement que E_1 et E_2 sont conditionnellement indépendants selon F si

$$P(E_1|E_2F) = P(E_1|F) \quad (3.22)$$

ou, de manière équivalente

$$P(E_1E_2|F) = P(E_1|F)P(E_2|F) \quad (3.23)$$

On peut facilement étendre cette notion d'indépendance conditionnelle à plus de deux événements. Ce travail est laissé au lecteur à titre d'exercice.

On aura remarqué que ce concept a été employé implicitement dans l'exemple 3.26 où l'on a admis que les événements «le signataire du contrat d'assurances a un accident durant la i -ème année de contrat» ($i = 1, 2$) étaient conditionnellement

indépendants selon que cet individu était à faible ou haut risque. [On avait utilisé cette hypothèse pour pouvoir attribuer à $P(A_2|AA_1)$ et $P(A_2|A^cA_1)$ respectivement les valeurs 0,4 et 0,2]. L'exemple suivant illustre la notion d'indépendance conditionnelle. On l'intitule parfois *règle de succession de Laplace*.

Exemple 3.31 Règle de succession de Laplace.

Une boîte contient $k + 1$ pièces. Pour la i -ème pièce, la probabilité de montrer pile lors d'un jet est i/k , $i = 0, 1, \dots, k$. On tire une pièce au hasard de la boîte pour la jeter ensuite un grand nombre de fois. Quelle est la probabilité conditionnelle que le $(n + 1)$ -ième jet donne pile sachant que les n premiers l'ont fait?

SOLUTION. Désignons par E_i l'événement «la pièce sélectionnée était la i -ème», $i = 0, 1, \dots, k$, par F_n l'événement «les n premiers jets donnent tous pile», enfin par F «le $(n + 1)$ -ième jet donne pile». La probabilité $P(F|F_n)$ que l'on cherche peut être calculée ainsi:

$$P(F|F_n) = \sum_{i=0}^k P(F|F_n E_i) P(E_i|F_n)$$

Si l'on admet que c'est la i -ème pièce qui a été tirée, les issues des jets seront conditionnellement indépendantes, pile apparaissant avec la probabilité i/k . Donc,

$$P(F|F_n E_i) = P(F|E_i) = \frac{i}{k}$$

Aussi

$$\begin{aligned} P(E_i|F_n) &= \frac{P(E_i F_n)}{P(F_n)} \\ &= \frac{P(F_n|E_i)P(E_i)}{\sum_{j=0}^k P(F_n|E_j)P(E_j)} \\ &= \frac{(i/k)^n [1/(k+1)]}{\sum_{j=0}^k (j/k)^n [1/(k+1)]} \\ P(F|F_n) &= \frac{\sum_{i=0}^k (i/k)^{n+1}}{\sum_{j=0}^k (j/k)^n} \end{aligned}$$

Lorsque k est grand on peut raisonnablement admettre les approximations intégrales

$$\begin{aligned} \frac{1}{k} \sum_{i=0}^k \left(\frac{i}{k}\right)^{n+1} &\approx \int_0^1 x^{n+1} dx = \frac{1}{n+2} \\ \frac{1}{k} \sum_{j=0}^k \left(\frac{j}{k}\right)^n &\approx \int_0^1 x^n dx = \frac{1}{n+1} \end{aligned}$$

et ainsi pour k grand

$$P(F|F_n) \approx \frac{n+1}{n+2}$$

■

3.6 EXERCICES THÉORIQUES

3.6.1 Une boule peut se trouver dans n'importe laquelle de n boîtes. Elle se trouve dans la boîte i avec probabilité P_i . Si elle se trouve dans la boîte i , elle ne sera détectée au cours d'une fouille de cette boîte qu'avec la probabilité α_i . Montrer que la probabilité conditionnelle que la boule se trouve dans la boîte j , sachant qu'une fouille de la boîte i n'a rien donné, est:

$$\frac{P_j}{1 - \alpha_i P_i} \quad \text{si } j \neq i$$

$$\frac{(1 - \alpha_i) P_i}{1 - \alpha_i P_i} \quad \text{si } j = i$$

3.6.2 Pour chacune des assertions suivantes donner soit une preuve, soit un contre-exemple:

- si E est indépendant de F et G , il l'est de $F \cup G$;
- si E est indépendant de F et G et si $FG = \emptyset$, E est indépendant de $F \cup G$;
- si E est indépendant de F et si F l'est de G et si en plus E l'est de FG , alors G est indépendant de EF .

3.6.3 L'événement F est porteur d'une information négative sur E , ce que nous écrirons $F \searrow E$, si

$$P(E|F) \leq P(E)$$

Pour chacune des assertions suivantes, donner soit une preuve soit un contre-exemple:

- si $F \searrow E$ alors $E \searrow F$;
- si $F \searrow E$ et $E \searrow G$, alors $F \searrow G$;
- si $F \searrow E$ et $G \searrow E$, alors $FG \searrow E$.

Répondre aux mêmes questions lorsqu'on remplace \searrow par \nearrow , $F \nearrow E$ signifiant $P(E|F) \geq P(E)$: autrement dit, F est porteur d'information positive sur E .

3.6.4 Soient $\{E_n, n \geq 1\}$ et $\{F_n, n \geq 1\}$ des suites croissantes d'événements ayant pour limites E et F respectivement. Montrer que E est indépendant de F si, pour tout n , E_n l'est de F_n .

3.6.5 Montrer que si E_1, E_2, \dots, E_n sont des événements totalement indépendants on aura

$$P(E_1 \cup E_2 \cup \dots \cup E_n) = 1 - \prod_{i=1}^n [1 - P(E_i)]$$

3.6.6

a) Une urne contient n boules blanches et m noires. On retire les boules une à une, jusqu'à ce que l'on soit sûr que toutes celles qui restent sont de la même couleur. Montrer que la probabilité que cette couleur soit le blanc est $n/(n+m)$. On imaginera que l'expérience est poursuivie jusqu'à la dernière boule et on considèrera la couleur de celle-ci.

b) Un étang est peuplé de trois espèces de poissons, qu'on appellera poissons rouges, bleus et verts. Les nombres des poissons de chaque espèce sont respectivement R , B et V . On retire les poissons un à un, au hasard (ce qui signifie qu'à chaque étape tout poisson restant a la même probabilité d'être tiré). Quelle est la probabilité que la première espèce à disparaître de l'étang soit le poisson rouge?

On partira de l'égalité $P(\{R\}) = P(\{RBV\}) + P(\{RVB\})$, puis on calculera les probabilités du membre de droite en conditionnant suivant la couleur de la dernière espèce à disparaître.

3.6.7 On considère des nombres a_i , $i = 1, 2, \dots$; de plus $0 \leq a_i \leq 1 \forall i$. Montrer que

$$\sum_{i=1}^{\infty} \left[a_i \prod_{j=1}^{i-1} (1 - a_j) \right] + \prod_{i=1}^{\infty} (1 - a_i) = 1$$

On peut pour cela imaginer qu'on lance un nombre infini de pièces et on considère a_i comme la probabilité que la i -ième pièce montre pile et considérer la première apparition du pile.

3.6.8 Une pièce tombe sur face avec probabilité p . Un joueur A commence à la lancer et poursuit jusqu'à la première apparition de pile. A ce moment B la lance jusqu'à ce que pile apparaisse pour la première fois, ce qui fait passer la pièce à A et ainsi de suite. On note par $P_{n,m}$ la probabilité que A ait pu accumuler un total de n faces avant que B en ait eu m . Montrer que

$$P_{n,m} = pP_{n-1,m} + (1-p)(1 - P_{m,n})$$

3.6.9 Vous jouez contre un adversaire infiniment riche, le jeu est divisé en parties. A chaque partie, vous pouvez soit gagner soit perdre une unité avec pour probabilité respectivement p et $1-p$. Montrer que la probabilité que vous finissiez par vous ruiner est

$$\begin{cases} 1 & \text{si } p \leq \frac{1}{2} \\ (q/p)^i & \text{si } p > \frac{1}{2} \text{ où } q = 1 - p \end{cases}$$

où i désigne votre fortune initiale.

3.6.10 On réalise des épreuves indépendantes jusqu'à obtenir r succès. La probabilité d'un succès est p . Montrer que la probabilité qu'il faille n épreuves est

$$\binom{n-1}{r-1} p^r (1-p)^{n-r}$$

Utiliser ce résultat pour résoudre le problème des points (exemple 3.25). On peut d'abord se demander le nombre de succès qu'on doit avoir lors des $n - 1$ premiers essais pour obtenir r succès en n épreuves.

3.6.11 On appelle épreuves de Bernoulli des épreuves indépendantes ayant p pour probabilité de succès et $1 - p$ pour probabilité d'échec. Appelons P_n la probabilité que n épreuves de Bernoulli successives débouchent sur un nombre pair de succès; 0 est considéré pair. Montrer que

$$P_n = p(1 - P_{n-1}) + (1 - p)P_{n-1} \quad n \geq 1$$

et utiliser ce résultat pour démontrer par induction que

$$P_n = \frac{1 + (1 - 2p)^n}{2}$$

3.6.12 Soit Q_n la probabilité qu'il n'apparaisse aucune série de trois piles consécutifs lors de l'expérience consistant à jeter n fois une pièce équilibrée. Montrer que

$$\begin{aligned} Q_n &= \frac{1}{2}Q_{n-1} + \frac{1}{4}Q_{n-2} + \frac{1}{8}Q_{n-3} \\ Q_0 &= Q_1 = Q_2 = 1 \end{aligned}$$

Calculer Q_8 . Pour cela, conditionner sur le premier face.

3.6.13 Le joueur A possède $n + 1$ pièces équilibrées, le joueur B en a n . Chacun lance toutes ses pièces. Montrer que la probabilité pour A d'avoir plus souvent pile que B est $\frac{1}{2}$.

Pour cela, conditionner selon que A ou B a plus souvent pile que l'autre lorsque les joueurs en sont au jet de leur n -ème pièce chacun (il y a trois éventualités).

3.6.14 Considérons le problème de la ruine du joueur à cela près que A et B décident de ne pas jouer plus de n parties. On désigne par $P_{n,i}$ la probabilité que A termine avec tout l'argent alors qu'il possédait i unités au départ, contre $N - i$ pour B . Exprimer $P_{n,i}$ en fonction de $P_{n-1,i+1}$ et $P_{n-1,i-1}$ et calculer $P_{7,3}$ lorsque $N = 5$.

3.6.15 On considère deux urnes contenant chacune des boules blanches et des boules noires. Les probabilités de tirer une boule blanche sont de p et p' , respectivement pour l'urne 1 et l'urne 2. On tire avec remplacement des boules, une à une, de la manière suivante: on détermine d'abord l'urne de laquelle la première boule sera tirée; l'urne 1 est choisie avec probabilité α , l'autre avec $1 - \alpha$. Pour la suite, les tirages obéissent à la règle suivante: lorsque la boule tirée est blanche, on la replace dans son urne, de laquelle on tire également la boule suivante; lorsqu'elle est noire au contraire, le tirage suivant est fait dans l'autre urne. Soit α_n la probabilité que la n -ème boule soit choisie dans l'urne 1. Montrer que

$$\alpha_{n+1} = \alpha_n(p + p' - 1) + 1 - p' \quad n \geq 1$$

et utiliser ce résultat pour montrer que

$$\alpha_n = \frac{1-p'}{2-p-p'} + \left(\alpha - \frac{1-p'}{2-p-p'} \right) (p+p'-1)^{n-1}$$

Soit P_n la probabilité que la n -ième boule tirée soit blanche.

Calculer P_n , $\lim_{n \rightarrow \infty} \alpha_n$ et $\lim_{n \rightarrow \infty} P_n$.

3.6.16 Problème d'élection.

Lors d'une élection le candidat A reçoit n voix contre seulement m ($m < n$) au candidat B . On admet que lors du dépouillement chacun des $(n+m)!/n!m!$ ordres de dépouillement est de même probabilité. $P_{n,m}$ désigne la probabilité que A reste du début à la fin en tête du scrutin.

a) Calculer $P_{2,1}$, $P_{3,2}$, $P_{4,1}$, $P_{4,2}$, $P_{4,3}$.

b) Trouver $P_{n,1}$, $P_{n,2}$.

c) En s'appuyant sur ces résultats, formuler une conjecture pour l'expression de $P_{n,m}$.

d) Etablir une formule récursive donnant $P_{n,m}$ en fonction de $P_{n-1,m}$ et $P_{n,m-1}$ en conditionnant selon que tel ou tel candidat a reçu la \dots -ième voix (compléter vous-même).

e) Utiliser ce dernier résultat pour démontrer votre conjecture faite en b) grâce à un raisonnement par induction portant sur $n+m$.

3.6.17 On construit un modèle simplifié de prévision météorologique en disant que le temps sera demain le même qu'aujourd'hui avec probabilité p . Le temps est sec (il ne peut être que sec ou humide) le 1er janvier. Montrer que la probabilité P_n qu'il soit sec n jours plus tard est donnée par

$$\begin{aligned} P_n &= (2p-1)P_{n-1} + (1-p) & n \geq 1 \\ P_0 &= 1 \end{aligned}$$

Montrer que

$$P_n = \frac{1}{2} + \frac{1}{2}(2p-1)^n \quad n \geq 0$$

3.6.18 Un sac contient a boules blanches et b noires. On tire des boules du sac selon le principe suivant:

a) une boule est tirée au hasard et mise de côté;

b) une seconde est également tirée au hasard. Si sa couleur diffère de celle de la première, on la réintroduit dans le sac et on recommence le processus depuis le début. Dans le cas contraire, on la met de côté et répète le point b).

En d'autres termes, les boules sont tirées et mises de côté jusqu'à ce qu'un changement de couleur intervienne, à partir de quoi la dernière boule tirée est réintroduite dans le sac et le processus réinitialisé. Désignons par $P_{a,b}$ la probabilité que la dernière boule du sac soit blanche. Montrer que

$$P_{a,b} = \frac{1}{2}$$

Raisonnement par induction sur $k \equiv a + b$, le nombre de boules dans le sac aux différents stades de l'expérience.

3.6.19 Un tournoi avec n participants se déroule de la manière suivante: chaque joueur joue une fois contre chaque autre joueur (il y a $\binom{n}{2}$ paires de joueurs). Le résultat de chaque match consiste en la victoire d'un joueur et la défaite de l'autre. Pour un entier k fixé, $k < n$, est-il possible que le résultat du tournoi soit tel que pour tous les ensembles de k joueurs, un joueur batte chaque membre de cet ensemble? Montrer que si

$$\binom{n}{k} \left[1 - \left(\frac{1}{2} \right)^k \right]^{n-k} < 1$$

un tel résultat est possible.

Pour cela, supposer que les résultats des matchs sont indépendants et que chaque match a la même probabilité d'être gagné par n'importe quel joueur. Numérotter les $\binom{n}{k}$ ensembles de k joueurs et soit B_i l'événement qu'aucun joueur ne batte les k joueurs du i -ième ensemble. Utiliser alors l'inégalité de Boole pour borner $P(\bigcup_i B_i)$.

3.6.20 Démontrer de manière directe que

$$P(E|F) = P(E|FG)P(G|F) + P(E|FG^c)P(G^c|F)$$

3.6.21 Etablir l'équivalence de (3.22) et de (3.23).

3.6.22 Généraliser la définition d'indépendance conditionnelle à plus de deux événements.

3.6.23 Démontrer ou infirmer par contre-exemple la proposition suivante: si E_1 et E_2 sont indépendants, $E_1|F$ et $E_2|F$ le sont aussi.

3.6.24 On considère la loi de succession de Laplace (exemple 3.31). Montrer que si les n premiers jets livrent tous pile, la probabilité conditionnelle que les m jets suivants donnent également pile seulement est $(n+1)/(n+m+1)$.

3.6.25 Dans le cas de la loi de succession de Laplace, on suppose que les n premiers jets ont donné r fois pile et $n-r$ fois face au total. Montrer que la probabilité que le $(n+1)$ -ième jet livre pile est $(r+1)/(n+2)$. Pour ce faire, démontrer puis utiliser l'identité

$$\int_0^1 y^n (1-y)^m dy = \frac{n! m!}{(n+m+1)!}$$

Pour démontrer cette identité, poser $C(n,m) = \int_0^1 y^n (1-y)^m dy$ et intégrer cette quantité par parties pour obtenir

$$C(n, m) = \frac{m}{n+1} C(n+1, m-1)$$

Démontrer l'identité proposée par induction sur m en partant de $C(n,0) = 1/(n+1)$.

3.6.26 L'un de vos amis, d'esprit peu mathématicien mais plutôt philosophe, soutient que la loi de succession de Laplace est infondée, car elle peut mener à ces conclusions ridicules. «Par exemple», dit-il, «cette loi prétend qu'un enfant de 10 ans vivra une onzième année avec une probabilité $\frac{11}{12}$. La même loi appliquée au grand-père de cet enfant et qui a déjà 80 ans lui donne 81 chances sur 82 de vivre un an de plus. Ce résultat est ridicule, l'enfant ayant clairement plus de chances de vivre un an de plus que son grand-père». Que répondriez-vous à votre ami?

3.7 PROBLÈMES

3.7.1 On jette deux dés équilibrés. Quelle est la probabilité qu'au moins l'un d'entre eux montre 6, sachant que les deux résultats sont différents?

3.7.2 On jette deux dés équilibrés. Quelle est la probabilité que le premier montre 6, sachant que la somme des deux est i ? Calculer le résultat pour toutes les valeurs de i comprises entre 2 et 12.

3.7.3 Utiliser l'équation (3.1) pour calculer la probabilité que, lors d'une partie de bridge, Est ait 3 piques, sachant que Nord et Sud en possèdent 8.

3.7.4 On jette deux dés équilibrés. Quelle est la probabilité qu'au moins l'un d'entre eux montre 6, sachant que la somme des deux est i , $i = 2, 3, \dots, 12$?

3.7.5 Une urne contient 6 boules blanches et 9 noires. On en tire 4 sans remise et au hasard. Quelle est la probabilité que les deux premières soient blanches et les deux autres noires?

3.7.6 On considère une urne contenant 12 boules desquelles 8 sont blanches. On tire un échantillon de 4 boules avec remise (respectivement sans remise). Quelle est dans chaque cas la probabilité conditionnelle que la première et la troisième boule soient blanches, sachant que l'échantillon contient exactement 3 boules blanches?

3.7.7 Le roi vient d'une famille de 2 enfants. Quelle est la probabilité que l'autre soit une sœur?

3.7.8 Un couple a deux enfants. Quelle est la probabilité que les deux soient des filles sachant que l'aînée en est une ?

3.7.9 On considère trois urnes. L'urne A contient 2 boules blanches et 4 rouges; l'urne B , 8 blanches et 4 rouges; l'urne C , 1 blanche et 3 rouges. On tire une boule de chacune des urnes. Quelle est la probabilité que la boule tirée de l'urne A soit blanche, si l'on sait que le tirage a livré deux boules blanches exactement?

3.7.10 Lors d'une partie de bridge, Ouest n'a reçu aucun as. Quelle est la probabilité que son partenaire

- n'ait reçu aucun as non plus;
- en ait reçu deux ou plus?
- Quels seraient ces résultats si Ouest avait reçu 1 as lors de la donne?

3.7.11 On choisit trois cartes au hasard et sans remise dans un jeu ordinaire de 52 cartes. Calculer la probabilité que la première carte tirée soit un pique, sachant que les deux dernières en sont?

3.7.12 Une grossesse ectopique a deux fois plus de chance de se développer lorsque la femme enceinte fume que lorsqu'elle est non-fumeuse. Si 32% des femmes en âge de maternité fument, quel pourcentage de femmes ayant une grossesse ectopique sont fumeuses?

3.7.13 98% des bébés survivent à l'accouchement. Cependant, 15% des naissances nécessitent une césarienne et lorsqu'une césarienne est pratiquée, les bébés survivent à 96%. Si une femme enceinte choisie aléatoirement ne fait pas de césarienne, quelle est la probabilité que son bébé survive?

3.7.14 Dans une certaine ville, 36% des familles possèdent un chien et 22% de celles qui ont un chien possèdent aussi un chat. De plus, 30% des familles ont un chat. Quelle est

- la probabilité qu'une famille sélectionnée au hasard possède un chien et un chat;
- la probabilité conditionnelle qu'une famille choisie au hasard possède un chien sachant qu'elle a un chat?

3.7.15 46% des électeurs d'une ville se déclarent indépendants alors que 30% se déclarent libéraux et 24% conservateurs. Lors d'une récente élection locale, 35% des indépendants, 62% des libéraux et 58% des conservateurs ont voté. Un électeur est choisi au hasard. Sachant qu'il a voté lors de l'élection locale, quelle est la probabilité qu'il soit

- indépendant;
- libéral;
- conservateur?
- Quelle fraction d'électeurs a participé à l'élection locale?

3.7.16 48% des femmes et 37% des hommes ayant suivi un programme pour arrêter de fumer sont restés non-fumeurs pendant au moins un an après la fin du programme. Ces personnes organisent une fête pour célébrer leur année sans fumer. Si 62% de tous les gens ayant suivi le programme étaient des hommes,

- quel pourcentage de femmes y aura-t-il à la fête?
- quel pourcentage de gens ayant suivi le programme se rendront à la fête?

3.7.17 52% des élèves d'un collège sont des filles. 5% des élèves de ce collège sont doués en informatique. 2% des élèves sont des filles douées en informatique. Si un élève est choisi au hasard, trouver la probabilité conditionnelle que

- cet élève soit une fille, sachant qu'il est doué en informatique;
- cet élève soit doué en informatique, sachant que c'est une fille.

3.7.18 500 couples mariés actifs ont été sondés sur leur salaire annuel, donnant les résultats suivants.

Femme	Mari	
	Moins de \$25 000	Plus de \$25 000
Moins de \$25 000	212	198
Plus de \$25 000	36	54

Ainsi, par exemple, dans 36 de ces couples la femme gagne plus et le mari moins de \$25 000. Si un des couples est choisi au hasard, quelle est

- la probabilité que le mari gagne moins de \$25 000;
- la probabilité conditionnelle que la femme gagne plus de \$25 000 sachant que son mari gagne plus que cette somme;
- la probabilité conditionnelle que la femme gagne plus de \$25 000 sachant que son mari gagne moins que cette somme?

3.7.19 La probabilité que la batterie d'une voiture neuve fonctionne plus de 10 000 miles est .8, la probabilité qu'elle fonctionne plus de 20 000 miles est .4 et la probabilité qu'elle fonctionne plus de 30 000 miles est .1. Si la batterie d'une voiture neuve fonctionne toujours après 10 000 miles, quelle est la probabilité que

- sa durée de vie dépasse 20 000 miles;
- sa durée de vie supplémentaire dépasse 20 000 miles?

3.7.20 Supposons qu'on divise aléatoirement un jeu ordinaire de 52 cartes (contenant 4 as) en 4 mains de 13 cartes chacune. Nous voulons déterminer la probabilité p que chaque main contienne un as. Soit E_i l'événement que la i -ième main ait exactement un as. Déterminer $p = P(E_1 E_2 E_3 E_4)$ en utilisant la règle de multiplication.

3.7.21 Une urne contient au départ 5 boules blanches et 7 noires. Chaque fois que l'on tire une boule, on note sa couleur, puis on la réintroduit ainsi que deux nouvelles boules de la même couleur qu'elle.

- Quelle est la probabilité que les deux premières boules tirées soient noires, puis les deux suivantes blanches?
- Quelle est la probabilité que deux exactement des 4 premières boules tirées soient noires?

3.7.22 Une urne I contient 2 boules blanches et 4 rouges, tandis qu'une urne II contient une boule de chacune de ces couleurs. Une boule est tirée au hasard de l'urne I et placée dans l'urne II, puis on tire une boule de cette dernière urne.

- Quelle est la probabilité que cette deuxième boule soit blanche?
- Quelle est la probabilité que la boule transférée soit blanche, sachant que la dernière boule était blanche?

3.7.23 Comment placer 20 boules, dont 10 sont blanches et 10 noires, dans deux urnes de manière à maximiser la probabilité de tirer une boule blanche dans l'expérience suivante: on choisit d'abord une urne au hasard, puis une boule dans cette urne?

3.7.24 On peint deux boules, soit en noir soit en rouge, au hasard; chaque boule est peinte indépendamment de l'autre, le noir ayant une chance sur deux d'être utilisé.

Les deux boules sont placées dans une urne.

- a) On parvient à apprendre que la peinture rouge a été utilisée, donc qu'au moins une des boules est rouge. Calculer la probabilité que dans ce cas les deux boules soient rouges.
- b) L'urne se renverse et une boule rouge en sort. Quelle est alors la probabilité que les deux boules soient rouges? Expliquer ce résultat.

3.7.25 On a utilisé la méthode suivante pour estimer le nombre de personnes de plus de 50 ans, dans une ville dont la population s'élève à 100 000 âmes. Elle consiste, pour l'expérimentateur, à enregistrer le pourcentage des gens de plus de 50 ans, lors de ses déplacements dans la rue. L'expérience s'étend sur quelques jours. Discuter cette méthode.

A titre d'indication, soit p la vraie proportion des gens de plus de 50 ans dans la ville considérée. De plus, α_1 désigne la proportion sur le temps total du temps qu'une personne de 50 ans ou plus passe dans la rue, α_2 étant cette proportion pour les moins de 50 ans. Quelle est la grandeur que la méthode utilisée estime? Dans quelles conditions l'estimation convient-elle pour p ?

3.7.26 On admet que 5 % des hommes et 0,25 % des femmes sont daltoniens. On sélectionne une personne daltonienne au hasard. Quelle est la probabilité qu'il s'agisse d'un homme ? On admettra que les hommes sont aussi nombreux que les femmes. Si au contraire il y en avait deux fois plus que de femmes, que deviendrait le résultat?

3.7.27 On considère deux boîtes, l'une contient une bille noire et une blanche, et l'autre deux noires et une blanche. On désigne une boîte au hasard, de laquelle on tire une bille. Quelle est la probabilité qu'elle soit noire? Si l'on sait que la bille est blanche, quelle est la probabilité que ce soit la première boîte qui ait été désignée?

3.7.28 Les Anglais et les Américains orthographient le mot *rigueur*, respectivement, *rigour* et *rigor*. Un homme ayant pris une chambre dans un hôtel parisien a écrit ce mot sur un bout de papier. Une lettre est prise au hasard dans ce mot, c'est une voyelle. Or 40% des anglophones de l'hôtel sont des Anglais et les 60% restants sont Américains. Quelle est la probabilité que l'auteur du mot soit anglais?

3.7.29 Deux urnes A et B contiennent respectivement deux boules blanches plus une noire et une blanche plus cinq noires. On tire au hasard une boule dans l'urne A et on la place dans B . On tire alors une boule de B , elle est blanche. Quelle est la probabilité que la boule transférée ait aussi été blanche?

3.7.30 Dans l'exemple 3.12, on doit admettre que le nouvel indice est sujet à différentes interprétations et n'apporte qu'une certitude à 90 % que le criminel possède la caractéristique mentionnée. Quelle est alors la probabilité que le suspect soit coupable (on admettra comme dans l'exemple qu'il porte cette caractéristique)?

3.7.31 Une classe d'étudiants en probabilité comprend 30 étudiants, dont 15 sont bons, 10 moyens et 5 mauvais. Une seconde classe de même effectif compte 5 bons étudiants, 10 moyens et 15 mauvais. L'examineur connaît cette situation, à la fin de l'année, mais ignore à quelle classe il a affaire. Il interroge un étudiant pris au

hasard dans chaque classe et constate que l'étudiant de la classe A est moyen, tandis que l'autre est mauvais. Quelle est la probabilité que la classe A soit la meilleure?

3.7.32 On compte respectivement 50, 75 et 100 employés dans trois entrepôts A , B et C , les proportions de femmes étant respectivement également 50, 60 et 70 pour cent. Une démission a autant de chance de se produire chez tous les employés, indépendamment de leur sexe. Une employée donne sa démission. Quelle est la probabilité qu'elle vienne de l'entrepôt C ?

3.7.33 Un joueur professionnel garde dans sa poche 2 pièces, l'une normale et l'autre ayant ses deux faces identiques, disons deux fois pile. Il en prend une au hasard et la lance; elle montre pile.

- Quelle est la probabilité qu'il s'agisse de la pièce normale?
- Il jette la même pièce une seconde fois, elle montre de nouveau pile. Que devient la probabilité précédente?
- Il la lance une 3ème fois, mais obtient face cette fois. Quelle est maintenant la probabilité qu'il s'agisse de la pièce normale?

3.7.34 Une urne A contient 5 boules blanches et 7 noires. L'urne B en contient 3 et 12 respectivement. On jette par ailleurs une pièce de monnaie équilibrée. Dans le cas où pile sort, on tire une boule de l'urne A , tandis qu'on en prend une dans l'urne B si la pièce montre face. Or, une boule blanche est tirée. Quelle est la probabilité que le jet qui a précédé le tirage de cette boule ait donné face?

3.7.35 Dans l'exemple 3.8, quelle est la probabilité qu'une personne ait un accident durant la deuxième année, sachant qu'elle n'en a pas eu lors de la première?

3.7.36 On considère un échantillon de taille 3 tiré de la manière suivante: on dispose au départ d'une urne contenant 5 boules blanches et 7 rouges. A chaque tirage, une boule est tirée et sa couleur enregistrée. La boule est alors réintroduite dans l'urne ainsi qu'une nouvelle boule de même couleur. Trouver la probabilité que l'échantillon comprenne précisément

- 0 boule blanche;
- 1 boule blanche;
- 3 boules blanches;
- 2 boules blanches.

3.7.37 Une urne contient b boules blanches et r rouges. L'une de ces boules est tirée au hasard. Quand on la remet dans l'urne, on l'accompagne de c nouvelles boules de la même couleur qu'elle. On tire une deuxième boule. Montrer que la probabilité pour la première boule tirée d'être blanche, sachant que la deuxième est rouge, est $b/(b + r + c)$.

3.7.38 On mélange bien un paquet de 52 cartes puis on le divise en deux parties égales. On choisit une carte de l'une des moitiés, qui se trouve être un as. On place cet as dans le second paquet qui est alors mélangé. On tire alors de ce paquet augmenté une carte. Calculer la probabilité que cette carte soit un as. Conditionner sur le fait que la carte qui a changé de paquet est ou n'est pas tirée.

3.7.39 Trois cuisiniers A , B et C sont chacun capables de préparer une spécialité de gâteau. Ce gâteau doit être cuit et risque de ne pas monter avec des probabilités de 0,02, 0,03 et 0,05 selon les cuisiniers. Dans le restaurant où ils travaillent, A cuit 50% de ces gâteaux, B 30% et C 20%. Quelle est la proportion des gâteaux ratés attribuables à A ?

3.7.40 Une boîte contient 3 pièces. La première est normale, la seconde porte deux fois pile et la troisième est biaisée de telle manière que pile sorte trois fois sur 4. Une pièce est tirée puis lancée et donne pile. Quelle est la probabilité qu'il s'agisse de celle aux deux faces identiques?

3.7.41 Le geôlier informe trois prisonniers que l'un d'entre eux a été choisi au hasard pour être exécuté, tandis que les deux autres seront libérés. Le prisonnier A lui demande de lui dire discrètement lequel de ses camarades d'infortune sera libéré, prétendant qu'il n'y a pas de mal à communiquer cette information puisqu'il sait déjà qu'au moins l'un des deux sera libéré. Le geôlier refuse, argumentant que si A sait lequel de ses camarades va être libéré, la probabilité que lui-même soit le condamné augmentera de $\frac{1}{3}$ à $\frac{1}{2}$, car il saura alors qu'il est parmi les deux personnes encore menacées. Que pensez-vous du raisonnement du geôlier?

3.7.42 On dispose de 10 pièces telles que, pour la i -ème d'entre elles, la probabilité de montrer pile lorsqu'on la lance est $i/10$, $i = 1, 2, \dots, 10$. Une pièce est tirée au hasard, lancée, elle donne pile. Quelle est la probabilité qu'il s'agisse de la cinquième pièce?

3.7.43 Une urne contient 5 boules blanches et 10 noires. Le chiffre donné par le jet d'un dé équilibré indique le nombre de boules que l'on va tirer de l'urne. Quelle est la probabilité que toutes les boules tirées soient blanches? Si toutes les boules sont blanches, quelle est la probabilité que le dé ait livré un 3?

3.7.44 Chacun de deux petits meubles identiques a deux tiroirs. Le meuble A contient une pièce d'argent dans chaque tiroir, le meuble B ayant une pièce d'argent dans un tiroir et une en or dans l'autre. On désigne l'un des petits meubles au hasard, ouvre l'un de ses tiroirs et y trouve une pièce d'argent.

- Quelle est la probabilité qu'il y ait une pièce d'argent dans l'autre tiroir?
- Quelle est la probabilité que ce soit le meuble A qui ait été choisi?

3.7.45 On suppose qu'un test de dépistage de cancer est caractérisé par une fiabilité de 95% aussi bien pour ceux qui portent que ceux qui n'ont pas cette maladie. Dans la population, 0,4% des gens ont le cancer. Quelle est la probabilité qu'une personne ait le cancer, sachant que son test l'indique?

3.7.46 Une compagnie d'assurances répartit les gens en trois classes: personnes à bas risque, risque moyen et haut risque. Ses statistiques indiquent que la probabilité que des gens soient impliqués dans un accident sur une période d'un an est respectivement 0,05, 0,15 et 0,30. On estime que 20% de la population est à bas risque, 50% est à risque moyen et 30% à haut risque. Quelle proportion des gens ont un accident ou plus au cours d'une année donnée? Si l'assuré A n'a pas eu d'accident en 1972, quelle

est la probabilité qu'il fasse partie de la classe à bas risque (respectivement à moyen risque)?

3.7.47 Si vous aviez à construire un modèle mathématique autour des événements E et F suivants, les supposeriez-vous indépendants? Expliquer votre choix.

- E est l'événement «une femme d'affaires a les yeux bleus» et F «sa secrétaire a les yeux bleus»;
- E est l'événement «un professeur possède une voiture» et F «il figure dans le bottin du téléphone»;
- E est «un homme mesure moins de 1,75 m» et F «il pèse plus de 100 kg»;
- E est «une femme vit aux Etats-Unis» et F «elle vit dans l'hémisphère occidental» (longitude 180° – 360°);
- E est «il pleuvra demain» et F «il pleuvra après-demain».

3.7.48 Une classe compte 4 garçons et 6 filles de première année, 6 garçons de seconde année. Combien doit-il y avoir de filles de deuxième année si l'on veut que sexe et année soient des facteurs indépendants lors du choix au hasard d'un étudiant?

3.7.49 Supposons qu'on collecte des coupons de m types différents. Supposons aussi qu'à chaque fois qu'on obtient un nouveau coupon, il est de type i avec probabilité p_i , $i = 1, \dots, m$. Supposons qu'on vienne de collecter le n -ième coupon. Quelle est la probabilité qu'il soit d'un nouveau type? Pour cela, conditionner sur le type de ce coupon.

3.7.50 Un modèle simplifié pour la variation du prix d'un stock suppose que chaque jour, le prix du stock grimpe d'une unité avec probabilité p ou chute d'une unité avec probabilité $1 - p$. Les variations quotidiennes sont indépendantes.

- Quelle est la probabilité qu'après 2 jours, le stock soit au prix original?
- Quelle est la probabilité qu'après 3 jours, le prix du stock ait augmenté d'une unité?
- Sachant qu'après 3 jours le prix du stock a augmenté d'une unité, quelle est la probabilité qu'il ait grimpé le premier jour?

3.7.51 La couleur des yeux d'une personne est déterminée par une unique paire de gènes. Si les deux sont des gènes yeux bleus, la personne aura les yeux bleus; si les deux sont des gènes yeux marrons, la personne aura les yeux marrons; si l'un est un gène œil bleu et l'autre un gène œil marron, la personne aura les yeux marrons. (A cause du fait que le gène œil marron est dominant par rapport au gène œil bleu.) Un nouveau-né reçoit indépendamment un gène œil de chacun de ses parents et le gène qu'il reçoit d'un de ses parents a autant de chances d'être l'un des deux gènes œil de ce parent. Supposons que Smith et ses deux parents ont les yeux marrons, mais que la sœur de Smith a les yeux bleus.

- Quelle est la probabilité que Smith ait un gène œil bleu?
Supposons que la femme de Smith a les yeux bleus.
- Quelle est la probabilité que leur premier enfant ait les yeux bleus?
- Si leur premier enfant a les yeux marrons, quelle est la probabilité que leur prochain enfant ait aussi les yeux marrons?

3.7.52 Barbara et Dianne vont faire du tir. Supposons que chaque tir de Barbara touche la cible avec une probabilité p_1 tandis que chaque tir de Dianne la touche avec une probabilité p_2 . Supposons qu'elles tirent ensemble sur la même cible. Si la cible est touchée, quelle est la probabilité que

- les deux tirs l'aient touchée?
- Barbara l'ait touchée?

Quelle hypothèse d'indépendance avez-vous faite?

3.7.53 A et B se battent en duel. Les règles du duel sont les suivantes. Ils ramassent leur pistolet et se tirent dessus simultanément. Si l'un ou l'autre est touché, le duel est fini. Si les deux tirs sont manqués, ils répètent le processus. Supposons que les résultats des tirs sont indépendants, que chaque tir de A touche B avec une probabilité p_A et que chaque tir de B touche A avec une probabilité p_B . Quelle est

- la probabilité que A ne soit pas touché;
- la probabilité que les deux duellistes soient touchés;
- la probabilité que le duel s'arrête après le n -ième tir;
- la probabilité conditionnelle que le duel s'arrête après le n -ième tir sachant que A n'est pas touché;
- la probabilité conditionnelle que le duel s'arrête après le n -ième tir sachant que les deux duellistes sont touchés?

3.7.54 Une question «vrai ou faux» est posée à un couple lors d'un jeu. Le mari et la femme donneront, indépendamment, la bonne réponse avec une probabilité p . Parmi les deux stratégies suivantes, quelle est la meilleure pour le couple?

- Choisir l'un d'eux et le laisser répondre à la question; ou
- Considérer la question tous les deux et alors soit donner la réponse commune s'ils sont d'accord, soit, s'ils ne sont pas d'accord, lancer une pièce pour déterminer la réponse à donner.

3.7.55 Dans le problème 54, si $p = .6$ et que le couple utilise la stratégie de la partie

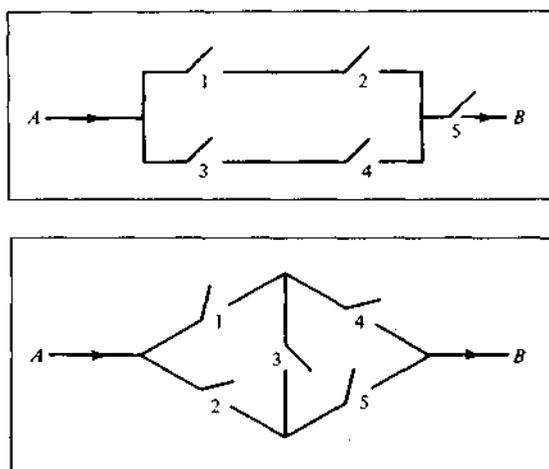
-), quelle est la probabilité conditionnelle que le couple donne la bonne réponse s'ils
- a) sont d'accord;
- b) ne sont pas d'accord?

3.7.56 $M. Jones$ a élaboré une stratégie pour gagner à la roulette: il ne mise que sur rouge et seulement si les dix numéros sortis précédemment ont été noirs. Etant donné la rareté des séquences de 11 numéros noirs, il pense que ses chances de gagner sont grandes. Que pensez-vous de sa stratégie?

3.7.57 On dit qu'un système mécanique est un système k sur n s'il faut et suffit que k des n composants du système soient fonctionnels pour que le système entier le soit. On admettra que les composants travaillent indépendamment les uns des autres. On désigne par P_i la probabilité que le i -ème composant soit opérationnel.

- Calculer la probabilité qu'un système 2 sur 4 fonctionne;
- même question pour un système 3 sur 5;
- même question pour un système k sur n où $P_i = p, \forall i = 1, 2, \dots, n$.

3.7.58 Dans les schémas qui suivent la probabilité que le relais i soit fermé est p_i , $i = 1, 2, 3, 4, 5$. Les relais fonctionnent indépendamment les uns des autres. Quel est, dans ce cas, la probabilité que le courant passe entre A et B ?



Conditionner sur le fait que le relais 3 est fermé ou non.

3.7.59 Un organisme possède 5 paires de gènes; dans chaque paire, les deux gènes sont presque identiques, aussi les désigne-t-on les deux par la même lettre, prise parmi les cinq premières de l'alphabet. Les deux formes possibles de chaque gène ne seront distinguées que par le détail suivant: le gène dominant sera écrit en majuscule, le gène récessif en minuscule. Un gène X est dit dominant si, lorsqu'un organisme possède la paire xX , son apparence extérieure est commandée par X . Par exemple, si X commande la couleur brune pour les yeux et x la couleur bleue, une personne ayant XX ou Xx aura les yeux bruns, tandis que seule une personne portant xx aura les yeux bleus. L'apparence extérieure est qualifiée de phénotype, tandis que la configuration génétique est dite génotype. Ainsi deux organismes porteurs des génotypes aA, bB, cc, dD, ee et AA, BB, cc, DD, ee sont pourtant du même phénotype.

Lors de l'accouplement de deux individus, chacun apporte, au hasard, un gène d'un certain type parmi les deux qu'il possède. Les cinq contributions d'un organisme du genre considéré sont indépendantes entre elles et indépendantes de celles de son partenaire. Consécutivement à l'accouplement de deux organismes de génotypes aA, bB, cC, dD, eE et aa, bb, cc, Dd, ee , quelle est la probabilité que leur progéniture soit (1) phénotypiquement et (2) génotypiquement identique

- au premier parent;
- au second;
- à l'un des deux parents;
- à aucun des deux parents.

3.7.60 La reine porte le gène de l'hémophilie avec une probabilité de 0.5. Si elle est porteuse, chaque prince aura une chance sur deux de souffrir de cette maladie. La

reine a eu trois fils non hémophiles. Quelle est la probabilité qu'elle soit porteuse du gène? S'il naît un quatrième prince, avec quelle probabilité sera-t-il hémophile?

3.7.61 Le 30 septembre 1982 au matin, les scores des trois meilleures équipes de baseball de la division occidentale de la Ligue Nationale des Etats-Unis étaient les suivants:

Equipe	Victoires	Défaites
Atlanta Braves	87	72
San Francisco Giants	86	73
Los Angeles Dodgers	86	73

Chaque équipe doit encore disputer 3 parties. Les Giants doivent jouer leurs trois parties contre les Dodgers, tandis que les Braves joueront les leurs contre les Padres de San Diego. On admettra que les résultats de ces jeux sont indépendants et que toutes les équipes ont la même probabilité de gagner. Quelle est la probabilité pour chacune de ces trois équipes d'emporter la première place de la division? Si deux équipes se retrouvent à égalité pour cette place, elles disputent une unique partie qui les départagera, chacune ayant une chance sur deux de la gagner.

3.7.62 Le conseil municipal d'une ville, composé de 7 membres, comprend un comité directeur de 3 membres. Un projet de loi est d'abord examiné par le comité directeur, puis par le conseil complet si au moins deux des trois membres du comité l'ont approuvé. Devant le conseil complet, le projet de loi doit être approuvé à la majorité (au moins 4 voix) pour prendre force de loi. On considère un projet et admet que chaque conseiller s'exprimera indépendamment des autres et en faveur du projet avec une probabilité p . Quelle est la probabilité que le choix d'un membre du comité directeur soit décisif, ce qui signifie que si ce membre change d'avis, le sort du projet change aussi? Qu'en est-il pour le choix d'un conseiller ne siégeant pas au comité directeur?

3.7.63 On admet que le sexe du dernier enfant d'un couple est indépendant de celui des autres enfants de la famille et qu'il y a autant de chances d'être masculin que féminin. Calculer, pour un couple ayant 5 enfants, les probabilités des événements suivants:

- tous les enfants sont du même sexe;
- les trois aînés sont des garçons, les deux autres des filles;
- il y a exactement 3 garçons;
- les deux aînés sont des garçons;
- il y a au moins une fille.

3.7.64 La probabilité de gagner lors du jet d'un seul dé est p . Le joueur A commence, puis passe le dé à B s'il n'a pas gagné. B joue à son tour et rend le dé à A si lui non plus n'a pas gagné. Les deux joueurs alternent ainsi tant qu'aucun n'a encore gagné. Quelles sont leurs probabilités de victoire respectives? Que devient ce résultat lorsqu'on admet k joueurs?

3.7.65 Mêmes questions que dans le problème 3.7.64 en admettant cependant que A gagne avec probabilité P_1 tandis que cette probabilité est P_2 pour B .

3.7.66 Chacun des trois joueurs lancent une pièce simultanément. La pièce lancée par A , respectivement (B) et (C) montre pile avec probabilité P_1 , (P_2), [P_3]. Si l'un des joueurs obtient un résultat différent de celui commun aux deux autres, il est exclu. Si personne n'est exclu, le jeu recommence jusqu'à ce qu'enfin quelqu'un le soit. Quelle est la probabilité que ce soit A qui se voie exclure?

3.7.67 Dans une expérience donnée, les événements E et F sont mutuellement exclusifs. Montrer que lors d'une suite d'épreuves indépendantes basée sur cette expérience, E apparaîtra avant F avec probabilité $P(E)/[P(E) + P(F)]$.

3.7.68 Une ligne est tracée sur le sol et deux joueurs jettent leur pièce dans sa direction. Celui qui parvient le plus près gagne un sou de la part de l'autre. Le joueur A débute avec 3 sous contre 7 au joueur B . Quelle est la probabilité que A finisse avec tous les sous si aucun n'est plus habile que l'autre? Qu'en serait-il si A était un joueur plus adroit gagnant 6 fois sur 10?

3.7.69 Quelle est la probabilité d'obtenir deux sommes de 7 avant que n'apparaissent 6 nombres pairs lorsqu'on jette plusieurs fois une paire de dés équilibrés?

3.7.70 On considère 2^n joueurs, tous de la même force, que l'on organise en paires tirées au hasard. A l'issue du premier tour de jeu, on compose à nouveau des paires au hasard avec les 2^{n-1} gagnants, et ainsi de suite, jusqu'à ce qu'il n'en reste qu'un. Considérons deux participants donnés, A et B , ainsi que les événements A_i et E définis comme suit:

A_i : « A joue exactement i fois», $i = 1, 2, \dots, n$,

E : « A et B ne jouent jamais l'un contre l'autre».

Trouver:

a) $P(A_i)$, $i = 1, \dots, n$.

b) $P(E)$.

c) On pose $P_n = P(E)$. Montrer que

$$P_n = \frac{1}{2^n - 1} + \frac{2^n - 2}{2^n - 1} \left(\frac{1}{2}\right)^2 P_{n-1}$$

et utiliser ce dernier résultat pour vérifier la réponse donnée en b).

Pour ce faire, trouver $P(E)$ en conditionnant sur l'événement A_i , $i = 1, \dots, n$ qui se produit. Pour simplifier votre réponse, utiliser l'identité algébrique

$$\sum_{i=1}^{n-1} ix^{i-1} = \frac{1 - nx^{n-1} + (n-1)x^n}{(1-x)^2}$$

Pour résoudre ce problème selon une autre approche, noter qu'il y a un total de $2^n - 1$ matchs joués.

d) Expliquer pourquoi $2^n - 1$ matchs sont joués au total.

Numéroter ces matchs et soit B_i l'événement que A et B jouent l'un contre l'autre au i -ième match, $i = 1, \dots, 2^n - 1$.

e) Quelle est $P(B_i)$?

f) Utiliser la question e) pour trouver $P(E)$.

3.7.71 Un spéculateur travaillant sur le marché boursier possède des actions cotées 25. Il a décidé de vendre ses titres si la cote tombe à 10 ou monte au-delà de 40. Les variations de prix de l'action sont de 1 point, avec probabilité 0,55 vers le haut et 0,45 vers le bas. Ces variations dans le temps sont indépendantes. Quelle est la probabilité que cet investisseur fasse une bonne affaire?

3.7.72 Deux joueurs A et B jettent une pièce. A commence et continue jusqu'à l'apparition d'un face. B prend alors la place de A et continue également jusqu'à obtenir face, et ainsi de suite. Désignons par P_1 la probabilité d'obtenir pile pour A , cette probabilité devenant P_2 pour B . Le gagnant du jeu est, dans quatre versions différentes, le premier qui obtiendra

a) deux piles de suite

b) un total de deux piles

c) trois piles de suite

d) un total de 3 piles.

Dans chaque version donner la probabilité que A soit le gagnant.

3.7.73 Un dé A a quatre faces rouges et deux blanches, tandis qu'un dé B en a deux rouges et quatre blanches. Une pièce équilibrée est lancée une fois. Si pile sort, le jeu continue avec le dé A , tandis que si c'est face, on utilise B .

a) Montrer que la probabilité qu'une face rouge apparaisse est $\frac{1}{2}$.

b) Si les deux premiers jets de dé donnent rouge, quelle est la probabilité que le 3ème en fasse autant?

c) Si les deux premiers jets donnent rouge, quelle est la probabilité que l'on soit en train d'utiliser le dé A ?

3.7.74 Dans une urne, on dispose 12 boules dont 4 sont blanches. Les trois joueurs A , B , et C tirent dans l'ordre une boule chacun, puis A recommence et ainsi de suite. Le gagnant est le premier à tirer une boule blanche. Trouver la probabilité pour chaque joueur de gagner dans les deux cas suivants: avec remise, sans remise.

3.7.75 Refaire le problème 3.7.74 lorsqu'il y a trois urnes de 12 boules au lieu d'une, chaque joueur tirant toujours dans la même.

3.7.76 Soit $S = \{1, 2, \dots, n\}$ et supposer que A et B ont, indépendamment, autant de chance d'être l'un des 2^n sous-ensembles de S (comprenant l'ensemble vide et S lui-même). Montrer que

$$P\{A \subset B\} = \left(\frac{3}{4}\right)^n$$

Pour cela, soit $N(B)$ le nombre d'éléments dans B . Utiliser

$$P\{A \subset B\} = \sum_{i=0}^n P\{A \subset B \mid N(B) = i\} P\{N(B) = i\}$$

3.7.77 On se place dans la situation de l'exemple 3.31. Quelle est la probabilité que la i -ème pièce ait été choisie, sachant que les n premières épreuves ont toutes donné pile?

3.7.78 Dans le cas de la loi de succession de Laplace, exemple 3.31, peut-on dire que les résultats des différents jets sont indépendants? Expliquer votre réponse.

3.7.79 Un tribunal de 3 juges déclare un individu coupable lorsque deux au moins des trois juges estiment que cette décision est fondée. On admettra que si l'accusé est effectivement coupable, chaque juge se prononcera dans ce sens avec probabilité 0,7, ceci indépendamment des 2 autres. Cette probabilité tombe à 0,2 dans le cas où l'accusé est innocent. 70% des accusés sont coupables. Calculer la probabilité que le juge n° 3 vote coupable dans chacune des situations suivantes:

- les juges 1 et 2 l'ont fait;
- les juges 1 et 2 sont partagés;
- les juges 1 et 2 votent tous deux non coupable.

On désigne par E_i , $i = 1, 2, 3$ l'événement «le juge i vote coupable». Ces événements sont-ils indépendants? Conditionnellement indépendants? Expliquer votre position.

Variables aléatoires

4.1 VARIABLES ALÉATOIRES

4.1.1 Définition

Après avoir réalisé une expérience, il arrive bien souvent qu'on s'intéresse plus à une fonction du résultat qu'au résultat lui-même. Expliquons ceci au moyen des exemples suivants: lorsqu'on joue aux dés, certains jeux accordent de l'importance à la somme obtenue sur deux dés, 7 par exemple, plutôt qu'à la question de savoir si c'est la paire (1,6) qui est apparue, ou (2,5), (3,4), (4,3), (5,2) ou plutôt (6,1). Dans le cas du jet d'une pièce, il peut être plus intéressant de connaître le nombre de fois où pile est apparu plutôt que la séquence détaillée des piles et faces. Ces grandeurs auxquelles on s'intéresse sont en fait des fonctions réelles définies sur l'ensemble fondamental et sont appelées *variables aléatoires*.

Du fait que la valeur d'une variable aléatoire est déterminée par le résultat de l'expérience, il est possible d'attribuer une probabilité aux différentes valeurs que la variable aléatoire peut prendre.

4.1.2 Exemples de variables aléatoires

Exemple 4.1 Notre expérience consiste à jeter trois pièces équilibrées. Si l'on désigne le nombre de piles par Y , Y est une variable aléatoire et peut prendre les valeurs 0, 1, 2, 3 avec pour probabilité respectivement

$$P\{Y = 0\} = P\{(P, P, P)\} = \frac{1}{8}$$

$$P\{Y = 1\} = P\{(P, P, F), (P, F, P), (F, P, P)\} = \frac{3}{8}$$

$$P\{Y = 2\} = P\{(P, F, F), (F, P, F), (F, F, P)\} = \frac{3}{8}$$

$$P\{Y = 3\} = P\{(F, F, F)\} = \frac{1}{8}$$

Du fait que Y doit nécessairement prendre l'une des valeurs 0, 1, 2, 3 on aura

$$1 = P\left(\bigcup_{i=0}^3 \{Y = i\}\right) = \sum_{i=0}^3 P\{Y = i\}$$

ce qui est évidemment confirmé par les probabilités calculées ci-dessus. ■

Exemple 4.2 D'une urne contenant 20 boules numérotées de 1 à 20, on tire sans remplacement 3 des boules. Quelqu'un parie qu'au moins une des boules tirées portera un numéro égal ou supérieur à 17. Quelle est la probabilité qu'il gagne ?

SOLUTION. Disons que X représente le plus grand numéro tiré. X est une variable aléatoire pouvant prendre les valeurs 3, 4, ..., 19 ou 20. En supposant que les $\binom{20}{3}$ tirages sont tous équiprobables, on a:

$$P\{X = i\} = \frac{\binom{i-1}{2}}{\binom{20}{3}} \quad i = 3, \dots, 20 \quad (4.1)$$

L'équation (4.1) s'explique par le fait que l'événement $\{X = i\}$ correspond au tirage de la boule i et de deux des boules portant les numéros 1 à $i - 1$. Or il y a clairement $\binom{i-1}{2}$ de ces tirages, d'où le résultat. Grâce à cette équation on peut calculer

$$P\{X = 20\} = \frac{\binom{19}{2}}{\binom{20}{3}} = \frac{3}{20} = .150$$

$$P\{X = 19\} = \frac{\binom{18}{2}}{\binom{20}{3}} = \frac{51}{380} = .134$$

$$P\{X = 18\} = \frac{\binom{17}{2}}{\binom{20}{3}} = \frac{34}{285} = .119$$

$$P\{X = 17\} = \frac{\binom{16}{2}}{\binom{20}{3}} = \frac{2}{19} = .105$$

Or, $\{X \geq 17\}$ est la réunion disjointe des événements $\{X = i\}$, $i = 17, 18, 19, 20$. La probabilité de gagner le pari est donc

$$P\{X \geq 17\} = .105 + .119 + .134 + .150 = .508 \quad \blacksquare$$

Exemple 4.3 On répète le jet d'une pièce jusqu'à ce que face apparaisse, mais au plus n fois. Les jets sont indépendants et face apparaît avec probabilité p . X désigne le nombre de jets réalisés jusqu'à l'arrêt de l'expérience. C'est donc une variable aléatoire et elle prendra les valeurs $1, 2, 3, \dots, n$ avec les probabilités respectives suivantes:

$$\begin{aligned} P\{X = 1\} &= P\{F\} = p \\ P\{X = 2\} &= P\{(P, F)\} = (1 - p)p \\ P\{X = 3\} &= P\{(P, P, F)\} = (1 - p)^2 p \\ &\vdots \\ P\{X = n - 1\} &= P\{(\underbrace{P, P, \dots, P}_{n-2}, F)\} = (1 - p)^{n-2} p \\ P\{X = n\} &= P\{(\underbrace{P, P, \dots, P}_{n-1}, P), (\underbrace{P, P, \dots, P}_{n-1}, F)\} = (1 - p)^{n-1} \end{aligned}$$

A titre de vérification, on remarquera que

$$\begin{aligned} P\left(\bigcup_{i=1}^n \{X = i\}\right) &= \sum_{i=1}^n P\{X = i\} \\ &= \sum_{i=1}^{n-1} p(1-p)^{i-1} + (1-p)^{n-1} \\ &= p \left[\frac{1 - (1-p)^{n-1}}{1 - (1-p)} \right] + (1-p)^{n-1} \\ &= 1 - (1-p)^{n-1} + (1-p)^{n-1} \\ &= 1 \end{aligned}$$

Exemple 4.4 D'une urne contenant 3 boules blanches, 3 rouges et 5 noires, on tire 3 boules. Supposons que l'on reçoive 1 franc pour chaque boule blanche tirée et que l'on doive au contraire payer 1 franc pour toute boule rouge. On désigne le bénéfice net laissé par le tirage par X . X est une variable aléatoire pouvant prendre les valeurs $0, \pm 1, \pm 2, \pm 3$ avec pour probabilités respectives

$$\begin{aligned} P\{X = 0\} &= \frac{\binom{5}{3} + \binom{3}{1}\binom{3}{1}\binom{5}{1}}{\binom{11}{3}} = \frac{55}{165} \\ P\{X = 1\} = P\{X = -1\} &= \frac{\binom{3}{1}\binom{5}{2} + \binom{3}{2}\binom{3}{1}}{\binom{11}{3}} = \frac{39}{165} \end{aligned}$$

$$P\{X = 2\} = P\{X = -2\} = \frac{\binom{3}{2}\binom{5}{1}}{\binom{11}{3}} = \frac{15}{165}$$

$$P\{X = 3\} = P\{X = -3\} = \frac{\binom{3}{3}}{\binom{11}{3}} = \frac{1}{165}$$

Pour expliquer comment on obtient ces probabilités, prenons le cas $X = 0$. On remarque qu'il faut pour cela tirer uniquement des boules noires ou autrement une boule de chaque couleur. Pour $X = 1$ il faudra tirer 1 boule blanche et 2 noires ou 2 blanches et 1 rouge. A titre de vérification, on peut s'assurer que

$$\sum_{i=0}^3 P\{X = i\} + \sum_{i=1}^3 P\{X = -i\} = \frac{55 + 39 + 15 + 1 + 39 + 15 + 1}{165} = 1$$

La probabilité de gagner de l'argent est

$$\sum_{i=1}^3 P\{X = i\} = \frac{55}{165} = \frac{1}{3} \quad \blacksquare$$

Exemple 4.5 On collectionne des coupons. Il y a N sortes de coupons et lorsqu'on en obtient un, il a autant de chances d'être d'une sorte plutôt que d'une autre, indépendamment de ceux qu'on a pu se procurer auparavant. La variable aléatoire T qui nous intéresse est celle qui compte le nombre de coupons qu'il est nécessaire de réunir pour obtenir une collection complète comprenant au moins un coupon de chaque type. Mais plutôt que de calculer $P\{T = n\}$ directement, commençons par calculer la probabilité que T soit plus grand que n . Pour cela, on fixe n et on définit les événements suivants: A_j est «le type j n'est pas représenté parmi les n premiers coupons rassemblés», $j = 1, 2, \dots, N$. Alors,

$$\begin{aligned} P\{T > n\} &= P\left(\bigcup_{j=1}^N A_j\right) \\ &= \sum_j P(A_j) - \sum_{j_1 < j_2} P(A_{j_1} A_{j_2}) + \dots \\ &\quad + (-1)^{k+1} \sum_{j_1 < j_2 < \dots < j_k} P(A_{j_1} A_{j_2} \dots A_{j_k}) \dots \\ &\quad + (-1)^{N+1} P(A_1 A_2 \dots A_N) \end{aligned}$$

Or, A_j ne se produira que si chacun des n coupons n'est pas du type j . Comme cette probabilité est $(N-1)/N$ pour chaque coupon, notre hypothèse d'indépendance sur les types obtenus successivement nous permet d'écrire

$$P(A_j) = \left(\frac{N-1}{N}\right)^n$$

Par ailleurs, l'événement $A_{j_1 j_2}$ ne se produira que si aucun des n premiers coupons n'est du type j_1 ni du type j_2 . La même hypothèse d'indépendance permet d'écrire

$$P(A_{j_1} A_{j_2}) = \left(\frac{N-2}{N}\right)^n$$

Le même raisonnement donne

$$P(A_{j_1} A_{j_2} \cdots A_{j_k}) = \left(\frac{N-k}{N}\right)^n$$

et on voit que, pour $m > 0$,

$$\begin{aligned} P\{T > n\} &= N \left(\frac{N-1}{N}\right)^n - \binom{N}{2} \left(\frac{N-2}{N}\right)^n + \binom{N}{3} \left(\frac{N-3}{N}\right)^n - \cdots \\ &\quad + (-1)^N \binom{N}{N-1} \left(\frac{1}{N}\right)^n \\ &= \sum_{i=1}^{N-1} \binom{N}{i} \left(\frac{N-i}{N}\right)^n (-1)^{i+1} \end{aligned} \quad (4.2)$$

La probabilité que T soit égale à n peut maintenant être déduite de ce qui précède du fait que

$$P\{T > n-1\} = P\{T = n\} + P\{T > n\}$$

ou, de façon équivalente,

$$P\{T = n\} = P\{T > n-1\} - P\{T > n\}$$

Une autre variable aléatoire qui nous intéresse est le nombre de types représentés dans les n premiers coupons rassemblés. Appelons-la D_n . Pour calculer $P\{D_n = k\}$, considérons d'abord un jeu bien particulier de k types et déterminons la probabilité que ce jeu soit celui des types représentés dans les n premiers coupons. Pour que cela soit le cas, il est nécessaire et suffisant que ces coupons vérifient les deux conditions suivantes:

A: chacun des coupons appartient à l'un des k types fixés

B: chacun de ces k types est représenté.

Or, un nouveau coupon sera de l'un de ces k types avec probabilité k/N et de ce fait $P(A) = (k/N)^n$. Par ailleurs, si l'on sait qu'un coupon est de l'un de ces k types, il est facile de voir qu'il a autant de chances d'être de n'importe lequel de ces k types. Aussi la probabilité que *B* soit vrai, sachant que *A* l'est, est-elle simplement la probabilité de l'événement suivant: «un groupe de n coupons contient un jeu complet des k types», chaque coupon pouvant être de n'importe quel type avec la même probabilité. Mais ceci n'est autre que la probabilité qu'il faille n coupons au moins pour former un jeu complet de k types, calculée dans (4.2) à condition de substituer k à N . Aussi avons-nous

$$P(A) = \left(\frac{k}{N}\right)^n$$

$$P(B|A) = 1 - \sum_{i=1}^{k-1} \binom{k}{i} \left(\frac{k-i}{k}\right)^n (-1)^{i+1}$$

Finalement, comme il y a $\binom{N}{k}$ choix possibles du jeu de k types initialement considéré, on obtient

$$P\{D_n = k\} = \binom{N}{k} P(AB)$$

$$= \binom{N}{k} \left(\frac{k}{N}\right)^n \left[1 - \sum_{i=1}^{k-1} \binom{k}{i} \left(\frac{k-i}{k}\right)^n (-1)^{i+1} \right] \quad \blacksquare$$

4.2 FONCTIONS DE RÉPARTITION

4.2.1 Définition

La *fonction de répartition* F d'une variable aléatoire X est définie pour tout nombre réel b , $-\infty < b < \infty$, par

$$F(b) = P\{X \leq b\}$$

En d'autres termes, $F(b)$ est la probabilité que la variable aléatoire X prenne une valeur inférieure ou égale à b .

4.2.2 Propriétés des fonctions de répartition

Voici quelques propriétés de ces fonctions:

- F est une fonction non décroissante; autrement dit si $a < b$, alors $F(a) \leq F(b)$,
- $\lim_{b \rightarrow \infty} F(b) = 1$,
- $\lim_{b \rightarrow -\infty} F(b) = 0$,
- F est continue à droite, c'est-à-dire que, quel que soit b et quelle que soit une suite décroissante b_n , $n \geq 1$ convergeant vers b , on a $\lim_{n \rightarrow \infty} F(b_n) = F(b)$.

La première propriété repose sur le fait que si $a < b$, l'événement $\{X \leq a\}$ est inclus dans $\{X \leq b\}$; la probabilité du premier est donc nécessairement plus petite que celle du second. Quant aux propriétés suivantes, elles résultent toutes de la propriété de continuité des probabilités (section 2.6).

Pour démontrer la deuxième propriété, on remarquera que si b_n tend vers l'infini, les événements $\{X \leq b_n\}$, $n \geq 1$ sont croissants emboîtés et que leur union est $\{X < \infty\}$. Par conséquent, en application de la propriété de continuité mentionnée

$$\lim_{n \rightarrow \infty} P\{X \leq b_n\} = P\{X < \infty\} = 1$$

ce qui établit la deuxième propriété.

La démonstration de la troisième propriété étant similaire, elle est laissée en exercice au lecteur. Quant à celle de la quatrième propriété, on part d'une suite $\{b_n\}$ convergeant vers b et décroissante. $\{X_n \leq b_n\}$, $n \geq 1$ est une suite emboîtée décroissante d'événements dont l'intersection est $\{X \leq b\}$. Toujours par continuité

$$\lim_n P\{X \leq b_n\} = P\{X \leq b\}$$

ce qui établit la quatrième propriété.

4.2.3 Fonction de répartition et probabilités sur X

Tous les calculs de probabilité concernant X peuvent être traités en termes de fonction de répartition. Par exemple,

$$P\{a < X \leq b\} = F(b) - F(a) \quad \text{pour tout } a < b \quad (4.3)$$

On peut s'en rendre mieux compte en écrivant $\{X \leq b\}$ comme union des deux événements mutuellement exclusifs $\{X \leq a\}$ et $\{a < X \leq b\}$, soit

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$$

et ainsi

$$P\{X \leq b\} = P\{X \leq a\} + P\{a < X \leq b\}$$

ce qui établit (4.3).

Pour obtenir $P\{X < b\}$ on peut écrire, en utilisant encore une fois la propriété de continuité

$$\begin{aligned} P\{X < b\} &= P\left(\lim_{n \rightarrow \infty} \left\{X \leq b - \frac{1}{n}\right\}\right) \\ &= \lim_{n \rightarrow \infty} P\left(X \leq b - \frac{1}{n}\right) \\ &= \lim_{n \rightarrow \infty} F\left(b - \frac{1}{n}\right) \end{aligned}$$

On remarquera que $P\{X < b\}$ n'est pas nécessairement égal à $F(b)$ puisque cette valeur comprend également la probabilité $P\{X = b\}$.

4.2.4 Exemple de fonction de répartition

Exemple 4.6 La fonction de répartition de la variable aléatoire X est donnée par

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \leq x < 1 \\ \frac{2}{3} & 1 \leq x < 2 \\ \frac{11}{12} & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

La figure 4.1 représente son graphe.

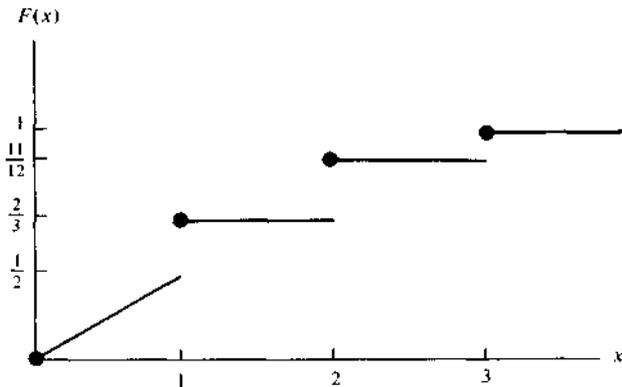


Fig. 4.1

- Calculer
- | | |
|---------------------------|--------------------------|
| a) $P\{X < 3\}$ | b) $P\{X = 1\}$ |
| c) $P\{X > \frac{1}{2}\}$ | d) $P\{2 < X \leq 4\}$. |

SOLUTION.

$$\text{a) } P\{X < 3\} = \lim_n P\left\{X \leq 3 - \frac{1}{n}\right\} = \lim_n F\left(3 - \frac{1}{n}\right) = \frac{11}{12}$$

$$\text{b) } P\{X = 1\} = P\{X \leq 1\} - P\{X < 1\}$$

$$= F(1) - \lim_n F\left(1 - \frac{1}{n}\right) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$$

$$\begin{aligned} \text{c) } P\left\{X > \frac{1}{2}\right\} &= 1 - P\left\{X \leq \frac{1}{2}\right\} \\ &= 1 - F\left(\frac{1}{2}\right) = \frac{3}{4} \end{aligned}$$

$$\begin{aligned} \text{d) } P\{2 < X \leq 4\} &= F(4) - F(2) \\ &= \frac{1}{12} \end{aligned} \quad \blacksquare$$

4.3 VARIABLES ALÉATOIRES DISCRÈTES

4.3.1 Définition, loi de probabilité

Une variable aléatoire ne pouvant prendre qu'une quantité dénombrable de valeurs est dite *discrète*. Pour une telle variable aléatoire X , on définit sa *loi de probabilité* p par

$$p(a) = P\{X = a\}$$

Cette loi de probabilité ne peut être positive que pour un ensemble au plus dénombrable d'arguments. En d'autres termes, si X peut prendre les valeurs x_1, x_2, \dots , alors

$$\begin{aligned} p(x_i) &\geq 0 & i = 1, 2, \dots \\ p(x) &= 0 & \text{pour toutes les autres valeurs de } x \end{aligned}$$

Du fait que X doit bien prendre l'une de ces valeurs x_i , on aura

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

Il est souvent instructif de représenter la fonction de densité de probabilité sur un graphique, en reportant $p(x_i)$ sur l'axe des y et x_i sur l'axe des x . A titre d'illustration, la loi de probabilité suivante est représentée à la figure 4.2:

$$p(0) = \frac{1}{4} \quad p(1) = \frac{1}{2} \quad p(2) = \frac{1}{4}$$

La figure 4.3 représente le graphe de la loi de probabilité d'une variable aléatoire comptant la somme des nombres obtenus lors du jet de deux dés équilibrés.

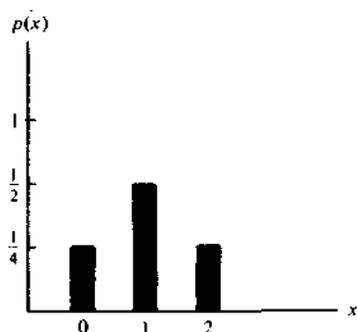


Figure 4.2

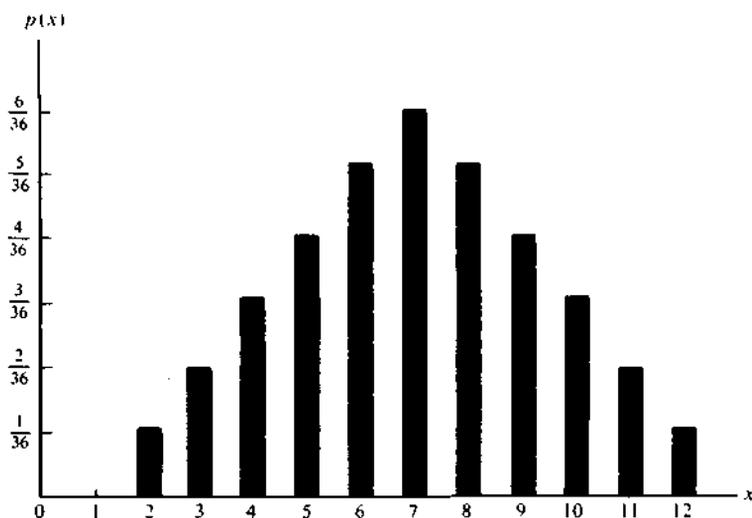


Figure 4.3

Exemple 4.7 La loi de probabilité d'une variable aléatoire X est donnée par $p(i) = c\lambda^i/i!$, $i = 0, 1, 2, \dots$, où λ est un réel positif. Trouver

- $P\{X = 0\}$;
- $P\{X > 2\}$.

SOLUTION. Puisque $\sum_{i=0}^{\infty} p(i) = 1$, nous avons que

$$c \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1$$

ce qui implique, puisque $e^x = \sum_{i=0}^{\infty} x^i/i!$, que

$$ce^{\lambda} = 1 \quad \text{ou} \quad c = e^{-\lambda}$$

Donc

$$(a) \quad P\{X = 0\} = e^{-\lambda} \lambda^0 / 0! = e^{-\lambda}$$

$$(b) \quad P\{X > 2\} = 1 - P\{X \leq 2\} = 1 - P\{X = 0\} - P\{X = 1\} - P\{X = 2\}$$

$$= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}$$

4.3.2 Fonction de répartition d'une variable aléatoire discrète

On peut exprimer la fonction de répartition F d'une variable aléatoire discrète en fonction des valeurs prises par sa loi de probabilité p :

$$F(a) = \sum_{x \leq a} p(x)$$

Dans le cas précis où les valeurs possibles de la variable aléatoire sont x_1, x_2, x_3, \dots , avec $x_1 < x_2 < x_3 < \dots$, la fonction F de répartition est une fonction en escalier. Ses valeurs seront constantes sur les intervalles $[x_{i-1}, x_i)$ et elle aura un saut de taille $p(x_i)$ en x_i , $i = 1, 2, \dots$. Dans le cas par exemple d'une variable aléatoire X dont la loi est donnée par

$$p(1) = \frac{1}{4} \quad p(2) = \frac{1}{2} \quad p(3) = \frac{1}{8} \quad p(4) = \frac{1}{8}$$

sa fonction de répartition sera

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{4} & 1 \leq a < 2 \\ \frac{3}{4} & 2 \leq a < 3 \\ \frac{7}{8} & 3 \leq a < 4 \\ 1 & 4 \leq a \end{cases}$$

Le graphe de cette dernière est représenté à la figure 4.4.

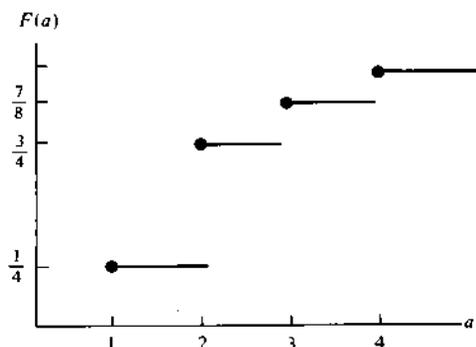


Figure 4.4

Le lecteur remarquera que la taille du saut aux abscisses 1, 2, 3 et 4 est égale à la probabilité que X prenne ces valeurs.

4.4 ESPÉRANCE

4.4.1 Définition et interprétation

L'espérance d'une variable aléatoire est l'un des concepts les plus importants en théorie des probabilités. Pour une variable aléatoire discrète X de loi de probabilité $p(\cdot)$, on définit l'*espérance* de X , notée $E[X]$, par l'expression

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

En termes concrets, l'espérance de X est la moyenne pondérée des valeurs que X peut prendre, les poids étant les probabilités que ces valeurs soient prises. Si, par exemple, la loi de probabilité de X est

$$p(0) = \frac{1}{2} = p(1)$$

alors

$$E[X] = 0\left(\frac{1}{2}\right) + 1\left(\frac{1}{2}\right) = \frac{1}{2}$$

n'est autre que la simple moyenne des deux valeurs 0 et 1 que X peut prendre. Si, par contre,

$$p(0) = \frac{1}{3} \quad p(1) = \frac{2}{3}$$

alors

$$E[X] = 0\left(\frac{1}{3}\right) + 1\left(\frac{2}{3}\right) = \frac{2}{3}$$

sera la moyenne pondérée des deux valeurs possibles 0 et 1, la valeur 1 recevant un poids deux fois plus important que la valeur 0; ceci s'explique du fait que $p(1) = 2p(0)$.

On peut trouver une autre origine de la définition de l'espérance dans l'interprétation des probabilités comme mesures de fréquences relatives. Aux termes de cette interprétation (partiellement étayée par la loi forte des grands nombres qui sera présentée au chapitre 8), la proportion du nombre d'apparitions d'un résultat E dans une séquence infiniment longue d'expériences identiques est $P(E)$. Imaginons maintenant qu'une variable aléatoire X prenne les valeurs x_1, x_2, \dots, x_n avec probabilités respectives $p(x_1), p(x_2), \dots, p(x_n)$. On admettra que X représente des gains nets lors d'un jeu. Le joueur gagnera donc la somme x_i avec probabilité $p(x_i)$, $i = 1, 2, \dots, n$. Selon l'hypothèse de l'interprétation discutée ici, au bout d'un nombre infiniment grand de parties, le joueur aura gagné x_i sur une proportion $p(x_i)$ de l'ensemble des tirages. Ceci étant vrai pour tout i , $i = 1, 2, \dots, n$, le gain moyen par jeu sera

$$\sum_{i=1}^n x_i p(x_i) = E[X]$$

4.4.2 Exemples d'espérances de variables discrètes

Exemple 4.8 On cherche l'espérance $E[X]$ de la variable X , résultat du lancer d'un dé équilibré.

SOLUTION. Comme $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$, on aura

$$E[X] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = \frac{7}{2} \quad \blacksquare$$

Exemple 4.9 I est une variable indicatrice pour l'événement A si

$$I = \begin{cases} 1 & \text{si } A \text{ se produit} \\ 0 & \text{si } A^c \text{ se produit} \end{cases}$$

Trouver $E[I]$.

SOLUTION. Puisque $p(1) = P\{A\}$, $p(0) = 1 - P\{A\}$, on a

$$E[I] = P\{A\}$$

Ceci signifie que l'espérance de la variable indicatrice pour l'événement A est égale à la probabilité que A se produise. ■

Exemple 4.10 On pose deux questions au participant d'un jeu télévisé. Il peut choisir l'ordre dans lequel il va répondre à ces questions, numérotées 1 et 2. S'il répond juste à la première, il est autorisé à continuer avec la seconde, sinon il doit s'arrêter. Il recevra V_i francs pour une bonne réponse à la question i , $i = 1, 2$. Par exemple, s'il répond aux deux questions, il gagnera $V_1 + V_2$ francs. Supposons qu'il connaisse la probabilité P_i , $i = 1, 2$, avec laquelle il répondra juste à la question i . A quelle question doit-il répondre d'abord pour maximiser son gain prospectif? On admet que les deux questions sont indépendantes.

SOLUTION. S'il commence par la question 1, ses gains seront

$$\begin{array}{ll} 0 & \text{avec probabilité } 1 - P_1 \\ V_1 & \text{avec probabilité } P_1(1 - P_2) \\ V_1 + V_2 & \text{avec probabilité } P_1P_2. \end{array}$$

Son gain moyen sera donc

$$V_1P_1(1 - P_2) + (V_1 + V_2)P_1P_2$$

Si, par contre, il commence avec la question 2, le gain moyen devient

$$V_2P_2(1 - P_1) + (V_1 + V_2)P_1P_2$$

Il est alors préférable de commencer par la question 1 si

$$V_1 P_1 (1 - P_2) \geq V_2 P_2 (1 - P_1)$$

ce qui équivaut à

$$\frac{V_1 P_1}{1 - P_1} \geq \frac{V_2 P_2}{1 - P_2}$$

A titre d'illustration, prenons le cas où il a 60 chances sur 100 de répondre juste à la question 1, qui lui rapporte 200 francs, contre 80 chances sur 100 pour la question 2 qui ne lui rapporte que 100 francs en cas de réponse correcte. Il doit ici répondre d'abord à la question 2 car

$$\frac{(200)(.6)}{(.4)} = 300 < \frac{(100)(.8)}{(.2)} = 400 \quad \blacksquare$$

4.4.3 Analogie avec une notion de mécanique

Le concept d'espérance est à rapprocher de la notion de *centre de gravité* d'un groupe de masses, au sens de la mécanique. Considérons en effet une variable X de loi de probabilité $P(x_i)$, $i \geq 1$. On sait que si des masses $P(x_i)$, $i \geq 1$ sont réparties sur une barre sans poids aux abscisses x_i , $i \geq 1$, le point sur lequel la barre pourra être posée et rester en équilibre est appelé centre de gravité (figure 4.5). Il est facile de voir, pour les lecteurs ayant quelques connaissances élémentaires en statique, que l'abscisse du centre de gravité est $E[X]$ ¹.

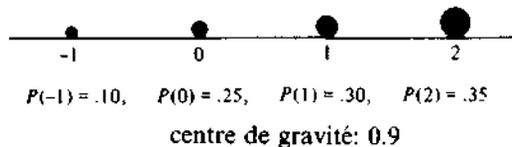


Figure 4.5

4.5 ESPÉRANCE D'UNE FONCTION D'UNE VARIABLE ALÉATOIRE

4.5.1 Première approche

Considérons une variable aléatoire discrète X et sa distribution; il arrive qu'on cherche à calculer l'espérance d'une fonction de X , disons $g(X)$. Comment peut-on s'y prendre? Un moyen est de remarquer que $g(X)$ étant une variable aléatoire discrète

¹ Pour le montrer, il suffit d'établir que la somme des moments des forces gravitationnelles par rapport au point d'abscisse $E[X]$ est 0. En d'autres termes, il suffit de montrer que $0 = \sum_i (x_i - E[X]) P(x_i)$, ce qui est immédiat.

elle aussi, elle doit avoir une distribution que l'on devrait pouvoir déduire de celle de X . Si on a pu déterminer la distribution de $g(X)$, il suffit d'appliquer la définition de l'espérance pour obtenir $E[g(X)]$.

Exemple 4.11 Soit X une variable aléatoire qui prend une des trois valeurs $-1, 0, 1$ avec les probabilités respectives

$$P\{X = -1\} = .2, \quad P\{X = 0\} = .5, \quad P\{X = 1\} = .3$$

Calculer $E[X^2]$.

SOLUTION. Soit $Y = X^2$. La distribution de Y est donnée par

$$P\{Y = 1\} = P\{X = -1\} + P\{X = 1\} = .5$$

$$P\{Y = 0\} = P\{X = 0\} = .5$$

Donc

$$E[X^2] = E[Y] = 1(.5) + 0(.5) = .5$$

Le lecteur remarquera que

$$.5 = E[X^2] \neq (E[X])^2 = .01 \quad \blacksquare$$

4.5.2 Théorème de calcul

Bien que la procédure précédente permette toujours de calculer l'espérance de n'importe quelle fonction de X si on connaît la distribution de X , il existe une autre façon de voir $E[g(X)]$. En effet, en notant que $g(X)$ vaut $g(x)$ lorsque X vaut x , il est raisonnable de penser que $E[g(X)]$ puisse être la moyenne pondérée des valeurs $g(x)$, avec $g(x)$ pondérée par la probabilité que X égale x . Donc, on obtient intuitivement le résultat suivant:

Théorème 4.1

Si X est une variable aléatoire discrète pouvant prendre ses valeurs parmi les valeurs x_i , $i \geq 1$, avec des probabilités respectives $p(x_i)$, alors pour toute fonction réelle g on aura

$$E[g(X)] = \sum_i g(x_i)p(x_i)$$

Avant de donner la démonstration de ce théorème, vérifions rapidement que son application à l'exemple 4.11 confirme bien les résultats trouvés. L'application du théorème à cet exemple donne

$$\begin{aligned}
 E[X^2] &= (-1)^2(.2) + 0^2(.5) + 1^2(.3) \\
 &= 1(.2 + .3) + 0(.5) \\
 &= .5
 \end{aligned}$$

ce qui correspond au résultat obtenu dans l'exemple 4.11.

DÉMONSTRATION. La preuve du théorème 4.1 s'effectue, comme dans la vérification précédente, en groupant dans $\sum g(x_i)p(x_i)$ tous les termes ayant la même valeur de $g(x_i)$. Spécifiquement, supposons que $y_j, j \geq 1$, représente les différentes valeurs de $g(x_i), i \geq 1$. Alors, en groupant tous les $g(x_i)$ ayant la même valeur, on obtient

$$\begin{aligned}
 \sum_i g(x_i)p(x_i) &= \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i) \\
 &= \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i) \\
 &= \sum_j y_j P\{g(X) = y_j\} \\
 &= E[g(X)] \quad \blacksquare
 \end{aligned}$$

4.5.3 Exemple de calcul d'espérance de fonction de variable aléatoire

Exemple 4.12 Un produit de saison rapporte un bénéfice net de b francs par unité vendue mais, inversement, chaque unité invendue à la fin de la saison engendre une perte de d francs. Le nombre X d'unités commandées auprès d'un certain magasin au cours des saisons de vente successives suit une loi de probabilité $p(\cdot)$ à valeurs non négatives. On admet que le magasin doit avoir constitué tout son stock avant la saison. Quelle doit être la taille de ce stock si l'on veut maximiser le résultat net moyen de l'opération?

SOLUTION. Désignons par s la taille du stock. Le résultat net sera noté $P(s)$ et son expression est

$$\begin{aligned}
 P(s) &= bX - (s - X)d && \text{si } X \leq s \\
 &= sb && \text{si } X > s
 \end{aligned}$$

Le résultat moyen de l'opération sera donc

$$\begin{aligned}
 E[P(s)] &= \sum_{i=0}^s [bi - (s - i)d]p(i) + \sum_{i=s+1}^{\infty} sbp(i) \\
 &= (b + d) \sum_{i=0}^s ip(i) - sd \sum_{i=0}^s p(i) + sb \left[1 - \sum_{i=0}^s p(i) \right] \\
 &= (b + d) \sum_{i=0}^s ip(i) - (b + d)s \sum_{i=0}^s p(i) + sb
 \end{aligned}$$

$$= sb + (b + d) \sum_{i=0}^s (i - s)p(i)$$

Pour déterminer la valeur optimale de s , voyons comment varie notre profit lorsque s augmente d'une unité. Par substitution, on obtient:

$$\begin{aligned} E[P(s + 1)] &= b(s + 1) + (b + d) \sum_{i=0}^{s+1} (i - s - 1)p(i) \\ &= b(s + 1) + (b + d) \sum_{i=0}^s (i - s - 1)p(i) \end{aligned}$$

D'où

$$E[P(s + 1)] - E[P(s)] = b - (b + d) \sum_{i=0}^s p(i)$$

Il sera préférable de stocker $s + 1$ unités plutôt que s tant que

$$\sum_{i=0}^s p(i) < \frac{b}{b + d} \quad (4.4)$$

Comme le membre de gauche de (4.4) est croissant pour s tandis que celui de droite est constant, l'inégalité (4.4) sera satisfaite pour toutes les valeurs de s inférieures ou égales à s^* , où s^* est la plus grande des valeurs vérifiant encore (4.4). Ce qui donne

$$E[P(0)] < \dots < E[P(s^*)] < E[P(s^* + 1)] > E[P(s^* + 2)] > \dots$$

On constate que le stockage de $s^* + 1$ unités conduira au résultat net moyen le plus élevé. ■

4.5.4 Linéarité de l'espérance

Le théorème suivant est une conséquence immédiate du théorème 4.1:

Théorème 4.2

Pour toute paire a, b de constantes, on peut écrire

$$E[aX + b] = aE[X] + b$$

DÉMONSTRATION.

$$\begin{aligned} E[aX + b] &= \sum_{x:p(x)>0} (ax + b)p(x) \\ &= a \sum_{x:p(x)>0} xp(x) + b \sum_{x:p(x)>0} p(x) \\ &= aE[X] + b \end{aligned} \quad \blacksquare$$

4.5.5 Autres moments à l'origine

L'espérance d'une variable X , notée $E[X]$, est parfois nommée *premier moment par rapport à l'origine*. La quantité $E[X^n]$, $n \geq 1$ est de manière générale appelée *n-ième moment de X par rapport à l'origine*. D'après le théorème 4.1, on peut calculer ainsi ces moments:

$$E[X^n] = \sum_{x:p(x)>0} x^n p(x)$$

4.6 VARIANCE

4.6.1 Introduction et définition

Une variable X et sa fonction F de répartition étant données, il serait pratique de pouvoir résumer les propriétés de F en deux ou trois mesures bien choisies. L'espérance $E[X]$ est une telle mesure. Cependant, si $E[X]$ nous donne une moyenne pondérée des valeurs possibles de X , elle ne nous dit rien des variations de X autour de l'espérance. On peut s'en rendre compte grâce aux exemples suivants. Soit les variables

$$\begin{aligned} W &= 0 && \text{avec probabilité } 1 \\ Y &= \begin{cases} -1 & \text{avec probabilité } \frac{1}{2} \\ +1 & \text{avec probabilité } \frac{1}{2} \end{cases} \\ Z &= \begin{cases} -100 & \text{avec probabilité } \frac{1}{2} \\ +100 & \text{avec probabilité } \frac{1}{2} \end{cases} \end{aligned}$$

Si toutes ont la même espérance – à savoir 0 –, il y a de bien plus grands écarts entre les différentes valeurs de Y qu'entre celle de W (qui est constante) et de plus grands écarts entre celles de Z qu'entre celles de Y .

Comme on s'attend à voir toute variable X prendre ses valeurs autour de son espérance $E[X]$, il paraît raisonnable de mesurer les variations de X en considérant l'écart moyen entre X et son espérance. Cela reviendrait à s'intéresser à la grandeur $E[|X - \mu|]$, où $\mu = E[X]$. Techniquement, cependant, il n'est pas facile de manipuler cette quantité, aussi lui préfère-t-on d'habitude l'espérance du carré de l'écart entre X et son espérance. On appelle *variance* de X , que l'on note $\text{Var}(X)$, la quantité

$$\text{Var}(X) = E[(X - \mu)^2]$$

où μ représente l'espérance de X .

On peut établir une autre formule pour le calcul de $\text{Var}(X)$ en procédant ainsi:

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - E[2\mu X] + E[\mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned}$$

ce qui revient à écrire

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

ou à dire que la variance de X peut être calculée en soustrayant à l'espérance de X^2 le carré de l'espérance de X . Dans la pratique, cette méthode de calcul est en général plus commode.

4.6.2 Exemple de calcul de variance

Exemple 4.13 On cherche $\text{Var}(X)$ où X est le nombre obtenu lors du jet d'un dé équilibré.

SOLUTION. On a vu dans l'exemple 4.8 que $E[X] = \frac{7}{2}$. De plus

$$\begin{aligned} E[X^2] &= 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{6}\right) + 3^2\left(\frac{1}{6}\right) + 4^2\left(\frac{1}{6}\right) + 5^2\left(\frac{1}{6}\right) + 6^2\left(\frac{1}{6}\right) \\ &= \left(\frac{1}{6}\right)(91) \end{aligned}$$

Et donc

$$\text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \quad \blacksquare$$

4.6.3 Propriété de la variance

Pour toute paire (a, b) de constantes, on peut établir l'identité suivante, fort utile:

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

Pour démontrer ceci, on utilise le résultat du théorème 4.2, à savoir que $E[aX + b] = a E[X] + b$. Donc

$$\begin{aligned} \text{Var}(aX + b) &= E[(aX + b - (aE[X] + b))^2] \\ &= E[(aX - aE[X])^2] \\ &= E[a^2(X - E[X])^2] \\ &= a^2 E[(X - E[X])^2] \\ &= a^2 \text{Var}(X) \end{aligned}$$

4.6.4 Interprétation grâce à une analogie tirée de la mécanique

De même que l'espérance avait été comparée au centre de gravité d'un ensemble de masses, la variance peut être rapprochée du concept mécanique de moment d'inertie (par rapport à l'espérance).

4.6.5 Ecart-type

La racine carrée de $\text{Var}(X)$ est appelée l'écart-type de X , qui se note σ . On a

$$\sigma = \sqrt{\text{Var}(X)}$$

Les variables aléatoires discrètes sont souvent réparties en catégories selon le type de leur loi. Les sections suivantes présentent quelques-uns de ces types.

4.7 VARIABLE DE BERNOULLI ET VARIABLE BINOMIALE

4.7.1 Variable de Bernoulli

On réalise une expérience dont le résultat sera interprété soit comme un succès soit comme un échec. On définit alors la variable aléatoire X en lui donnant la valeur 1 lors d'un succès et 0 lors d'un échec. La loi de probabilité de X est alors

$$\begin{aligned} p(0) &= P\{X = 0\} = 1 - p \\ p(1) &= P\{X = 1\} = p \end{aligned} \quad (4.5)$$

où p est la probabilité d'un succès, $0 \leq p \leq 1$.

Une variable aléatoire X est dite *de Bernoulli* (du nom du mathématicien suisse Jacques Bernoulli) s'il existe un nombre $p \in (0,1)$ tel que la loi de probabilité de X soit donnée par (4.5).

4.7.2 Variables binomiales

Supposons qu'on exécute maintenant n épreuves indépendantes, chacune ayant p pour probabilité de succès et $1 - p$ pour probabilité d'échec. La variable aléatoire X qui compte le nombre de succès sur l'ensemble des n épreuves est dite variable aléatoire *binomiale* de paramètres (n, p) . Une variable de Bernoulli n'est donc qu'une variable binomiale de paramètres $(1, p)$.

La loi de probabilité d'une variable aléatoire binomiale de paramètres (n, p) est donnée par

$$p(i) = \binom{n}{i} p^i (1 - p)^{n-i} \quad i = 0, 1, \dots, n \quad (4.6)$$

Pour établir (4.6) il faut tout d'abord remarquer que toute séquence donnée comportant i succès et $n - i$ échecs pour une longueur totale de n épreuves a pour probabilité $p^i(1 - p)^{n-i}$, en vertu de l'indépendance de ces épreuves. Comme il y a $\binom{n}{i}$ de ces séquences comptant i succès et $n - i$ échecs, on aboutit bien à (4.6). On peut le voir encore plus facilement si l'on remarque qu'il y a $\binom{n}{i}$ choix différents des i épreuves donnant un succès. Plaçons-nous par exemple dans le cas où $n = 4$ et $i = 2$. Il y a bien $\binom{4}{2} = 6$ manières d'obtenir deux succès parmi les 4 résultats, à savoir (s, s, e, e) , (s, e, s, e) , (s, e, e, s) , (e, s, s, e) , (e, s, e, s) et (e, e, s, s) . Par (s, s, e, e) on veut dire que les deux premières épreuves ont donné des succès, au contraire des deux dernières. Chacune de ces séquences ayant pour probabilité $p^2(1 - p)^2$, la probabilité cherchée est bien $\binom{4}{2} p^2(1 - p)^2$.

On remarquera qu'en application du théorème du binôme, la somme de tous les $p(i)$ est 1:

$$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^n \binom{n}{i} p^i (1-p)^{n-i} = [p + (1-p)]^n = 1$$

4.7.3 Exemples de variables binomiales et de Bernoulli

Exemple 4.14 On jette cinq pièces équilibrées. Les résultats sont supposés indépendants. Donner la loi de probabilité de la variable X qui compte le nombre de piles obtenus.

SOLUTION. Soit X le nombre de piles (donc de succès) au total. X est une variable aléatoire binomiale de paramètres ($n = 5, p = \frac{1}{2}$). Aussi a-t-on, en application de (4.6),

$$P\{X = 0\} = \binom{5}{0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

$$P\{X = 1\} = \binom{5}{1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4 = \frac{5}{32}$$

$$P\{X = 2\} = \binom{5}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = \frac{10}{32}$$

$$P\{X = 3\} = \binom{5}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = \frac{10}{32}$$

$$P\{X = 4\} = \binom{5}{4} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^1 = \frac{5}{32}$$

$$P\{X = 5\} = \binom{5}{5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^0 = \frac{1}{32} \quad \blacksquare$$

Exemple 4.15 On sait que les vis fabriquées par une certaine société sont affectées d'un défaut avec probabilité 0,01; l'état d'une vis est indépendant de celui des précédentes ou suivantes. Or, la société accepte de rembourser les paquets de 10 vis qu'elle vend si plus d'une des vis présente un défaut. Quelle proportion des paquets vendus la société s'expose-t-elle à devoir rembourser?

SOLUTION. Désignons par X le nombre de vis malformées d'un paquet donné. X est une variable aléatoire binomiale de paramètres (10, 0,01). La probabilité qu'il faille remplacer un paquet est

$$\begin{aligned} 1 - P\{X = 0\} - P\{X = 1\} &= 1 - \binom{10}{0} (.01)^0 (.99)^{10} - \binom{10}{1} (.01)^1 (.99)^9 \\ &\approx .004 \end{aligned}$$

Ainsi ne faudra-t-il remplacer que 0,4 pour cent seulement des paquets. ■

Exemple 4.16 Le jeu d'argent décrit ci-dessous est appelé «roue de la fortune» et est très populaire lors de bien des carnivals et dans les casinos; un joueur parie et mise sur un numéro compris entre 1 et 6 inclusivement. On jette ensuite trois dés. Si le nombre choisi par le joueur apparaît i fois ($i = 1, 2, 3$), celui-ci gagne i unités. Dans le cas où ce nombre n'apparaît pas, le joueur perd une unité. Ce jeu est-il honnête vis-à-vis du joueur? (En fait, on joue en lançant une roue qui s'immobilise en laissant apparaître un jeu de trois nombres compris entre 1 et 6, mais tout revient du point de vue mathématique à jeter trois dés).

SOLUTION. Admettons que les dés sont équilibrés et que leurs résultats sont indépendants les uns des autres. Le nombre de fois qu'apparaît le nombre sur lequel le joueur a misé est une variable aléatoire binomiale de paramètres $(3, \frac{1}{6})$. Désignons par X les gains du joueur lors d'une partie. On aura

$$P\{X = -1\} = \binom{3}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^3 = \frac{125}{216}$$

$$P\{X = 1\} = \binom{3}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^2 = \frac{75}{216}$$

$$P\{X = 2\} = \binom{3}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^1 = \frac{15}{216}$$

$$P\{X = 3\} = \binom{3}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^0 = \frac{1}{216}$$

Pour déterminer si le jeu est équilibré, déterminons $E[X]$. D'après les probabilités précédentes, on obtient

$$\begin{aligned} E[X] &= \frac{-125 + 75 + 30 + 3}{216} \\ &= \frac{-17}{216} \end{aligned}$$

Ainsi, sur un nombre infini de parties, le joueur perdra 17 unités par groupe de 216 parties. ■

Dans l'exemple suivant nous allons étudier une version simple de la théorie de l'hérédité développée par Gregor Mendel (1822–1884).

Exemple 4.17 On admet qu'un trait physique (telle la couleur des yeux ou le fait d'être gaucher) chez un homme est déterminé par une paire de gènes. On désignera par d le gène de la paire qui est dominant, et par r celui qui est récessif. Une personne portant dd sera ainsi à dominance pure, une autre portant rr sera à caractère récessif, alors que rd entraînera une dominance hybride. Les dominances pure et hybride ne se distinguent pas extérieurement. Un enfant recevra un gène de chacun de ses parents.

Si, par un trait particulier, les deux parents sont hybrides et s'ils ont 4 enfants, quelle est la probabilité que 3 de ceux-ci manifestent extérieurement le trait dominant?

SOLUTION. Admettons que chaque enfant a autant de chances de recevoir chacun des deux gènes de chacun de ses parents. Les probabilités que l'enfant de deux parents hybrides porte les gènes dd , rr ou rd sont respectivement $\frac{1}{4}$, $\frac{1}{4}$ et $\frac{1}{2}$. Comme un descendant aura le trait dominant s'il porte les paires de gènes dd ou rd , le nombre d'enfants ainsi conformés est réparti selon la loi binomiale avec pour paramètres $(4, \frac{3}{4})$ dans notre cas. La probabilité cherchée est donc

$$\binom{4}{3} \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right)^1 = \frac{27}{64}$$

■

Exemple 4.18 On considère un jugement pour lequel la condamnation doit être votée par 8 des 12 membres du jury au moins pour devenir exécutoire. En admettant que les jurés se déterminent indépendamment les uns des autres et que la probabilité de décision correcte est θ pour chacun d'entre eux, quelle est la probabilité que la décision du jury entier soit correcte?

SOLUTION. Il est impossible de donner une solution au problème tel qu'il est énoncé ci-dessus, par manque d'information. Si l'accusé est innocent par exemple, la probabilité que le jury rende une sentence correcte est

$$\sum_{i=5}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$$

alors que s'il est coupable cette probabilité devient

$$\sum_{i=8}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$$

On pourra ainsi obtenir la probabilité que le jury se détermine correctement en conditionnant selon que l'accusé est coupable ou innocent, α désignant la probabilité qu'il soit coupable:

$$\alpha \sum_{i=8}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i} + (1-\alpha) \sum_{i=5}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$$

■

Exemple 4.19 Un système de communication comporte n composants; chacun d'entre eux fonctionnera, indépendamment des autres, avec une probabilité p . Le système total pourra effectivement fonctionner si au moins la moitié de ses composants sont opérationnels.

- Pour quelles valeurs de p un système à 5 composants est-il plus souvent en état de fonctionnement que celui à 3 composants?
- De manière générale, dans quel cas un système à $2k + 1$ composants est-il préférable à un système à $2k - 1$ composants?

SOLUTION.

a) Comme le nombre de composants en bon état est une variable aléatoire binomiale de paramètres (n, p) , la probabilité qu'un système à 5 composants fonctionne est

$$\binom{5}{3} p^3 (1-p)^2 + \binom{5}{4} p^4 (1-p) + p^5$$

tandis que la probabilité correspondante pour un système à 3 composants est

$$\binom{3}{2} p^2 (1-p) + p^3.$$

Par conséquent, le système à 5 composants est préférable si

$$10p^3 (1-p)^2 + 5p^4 (1-p) + p^5 > 3p^2 (1-p) + p^3$$

qui se réduit à

$$3(p-1)^2 (2p-1) > 0$$

ou

$$p > 1/2.$$

b) En général, un système à $2k+1$ composants est préférable à un système à $2k-1$ composants si (et seulement si) $p \geq 1/2$. Pour le montrer, considérons un système à $2k+1$ composants et notons X le nombre de ceux qui fonctionnent parmi les $2k-1$ premiers composants. Alors

$$P_{2k+1} \text{ (le système fonctionne)} = P\{X \geq k+1\} + P\{X = k\} (1 - (1-p)^2) \\ + P\{X = k-1\} p^2$$

car le système à $2k+1$ composants fonctionnera dans les cas suivants:

- $X \geq k+1$,
- $X = k$ et au moins l'un des 2 composants restants fonctionne ou
- $X = k-1$ et tous les deux composants suivants fonctionnent.

Comme

$$P_{2k-1} \text{ (le système fonctionne)} = P\{X \geq k\} \\ = P\{X = k\} + P\{X \geq k+1\}$$

on obtient que

$$P_{2k+1} \text{ (le système fonctionne)} - P_{2k-1} \text{ (le système fonctionne)} \\ = P\{X = k-1\} p^2 - (1-p)^2 P\{X = k\} \\ = \binom{2k-1}{k-1} p^{k-1} (1-p)^k p^2 - (1-p)^2 \binom{2k-1}{k} p^k (1-p)^{k-1} \\ = \binom{2k-1}{k} p^k (1-p)^k [p - (1-p)] \text{ puisque } \binom{2k-1}{k-1} = \binom{2k-1}{k} \\ > 0 \Leftrightarrow p > \frac{1}{2}. \quad \blacksquare$$

4.7.4 Propriétés des variables aléatoires binomiales

Nous allons examiner à présent les propriétés d'une variable aléatoire binomiale de paramètres n et p . Pour commencer, calculons son espérance et sa variance.

$$\begin{aligned} E[X^k] &= \sum_{i=0}^n i^k \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=1}^n i^k \binom{n}{i} p^i (1-p)^{n-i} \end{aligned}$$

En utilisant l'identité

$$i \binom{n}{i} = n \binom{n-1}{i-1}$$

on obtient

$$\begin{aligned} E[X^k] &= np \sum_{i=1}^n i^{k-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \\ &= np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \quad \text{avec } j = i-1 \\ &= np E[(Y+1)^{k-1}] \end{aligned}$$

où Y est une variable aléatoire binomiale de paramètres $n-1, p$. En posant $k=1$ dans l'équation précédente, on obtient

$$E[X] = np$$

Autrement dit, le nombre espéré de succès lors de n épreuves indépendantes où chacune a une probabilité p de succès est égal à np . En posant $k=2$ dans cette même équation et en utilisant la formule précédente pour l'espérance d'une variable aléatoire binomiale, on obtient

$$\begin{aligned} E[X^2] &= np E[Y+1] \\ &= np[(n-1)p+1] \end{aligned}$$

Comme $E[X] = np$, on obtient

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= np[(n-1)p+1] - (np)^2 \\ &= np(1-p) \end{aligned}$$

En résumé, nous avons montré que:

Si X est une variable aléatoire binomiale de paramètres n et p , alors

$$E[X] = np$$

$$\text{Var}(X) = np(1-p)$$

La proposition suivante montre la façon dont la distribution binomiale croît d'abord puis décroît.

Théorème 4.3

Soit X une variable aléatoire binomiale de paramètres (n, p) avec $0 < p < 1$. Lorsque k croît de 0 à n , $P(X = k)$ grandit d'abord de manière monotone, puis décroît également de manière monotone, le pic étant atteint lorsque k est égal à la partie entière de $(n + 1)p$.

DÉMONSTRATION. On démontre ce théorème en considérant le rapport $P\{X = k\} / P\{X = k - 1\}$ et en étudiant les valeurs de k pour lesquelles il est plus grand ou plus petit que 1. Or

$$\begin{aligned} \frac{P\{X = k\}}{P\{X = k - 1\}} &= \frac{\frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}}{\frac{n!}{(n-k+1)!(k-1)!} p^{k-1} (1-p)^{n-k+1}} \\ &= \frac{(n-k+1)p}{k(1-p)} \end{aligned}$$

Donc $P\{X = k\} \geq P\{X = k - 1\}$ si et seulement si

$$(n - k + 1)p \geq k(1 - p)$$

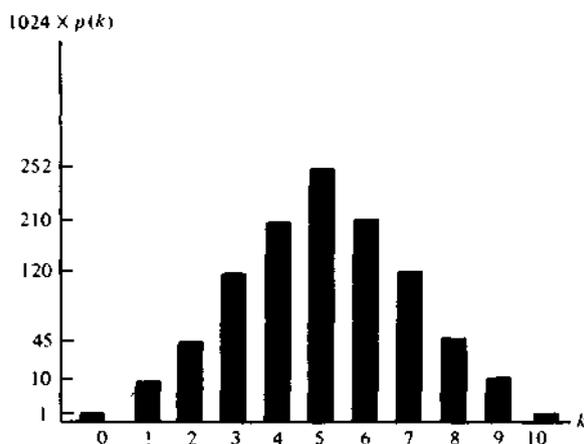


Figure 4.6 Graphe de $p(k) = \binom{10}{k} \left(\frac{1}{2}\right)^{10}$

ou de façon équivalente, si et seulement si

$$k \leq (n + 1)p$$

Le théorème est ainsi démontré. ■

En illustration à ce théorème, on donne à la figure 4.6 le graphe de la loi d'une variable aléatoire binomiale ayant pour paramètres $(10, \frac{1}{2})$.

4.7.5 Application de la loi binomiale

Exemple 4.20 Lors d'une élection présidentielle américaine, le candidat qui recueille le plus grand nombre de voix dans un Etat emporte la totalité de celles du collège électoral allouées à cet Etat¹. Ce nombre de voix est approximativement proportionnel à la population de cet Etat, ce qui revient à dire qu'un Etat de population n dispose d'environ nc voix au collège électoral. (En fait, il dispose plutôt d'environ $nc + 2$ voix puisqu'il reçoit une voix par député qu'il envoie à la Chambre des Représentants – et le nombre de ces députés est proportionnel à la population de cet Etat – plus une voix par sénateur – et chaque Etat dispose de 2 sénateurs –.) Nous allons déterminer la puissance de vote moyenne d'un citoyen établi dans un Etat de population n lors d'une élection présidentielle très disputée; par puissance de vote moyenne lors d'une élection serrée, on entend le produit ncP , P étant la probabilité que l'électeur arbitre le scrutin dans son Etat. Il arbitrera le scrutin si les $n - 1$ autres électeurs de l'Etat ont attribué exactement $(n - 1)/2$ voix à chacun des deux candidats. Ceci suppose que le nombre total n d'électeurs dans cet Etat soit impair, mais le cas où n est pair peut être traité de manière très similaire. Le fait que l'élection soit serrée se traduit par les hypothèses que les $n - 1$ autres électeurs se déterminent indépendamment les uns des autres, avec probabilité $\frac{1}{2}$ en faveur de chaque candidat. De ce fait, la probabilité qu'un électeur établi dans un Etat de taille $n = 2k + 1$ puisse arbitrer le scrutin est égale à la probabilité que lors d'une séquence de $2k$ jets d'une pièce équilibrée il sorte aussi souvent pile que face, à savoir

$$P = P(\text{l'électeur soit arbitre dans un Etat de taille } 2k + 1)$$

$$\begin{aligned} &= \binom{2k}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^k \\ &= \frac{(2k)!}{k! k! 2^{2k}} \end{aligned}$$

On obtient une approximation de cette quantité grâce à la *formule de Stirling* qui affirme que lorsque k est grand

$$k! \sim k^{k+1/2} e^{-k} \sqrt{2\pi}$$

¹ ndt: l'élection présidentielle américaine est organisée sur un mode particulier de suffrage indirect, avec en général deux candidats seulement.

où $a_k \sim b_k$ signifie que le rapport a_k/b_k tend vers 1 lorsque k tend vers l'infini. De ce fait

$$P = P(\text{l'électeur soit arbitre dans un Etat de taille } 2k + 1)$$

$$\sim \frac{(2k)^{2k+1/2} e^{-2k} \sqrt{2\pi}}{k^{2k+1} e^{-2k} (2\pi) 2^{2k}} = \frac{1}{\sqrt{k\pi}}$$

Comme cet électeur fait basculer nc votes du collège électoral s'il arbitre l'élection dans son Etat, sa puissance moyenne de vote ncP est approximativement

$$\begin{aligned} &\sim \frac{nc}{\sqrt{n\pi/2}} \\ &= c\sqrt{2n/\pi} \end{aligned}$$

Cette puissance moyenne de vote étant proportionnelle à la racine carrée de n , l'exemple traité montre que les électeurs de grands Etats sont plus influents que ceux des plus petits. ■

NOTE HISTORIQUE

Les épreuves indépendantes ayant une probabilité p de succès ont été étudiées pour la première fois par le mathématicien suisse Jacques Bernoulli (1654 - 1705). Dans son livre «Ars Conjectandi» (l'Art de la Conjecture), publié en 1713 par son neveu Nicholas huit ans après sa mort, Bernoulli montre que si on réalise un grand nombre de telles épreuves, la proportion de celles où un succès se produit est proche de p avec une probabilité proche de 1.

Jacques Bernoulli fait partie de la première génération de la famille de mathématiciens la plus célèbre de tous les temps. Somme toute, entre huit et douze Bernoulli, répartis sur trois générations, apportèrent des contributions fondamentales aux probabilités, statistiques et mathématiques. Une difficulté pour connaître leur nombre exact provient du fait que plusieurs d'entre eux portaient le même prénom. (Par exemple, deux des fils de Jean, le frère de Jacques, s'appelaient Jacques et Jean.) Une autre difficulté provient du fait que plusieurs Bernoulli étaient connus sous des noms différents à des endroits différents. Notre Jacques (parfois écrit Jaques) était, par exemple, aussi connu sous les noms de Jakob (parfois écrit Jacob) et de James Bernoulli. Mais quel que soit leur nombre, leur influence et leur production sont prodigieuses. Comme les Bach pour la musique, les Bernoulli resteront pour l'éternité une grande famille pour les mathématiques!

4.7.6 Calcul de la fonction de répartition binomiale

Supposons que X est une variable aléatoire binomiale de paramètres (n, p) . L'idée clé pour le calcul de sa fonction de répartition

$$P\{X \leq i\} = \sum_{k=0}^i \binom{n}{k} p^k (1-p)^{n-k}, \quad i = 0, 1, \dots, n$$

consiste à utiliser la relation suivante entre $P\{X = k + 1\}$ et $P\{X = k\}$, formule qui a été établie lors de la démonstration du théorème 4.3:

$$P\{X = k + 1\} = \frac{p}{1-p} \frac{n-k}{k+1} P\{X = k\} \quad (4.7)$$

Exemple 4.21 Soit X une variable aléatoire binomiale de paramètres $n = 6$, $p = 0,4$. En partant de $P\{X = 0\} = (0,6)^6$ et en utilisant l'équation (4.7) de façon récursive, on obtient

$$\begin{aligned} P\{X = 0\} &= (.6)^6 \approx .0467 \\ P\{X = 1\} &= \frac{4}{3} P\{X = 0\} \approx .1866 \\ P\{X = 2\} &= \frac{4}{3} \frac{5}{2} P\{X = 1\} \approx .3110 \\ P\{X = 3\} &= \frac{4}{3} \frac{4}{3} P\{X = 2\} \approx .2765 \\ P\{X = 4\} &= \frac{4}{3} \frac{2}{3} P\{X = 3\} \approx .1382 \\ P\{X = 5\} &= \frac{4}{3} \frac{1}{3} P\{X = 4\} \approx .0369 \\ P\{X = 6\} &= \frac{4}{3} \frac{1}{6} P\{X = 5\} \approx .0041. \quad \blacksquare \end{aligned}$$

On peut facilement écrire un programme qui utilise la formule de récurrence (4.7) pour calculer la fonction de répartition binomiale. Pour trouver $P\{X \leq i\}$, le programme calcule d'abord $P\{X = i\}$ puis utilise la récurrence pour calculer successivement $P\{X = i-1\}$, $P\{X = i-2\}$, etc. Le calcul de

$$P\{X = i\} = \frac{n(n-1)\dots(n-i+1)}{i!} p^i (1-p)^{n-i}$$

peut s'effectuer en prenant d'abord les logarithmes pour obtenir

$$\ln(P\{X = i\}) = \sum_{k=1}^i \ln(n+1-k) - \sum_{k=1}^i \ln(k) + i \ln(p) + (n-i) \ln(1-p)$$

puis en prenant

$$P\{X = i\} = \exp\{\ln(P\{X = i\})\}$$

Exemple 4.22

- Déterminer $P\{X \leq 145\}$ où X est une variable aléatoire binomiale de paramètres (250, 0,5).
- Déterminer $P\{X \leq 90\}$ où X est une variable aléatoire binomiale de paramètres (1000, 0,1).

SOLUTION. Exécuter le programme de la distribution binomiale:

```

RUN
THE DISTRIBUTION FUNCTION OF A BINOMIAL(n,p) RANDOM VARIABLE
ENTER n
? 250
ENTER p
? .5
ENTER i
? 145
THE PROBABILITY IS .995255
OK

```

```

RUN
THE DISTRIBUTION FUNCTION OF A BINOMIAL(n,p) RANDOM VARIABLE
ENTER n
? 1000
ENTER p
? .1
ENTER i
? 90
THE PROBABILITY IS .1582189
OK

```

4.8 VARIABLE ALÉATOIRE DE POISSON

4.8.1 Définition

Une variable aléatoire X pouvant prendre pour valeur $0, 1, 2, \dots$ est dite *de Poisson* avec paramètre λ s'il existe un réel $\lambda > 0$ tel que

$$p(i) = P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!} \quad i = 0, 1, 2, \dots \quad (4.8)$$

L'équation (4.8) définit bien une loi de probabilité puisque

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1$$

La distribution poissonnienne fut introduite par Siméon Denis Poisson dans un ouvrage traitant des applications de la théorie des probabilités aux problèmes juridiques tels que des procès, des jugements en matière criminelle, etc. Son livre, publié en 1837, était intitulé *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*.¹

¹ ndt: titre en français dans le texte anglais.

4.8.2 Approximation poissonnienne de lois binomiales

Les variables aléatoires de Poisson ont un champ d'application fort vaste, en particulier du fait qu'on peut les utiliser pour approximer des variables aléatoires binomiales de paramètres (n, p) pour autant que n soit grand et p assez petit pour que np soit d'ordre de grandeur moyen. Pour s'en convaincre, admettons que X soit une variable aléatoire binomiale de paramètres (n, p) et posons $\lambda = np$. On aura

$$\begin{aligned} P\{X = i\} &= \frac{n!}{(n-i)! i!} p^i (1-p)^{n-i} \\ &= \frac{n!}{(n-i)! i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1) \cdots (n-i+1)}{n^i} \frac{\lambda^i (1 - \lambda/n)^n}{i! (1 - \lambda/n)^i} \end{aligned}$$

Maintenant, pour n grand et λ modéré

$$\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda} \quad \frac{n(n-1) \cdots (n-i+1)}{n^i} \approx 1 \quad \left(1 - \frac{\lambda}{n}\right)^i \approx 1$$

Donc, pour n grand et λ modéré,

$$P\{X = i\} \approx e^{-\lambda} \frac{\lambda^i}{i!}$$

En d'autres termes, lorsqu'on réalise n épreuves indépendantes ayant p pour probabilité de succès et si n est grand et p assez petit pour rendre np moyen, le nombre de succès est une variable aléatoire de répartition approximativement poissonnienne avec paramètre $\lambda = np$. La détermination de cette grandeur λ sera en général empirique. On montrera plus tard qu'elle représente d'ailleurs un nombre moyen de succès.

4.8.3 Applications de la loi de Poisson

On cite ci-dessous quelques exemples de variables aléatoires qui obéissent en règle générale à la loi de probabilité de Poisson (c'est-à-dire qui satisfont (4.8)):

1. le nombre de coquilles par page ou groupe de pages d'un livre
2. le nombre d'individus dépassant l'âge de 100 ans dans une communauté humaine
3. le nombre de faux numéros téléphoniques composés en un jour
4. le nombre de paquets de biscuits pour chien vendus dans un magasin donné en l'espace d'un jour
5. le nombre de clients pénétrant dans un bureau de poste donné en l'espace d'un jour
6. le nombre de charges devenues vacantes à la Cour suprême en l'espace d'un an
7. le nombre de particules α émises par un matériau radioactif pendant un certain laps de temps.

Dans chacun de ces exemples – et dans bien d'autres – la variable aléatoire est toujours répartie de manière approximativement poissonnienne pour la même raison:

parce qu'on approxime par là une variable binomiale. Dans le premier cas par exemple, on peut supposer que chacun des caractères composant une page a une probabilité p d'être mal rendu. Aussi le nombre de coquilles par page sera-t-il distribué approximativement suivant la loi de Poisson avec paramètre $\lambda = np$ où n est le nombre de caractères par page. De la même manière, on peut supposer que toute personne dans une communauté a la même probabilité de devenir centenaire. On peut aussi attribuer une probabilité d'acheter des biscuits pour chien à toute personne entrant dans une épicerie, et ainsi de suite.

4.8.4 Exemples de variables aléatoires de Poisson

Exemple 4.23 Admettons que le nombre d'erreurs par page dans ce livre suive une distribution de Poisson avec paramètre $\lambda = \frac{1}{2}$. Calculer la probabilité qu'il y ait au moins une erreur sur cette page.

SOLUTION. Désignons par X le nombre d'erreurs sur cette page. On aura

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-1/2} \approx .393 \quad \blacksquare$$

Exemple 4.24 On admet que la probabilité de défaut pour un objet fabriqué à la machine est 0,1. Trouver la probabilité qu'un lot de 10 objets comprenne au plus un élément affecté d'un défaut.

SOLUTION. La probabilité cherchée est exactement

$$\binom{10}{0}(0,1)^0(0,9)^{10} + \binom{10}{1}(0,1)^1(0,9)^9 = 0,7361$$

alors que l'approximation donnée par la loi de Poisson mène à $e^{-1} + e^{-1} \approx 0,7358$. ■

Exemple 4.25 On considère l'expérience qui consiste à mesurer le nombre de particules α émises dans l'espace d'une seconde par un gramme de matière radioactive. Des expériences ont montré dans le passé qu'en moyenne le nombre de particules α émises est 3,2. Donner une bonne approximation pour la probabilité qu'au plus deux particules α seront enregistrées.

SOLUTION. Représentons-nous le gramme de matière radioactive comme une collection de n atomes (n est grand). Chacun peut se désintégrer, ceci avec une probabilité de $3,2/n$ pour la durée de mesure, et donner une particule α . On peut alors dire que le nombre de particules α émises sera approximativement une variable aléatoire de Poisson de paramètre $\lambda = 3,2$ et l'approximation est ici très bonne. La probabilité cherchée sera ainsi

$$P\{X \leq 2\} = e^{-3.2} + 3.2e^{-3.2} + \frac{(3.2)^2}{2}e^{-3.2} \\ \approx .3799 \quad \blacksquare$$

4.8.5 Espérance et variance d'une variable aléatoire de Poisson

Avant de calculer l'espérance et la variance d'une variable aléatoire de Poisson de paramètre λ , rappelons que cette variable aléatoire est une approximation d'une variable aléatoire binomiale de paramètres n et p lorsque n est grand, p est petit et $\lambda = np$. Puisqu'une variable aléatoire binomiale a pour espérance $np = \lambda$ et pour variance $np(1-p) = \lambda(1-p) \approx \lambda$ (comme p est petit), il semblerait que l'espérance et la variance d'une variable aléatoire de Poisson soient toutes les deux égales au paramètre λ . Vérifions ce résultat:

$$\begin{aligned} E[X] &= \sum_{i=0}^{\infty} \frac{ie^{-\lambda} \lambda^i}{i!} \\ &= \lambda \sum_{i=1}^{\infty} \frac{e^{-\lambda} \lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \quad \text{en posant } j = i - 1 \\ &= \lambda \quad \text{puisque } \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda} \end{aligned}$$

Donc, l'espérance d'une variable aléatoire de Poisson est en effet égale au paramètre λ . Pour déterminer sa variance, calculons d'abord $E[X^2]$:

$$\begin{aligned} E[X^2] &= \sum_{i=0}^{\infty} \frac{i^2 e^{-\lambda} \lambda^i}{i!} \\ &= \lambda \sum_{i=1}^{\infty} \frac{ie^{-\lambda} \lambda^{i-1}}{(i-1)!} \\ &= \lambda \sum_{j=0}^{\infty} \frac{(j+1)e^{-\lambda} \lambda^j}{j!} \quad \text{en posant } j = i - 1 \\ &= \lambda \left[\sum_{j=0}^{\infty} \frac{je^{-\lambda} \lambda^j}{j!} + \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} \right] \\ &= \lambda(\lambda + 1) \end{aligned}$$

L'égalité finale provient du fait que la première somme est l'espérance d'une variable aléatoire de Poisson de paramètre λ et que la seconde est la somme des probabilités de cette variable aléatoire. Donc, puisque nous avons montré que $E[X] = \lambda$, on obtient

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= \lambda \end{aligned}$$

L'espérance et la variance d'une variable aléatoire de Poisson sont donc toutes les deux égales à son paramètre λ .

4.8.6 Autres cas d'approximation poissonnienne

Nous avons montré que la loi de Poisson de paramètre np est une très bonne approximation de la distribution du nombre de succès obtenus dans n épreuves indépendantes, où chaque épreuve a la probabilité p d'aboutir à un succès, à condition que n soit grand et p petit. En fait, cette approximation reste valable même lorsque les épreuves ne sont pas indépendantes, pourvu que leur dépendance soit «faible».

Exemple 4.26 Reprenons le problème de rencontre de l'exemple 2.13 du chapitre 2. n hommes tirent au hasard chacun un chapeau dans l'ensemble de leurs chapeaux. Si l'on s'intéresse au nombre d'hommes qui choisissent leur propre chapeau, on peut considérer le tirage aléatoire comme le résultat des n épreuves telles que chaque épreuve i est un succès si la personne i a tiré son propre chapeau, $i = 1, \dots, n$. En définissant les événements E_i , $i = 1, \dots, n$ par

$$E_i = \{\text{l'épreuve } i \text{ est un succès}\}$$

il est facile de voir que

$$P\{E_i\} = 1/n \quad \text{et} \quad P\{E_i | E_j\} = 1/(n-1), \quad j \neq i$$

Ainsi, bien que les événements E_i , $i = 1, \dots, n$ ne sont pas indépendants, on voit que, pour n grand, leur dépendance est plutôt faible. De ce fait, on peut raisonnablement espérer que le nombre de succès suit approximativement une loi de Poisson de paramètre $n \times 1/n = 1$. L'exemple 2.13 du chapitre 2 le vérifie effectivement. ■

Exemple 4.27 Pour une deuxième illustration de l'efficacité de l'approximation poissonnienne lorsque les épreuves sont faiblement dépendantes, considérons de nouveau le problème des anniversaires présenté dans l'exemple 2.10 du chapitre 2. Dans cet exemple, nous avons supposé que chacune des n personnes a, de façon équiprobable, l'un des 365 jours de l'année comme jour d'anniversaire. Le problème est de déterminer la probabilité que, dans un ensemble de n personnes indépendantes, toutes ont leurs jours d'anniversaire différents. Nous avons utilisé un argument combinatoire pour évaluer cette probabilité et nous avons calculé que, lorsque $n = 23$, elle est inférieure à $1/2$.

Nous pouvons estimer cette probabilité en utilisant l'approximation poissonnienne de la manière suivante. Imaginons que l'on a une épreuve pour chacune des $\binom{n}{2}$ paires d'individus i et j , $i \neq j$ et disons que l'épreuve $i - j$ est un succès si les personnes i et j ont le même jour d'anniversaire. Soit E_{ij} l'événement «l'épreuve $i - j$ est un succès». Bien que les $\binom{n}{2}$ événements, $1 \leq i < j \leq n$ ne sont pas indépendants (voir l'exercice théorique 4.10.21), leur dépendance est plutôt faible. (En fait, ces événements sont même «indépendants deux à deux», dans le sens que toute paire d'événements E_{ij} et E_{kl} sont indépendants – voir de nouveau l'exercice 4.10.21). Comme $P(E_{ij}) = 1/365$, il est raisonnable de penser que le nombre de succès doit suivre approximativement une loi de Poisson de paramètre

$$\binom{n}{2}/365 = n(n-1)/730.$$

Par conséquent,

$$P\{2 \text{ personnes différentes n'ont pas le même jour d'anniversaire}\} = P\{0 \text{ succès}\} \\ \approx \exp\{-n(n-1)/730\}$$

Pour déterminer le plus petit entier n pour lequel cette probabilité est inférieure à $1/2$, notons que l'inégalité

$$\exp\{-n(n-1)/730\} \leq \frac{1}{2}$$

est équivalente à

$$\exp\{n(n-1)/730\} \geq 2$$

ou, en prenant le logarithme dans les deux membres, à

$$n(n-1) \geq 730 \ln 2 \\ \approx 505.997$$

ce qui donne la solution $n = 23$, en accord avec le résultat de l'exemple 2.10.

Supposons maintenant que l'on demande la probabilité qu'aucun groupe de 3 personnes parmi n n'ont leur anniversaire le même jour. Alors que l'on est maintenant en présence d'un problème combinatoire difficile, une bonne approximation est facile à obtenir. Pour commencer, faisons correspondre une épreuve à chacun des $\binom{n}{3}$ triplets i, j, k tels que $1 \leq i < j < k \leq n$, et disons que l'épreuve $i - j - k$ est un succès si les personnes i, j et k ont leur anniversaire le même jour. Comme auparavant, nous pouvons conclure que le nombre de succès est approximativement une variable aléatoire de Poisson de paramètre

$$\binom{n}{3} P\{i, j, k \text{ ont leur anniversaire le même jour}\} = \binom{n}{3} \left(\frac{1}{365}\right)^2 \\ = \frac{n(n-1)(n-2)}{6 \times (365)^2}$$

Par conséquent,

$$P\left\{ \begin{array}{l} 3 \text{ personnes différentes} \\ \text{n'ont pas le même jour d'anniversaire} \end{array} \right\} \approx \exp\{-n(n-1)(n-2)/799350\}$$

Cette probabilité sera inférieure à $1/2$ lorsque n est tel que

$$n(n-1)(n-2) \geq 799350 \ln 2 \approx 554067.1$$

qui est équivalent à $n \geq 84$. Ainsi, la probabilité qu'au moins 3 personnes d'un groupe de 84 personnes ou plus ont le même jour d'anniversaire dépasse $1/2$. ■

4.8.7 Introduction au processus de Poisson

Les situations où un événement particulier se reproduit à intervalles réguliers au cours du temps peuvent fournir des cas d'application de la loi de Poisson. On peut citer comme exemple d'un tel événement un tremblement de terre, ou l'entrée d'une personne dans un établissement donné (banque, poste, station d'essence, etc.), ou encore l'apparition d'une guerre. Supposons que l'on ait affaire à de tels événements et qu'en plus il existe une constante positive λ pour laquelle les conditions suivantes soient vérifiées:

- **Condition 1:** la probabilité qu'il adienne exactement 1 événement dans un intervalle de temps de durée h vaut $\lambda h + o(h)$, où $o(h)$ désigne toute fonction $f(h)$ telle que $\lim_{h \rightarrow 0} f(h)/h = 0$. (Par exemple $f(h) = h^2$ est $o(h)$, mais $f(h) = h$ ne l'est pas.)
- **Condition 2:** la probabilité qu'il survienne deux événements ou plus dans un laps de temps de durée h est $o(h)$.
- **Condition 3:** soit des entiers quelconques n, j_1, \dots, j_n et un ensemble quelconque de n intervalles sans intersection. Soient E_i les événements «il survient exactement j_i événements durant l'intervalle i ». Les événements E_1, E_2, \dots, E_n seront toujours indépendants.

En termes approximatifs, les conditions 1 et 2 établissent que lorsque h est petit, la probabilité d'observer exactement 1 événement durant un intervalle de longueur h est λh plus quelque chose de petit comparé à h , tandis que celle d'observer deux événements ou plus est petite comparée à h . La condition 3 garantit que ce qui se passe au cours d'un intervalle n'a pas d'influence sur ce qui arrive durant tout autre intervalle disjoint du premier.

4.8.8 Processus de Poisson et variable aléatoire de Poisson

Nous allons montrer que sous les trois conditions précitées, le nombre d'événements survenant dans un laps de temps d'origine quelconque et de durée t est une variable aléatoire de Poisson avec paramètre λt . Pour plus de clarté, on notera l'intervalle $[0, t]$ et le nombre d'occurrences de l'événement $N(t)$. Dans le but d'obtenir une expression de $P\{N(t) = k\}$ on va partitionner l'intervalle $[0, t]$ en n intervalles disjoints de longueur t/n chacun (cf. fig. 4.7).



Figure 4.7

Or

$$\begin{aligned}
 P\{N(t) = k\} &= P\{k \text{ des } n \text{ sous-intervalles contiennent exactement 1 événement et} \\
 &\quad \text{les } n - k \text{ autres } 0\} \\
 &+ P\{N(t) = k \text{ et au moins un sous-intervalle contient 2 événe-} \\
 &\quad \text{ments ou plus}\}.
 \end{aligned} \tag{4.9}$$

Ceci résulte du fait que les deux événements apparaissant dans le membre de droite de (4.9) sont mutuellement exclusifs. Désignons-les par A et B respectivement. On aura:

$$\begin{aligned}
 P(B) &\leq P\{\text{au moins l'un des sous-intervalles contient 2 occurrences ou plus de l'événement}\} \\
 &= P\left(\bigcup_{i=1}^n \{\text{le } i\text{-ème sous-intervalle contient 2 occurrences ou plus}\}\right) \\
 &\leq \sum_{i=1}^n P\{\text{le } i\text{-ème sous-intervalle contient 2 occurrences ou plus}\} \\
 &\hspace{15em} (\text{en vertu de l'inégalité de Boole}) \\
 &= \sum_{i=1}^n o\left(\frac{t}{n}\right) \\
 &\hspace{15em} (\text{en vertu de la condition 2}) \\
 &= no\left(\frac{t}{n}\right) \\
 &= t \left[\frac{o(t/n)}{t/n} \right]
 \end{aligned}$$

Or, pour tout t , t/n tend vers 0 lorsque n tend vers l'infini et donc $o(t/n)/(t/n)$ tend vers 0 lorsque n tend vers l'infini, par définition de $o(h)$. Par conséquent

$$P(B) \rightarrow 0 \text{ lorsque } n \rightarrow \infty \quad (4.10)$$

D'autre part, du fait qu'on peut écrire grâce aux conditions 1 et 2

$$\begin{aligned}
 P \{\text{il ne survient aucun élément dans un intervalle de durée } h\} \\
 = 1 - [\lambda h + o(h) + o(h)] = 1 - \lambda h - o(h)^1
 \end{aligned}$$

on peut encore écrire en utilisant la condition 3 d'indépendance

$$\begin{aligned}
 P(A) &= P\{k \text{ des sous-intervalles contiennent exactement 1 occurrence et les } \\
 &\quad n - k \text{ autres aucune}\} \\
 &= \binom{n}{k} \left[\frac{\lambda t}{n} + o\left(\frac{t}{n}\right) \right]^k \left[1 - \left(\frac{\lambda t}{n}\right) - o\left(\frac{t}{n}\right) \right]^{n-k}
 \end{aligned}$$

Cependant, puisque

$$n \left[\frac{\lambda t}{n} + o\left(\frac{t}{n}\right) \right] = \lambda t + t \left[\frac{o(t/n)}{t/n} \right] \rightarrow \lambda t \text{ quand } n \rightarrow \infty$$

Il en résulte l'équation suivante, en utilisant l'argument déjà connu qui établit que la distribution binomiale peut être approximée par une distribution poissonnienne:

¹ La somme des deux fonctions $o(h)$ est encore $o(h)$. Il en est ainsi du fait que si $\lim_{h \rightarrow 0} f(h)/h = \lim_{h \rightarrow 0} g(h)/h = 0$, alors $\lim_{h \rightarrow 0} [f(h) + g(h)]/h = 0$.

$$P(A) \rightarrow e^{-\lambda t} \frac{(\lambda t)^k}{k!} \text{ lorsque } n \rightarrow \infty \quad (4.11)$$

En exploitant (4.9), (4.10) et (4.11) on obtient donc, lorsque n tend vers l'infini

$$P\{N(t) = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad k = 0, 1, \dots \quad (4.12)$$

Ainsi le nombre d'occurrences d'événements est-il, sous les conditions 1, 2 et 3, une variable aléatoire de Poisson de paramètre λt ; on dit alors que les événements se réalisent selon un *processus de Poisson* de paramètre λ . Cette grandeur doit être déterminée de manière empirique et on peut montrer qu'elle représente le taux d'occurrence d'événements par unité de temps.

Ce qui précède aide à comprendre pourquoi les variables aléatoires de Poisson donnent généralement de bonnes approximations de phénomènes très divers tels que par exemple

1. le nombre de tremblements de terre survenant pendant une période de longueur donnée
2. le nombre de guerres se déclarant chaque année
3. le nombre d'électrons libérés par une cathode surchauffée durant une période de longueur donnée
4. le nombre de décès parmi les assurés d'une compagnie d'assurance-vie, sur une période de longueur donnée.

4.8.9 Exemple d'application du processus de Poisson

Exemple 4.28 Supposons que les secousses sismiques dans la moitié ouest des Etats-Unis surviennent de manière telle que les conditions 1, 2 et 3 soient satisfaites, λ valant 2 et l'unité de temps étant la semaine. (Ceci revient à dire que des secousses se produisent, en accord avec les trois conditions précitées, au rythme de 2 par semaine).

a) Trouver d'abord la probabilité qu'au moins 3 secousses aient lieu durant les 2 prochaines semaines. b) Trouver ensuite la distribution de la durée entre maintenant et la prochaine secousse.

SOLUTION. a) D'après (4.12) nous aurons

$$\begin{aligned} P\{N(2) \geq 3\} &= 1 - P\{N(2) = 0\} - P\{N(2) = 1\} - P\{N(2) = 2\} \\ &= 1 - e^{-4} - 4e^{-4} - \frac{4^2}{2}e^{-4} \\ &= 1 - 13e^{-4} \end{aligned}$$

b) Désignons par X la durée d'attente jusqu'à la prochaine secousse, mesurée en semaines. Du fait que X sera supérieure à t si et seulement si il ne survient aucune secousse durant les t prochaines semaines, (4.12) donne

$$P\{X > t\} = P\{N(t) = 0\} = e^{-\lambda t}$$

et ainsi la fonction de répartition F de X sera

$$\begin{aligned} F(t) = P\{X \leq t\} &= 1 - P\{X > t\} = 1 - e^{-\lambda t} \\ &= 1 - e^{-2t} \end{aligned}$$

■

4.8.10 Calcul de la fonction de répartition de Poisson

Si X est une variable aléatoire de Poisson de paramètre λ , alors

$$\frac{P\{X = i + 1\}}{P\{X = i\}} = \frac{e^{-\lambda} \lambda^{i+1} / (i + 1)!}{e^{-\lambda} \lambda^i / i!} = \frac{\lambda}{i + 1}. \quad (4.13)$$

En commençant par $P\{X = 0\} = e^{-\lambda}$, nous pouvons utiliser la formule (4.13) pour calculer successivement

$$P\{X = 1\} = \lambda P\{X = 0\}$$

$$P\{X = 2\} = \frac{\lambda}{2} P\{X = 1\}$$

$$\vdots$$

$$P\{X = i + 1\} = \frac{\lambda}{i + 1} P\{X = i\}.$$

Un programme en langage Basic, qui utilise l'équation (4.13) pour calculer la fonction de répartition de Poisson, est présenté à la fin de ce chapitre. Pour calculer $P\{X \leq i\}$, le programme calcule d'abord $P\{X = i\}$ puis utilise la récurrence pour calculer successivement $P\{X = i - 1\}$, $P\{X = i - 2\}$, etc. Le programme calcule $P\{X = i\}$ en évaluant d'abord

$$\ln(P\{X = i\}) = -\lambda + i \ln \lambda - \sum_{k=1}^i \ln k$$

puis en utilisant l'égalité

$$P\{X = i\} = \exp\{\ln(P\{X = i\})\}.$$

Exemple 4.29

- Déterminer $P\{X \leq 100\}$ quand X est une variable aléatoire de Poisson de moyenne 90.
- Déterminer $P\{X \leq 1075\}$ quand X est une variable aléatoire de Poisson de moyenne 1000.

SOLUTION. Exécuter le programme suivant:

```

RUN
THIS PROGRAM COMPUTES THE PROBABILITY THAT A POISSON RANDOM VARIABLE
IS LESS THAN OR EQUAL TO i
ENTER THE MEAN OF THE RANDOM VARIABLE
? 100
ENTER THE DESIRED VALUE OF i
? 90
THE PROBABILITY THAT A POISSON RANDOM VARIABLE WITH MEAN 100
IS LESS THAN OR EQUAL TO 90 IS .1713914
OK
RUN
THIS PROGRAM COMPUTES THE PROBABILITY THAT A POISSON RANDOM VARIABLE
IS LESS THAN OR EQUAL TO i
ENTER THE MEAN OF THE RANDOM VARIABLE
? 1000
ENTER THE DESIRED VALUE OF i
? 1075
THE PROBABILITY THAT A POISSON RANDOM VARIABLE WITH MEAN 1000
IS LESS THAN OR EQUAL TO 1075 IS .989354
OK

```

4.9 AUTRES LOIS DISCRÈTES

4.9.1 Variables aléatoires géométriques

On exécute une série d'épreuves indépendantes ayant chacune la probabilité p d'être un succès, $0 < p < 1$, jusqu'à obtenir le premier succès. Si l'on désigne le nombre d'épreuves nécessaires jusqu'à ce résultat par X on aura

$$P\{X = n\} = (1 - p)^{n-1}p \quad n = 1, 2, \dots \quad (4.14)$$

En effet, pour que X prenne n pour valeur, il faut et suffit que les $n - 1$ premières épreuves soient des échecs tandis que la n -ième devra être un succès. (4.14) est alors immédiate puisque les épreuves sont indépendantes.

Du fait que

$$\sum_{n=1}^{\infty} P\{X = n\} = p \sum_{n=1}^{\infty} (1 - p)^{n-1} = \frac{p}{1 - (1 - p)} = 1$$

il est établi qu'avec probabilité 1 un succès finira par se produire.

Les variables aléatoires dont la loi est donnée par (4.14) sont appelées *variables aléatoires géométriques* de paramètres p .

Exemple 4.30 Une urne contient N boules blanches et M noires. On tire des boules une par une avec remise jusqu'à l'apparition d'une noire. a) Quelle est la probabilité qu'il faille exactement n tirages? b) Quelle est la probabilité qu'il faille au moins k tirages?

SOLUTION. Désignons par X le nombre de tirages nécessaires jusqu'à l'apparition de la première boule noire. X est régie par (4.14) avec $p = M/(M + N)$. D'où les deux résultats demandés:

$$\text{a) } P\{X = n\} = \left(\frac{N}{M+N}\right)^{n-1} \frac{M}{M+N} = \frac{MN^{n-1}}{(M+N)^n}$$

$$\begin{aligned} \text{b) } P\{X \geq k\} &= \frac{M}{M+N} \sum_{n=k}^{\infty} \left(\frac{N}{M+N}\right)^{n-1} \\ &= \left(\frac{M}{M+N}\right) \left(\frac{N}{M+N}\right)^{k-1} \left/ \left[1 - \frac{N}{M+N}\right]\right. \\ &= \left(\frac{N}{M+N}\right)^{k-1} \end{aligned}$$

On aurait bien sûr pu obtenir le deuxième directement puisque la probabilité qu'il faille au moins k essais pour obtenir un premier succès est égale à celle de n'avoir que des échecs sur les $k - 1$ premières épreuves. Cette probabilité est, pour une variable géométrique

$$P\{X \geq k\} = (1 - p)^{k-1} \quad \blacksquare$$

Exemple 4.31 Trouver l'espérance d'une variable aléatoire géométrique.

SOLUTION. En posant $q = 1 - p$, on a

$$\begin{aligned} E[X] &= \sum_{n=1}^{\infty} nq^{n-1} p \\ &= p \sum_{n=0}^{\infty} \frac{d}{dq} (q^n) \\ &= p \frac{d}{dq} \left(\sum_{n=0}^{\infty} q^n \right) \\ &= p \frac{d}{dq} \left(\frac{1}{1-q} \right) \\ &= \frac{p}{(1-q)^2} \\ &= \frac{1}{p} \end{aligned}$$

En d'autres termes, si des épreuves indépendantes ayant une probabilité p d'obtenir un succès sont réalisées jusqu'à ce que le premier succès se produise, le nombre espéré d'essais nécessaires est égal à $1/p$. Par exemple, le nombre espéré de jets d'un dé équilibré qu'il faut pour obtenir la valeur 1 est 6. \blacksquare

Exemple 4.32 Trouver la variance d'une variable aléatoire géométrique.

SOLUTION. Pour trouver $\text{Var}(X)$, calculons d'abord $E[X^2]$. En posant $q = 1 - p$, on a

$$\begin{aligned} E[X^2] &= \sum_{n=1}^{\infty} n^2 q^{n-1} p \\ &= p \sum_{n=1}^{\infty} \frac{d}{dq} (nq^n) \\ &= p \frac{d}{dq} \left(\sum_{n=1}^{\infty} nq^n \right) \\ &= p \frac{d}{dq} \left(\frac{q}{1-q} E[X] \right) \\ &= p \frac{d}{dq} [q(1-q)^{-2}] \\ &= p \left[\frac{1}{p^2} + \frac{2(1-p)}{p^3} \right] \\ &= \frac{2}{p^2} - \frac{1}{p} \end{aligned}$$

Donc, comme $E[X] = 1/p$,

$$\text{Var}(X) = \frac{1-p}{p^2}$$

4.9.2 Variables aléatoires binomiales négatives

On exécute une série d'épreuves indépendantes ayant chacune une probabilité p de donner un succès, $0 < p < 1$, jusqu'à obtenir un total de r succès. Désignons par X le nombre d'épreuves nécessaires pour atteindre ce résultat. On aura

$$P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r} \quad n = r, r+1, \dots \quad (4.15)$$

En effet, pour obtenir un r -ième succès lors de la n -ième épreuve il a fallu $r-1$ succès lors des $n-1$ premières épreuves et il faut que la n -ième épreuve soit un succès. La probabilité de la première condition est

$$\binom{n-1}{r-1} p^{r-1} (1-p)^{n-r}$$

et celle de la seconde est p . De ce fait (4.15) est établie puisque les épreuves sont indépendantes.

On peut établir que la probabilité d'obtenir r succès est 1. Il existe une démonstration analytique de l'équation

$$\sum_{n=r}^{\infty} P\{X = n\} = \sum_{n=r}^{\infty} \binom{n-1}{r-1} p^r (1-p)^{n-r} = 1 \quad (4.16)$$

mais on peut donner l'argument probabiliste suivant: le nombre d'épreuves nécessaires à l'obtention de r succès peut être écrit $Y_1 + Y_2 + \dots + Y_r$, Y_1 étant le nombre d'épreuves nécessaires jusqu'au premier succès, Y_2 le nombre d'épreuves supplémentaires nécessaires pour obtenir un deuxième succès, Y_3 celui menant au 3ème et ainsi de suite. Les tirages étant indépendants et ayant toujours la même probabilité de succès, chacune des variables Y_1, Y_2, \dots, Y_r est géométrique. On a vu que chacune est finie avec probabilité 1 et par conséquent leur somme X l'est aussi, ce qui établit (4.16).

Une variable aléatoire dont la loi est donnée par (4.15) est dite *variable aléatoire binomiale négative* de paramètres (r, p) . On remarquera qu'une variable géométrique est binomiale négative de paramètre $(1, p)$.

Dans l'exemple suivant, le problème des points trouve une autre solution grâce à l'emploi d'une variable binomiale négative.

Exemple 4.33 On exécute une série d'épreuves indépendantes, chacune aboutissant à un succès avec la même probabilité p . Quelle est la probabilité que r succès apparaissent avant que le m -ième échec ne survienne?

SOLUTION. On parvient à cette solution en remarquant que r succès n'apparaissent avant le m -ième échec que si le r -ième succès survient au plus tard à la $(r+m-1)$ -ième épreuve. En effet, si le r -ième succès a lieu avant cette $(r+m-1)$ -ième épreuve ou au plus tard lors même de celle-ci, elle intervient avant le m -ième échec et l'implication inverse est vraie. On tire alors de (4.15) la probabilité voulue

$$\sum_{n=r}^{r+m-1} \binom{n-1}{r-1} p^r (1-p)^{n-r} \quad \blacksquare$$

Exemple 4.34 Le problème des allumettes de Banach

Un mathématicien se trouve être également fumeur de pipe et il porte à tout moment deux boîtes d'allumettes, une dans chacune de ses poches gauche et droite. Chaque fois qu'il a besoin d'une allumette, il a une chance sur deux d'aller la chercher dans sa poche gauche et autant pour l'autre. Il découvre subitement que la boîte tirée est vide. Les deux boîtes contenaient au départ N allumettes chacune. Quelle est la probabilité qu'il lui reste k allumettes dans l'autre boîte, $k = 0, 1, \dots, N$?

SOLUTION. Désignons par E l'événement «le mathématicien découvre que sa boîte droite est vide alors qu'il reste k allumettes dans l'autre». Cet événement n'aura lieu que s'il choisit la boîte droite pour la $(N+1)$ -ième fois lors du $N+1+N-k$ -ième tirage. Grâce à (4.15) on peut alors écrire, en prenant $p = \frac{1}{2}$, $r = N+1$, $n = 2N-k+1$,

$$P(E) = \binom{2N-k}{N} \left(\frac{1}{2}\right)^{2N-k+1}$$

Comme la probabilité est la même que ce soit la poche gauche qui se vide tandis qu'il reste k allumettes dans la droite, la probabilité voulue est

$$2P(E) = \binom{2N-k}{N} \left(\frac{1}{2}\right)^{2N-k} \quad \blacksquare$$

Exemple 4.35 Calculer l'espérance et la variance d'une variable aléatoire binomiale négative de paramètres r et p .

SOLUTION.

$$\begin{aligned} E[X^k] &= \sum_{n=r}^{\infty} n^k \binom{n-1}{r-1} p^r (1-p)^{n-r} \\ &= \frac{r}{p} \sum_{n=r}^{\infty} n^{k-1} \binom{n}{r} p^{r+1} (1-p)^{n-r} \quad \text{comme } n \binom{n-1}{r-1} = r \binom{n}{r} \\ &= \frac{r}{p} \sum_{m=r+1}^{\infty} (m-1)^{k-1} \binom{m-1}{r} p^{r+1} (1-p)^{m-(r+1)} \quad \text{en posant } m = n+1 \\ &= \frac{r}{p} E[(Y-1)^{k-1}] \end{aligned}$$

où Y est une variable aléatoire binomiale négative de paramètres $r+1$, p . En posant $k=1$ dans l'équation précédente, on obtient

$$E[X] = \frac{r}{p}$$

En posant $k=2$ dans cette même équation et en utilisant la formule ci-dessus pour l'espérance d'une variable aléatoire binomiale négative, on obtient

$$\begin{aligned} E[X^2] &= \frac{r}{p} E[Y-1] \\ &= \frac{r}{p} \left(\frac{r+1}{p} - 1 \right) \end{aligned}$$

Donc

$$\begin{aligned} \text{Var}(X) &= \frac{r}{p} \left(\frac{r+1}{p} - 1 \right) - \left(\frac{r}{p} \right)^2 \\ &= \frac{r(1-p)}{p^2} \quad \blacksquare \end{aligned}$$

L'exemple 4.35 nous montre que si on réalise des épreuves indépendantes, chacune ayant une probabilité p de succès, alors l'espérance et la variance du nombre d'essais nécessaires pour obtenir r succès sont r/p et $r(1-p)/p^2$ respectivement.

Puisqu'une variable aléatoire géométrique est une binomiale négative de paramètre $n = 1$, l'exemple précédent entraîne que la variance d'une variable aléatoire géométrique de paramètre p est égale à $(1-p)/p^2$, ce qui correspond au résultat de l'exemple 4.32.

Exemple 4.36 Calculer l'espérance et la variance du nombre de jets d'un dé nécessaires pour obtenir 4 fois la valeur 1.

SOLUTION. Puisque la variable aléatoire étudiée est une binomiale négative de paramètres $r = 4$ et $p = 1/6$, on a

$$E[X] = 24$$

$$\text{Var}(X) = \frac{4 \left(\frac{5}{6}\right)}{\left(\frac{1}{6}\right)^2} = 120$$

4.9.3 Variables aléatoires hypergéométriques

On tire sans remise un échantillon de n boules d'une urne en contenant N , dont m sont blanches et $N - m$ sont noires. Désignons par X le nombre de boules blanches tirées. On aura

$$P\{X = i\} = \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}}, \quad i = 0, 1, \dots, n \quad (4.17)$$

S'il existe certaines valeurs de n , N et m pour lesquelles la loi d'une variable aléatoire vérifie (4.17), cette variable est dite *variable aléatoire hypergéométrique*.

REMARQUE. Bien que nous ayons écrit la distribution hypergéométrique avec i variant de 0 à n , $P\{X = i\}$ vaudra 0 à moins que i satisfasse les inégalités $n - (N - m) \leq i \leq \min(n, m)$. Cependant, l'équation (4.17) est toujours valable à cause de notre convention qui pose $\binom{r}{k}$ égal à 0 lorsque $k < 0$ ou $r < k$.

Exemple 4.37 Le nombre d'animaux d'une certaine espèce habitant un territoire donné est N , inconnu. Pour obtenir une information sur la taille de cette population, les écologues ont souvent recours à l'expérience suivante: ils capturent en premier lieu une certaine quantité de ces animaux, mettons m . Ils les marquent puis les relâchent et leur laissent le temps de se disperser sur l'ensemble du territoire étudié. Dans un deuxième temps, ils font un certain nombre n de nouvelles captures. Désignons par X le nombre d'animaux marqués figurant parmi ces nouvelles captures. On admet que la population animale n'a pas changé entre les dates des deux séries de captures et

que lors de chaque capture tous les animaux restant à ce stade ont la même probabilité d'être pris. X est alors une variable aléatoire hypergéométrique et

$$P(X=i) = \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}} = P_i(N)$$

Supposons maintenant que la valeur observée de X soit i . Désignons par $P_i(N)$ la probabilité que ceci arrive alors que la population totale est N . Une estimation intuitivement raisonnable de N est donnée par la valeur pour laquelle $P_i(N)$ est maximale. (Une telle estimation est dite *estimation du maximum de vraisemblance*. Les exercices théoriques 4.10.8 et 4.10.13 sont d'autres exemples de ce type de procédé d'estimation).

Le moyen le plus simple de trouver le maximum de $P_i(N)$ se base sur le calcul du rapport

$$\frac{P_i(N)}{P_i(N-1)} = \frac{(N-m)(N-n)}{N(N-m-n+i)}$$

Or, le rapport est supérieur à 1 si et seulement si

$$(N-m)(N-n) \geq N(N-m-n+i)$$

ou, ce qui est équivalent, si et seulement si,

$$N \leq \frac{m n}{i}$$

Ainsi, $P_i(N)$ est-elle d'abord croissante puis décroissante, prenant son maximum pour la partie entière de mn/i . Cette valeur est donc l'estimation du maximum de vraisemblance de N . Si par exemple la première campagne de capture a livré $m = 50$ animaux marqués puis relâchés et la 2ème campagne a permis de capturer $n = 40$ animaux dont $i = 4$ sont marqués, on estimera que la population de ces animaux se chiffre à 500 sur le territoire étudié. (On aurait aussi pu obtenir cette estimation en faisant l'hypothèse que la proportion i/n d'animaux marqués lors de la 2ème campagne de captures est égale à celle des animaux marqués dans toute la population, soit m/N). ■

Exemple 4.38 Un électricien achète des composants par paquets de 10. Sa technique de contrôle est de n'examiner que 3 des composants, tirés au hasard dans le paquet, et de n'accepter le lot des 10 que si les 3 composants examinés sont sans défaut. Si 30% des paquets contiennent 4 composants à malfaçon tandis que les 70% restants n'en contiennent qu'un, quelle proportion des paquets notre électricien rejettera-t-il?

SOLUTION. Désignons par A l'événement «l'électricien accepte un paquet». On sait que

$$\begin{aligned}
 P(A) &= P(A|\text{le paquet contient 4 mauvais composants}) \frac{3}{10} + \\
 &\quad P(A|\text{le paquet contient 1 mauvais composant}) \frac{7}{10} \\
 &= \frac{\binom{4}{0} \binom{6}{3}}{\binom{10}{3}} \left(\frac{3}{10} \right) + \frac{\binom{1}{0} \binom{9}{3}}{\binom{10}{3}} \left(\frac{7}{10} \right) \\
 &= \frac{54}{100}
 \end{aligned}$$

46% des paquets seront donc refusés. ■

Si n balles sont choisies au hasard sans remise dans un ensemble de N balles dont la fraction $p = m/N$ est blanche, alors le nombre de balles blanches choisies est hypergéométrique. A présent, il semblerait que lorsque m et N sont grands par rapport à n , il ne devrait pas y avoir trop de différence entre les tirages avec et sans remise. Du fait que les balles tirées précédemment n'ont pas une grosse importance, chaque nouveau tirage, lorsque m et N sont grands, sera blanc avec une probabilité approximativement égale à p . En d'autres termes, on sent intuitivement lorsque m et N sont grands par rapport à n , que la distribution de X doit approximativement être celle d'une variable aléatoire binomiale de paramètres n et p . Pour vérifier cette intuition, remarquons que si X est hypergéométrique, alors pour $i \leq n$,

$$\begin{aligned}
 P\{X = i\} &= \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}} \\
 &= \frac{m!}{(m-i)! i!} \frac{(N-m)!}{(N-m-n+i)! (n-i)!} \frac{(N-n)! n!}{N!} \\
 &= \binom{n}{i} \frac{m}{N} \frac{m-1}{N-1} \dots \frac{m-i+1}{N-i+1} \frac{N-m}{N-i} \frac{N-m-1}{N-i-1} \dots \frac{N-m-(n-i-1)}{N-i-(n-i-1)} \\
 &\approx \binom{n}{i} p^i (1-p)^{n-i} \quad \text{lorsque } p = \frac{m}{N} \text{ et } m \text{ et } N \text{ sont} \\
 &\quad \text{grands par rapport à } n \text{ et } i
 \end{aligned}$$

Exemple 4.39 Déterminer l'espérance et la variance d'une variable aléatoire hypergéométrique X de paramètres n, N, m .

SOLUTION.

$$\begin{aligned}
 E[X^k] &= \sum_{i=0}^n i^k P\{X = i\} \\
 &= \sum_{i=1}^n i^k \binom{m}{i} \binom{N-m}{n-i} / \binom{N}{n}
 \end{aligned}$$

En utilisant les identités

$$i \binom{m}{i} = m \binom{m-1}{i-1} \quad \text{et} \quad n \binom{N}{n} = N \binom{N-1}{n-1}$$

on obtient

$$\begin{aligned} E[X^k] &= \frac{nm}{N} \sum_{i=1}^n i^{k-1} \binom{m-1}{i-1} \binom{N-m}{n-i} / \binom{N-1}{n-1} \\ &= \frac{nm}{N} \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{m-1}{j} \binom{N-m}{n-1-j} / \binom{N-1}{n-1} \\ &= \frac{nm}{N} E[(Y+1)^{k-1}] \end{aligned}$$

où Y est une variable aléatoire hypergéométrique de paramètres $n-1$, $N-1$, $m-1$.
Donc, en posant $k=1$, on obtient

$$E[X] = \frac{nm}{N}$$

En d'autres termes, si n balles sont choisies aléatoirement parmi un ensemble de N balles dont m sont blanches, le nombre espéré de balles blanches sélectionnées est nm/N .

En posant $k=2$ dans l'équation de $E[X^k]$, on obtient

$$\begin{aligned} E[X^2] &= \frac{nm}{N} E[Y+1] \\ &= \frac{nm}{N} \left[\frac{(n-1)(m-1)}{N-1} + 1 \right] \end{aligned}$$

où l'égalité finale utilise le résultat précédent pour calculer l'espérance de la variable aléatoire hypergéométrique Y . Comme $E[X] = nm/N$, on peut conclure que

$$\text{Var}(X) = \frac{nm}{N} \left[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N} \right] \quad (4.18)$$

Si $p = m/N$ est la fraction de balles blanches, l'équation (4.18) donne, après un peu d'algèbre

$$\text{Var}(X) = \frac{N-n}{N-1} np(1-p) \quad (4.19)$$

REMARQUE. Nous avons montré dans l'exemple 4.39 que si n balles sont sélectionnées au hasard sans remise parmi un ensemble de N balles dont une fraction p est blanche,

le nombre espéré de balles blanches choisies est np . De plus, si N est grand par rapport à n [et donc si $(N - n) / (N - 1)$ est approximativement égal à 1], alors

$$\text{Var}(X) \approx np(1 - p)$$

En d'autres termes, $E[X]$ est la même que lorsque la sélection des balles se fait avec remise (le nombre de balles blanches est alors binomial de paramètres n et p) et si le nombre total de balles est grand, $\text{Var}(X)$ est approximativement égale à ce qu'elle vaudrait si la sélection était effectuée avec remise. Naturellement, ceci est exactement ce que nous avons deviné à partir du résultat précédent, qui établissait que lorsque le nombre de balles dans l'urne est grand, le nombre de balles blanches choisies suit approximativement la distribution d'une variable aléatoire binomiale.

4.94 Variables aléatoires Zéta (ou de Zipf)

On dit qu'une variable aléatoire suit une *distribution zéta* (parfois aussi dite *de Zipf*) si sa loi de probabilité est

$$P\{X = k\} = \frac{C}{k^{\alpha+1}}, \quad k = 1, 2, \dots$$

où $\alpha > 0$. Du fait que la somme de ces probabilités doit donner 1, on peut calculer C :

$$C = \left[\sum_{k=1}^{\infty} \left(\frac{1}{k} \right)^{\alpha+1} \right]^{-1}$$

La loi zéta doit son nom au fait que la fonction

$$\zeta(s) = 1 + \left(\frac{1}{2} \right)^s + \left(\frac{1}{3} \right)^s + \dots + \left(\frac{1}{k} \right)^s + \dots$$

est appelée par les mathématiciens fonction zéta de Riemann (du nom du mathématicien allemand G.F.B. Riemann). Cette distribution zéta a été utilisée par le célèbre économiste italien Pareto pour décrire la répartition des revenus familiaux à travers un pays donné. Ce fut G.K. Zipf cependant qui popularisa leur usage en les appliquant à des domaines très variés.

4.10 EXERCICES THÉORIQUES

4.10.1 Il existe des coupons de N sortes. On les obtient à raison d'un à la fois et dans chaque cas, le coupon reçu sera du type i avec probabilité P_i , $i = 1, 2, \dots, N$ indépendamment du type des coupons reçus auparavant. Soit T le nombre de coupons

qu'il faut collectionner pour obtenir un assortiment complet comprenant un coupon de chaque type au moins. Calculer $P\{T = n\}$.

On pourra utiliser un argument similaire à celui utilisé dans l'exemple 4.5.

4.10.2 Etablir la propriété 3 des fonctions de répartition.

4.10.3 Exprimer $P\{X \geq a\}$ grâce à la fonction de répartition de X .

4.10.4 Démontrer ou infirmer au moyen d'un contre-exemple l'égalité suivante:

$$P\{X < b\} = \lim_{b_n \rightarrow b} P\{X < b_n\}$$

4.10.5 Soit F la fonction de répartition de X . Quelle est la fonction de répartition de $\alpha X + \beta$, où α et β sont des constantes, $\alpha \neq 0$?

4.10.6 Pour une variable aléatoire N à valeurs entières non négatives, montrer que

$$E[N] = \sum_{i=1}^{\infty} P\{N \geq i\}$$

Utiliser pour cela $\sum_{i=1}^{\infty} P\{N \geq i\} = \sum_{i=1}^{\infty} \sum_{k=i}^{\infty} P\{N = k\}$ en changeant l'ordre de sommation.

4.10.7 Soit N une variable aléatoire entière non négative. Montrer que

$$\sum_{i=0}^{\infty} iP\{N > i\} = \frac{1}{2} (E[N^2] - E[N])$$

Pour cela, utiliser

$$\sum_{i=0}^{\infty} iP\{N > i\} = \sum_{i=0}^{\infty} i \sum_{k=i+1}^{\infty} P\{N = k\}$$

puis changer l'ordre des sommations.

4.10.8 Soit X tel que

$$P\{X = 1\} = p = 1 - P\{X = -1\}$$

Trouver $c \neq 1$ tel que $E[c^X] = 1$.

4.10.9 Soit X une variable aléatoire d'espérance μ et de variance σ^2 . Trouver l'espérance et la variance de

$$Y = \frac{X - \mu}{\sigma}$$

4.10.10 Soit X une variable aléatoire binomiale de paramètres n et p . Montrer que

$$E\left[\frac{1}{X+1}\right] = \frac{1 - (1-p)^{n+1}}{(n+1)p}$$

4.10.11 On exécute une série de n épreuves indépendantes ayant chacune une probabilité p d'aboutir à un succès. Montrer que les $n! / [k!(n-k)!]$ séquences contenant exactement k succès sont toutes équiprobables.

4.10.12 On considère un alignement de n composants. Chacun d'entre eux fonctionne avec probabilité p . Quelle est la probabilité de ne pas rencontrer dans cet alignement des composants hors service voisins?

On pourra conditionner sur le nombre de composants défectueux et utiliser les résultats de l'exemple 1.14 du chapitre 1.

4.10.13 On considère une variable aléatoire binomiale de X de paramètres (n, p) . Pour quelle valeur de p la probabilité $P\{X = k\}$ est-elle maximale, dans le cas où $k = 1, 2, \dots, n$? Ce résultat est utilisé en statistique pour estimer p lorsqu'on a observé que $X = k$. Le paramètre n étant connu, cette valeur de p qui rend maximale $P\{X = k\}$ est appelée estimation de p par la méthode du maximum de vraisemblance.

4.10.14 On admet que la probabilité qu'une famille ait n enfants est αp^n , $n \geq 1$, $\alpha \leq (1-p)/p$.

a) Quelle est la proportion parmi toutes les familles de celles qui n'auront aucun enfant?

b) Les enfants ont autant de chances d'être des garçons que des filles, indépendamment du sexe de leurs aînés. Quelle est la proportion parmi toutes les familles de celles ayant exactement k filles, le nombre de garçons n'étant pas fixé?

4.10.15 On jette n fois une pièce, pile sortant avec probabilité p à chaque tirage indépendamment de ce qui a précédé. Montrer que la probabilité d'obtenir un nombre pair de piles est $\frac{1}{2}[1 + (q-p)^n]$, où $q = 1-p$. Utiliser pour cela l'identité suivante, après l'avoir démontrée:

$$\sum_{i=0}^{[n/2]} \binom{n}{2i} p^{2i} q^{n-2i} = \frac{1}{2}[(p+q)^n + (q-p)^n]$$

où $[n/2]$ désigne la partie entière de $n/2$. Comparer cet exercice à l'exercice 3.6.11.

4.10.16 Soit X une variable aléatoire de Poisson avec paramètre λ . Montrer que $P\{X = i\}$ est une fonction monotone croissante puis décroissante de i prenant son maximum lorsque $i = [\lambda]$.

Etudier pour cela $P\{X = i\} / P\{X = i-1\}$.

4.10.17 Soit X une variable aléatoire de Poisson de paramètre λ . a) Montrer que

$$P\{X \text{ est paire}\} = \frac{1}{2}[1 + e^{-2\lambda}]$$

en utilisant les résultats du problème 4.10.15 et la relation entre les variables aléatoires poissonnienne et binomiale. b) Faire une vérification directe en s'aidant du développement de $e^{-\lambda} + e^{\lambda}$.

4.10.18 Soit X une variable aléatoire de Poisson avec paramètre λ . Quelle est la valeur de λ qui maximise $P\{X = k\}$, $k \geq 0$?

4.10.19 Si X est une variable aléatoire de Poisson de paramètre λ , montrer que

$$E[X^n] = \lambda E[(X+1)^{n-1}]$$

Utiliser ce résultat pour calculer $E[X^3]$.

4.10.20 Soit X une variable aléatoire de Poisson de paramètre λ , où $0 < \lambda < 1$. Trouver $E[X!]$.

4.10.21 On considère un ensemble de n personnes choisies au hasard. Soit E_{ij} l'événement «les personnes i et j ont leur anniversaire le même jour». On suppose que chaque personne a son anniversaire qui tombe sur un des 365 jours de l'année, de façon équiprobable. Trouver

- a) $P(E_{3,4} | E_{1,2})$
- b) $P(E_{1,3} | E_{1,2})$
- c) $P(E_{2,3} | E_{1,2} \cap E_{1,3})$

Que peut-on conclure sur l'indépendance des (?) événements E_{ij} ?

4.10.22 Une urne contient $2n$ boules, dont 2 sont numérotées 1, 2 numérotées 2, ..., 2 numérotées n . Des couples de boules sont tirées successivement sans remise. Soit T le premier tirage lors duquel les boules obtenues ont le même numéro (T vaut infini si aucun couple de boules tirées n'ont le même numéro). Pour $0 < \alpha < 1$, on veut montrer que

$$\lim_n P\{T > \alpha n\} = e^{-\alpha/2}$$

Pour le vérifier, posons M_k le nombre de paires obtenues dans les k premiers tirages, $k = 1, \dots, n$.

- a) Justifier pourquoi, lorsque n est grand, M_k peut être associé au nombre de succès dans k épreuves (approximativement) indépendantes.
- b) Dans le cas où n est grand, approximer $P\{M_k = 0\}$.
- c) Décrire l'événement « $T > \alpha n$ » en fonction de la valeur prise par l'une des variables M_k .
- d) Vérifier la probabilité limite ci-dessus.

4.10.23 On admet que le nombre d'événements d'un certain type survenant pendant un laps de temps donné est une variable aléatoire de paramètre λ . Ces événements sont enregistrés avec probabilité p (certains passent inaperçus), les enregistrements étant indépendants les uns des autres. Montrer que le nombre d'événements enregistrés est une variable aléatoire de Poisson de paramètre λp . Donner une explication intuitive soutenant ce résultat.

En application de ce qui précède, on étudie une campagne de prospection pour découvrir des gisements d'uranium supposés clairement séparés les uns des autres. Dans une région donnée, le nombre de tels gisements distincts est une variable aléatoire de Poisson avec paramètre $\lambda = 10$. Pendant la durée de la campagne la probabilité qu'un gisement donné soit découvert est $\frac{1}{50}$. Trouver la probabilité que

- a) exactement un gisement,
 b) au moins un gisement,
 c) au plus un gisement
 soit découvert lors de cette campagne.

4.10.24 Démontrer l'égalité

$$\sum_{i=0}^n e^{-\lambda} \frac{\lambda^i}{i!} = \frac{1}{n!} \int_{\lambda}^{\infty} e^{-x} x^n dx$$

en intégrant par parties.

4.10.25 Soit X une variable aléatoire géométrique. Montrer par un calcul analytique que

$$P\{X = n + k | X > n\} = P\{X = k\}$$

Formuler un argument intuitif en faveur de cette équation en se basant sur le modèle général auquel s'appliquent les variables géométriques.

4.10.26 Soit X une variable aléatoire binomiale négative de paramètres r et p et soit Y une variable aléatoire binomiale de paramètres n et p . Montrer que

$$P\{X > n\} = P\{Y < r\}$$

On pourra tenter de démontrer analytiquement l'expression ci-dessus en montrant l'identité suivante

$$\sum_{i=n+1}^{\infty} \binom{i-1}{r-1} p^r (1-p)^{i-r} = \sum_{i=0}^{r-1} \binom{n}{i} p^i (1-p)^{n-i}$$

ou on pourra tenter une preuve utilisant l'interprétation probabiliste de ces variables aléatoires. Autrement dit, dans ce dernier cas, on commencera par considérer une séquence d'épreuves indépendantes ayant toutes une probabilité p de succès. Essayer alors d'exprimer les événements $\{X > n\}$ et $\{Y < r\}$ en termes de résultats de cette séquence.

4.10.27 Pour une variable aléatoire X hypergéométrique, déterminer

$$P\{X = k + 1\} / P\{X = k\}$$

4.10.28 Une urne contient des boules numérotées de 1 à N . Supposons qu'on en tire n , $n \leq N$, au hasard et sans remise. Y désigne le plus grand numéro tiré. Donner la loi de Y .

4.10.29 Un bocal contient $m + n$ jetons numérotés de 1 à $n + m$. On en tire n . X désigne le nombre de jetons dont le numéro est supérieur à celui de tous les jetons restés dans le bocal. Donner la loi de X .

4.10.30 Un bocal contient n jetons. Un garçon tire les jetons un à un avec remise jusqu'à ce qu'il obtienne un jeton déjà tiré. X désigne le nombre de tirages. Donner la loi de probabilité de X .

4.10.31 Montrer que l'équation (4.19) découle de l'équation (4.18).

4.10.32 Partant d'un ensemble de n éléments, on choisit de manière aléatoire un sous-ensemble non vide de manière que tous les sous-ensembles non vides aient la même probabilité d'être choisis. Soit X le cardinal du sous-ensemble choisi. En utilisant les identités données dans l'exercice théorique 1.7.13, montrer que

$$E[X] = \frac{n}{2 - \left(\frac{1}{2}\right)^{n-1}}$$

$$\text{Var}(X) = \frac{n2^{2n-2} - n(n+1)2^{n-2}}{(2^n - 1)^2}$$

Montrer aussi que lorsque n est grand, $\text{Var}(X)$ est d'ordre $\frac{n}{4}$ en ce sens que le rapport donnant la variance s'approche de 1 lorsque n tend vers l'infini. Comparer ce résultat avec la forme limite que prend $\text{Var}(Y)$ quand $P\{Y = i\} = 1/n, i = 1, \dots, n$.

4.11 PROBLÈMES

4.11.1 On choisit deux boules au hasard d'une urne en contenant 8 blanches, 4 noires et 2 oranges. Supposons que l'on reçoive 2 \$ pour chaque boule noire tirée et que l'on perde 1 \$ pour chaque boule blanche tirée. Désignons les gains nets par X . Quelles sont les valeurs possibles pour X et quelles sont les probabilités associées à ces valeurs?

4.11.2 On jette deux dés équilibrés et X désigne le produit des deux nombres obtenus. Calculer $P\{X = i\}, i = 1, 2, \dots$

4.11.3 On jette 3 dés et on admet que les $6^3 = 216$ résultats possibles sont tous équiprobables. X désigne la somme des 3 nombres obtenus. Donner les probabilités attachées aux différentes valeurs que X peut prendre.

4.11.4 On classe cinq hommes et cinq femmes selon leurs résultats lors d'un examen. On fait l'hypothèse que tous les scores sont différents et que les $10!$ classements possibles ont tous la même probabilité. On désigne le rang de la meilleure femme par X (par exemple X vaudra 2 si le meilleur résultat a été obtenu par un homme et le suivant par une femme). Trouver $P\{X = i\}, i = 1, 2, \dots, 10$.

4.11.5 Soit X la variable aléatoire comptant la différence entre les nombres de faces et de piles lors d'une répétition de n jets d'une pièce. Quelles sont les valeurs que peut prendre X ?

4.11.6 En admettant que dans le problème 4.11.5 la pièce ait été équilibrée, quelles sont les probabilités associées aux valeurs que X peut prendre lorsque $n = 3$?

4.11.7 On jette deux fois un dé. Quelles sont les valeurs que peuvent prendre les variables aléatoires suivantes:

- le plus grand des deux chiffres obtenus,
- le plus petit des deux chiffres obtenus,
- la somme des deux chiffres,
- la différence entre le premier chiffre et le second?

4.11.8 Si le dé utilisé pour le problème 4.11.7 est biaisé, calculer les probabilités associées aux valeurs des quatre variables aléatoires citées à ce même exercice.

4.11.9 Traiter l'exemple 4.2 dans le cas où les boules sont tirées avec remise.

4.11.10 Reprendre l'exemple 4.4 et calculer la probabilité de gagner i francs sachant que l'on gagne quelque chose; faire le calcul pour $i = 1, 2, 3$.

4.11.11

- On choisit au hasard un nombre entier N entre 1 et 1 000. Quelle est la probabilité que le nombre tiré soit divisible par 3? par 5? par 7? par 15? par 105? Que devient la réponse lorsqu'on remplace 1 000 par 10^k et que k devient de plus en plus grand?
- La fonction $\mu(n)$ de Möbius est importante en théorie des nombres. On peut montrer que ses propriétés sont en relation avec le problème non résolu sans doute le plus important en mathématiques, à savoir l'hypothèse de Riemann. Cette fonction est définie comme suit sur l'ensemble des entiers positifs: pour tout tel entier n , on considère sa décomposition en facteurs premiers. Si dans cette décomposition un facteur se répète, comme dans $12 = 2 \cdot 2 \cdot 3$ ou $49 = 7 \cdot 7$, $\mu(n)$ est déclaré nul. Si tous les facteurs sont distincts on affecte à $\mu(n)$ la valeur 1 dans le cas où ces facteurs sont en nombre impair et -1 s'ils sont en nombre pair. Par exemple $\mu(6) = -1$ car $6 = 2 \cdot 3$, mais $\mu(30) = 1$ car $30 = 2 \cdot 3 \cdot 5$. Soit donc un entier N choisi au hasard entre 1 et 10^k où k est grand. Donner la loi de probabilité de $\mu(N)$ lorsque k tend vers l'infini.

Pour calculer $P\{\mu(N) \neq 0\}$ on peut utiliser l'identité

$$\prod_{i=1}^{\infty} \frac{P_i^2 - 1}{P_i^2} = \left(\frac{3}{4}\right)\left(\frac{8}{9}\right)\left(\frac{24}{25}\right)\left(\frac{48}{49}\right) \cdots = \frac{6}{\pi^2}$$

où P_i est le i -ème nombre premier en partant des plus petits, 1 n'étant pas considéré comme un nombre premier.

4.11.12 Dans le jeu «pair – impair» les deux participants montrent chacun un ou deux doigts et en même temps annoncent combien de doigts ils pensent que leur adversaire va montrer. Si l'un seulement des joueurs devine juste il gagne un nombre de francs égal au total des doigts montrés par lui et son concurrent. Si les deux devinent correctement ou si les deux se trompent aucun argent n'est échangé. On considère l'un des deux joueurs et désigne par X le montant qu'il gagnera lors d'une unique partie de «pair – impair».

- Si les joueurs agissent indépendamment l'un de l'autre et si les 4 issues possibles au jeu sont équiprobables, quelles valeurs X peut-elle prendre et quelles sont les probabilités qui leur sont associées?

b) On admet toujours que les deux personnes jouent indépendamment l'une de l'autre. Mais chaque joueur décide maintenant de lever autant de doigts qu'il pense en voir apparaître chez l'autre. Chaque joueur a autant de chances de montrer 1 doigt que deux. Quelles sont les valeurs possibles de X et les probabilités qui leur sont associées?

4.11.13 Un vendeur a fixé deux rendez-vous pour vendre des encyclopédies. Au premier rendez-vous, il vendra un livre avec une probabilité .3 alors qu'au second, il en vendra un avec une probabilité .6. A chaque vente, il y a autant de chances de vendre le modèle de luxe qui coûte 1 000 Frs que le modèle standard qui coûte 500 Frs. Déterminer la distribution de X , la valeur totale en francs de toutes les ventes.

4.11.14 Cinq nombres distincts sont distribués aléatoirement à des joueurs numérotés de 1 à 5. Lorsque deux joueurs comparent leur numéro, celui qui a le plus grand est déclaré vainqueur. Au départ, les joueurs 1 et 2 comparent leur numéro; le vainqueur compare le sien avec le joueur 3, etc. Soit X le nombre de fois où le joueur 1 gagne. Trouver $P\{X = i\}$, $i = 0, 1, 2, 3, 4$.

4.11.15 La loterie de la «National Basketball Association (NBA)» implique les 11 équipes ayant les plus mauvais rapports victoire / défaite de l'année. 66 balles sont placées dans une urne. Sur chacune des balles est inscrit le nom d'une équipe; 11 possèdent le nom de l'équipe ayant le plus mauvais rapport, 10 possèdent le nom de l'équipe ayant le deuxième plus mauvais rapport, 9 possèdent le nom de l'équipe ayant le troisième plus mauvais rapport, etc. (1 balle ayant le nom de l'équipe au onzième plus mauvais rapport). Une balle est choisie au hasard et l'équipe dont le nom est sur la balle peut choisir en premier un joueur sur la liste de recrutement. Une autre balle est tirée et si elle représente une équipe différente de celle choisie en premier, alors l'équipe dont le nom est sur cette deuxième balle peut choisir à son tour un joueur sur la liste de recrutement. (Si la balle représente l'équipe qui a eu le premier choix, elle est mise de côté et on en choisit une autre; ceci continue jusqu'à ce qu'une balle d'une autre équipe soit choisie.) Finalement, une autre balle est choisie et l'équipe marquée sur la balle (à condition qu'elle soit différente des deux équipes précédentes) peut choisir un joueur en troisième. Les joueurs 4 à 11 restant dans la liste de recrutement sont attribués aux 8 équipes qui n'ont pas «gagné à la loterie» selon l'ordre inverse de leurs résultats. Par exemple, si l'équipe au pire résultat ne fait pas partie des 3 vainqueurs de la loterie, elle pourra choisir en quatrième. Soit X la position du choix de l'équipe au plus mauvais résultat. Quelle est la distribution de X ?

4.11.16 Dans le problème 4.11.15, soit l'équipe 1 celle ayant le plus mauvais score, soit l'équipe 2 celle ayant le deuxième plus mauvais score, etc. Soit Y_i l'équipe qui peut choisir en i -ième position. Donc $Y_1 = 3$ si la première balle choisie représente l'équipe 3. Trouver la distribution de

- a) Y_1 ;
- b) Y_2 ;
- c) Y_3 .

4.11.17 La fonction de répartition d'une variable X est la suivante:

$$F(b) = \begin{cases} 0 & b < 0 \\ \frac{b}{4} & 0 \leq b < 1 \\ \frac{1}{2} + \frac{b-1}{4} & 1 \leq b < 2 \\ \frac{11}{12} & 2 \leq b < 3 \\ 1 & 3 \leq b \end{cases}$$

a) Trouver $P\{X = i\}$, $i = 1, 2, 3$. b) Trouver $P\{\frac{1}{2} < X < \frac{3}{2}\}$.

4.11.18 On lance quatre fois une pièce équilibrée. X désigne le nombre de piles obtenus. Représenter graphiquement la loi de probabilité de $X - 2$.

4.11.19 La fonction de répartition de X est donnée par

$$F(b) = \begin{cases} 0 & b < 0 \\ \frac{1}{2} & 0 \leq b < 1 \\ \frac{3}{5} & 1 \leq b < 2 \\ \frac{4}{5} & 2 \leq b < 3 \\ \frac{9}{10} & 3 \leq b < 3.5 \\ 1 & b \geq 3.5 \end{cases}$$

Calculer la loi de probabilité de X .

4.11.20 Un livre de jeux recommande la «stratégie gagnante» suivante pour le jeu de la roulette. Il recommande de miser 1 Fr sur le rouge. Si le rouge apparaît (ce qui a une probabilité $18/38$), le joueur doit prendre son profit de 1 Fr et arrêter. Si le joueur perd (ce qui a une probabilité $20/38$), il doit à nouveau miser 1 Fr sur le rouge lors des deux prochains tours puis arrêter. Soit X le gain du joueur quand il s'arrête.

a) Trouver $P\{X > 0\}$.

b) Etes-vous convaincus que cette stratégie est vraiment une stratégie «gagnante»? Expliquer votre réponse!

c) Calculer $E[X]$.

4.11.21 Quatre bus transportant 148 élèves de la même école arrivent à un stade de football. Les bus transportent respectivement 40, 33, 25 et 50 élèves. Un des étudiants est choisi au hasard. Soit X le nombre d'étudiants qui était dans le bus de cet élève choisi aléatoirement. Un des quatre chauffeurs de bus est également choisi au hasard. Soit Y le nombre d'élèves dans son bus.

a) Entre $E[X]$ et $E[Y]$, de laquelle diriez-vous qu'elle est la plus grande? Pourquoi?

b) Calculer $E[X]$ et $E[Y]$.

4.11.22 Supposer que deux équipes jouent une série de matchs qui se termine lorsqu'une des deux a gagné i matchs. Supposer que chaque match joué est, indépendamment, gagné par le joueur A avec probabilité p . Trouver l'espérance du nombre de matchs joués lorsque

a) $i = 2$;

b) $i = 3$.

Montrer aussi dans les deux cas que ce nombre est maximisé lorsque $p = 1/2$.

4.11.23 On sait qu'une boîte de 5 composants électriques en comporte deux qui sont défectueux. Les composants sont choisis au hasard et testés l'un après l'autre. Trouver l'espérance du nombre de test qu'il faudra effectuer pour trouver les deux éléments défectueux.

4.11.24 A et B jouent au jeu suivant: A écrit soit le nombre 1, soit le nombre 2 et B doit deviner lequel a été écrit. Si i est le nombre écrit par A et que B le devine correctement, B reçoit i unités de la part de A . Si B se trompe, alors B paie $\frac{3}{4}$ d'unité à A . Si B prend sa décision de façon aléatoire mais en accordant à 1 le poids p et 2 le poids $(1 - p)$, déterminer l'espérance de son gain dans les cas suivants:

a) A a écrit le nombre 1;

b) A a écrit le nombre 2.

Quelle est la valeur de p qui rend maximal le minimum des espérances de gain de B et combien vaut ce maximum? (On remarquera que l'espérance du gain de B dépend non seulement de p , mais aussi de ce que A fait).

Considérons maintenant le joueur A . Supposons que lui aussi prend sa décision au hasard en écrivant le nombre 1 avec probabilité q . Quelle est l'espérance de perte de A si

c) B choisit le nombre 1;

d) B choisit le nombre 2.

Quelle est la valeur de q rendant minimum le maximum des espérances de perte de A ? Montrer que le minimum du maximum des espérances de perte de A est égal au maximum du minimum des espérances de perte de B . Ce résultat, connu sous le nom de théorème du minimax, fut démontré dans toute sa généralité à l'origine par le mathématicien John von Neumann. Il constitue le résultat fondamental de la discipline mathématique qu'est la théorie des jeux. La valeur commune est appelée valeur du jeu pour le joueur B .

4.11.25 On choisit au hasard un nombre compris entre 1 et 10. Vous devez deviner ce nombre en posant des questions auxquelles il ne sera répondu que par oui ou non. Calculer l'espérance du nombre de questions nécessaires dans les deux cas suivants:

a) votre i -ème question est du type «Est-ce i ?», $i = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$;

b) avec chaque question, vous essayez d'éliminer à peu près la moitié des nombres encore possibles.

4.11.26 Une compagnie d'assurance établit un contrat stipulant qu'une somme d'argent A doit être versée si un événement E se produit dans un intervalle d'un an. La compagnie estime que la probabilité que E se produise en l'espace d'un an est p . Comment calculer la prime d'assurance de façon que le bénéfice représente 10% de A ?

4.11.27 Le type le plus répandu de machines à sous possède trois roues munies chacune de 20 symboles (cerises, citrons, prunes, oranges, cloches et barres). Voici la description d'un jeu typique de ces roues:

	Roue 1	Roue 2	Roue 3
Cerises	7	7	0
Oranges	3	7	6
Citrons	3	0	4
Prunes	4	1	6
Cloches	2	2	3
Barres	1	3	1
	20	20	20

Ce tableau indique que des 20 symboles de la roue n° 1, 7 sont des cerises, 3 des oranges, etc... Le gain ordinaire par pièce mise est indiqué dans le tableau suivant:

Roue 1	Roue 2	Roue 3	Gain
barre	barre	barre	60
cloche	cloche	cloche	20
cloche	cloche	barre	18
prune	prune	prune	14
orange	orange	orange	10
orange	orange	barre	8
cerise	cerise	n'importe quoi	4
cerise	tout sauf cerise	n'importe quoi	2
autres combinaisons			-1

Calculer le gain que l'on peut espérer en jouant une partie avec une telle machine. On admet que les roues se meuvent de manière indépendante.

4.11.28 Un échantillon de trois objets est choisi au hasard d'une boîte en contenant 20, dont 4 sont défectueux. Trouver l'espérance du nombre des objets défectueux dans l'échantillon.

4.11.29 Une machine peut tomber en panne pour deux raisons. Le diagnostic de la première cause coûte C_1 francs; s'il est positif, la réparation coûte alors R_1 francs. De façon analogue, la seconde cause de panne occasionne des coûts C_2 et R_2 . Soient p et $(1 - p)$ respectivement les probabilités d'occurrence de la première et de la seconde pannes. Quelles sont les conditions que doivent satisfaire $p, C_p, R_p, i = 1, 2$ pour qu'il revienne en moyenne moins cher d'examiner la première cause de panne d'abord, plutôt que de conduire l'examen de façon inverse? On admettra que si le premier examen est négatif, le deuxième devra malgré tout être fait.

4.11.30 Un individu jette une pièce de monnaie équilibrée jusqu'à ce que pile apparaisse pour la première fois. Si pile apparaît au n -ième jet, l'individu gagne 2^n francs. Soit X , le gain du joueur. Montrer que $E[X] = +\infty$. Ce problème porte le nom de paradoxe de St-Petersbourg.

- Seriez-vous disposé à payer 1 million pour jouer une fois à ce jeu?
- Seriez-vous disposé à payer 1 million par partie en admettant que vous puissiez jouer aussi longtemps que vous le désirez et que vous n'ayez à régler les comptes qu'au moment de l'arrêt du jeu?

4.11.31 Chaque nuit, différents météorologues nous donnent la probabilité qu'il pleuve le lendemain. Pour juger leurs prédictions, nous allons les noter de la manière suivante: si un météorologue prévoit qu'il pleuvra avec probabilité p , il recevra une note de

$$\begin{aligned} 1 - (1 - p)^2 & \text{ s'il pleut} \\ (1 - p)^2 & \text{ s'il ne pleut pas} \end{aligned}$$

Nous regarderons les notes sur un certain laps de temps et conclurons que le météorologue ayant la meilleure moyenne est le meilleur prédicteur. Supposons à présent qu'un météorologue donné est au courant de ce procédé et souhaite maximiser l'espérance de son score. Si cette personne croit vraiment qu'il pleuvra demain avec probabilité p^* , quelle valeur de p doit-elle proposer pour maximiser l'espérance de son score?

4.11.32 Cent personnes subissent une analyse de sang pour qu'on puisse déterminer si oui ou non elles souffrent d'une certaine maladie. Cependant, plutôt que de tester chaque personne individuellement, il a été décidé de former des groupes de dix personnes. Les échantillons de sang des dix personnes de chaque groupe seront mélangés et analysés ensemble. Si le test est négatif, un seul test suffira pour ces dix personnes; cependant, si le test est positif, chacune des dix personnes sera examinée individuellement et en tout, 11 tests seront effectués pour ce groupe. On suppose que la probabilité qu'une personne soit atteinte de la maladie est 0.1 et que la maladie frappe les gens indépendamment les uns des autres. Calculer l'espérance du nombre de tests qu'il faudra faire sur les cent personnes. On admet ici que l'échantillon commun de 10 personnes sera positif dès qu'au moins une de ces personnes est malade.

4.11.33 Un vendeur de journaux achète ses journaux 10 centimes et les revend 15 centimes. Cependant, il ne peut pas se faire rembourser les exemplaires invendus. Si la demande journalière est une variable aléatoire binomiale de paramètres $n = 300$ et $p = 1/3$, quel est approximativement le nombre de journaux qu'il doit acheter afin de maximiser l'espérance de son bénéfice?

4.11.34 Supposons que le grand magasin décrit dans l'exemple 4.12 encoure un coût additionnel c pour chaque demande non satisfaite (ceci est fréquemment appelé un coût en «goodwill» car le magasin perd un peu de la confiance des clients dont la demande n'est pas satisfaite). Calculer l'espérance de profit si le stock est de s unités et déterminer la valeur de s qui maximise ce profit.

4.11.35 Une boîte contient 5 billes rouges et 5 billes bleues. Deux billes sont tirées au hasard. Si elles sont de la même couleur, vous gagnez 1.10 Frs; si elles sont de couleurs différentes, vous perdez 1.00 Fr. Calculer

- l'espérance du gain;
- la variance du gain.

4.11.36 Considérer le problème 4.11.22 avec $i = 2$. Trouver la variance du nombre de matchs joués et montrer que ce nombre est maximisé lorsque $p = 1/2$.

4.11.37 Trouver $\text{Var}(X)$ et $\text{Var}(Y)$ pour X et Y donnés dans le problème 4.11.21.

4.11.38 Si $E[X] = 1$ et $\text{Var}(X) = 5$, trouver

- $E[(2 + X)^2]$;
- $\text{Var}(4 + 3X)$.

4.11.39 On tire une boule d'une urne en contenant 3 blanches et 3 noires. On la replace après tirage, pour recommencer indéfiniment cette séquence d'opérations. Quelle est la probabilité de trouver exactement deux boules blanches parmi les quatre premières boules tirées?

4.11.40 Un examen est administré sous forme d'un questionnaire de 5 questions à 3 choix multiples chacune. Quelle est la probabilité qu'un étudiant obtienne 4 bonnes réponses ou plus en devinant?

4.11.41 Un homme prétend avoir des capacités de perception extrasensorielle. Le test qu'on lui administre consiste à lui faire deviner les 10 résultats des 10 jets d'une pièce équilibrée. Il donne 7 bonnes réponses. Quelle est la probabilité qu'il obtienne un résultat aussi bon ou meilleur s'il n'a aucune capacité de perception extrasensorielle?

4.11.42 Les moteurs d'un avion ont une probabilité $1 - p$ de défaillance en cours de vol, et ce indépendamment les uns des autres. Un avion a besoin d'une majorité de ses moteurs pour pouvoir terminer son vol. Pour quelles valeurs de p un avion à cinq moteurs est-il préférable à un trimoteur?

4.11.43 Un canal de transmission d'information ne peut traiter que des 0 et des 1. A cause de perturbations dues à l'électricité statique chaque chiffre transmis l'est avec une probabilité d'erreur de 0,2. Admettons que l'on veuille transmettre un message important limité à 1 signal binaire. Pour éviter une erreur on transmettra 00000 au lieu de 0 et 11111 au lieu de 1. Si le récepteur décode suivant la règle de majorité, quelle est la probabilité que le message soit mal interprété? Quelles hypothèses d'indépendance devez-vous faire?

4.11.44 Un système de communication par satellite se compose de n sous-systèmes et fonctionne un jour donné si ce jour-là au moins k des sous-systèmes sont opérationnels. Par temps pluvieux, chaque sous-système fonctionne avec probabilité p_1 , indé-

pendamment des autres. De même par temps sec, mais avec une probabilité p_2 . Si α désigne la probabilité qu'il pleuve demain, quelle est la probabilité que le système total fonctionne alors?

4.11.45 Un étudiant se prépare à passer un examen oral important. Il se préoccupe de la question de savoir s'il sera en forme ou non. Son opinion est que s'il est en forme chacun de ses examinateurs le jugera suffisant avec une probabilité de 0,8 et indépendamment des autres examinateurs. Dans le cas contraire, cette probabilité tombe à 0,4. L'étudiant est promu si une majorité de ses examinateurs le juge suffisant. Par ailleurs, il pense avoir deux fois plus de chances d'être en méforme qu'en forme. A-t-il plus intérêt à demander un contrôle par 3 ou 5 examinateurs?

4.11.46 Au moins 9 des 12 jurés réunis doivent estimer l'accusé coupable pour rendre le jugement exécutoire. Supposons que la probabilité pour un juré d'estimer un coupable innocent est 0,2 tandis qu'elle est de 0,1 de commettre l'erreur contraire. Les jurés décident en toute indépendance et 65% des accusés sont coupables. Trouver la probabilité que le jury rende une sentence correcte. Quel pourcentage des accusés sera condamné?

4.11.47 Dans certains tribunaux militaires on désigne 9 juges pour une affaire. Cependant le procureur, autant que l'avocat de la défense, peuvent faire opposition à la désignation de tout juge, auquel cas le juge écarté n'est pas remplacé. Un accusé est déclaré coupable si la majorité des juges le déclarent coupable et est considéré comme innocent sinon. On suppose que dans le cas d'un accusé réellement coupable chaque juge votera la culpabilité (indépendamment des autres) avec probabilité 0,7; cette probabilité n'est que 0,3 lorsque l'accusé est innocent.

- Quelle est la probabilité qu'un accusé coupable soit jugé tel s'il y a 9 juges? 8 juges? 7 juges?
- Qu'en est-il si l'accusé est innocent?
- Dans un cas le procureur n'exerce pas son droit d'opposition. Par ailleurs l'avocat de la défense est limité à 2 oppositions. Combien d'oppositions a-t-il intérêt à faire s'il pense que son client a 60% de risques d'être coupable?

4.11.48 On sait que les disquettes produites par une certaine firme sont défectueuses avec une probabilité de 0,01, indépendamment les unes des autres. La compagnie vend les disquettes par lots de 10 et garantit contre remboursement qu'au plus 1 des 10 disquettes du lot est défectueuse. A l'achat de 3 lots, quelle est la probabilité qu'un lot exactement doive être retourné?

4.11.49 On suppose que 10 % des puces produites par une usine de matériel d'ordinateurs sont défectueuses. Si l'on commande 100 puces, le nombre de puces défectueuses suit-il une loi binomiale?

4.11.50 Supposer qu'on lance dix fois une pièce biaisée qui tombe sur face avec probabilité p . Sachant qu'on a obtenu 6 faces, trouver la probabilité conditionnelle que les 3 premiers tirages soient

- F, P, P (signifiant que le premier tirage est face et que les deux autres sont pile);
- P, F, P .

4.11.51 L'espérance du nombre d'erreurs typographiques sur une page d'un certain magazine est .2. Quelle est la probabilité que la prochaine page lue contienne a) 0, b) 2 ou plus d'erreurs typographiques? Expliquer votre raisonnement!

4.11.52 Le nombre moyen d'accidents d'avions commerciaux par mois dans le monde est 3.5. Quelle est la probabilité qu'il y ait

a) au moins 2 accidents le mois prochain;

b) au plus 1 accident le mois prochain?

Expliquer votre raisonnement!

4.11.53 On a célébré environ 80 000 mariages l'an dernier dans l'état de New-York. Estimer la probabilité que pour au moins un de ces couples

a) les deux époux soient nés le 30 avril;

b) les deux époux célèbrent leurs anniversaires le même jour de l'année.

Poser vos hypothèses.

4.11.54 Supposer que le nombre moyen de voitures abandonnées chaque semaine sur une certaine autoroute est 2.2. Calculer la probabilité qu'il y ait

a) aucune voiture abandonnée la semaine prochaine;

b) au moins 2 voitures abandonnées la semaine prochaine.

4.11.55 Une agence de dactylographie emploie 2 dactylos. Le nombre d'erreurs par article est 3 pour le premier dactylo et 4.2 pour le second. Si un article a autant de chance d'être tapé par l'un ou l'autre dactylo, quelle est la probabilité qu'il ne contienne pas d'erreur?

4.11.56 Combien de personnes faut-il pour que la probabilité qu'au moins une d'entre elles soit née le même jour que vous soit supérieure à $1/2$?

4.11.57 On admet que le nombre d'accidents survenant sur une autoroute quotidiennement est une variable aléatoire de Poisson de paramètre $\lambda = 3$.

a) Quelle est la probabilité qu'il survienne 3 accidents ou plus lors d'un jour donné?

b) Même question si l'on sait qu'un accident au moins a eu lieu.

4.11.58 Comparer l'approximation poissonnienne aux probabilités exactes données par la loi binomiale dans les cas suivants:

a) $P\{X = 2\}$ lorsque $n = 8, p = 0,1$;

b) $P\{X = 9\}$ lorsque $n = 10, p = 0,95$;

c) $P\{X = 0\}$ lorsque $n = 10, p = 0,1$;

d) $P\{X = 4\}$ lorsque $n = 9, p = 0,2$.

4.11.59 Vous participez à 50 tirages consécutifs d'une loterie. A chaque tirage la probabilité que vous gagniez un prix est $\frac{1}{100}$. Quelle est la probabilité (approximative) que vous gagniez un prix

a) au moins une fois;

b) exactement une fois;

c) au moins deux fois?

4.11.60 Le nombre de rhumes attrapés par un individu en l'espace d'un an est une variable aléatoire de Poisson de paramètre $\lambda = 5$. Admettons qu'un remède miracle (basé sur l'effet de vitamine C à haute dose) ait été lancé sur le marché et qu'il abaisse le paramètre λ à 3 pour 75% de la population. Pour les 25 derniers pourcents de la population le remède n'a pas d'effet appréciable. Un individu essaie ce médicament pendant un an et attrape deux rhumes. Quelle est la probabilité que le remède ait un effet sur lui?

4.11.61 Au poker, la probabilité de se voir distribuer une main pleine est approximativement 0,0014. Calculer une approximation de la probabilité d'obtenir au moins deux mains pleines sur 1 000 donnes.

4.11.62 Si l'on place n couples mariés autour d'une table, calculer la probabilité approximative qu'aucune femme ne se trouve à côté de son mari. Lorsque $n = 10$ comparer le résultat avec la valeur exacte donnée dans l'exemple 2.14 du chapitre 2.

4.11.63 Les gens entrent dans un casino au rythme d'une personne toutes les deux minutes.

- Quelle est la probabilité qu'il n'entre personne entre 12 h et 12 h 05?
- Quelle est la probabilité que 4 personnes au moins se présentent durant cette même période?

4.11.64 Le taux de suicide pour un pays donné est de 1 personne pour 100 000 habitants et par mois.

- Quelle est la probabilité qu'il y ait 8 suicides ou plus en un mois dans une ville de 400 000 âmes?
 - Quelle est la probabilité qu'au cours d'une année le nombre de suicides mensuels dépasse deux fois ou plus le niveau de 8?
 - Le mois en cours étant appelé mois 1, quelle est la probabilité que le premier mois où l'on enregistre 8 suicides ou plus soit le mois i , $i \geq 1$?
- Quelles hypothèses faites-vous?

4.11.65 Chacun des soldats d'une troupe de 500 hommes est porteur d'une certaine maladie avec probabilité $\frac{1}{1000}$. Cette maladie est détectable à l'aide d'un test sanguin et, pour faciliter les choses, on ne teste qu'un mélange du sang de chacun des 500 soldats.

- Quelle est la probabilité (approximative) que le test soit positif, indiquant par là qu'au moins une des personnes est malade?

On suppose par la suite que le test a été positif.

- Quelle est la probabilité que dans ce cas plus d'une personne soit malade?
- L'une de ces 500 personnes s'appelle Jones, et Jones sait qu'il est porteur de la maladie. Quelle doit être, de son point de vue, la probabilité qu'une autre personne au moins soit porteuse de la maladie?
- Le test étant positif, il est décidé que des tests individuels seront menés. Les $i - 1$ premiers de ces tests sont négatifs. Le i -ème est positif – c'est celui de Jones. Quelle est la probabilité qu'une des personnes restantes au moins soit encore malade, en fonction de i ?

4.11.66 On considère une roue de roulette comprenant 38 cases numérotées 0, 00 et de 1 à 36. Smith parie régulièrement sur la sortie des numéros 1 à 12. a) Quelle est la probabilité qu'il perde ses 5 premiers paris? b) Quelle est la probabilité que son premier gain survienne lors du quatrième tirage?

4.11.67 Deux équipes de sportifs disputent une série de matchs. La première équipe à enregistrer 4 victoires est déclarée gagnante de la série. On admet que l'une d'elles est plus forte que l'autre et remporte un match avec probabilité 0,6, indépendamment de l'issue des autres parties. Trouver la probabilité que cette équipe remporte la série en i jeux exactement. Calculer ce résultat pour $i = 4, 5, 6, 7$. Comparer ces résultats avec celui obtenu sous l'hypothèse que l'équipe gagnante est la première à enregistrer 2 victoires seulement.

4.11.68 Supposer dans le problème 4.11.67 que les deux équipes sont de force égale et que chacune a une probabilité $1/2$ de gagner chaque match. Trouver l'espérance du nombre de matchs joués.

4.11.69 Un journaliste se voit remettre une liste de personnes à interviewer. Il doit interroger 5 personnes au moins. Les interviewés potentiels n'acceptent de parler qu'avec une probabilité de $\frac{2}{3}$, indépendamment les uns des autres. Quelle est la probabilité qu'il puisse réaliser ses 5 entretiens si la liste compte

a) 5 noms?

b) 8 noms?

Dans ce dernier cas quelle est la probabilité qu'il puisse parler à

c) 6 personnes exactement?

d) 7 exactement?

4.11.70 On jette une pièce de monnaie jusqu'à obtenir pile pour la deuxième fois. La variable X compte le nombre d'apparitions de face. Quelle est la loi de X ?

4.11.71 Résoudre le problème des allumettes de Banach (exemple 4.34) lorsque la boîte de gauche contient initialement N_1 allumettes contre N_2 pour la boîte de droite.

4.11.72 Dans le problème des boîtes d'allumettes de Banach, trouver la probabilité que lorsque la première boîte est vidée (plutôt que trouvée vide), l'autre boîte contienne exactement k allumettes.

4.11.73 Une urne contient 4 boules blanches et 4 noires. On tire 4 boules au hasard. Si deux sont blanches et deux sont noires on s'arrête. Sinon on remet les boules dans l'urne et recommence le tirage, jusqu'à obtenir deux blanches et deux noires. Quelle est la probabilité qu'il faille exactement n tirages avant de s'arrêter?

4.11.74 Dans le problème 4.11.67, trouver la probabilité conditionnelle que l'équipe la plus forte

a) gagne la compétition sachant qu'elle a gagné le premier match;

b) gagne le premier match sachant qu'elle a gagné la compétition.

4.11.75 Keno est le nom d'un jeu populaire dans les maisons de jeux du Nevada. On y joue comme suit: la banque choisit au hasard 20 nombres parmi l'ensemble des nombres compris entre 1 et 80. Un joueur peut choisir entre 1 et 15 de ces 80 nombres. Un gain survient lorsqu'une certaine fraction des nombres du joueur correspond à certains de ceux choisis par la banque. Le montant du gain dépend du nombre d'éléments dans le jeu du joueur et du nombre de correspondances. Par exemple, si le joueur ne prend qu'un nombre il gagnera si ce nombre est dans le lot des 20 nombres de la banque; le rapport sera dans ce cas $2,2 : 1$, soit 2,20 dollars par dollar de mise. (Comme la probabilité de gagner dans cette situation est $\frac{3}{4}$, le rapport juste serait $3 : 1$). Lorsqu'un joueur prend deux nombres, le rapport est $12 : 1$ dans le cas où les deux nombres sont gagnants.

a) Quel serait le juste rapport dans ce dernier cas?

On note par $P_{n,k}$ la probabilité que k exactement des n nombres pris par le joueur soient gagnants.

b) Calculer $P_{n,k}$.

c) La mise la plus courante au Keno consiste à prendre 10 nombres. Le tableau 4.7 indique les rapports payés par la banque. Construire la dernière colonne de ce tableau.

<i>Nombre de concordances</i>	<i>Rapport brut par dollar de mise</i>	<i>Rapport équitable</i>
0-4	-1	
5	1	
6	17	
7	179	
8	1,299	
9	2,599	
10	24,999	

Tableau 4.8 Rapports au Keno pour des prises de 10 nombres

4.11.76 On considère la situation présentée dans l'exemple 4.38. Quel pourcentage de lots à i composants défectueuses sera-t-il rejeté par l'acheteur? Calculer le résultat pour $i = 1, \dots, 4$. Si un lot est rejeté, quelle est la probabilité qu'il contienne 4 composants défectueux?

4.11.77 Un industriel achète les transistors par lots de 20. Sa stratégie consiste à tester seulement 4 transistors par lot, pris au hasard, et à n'accepter le lot que si tous sont en bon état. Si la probabilité pour un transistor isolé d'être malformé est 0,1, ceci indépendamment de l'état des autres transistors, quelle proportion des lots sera refusée par l'industriel?

Calcul de la fonction de répartition binomiale (voir section 4.7.6)

```

10 PRINT "THE DISTRIBUTION FUNCTION OF A BINOMIAL(n,p) RANDOM VARIABLE"
20 PRINT "ENTER n"
30 INPUT N
40 PRINT "ENTER p"
50 INPUT P
60 PRINT "ENTER i"
70 INPUT I
80 S=(1-P)^N
90 IF S=0 GOTO 180
100 A=P/(1-P)
110 T=S
120 IF I=0 GOTO 390
130 FOR K=0 TO I-1
140 S=S*A*(N-K)/(K+1)
150 T=T+S
160 NEXT K
170 GOTO 390
180 J=I
190 IF J>N*P THEN J=INT(N*P)
200 FOR K=1 TO J
210 L=L+LOG(N+1-K)-LOG(J+1-K)
220 NEXT K
230 L=L+J*LOG(P)+(N-J)*LOG(1-P)
240 L=EXP(L)
250 B=(1-P)/P
260 F=1
270 FOR K=1 TO J
280 F=F*B*(J+1-K)/(N-J+K)
290 T=T+F
300 NEXT K
310 IF J=I GOTO 380
320 C=1/B
330 F=1
340 FOR K=1 TO I-J
350 F=F*C*(N+1-J-K)/(J+K)
360 T=T+F
370 NEXT K
380 T=(T+1)*L
390 PRINT "THE PROBABILITY IS";T
400 END

```

Calcul de la fonction de répartition de Poisson (voir section 4.8.10)

```

10 PRINT "THE PROBABILITY THAT A POISSON VARIABLE IS LESS THAN OR EQUAL TO i"
20 PRINT "ENTER THE MEAN OF THE RANDOM VARIABLE"
30 INPUT C
40 PRINT "ENTER THE DESIRED VALUE OF i"
50 INPUT I
60 S=EXP(-C)
70 IF S=0 GOTO 150
80 T=S
90 IF I=0 GOTO 340
100 FOR K=0 TO I-1
110 S=S*C/(K+1)
120 T=T+S
130 NEXT K
140 GOTO 340
150 J=I
160 IF J>C THEN J=INT(C)
170 FOR K=1 TO J
180 FAC=FAC+LOG(K)
190 NEXT K
200 L=-C-FAC+J*LOG(C)

```

```
210 L=EXP(L)
220 F=1
230 FOR K=1 TO J
240 F=F*(J+1-K)/C
250 T=T+F
260 NEXT K
270 IF J=I GOTO 330
280 F=1
290 FOR K=1 TO I-J
300 F=F*C/(K+J)
310 T=T+F
320 NEXT K
330 T=(T+1)*L
340 PRINT "THE PROBABILITY IS";T
350 END
```

Variables aléatoires continues

5.1 INTRODUCTION

5.1.1 Définitions

Dans le chapitre 4 nous avons traité des variables aléatoires discrètes, c'est-à-dire de variables dont l'ensemble des états est fini ou infini dénombrable. Il existe cependant des variables dont l'ensemble des états possibles est infini non dénombrable. On peut citer par exemple l'heure d'arrivée d'un train à une gare donnée ou encore la durée de vie d'un transistor. Désignons par X une telle variable. On qualifiera X de *variable aléatoire continue*¹ s'il existe une fonction f non négative définie pour tout $x \in \mathbb{R}$ et vérifiant pour tout ensemble B de nombres réels la propriété:

$$P\{X \in B\} = \int_B f(x) dx \quad (5.1)$$

La fonction f est appelée *densité de probabilité* de la variable aléatoire X .

En d'autres termes, (5.1) signifie que la probabilité que X prenne une valeur de B peut être obtenue en intégrant la densité de probabilité sur B . Du fait que X doit bien prendre une valeur, il résulte la contrainte suivante pour f :

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x) dx$$

Tous les problèmes de probabilité relatifs à X peuvent être traités grâce à f . Par exemple pour $B = [a, b]$ on obtient grâce à (5.1):

¹ On dit parfois absolument continue.

$$P\{a \leq X \leq b\} = \int_a^b f(x) dx \quad (5.2)$$

Si l'on pose $a = b$ dans (5.2), il résulte

$$P\{X = a\} = \int_a^a f(x) dx = 0$$

Ceci signifie en clair que la probabilité qu'une variable aléatoire continue prenne une valeur isolée fixe est toujours nulle. Aussi peut-on écrire pour une telle variable

$$P\{X < a\} = P\{X \leq a\} = F(a) = \int_{-\infty}^a f(x) dx$$

5.1.2 Exemples de variables aléatoires continues

Exemple 5.1 Supposons que X soit une variable aléatoire continue dont la densité est

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{sinon} \end{cases}$$

a) Quelle est la valeur de C ? b) Que vaut $P\{X > 1\}$?

SOLUTION. a) Du fait que f est une densité, elle vérifie la relation $\int_{-\infty}^{+\infty} f(x) dx = 1$, ce qui entraîne à son tour que

$$C \int_0^2 (4x - 2x^2) dx = 1$$

ou

$$C \left[2x^2 - \frac{2x^3}{3} \right] \Big|_{x=0}^{x=2} = 1$$

ou

$$C = \frac{3}{8}$$

Donc

$$b) \quad P\{X > 1\} = \int_1^{\infty} f(x) dx = \frac{3}{8} \int_1^2 (4x - 2x^2) dx = \frac{1}{2} \quad \blacksquare$$

Exemple 5.2 La durée de fonctionnement d'un ordinateur avant sa première panne est une variable aléatoire continue de densité donnée par

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

a) Quelle est la probabilité que cette durée de fonctionnement soit comprise entre 50 et 150 heures? b) Quelle est la probabilité que l'ordinateur fonctionne moins de 100 heures?

SOLUTION. a) Comme

$$1 = \int_{-\infty}^{\infty} f(x) dx = \lambda \int_0^{\infty} e^{-x/100} dx$$

on obtient

$$1 = -\lambda(100)e^{-x/100} \Big|_0^{\infty} = 100\lambda \quad \text{ou} \quad \lambda = \frac{1}{100}$$

Ainsi la probabilité que la durée de fonctionnement de l'ordinateur soit comprise entre 50 et 150 heures est donnée par

$$\begin{aligned} P\{50 < X < 150\} &= \int_{50}^{150} \frac{1}{100} e^{-x/100} dx = -e^{-x/100} \Big|_{50}^{150} \\ &= e^{-1/2} - e^{-3/2} \approx .383 \end{aligned}$$

b) De la même manière

$$P\{X < 100\} = \int_0^{100} \frac{1}{100} e^{-x/100} dx = -e^{-x/100} \Big|_0^{100} = 1 - e^{-1} \approx .632$$

En d'autres termes, l'ordinateur tombera en panne avant sa 100-ième heure de service 63,3 fois sur 100 en moyenne. ■

Exemple 5.3 La durée de vie d'un certain type de diode de radio est une variable aléatoire de densité donnée par

$$f(x) = \begin{cases} 0 & x \leq 100 \\ \frac{100}{x^2} & x > 100 \end{cases}$$

Quelle est la probabilité qu'exactly 2 des 5 diodes de ce type doivent être remplacées lors des 150 premières heures de service de la radio? On admettra que les événements E_i : «la i -ème diode doit être remplacée avant la 150-ième heure de service», $i = 1, 2, 3, 4, 5$, sont indépendants.

SOLUTION. On a

$$\begin{aligned} P(E_i) &= \int_0^{150} f(x) dx \\ &= 100 \int_{100}^{150} x^{-2} dx \\ &= \frac{1}{3} \end{aligned}$$

L'indépendance des E_i permet alors d'écrire la probabilité cherchée

$$\binom{5}{2} \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^3 = \frac{80}{243} \quad \blacksquare$$

5.1.3 Fonction de répartition d'une variable aléatoire continue

La relation entre la fonction de répartition F et la densité f d'une variable aléatoire continue X est donnée par

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^a f(x) dx$$

La dérivation des deux membres dans l'équation ci-dessus livre

$$\frac{d}{da} F(a) = f(a)$$

Autrement dit, la densité d'une variable aléatoire continue est la dérivée de la fonction de répartition.

5.1.4 Interprétation intuitive de la densité

On peut dériver de (5.2) une interprétation plus intuitive de la notion de densité. En effet,

$$P\left\{a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}\right\} = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x) dx \approx \varepsilon f(a)$$

lorsque ε est petit et si $f(\cdot)$ est continue en $x = a$. En d'autres termes, la probabilité que X prenne une valeur dans un intervalle de longueur ε centré en a est approximativement $\varepsilon f(a)$. On en conclut que $f(a)$ est une sorte de mesure de la probabilité que X soit proche de a .

5.2 ESPÉRANCE ET VARIANCE DE VARIABLES ALÉATOIRES CONTINUES

5.2.1 Définition de l'espérance d'une variable aléatoire continue

Dans le chapitre 4, nous avons défini l'espérance d'une variable aléatoire discrète X par

$$E\{X\} = \sum_x xP\{X = x\}$$

Si X est une variable aléatoire continue ayant pour densité $f(x)$, alors comme

$$f(x)dx \approx P\{x \leq X \leq x + dx\} \text{ pour } dx \text{ petit}$$

il est facile de voir que la définition analogue de l'espérance de X est

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

5.2.2 Exemples d'espérance de variables continues

Exemple 5.4 Trouver $E[X]$ lorsque la densité de X est

$$f(x) = \begin{cases} 2x & \text{si } 0 \leq x \leq 1 \\ 0 & \text{sinon} \end{cases}$$

SOLUTION.

$$\begin{aligned} E[X] &= \int xf(x)dx \\ &= \int_0^1 2x^2 dx \\ &= \frac{2}{3} \end{aligned}$$

Exemple 5.5 La densité de X est donnée par

$$f(x) = \begin{cases} 1 & \text{si } 0 \leq x \leq 1 \\ 0 & \text{sinon} \end{cases}$$

Trouver $E[e^X]$.

SOLUTION. Soit $Y = e^X$. Commençons par déterminer la fonction de distribution F_Y de Y . Pour $1 \leq x \leq e$,

$$\begin{aligned} F_Y(x) &= P\{Y \leq x\} \\ &= P\{e^X \leq x\} \\ &= P\{X \leq \log(x)\} \\ &= \int_0^{\log(x)} f(y)dy \\ &= \log(x) \end{aligned}$$

En dérivant $F_Y(x)$, on obtient la densité de Y :

$$f_Y(x) = \frac{1}{x}, \quad 1 \leq x \leq e.$$

Donc

$$\begin{aligned} E[e^X] &= E[Y] = \int_{-\infty}^{\infty} x f_Y(x) dx \\ &= \int_0^1 dx \\ &= e - 1 \end{aligned}$$

■

5.2.3 Espérance d'une fonction d'une variable aléatoire continue

Bien que la méthode employée dans l'exemple 5.5 pour calculer l'espérance d'une fonction de X soit toujours applicable, il existe, comme dans le cas discret, une autre façon de procéder. Le théorème suivant est une analogie directe du théorème 4.1 du chapitre 4.

Théorème 5.1

Si X est une variable aléatoire continue de densité $f(x)$, alors pour toute fonction réelle g on aura

$$E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Une application du théorème 5.1 à l'exemple 5.5 donne

$$\begin{aligned} E[e^X] &= \int_0^1 e^x dx \quad \text{comme } f(x) = 1, \quad 0 < x < 1 \\ &= e - 1 \end{aligned}$$

ce qui correspond au résultat de l'exemple.

La preuve du théorème 5.1 est plus compliquée que dans le cas discret et nous en présenterons une sous la condition que la variable aléatoire $g(X)$ est non négative. (Des indices pour la preuve générale sont donnés dans les exercices théoriques 5.8.2 et 5.8.3.) Pour la démonstration, nous aurons besoin du théorème suivant.

Théorème 5.2

Pour une variable aléatoire Y non négative,

$$E[Y] = \int_0^{\infty} P\{Y > y\} dy$$

DÉMONSTRATION. Nous présentons ici une preuve pour le cas où Y est une variable aléatoire continue de densité f_Y . On a

$$\int_0^{\infty} P\{Y > y\} dy = \int_0^{\infty} \int_y^{\infty} f_Y(x) dx dy$$

où nous avons utilisé le fait que $P\{Y > y\} = \int_y^{\infty} f_Y(x) dx$. En changeant l'ordre d'intégration dans l'équation précédente, on a

$$\begin{aligned} \int_0^{\infty} P\{Y > y\} dy &= \int_0^{\infty} \left(\int_0^x dy \right) f_Y(x) dx \\ &= \int_0^{\infty} x f_Y(x) dx \\ &= E[Y] \end{aligned}$$

■

DÉMONSTRATION DU THÉORÈME 5.1. Pour toute fonction g telle que $g(x) \geq 0$, on a d'après le théorème 5.2 que

$$\begin{aligned} E[g(x)] &= \int_0^{\infty} P\{g(x) > y\} dy \\ &= \int_0^{\infty} \int_{x:g(x)>y} f(x) dx dy \\ &= \int_{x:g(x)>0} \int_0^{g(x)} dy f(x) dx \\ &= \int_{x:g(x)>0} g(x) f(x) dx \end{aligned}$$

ce qui termine la démonstration. ■

5.2.4 Exemple d'espérance d'une fonction d'une variable aléatoire continue

Exemple 5.6 Supposer que si vous arrivez s minutes en avance à un rendez-vous, vous devez payer cs , et que si vous arrivez s minutes en retard, vous devez payer ks . Supposer que le temps de trajet du lieu où vous êtes au lieu de rendez-vous est une variable aléatoire continue de fonction de densité f . Déterminer le temps auquel vous devez partir pour minimiser l'espérance de votre coût.

SOLUTION. Soit X le temps de trajet. Si vous partez t minutes avant votre rendez-vous, alors votre coût $C_t(X)$ est donné par

$$C_t(X) = \begin{cases} c(t - X) & \text{si } X \leq t \\ k(X - t) & \text{si } X \geq t \end{cases}$$

Par conséquent,

$$\begin{aligned}
 E[C_t(X)] &= \int_0^{\infty} C_t(x) f(x) dx \\
 &= \int_0^t c(t-x) f(x) dx + \int_t^{\infty} k(x-t) f(x) dx \\
 &= ct \int_0^t f(x) dx - c \int_0^t xf(x) dx + k \int_t^{\infty} xf(x) dx - kt \int_t^{\infty} f(x) dx
 \end{aligned}$$

La valeur de t qui minimise $E[C_t(X)]$ peut être obtenue par calcul. On obtient en dérivant

$$\begin{aligned}
 \frac{d}{dt} E[C_t(X)] &= ct f(t) + cF(t) - ct f(t) - kf(t) + kf(t) - k[1 - F(t)] \\
 &= (k + c)F(t) - k
 \end{aligned}$$

En égalant ceci à zéro, on voit que l'espérance minimale du coût est obtenue en partant t^* minutes avant le rendez-vous, où t^* satisfait

$$F(t^*) = \frac{k}{k + c}$$

5.2.5 Linéarité de l'espérance

Comme dans le chapitre 4, nous pouvons utiliser le théorème 5.1 pour montrer le théorème suivant.

Théorème 5.3

Pour toute paire a, b de constantes, on a

$$E[aX + b] = aE[X] + b$$

La démonstration du théorème 5.3 pour une variable aléatoire continue X est la même que celle donnée pour une variable aléatoire discrète. La seule modification est le remplacement de la somme par une intégrale et de la fonction de répartition par la densité.

5.2.6 Définition de la variance d'une variable aléatoire continue

La variance d'une variable aléatoire continue est définie exactement comme celle d'une variable discrète. C'est à dire que si X est une variable aléatoire d'espérance μ , la variance de X est définie (pour tous les types de variables aléatoires) par

$$\text{Var}(X) = E[(X - \mu)^2]$$

L'autre formule

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

est établie de la même manière que dans le cas discret.

5.2.7 Exemple de variance d'une variable aléatoire continue

Exemple 5.7 Trouver $\text{Var}(X)$ pour X donnée dans l'exemple 5.4.

SOLUTION. Calculons d'abord $E[X^2]$.

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \int_0^1 2x^3 dx \\ &= \frac{1}{2} \end{aligned}$$

Donc, puisque $E[X] = \frac{2}{3}$, on obtient

$$\text{Var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

■

On peut montrer, par une preuve imitant celle donnée pour les variables aléatoires discrètes, que pour les constantes a et b

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

En théorie des probabilités, il existe plusieurs classes importantes de variables aléatoires continues. Les sections suivantes sont consacrées à l'étude de quelques-unes de ces classes.

5.3 VARIABLE ALÉATOIRE UNIFORME

5.3.1 Variable uniforme sur $(0, 1)$

Une variable aléatoire est dite *uniformément* distribuée sur l'intervalle $(0, 1)$ si sa densité est

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{sinon} \end{cases} \quad (5.3)$$

On vérifie que (5.3) correspond bien à une densité puisque $f(x) \geq 0$ et $\int_{-\infty}^{+\infty} f(x) dx = \int_0^1 dx = 1$. Comme $f(x) > 0$ seulement lorsque $x \in (0, 1)$, on en déduit que X ne prend des valeurs que dans cet intervalle. De plus, X a autant de chances d'être près de n'importe quelle valeur de $(0, 1)$ plutôt que de n'importe quelle autre puisque $f(x)$ est constante sur cet intervalle. On le vérifie en prenant deux nombres a et b quelconques tels que $0 < a < b < 1$; alors,

$$P\{a \leq X \leq b\} = \int_a^b f(x) dx = b - a$$

En d'autres termes, la probabilité que X prenne une valeur dans un sous-intervalle de $(0, 1)$ est égale à la longueur de ce sous-intervalle.

5.3.2 Variable uniforme quelconque

En généralisant, une variable aléatoire est uniforme sur l'intervalle (α, β) si sa densité est

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{si } \alpha < x < \beta \\ 0 & \text{sinon} \end{cases} \quad (5.4)$$

5.3.3 Fonction de répartition de variable aléatoire uniforme

A partir de $F(a) = \int_{-\infty}^a f(x) dx$ et de (5.4) on obtient la fonction de répartition d'une variable aléatoire uniforme sur l'intervalle (α, β) :

$$F(a) = \begin{cases} 0 & a \leq \alpha \\ \frac{a - \alpha}{\beta - \alpha} & \alpha < a < \beta \\ 1 & a \geq \beta \end{cases}$$

La figure 5.1 représente les graphes de f et F dans le cas général.

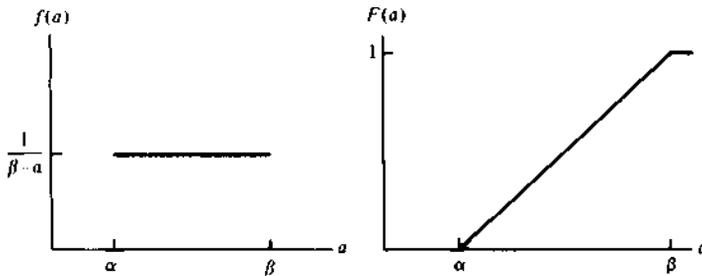


Figure 5.1 Graphes de f (à gauche) et F (à droite) pour une variable aléatoire uniforme sur (α, β) .

5.3.4 Exemples de variables aléatoires uniformes

Exemple 5.8 Soit X uniformément distribuée sur (α, β) . Trouver

- a) $E[X]$;
b) $\text{Var}(X)$.

SOLUTION a)

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha} dx \\ &= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\ &= \frac{\beta + \alpha}{2} \end{aligned}$$

En d'autres termes, l'espérance d'une variable aléatoire uniformément distribuée sur un intervalle est égale au milieu de l'intervalle.

b) Pour trouver $\text{Var}(X)$, calculons d'abord $E[X^2]$.

$$\begin{aligned} E[X^2] &= \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} x^2 dx \\ &= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} \\ &= \frac{\beta^2 + \alpha\beta + \alpha^2}{3} \end{aligned}$$

Donc

$$\begin{aligned} \text{Var}(X) &= \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \frac{(\alpha + \beta)^2}{4} \\ &= \frac{(\beta - \alpha)^2}{12} \end{aligned}$$

Par conséquent, la variance d'une variable aléatoire uniformément distribuée sur un intervalle est le carré de la longueur de l'intervalle divisé par 12. ■

Exemple 5.9 Soit X une variable uniforme sur $(0, 10)$. Calculer les probabilités suivantes: a) $P\{X < 3\}$, b) $P\{X > 6\}$, c) $P\{3 < X < 8\}$.

SOLUTION.

$$\text{a) } P\{X < 3\} = \int_0^3 \frac{1}{10} dx = \frac{3}{10}$$

$$\text{b) } P\{X > 6\} = \int_6^{10} \frac{1}{10} dx = \frac{4}{10}$$

$$\text{c) } P\{3 < X < 8\} = \int_3^8 \frac{1}{10} dx = \frac{1}{2} \quad \blacksquare$$

Exemple 5.10 A partir de 7 heures, les bus passent toutes les 15 minutes à un arrêt donné. Ils passent donc à 7 h 00, 7 h 15, 7 h 30 et ainsi de suite. Un usager se présente entre 7 h 00 et 7 h 30 à cet arrêt, l'heure exacte de son arrivée étant une variable uniforme sur cette période. a) Trouver la probabilité qu'il doive attendre moins de 5 minutes, b) puis plus de 10 minutes.

SOLUTION. a) Désignons par X le nombre de minutes s'écoulant à partir de 7 h 00 jusqu'à l'arrivée de l'usager. X est une variable uniforme sur $(0, 30)$, d'une part. D'autre part, l'attente n'est inférieure à 5 minutes que si l'usager arrive entre 7 h 10 et 7 h 15 ou entre 7 h 25 et 7 h 30. La probabilité d'attendre moins de 5 minutes est ainsi

$$P\{10 < X < 15\} + P\{25 < X < 30\} = \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = \frac{1}{3}$$

b) De même, il n'attendra plus de 10 minutes que s'il arrive entre 7 h 00 et 7 h 05 ou entre 7 h 15 et 7 h 20, ce qui livre la probabilité de cet événement:

$$P\{0 < X < 5\} + P\{15 < X < 20\} = \frac{1}{3} \quad \blacksquare$$

L'exemple suivant fut étudié pour la première fois par le mathématicien français L.F. Bertrand en 1889. Il est souvent appelé *paradoxe de Bertrand*. Il servira d'introduction à la notion de probabilité géométrique.

Exemple 5.11 Choisissons au hasard une corde dans un cercle. Quelle est la probabilité que la longueur de cette corde dépasse le côté du triangle équilatéral inscrit dans le même cercle?

SOLUTION. Le problème tel qu'énoncé ne peut être résolu car l'expression «choisir une corde au hasard» n'est pas claire. Pour qu'elle le devienne il faut reformuler le problème, ce que nous ferons de deux manières différentes.

Voici la première: ce n'est pas la corde, mais la distance de la corde au centre du cercle de rayon r qui est choisie au hasard. Si cette distance est inférieure à $r/2$, la corde sera d'une longueur supérieure à celle du côté du triangle équilatéral inscrit dans le cercle. Admettons maintenant que D , la distance de la corde au centre, soit une variable uniformément distribuée entre 0 et r . La probabilité cherchée est

$$P\left\{D < \frac{r}{2}\right\} = \frac{r/2}{r} = \frac{1}{2}$$

Dans la seconde formulation du problème, c'est l'angle θ entre la corde et la tangente à l'une de ses extrémités qui est choisi au hasard, θ variant de 0 à 180° (voir figure 5.2). Dans ce cas, la longueur de la corde sera supérieure au côté du triangle équilatéral inscrit si l'angle θ est compris entre 60° et 120° . Si l'on admet que cet angle est une variable uniforme entre 0° et 180° , la réponse correspondant à cette formulation est

$$P\{60 < \theta < 120\} = \frac{120 - 60}{180} = \frac{1}{3}$$

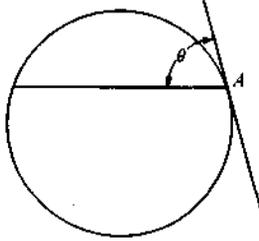


Figure 5.2

On remarquera qu'il est possible de construire des expériences aléatoires pour lesquelles les probabilités résultantes sont alternativement $\frac{1}{2}$ et $\frac{1}{3}$. Pour la première formulation par exemple, l'expérience consisterait à lancer au hasard un disque de rayon r sur une table portant des droites parallèles distantes de $2r$ les unes des autres. Il y aura dans ce cas nécessairement une et une seule droite en intersection avec le disque, ce qui définira une corde. Toutes les distances de cette corde au centre du disque ont la même probabilité d'apparaître, ainsi la probabilité cherchée sera-t-elle $\frac{1}{2}$. D'autre part, l'expérience peut consister à faire tourner au hasard une aiguille à partir d'un point A du cercle (voir figure 5.2). Dans ce cas la réponse est $\frac{1}{3}$. ■

5.4 VARIABLES ALÉATOIRES NORMALES

5.4.1 Définition

Une variable aléatoire X est dite *normale* – ou parfois distribuée normalement – avec paramètres μ et σ^2 si la densité de X est donnée par

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty$$

Le graphe de cette densité est une courbe en forme de cloche avec un axe de symétrie vertical en μ (voir figure 5.3).

La distribution normale fut introduite par le mathématicien français De Moivre en 1733; celui-ci l'utilisa pour approximer les probabilités associées à toute variable aléatoire binomiale, pourvu que le paramètre n de celle-ci soit assez grand. Ce résultat

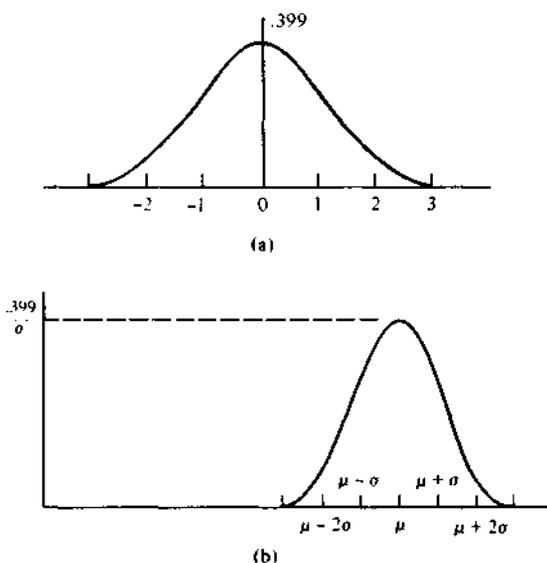


Figure 5.3 Densité de variable normale

(a) avec $\mu = 0$, $\sigma = 1$

(b) μ et σ^2 quelconques

fut ensuite progressivement généralisé par Laplace et d'autres pour devenir le théorème actuellement connu sous le nom de théorème central limite, qui sera discuté au chapitre 8. Ce théorème, l'un des deux résultats les plus importants de la théorie des probabilités¹, sert de base théorique pour expliquer un fait empirique souvent relevé, à savoir qu'en pratique de très nombreux phénomènes aléatoires suivent approximativement une distribution normale. On peut citer à titre d'exemple de variables qui illustrent ce comportement la taille d'un individu choisi au hasard, la vitesse en norme d'une molécule de gaz ou encore l'erreur lors de la mesure d'une quantité physique.

5.4.2 Validation de la définition

Il faut, en fait, prouver que f est bien une densité de probabilité, c'est-à-dire montrer que

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx = 1$$

En effectuant le changement de variable $y = (x - \mu)/\sigma$, on obtient

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy$$

¹ L'autre étant la loi forte des grands nombres.

et il reste donc à montrer que

$$\int_{-\infty}^{\infty} e^{-y^2/2} dy = \sqrt{2\pi}$$

Pour ce faire, posons

$$I = \int_{-\infty}^{\infty} e^{-y^2/2} dy.$$

On aura

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} e^{-y^2/2} dy \int_{-\infty}^{\infty} e^{-x^2/2} dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(y^2+x^2)/2} dy dx \end{aligned}$$

En passant à un système de coordonnées polaires, on peut évaluer cette intégrale double. Il suffit de poser $x = r \cos \theta$, $y = r \sin \theta$, $dy dx = r d\theta dr$. Dès lors,

$$\begin{aligned} I^2 &= \int_0^{\infty} \int_0^{2\pi} e^{-r^2/2} r d\theta dr \\ &= 2\pi \int_0^{\infty} r e^{-r^2/2} dr \\ &= -2\pi e^{-r^2/2} \Big|_0^{\infty} \\ &= 2\pi \end{aligned}$$

Ceci établit bien que I vaut $\sqrt{2\pi}$, et le résultat annoncé est ainsi démontré.

5.4.3 Espérance et variance d'une variable aléatoire normale

Nous allons montrer à présent que les paramètres μ et σ^2 d'une variable aléatoire normale représentent respectivement son espérance et sa variance.

Exemple 5.12 Si X est une variable aléatoire normale de paramètres μ et σ^2 , trouver

- $E[X]$;
- $\text{Var}(X)$.

SOLUTION. a)

$$E[X] = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/2\sigma^2} dx$$

En écrivant x comme $(x - \mu) + \mu$, on a

$$E[X] = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} (x - \mu) e^{-(x-\mu)^2/2\sigma^2} dx + \mu \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx$$

En posant $y = x - \mu$ dans la première intégrale, on obtient

$$E[X] = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} y e^{-y^2/2\sigma^2} dy + \mu \int_{-\infty}^{\infty} f(x) dx$$

où $f(x)$ est la densité normale. Par symétrie, la première intégrale est nulle, d'où

$$E[X] = \mu \int_{-\infty}^{\infty} f(x) dx = \mu$$

b) Puisque $E[X] = \mu$, on a

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] \\ &= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx \end{aligned} \quad (5.5)$$

En substituant $y = (x - \mu)/\sigma$ dans l'équation (5.5), on obtient

$$\begin{aligned} \text{Var}(X) &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \left[-y e^{-y^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2} dy \right] \quad \text{en intégrant par parties} \\ &= \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \\ &= \sigma^2 \end{aligned}$$

5.4.4 Propriété des variables aléatoires normales

Une propriété importante de la famille des variables normales est que si X est normalement distribuée avec paramètres μ et σ^2 , alors $Y = \alpha X + \beta$ est normalement distribuée avec paramètres $\alpha\mu + \beta$ et $\alpha^2\sigma^2$. Ceci résulte du fait que F_Y ¹, la fonction de répartition de la variable Y , est donnée par (lorsque $\alpha > 0$):

$$\begin{aligned} F_Y(a) &= P\{Y \leq a\} \\ &= P\{\alpha X + \beta \leq a\} \end{aligned}$$

¹ La fonction de répartition de Z sera notée F_Z dès qu'il y aura plus d'une variable aléatoire en considération. De même, la densité de Z sera notée f_Z .

$$\begin{aligned}
&= P\left\{X \leq \frac{a - \beta}{\alpha}\right\} \\
&= F_X\left(\frac{a - \beta}{\alpha}\right) \\
&= \int_{-\infty}^{[(a-\beta)/\alpha]} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx \\
&= \int_{-\infty}^a \frac{1}{\sqrt{2\pi}\alpha\sigma} \exp\left\{-\frac{[y - (\alpha\mu + \beta)]^2}{2\alpha^2\sigma^2}\right\} dy \quad (5.6)
\end{aligned}$$

où l'équation (5.6) est obtenue grâce au changement de variables $y = \alpha x + \beta$. Mais comme $F_Z(a) = \int_{-\infty}^a f_Z(y) dy$, il résulte de (5.6) que f_Y , la densité de Y , est donnée par

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\alpha\sigma} \exp\left\{-\frac{[y - (\alpha\mu + \beta)]^2}{2(\alpha\sigma)^2}\right\}$$

La variable aléatoire Y est donc bien distribuée normalement avec paramètres $\alpha\mu + \beta$ et $(\alpha\sigma)^2$.

5.4.5 Variable normale centrée réduite

La conséquence importante du résultat précédent est que si X est une variable normalement distribuée et de paramètres μ et σ^2 , la variable $Z = (X - \mu)/\sigma$ est normalement distribuée de paramètres 0 et 1. Une variable normale ayant ces deux paramètres est dite *standard* ou *centrée réduite*.

L'usage s'est établi de noter la fonction de répartition d'une variable normale centrée réduite par le symbole Φ . En clair,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$$

Les valeurs $\Phi(x)$ pour des arguments x non négatifs sont données dans le tableau 5.4. Pour les arguments x négatifs, on calcule $\Phi(x)$ grâce à l'équation

$$\Phi(-x) = 1 - \Phi(x) \quad -\infty < x < \infty \quad (5.7)$$

La démonstration de (5.7) est laissée en exercice. Elle résulte de la symétrie de la densité normale standard. Grâce à (5.7) on peut encore écrire, pour une variable normale Z centrée réduite:

$$P\{Z \leq -x\} = P\{Z > x\} \quad -\infty < x < \infty$$

Si l'on considère maintenant une variable normale X de paramètres μ et σ^2 quelconques, la variable $Z = (X - \mu)/\sigma$ sera normale centrée réduite. Par conséquent, on peut exprimer la fonction de répartition de X de la manière suivante:

$$\begin{aligned} F_X(a) &= P\{X \leq a\} \\ &= P\left\{\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right\} \\ &= \Phi\left(\frac{a - \mu}{\sigma}\right) \end{aligned}$$

5.4.6 Exemples d'application de variables normales

Exemple 5.13 Soit X une variable aléatoire de paramètres $\mu = 3$ et $\sigma^2 = 9$. Calculer

- $P\{2 < X < 5\}$,
- $P\{X > 0\}$,
- $P\{|X - 3| > 6\}$.

SOLUTION.

$$\begin{aligned} \text{a) } P\{2 < X < 5\} &= P\left\{\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right\} = P\left\{-\frac{1}{3} < Z < \frac{2}{3}\right\} \\ &= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right) \\ &= \Phi\left(\frac{2}{3}\right) - \left[1 - \Phi\left(\frac{1}{3}\right)\right] \approx .3779 \end{aligned}$$

$$\begin{aligned} \text{b) } P\{X > 0\} &= P\left\{\frac{X-3}{3} > \frac{0-3}{3}\right\} = P\{Z > -1\} \\ &= 1 - \Phi(-1) \\ &= \Phi(1) \\ &\approx .8413 \end{aligned}$$

$$\begin{aligned} \text{c) } P\{|X - 3| > 6\} &= P\{X > 9\} + P\{X < -3\} \\ &= P\left\{\frac{X-3}{3} > \frac{9-3}{3}\right\} + P\left\{\frac{X-3}{3} < \frac{-3-3}{3}\right\} \\ &= P\{Z > 2\} + P\{Z < -2\} \\ &= 1 - \Phi(2) + \Phi(-2) \\ &= 2[1 - \Phi(2)] \\ &\approx .0456 \end{aligned}$$

■

Exemple 5.14 Il est courant d'admettre qu'un examen est bien construit (dans le sens où il permet de construire une fourchette serrée et fiable pour la note d'un candidat) si la répartition des scores obtenus par les participants se rapproche de la densité d'une variable normale. En d'autres mots, cette répartition devrait affecter la forme en cloche des densités normales. L'enseignant utilise alors les scores pour évaluer les paramètres μ et σ^2 puis assigne souvent des notes selon le principe que voici: ceux dont le score est supérieur à $\mu + \sigma$ reçoivent la note A; ceux dont le score est compris entre μ et $\mu + \sigma$ reçoivent B; ceux dont le score est entre $\mu - \sigma$ et μ reçoivent C, tandis que ceux qui tombent entre $\mu - 2\sigma$ et $\mu - \sigma$ reçoivent D. En dessous de $\mu - 2\sigma$ la note est F. Il s'agit d'une espèce d'évaluation «à échelle mobile» basée sur des divisions fixes de la courbe de répartition. On aura:

$$P\{X > \mu + \sigma\} = P\left\{\frac{X - \mu}{\sigma} > 1\right\} = 1 - \Phi(1) \approx .1587$$

$$P\{\mu < X < \mu + \sigma\} = P\left\{0 < \frac{X - \mu}{\sigma} < 1\right\} = \Phi(1) - \Phi(0) \approx .3413$$

$$\begin{aligned} P\{\mu - \sigma < X < \mu\} &= P\left\{-1 < \frac{X - \mu}{\sigma} < 0\right\} \\ &= \Phi(0) - \Phi(-1) \approx .3413 \end{aligned}$$

$$\begin{aligned} P\{\mu - 2\sigma < X < \mu - \sigma\} &= P\left\{-2 < \frac{X - \mu}{\sigma} < -1\right\} \\ &= \Phi(-1) - \Phi(-2) \approx .1359 \end{aligned}$$

$$P\{X < \mu - 2\sigma\} = P\left\{\frac{X - \mu}{\sigma} < -2\right\} = \Phi(-2) = .0228$$

Il en résulte que 16% des candidats recevront la note A, 34% recevront B, autant auront C, 14% recevant D et 2% F. ■

Exemple 5.15 Lors d'un procès en attribution de paternité, un expert témoigne que la durée de la grossesse, en jours, c'est-à-dire le laps de temps entre la conception et la naissance de l'enfant, est de distribution approximativement normale avec paramètres $\mu = 270$ et $\sigma^2 = 100$. L'un des pères putatifs est en mesure de prouver son absence du pays pendant une période s'étendant entre le 290-ième et le 240-ième jour précédant l'accouchement. Quelle est la probabilité que la conception de l'enfant ait eu lieu plus de 290 jours avant sa naissance ou moins de 240 jours avant?

SOLUTION. Soit X la durée de la grossesse et admettons que le père putatif soit bien le géniteur. La probabilité cherchée est alors

$$\begin{aligned} P\{X > 290 \text{ ou } X < 240\} &= P\{X > 290\} + P\{X < 240\} \\ &= P\left\{\frac{X - 270}{10} > 2\right\} + P\left\{\frac{X - 270}{10} < -3\right\} \\ &= 1 - \Phi(2) + 1 - \Phi(3) \\ &\approx .0241 \end{aligned}$$

■

Exemple 5.16 On désire envoyer un signal binaire - c'est-à-dire valant 0 ou 1 - par câble électrique d'un point A à un point B . Cependant, la transmission est affectée par des perturbations, dites bruit. Aussi émet-on un signal d'intensité 2 lorsqu'on veut communiquer 1 et d'intensité -2 lorsqu'on veut indiquer 0. Si on désigne par x (où $x = \pm 2$) la valeur émise en A et par R la valeur enregistrée en B , on aura $R = x + N$ où N représente l'erreur due au bruit du canal de transmission. Le décodage du signal en B obéit à la règle suivante:

$R \geq 0,5$ est interprété comme signifiant 1

$R < 0,5$ est interprété comme signifiant 0.

Le bruit N du canal est souvent de distribution normale. On supposera ici que sa répartition est normale centrée réduite.

Deux types d'erreurs peuvent survenir: un signal 1 peut être faussement compris comme un 0, ou l'inverse. Le premier type d'erreur sera observé si le signal est 1 et $2 + N < 0,5$. Le second le sera lorsque le signal est 0 et $-2 + N \geq 0,5$. Ainsi,

$$\begin{aligned} P\{\text{erreur} \mid \text{le message est 1}\} &= P\{N < -1,5\} \\ &= 1 - \Phi(1,5) \approx .0668 \end{aligned}$$

et

$$\begin{aligned} P\{\text{erreur} \mid \text{le message est 0}\} &= P\{N \geq 2,5\} \\ &= 1 - \Phi(2,5) \approx .0062 \end{aligned} \quad \blacksquare$$

5A.7 Approximation de Φ

L'inégalité suivante sur $\Phi(x)$ est de portée théorique:

$$\frac{1}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^3} \right) e^{-x^2/2} < 1 - \Phi(x) < \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2} \quad \text{pour tout } x > 0 \quad (5.8)$$

Pour démontrer (5.8), on partira de l'inégalité évidente:

$$(1 - 3y^{-4})e^{-y^2/2} < e^{-y^2/2} < (1 + y^{-2})e^{-y^2/2}$$

ce qui implique que

$$\int_x^\infty (1 - 3y^{-4})e^{-y^2/2} dy < \int_x^\infty e^{-y^2/2} dy < \int_x^\infty (1 + y^{-2})e^{-y^2/2} dy \quad (5.9)$$

Or

$$\frac{d}{dy} [(y^{-1} - y^{-3})e^{-y^2/2}] = -(1 - 3y^{-4})e^{-y^2/2}$$

$$\frac{d}{dy} [y^{-1}e^{-y^2/2}] = -(1 + y^{-2})e^{-y^2/2}$$

et ainsi, pour $x > 0$,

$$-(y^{-1} - y^{-3})e^{-y^2/2} \Big|_x^\infty < \int_x^\infty e^{-y^2/2} dy < -y^{-1}e^{-y^2/2} \Big|_x^\infty$$

ou

$$(x^{-1} - x^{-3})e^{-x^2/2} < \int_x^\infty e^{-y^2/2} dy < x^{-1}e^{-x^2/2}$$

ce qui établit (5.8).

Il résulte en plus de (5.8) la relation

$$1 - \Phi(x) \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$$

pour de grandes valeurs de x [la notation $a(x) \sim b(x)$ lorsque x est grand signifie que $\lim_{x \rightarrow \infty} a(x)/b(x) = 1$].

5.4.8 Approximation normale d'une répartition binomiale

Le théorème présenté ci-dessous est connu sous le nom de théorème limite de De Moivre-Laplace. De Moivre fut le premier à l'établir dans le cas particulier $p = \frac{1}{2}$ en 1733, tandis que Laplace a pu le généraliser pour toute valeur de p en 1812. Ce théorème énonce que si «on standardise» une variable aléatoire binomiale de paramètres n et p en soustrayant d'abord sa moyenne np puis en divisant le résultat par son écart-type $\sqrt{np(1-p)}$, alors la variable aléatoire standardisée (de moyenne 0 et de variance 1) suivra approximativement, lorsque n est grand, une distribution normale standard.

Théorème limite de De Moivre-Laplace

Soit S_n le nombre de succès lors de la réalisation de n épreuves indépendantes, la probabilité de réussite pour chaque épreuve étant p . Alors, pour tout $a < b$ on peut écrire

$$P\left\{a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right\} \rightarrow \Phi(b) - \Phi(a)$$

lorsque $n \rightarrow \infty$.

Ce théorème ne constituant qu'un cas particulier du théorème central limite, lequel sera présenté au chapitre 8, nous n'en donnerons pas de démonstration.

On remarquera qu'à ce stade deux approximations de la répartition binomiale ont été proposées: l'approximation de Poisson, satisfaisante lorsque n est grand et lorsque np n'est pas extrême; l'approximation normale, de laquelle on peut montrer qu'elle est de bonne qualité lorsque $np(1-p)$ est grand (voir figure 5.5). [En règle générale, cette approximation est tout à fait satisfaisante dès que $np(1-p)$ dépasse 10].

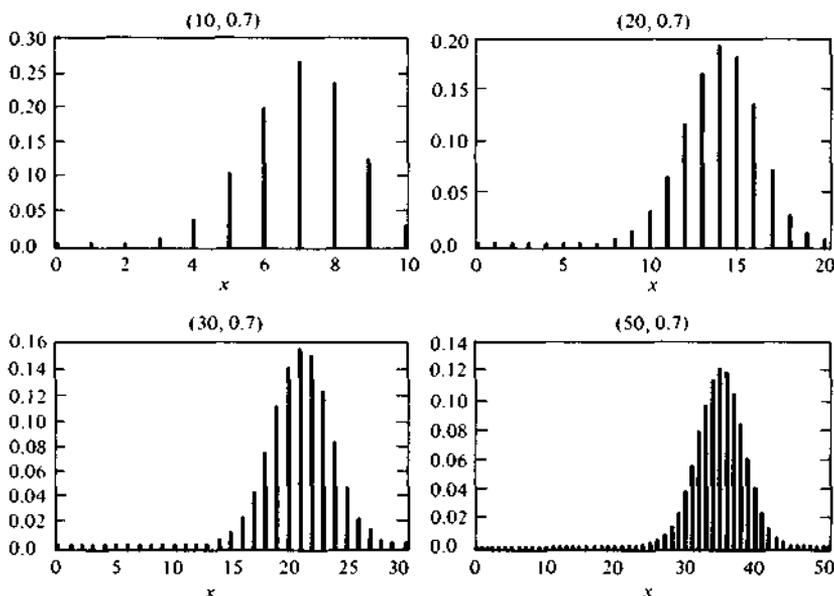


Figure 5.5 La loi de probabilité d'une variable aléatoire binomiale (n, p) devient de plus en plus «normale» à mesure que n augmente.

5.4.9 Exemples d'approximation normale de répartition binomiales

Exemple 5.17 Soit X la variable aléatoire comptant le nombre d'occurrences de pile lors d'une série de 40 jets. On veut calculer $P\{X = 20\}$ par approximation normale puis comparer le résultat à la valeur exacte.

SOLUTION. Comme X est une variable discrète tandis qu'une variable normale est continue, la meilleure approximation de la probabilité cherchée sera

$$\begin{aligned}
 P\{X = 20\} &= P\{19.5 < X < 20.5\} \\
 &= P\left\{\frac{19.5 - 20}{\sqrt{10}} < \frac{X - 20}{\sqrt{10}} < \frac{20.5 - 20}{\sqrt{10}}\right\} \\
 &= P\left\{-.16 < \frac{X - 20}{\sqrt{10}} < .16\right\} \\
 &\approx \Phi(.16) - \Phi(-.16) \approx .1272
 \end{aligned}$$

Le résultat exact est

$$P\{X = 20\} = \binom{40}{20} \left(\frac{1}{2}\right)^{40}$$

qui peut être évalué à 0,1254. ■

Exemple 5.18 La taille idéale pour une classe de première année dans un collège donné est de 150 étudiants. La politique de ce collège est d'admettre 450 étudiants et est basée sur la constatation expérimentale que 30% seulement des étudiants admis suivront vraiment le cours. Quelle est la probabilité que le collège se retrouve avec une première classe de plus de 150 étudiants lors d'une année donnée ?

SOLUTION. On désigne par X le nombre d'étudiants qui suivent effectivement le cours. Cette variable X est donc binomiale de paramètres $n = 450$ et $p = 0,3$. L'approximation normale livre

$$P\{X \geq 150,5\} = P\left\{\frac{X - (450)(.3)}{\sqrt{450(.3)(.7)}} \geq \frac{150,5 - (450)(.3)}{\sqrt{450(.3)(.7)}}\right\} \approx 1 - \Phi(1,59) \\ = .0559$$

Ainsi, dans moins de 6% des cas seulement la première année aura un effectif supérieur à l'optimum. (On remarque que ce calcul est basé sur une hypothèse d'indépendance. Laquelle?). ■

Exemple 5.19 On impose à 100 personnes un régime alimentaire thérapeutique pour évaluer l'effet de ce régime sur la concentration en cholestérol du sang. Leur taux de cholestérol est mesuré après une période suffisante d'application du régime. Le spécialiste en nutrition qui réalise l'expérience a décidé de recommander ce régime si 65% au moins des sujets montrent une baisse du taux de cholestérol. Quelle est la probabilité qu'il prenne une décision erronée et recommande le régime alors que celui-ci est sans effet?

SOLUTION. Admettons que dans le cas où le régime est sans effet, une personne donnée verra son taux de cholestérol baisser lors du régime sous le seul effet du hasard et avec une probabilité de $\frac{1}{2}$. Désignons par X le nombre de personnes dont le taux s'est abaissé. La probabilité de recommander un régime qui n'a en réalité pas d'effet du tout est alors:

$$\sum_{i=65}^{100} \binom{100}{i} \left(\frac{1}{2}\right)^{100} = P\{X \geq 64,5\} \\ = P\left\{\frac{X - (100)(\frac{1}{2})}{\sqrt{100(\frac{1}{2})(\frac{1}{2})}} \geq 2,9\right\} \\ \approx 1 - \Phi(2,9) \\ \approx .0019 \quad \blacksquare$$

NOTES HISTORIQUES AU SUJET DE LA DISTRIBUTION NORMALE

La distribution normale a été introduite en 1733 par le mathématicien français Abraham De Moivre, qui a utilisé cette distribution pour approximer les probabilités du lancer d'une pièce et qui l'a appelée la courbe exponentielle en forme de cloche. Son utilité n'est cependant apparue claire qu'en 1809, lorsque le fameux mathématicien allemand K. F. Gauss l'a utilisée comme partie intégrante de son approche pour prédire l'emplacement des entités atomiques. Il devint alors commun d'appeler cette distribution la distribution Gaussienne.

Durant la deuxième partie du dix-neuvième siècle, de nombreux mathématiciens crurent que la majorité des ensembles de données avait des histogrammes conformes à la distribution Gaussienne en forme de cloche. En effet, on accepta qu'il était «normal» que n'importe quel ensemble de données au bon comportement suive cette courbe. Ainsi, suivant le statisticien britannique Karl Pearson, les gens commencèrent à appeler la distribution Gaussienne la distribution normale. (Une explication partielle du fait que tant d'ensembles de données sont conformes à la courbe normale est donnée par le théorème central limite, présenté dans le chapitre 8.)

Abraham De Moivre (1667-1754)

Aujourd'hui, on ne manque pas de consultants en statistiques et la plupart d'entre eux exercent leur métier dans des cadres très élégants. Cependant, le premier de leur profession travaillait à Long Acres, à Londres, durant les premières années du dix-huitième siècle, dans une maison de jeux sombre et sale connue sous le nom de Slaughter's Coffee House. Il s'appelait Abraham De Moivre, était réfugié protestant de la France catholique et, contre salaire, il calculait les probabilités pour parier à tous les types de jeux de chance.

Bien que De Moivre, qui a découvert la courbe normale, vive dans ce café, il était un mathématicien compétent reconnu. En effet, il était membre de la Royal Society et un intime d'Isaac Newton. Karl Pearson a imaginé De Moivre au travail au Slaughter's Coffee House: «J'imagine De Moivre travaillant sur une table sale dans le café avec un joueur paumé derrière lui et Isaac Newton marchant dans la foule pour aller chercher son ami. Cela ferait un beau tableau pour un artiste inspiré.»

Karl Friedrich Gauss (1777-1855)

Karl Friedrich Gauss, un des premiers utilisateurs de la courbe normale, est l'un des plus grands mathématiciens de tous les temps. Voyons ce que l'historien des mathématiques bien connu E. T. Bell écrit dans son livre *Men of Mathematics* en 1954. Dans un chapitre intitulé «Le Prince des Mathématiciens», il écrit: «Archimedes, Newton et Gauss; ces trois-là forment une classe parmi les grands mathématiciens et ce n'est pas au commun des mortels d'essayer de les classer selon leurs mérites. Tous les trois ont fait avancer de manière considérable les recherches en mathématiques pures et appliquées. Archimedes estimait plus ses mathématiques pures que leurs applications; Newton semble avoir trouvé la principale justification à ses inventions mathématiques dans l'utilisation scientifique qu'il en faisait; Gauss déclarait que ses recherches pures et appliquées ne faisaient qu'un.»

5.5 VARIABLES ALÉATOIRES EXPONENTIELLES

5.5.1 Définition

Une variable aléatoire dont la densité est donnée par l'équation suivante, où λ est positif,

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

est dite variable aléatoire *exponentielle* (ou plus simplement est dite de distribution exponentielle) de paramètre λ . La fonction de répartition F d'une variable exponentielle est donnée par

$$\begin{aligned} F(a) &= P\{X \leq a\} \\ &= \int_0^a \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x} \Big|_0^a \\ &= 1 - e^{-\lambda a} \quad a \geq 0 \end{aligned}$$

On remarquera que $F(\infty) = \int_0^{\infty} \lambda e^{-\lambda x} dx = 1$ comme il se doit. Nous allons montrer à présent que le paramètre λ est égal à l'inverse de l'espérance de la variable.

Exemple 5.20 Soit X une variable aléatoire exponentielle de paramètre λ . Calculer

- a) $E[X]$;
b) $\text{Var}(X)$.

SOLUTION. a) Comme la densité de X est donnée par

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

on obtient

$$E[X] = \int_0^{\infty} x \lambda e^{-\lambda x} dx$$

Une intégration par parties ($\lambda e^{-\lambda x} dx = dv$, $u = x$) donne

$$\begin{aligned} E[X] &= -xe^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx \\ &= 0 - \frac{e^{-\lambda x}}{\lambda} \Big|_0^{\infty} \\ &= \frac{1}{\lambda} \end{aligned}$$

b) Pour obtenir la variance de X , calculons d'abord $E[X^2]$.

$$E[X^2] = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx$$

Une intégration par parties ($\lambda e^{-\lambda x} dx = dv$, $u = x^2$) donne

$$\begin{aligned} E[X^2] &= -x^2 e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} 2xe^{-\lambda x} dx \\ &= 0 + \frac{2}{\lambda} E[X] \\ &= \frac{2}{\lambda^2} \end{aligned}$$

Donc

$$\begin{aligned} \text{Var}(X) &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \\ &= \frac{1}{\lambda^2} \end{aligned}$$

Donc la moyenne d'une exponentielle est l'inverse de son paramètre λ et la variance est le carré de la moyenne. ■

Dans la pratique, on rencontre souvent la distribution exponentielle lorsqu'il s'agit de représenter le temps d'attente avant l'arrivée d'un événement spécifié. Par exemple, le temps qui nous sépare du prochain tremblement de terre, de la prochaine guerre ou du prochain appel téléphonique mal aiguillé sont toutes des variables aléatoires dont les distributions tendent à toutes fins utiles à se rapprocher de distributions exponentielles. (Le lecteur trouvera une explication de nature théorique à ces faits à la section 4.8, en particulier dans l'exemple 4.26).

Exemple 5.21 On suppose que la durée d'une conversation téléphonique, mesurée en minutes, est une variable aléatoire exponentielle de paramètre $\lambda = \frac{1}{10}$. Vous arrivez à une cabine téléphonique et quelqu'un passe juste devant vous. Avec quelle probabilité devrez-vous attendre

- plus de 10 minutes;
- entre 10 et 20 minutes?

SOLUTION. X désignera la durée de la conversation de la personne qui vous a devancé. Les probabilités cherchées seront respectivement:

$$\text{a) } P\{X > 10\} = \int_{10}^{\infty} \frac{1}{10} e^{-x/10} dx = -e^{-x/10} \Big|_{10}^{\infty} = e^{-1} \approx .368$$

$$\begin{aligned} \text{b)} \quad P\{10 < X < 20\} &= \int_{10}^{20} \frac{1}{10} e^{-x/10} dx = -e^{-x/10} \Big|_{10}^{20} \\ &= e^{-1} - e^{-2} \approx .233 \end{aligned} \quad \blacksquare$$

5.5.2 Propriété d'absence de mémoire

On dira qu'une variable aléatoire non négative X est *sans mémoire* lorsque

$$P\{X > s + t | X > t\} = P\{X > s\} \text{ pour tous } s, t \geq 0 \quad (5.10)$$

Représentons-nous X comme la durée de vie d'un certain instrument. Dans ce cas (5.10) signifie que la probabilité pour l'instrument de durer au moins $s + t$ heures sachant qu'il en a déjà vécu t est la même que la probabilité non conditionnelle qu'il dure s heures à partir de la mise en fonction initiale. En d'autres termes, si l'instrument fonctionne encore après t heures de service, la distribution de sa durée de vie à partir de là est la même que la distribution de la durée de vie de l'appareil neuf. On peut dire que l'appareil fonctionne sans mémoire du temps d'usage déjà écoulé.

Or, (5.10) est équivalent à

$$\frac{P\{X > s + t, X > t\}}{P\{X > t\}} = P\{X > s\}$$

ou

$$P\{X > s + t\} = P\{X > s\}P\{X > t\} \quad (5.11)$$

Comme (5.11) est vérifiée par toute variable exponentielle X (puisque $e^{-\lambda(s+t)} = e^{-\lambda s} e^{-\lambda t}$) la classe des variables exponentielles est sans mémoire.

5.5.3 Exemples d'application de variables exponentielles

Exemple 5.22 Dans un bureau de poste, le service est assuré par deux employés. Lorsque Smith y entre, l'un des employés sert Jones tandis que l'autre répond à Brown. On admettra que Smith sera à son tour servi dès le départ de Jones ou de Brown. Le temps passé par un employé de poste pour chaque client est distribué exponentiellement avec paramètre λ . Quelle est la probabilité que Smith soit le dernier des trois clients à sortir de ce bureau de poste?

SOLUTION. On peut adopter l'approche suivante: au moment où Smith trouve un postier libre, l'un des deux autres clients vient de partir tandis que l'autre est encore au guichet. Mais ce dernier, en vertu de l'absence de mémoire des variables exponentielles, va encore rester pendant un temps qui est toujours exponentiellement distribué avec paramètre λ . Tout se passe comme si ce client venait d'arriver au guichet. Par conséquent Smith a une chance sur deux de sortir le dernier, du fait de la symétrie des deux situations. \blacksquare

Exemple 5.23 Le nombre de miles couvert par une batterie de voiture avant défaillance est distribuée exponentiellement et sa valeur moyenne est de 10 000 miles. Une personne souhaite se lancer dans un voyage de 5 000 miles. Avec quelle probabilité terminera-t-elle son voyage sans avarie de batterie? Que devient cette probabilité si la distribution n'est pas exponentielle?

SOLUTION. Du fait que les variables exponentielles sont sans mémoire, la durée de vie résiduelle de la batterie sera distribuée exponentiellement avec paramètre $\lambda = \frac{1}{10}$ en prenant comme unité le millier de miles. La probabilité voulue est donc

$$P\{\text{durée de vie résiduelle} > 5\} = 1 - F(5) = e^{-5\lambda} = e^{-1/2} \approx .607$$

Si, par contre, la distribution de cette durée résiduelle n'est pas exponentielle, ce que l'on cherche peut être écrit

$$P\{\text{durée résiduelle} > t + 5 \mid \text{durée résiduelle} > t\} = \frac{1 - F(t + 5)}{1 - F(t)}$$

où t est la durée de service de la batterie jusqu'au moment où le voyage commence. Et l'on voit que dans ce cas une information supplémentaire (à savoir t) est nécessaire au calcul de la réponse. ■

5.5.4 Unicité de la classe des variables aléatoires sans mémoire

Non seulement les variables exponentielles n'ont pas de mémoire, elles sont de plus la seule classe de variables à jouir de cette propriété. Pour la vérifier, supposons qu'une variable X soit sans mémoire et posons $\bar{F}(x) = P\{X > x\}$. De (5.11), on tire

$$\bar{F}(s + t) = \bar{F}(s)\bar{F}(t)$$

Cela veut dire que \bar{F} satisfait l'équation fonctionnelle

$$g(s + t) = g(s)g(t)$$

Cependant, la seule solution continue à droite et non triviale de cette équation fonctionnelle est¹

$$g(x) = e^{-\lambda x} \quad (5.12)$$

¹ On obtient (5.12) de la manière suivante: si $g(s + t) = g(s) \cdot g(t)$, alors

$$g\left(\frac{2}{n}\right) = g\left(\frac{1}{n} + \frac{1}{n}\right) = g^2\left(\frac{1}{n}\right)$$

et par itération on obtient $g(m/n) = g^m(1/n)$. D'autre part,

$$g(1) = g\left(\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}\right) = g^n\left(\frac{1}{n}\right) \quad \text{ou} \quad g\left(\frac{1}{n}\right) = (g(1))^{1/n}$$

et par conséquent $g(m/n) = (g(1))^{m/n}$. Comme g est continue à droite, on peut écrire $g(x) = (g(1))^x$. On peut encore poser $\lambda = -\ln(g(1))$ puisque $g(1) = (g(1/2))^2 \geq 0$, ce qui laisse $g(x) = e^{-\lambda x}$.

La continuité à droite des solutions peut être supposée puisque toutes les fonctions de répartition le sont. Il reste que

$$\bar{F}(x) = e^{-\lambda x} \quad \text{ou} \quad F(x) = P\{X \leq x\} = 1 - e^{-\lambda x}$$

ce qui établit que X est nécessairement exponentielle.

5.5.5 Distribution de Laplace

La *distribution de Laplace* est une variante de la distribution exponentielle. Une telle distribution est symétrique par rapport à l'origine et la valeur absolue d'une variable laplacienne est exponentiellement distribuée². Si le paramètre de cette exponentielle est $\lambda > 0$, λ sera également le paramètre de la variable laplacienne. La densité d'une telle variable est

$$f(x) = \frac{1}{2}\lambda e^{-\lambda|x|} \quad -\infty < x < \infty$$

Sa fonction de répartition est donnée par

$$F(x) = \begin{cases} \frac{1}{2} \int_{-\infty}^x \lambda e^{\lambda x} dx & x < 0 \\ \frac{1}{2} \int_{-\infty}^0 \lambda e^{\lambda x} dx + \frac{1}{2} \int_0^x \lambda e^{-\lambda x} dx & x > 0 \end{cases}$$

$$= \begin{cases} \frac{1}{2} e^{\lambda x} & x < 0 \\ 1 - \frac{1}{2} e^{-\lambda x} & x > 0 \end{cases}$$

Exemple 5.24 Revenons à l'exemple 5.16 qui traitait de l'émission d'un signal binaire de A à B ; lorsque le signal valait 1 on émettait avec une intensité de $+2$, tandis que s'il valait 0 on émettait avec une intensité de -2 . On suppose cependant maintenant que le bruit du canal de transmission est une variable aléatoire laplacienne de paramètre $\lambda = 1$. Comme auparavant, R désigne le signal reçu en B et la convention de décodage reste

$R \geq 0,5$: on admet que 1 a été émis

$R < 0,5$: on admet que 0 a été émis.

² Pour cette raison, une variable laplacienne est parfois dite exponentielle double.

Dans le cas présent, où le bruit est laplacien de paramètre $\lambda = 1$, les probabilités des deux types d'erreurs sont respectivement

$$\begin{aligned} P\{\text{erreur}|\text{message à transmettre} = 1\} &= P\{N < -1.5\} \\ &= \frac{1}{2}e^{-1.5} \\ &= .1116 \end{aligned}$$

$$\begin{aligned} P\{\text{erreur}|\text{message à transmettre} = 0\} &= P\{N \geq 2.5\} \\ &= \frac{1}{2}e^{-2.5} \\ &= .041 \end{aligned}$$

En comparant ces résultats avec ceux de l'exemple 5.16, on constate que les probabilités d'erreur sont plus élevées si le bruit est laplacien de paramètre $\lambda = 1$, plutôt que de distribution normale centrée réduite. ■

5.5.6 Fonctions taux de panne

Considérons une variable aléatoire continue à valeurs positives qui puisse représenter la durée de vie d'un certain composant, de fonction de répartition F et de densité f . La fonction *taux de panne* $\lambda(t)$ de F est définie par l'équation

$$\lambda(t) = \frac{f(t)}{\bar{F}(t)}, \quad \bar{F} = 1 - F$$

Pour obtenir une interprétation de $\lambda(t)$, supposons que notre composant ait déjà t heures de service et que l'on veuille calculer la probabilité de sa défaillance dans l'espace de temps dt qui suit. En d'autres termes, on cherche à calculer $P\{X \in (t, t + dt) | X > t\}$. Or,

$$\begin{aligned} P\{X \in (t, t + dt) | X > t\} &= \frac{P\{X \in (t, t + dt), X > t\}}{P\{X > t\}} \\ &= \frac{P\{X \in (t, t + dt)\}}{P\{X > t\}} \\ &\approx \frac{f(t)}{\bar{F}(t)} dt \end{aligned}$$

ce qui peut s'interpréter comme suit: $\lambda(t)$ représente un taux de panne conditionnel instantané, la condition étant que le composant ait pu assurer déjà t heures de service.

Dans le cas d'une durée de vie exponentielle, l'absence de mémoire de cette distribution signifie que la durée de vie résiduelle d'un composant conditionnellement à une durée de service t jusque-là est de même distribution que la durée de vie initiale. Aussi λ devrait-elle être une fonction constante. Cela est confirmé par le calcul suivant:

$$\begin{aligned}\lambda(t) &= \frac{f(t)}{\bar{F}(t)} \\ &= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} \\ &= \lambda\end{aligned}$$

Ainsi, le taux de panne d'une variable exponentielle est-il constant. C'est la raison pour laquelle le paramètre λ est souvent appelé le *taux* d'une telle distribution.

On peut par ailleurs établir que la fonction taux de panne λ détermine de manière univoque la fonction de répartition F des variables aléatoires obéissant à ce taux. En effet, par définition

$$\lambda(t) = \frac{\frac{d}{dt} F(t)}{1 - F(t)}$$

L'intégration des deux membres donne

$$\ln(1 - F(t)) = - \int_0^t \lambda(t) dt + k$$

ou

$$1 - F(t) = e^k \exp \left\{ - \int_0^t \lambda(t) dt \right\}$$

Or $k = 0$, ce que l'on voit en posant $t = 0$. Donc,

$$F(t) = 1 - \exp \left\{ - \int_0^t \lambda(t) dt \right\} \quad (5.13)$$

Par conséquent, la fonction de répartition d'une variable aléatoire continue à valeurs positives est entièrement déterminée par la donnée d'une fonction taux de panne. A titre d'exemple, si ce taux de panne est une fonction linéaire, donc du type

$$\lambda(t) = a + bt$$

sa fonction de répartition sera

$$F(t) = 1 - e^{-at - bt^2/2}$$

et sa densité, par dérivation

$$f(t) = (a + bt)e^{-(at + bt^2/2)}, \quad t \geq 0$$

Pour $a = 0$ la densité ci-dessus porte le nom de *densité de Rayleigh*.

Exemple 5.25 On entend souvent dire que le taux de mortalité chez les fumeurs, à tout âge, est le double de celui des non-fumeurs. Qu'est-ce que cela veut dire? Cela signifie-t-il qu'un non-fumeur a une probabilité deux fois plus grande de survivre à un nombre d'années donné qu'un fumeur du même âge?

SOLUTION. Soit $\lambda_f(t)$ le taux de mortalité pour un fumeur âgé de t années et $\lambda_n(t)$ celui d'un non-fumeur du même âge. Alors l'affirmation ci-dessus équivaut à dire que

$$\lambda_f(t) = 2\lambda_n(t).$$

La probabilité qu'un non-fumeur âgé de A années survive à l'âge B , $A < B$ est $P\{\text{un non-fumeur âgé de } A \text{ atteint l'âge } B\}$

$= P\{\text{la durée de vie des non-fumeurs} > B \mid \text{la durée de vie des non-fumeurs} > A\}$

$$\begin{aligned} &= \frac{1 - F_{\text{non}}(B)}{1 - F_{\text{non}}(A)} \\ &= \frac{\exp\left\{-\int_0^B \lambda_n(t) dt\right\}}{\exp\left\{-\int_0^A \lambda_n(t) dt\right\}} && \text{grâce à (5.13)} \\ &= \exp\left\{-\int_A^B \lambda_n(t) dt\right\} \end{aligned}$$

alors que la probabilité correspondante pour un fumeur est, par le même raisonnement,

$$\begin{aligned} P\{\text{un fumeur âgé de } A \text{ atteint l'âge } B\} &= \exp\left\{-\int_A^B \lambda_f(t) dt\right\} \\ &= \exp\left\{-2 \int_A^B \lambda_n(t) dt\right\} \\ &= \left[\exp\left\{-\int_A^B \lambda_n(t) dt\right\}\right]^2 \end{aligned}$$

En d'autres termes, si l'on a deux individus de même âge, dont l'un est fumeur et l'autre pas, la probabilité que le fumeur survive à un âge donné est le carré (non la moitié) de celle du non-fumeur. Par exemple, si $\lambda_n(t) = \frac{1}{30}$, $50 \leq t \leq 60$, alors la probabilité qu'un non-fumeur âgé de 50 ans atteigne l'âge de 60 ans est $e^{-1/3} \approx 0,7165$ alors que pour un fumeur elle vaut $e^{-2/3} \approx 0,5134$. ■

5.6 AUTRES DISTRIBUTIONS CONTINUES

5.6.1 Distribution gamma

On dira qu'une variable aléatoire suit une *loi gamma* de paramètres (t, λ) , $\lambda > 0$ et $t > 0$ si sa densité est

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{t-1}}{\Gamma(t)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

où $\Gamma(t)$, appelée fonction gamma, est définie par

$$\Gamma(t) = \int_0^{\infty} e^{-y} y^{t-1} dy$$

Une intégration par parties de l'intégrale donnant $\Gamma(t)$ livre

$$\begin{aligned} \Gamma(t) &= -e^{-y} y^{t-1} \Big|_0^{\infty} + \int_0^{\infty} e^{-y} (t-1) y^{t-2} dy \\ &= (t-1) \int_0^{\infty} e^{-y} y^{t-2} dy \\ &= (t-1) \Gamma(t-1) \end{aligned} \tag{5.14}$$

Pour des valeurs n entières de t , l'utilisation itérée de (5.14) donne

$$\begin{aligned} \Gamma(n) &= (n-1) \Gamma(n-1) \\ &= (n-1)(n-2) \Gamma(n-2) \\ &= \dots \\ &= (n-1)(n-2) \dots 3 \cdot 2 \Gamma(1) \end{aligned}$$

Comme $\Gamma(1) = \int_0^{\infty} e^{-x} dx = 1$, on aura pour tout entier n

$$\Gamma(n) = (n-1)!$$

Lorsque t prend des valeurs entières positives, disons $t = n$, la loi gamma de paramètres (t, λ) représente fréquemment la distribution du temps d'attente avant la n -ième occurrence d'un certain type d'événements. Plus précisément, supposons que des événements se répètent au hasard dans le temps et satisfassent aux trois conditions exposées dans la section 4.8. Dans ce cas, le temps d'attente avant la n -ième occurrence suivra une répartition gamma de paramètres (n, λ) . Pour s'en assurer, notons par T_n l'heure à laquelle le n -ième événement se produit. On remarquera que $T_n \leq t$ si et seulement si le nombre d'occurrences enregistrées au temps t est de n au moins. En posant que ce nombre d'événements dans l'intervalle $[0, t]$ est $N(t)$, on aura

$$\begin{aligned} P\{T_n \leq t\} &= P\{N(t) \geq n\} \\ &= \sum_{j=n}^{\infty} P\{N(t) = j\} \\ &= \sum_{j=n}^{\infty} \frac{e^{-\lambda t} (\lambda t)^j}{j!} \end{aligned}$$

La dernière de ces équations résulte du fait que le nombre d'événements dans $[0, t]$ suit une loi de Poisson de paramètre λt , et sa dérivation donne la densité f de T_n :

$$\begin{aligned} f(t) &= \sum_{j=n}^{\infty} \frac{e^{-\lambda t} j(\lambda t)^{j-1} \lambda}{j!} - \sum_{j=n}^{\infty} \frac{\lambda e^{-\lambda t} (\lambda t)^j}{j!} \\ &= \sum_{j=n}^{\infty} \frac{\lambda e^{-\lambda t} (\lambda t)^{j-1}}{(j-1)!} - \sum_{j=n}^{\infty} \frac{\lambda e^{-\lambda t} (\lambda t)^j}{j!} \\ &= \frac{\lambda e^{-\lambda t} (\lambda t)^{n-1}}{(n-1)!} \end{aligned}$$

On constate bien que T_n suit une loi gamma de paramètres (n, λ) . (Dans la littérature mathématique, cette loi est aussi fréquemment appelée *loi d'Erlang* de paramètre n .) On remarquera que pour $n = 1$ cette distribution n'est autre que l'exponentielle.

Une loi gamma de paramètres $\lambda = \frac{1}{2}$ et $t = n/2$ (n étant ici entier positif) est aussi appelée loi χ_n^2 (lire chi-carré à n degrés de liberté). A titre d'exemple de variable de loi χ_n^2 , on peut citer l'erreur commise lors d'un tir sur une cible à n dimensions lorsque l'erreur dans chaque dimension est normalement distribuée. La loi χ_n^2 sera étudiée au chapitre 6 où son lien avec la loi normale sera expliqué.

Exemple 5.26 Soit X une variable aléatoire gamma de paramètres t et λ . Calculer

- a) $E[X]$;
b) $\text{Var}(X)$.

SOLUTION a)

$$\begin{aligned} E[X] &= \frac{1}{\Gamma(t)} \int_0^{\infty} \lambda x e^{-\lambda x} (\lambda x)^{t-1} dx \\ &= \frac{1}{\lambda \Gamma(t)} \int_0^{\infty} \lambda e^{-\lambda x} (\lambda x)^t dx \\ &= \frac{\Gamma(t+1)}{\lambda \Gamma(t)} \\ &= \frac{t}{\lambda} \quad \text{selon l'équation (5.14)} \end{aligned}$$

- b) En calculant tout d'abord $E[X^2]$, on peut montrer que

$$\text{Var}(X) = \frac{t}{\lambda^2}$$

Les détails sont laissés en exercice. ■

5.6.2 Loi de Weibull

Cette loi possède de nombreuses applications en sciences de l'ingénieur. Originellement utilisée pour représenter les effets de l'usure de pièces, son usage s'est ensuite beaucoup étendu. En biologie, en particulier, on l'utilise pour modéliser la durée de vie d'un organisme, surtout lorsque la vie de cet organisme dépend de son «maillon le plus faible». En d'autres termes, considérons un organisme constitué de plusieurs composants et admettons que cet organisme meure dès la défaillance de n'importe lequel de ses composants. Il a été établi (au niveau théorique autant qu'empirique) que dans ce modèle la loi de Weibull donne une bonne approximation de la durée de vie de l'organisme étudié.

La fonction de répartition d'une variable de Weibull est

$$F(x) = \begin{cases} 0 & x \leq v \\ 1 - \exp \left\{ - \left(\frac{x-v}{\alpha} \right)^\beta \right\} & x > v \end{cases} \quad (5.15)$$

Une variable ayant une telle fonction de répartition est dite *variable de Weibull* avec paramètres v , α et β . Par dérivation on obtient sa densité:

$$f(x) = \begin{cases} 0 & x \leq v \\ \left(\frac{\beta}{\alpha} \right) \left(\frac{x-v}{\alpha} \right)^{\beta-1} \exp \left\{ - \left(\frac{x-v}{\alpha} \right)^\beta \right\} & x > v \end{cases}$$

5.6.3 Loi de Cauchy

Une *variable de Cauchy* de paramètre θ , $-\infty < \theta < +\infty$, a pour densité:

$$f(x) = \frac{1}{\pi} \frac{1}{[1 + (x - \theta)^2]} \quad -\infty < x < \infty$$

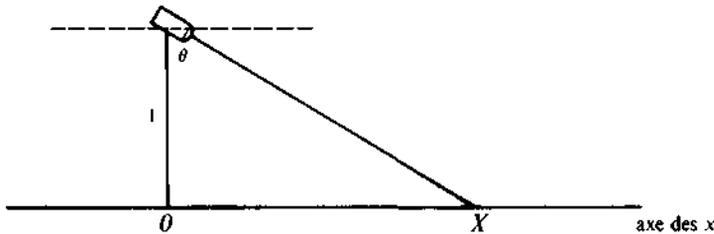


Figure 5.6

Exemple 5.27 Un projecteur à faisceau fin est mis en rotation autour de son axe, lequel est situé à une unité de distance de l'axe des abscisses (figure 5.6). X représente l'abscisse de l'intersection du faisceau avec l'axe $0x$ une fois que le projecteur s'est arrêté de tourner. (Si le faisceau n'est pas dirigé vers l'axe, on recommence l'expérience).

Comme on peut le constater sur la figure 5.6, X est déterminée par la valeur de l'angle θ entre le faisceau lumineux et l'axe Oy . On admettra l'hypothèse de nature physique que θ est uniformément distribuée entre $-\pi/2$ et $\pi/2$. La fonction de répartition de X sera alors

$$\begin{aligned} F(x) &= P\{X \leq x\} \\ &= P\{\text{tg } \theta \leq x\} \\ &= P\{\theta \leq \text{Arc tg } x\} \\ &= \frac{1}{2} + \frac{1}{\pi} \text{Arc tg } x \end{aligned}$$

Cette dernière équation résulte de la suivante, valide en raison de l'uniformité de la distribution de θ entre $-\pi/2$ et $\pi/2$:

$$P\{\theta \leq a\} = \frac{a - (-\pi/2)}{\pi} = \frac{1}{2} + \frac{a}{\pi}, \quad -\frac{\pi}{2} < a < \frac{\pi}{2}$$

Aussi la densité de X est-elle

$$f(x) = \frac{d}{dx} F(x) = \frac{1}{\pi(1+x^2)} \quad -\infty < x < \infty$$

ce qui établit que X suit une loi de Cauchy de paramètre 0^1 . ■

5.6.4 Loi bêta

On dit qu'une variable aléatoire suit une *loi bêta* si sa densité est

$$f(x) = \begin{cases} \frac{1}{B(a, b)} x^{a-1}(1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{sinon} \end{cases}$$

où

$$B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx$$

¹ La relation $d/dx \text{Arc tg } x = 1/(1+x^2)$ est établie comme suit: si $y = \text{Arc tg } x$, alors $\text{tg } y = x$ et par conséquent

$$\begin{aligned} 1 &= \frac{d}{dx} (\text{tg } y) = \frac{d}{dy} (\text{tg } y) \frac{dy}{dx} = \frac{d}{dy} \left(\frac{\sin y}{\cos y} \right) \frac{dy}{dx} \\ &= \left(\frac{\cos^2 y + \sin^2 y}{\cos^2 y} \right) \frac{dy}{dx} \end{aligned}$$

ou

$$\frac{dy}{dx} = \cos^2 y = \frac{\cos^2 y}{\sin^2 y + \cos^2 y} = \frac{1}{\text{tg}^2 y + 1} = \frac{1}{x^2 + 1}$$

La loi bêta s'applique à des phénomènes lors desquels la variable aléatoire peut prendre les valeurs comprises dans un certain intervalle $[c, d]$. En déclarant que c représente l'origine et en adoptant $d - c$ comme unité, cela revient à s'intéresser à l'intervalle $[0, 1]$.

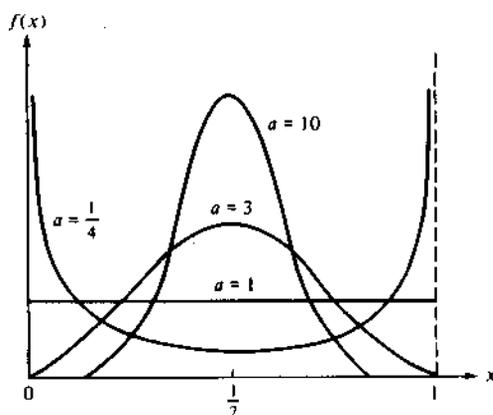


Figure 5.7

Densités bêta lorsque les paramètres a et b sont égaux

Comme le montre la figure 5.7, les densités bêta sont symétriques autour de $\frac{1}{2}$ lorsque $a = b$ et donnent de plus en plus de poids à cette région centrale au fur et à mesure que a augmente. La figure 5.8 montre que lorsque $b > a$ les densités sont ramassées à gauche, ce qui signifie que les petites valeurs de la variable sont plus probables. Lorsque $a > b$, les densités, inversement, sont ramassées à droite.

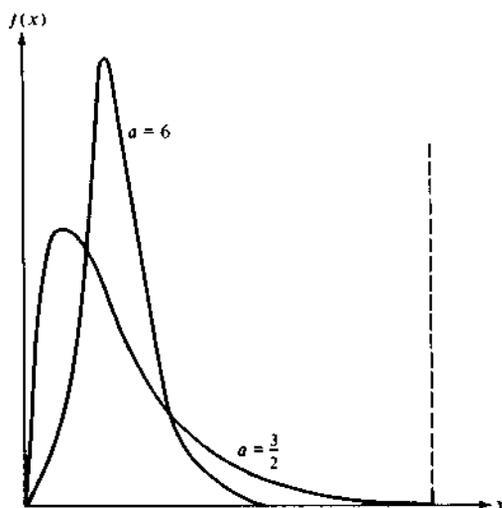


Figure 5.8

Densités bêta de paramètres (a, b) tels que

$$\frac{a}{a+b} = \frac{1}{20}$$

On peut montrer qu'il existe la relation suivante entre

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$$

et la fonction gamma:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad (5.16)$$

En utilisant l'équation (5.14) avec l'identité ci-dessus, il est facile de montrer que si X est une variable aléatoire bêta de paramètres a et b , alors

$$E[X] = \frac{a}{a+b}$$

$$\text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

REMARQUE. Une vérification de l'équation (5.16) est donnée dans l'exemple 6.29 du chapitre 6.

5.7 DISTRIBUTION D'UNE FONCTION DE VARIABLE ALÉATOIRE

5.7.1 Illustration du problème

Il arrive souvent que la distribution d'une variable aléatoire soit connue mais que l'on s'intéresse plutôt à celle d'une fonction de cette variable. En d'autres termes, on connaîtrait la distribution de X mais souhaiterait connaître celle de $g(X)$. Pour y parvenir, il faut exprimer l'événement $g(X) \leq y$ sous forme d'une condition où X appartient à un certain ensemble, comme l'illustrent les exemples suivants.

Exemple 5.28 Soit X uniformément distribuée dans l'intervalle $(0, 1)$. On obtiendra la distribution de $Y = X^n$ de la manière suivante: pour $0 \leq y \leq 1$,

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{X^n \leq y\} \\ &= P\{X \leq y^{1/n}\} \\ &= F_X(y^{1/n}) \\ &= y^{1/n} \end{aligned}$$

Aussi peut-on donner à la densité de Y la forme suivante, par exemple:

$$f_Y(y) = \begin{cases} \frac{1}{n} y^{1/n-1} & 0 \leq y \leq 1 \\ 0 & \text{sinon} \end{cases} \quad \blacksquare$$

Exemple 5.29 Pour toute variable aléatoire continue X de densité f_X , on obtient la fonction de répartition de $Y = X^2$ de la manière suivante: pour $y \geq 0$,

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{X^2 \leq y\} \\ &= P\{-\sqrt{y} \leq X \leq \sqrt{y}\} \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

Une dérivation livre la densité

$$f_Y(y) = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})] \quad \blacksquare$$

Exemple 5.30 Soit X de densité f_X . La densité de $Y = |X|$ peut être calculée ainsi: pour $y \geq 0$

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{|X| \leq y\} \\ &= P\{-y \leq X \leq y\} \\ &= F_X(y) - F_X(-y) \end{aligned}$$

et par dérivation de nouveau on obtient

$$f_Y(y) = f_X(y) + f_X(-y) \quad y \geq 0 \quad \blacksquare$$

5.7.2 Densité d'une fonction de variable aléatoire

La méthode utilisée pour traiter les exemples 5.28 à 5.30 peut servir de démonstration au théorème général suivant:

Théorème 5.4

Soit X une variable aléatoire continue de densité f_X . Soit g une fonction strictement monotone (croissante ou décroissante) et dérivable, donc continue. La densité de la variable aléatoire $Y = g(X)$ est alors

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)] \left| \frac{d}{dy} g^{-1}(y) \right| & \text{si } y = g(x) \text{ pour un } x \text{ quelconque} \\ 0 & \text{si } y \neq g(x) \text{ pour tout } x \end{cases}$$

où $g^{-1}(y)$ est défini comme étant égal à x tel que $g(x) = y$.

La démonstration de ce théorème est laissée en exercice.

Exemple 5.31 Soit X une variable aléatoire continue non-négative de densité f et soit $Y = X^n$. Trouver la densité f_Y de Y .

SOLUTION. Si $g(x) = x^n$ alors

$$g^{-1}(y) = y^{1/n}$$

et

$$\frac{d}{dy}\{g^{-1}(y)\} = \frac{1}{n}y^{\frac{1}{n}-1}$$

Donc, d'après le théorème 5.4, on obtient

$$f_Y(y) = \frac{1}{n}y^{\frac{1}{n}-1}f(y^{1/n})$$

Si $n = 2$, cela donne

$$f_Y(y) = \frac{1}{2\sqrt{y}}f(\sqrt{y})$$

ce qui correspond (puisque $X \geq 0$) au résultat de l'exemple 5.29. ■

5.8 EXERCICES THÉORIQUES

5.8.1 La vitesse d'une molécule au sein d'un gaz homogène en état d'équilibre est une variable aléatoire, dont la fonction de densité est donnée par

$$f(x) = \begin{cases} ax^2 e^{-bx^2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

où $b = m/2kT$ et k , T , m sont respectivement la constante de Boltzmann, la température absolue et la masse de la molécule. Evaluer a en termes de b .

5.8.2 Montrer que

$$E[Y] = \int_0^{\infty} P\{Y > y\} dy - \int_0^{\infty} P\{Y < -y\} dy$$

Pour cela, montrer que

$$\int_0^{\infty} P\{Y < -y\} dy = - \int_{-\infty}^0 xf_Y(x) dx$$

$$\int_0^{\infty} P\{Y > y\} dy = \int_0^{\infty} xf_Y(x) dx$$

5.8.3 Si X a pour densité f , montrer que

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Utiliser pour cela l'exercice théorique 5.8.2 en commençant avec

$$E[g(X)] = \int_0^{\infty} P\{g(X) > y\}dy - \int_0^{\infty} P\{g(X) < -y\}dy$$

et procéder comme dans la preuve donnée dans le texte lorsque $g(X) \geq 0$.

5.8.4 Démontrer le théorème 5.3.

5.8.5 Utiliser le fait qu'une variable aléatoire non négative Y ait pour espérance

$$E[Y] = \int_0^{\infty} P\{Y > t\}dt$$

pour montrer que pour une variable aléatoire non négative X

$$E[X^n] = \int_0^{\infty} nx^{n-1}P\{X > x\}dx$$

Commencer avec

$$E[X^n] = \int_0^{\infty} P\{X^n > t\}dt$$

et effectuer le changement de variable $t = x^n$.

5.8.6 Donner un ensemble d'événements E_a , $0 < a < 1$, ayant la propriété que

$$P\{E_a\} = 1 \text{ pour tout } a \text{ mais que } P\left\{\bigcap_a E_a\right\} = 0.$$

Pour cela, prendre X uniforme sur $(0, 1)$ et définir chaque E_a en fonction de X .

5.8.7 L'écart-type de X , noté σ , est donné par

$$\sigma = \sqrt{\text{Var}(X)}$$

Trouver l'écart-type de la transformation linéaire $(aX + b)$ si X a pour variance σ^2 .

5.8.8 Soit X une variable aléatoire prenant ses valeurs entre 0 et c , c'est à dire $P(0 \leq X \leq c) = 1$. Montrer que

$$\text{Var}(X) \leq \frac{c^2}{4}$$

Une approche consiste tout d'abord à voir que

$$E[X^2] \leq cE[X]$$

Utiliser ensuite cette inégalité pour montrer que

$$\text{Var}(X) \leq c^2[\alpha(1-\alpha)] \quad \text{où } \alpha = \frac{E[X]}{c}$$

5.8.9 Si Z est une variable aléatoire normale standard, montrer que pour $x > 0$,

a) $P\{Z > x\} = P\{Z < -x\}$;

b) $P\{|Z| > x\} = 2P\{Z > x\}$;

c) $P\{|Z| < x\} = 2P\{Z < x\} - 1$.

5.8.10 Soit $f(x)$ la densité d'une variable aléatoire normale de moyenne μ et de variance σ^2 . Montrer que $\mu - \sigma$ et $\mu + \sigma$ sont les points d'inflexion de cette fonction. Autrement dit montrer que $f''(x) = 0$ quand $x = \mu - \sigma$ ou $x = \mu + \sigma$.

5.8.11 Utiliser l'identité de l'exercice théorique 5.8.5 pour dériver $E[X^2]$ lorsque X est une variable aléatoire exponentielle de paramètre λ .

5.8.12 La médiane d'une variable aléatoire continue ayant une fonction de répartition F est la valeur m , telle que $F(m) = \frac{1}{2}$. En d'autres termes une variable aléatoire a autant de chances d'être plus grande ou plus petite que sa médiane. Trouver la médiane de X si X est une variable:

- a) uniformément distribuée sur (a, b) ;
- b) normale de paramètres μ, σ^2 ;
- c) exponentielle de paramètre λ .

5.8.13 Le mode d'une variable aléatoire continue ayant une densité f est la valeur x pour laquelle $f(x)$ atteint son maximum. Calculer le mode de X dans les cas a), b) et c) de l'exercice théorique 5.8.12.

5.8.14 Si X est une variable aléatoire exponentielle de paramètre λ , et $c > 0$, montrer que cX est exponentielle de paramètre λ/c .

5.8.15 Calculer le taux de panne de X quand X est uniformément distribuée sur l'intervalle $(0, a)$.

5.8.16 Si X a un taux de panne $\lambda_X(t)$, calculer le taux de panne de aX , où a est une constante positive.

5.8.17 Vérifier que l'intégrale de la fonction de densité de la loi gamma donne 1.

5.8.18 Si X est une variable aléatoire de moyenne $1/\lambda$, montrer que

$$E[X^k] = \frac{k!}{\lambda^k} \quad k = 1, 2, \dots$$

Utiliser la densité de gamma pour évaluer ceci.

5.8.19 Vérifier que $\text{Var}(X) = \frac{t}{\lambda^2}$ lorsque X est une variable aléatoire gamma de paramètres t et λ .

5.8.20 Montrer que $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. Procéder de la façon suivante: on a

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} e^{-x} x^{-\frac{1}{2}} dx$$

Effectuer alors le changement de variable $y = \sqrt{2x}$ et relier l'expression obtenue à la distribution normale.

5.8.21 Calculer la fonction taux de panne d'une variable aléatoire gamma de paramètre (t, λ) et montrer qu'elle est croissante quand $t \geq 1$ et décroissante quand $t \leq 1$.

5.8.22 Calculer la fonction taux de panne d'une variable aléatoire de Weibull et montrer qu'elle est croissante quand $\beta \geq 1$ et décroissante quand $\beta \leq 1$.

5.8.23 Montrer que le graphe de $\ln(\ln(1 - F(x))^{-1})$ en fonction de $\ln(x)$ est une droite de pente β quand $F(\cdot)$ est une fonction de répartition de Weibull. Montrer également qu'approximativement 63,2% de toutes les observations d'une telle distribution sont inférieures à α . Supposer que $v = 0$.

5.8.24 Soit

$$Y = \left(\frac{X - v}{\alpha}\right)^\beta$$

Montrer que si X est une variable aléatoire de Weibull dont les paramètres sont v , α , β , alors Y est une variable aléatoire exponentielle de paramètre $\lambda = 1$ et vice versa.

5.8.25 Si X est une variable aléatoire bêta de paramètres a et b , montrer que

$$E[X] = \frac{a}{a+b}, \quad \text{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

5.8.26 Si X est uniformément distribuée sur l'intervalle (a, b) , quelle est la variable aléatoire dépendant linéairement de X qui admet une distribution uniforme sur l'intervalle $(0, 1)$?

5.8.27 On considère la distribution bêta de paramètres (a, b) . Montrer que:

- a) quand $a > 1, b > 1$, la densité est unimodale (c'est-à-dire qu'elle a un mode unique) de mode égal à $(a - 1)/(a + b - 2)$;
 b) quand $a \leq 1, b \leq 1, a + b < 2$, la densité est soit unimodale, avec un mode en 0 ou en 1, soit en forme de U avec des modes à la fois en 0 et en 1;
 c) quand $a = 1 = b$ tous les points dans l'intervalle $[0, 1]$ sont des modes.

5.8.28 Soit X une variable aléatoire continue ayant une fonction de répartition F . On définit la variable aléatoire Y par $Y = F(X)$. Montrer que Y est uniformément distribuée sur l'intervalle $(0, 1)$.

5.8.29 Soit X une variable aléatoire de densité de probabilité f_X . Trouver la fonction de densité de la variable aléatoire définie par $Y = aX + b$.

5.8.30 Trouver la fonction de densité de $Y = e^X$ quand X est distribuée selon une loi normale de paramètres μ et σ^2 . La variable aléatoire Y suit une distribution dite log-normale (puisque $\ln Y$ a une distribution normale) de paramètres μ et σ^2 .

5.8.31 Soit X et Y des variables aléatoires indépendantes qui toutes deux ont la même chance d'être égales à $1, 2, \dots, (10)^N$, où N est très grand. Appelons D le plus grand commun diviseur de X et Y et soit $Q_k = P\{D = k\}$.

- a) Montrer par un argument heuristique que $Q_k = 1/k^2 Q_1$. On peut remarquer que pour que D soit égal à k , k doit diviser X et Y et X/k et Y/k doivent être premiers entre eux (c'est-à-dire que leur plus grand commun diviseur doit être égal à 1).

- b) Utiliser a) pour montrer que

$$Q_1 = P\{X \text{ et } Y \text{ soient premiers entre eux}\} = \frac{1}{\sum_{k=1}^{\infty} 1/k^2}$$

En vertu d'une identité bien connue selon laquelle $\sum_{k=1}^{\infty} 1/k^2 = \pi^2/6$, il s'ensuit que $Q_1 = 6/\pi^2$.

(En théorie des nombres ceci est connu sous le nom de théorème de Legendre).

- c) Maintenant établir que

$$Q_1 = \prod_{i=1}^{\infty} \left(\frac{P_i^2 - 1}{P_i^2} \right)$$

où P_i est le i -ème des nombres premiers rangés dans l'ordre de croissance, P_1 étant 2. A titre d'indication, on peut noter que X et Y seront premiers entre eux s'ils n'ont pas de facteurs premiers communs.

Ainsi de b) on voit que $\prod_{i=1}^{\infty} \left(\frac{P_i^2 - 1}{P_i^2} \right) = 6/\pi^2$; ce résultat a déjà été mentionné sans

explication dans le problème 4.11.11 (la relation entre ce problème et le problème 4.11.11 réside dans le fait que X et Y sont premiers entre eux si leur produit XY n'a pas de facteur premier répété).

5.9 PROBLÈMES

5.9.1 Soit X une variable aléatoire dont la fonction de probabilité est:

$$f(x) = \begin{cases} c(1 - x^2) & -1 < x < 1 \\ 0 & \text{sinon} \end{cases}$$

- a) Quelle est la valeur de c ?
 b) Quelle est la fonction de répartition de X ?

5.9.2 Pour fonctionner, un système utilise une cellule interchangeable. On dispose de la pièce originale et d'une cellule de rechange. Si le système a une durée de vie aléatoire X et que sa densité est donnée (en mois) par:

$$f(x) = \begin{cases} cxe^{-x/2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

quelle est la probabilité que le système fonctionne pendant au moins 5 mois?

5.9.3 On considère la fonction

$$f(x) = \begin{cases} c(2x - x^3) & 0 < x < \frac{5}{2} \\ 0 & \text{sinon} \end{cases}$$

La fonction f représente-t-elle une fonction de densité? Si oui, déterminer c . Même question avec $f(x)$ donnée par

$$f(x) = \begin{cases} c(2x - x^2) & 0 < x < \frac{5}{2} \\ 0 & \text{sinon} \end{cases}$$

5.9.4 La fonction de densité de X , variable aléatoire représentant la durée de vie en heures d'un certain composant électronique, est donnée par:

$$f(x) = \begin{cases} \frac{10}{x^2} & x > 10 \\ 0 & x \leq 10 \end{cases}$$

- a) Trouver $P\{X > 20\}$.
 b) Quelle est la fonction de répartition de X ?
 c) Quelle est la probabilité que parmi 6 composants au moins 3 d'entre eux fonctionnent durant au moins 15 heures? Quelles hypothèses faites-vous?

5.9.5 Une station service est approvisionnée en essence une fois par semaine. Si son volume de vente hebdomadaire, en milliers de litres, est une variable aléatoire de fonction de densité:

$$f(x) = \begin{cases} 5(1 - x)^4 & 0 < x < 1 \\ 0 & \text{sinon} \end{cases}$$

quelle est la capacité que doit avoir le réservoir pour que la probabilité d'épuiser l'approvisionnement d'une semaine soit égal à 0,01?

5.9.6 Calculer $E[X]$ si X a comme densité

$$\text{a) } f(x) = \begin{cases} \frac{1}{4} x e^{-\frac{x}{2}} & x > 0 \\ 0 & \text{sinon} \end{cases}$$

$$\text{b) } f(x) = \begin{cases} c(1-x^2) & -1 < x < 1 \\ 0 & \text{sinon} \end{cases}$$

$$\text{c) } f(x) = \begin{cases} \frac{5}{x^2} & x > 5 \\ 0 & x \leq 5 \end{cases}$$

5.9.7 La densité de X est donnée par

$$f(x) = \begin{cases} a + bx^2 & 0 \leq x \leq 1 \\ 0 & \text{sinon} \end{cases}$$

Si $E[X] = \frac{3}{5}$, trouver a et b .

5.9.8 La durée de vie en heures d'un tube électronique est une variable aléatoire ayant pour densité

$$f(x) = x e^{-x} \quad x \geq 0$$

Calculer l'espérance de la durée de vie d'un tel tube.

5.9.9 Considérer l'exemple 4.12 du chapitre 4 et supposer maintenant que la demande saisonnière est une variable aléatoire continue de densité f . Montrer que la quantité optimale à stocker est la valeur s^* satisfaisant

$$F(s^*) = \frac{b}{b+l}$$

où b est le profit net par unité vendue, l est la perte nette par unité invendue et F est la fonction de distribution cumulative de la demande saisonnière.

5.9.10 Les trains à destination de A arrivent à la gare toutes les 15 min. à partir de 7 h. du matin, et les trains à destination de B arrivent toutes les 15 min. également, mais à partir de 7 h 05 du matin. Un certain passager arrive à la gare à une

heure uniformément distribuée entre 7 h et 8 h du matin et prend le premier train qui se présente. a) Dans quelle proportion des cas se rendra-t-il à la destination A? b) Qu'en est-il si le passager arrive à une heure uniformément distribuée entre 7 h 10 et 8 h 10 du matin?

5.9.11 Un point est choisi au hasard sur un segment de longueur L . Interpréter cet énoncé et trouver la probabilité que le rapport entre le plus petit et le plus grand segment soit inférieure à $\frac{1}{4}$.

5.9.12 Un bus circule entre 2 villes A, B distantes de 100 miles. On admet que lorsque le bus tombe en panne, la distance de l'endroit de la panne à la ville A a une distribution uniforme sur l'intervalle $(0, 100)$. Il y a une station de réparation en A , une en B et une autre à mi-distance entre A et B . On suggère qu'il serait plus efficace d'avoir les 3 stations localisées respectivement à 25, 50 et 75 miles de A . Etes-vous de cet avis? Pourquoi?

5.9.13 Vous arrivez à un arrêt de bus à 10 h sachant que le bus arrivera à un certain instant qui est distribué uniformément entre 10 h et 10 h 30. a) Quelle est la probabilité que vous deviez attendre plus de 10 min.? b) Si à 10 h 15 le bus n'est pas encore arrivé, quelle est la probabilité que vous deviez attendre au moins 10 min. supplémentaires?

5.9.14 Soit X une variable aléatoire uniforme $(0, 1)$. Calculer $E[X^n]$ en utilisant le théorème 5.1 et vérifier le résultat en utilisant la définition de l'espérance.

5.9.15 Si X est une variable aléatoire normale de paramètres $\mu = 10$ et $\sigma^2 = 36$, calculer:

- (a) $P\{X > 5\}$; (b) $P\{4 < X < 16\}$; (c) $P\{X < 8\}$;
 (d) $P\{X < 20\}$; (e) $P\{X > 16\}$.

5.9.16 La quantité annuelle de précipitations (en cm) dans une certaine région est distribuée selon une loi normale avec $\mu = 140$ et $\sigma^2 = 16$. Quelle est la probabilité qu'à partir de cette année, il faille attendre plus de 10 ans avant d'obtenir une année avec une quantité annuelle de pluie supérieure à 150 cm? Quelles hypothèses faites-vous?

5.9.17 Un homme tirant sur une cible reçoit 10 points si son coup est à moins de 1 cm du centre de la cible, 5 points s'il s'en éloigne de 1 à 3 cm et 3 points s'il s'en éloigne de 3 à 5 cm. Trouver l'espérance du nombre de points si la distance du tir au centre de la cible est uniformément distribuée entre 0 et 10.

5.9.18 Supposer que X est une variable aléatoire normale de moyenne 5. Si $P\{X > 9\} = .2$, quelle est approximativement $\text{Var}(X)$?

5.9.19 Soit X une variable aléatoire normale de moyenne 12 et de variance 4. Trouver la valeur de c telle que $P\{X > c\} = .10$.

5.9.20 Si 65% de la population d'une grande communauté est en faveur de l'augmentation des taxes scolaires, estimer la probabilité qu'un échantillon aléatoire de 100 personnes contienne

- a) au moins 50 personnes en faveur de l'augmentation;
- b) entre 60 et 70 personnes (inclus) en faveur de l'augmentation;
- c) moins de 75 personnes en faveur de l'augmentation.

5.9.21 On suppose que la taille, en centimètres, d'un homme âgé de 25 ans est une variable aléatoire normale de paramètres $\mu = 175$ et $\sigma^2 = 36$. Quel est le pourcentage d'hommes de 25 ans ayant une taille supérieure à 185 cm? Parmi les hommes mesurant plus de 180 cm, quel pourcentage d'entre eux dépassent 192 cm?

5.9.22 La largeur (en cm) d'une fente entaillée dans une pièce fabriquée en aluminium est distribuée selon une loi normale de paramètres $\mu = 2$ et $\sigma = 0,007$. Les limites de tolérance sont données comme étant $2,0000 \pm 0,0120$. a) Quel sera le pourcentage de pièces défectueuses? b) Quelle est la valeur maximale que peut prendre σ afin que le pourcentage de pièces défectueuses ne dépasse pas 1%, si la largeur des fentes suit une distribution normale de paramètres $\mu = 2$ et σ ?

5.9.23 On considère 1000 jets indépendants d'un dé homogène. Calculer une approximation de la probabilité que la face 6 apparaisse entre 150 et 200 fois. Si la face 6 apparaît exactement 200 fois, trouver la probabilité que la face 5 apparaisse moins de 150 fois.

5.9.24 La durée de vie des puces d'ordinateur interactif produites par un fabricant de semi-conducteurs est distribuée normalement de paramètres $\mu = 1,4 \times 10^6$ heures et $\sigma = 3 \times 10^5$ heures. Quelle est la probabilité approximative qu'un lot de 100 puces contienne au moins 20 puces dont la durée de vie ne dépasse pas $1,8 \times 10^6$?

5.9.25 Utiliser un programme pour calculer $P\{X \leq 25\}$, quand X est une variable aléatoire binomiale de paramètres $n = 300$, $p = 0,1$. Comparer le résultat avec son approximation

- a) poissonnienne et
- b) normale.

En utilisant l'approximation normale, écrire la probabilité cherchée sous la forme $P\{X \leq 25,5\}$ afin d'introduire la correction de continuité. (Vous aurez besoin d'un programme pour calculer l'approximation de Poisson.)

5.9.26 Deux types de pièces de monnaie sont produites dans une fabrique: des pièces homogènes et des pièces biaisées, lesquelles montrent la face pile dans 55% des cas. Supposons que nous possédions une pièce de cette fabrique et que nous ignorions si elle est homogène ou biaisée. Pour pouvoir déterminer de quelle pièce il s'agit, nous effectuons le test statistique suivant: la pièce est lancée 1000 fois; si l'on obtient pile 525 fois ou plus, alors on conclut que c'est une pièce biaisée, tandis que si l'on obtient pile moins de 525 fois, alors on conclut que c'est une pièce homogène. Si la pièce est réellement homogène, quelle est la probabilité que l'on aboutisse à une conclusion fautive? Qu'en est-il si la pièce est biaisée?

59.27 Sur 10 000 jets indépendants, une pièce de monnaie donne 5800 fois pile. Est-il raisonnable de présumer que la pièce n'est pas homogène? Expliquer pourquoi.

59.28 Une image est composée de 2 régions, l'une blanche et l'autre noire. Lors d'une lecture digitale, un point choisi aléatoirement dans la zone blanche donnera une valeur qui est distribuée selon une loi normale de paramètres (4, 4). Un point choisi aléatoirement dans la partie noire aura une valeur distribuée selon la même loi, mais de paramètres (6, 9). Considérons un point choisi aléatoirement sur l'image et qui présente une valeur égale à 5. Désignons par α la fraction de l'image qui est noire. Pour quelle valeur de α a-t-on la même probabilité de se tromper en concluant que le point choisi provient de la zone noire ou le contraire?

59.29 a) Une caserne de pompiers doit être construite sur une route de longueur A , $A < \infty$. Si un incendie se déclare en des points uniformément distribués entre (0, A), où doit être située la caserne pour minimiser l'espérance de la distance jusqu'au feu? Autrement dit, trouver a tel que $E[|X - a|]$ soit minimisée lorsque X est distribué uniformément sur (0, A).

b) Supposer à présent que la route soit de longueur infinie - partant du point 0 vers ∞ . Si la distance d'un incendie au point 0 est distribuée selon une loi exponentielle de paramètre λ , où doit se trouver la caserne? Ici, on cherche à minimiser $E[|X - a|]$ où X est exponentielle de paramètre λ .

59.30 Le temps (en heures) nécessaire pour réparer une machine est une variable aléatoire exponentiellement distribuée de paramètre $\lambda = \frac{1}{2}$.

a) Quelle est la probabilité que le temps de réparation excède 2 heures?

b) Quelle est la probabilité conditionnelle qu'une réparation prenne au moins 10 heures, étant donné que sa durée a déjà dépassé 9 heures?

59.31 Le nombre d'années de fonctionnement d'une radio est distribué selon une loi exponentielle de paramètre $\lambda = \frac{1}{8}$. Si Jones achète une radio d'occasion, quelle est la probabilité qu'elle fonctionne encore après 8 nouvelles années d'usage?

59.32 Jones estime que le nombre total de milliers de miles que peut parcourir une voiture avant qu'elle ne soit mise à la ferraille est une variable aléatoire exponentielle de paramètre $\lambda = \frac{1}{20}$. Smith a une voiture dont il prétend qu'elle n'a roulé que 10 000 miles. Si Jones achète la voiture, quelle est la probabilité qu'il puisse encore l'utiliser pendant au moins 20 000 miles? Refaire le problème en considérant l'hypothèse que la durée de vie de la voiture (exprimée en milliers de miles) suit une distribution uniforme sur l'intervalle (0, 40) et non plus une loi exponentielle.

59.33 Le taux de cancer des poumons chez les fumeurs âgés de t années, $\lambda(t)$, est tel que

$$\lambda(t) = .027 + .00025(t - 40)^2, \quad t \geq 40$$

En supposant qu'un fumeur de 40 ans survive à toute autre maladie, quelle est la probabilité qu'il survive à l'âge de

- a) 50 ans,
 - b) 60 ans
- sans contracter un cancer des poumons.

5.9.34 Supposons que la distribution de la durée de vie d'un élément a une fonction taux de panne $\lambda(t) = t^3$, $t > 0$.

- a) Quelle est la probabilité que l'élément survive à l'âge de 2 ans?
- b) Quelle est la probabilité que la durée de vie de l'élément se situe entre 0,4 et 1,4?
- c) Quelle est la probabilité qu'un élément âgé de 1 année survive à l'âge de 2 ans?

5.9.35 Si X est uniformément distribuée sur l'intervalle $(-1, 1)$, trouver:

- a) $P\{|X| > \frac{1}{2}\}$;
- b) la densité de la variable aléatoire $|X|$.

5.9.36 Si Y est de distribution uniforme sur l'intervalle $(0, 5)$, quelle est la probabilité que les racines de l'équation $4x^2 + 4xY + Y + 2 = 0$ soient toutes deux réelles?

5.9.37 Si X est une variable aléatoire exponentielle de paramètre $\lambda = 1$, calculer la densité de la variable aléatoire Y définie par $Y = \ln X$.

5.9.38 Si X est uniformément distribuée sur l'intervalle $(0,1)$, trouver la densité de $Y = e^X$.

5.9.39 Trouver la distribution de $R = A \sin \theta$, où A est une constante fixée et où θ est uniformément distribuée sur l'intervalle $(-\pi/2, \pi/2)$. Une telle variable aléatoire R apparaît en théorie de la balistique. Si un projectile est tiré de l'origine avec un angle α par rapport à la terre et avec une vitesse V , alors son point de chute R peut s'exprimer par $R = (v^2/g) \sin 2\alpha$ où g est la constante de gravitation égale à $9,81 \text{ m/s}^2$.

Variables aléatoires simultanées

6.1 DÉFINITION DES DISTRIBUTIONS SIMULTANÉES

6.1.1 Fonction de répartition conjointe

Nous n'avons traité jusqu'ici que de distributions de variables isolées. Or, il est souvent nécessaire de considérer des événements relatifs à deux variables simultanément, ou même à plus de deux variables. On définit pour traiter de tels problèmes une *fonction F de répartition simultanée*, ou *conjointe*, pour toute paire de variables aléatoires X et Y :

$$F(a, b) = P\{X \leq a, Y \leq b\} \quad -\infty < a, b < \infty$$

6.1.2 Fonction de répartition marginale

La fonction de répartition de X peut être déduite de la fonction de répartition conjointe de X et Y comme suit:

$$\begin{aligned} F_X(a) &= P\{X \leq a\} \\ &= P\{X \leq a, Y < \infty\} \\ &= P\left(\lim_{b \rightarrow \infty} \{X \leq a, Y \leq b\}\right) \\ &= \lim_{b \rightarrow \infty} P\{X \leq a, Y \leq b\} \\ &= \lim_{b \rightarrow \infty} F(a, b) \\ &= F(a, \infty) \end{aligned}$$

Le lecteur remarquera qu'il a une fois de plus été fait usage, dans les équations ci-dessus, de la propriété de continuité des fonctions de probabilité. On obtient par ailleurs de manière similaire la fonction de répartition de Y :

$$\begin{aligned} F_Y(b) &= P\{Y \leq b\} \\ &= \lim_{a \rightarrow -\infty} F(a, b) \\ &\equiv F(\infty, b) \end{aligned}$$

Les fonctions de répartition F_X et F_Y sont parfois dites *fonctions de répartition marginales* de X et Y .

6.1.3 Universalité des fonctions simultanées

La probabilité de tous les événements simultanément relatifs à X et Y peut théoriquement être calculée grâce aux fonctions conjointes définies plus haut. Supposons par exemple que l'on veuille calculer la probabilité que simultanément X soit supérieure à a et Y à b . On procédera par exemple comme suit:

$$\begin{aligned} P\{X > a, Y > b\} &= 1 - P(\{X > a, Y > b\}^c) \\ &= 1 - P(\{X > a\}^c \cup \{Y > b\}^c) \\ &= 1 - P(\{X \leq a\} \cup \{Y \leq b\}) \\ &= 1 - [P\{X \leq a\} + P\{Y \leq b\} - P\{X \leq a, Y \leq b\}] \\ &= 1 - F_X(a) - F_Y(b) + F(a, b) \end{aligned} \quad (6.1)$$

Cette dernière formule est un cas particulier de (6.2) ci-dessous, dont la démonstration est laissée en exercice.

$$\begin{aligned} P\{a_1 < X \leq a_2, b_1 < Y \leq b_2\} \\ = F(a_2, b_2) + F(a_1, b_1) - F(a_1, b_2) - F(a_2, b_1) \end{aligned} \quad (6.2)$$

sous les conditions $a_1 < a_2, b_1 < b_2$.

6.1.4 Loi discrète conjointe

Dans le cas où X et Y sont deux variables discrètes, il est commode de définir la fonction p suivante, dite *loi de probabilité simultanée ou conjointe* de X et Y :

$$p(x, y) = P\{X = x, Y = y\}$$

La loi de probabilité marginale de X s'en déduit ainsi:

$$\begin{aligned} p_X(x) &= P\{X = x\} \\ &= \sum_{y: p(x, y) > 0} p(x, y) \end{aligned}$$

De façon similaire

$$p_Y(y) = \sum_{x: p(x,y) > 0} p(x, y)$$

6.1.5 Exemples de variables aléatoires simultanées

Exemple 6.1 On tire au hasard 3 boules d'une urne en contenant 3 rouges, 4 blanches et 5 bleues. X et Y désignent respectivement le nombre de boules rouges et celui de boules blanches tirées. La loi de probabilité simultanée $p(i, j) = P(X = i, Y = j)$ de X et Y est alors:

$$p(0, 0) = \binom{5}{3} / \binom{12}{3} = \frac{10}{220}$$

$$p(0, 1) = \binom{4}{1} \binom{5}{2} / \binom{12}{3} = \frac{40}{220}$$

$$p(0, 2) = \binom{4}{2} \binom{5}{1} / \binom{12}{3} = \frac{30}{220}$$

$$p(0, 3) = \binom{4}{3} / \binom{12}{3} = \frac{4}{220}$$

$$p(1, 0) = \binom{3}{1} \binom{5}{2} / \binom{12}{3} = \frac{30}{220}$$

$$p(1, 1) = \binom{3}{1} \binom{4}{1} \binom{5}{1} / \binom{12}{3} = \frac{60}{220}$$

$$p(1, 2) = \binom{3}{1} \binom{4}{2} / \binom{12}{3} = \frac{18}{220}$$

$$p(2, 0) = \binom{3}{2} \binom{5}{1} / \binom{12}{3} = \frac{15}{220}$$

$$p(2, 1) = \binom{3}{2} \binom{4}{1} / \binom{12}{3} = \frac{12}{220}$$

$$p(3, 0) = \binom{3}{3} / \binom{12}{3} = \frac{1}{220}$$

Il est commode d'exprimer ces probabilités à l'aide d'un tableau à deux entrées tel que le tableau 6.1.

Tableau 6.1

$i \backslash j$	0	1	2	3	Somme de ligne $= P\{X = i\}$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Somme de colonne $= P\{Y = j\}$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

Le lecteur remarquera que la loi marginale de X est calculée en faisant les totaux par ligne, tandis que celle de Y l'est en faisant les totaux par colonne. C'est le fait que les lois de X et Y individuellement puissent être lues dans les marges du tableau qui leur vaut leur nom de lois marginales. ■

Exemple 6.2 On sait que 15% des familles d'une certaine localité n'ont pas d'enfant, 20% d'entre elles en ont 1, 35% en ont 2 et 30% en ont 3. On sait de plus que pour chaque famille un enfant a autant de chances d'être un garçon qu'une fille, indépendamment du sexe de ses frères et soeurs. La loi de probabilité conjointe de G , le nombre de garçons d'une famille tirée au hasard, et de F , le nombre de filles dans cette famille, est donnée dans le tableau 6.2.

Tableau 6.2

$i \backslash j$	0	1	2	3	Somme de ligne $= P\{G = i\}$
0	.15	.10	.0875	.0375	.3750
1	.10	.175	.1125	0	.3875
2	.0875	.1125	0	0	.2000
3	.0375	0	0	0	.0375
Somme de colonne $= P\{F = j\}$.375	.3875	.2000	.0375	

SOLUTION. On obtient ces probabilités de la manière suivante:

$$P\{G = 0, F = 0\} = P\{\text{n'avoir pas d'enfant}\} = 0,15$$

$$\begin{aligned} P\{G = 0, F = 1\} &= P\{\text{avoir 1 fille et 1 enfant au total}\} \\ &= P\{\text{avoir 1 enfant}\} \cdot P\{\text{avoir 1 fille} | \text{on a 1 enfant}\} \\ &= (0,20)\left(\frac{1}{2}\right) \end{aligned}$$

$$\begin{aligned} P\{G = 0, F = 2\} &= P\{\text{avoir 2 filles et un total de 2 enfants}\} \\ &= P\{\text{avoir 2 enfants}\} \cdot P\{\text{avoir 2 filles} | \text{on a 2 enfants}\} \\ &= (0,35)\left(\frac{1}{2}\right)^2. \end{aligned}$$

La vérification des autres résultats du tableau 6.2 est laissée au lecteur. ■

6.1.6 Densité conjointe

Les variables X et Y sont dites *conjointement continues* s'il existe une fonction f de deux arguments réels ayant pour tout sous-ensemble C du plan la propriété suivante:

$$P\{(X, Y) \in C\} = \iint_{(x,y) \in C} f(x, y) \, dx \, dy \quad (6.3)$$

La fonction f est appelée *densité conjointe* ou *simultanée* de X et Y . Notons par A et B deux ensembles de nombres réels. En définissant $C = \{(x, y) : x \in A, y \in B\}$, on obtient grâce à (6.3)

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x, y) \, dx \, dy$$

Comme

$$\begin{aligned} F(a, b) &= P\{X \in (-\infty, a], Y \in (-\infty, b]\} \\ &= \int_{-\infty}^b \int_{-\infty}^a f(x, y) \, dx \, dy \end{aligned} \quad (6.4)$$

il suffit de dériver pour obtenir

$$f(a, b) = \frac{\partial^2}{\partial a \partial b} F(a, b)$$

pour autant que les dérivées partielles soient définies. Au-delà de cette propriété on peut donner une interprétation intuitive à une densité conjointe en partant de (6.4) et grâce au calcul suivant:

$$\begin{aligned} P\{a < X < a + da, b < Y < b + db\} &= \int_b^{b+db} \int_a^{a+da} f(x, y) \, dx \, dy \\ &\approx f(a, b) \, da \, db \end{aligned}$$

à condition que da et db soient petits et que f soit continue au point (a,b) . Aussi $f(a,b)$ est-elle une indication de la probabilité avec laquelle (X, Y) sera dans le voisinage du point (a, b) .

Enfin, si X et Y sont des variables aléatoires conjointement continues, alors elles sont individuellement continues, également. On obtient leurs densités marginales ainsi:

$$\begin{aligned} P\{X \in A\} &= P\{X \in A, Y \in (-\infty, \infty)\} \\ &= \int_A \int_{-\infty}^{\infty} f(x, y) dy dx \\ &= \int_A f_X(x) dx \end{aligned}$$

où

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

Cette dernière fonction est clairement identifiable à la densité de X . On obtient de même l'expression de la densité de Y :

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

6.1.7 Exemples d'applications des densités conjointes

Exemple 6.3 La densité conjointe de X et Y est donnée par

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{sinon} \end{cases}$$

On veut calculer:

- $P\{X > 1, Y < 1\}$,
- $P\{X < Y\}$ et
- $P\{X < a\}$.

SOLUTION.

$$\begin{aligned}
 \text{a)} \quad P\{X > 1, Y < 1\} &= \int_0^1 \int_1^\infty 2e^{-x}e^{-2y} dx dy \\
 &= \int_0^1 2e^{-2y} \left(-e^{-x} \Big|_1^\infty \right) dy \\
 &= e^{-1} \int_0^1 2e^{-2y} dy \\
 &= e^{-1}(1 - e^{-2})
 \end{aligned}$$

$$\begin{aligned}
 \text{b)} \quad P\{X < Y\} &= \iint_{(x,y): x < y} 2e^{-x}e^{-2y} dx dy \\
 &= \int_0^\infty \int_0^y 2e^{-x}e^{-2y} dx dy \\
 &= \int_0^\infty 2e^{-2y}(1 - e^{-y}) dy \\
 &= \int_0^\infty 2e^{-2y} dy - \int_0^\infty 2e^{-3y} dy \\
 &= 1 - \frac{2}{3} \\
 &= \frac{1}{3}
 \end{aligned}$$

$$\begin{aligned}
 \text{c)} \quad P\{X < a\} &= \int_0^a \int_0^\infty 2e^{-2y}e^{-x} dy dx \\
 &= \int_0^a e^{-x} dx \\
 &= 1 - e^{-a}
 \end{aligned}$$

■

Exemple 6.4 On considère un cercle de rayon R et l'on y choisit un point au hasard, ce qui signifie que toutes les régions de taille donnée incluses dans le cercle ont la même probabilité de contenir ce point. En d'autres termes, la distribution de ce point dans le cercle est uniforme. On admet que le centre du cercle est à l'origine des axes de coordonnées. Les variables aléatoires X et Y représentent les coordonnées du point choisi (figure 6.3). La répartition conjointe de X et Y étant uniforme, il doit exister une constante c telle que

$$f(x, y) = \begin{cases} c & \text{si } x^2 + y^2 \leq R^2 \\ 0 & \text{si } x^2 + y^2 > R^2 \end{cases}$$

On veut

- trouver c ,
- trouver les densités marginales de X et Y et
- calculer la probabilité que la distance du point choisi au centre soit inférieure ou égale à a .

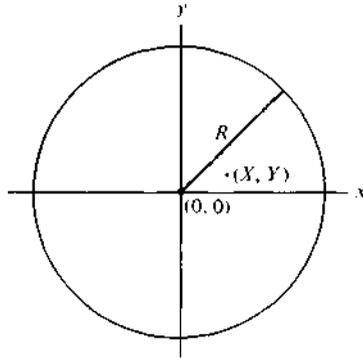


Figure 6.3 Densité conjointe

SOLUTION.

- Du fait que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx = 1$$

il suit que

$$c \iint_{x^2 + y^2 \leq R^2} dy dx = 1$$

Cette dernière intégrale peut être évaluée en passant à un système de coordonnées polaires ou plus simplement en utilisant le fait qu'elle représente l'aire du cercle de rayon R . Elle vaut donc πR^2 et par conséquent

$$c = \frac{1}{\pi R^2}$$

-

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\ &= \frac{1}{\pi R^2} \int_{x^2 + y^2 \leq R^2} dy \\ &= \frac{1}{\pi R^2} \int_{-c}^c dy \quad c = \sqrt{R^2 - x^2} \\ &= \frac{2}{\pi R^2} \sqrt{R^2 - x^2} \quad x^2 \leq R^2 \end{aligned}$$

Lorsque $x^2 > R^2$, cette densité est nulle. La densité marginale de Y est donnée pour des raisons de symétrie par

$$f_Y(y) = \begin{cases} \frac{2}{\pi R^2} \sqrt{R^2 - y^2} & y^2 \leq R^2 \\ 0 & y^2 > R^2 \end{cases}$$

c) La fonction de répartition de $Z = \sqrt{X^2 + Y^2}$, distance du point choisi à l'origine, peut être calculée comme suit: pour $0 \leq a \leq R$,

$$\begin{aligned} F_Z(a) &= P\{\sqrt{X^2 + Y^2} \leq a\} \\ &= P\{X^2 + Y^2 \leq a^2\} \\ &= \iint_{x^2 + y^2 \leq a^2} f(x, y) \, dy \, dx \\ &= \frac{1}{\pi R^2} \iint_{x^2 + y^2 \leq a^2} dy \, dx \\ &= \frac{\pi a^2}{\pi R^2} \\ &= \frac{a^2}{R^2} \end{aligned}$$

où il a de nouveau été fait usage de la formule de l'aire d'un cercle de rayon a pour calculer la dernière intégrale. ■

Exemple 6.5 La densité conjointe de X et Y est

$$f(x, y) = \begin{cases} e^{-(x+y)} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{sinon} \end{cases}$$

On veut calculer la densité de la variable aléatoire X/Y .

SOLUTION. Cherchons d'abord la fonction de répartition de X/Y . Pour $a > 0$,

$$\begin{aligned} F_{X/Y}(a) &= P\left\{\frac{X}{Y} \leq a\right\} \\ &= \iint_{x/y \leq a} e^{-(x+y)} \, dx \, dy \\ &= \int_0^\infty \int_0^{ay} e^{-(x+y)} \, dx \, dy \end{aligned}$$

$$\begin{aligned}
 &= \int_0^{\infty} (1 - e^{-ay}) e^{-y} dy \\
 &= \left[-e^{-y} + \frac{e^{-(a+1)y}}{a+1} \right] \Big|_0^{\infty} \\
 &= 1 - \frac{1}{a+1}
 \end{aligned}$$

On obtient la densité de X/Y par dérivation de ce dernier résultat, ce qui donne $f_{X/Y}(a) = 1/(a+1)^2$, $0 < a < \infty$. ■

6.1.8 Distribution conjointe de plusieurs variables

On peut définir les éléments de la distribution conjointe de n variables en suivant la même démarche que celle utilisée dans le cas $n = 2$. Par exemple, la *fonction de répartition conjointe F de n variables* X_1, X_2, \dots, X_n a pour définition

$$F(a_1, a_2, \dots, a_n) = P\{X_1 \leq a_1, X_2 \leq a_2, \dots, X_n \leq a_n\}$$

Par ailleurs, ces n variables seront dites *conjointement continues*, s'il existe une fonction f de n arguments, appelée *densité conjointe de ces variables*, telle que pour tout sous-ensemble C de l'espace à n dimensions

$$P\{(X_1, X_2, \dots, X_n) \in C\} = \iiint_{(x_1, \dots, x_n) \in C} f(x_1, \dots, x_n) dx_1 dx_2 \cdots dx_n$$

On pourra écrire en particulier que pour toute collection de n sous-ensembles A_1, A_2, \dots, A_n de l'ensemble des nombres réels

$$\begin{aligned}
 &P\{X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n\} \\
 &= \int_{A_n} \int_{A_{n-1}} \cdots \int_{A_1} f(x_1, \dots, x_n) dx_1 dx_2 \cdots dx_n
 \end{aligned}$$

Exemple 6.6 Distribution multinomiale

L'une des distributions conjointes les plus importantes est la distribution multinomiale. Celle-ci modélise l'expérience consistant à répéter n fois indépendamment une épreuve admettant r issues différentes, de probabilités respectives p_1, p_2, \dots, p_r , telles que $\sum_{i=1}^r p_i = 1$. Désignons par X_i le nombre d'épreuves ayant abouti au résultat de type i dans la série complète des n réalisations. On aura

$$P\{X_1 = n_1, X_2 = n_2, \dots, X_r = n_r\} = \frac{n!}{n_1! n_2! \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r} \quad (6.5)$$

pour tous les choix tels que $\sum_{i=1}^r n_i = n$.

On peut vérifier (6.5) de la manière suivante: toute séquence de n résultats composée de n_i occurrences du type i , $i = 1, 2, \dots, r$, aura pour probabilité $p_1^{n_1} p_2^{n_2} \dots p_r^{n_r}$ en raison de l'indépendance des épreuves successives. Or il y a $n!/n_1!n_2!\dots n_r!$ séquences de ce genre puisque le total des $n!$ permutations possibles doit être divisé par le nombre $n_1!$ des permutations des occurrences du type 1, indistinguables entre elles, puis par $n_2!$ pour les mêmes raisons et ainsi de suite. Ceci établit (6.5).

On appelle *distribution multinomiale* une distribution dont la loi de probabilité conjointe est donnée par (6.5). Le lecteur constatera que, pour $n = 2$, cette distribution n'est autre que la distribution binomiale.

A titre d'application de cette distribution multinomiale, prenons le cas d'un dé équilibré que l'on jette 9 fois. La probabilité que 1 apparaisse trois fois, que 2 et 3 apparaissent deux fois chacun, que 4 et 5 n'apparaissent qu'une fois et 6 jamais sera

$$\frac{9!}{3!2!2!1!1!0!} \left(\frac{1}{6}\right)^3 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^0 = \frac{9!}{3!2!2!} \left(\frac{1}{6}\right)^9$$

6.2 VARIABLES ALÉATOIRES INDÉPENDANTES

6.2.1 Définition, critères d'indépendance

Deux variables aléatoires X et Y sont dites *indépendantes* si, pour tout choix d'une paire d'ensembles A et B de nombres réels, on a

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\} \quad (6.6)$$

En d'autres termes, X et Y sont indépendantes si, quels que soient A et B , les événements $E_A = \{X \in A\}$ et $F_B = \{Y \in B\}$ sont indépendants.

On peut montrer, en s'appuyant sur les trois axiomes de la théorie des probabilités, que (6.6) est vraie si et seulement si pour tout couple a, b de réels

$$P\{X \leq a, Y \leq b\} = P\{X \leq a\}P\{Y \leq b\}$$

ce qui revient à écrire que X et Y sont indépendantes si pour tout couple a, b

$$F(a, b) = F_X(a)F_Y(b)$$

Lorsque X et Y sont discrètes, la condition (6.6) est équivalente à

$$p(x, y) = p_X(x)p_Y(y) \quad (6.7)$$

pour tout x et tout y . L'équivalence résulte d'une part du fait qu'en choisissant $A = \{x\}$ et $B = \{y\}$ dans (6.6), on obtient (6.7); d'autre part du fait qu'en supposant (6.7) vraie, on aura pour toute paire d'ensembles A et B

$$\begin{aligned}
 P\{X \in A, Y \in B\} &= \sum_{y \in B} \sum_{x \in A} p(x, y) \\
 &= \sum_{y \in B} \sum_{x \in A} p_X(x) p_Y(y) \\
 &= \sum_{y \in B} p_Y(y) \sum_{x \in A} p_X(x) \\
 &= P\{Y \in B\} P\{X \in A\}
 \end{aligned}$$

ce qui établit que (6.6) est un critère d'indépendance.

Lorsque X et Y sont des variables conjointement continues, le critère d'indépendance sera

$$f(x, y) = f_X(x) f_Y(y)$$

pour tout x et tout y .

Intuitivement parlant, on pourra donc dire que X et Y sont indépendantes si le fait de connaître la valeur de l'une n'influe pas sur la distribution de l'autre. Des variables qui ne sont pas indépendantes sont dites *dépendantes*.

6.2.2 Exemples de variables aléatoires indépendantes

Exemple 6.7 On réalise $n + m$ épreuves indépendantes ayant chacune p pour probabilité de succès. La variable X est le nombre de succès lors des n premières épreuves, Y étant le nombre de succès lors des m dernières. X et Y sont alors indépendantes puisque le fait de connaître le nombre de succès lors des n premières épreuves n'influe en rien sur celui des succès lors des m dernières (c'est là la traduction de l'indépendance des épreuves). On peut d'ailleurs écrire, pour toute paire d'entiers x et y

$$\begin{aligned}
 P\{X = x, Y = y\} &= \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{y} p^y (1-p)^{m-y}, & 0 \leq x \leq n \\
 & & 0 \leq y \leq m \\
 &= P\{X = x\} P\{Y = y\}
 \end{aligned}$$

Par contre X et Z sont des variables dépendantes si Z représente le nombre total de succès au cours des $n + m$ épreuves (pourquoi cela?). ■

Exemple 6.8 On admet que le nombre de clients d'un bureau de poste en l'espace d'un jour est une variable aléatoire poissonnienne de paramètre λ . On note par p la probabilité qu'une personne pénétrant dans ce bureau de poste soit un homme. On veut montrer que, dans ce cas, le nombre des hommes et celui des femmes parmi les clients quotidiens sont des variables aléatoires poissonniennes de paramètres respectifs λp et $\lambda(1 - p)$ et qu'elles sont indépendantes.

SOLUTION. Désignons par X et Y respectivement le nombre de clients masculins et féminins de ce bureau de poste. Ces variables X et Y seront indépendantes si (6.7) peut

être vérifiée. Pour obtenir une expression de $P\{X = i, Y = j\}$, on peut conditionner selon les valeurs prises par $X + Y$ de la manière suivante:

$$P\{X = i, Y = j\} = P\{X = i, Y = j | X + Y = i + j\}P\{X + Y = i + j\} \\ + P\{X = i, Y = j | X + Y \neq i + j\}P\{X + Y \neq i + j\}$$

[Le lecteur remarquera que cette équation n'est qu'une illustration de la formule $P(E) = P(E|F) \cdot P(F) + P(E|F^c) \cdot P(F^c)$. Comme manifestement $P\{X = i, Y = j | X + Y \neq i + j\} = 0$, il reste

$$P\{X = i, Y = j\} = P\{X = i, Y = j | X + Y = i + j\}P\{X + Y = i + j\} \quad (6.8)$$

Comme maintenant $X + Y$ n'est autre que le nombre total des clients, on peut écrire par hypothèse que

$$P\{X + Y = i + j\} = e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} \quad (6.9)$$

Par ailleurs, si l'on sait que $i + j$ personnes sont venues au bureau de poste, la probabilité que i d'entre elles soient des hommes et j des femmes n'est autre que la probabilité binomiale $\binom{i+j}{i} p^i (1-p)^j$. Donc,

$$P\{X = i, Y = j | X + Y = i + j\} = \binom{i+j}{i} p^i (1-p)^j \quad (6.10)$$

La substitution des membres appropriés de (6.9) et (6.10) dans (6.8) livre

$$P\{X = i, Y = j\} = \binom{i+j}{i} p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} \\ = e^{-\lambda} \frac{(\lambda p)^i}{i! j!} [\lambda(1-p)]^j \\ = \frac{e^{-\lambda p} (\lambda p)^i}{i!} e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^j}{j!} \quad (6.11)$$

Donc

$$P\{X = i\} = e^{-\lambda p} \frac{(\lambda p)^i}{i!} \sum_j e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^j}{j!} = e^{-\lambda p} \frac{(\lambda p)^i}{i!} \quad (6.12)$$

et de façon analogue

$$P\{Y = j\} = e^{-\lambda(1-p)} \frac{[\lambda(1-p)]^j}{j!} \quad (6.13)$$

Finalement (6.11), (6.12) et (6.13) établissent les résultats attendus. ■

Exemple 6.9 Un homme et une femme se donnent rendez-vous. L'heure d'arrivée de chacune de ces deux personnes sur les lieux du rendez-vous est une variable aléatoire uniforme entre midi et une heure. Ces deux variables sont indépendantes. Quelle est la probabilité que la première arrivée doive attendre plus de 10 minutes?

SOLUTION. Désignons par X et Y respectivement l'écart (en minutes) entre midi et l'arrivée de l'homme ou de la femme. X et Y sont alors des variables aléatoires indépendantes et uniformément distribuées dans l'intervalle $(0, 60)$. La probabilité cherchée, à savoir $P\{X + 10 < Y\} + P\{Y + 10 < X\}$, est par symétrie égale à $2P\{X + 10 < Y\}$. On l'obtient ainsi:

$$\begin{aligned} 2P\{X + 10 < Y\} &= 2 \cdot \iint_{x+10 < y} f(x, y) \, dx \, dy \\ &= 2 \iint_{x+10 < y} f_X(x)f_Y(y) \, dx \, dy \\ &= 2 \int_{10}^{60} \int_0^{y-10} \left(\frac{1}{60}\right)^2 \, dx \, dy \\ &= \frac{2}{(60)^2} \int_{10}^{60} (y - 10) \, dy \\ &= \frac{25}{36} \end{aligned}$$

■

L'exemple qui suit est le plus ancien des problèmes traitant de probabilité en relation avec des modèles géométriques. Le naturaliste français Buffon fut le premier au dix-huitième siècle à s'intéresser à ce problème et il lui donna une solution. Aussi ce problème est-il dit «de l'aiguille de Buffon».

Exemple 6.10 Problème de l'aiguille de Buffon.

Sur une table on trace des lignes parallèles espacées d'un écart D les unes des autres. On y jette une aiguille de longueur L , avec $L \leq D$. Quelle est la probabilité que l'aiguille coupe une ligne (l'alternative étant que l'aiguille soit complètement située dans une des bandes délimitées par les lignes)?

SOLUTION. On repérera la position de l'aiguille grâce à la distance X entre le milieu de celle-ci et la parallèle la plus proche, et grâce à l'angle θ entre l'aiguille et une perpendiculaire aux lignes (voir figure 6.4). L'aiguille chevauchera une parallèle si l'hypoténuse du triangle rectangle de la figure 6.4 est de longueur inférieure à $L/2$, c'est-à-dire si

$$\frac{X}{\cos \theta} < \frac{L}{2} \quad \text{ou} \quad X < \frac{L}{2} \cos \theta$$

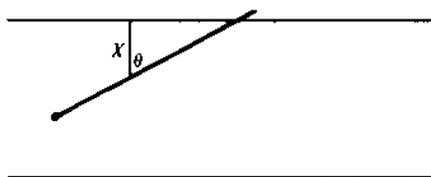


Figure 6.4

La variable X varie entre 0 et $D/2$ tandis que θ varie entre 0 et $\pi/2$. Il est raisonnable d'admettre que, dans ces limites, X et θ sont de distributions indépendantes et uniformes. Aussi aura-t-on

$$\begin{aligned}
 P\left\{X < \frac{L}{2} \cos \theta\right\} &= \iint_{x < L/2 \cos y} f_X(x) f_\theta(y) dx dy \\
 &= \frac{4}{\pi D} \int_0^{\pi/2} \int_0^{L/2 \cos y} dx dy \\
 &= \frac{4}{\pi D} \int_0^{\pi/2} \frac{L}{2} \cos y dy \\
 &= \frac{2L}{\pi D}
 \end{aligned}$$

■

Exemple 6.11 Caractérisation d'une distribution normale.

Appelons X et Y l'écart horizontal, respectivement vertical, entre le point d'impact d'une balle et le centre de la cible. On admettra que

- X et Y sont des variables indépendantes continues de densités dérivables
- la densité conjointe de X et Y , qui vérifie $f(x, y) = f_X(x) \cdot f_Y(y)$ pour tout couple (x, y) , ne dépend des valeurs x et y qu'à travers la fonction $\sqrt{x^2 + y^2}$.

En gros, la seconde hypothèse indique que la probabilité d'impact sur une zone minuscule entourant un point donné de la cible ne dépend que de la distance entre ce point et le centre de la cible mais pas de l'orientation de ce point. On peut encore exprimer cette propriété en disant que la densité conjointe est invariante par rotation.

Ces deux hypothèses entraînent une conséquence tout à fait remarquable: X et Y sont normalement distribuées. La démonstration s'appuie d'abord sur une conséquence directe des hypothèses, à savoir qu'il existe une fonction g telle que

$$f(x, y) = f_X(x) f_Y(y) = g(x^2 + y^2) \quad (6.14)$$

La dérivation des deux membres de (6.14) par rapport à x livre

$$f'_X(x) f_Y(y) = 2xg'(x^2 + y^2) \quad (6.15)$$

On divise ensuite (6.15) par (6.14) membre à membre

$$\frac{f'_X(x)}{f_X(x)} = \frac{2xg'(x^2 + y^2)}{g(x^2 + y^2)}$$

ou

$$\frac{f'_X(x)}{2xf_X(x)} = \frac{g'(x^2 + y^2)}{g(x^2 + y^2)} \quad (6.16)$$

Dans (6.16), le membre de gauche est constant car il ne dépend que de x , tandis que celui de droite dépend de $x^2 + y^2$; ceci permet en effet d'écrire, en choisissant pour tout couple x_1, x_2 deux valeurs y_1, y_2 telles que $x_1^2 + y_1^2 = x_2^2 + y_2^2$, et en utilisant (6.16):

$$\frac{f'_X(x_1)}{2x_1 f_X(x_1)} = \frac{g'(x_1^2 + y_1^2)}{g(x_1^2 + y_1^2)} = \frac{g'(x_2^2 + y_2^2)}{g(x_2^2 + y_2^2)} = \frac{f'_X(x_2)}{2x_2 f_X(x_2)}$$

Ce membre de gauche étant constant, on peut écrire

$$\frac{f'_X(x)}{xf_X(x)} = c \quad \text{ou} \quad \frac{d}{dx}(\ln f_X(x)) = cx$$

ce qui donne, après intégration des deux membres,

$$\ln f_X(x) = a + \frac{cx^2}{2} \quad \text{ou} \quad f_X(x) = ke^{cx^2/2}$$

Par ailleurs, f_X vérifie $\int_{-\infty}^{+\infty} f_X(x) dx = 1$. La constante c sera donc nécessairement négative et nous l'écrivons $c = -1/\sigma^2$. Par conséquent

$$f_X(x) = ke^{-x^2/2\sigma^2}$$

En conclusion, X est donc une variable aléatoire normale de paramètres $\mu = 0$ et σ^2 . Grâce à un raisonnement tout à fait similaire, on peut établir que

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-y^2/2\sigma^2}$$

La seconde hypothèse du problème entraîne par ailleurs que $\sigma^2 = \bar{\sigma}^2$. Les variables X et Y sont donc identiquement distribuées, en plus d'être indépendantes et normales de paramètres $\mu = 0$ et σ^2 . ■

6.2.3 Indépendance de plus de deux variables aléatoires

On peut évidemment étendre la notion d'indépendance de variables aléatoires à plus de deux variables. Les n variables X_1, X_2, \dots, X_n seront dites indépendantes si, pour tout choix de n ensembles de nombres réels A_1, A_2, \dots, A_n ,

$$P\{X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n\} = \prod_{i=1}^n P\{X_i \in A_i\}$$

On peut comme plus haut montrer que cette dernière égalité est équivalente à

$$\begin{aligned} P\{X_1 \leq a_1, X_2 \leq a_2, \dots, X_n \leq a_n\} \\ = \prod_{i=1}^n P\{X_i \leq a_i\} \text{ pour tous } a_1, a_2, \dots, a_n \end{aligned}$$

Une collection infinie de variables aléatoires est indépendante si tout sous-ensemble fini que l'on puisse en tirer est composé de variables indépendantes.

6.2.4 Exemples d'indépendance de plusieurs variables

Exemple 6.12 Comment faire engendrer par un ordinateur un sous-ensemble de composition aléatoire? La plupart des ordinateurs sont capables d'engendrer par simulation des «nombres aléatoires» qui sont, avec un certain degré d'approximation, de distribution uniforme dans $(0, 1)$. Il est facile de construire à partir de là un simulateur de variable aléatoire indicatrice, plus précisément, dans notre cas, un générateur de variable de Bernoulli; supposons que I soit une variable aléatoire indicatrice qui doit vérifier

$$P\{I = 1\} = p = 1 - P\{I = 0\}$$

L'ordinateur peut simuler I en tirant un nombre aléatoire U de distribution uniforme dans $(0, 1)$ puis en posant

$$I = \begin{cases} 1 & \text{si } U \leq p \\ 0 & \text{si } U > p \end{cases}$$

Admettons maintenant que nous nous intéressions à la sélection aléatoire d'un sous-ensemble de taille k , $k \leq n$, de l'ensemble $\{1, 2, \dots, n\}$ de telle manière que chacun des $\binom{n}{k}$ résultats possibles soit équiprobable aux autres. La méthode qui est exposée ci-après permet d'effectuer ce tirage. Pour ce, commençons par simuler le tirage séquentiel de n variables indicatrices I_1, I_2, \dots, I_n de telle manière qu'exactly k de ces indicatrices aient 1 pour valeur. Les indices i pour lesquels $I_i = 1$ seront les éléments du sous-ensemble à construire.

Pour engendrer les variables aléatoires I_1, I_2, \dots, I_n , commençons par simuler le tirage de n variables indépendantes uniformes sur $(0, 1)$, notées U_1, U_2, \dots, U_n . On définit ensuite

$$I_1 = \begin{cases} 1 & \text{si } U_1 < \frac{k}{n} \\ 0 & \text{sinon} \end{cases}$$

et subséquemment on pose de manière réursive

$$I_{i+1} = \begin{cases} 1 & \text{si } U_{i+1} < \frac{k - (I_1 + \dots + I_i)}{n - i} \\ 0 & \text{sinon} \end{cases}$$

En d'autres termes, on attribuera à I_{i+1} la valeur 1 avec une probabilité égale au nombre de places qu'il faut encore pourvoir dans le sous-ensemble à tirer (à savoir $k - \sum_{j=1}^i I_j$) divisé par le nombre restant $n - i$ d'occasions qu'il reste de pourvoir ces places. (On rappelle ici que si $I_{i+1} = 1$, le nombre $i + 1$ sera inclus dans le sous-ensemble à tirer). La loi conjointe des variables I_1, I_2, \dots, I_n est donc donnée par

$$P\{I_1 = 1\} = \frac{k}{n}$$

$$P\{I_{i+1} = 1 | I_1, \dots, I_i\} = \frac{k - \sum_{j=1}^i I_j}{n - i} \quad 1 < i < n$$

La preuve que les sous-ensembles ainsi choisis sont tous équiprobables peut être faite par induction sur $k + n$. Elle est immédiate lorsque $k + n = 2$, donc lorsque $k = 1$ et $n = 1$. Supposons donc qu'elle soit établie pour $k + n \leq l$. Montrons qu'elle le sera pour $k + n = l + 1$. On considère un sous-ensemble de taille k quelconque, $i_1 \leq i_2 \dots \leq i_k$ par exemple, et distingue les deux cas suivants:

- Cas 1: $i_1 = 1$.

$$P\{I_1 = I_2 = \dots = I_{i_k} = 1, I_j = 0 \text{ ailleurs}\}$$

$$= P\{I_1 = 1\}P\{I_2 = \dots = I_{i_k} = 1, I_j = 0 \text{ ailleurs} | I_1 = 1\}$$

Etant donné que $I_1 = 1$, le reste des éléments du sous-ensemble sera choisi comme s'il fallait tirer un sous-ensemble de taille $k - 1$ parmi $n - 1$ éléments, nommément les nombres $2, 3, \dots, n$. En vertu de l'hypothèse d'induction, la probabilité conditionnelle d'obtenir un sous-ensemble bien déterminé de taille $k - 1$ sera $1/\binom{n-1}{k-1}$. Aussi

$$P\{I_1 = I_2 = \dots = I_{i_k} = 1, I_j = 0 \text{ ailleurs}\}$$

$$= \frac{k}{n} \frac{1}{\binom{n-1}{k-1}} = \frac{1}{\binom{n}{k}}$$

- Cas 2: $i_1 \neq 1$.

$$\begin{aligned} & P\{I_{i_1} = I_{i_2} = \dots = I_{i_k} = 1, I_j = 0 \text{ ailleurs}\} \\ &= P\{I_{i_1} = \dots = I_{i_k} = 1, I_j = 0 \text{ ailleurs} \mid I_1 = 0\} P\{I_1 = 0\} \\ &= \frac{1}{\binom{n-1}{k}} \left(1 - \frac{k}{n}\right) = \frac{1}{\binom{n}{k}} \end{aligned}$$

où nous avons fait usage de l'hypothèse d'induction lors de l'évaluation de la probabilité conditionnelle que nous avons fait apparaître.

Aussi a-t-on l'assurance que tout sous-ensemble de taille k donné apparaîtra avec probabilité $1/\binom{n}{k}$. ■

REMARQUE. Cette méthode pour générer un sous-ensemble aléatoire nécessite une très faible quantité de mémoire. Un algorithme plus rapide mais qui requiert plus de mémoire est présenté dans la section 10.1.2 du chapitre 10. (Cette dernière méthode utilise les k derniers éléments d'une permutation aléatoire de $1, 2, \dots, n$.)

Exemple 6.13 Admettons que X , Y et Z sont trois variables aléatoires indépendantes et uniformément réparties sur $(0, 1)$. On souhaite calculer $P\{X \geq YZ\}$.

SOLUTION. On sait que

$$f_{X,Y,Z}(x, y, z) = f_X(x)f_Y(y)f_Z(z) = 1 \quad 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1$$

on a

$$\begin{aligned} P\{X \geq YZ\} &= \iiint_{x \geq yz} f_{X,Y,Z}(x, y, z) \, dx \, dy \, dz \\ &= \int_0^1 \int_0^1 \int_{yz}^1 dx \, dy \, dz \\ &= \int_0^1 \int_0^1 (1 - yz) \, dy \, dz \\ &= \int_0^1 \left(1 - \frac{z}{2}\right) dz \\ &= \frac{3}{4} \end{aligned}$$

■

Exemple 6.14 Une interprétation probabiliste de la demi-vie.

Soit $N(t)$ le nombre de noyaux contenus dans une masse radioactive d'un matériau, au temps t . Le concept de demi-vie est souvent défini de manière déterministe; en effet, c'est l'expérience qui a permis d'établir que pour une valeur h appelée la demi-vie

$$N(t) = 2^{-t/h} N(0), \quad t > 0.$$

(Remarquer que $N(h) = N(0)/2$.) Puisque l'égalité ci-dessus implique que pour tous s et t non négatifs

$$N(t + s) = 2^{-(s+t)/h} N(0) = 2^{-t/h} N(s)$$

il s'ensuit que, indépendamment du temps s écoulé, pendant un temps additionnel t , le nombre de noyaux restants sera diminué d'un facteur $2^{-t/h}$.

Comme la relation déterministe précédente résulte d'observations de masses radioactives contenant une grande quantité de noyaux, elle pourrait bien être consistante avec une interprétation probabiliste. La clé pour la déduction du modèle probabiliste approprié pour la demi-vie réside dans l'observation empirique suivante: la proportion de désintégration dans un intervalle de temps donné ne dépend ni du nombre total de noyaux au début de l'intervalle ni de la location de cet intervalle (vu que $N(t+s)/N(s)$ ne dépend ni de $N(s)$ ni de s). Par conséquent, il apparaît que chaque noyau individuel réagit indépendamment des autres et selon une distribution de durée de vie sans mémoire. Etant donné que l'unique distribution sans mémoire est la distribution exponentielle et qu'exactement la moitié de la quantité de masse donnée disparaît toutes les h unités de temps, on propose le modèle probabiliste suivant pour la désintégration radioactive.

Interprétation probabiliste de la demi-vie h : Les durées de vie des noyaux individuels sont des variables aléatoires indépendantes de distribution exponentielle dont la médiane est égale à h . En d'autres termes, si L représente la durée de vie d'un noyau donné, alors

$$P\{L < t\} = 1 - 2^{-t/h}$$

(Comme $P\{L < h\} = 1/2$ et que l'égalité précédente peut être écrite sous la forme $P\{L < t\} = 1 - \exp\left\{-t \frac{\ln 2}{h}\right\}$, on voit que L suit effectivement une distribution exponentielle de médiane h .)

On notera qu'avec cette interprétation probabiliste de la demi-vie, si l'on compte $N(0)$ noyaux au temps 0, alors $N(t)$, le nombre de noyaux restants au temps t , suivra une loi binomiale de paramètres $n = N(0)$ et $p = 2^{-t/h}$. Des résultats du chapitre 8 montreront que cette interprétation de la demi-vie est consistante avec le modèle déterministe lorsque l'on considère la proportion d'un grand nombre de noyaux qui se désintègrent pendant un laps de temps donné. Cependant, la différence entre l'interprétation déterministe et probabiliste devient apparente au moment où l'on considère le nombre actuel de noyaux désintégrés. Nous allons maintenant mentionner ce fait par rapport à la question relative à la désintégration des protons.

Il y a une controverse sur la désintégration ou non des protons. Une théorie prévoit la disparition des protons avec une demi-vie d'environ $h = 10^{30}$ années. Une vérification empirique consiste à suivre un grand nombre de protons pendant, disons, 1 ou 2 ans, et à déterminer s'il y a une diminution pendant cette période. (Il est clair qu'il n'est pas possible de suivre une masse de protons pendant 10^{30} années pour vérifier si la moitié d'entre eux disparaît.) Supposons que l'on puisse garder trace de 10^{30} protons pendant c années. Le nombre de désintégration prédit par le modèle déterministe serait donné par

$$\begin{aligned}
 N(0) - N(c) &= h(1 - 2^{-c/h}) \\
 &= \frac{1 - 2^{-c/h}}{1/h} \\
 &\approx \lim_{x \rightarrow 0} (1 - 2^{-cx})/x && \text{puisque } 1/h = 10^{-30} \approx 0 \\
 &= \lim_{x \rightarrow 0} (c2^{-cx} \ln 2) && \text{d'après la règle de l'Hôpital} \\
 &= c \ln 2 \approx .6931c
 \end{aligned}$$

Par exemple, le modèle déterministe prévoit 1,3863 disparus en 2 ans, et l'on entre donc sérieusement en conflit avec l'hypothèse que les protons meurent avec une demi-vie de 10^{30} années si aucune désintégration n'a été observée pendant ces 2 ans.

Comparons maintenant ces conclusions avec celles obtenues à partir du modèle probabiliste. Supposons de nouveau que la demi-vie des protons est $h = 10^{30}$ ans et suivons h protons pendant c années. Comme il y a un grand nombre de protons indépendants, chacun ayant une très petite probabilité de mourir pendant cette période, cela implique que le nombre de désintégration aura (avec une très forte approximation) une distribution de Poisson de paramètre égal à $h(1 - 2^{-c/h}) \approx c \ln 2$. Ainsi

$$\begin{aligned}
 P\{0 \text{ perte}\} &= e^{-c \ln 2} \\
 &= e^{-\ln(2^c)} = 1/2^c
 \end{aligned}$$

et, de façon générale,

$$P\{n \text{ pertes}\} = \frac{2^{-c} [c \ln 2]^n}{n!}, \quad n \geq 0$$

Ainsi, bien que le nombre moyen de disparitions sur 2 ans soit (selon le modèle déterministe) 1,3863, il y a une chance sur 4 qu'aucun proton ne meure, indiquant par là qu'un tel résultat ne peut en aucune façon valider l'hypothèse originale sur la désintégration des protons. ■

6.3 SOMMES DE VARIABLES ALÉATOIRES INDÉPENDANTES

6.3.1 Convolution

Il est très souvent nécessaire de déterminer la distribution de la somme $X + Y$ de deux variables aléatoires X et Y indépendantes en se basant sur leurs distributions marginales. Supposons que ces distributions soient données par les densités f_X et f_Y . Le calcul suivant résout le problème

$$\begin{aligned}
 F_{X+Y}(a) &= P\{X + Y \leq a\} \\
 &= \iint_{x+y \leq a} f_X(x) f_Y(y) \, dx \, dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) f_Y(y) \, dx \, dy \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) \, dx f_Y(y) \, dy \\
 &= \int_{-\infty}^{\infty} F_X(a-y) f_Y(y) \, dy
 \end{aligned} \tag{6.17}$$

La fonction obtenue dans (6.17), ici fonction de répartition de $X + Y$, est appelée *convolution* des fonctions F_X et f_Y . La densité f_{X+Y} d'une somme est obtenue par convolution également, tout comme sa fonction de répartition. Il suffit en effet de dériver (6.17) pour obtenir f_{X+Y} :

$$\begin{aligned}
 f_{X+Y}(a) &= \frac{d}{da} \int_{-\infty}^{\infty} F_X(a-y) f_Y(y) \, dy \\
 &= \int_{-\infty}^{\infty} \frac{d}{da} F_X(a-y) f_Y(y) \, dy \\
 &= \int_{-\infty}^{\infty} f_X(a-y) f_Y(y) \, dy
 \end{aligned} \tag{6.18}$$

qui est bien une convolution, celle de f_X et f_Y .

Exemple 6.15 Cas de la somme de deux variables aléatoires uniformes indépendantes. Admettons que X et Y soient uniformes sur $(0, 1)$ et indépendantes. Déterminons la densité de $X + Y$.

SOLUTION. On a

$$f_X(a) = f_Y(a) = \begin{cases} 1 & 0 < a < 1 \\ 0 & \text{sinon} \end{cases}$$

En application de (6.18), on peut écrire

$$f_{X+Y}(a) = \int_0^1 f_X(a-y) \, dy$$

Pour $0 \leq a \leq 1$, on obtient

$$f_{X+Y}(a) = \int_0^a dy = a$$

tandis que pour $1 < a < 2$, on aura

$$f_{X+Y}(a) = \int_{a-1}^1 dy = 2 - a$$

Ainsi

$$f_{X+Y}(a) = \begin{cases} a & 0 \leq a \leq 1 \\ 2 - a & 1 < a < 2 \\ 0 & \text{sinon} \end{cases}$$

6.3.2 Additivité de la loi Gamma

On se souvient que la densité d'une variable aléatoire suivant une loi gamma est de la forme

$$f(y) = \frac{\lambda e^{-\lambda y} (\lambda y)^{t-1}}{\Gamma(t)} \quad 0 < y < \infty$$

où (λ, t) est le couple des paramètres spécifiques à cette variable. La famille des distributions gamma possède une importante propriété d'additivité lorsque λ est constant. Enoncé autrement, lorsque λ est constant, la famille est stable pour la convolution; ce qu'explique le théorème suivant:

Théorème 6.1

Si X et Y sont deux variables indépendantes suivant des lois gamma de paramètres respectifs (s, λ) et (t, λ) , $X + Y$ sera également une variable de loi gamma avec paramètres $(s + t, \lambda)$.

DÉMONSTRATION. On utilise (6.18):

$$\begin{aligned} f_{X+Y}(a) &= \frac{1}{\Gamma(s)\Gamma(t)} \int_0^a \lambda e^{-\lambda(a-y)} [\lambda(a-y)]^{s-1} \lambda e^{-\lambda y} (\lambda y)^{t-1} dy \\ &= K e^{-\lambda a} \int_0^a (a-y)^{s-1} y^{t-1} dy \end{aligned}$$

on pose alors $x = y/a$, ce qui donne

$$\begin{aligned} &K e^{-\lambda a} a^{s+t-1} \int_0^1 (1-x)^{s-1} x^{t-1} dx \\ &= C e^{-\lambda a} a^{s+t-1} \end{aligned}$$

où la valeur de la constante C ne dépend pas de a . Cette dernière expression étant une densité, son intégration devra donner 1, ce qui déterminera la valeur de C . On trouve après calculs

$$f_{X+Y}(a) = \frac{\lambda e^{-\lambda a} (\lambda a)^{s+t-1}}{\Gamma(s+t)}$$

ce qui établit le théorème. ■

Il est à partir de là facile de démontrer par induction que si $X_i, i = 1, \dots, n$ sont des variables indépendantes suivant des lois gamma de paramètres respectifs (t_i, λ) , $i = 1, \dots, n$, alors $\sum_{i=1}^n X_i$ suivra aussi une loi gamma de paramètres $(\sum_{i=1}^n t_i, \lambda)$. Ce travail est laissé en exercice.

6.3.3 Première application: somme de variables exponentielles

On considère n variables aléatoires exponentielles indépendantes de même paramètre λ . Une variable exponentielle de paramètre λ étant en fait également une variable de loi gamma $(1, \lambda)$, le théorème 6.1 permet de conclure que la somme $X_1 + X_2 + \dots + X_n$ suit une loi gamma de paramètres (n, λ) .

6.3.4 Deuxième application: densité des lois χ^2

Soient n variables aléatoires Z_1, Z_2, \dots, Z_n normales centrées réduites et indépendantes. La distribution de la variable $Y = \sum_{i=1}^n Z_i^2$ est appelée *distribution chi-carré* à n degrés de liberté, que l'on note parfois distribution χ_n^2 .

Nous allons déterminer sa densité. Pour $n = 1$ et $Y = Z_1^2$, la densité de Y a été calculée à l'occasion de la résolution de l'exemple 5.29 où l'on avait trouvé

$$\begin{aligned} f_{Z^2}(y) &= \frac{1}{2\sqrt{y}} [f_Z(\sqrt{y}) + f_Z(-\sqrt{y})] \\ &= \frac{1}{2\sqrt{y}} 2 \left(\frac{1}{\sqrt{2\pi}} e^{-y/2} \right) \\ &= \frac{\frac{1}{2} e^{-1/2y} (\frac{1}{2}y)^{1/2-1}}{\sqrt{\pi}} \end{aligned}$$

On reconnaîtra ici la densité de la loi gamma de paramètres $(\frac{1}{2}, \frac{1}{2})$. [En comparant l'expression ci-dessus à la densité gamma mentionnée et en se souvenant que ces deux fonctions sont des densités dont l'intégrale vaut nécessairement 1, on obtient le résultat secondaire intéressant $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ et par suite l'identité annoncée]. Puisqu'alors Z_i^2 suit pour tout i une loi gamma $(\frac{1}{2}, \frac{1}{2})$, en vertu du théorème 6.1 la loi χ_n^2 n'est rien d'autre que la loi gamma $(n/2, \frac{1}{2})$. Sa densité est donc

$$\begin{aligned} f_{\chi^2}(y) &= \frac{\frac{1}{2} e^{-y/2} \left(\frac{y}{2}\right)^{n/2-1}}{\Gamma\left(\frac{n}{2}\right)} \quad y > 0 \\ &= \frac{e^{-y/2} y^{n/2-1}}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} \quad y > 0 \end{aligned}$$

Lorsque n est un entier pair, on calcule $\Gamma(n/2)$ en utilisant la relation

$\Gamma(n/2) = [(n/2) - 1]!$. Si n est impair, on utilise autant de fois que nécessaire la relation $\Gamma(t) = (t - 1)\Gamma(t - 1)$ et le résultat précédemment obtenu $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. [A titre d'exemple, $\Gamma(\frac{3}{2}) = (\frac{1}{2})\Gamma(\frac{1}{2}) = (\frac{1}{2})(\frac{1}{2})\Gamma(\frac{1}{2}) = (\frac{1}{4})\sqrt{\pi}$].

Dans la pratique, on rencontre la distribution χ^2 comme étant la répartition du carré de l'erreur obtenue lors d'un tir sur une cible à n dimensions, lorsque les erreurs le long de chaque axe sont de répartition normale standard. Au-delà de cet exemple, la distribution χ^2 joue également un rôle important dans le domaine de la statistique inférentielle.

6.3.5 Additivité de la loi normale

On peut grâce à (6.18) établir également l'important résultat suivant concernant des variables normales:

Théorème 6.2

Soient X_1, X_2, \dots, X_n des variables aléatoires indépendantes normales de paramètres $(\mu_i, \sigma_i^2), i = 1, \dots, n$. La variable $\sum_{i=1}^n X_i$ est alors normale de paramètres $\sum_{i=1}^n \mu_i$ et $\sum_{i=1}^n \sigma_i^2$.

La démonstration de ce théorème est laissée en exercice.

6.3.6 Additivité de lois discrètes: loi de Poisson, loi binomiale

Plutôt que d'essayer d'exhiber une formule générale donnant la distribution d'une somme de variables discrètes, deux cas particuliers seront traités à travers des exemples: celui de la somme de variables poissoniennes puis binomiales.

Exemple 6.16 Somme de deux variables aléatoires de Poisson indépendantes. On veut ici déterminer la loi de probabilité de $X + Y$ où X et Y suivent des lois de Poisson de paramètres λ_1 et λ_2 .

SOLUTION. L'événement $\{X + Y = n\}$ est l'union disjointe des événements $\{X = k, Y = n - k\}$ pour $k = 0, 1, \dots, n$. Donc

$$\begin{aligned} P\{X + Y = n\} &= \sum_{k=0}^n P\{X = k, Y = n - k\} \\ &= \sum_{k=0}^n P\{X = k\}P\{Y = n - k\} \\ &= \sum_{k=0}^n e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} \\ &= e^{-(\lambda_1 + \lambda_2)} \sum_{k=0}^n \frac{\lambda_1^k \lambda_2^{n-k}}{k!(n-k)!} \\ &= \frac{e^{-(\lambda_1 + \lambda_2)}}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k} \\ &= \frac{e^{-(\lambda_1 + \lambda_2)}}{n!} (\lambda_1 + \lambda_2)^n \end{aligned}$$

En d'autres termes, $X + Y$ suit une loi de Poisson de paramètre $\lambda_1 + \lambda_2$. ■

Exemple 6.17 Somme de variables binomiales indépendantes.

On cherche toujours la loi de $X + Y$, les variables X et Y étant ici binomiales de paramètres (n, p) et (m, p) respectivement.

SOLUTION. Sans faire le moindre calcul, on peut déterminer que $X + Y$ suit une loi binomiale de paramètre $(n + m, p)$. Il suffit de revenir à l'interprétation qui présente une variable binomiale X de loi notée $b(n, p)$ comme le nombre de succès obtenus lors de la répétition de n épreuves indépendantes ayant chacune une probabilité p de succès. Si Y compte le nombre de succès de probabilité p lors d'une suite de m épreuves, et si de plus X et Y sont indépendantes, alors $X + Y$ peut représenter le nombre de succès de probabilité p lors d'une suite de $n + m$ épreuves. Il s'agit bien là d'une variable de loi binomiale $b(n + m, p)$. Mais on peut établir ce même résultat de manière analytique:

$$\begin{aligned} P\{X + Y = k\} &= \sum_{i=0}^n P\{X = i, Y = k - i\} \\ &= \sum_{i=0}^n P\{X = i\}P\{Y = k - i\} \\ &= \sum_{i=0}^n \binom{n}{i} p^i q^{n-i} \binom{m}{k-i} p^{k-i} q^{m-k+i} \end{aligned}$$

où $q = 1 - p$ et où $\binom{r}{j} = 0$ lorsque $j > r$. Donc

$$P\{X + Y = k\} = p^k q^{n+m-k} \sum_{i=0}^n \binom{n}{i} \binom{m}{k-i}$$

et le résultat attendu apparaît après application de l'identité combinatoire

$$\binom{n+m}{k} = \sum_{i=0}^n \binom{n}{i} \binom{m}{k-i} \quad \blacksquare$$

6.4 DISTRIBUTIONS CONDITIONNELLES

6.4.1 Cas discret

On se souvient que, pour toute paire d'événements E et F , la probabilité de E sous condition que F soit réalisé est, pour autant que $P(F) > 0$,

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Il est naturel à partir de là de définir la *loi de probabilité de X sous la condition $Y = y$* :

$$\begin{aligned}
 p_{X|Y}(x|y) &= P\{X = x | Y = y\} \\
 &= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\
 &= \frac{p(x, y)}{p_Y(y)}
 \end{aligned}$$

pour tous les cas où $p_Y(y) > 0$. On définit également la *fonction de répartition conditionnelle* de X , sachant que $Y = y$, pour autant que $p_Y(y) > 0$ toujours:

$$\begin{aligned}
 F_{X|Y}(x|y) &= P\{X \leq x | Y = y\} \\
 &= \sum_{a \leq x} p_{X|Y}(a|y)
 \end{aligned}$$

On constate donc que les définitions sont exactement les mêmes que dans le cas où il n'existe pas de condition. Simplement, les probabilités sont toutes modifiées par le fait que l'on sache que $Y = y$.

Lorsque X et Y sont indépendantes, les lois conditionnelles et non conditionnelles sont identiques. On peut se convaincre de cette évidence grâce au calcul suivant:

$$\begin{aligned}
 p_{X|Y}(x|y) &= P\{X = x | Y = y\} \\
 &= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\
 &= \frac{P\{X = x\}P\{Y = y\}}{P\{Y = y\}} \\
 &= P\{X = x\}
 \end{aligned}$$

6.4.2 Exemples de distributions conditionnelles discrètes

Exemple 6.18 La loi conjointe de probabilité $p(x, y)$ de deux variables X et Y est donnée ainsi:

$$p(0, 0) = .4 \quad p(0, 1) = .2 \quad p(1, 0) = .1 \quad p(1, 1) = .3$$

On veut trouver la loi conditionnelle de X lorsque $Y = 1$.

SOLUTION. On calcule d'abord

$$p_Y(1) = \sum_x p(x, 1) = p(0, 1) + p(1, 1) = .5$$

Ainsi

$$p_{X|Y}(0|1) = \frac{p(0, 1)}{p_Y(1)} = \frac{.2}{.5}$$

et

$$p_{X|Y}(1|1) = \frac{p(1, 1)}{p_Y(1)} = \frac{.3}{.5}$$

■

Exemple 6.19 Soient deux variables poissonniennes X et Y indépendantes et de paramètres λ_1 et λ_2 . On souhaite connaître la loi conditionnelle de X lorsqu'on sait que $X + Y = n$.

SOLUTION. Cette loi sera obtenue ainsi:

$$\begin{aligned} P\{X = k | X + Y = n\} &= \frac{P\{X = k, X + Y = n\}}{P\{X + Y = n\}} \\ &= \frac{P\{X = k, Y = n - k\}}{P\{X + Y = n\}} \\ &= \frac{P\{X = k\}P\{Y = n - k\}}{P\{X + Y = n\}} \end{aligned}$$

où la dernière transformation se fonde sur l'indépendance de X et Y . La loi de la somme $X + Y$ ayant déjà été calculée (d'après l'exemple 6.16, il s'agit d'une loi de Poisson de paramètres $\lambda_1 + \lambda_2$), l'expression précédente devient

$$\begin{aligned} P\{X = k | X + Y = n\} &= \frac{e^{-\lambda_1} \lambda_1^k}{k!} \frac{e^{-\lambda_2} \lambda_2^{n-k}}{(n-k)!} \left[\frac{e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^n}{n!} \right]^{-1} \\ &= \frac{n!}{(n-k)! k!} \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n} \\ &= \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n-k} \end{aligned}$$

En d'autres termes, la variable X suit, sous la condition que $X + Y = n$, une loi binomiale de paramètres n et $\lambda_1/(\lambda_1 + \lambda_2)$. ■

6.4.3 Cas continu

Soient X et Y des variables de densité conjointe $f(x, y)$. On définit la *densité conditionnelle de X* sous la condition $Y = y$, et lorsque $f_Y(y) > 0$ par la relation

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

On peut donner un fondement intuitif à cette définition en multipliant le membre de gauche par dx et celui de droite par $(dx dy)/dy$ pour obtenir

$$\begin{aligned} f_{X|Y}(x|y) dx &= \frac{f(x, y) dx dy}{f_Y(y) dy} \\ &\approx \frac{P\{x \leq X \leq x + dx, y \leq Y \leq y + dy\}}{P\{y \leq Y \leq y + dy\}} \\ &= P\{x \leq X \leq x + dx | y \leq Y \leq y + dy\} \end{aligned}$$

En d'autres termes, lorsque dx et dy sont assez petits, la quantité $f_{X|Y}(x|y) dx$ représente la probabilité conditionnelle que X soit entre x et $x + dx$, sachant que Y est comprise entre y et $y + dy$.

L'usage de ces densités conditionnelles rend possible le calcul de probabilités d'événements relatifs à une variable, sous condition qu'une seconde ait pris une valeur connue. Nommément, lorsque X et Y sont conjointement continues et pour tout événement A relatif à X , on aura

$$P\{X \in A | Y = y\} = \int_A f_{X|Y}(x|y) dx$$

Si en particulier on choisit $A = (-\infty, a]$, on aboutit à la définition de la *fonction de répartition conditionnelle de X* sous la condition $Y = y$:

$$F_{X|Y}(a|y) \equiv P\{X \leq a | Y = y\} = \int_{-\infty}^a f_{X|Y}(x|y) dx$$

Il faut noter le fait que les notions qui viennent d'être présentées permettent le calcul de probabilités conditionnelles même dans les cas où la probabilité de la condition (à savoir $Y = y$) est nulle, ce qui est assez remarquable.

6.4.4 Exemples de distributions conditionnelles continues

Exemple 6.20 Soient X et Y des variables ayant pour densité conjointe

$$f(x, y) = \begin{cases} \frac{12}{5}x(2 - x - y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{sinon} \end{cases}$$

On cherche la densité conditionnelle de X , sachant que $Y = y$, où $0 < y < 1$.

SOLUTION. On aura, lorsque $0 < x < 1$ et $0 < y < 1$

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x, y)}{f_Y(y)} \\ &= \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx} \\ &= \frac{x(2 - x - y)}{\int_0^1 x(2 - x - y) dx} \\ &= \frac{x(2 - x - y)}{\frac{2}{3} - y/2} \\ &= \frac{6x(2 - x - y)}{4 - 3y} \end{aligned}$$

■

Exemple 6.21 Supposons que X et Y aient pour densité conjointe

$$f(x, y) = \begin{cases} \frac{e^{-x/y} e^{-y}}{y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{sinon} \end{cases}$$

On cherche à calculer $P\{X > 1 | Y = y\}$.

SOLUTION. Cherchons d'abord la densité conditionnelle de X lorsque $Y = y$:

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x, y)}{f_Y(y)} \\ &= \frac{e^{-x/y} e^{-y}/y}{e^{-y} \int_0^{\infty} (1/y) e^{-x/y} dx} \\ &= \frac{1}{y} e^{-x/y} \end{aligned}$$

Ainsi

$$\begin{aligned} P\{X > 1 | Y = y\} &= \int_1^{\infty} \frac{1}{y} e^{-x/y} dx \\ &= -e^{-x/y} \Big|_1^{\infty} \\ &= e^{-1/y} \end{aligned}$$

Comme dans le cas discret, X et Y ont une même densité conditionnelle et non conditionnelle en cas d'indépendance. En effet, sous cette hypothèse

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x) \quad \blacksquare$$

6.4.5 Autres cas de distributions conditionnelles

On peut parler de distribution conditionnelle même lorsque les variables mises en jeu ne sont ni conjointement continues ni conjointement discrètes. On peut par exemple imaginer le cas de deux variables, l'une notée X continue de densité f , l'autre discrète notée N , et s'intéresser à la densité conditionnelle de X sous la condition $N = n$. Tout d'abord

$$\begin{aligned} &\frac{P\{x < X < x + dx | N = n\}}{dx} \\ &= \frac{P\{N = n | x < X < x + dx\} P\{x < X < x + dx\}}{P\{N = n\} dx} \end{aligned}$$

Lorsque dx tend vers 0, on peut réécrire cette égalité ainsi:

$$\lim_{dx \rightarrow 0} \frac{P\{x < X < x + dx | N = n\}}{dx} = \frac{P\{N = n | X = x\}}{P\{N = n\}} f(x)$$

où l'on voit que la densité cherchée est

$$f_{X|N}(x|n) = \frac{P\{N = n | X = x\}}{P\{N = n\}} f(x)$$

Exemple 6.22 On considère une suite de $n + m$ épreuves indépendantes identiques. Leur probabilité de succès n'est cependant pas connue d'avance. On admet qu'elle est aléatoire et distribuée uniformément dans l'intervalle $(0, 1)$. Que devient cette distribution si l'on apprend qu'une réalisation des $m + n$ épreuves a donné n succès ?

SOLUTION. Désignons par X la probabilité de succès à l'occasion d'une épreuve isolée. X est uniforme sur $(0, 1)$. Si l'on sait que $X = x$, les $n + m$ épreuves sont indépendantes et la probabilité de succès de chacune est x . Le nombre N des succès est donc une variable binomiale de paramètres $(n + m, x)$. Dès lors, la densité conditionnelle de X sous la condition $N = n$ est

$$\begin{aligned} f_{X|N}(x|n) &= \frac{P\{N = n | X = x\} f_X(x)}{P\{N = n\}} \\ &= \frac{\binom{n+m}{n} x^n (1-x)^m}{P\{N = n\}} \quad 0 < x < 1 \\ &= cx^n (1-x)^m \end{aligned}$$

où c ne dépend pas de x . On reconnaît la densité de la loi bêta de paramètres $(n + 1, m + 1)$.

Ce dernier résultat présente un intérêt supplémentaire: si dans notre expérience la distribution a priori (c'est-à-dire avant toute réalisation) de la probabilité de succès d'un tirage est uniforme sur $(0, 1)$ – ce qui équivaut à dire qu'elle suit une loi bêta de paramètres $(1, 1)$ – la probabilité a posteriori (donc conditionnelle) suit une loi bêta $(1 + n, 1 + m)$ lorsqu'il y a eu n succès parmi les $n + m$ épreuves. Cet exemple donne donc un fondement intuitif qui permet de mieux sentir ce que représente l'hypothèse pour une variable de suivre une loi bêta. ■

6.5 STATISTIQUES D'ORDRE

6.5.1 Définition

Considérons X_1, X_2, \dots, X_n , un groupe de n variables aléatoires indépendantes identiquement distribuées, continues de densité f et ayant une fonction de répartition F . On définit les variables aléatoires suivantes:

$X_{(1)}(\omega_1, \omega_2, \dots, \omega_n)$ vaudra le plus petit des $X_i(\omega_i)$
 $X_{(2)}(\omega_1, \omega_2, \dots, \omega_n)$ vaudra le second plus petit des $X_i(\omega_i)$

·
·
·

$X_{(j)}(\omega_1, \omega_2, \dots, \omega_n)$ vaudra le j -ème plus petit des $X_i(\omega_i)$

·
·
·

$X_{(n)}(\omega_1, \omega_2, \dots, \omega_n)$ vaudra le plus grand des $X_i(\omega_i)$,

où $(\omega_1, \omega_2, \dots, \omega_n)$ est un événement de l'ensemble fondamental associé aux variables conjointes X_1, X_2, \dots, X_n . Les fonctions ordonnées $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ sont appelées **statistiques d'ordre** associées à X_1, X_2, \dots, X_n . En d'autres termes, ces variables font correspondre à un événement conjoint non pas les valeurs directes mais les valeurs classées par ordre croissant.

6.5.2 Densité conjointe

On peut obtenir la densité conjointe des statistiques d'ordre en prenant pour point de départ le fait que les variables $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ prendront les valeurs $x_1 \leq x_2 \leq \dots \leq x_n$ si et seulement s'il existe une permutation (i_1, i_2, \dots, i_n) de $(1, 2, \dots, n)$ telle que

$$X_1 = x_{i_1}, X_2 = x_{i_2}, \dots, X_n = x_{i_n}$$

Or, pour toute permutation semblable, on peut écrire

$$\begin{aligned} P \left\{ x_{i_1} - \frac{\varepsilon}{2} < X_1 < x_{i_1} + \frac{\varepsilon}{2}, \dots, x_{i_n} - \frac{\varepsilon}{2} < X_n < x_{i_n} + \frac{\varepsilon}{2} \right\} \\ &\approx \varepsilon^n f_{X_1, \dots, X_n}(x_{i_1}, \dots, x_{i_n}) \\ &= \varepsilon^n f(x_{i_1}) \cdots f(x_{i_n}) \\ &= \varepsilon^n f(x_1) \cdots f(x_n) \end{aligned}$$

De ce fait, en passant aux statistiques d'ordre, on aura

$$\begin{aligned} P \left\{ x_1 - \frac{\varepsilon}{2} < X_{(1)} < x_1 + \frac{\varepsilon}{2}, \dots, x_n - \frac{\varepsilon}{2} < X_{(n)} < x_n + \frac{\varepsilon}{2} \right\} \\ &\approx n! \varepsilon^n f(x_1) \cdots f(x_n) \end{aligned}$$

Lorsqu'on divise par ε^n et à condition que ε tende vers 0, on obtient

$$f_{X_{(1)}, \dots, X_{(n)}}(x_1, x_2, \dots, x_n) = n! f(x_1) \cdots f(x_n) \quad x_1 < x_2 < \dots < x_n \quad (6.19)$$

Il existe une justification intuitive à (6.19): dès que (X_1, X_2, \dots, X_n) prend pour valeur l'une des $n!$ permutations de (x_1, x_2, \dots, x_n) , les variables conjointes $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$

prennent pour valeur (x_1, x_2, \dots, x_n) précisément. Il y a d'ailleurs équivalence. Comme la densité de probabilité pour toute permutation de (x_1, x_2, \dots, x_n) est $f(x_1) \dots f(x_n)$, (6.19) se trouve expliquée.

Exemple 6.23 Trois personnes sont échelonnées «au hasard» sur une route de un kilomètre de longueur. On se demande la probabilité que les personnes soient espacées d'au moins d kilomètre, d étant inférieur ou égal à 0,5.

SOLUTION. Admettons que l'hypothèse «échelonnés au hasard» signifie que la position de chacune des trois personnes soit uniformément distribuée le long de ce bout de route et indépendante de celle des autres. On désigne par X_i la position de la i -ème personne. Ce que l'on cherche à calculer est donc $P\{X_{(i)} > X_{(i-1)} + d, i = 2, 3\}$. Comme

$$f_{X_{(1)}, X_{(2)}, X_{(3)}}(x_1, x_2, x_3) = 3! \quad 0 < x_1 < x_2 < x_3 < 1$$

on peut écrire

$$\begin{aligned} P\{X_{(i)} > X_{(i-1)} + d, i = 2, 3\} &= \iiint_{\substack{x_i > x_{i-1} + d \\ i=2,3}} f_{X_{(1)}, X_{(2)}, X_{(3)}}(x_1, x_2, x_3) dx_1 dx_2 dx_3 \\ &= 3! \int_0^{1-2d} \int_{x_1+d}^{1-d} \int_{x_2+d}^1 dx_3 dx_2 dx_1 \\ &= 6 \int_0^{1-2d} \int_{x_1+d}^{1-d} (1-d-x_2) dx_2 dx_1 \\ &= 6 \int_0^{1-2d} \int_0^{1-2d-x_1} y_2 dy_2 dx_1 \end{aligned}$$

où l'on a effectué le changement de variable $y_2 = 1 - d - x_2$. La dernière expression devient

$$\begin{aligned} &= 6 \int_0^{1-2d} \frac{(1-2d-x_1)^2}{2} dx_1 \\ &= 6 \int_0^{1-2d} \frac{y_1^2}{2} dy_1 \\ &= (1-2d)^3 \end{aligned}$$

La probabilité cherchée, à savoir celle que les personnes soient espacées d'au moins d kilomètre, est donc $(1-2d)^3$ lorsque $d \leq 0,5$. Grâce à un calcul analogue, on peut établir que la probabilité correspondante lorsqu'on place n personnes au hasard sur une route mesurant un kilomètre de longueur est

$$[1 - (n - 1)d]^n \quad \text{lorsque } d \leq \frac{1}{n - 1}$$

La preuve en est laissée en exercice. ■

6.5.3 Densités marginales

On peut obtenir la densité marginale de la j -ième statistique d'ordre $X_{(j)}$ en intégrant la densité conjointe donnée dans (6.19). Mais on peut alternativement faire le raisonnement suivant: pour que $X_{(j)}$ prenne la valeur x , il faut et suffit que $j - 1$ des n valeurs prises par X_1, \dots, X_n soient inférieures à x , que $n - j$ d'entre elles soient supérieures à x et que la dernière soit x précisément. La densité de probabilité calculée en un point (x_1, x_2, \dots, x_n) dont les coordonnées satisfont les conditions précitées est

$$[F(x)]^{j-1} [1 - F(x)]^{n-j} f(x)$$

Or il y a

$$\binom{n}{j-1, n-j, 1} = \frac{n!}{(n-j)!(j-1)!}$$

partitions de n nombres en trois groupes tels que ceux décrits. Par conséquent la densité de $X_{(j)}$ sera

$$f_{X_{(j)}}(x) = \frac{n!}{(n-j)!(j-1)!} [F(x)]^{j-1} [1 - F(x)]^{n-j} f(x) \quad (6.20)$$

Exemple 6.24 On appelle échantillon de taille $2n + 1$ tout groupe de $2n + 1$ variables aléatoires indépendantes et identiquement distribuées. La médiane d'un tel échantillon est $X_{(n+1)}$. Supposons que l'on observe un échantillon de taille 3 de variables uniformes sur $(0, 1)$. On s'intéresse à la probabilité que la médiane prenne une valeur comprise entre $\frac{1}{4}$ et $\frac{3}{4}$.

SOLUTION. La densité de $X_{(2)}$ est donnée selon (6.20) par

$$f_{X_{(2)}}(x) = \frac{3!}{1!1!} x(1-x) \quad 0 < x < 1$$

Ainsi

$$\begin{aligned} P\left(\frac{1}{4} < X_{(2)} < \frac{3}{4}\right) &= 6 \int_{1/4}^{3/4} x(1-x) dx \\ &= 6 \left\{ \frac{x^2}{2} - \frac{x^3}{3} \right\} \Bigg|_{x=1/4}^{x=3/4} = \frac{11}{16} \end{aligned}$$

■

6.5.4 Fonction de répartition

On peut calculer la fonction de répartition de $X_{(j)}$ par intégration de (6.20), ce qui donne

$$F_{X_{(j)}}(y) = \frac{n!}{(n-j)!(j-1)!} \int_{-\infty}^y [F(x)]^{j-1} [1-F(x)]^{n-j} f(x) dx \quad (6.21)$$

Mais il existe une autre méthode assez directe pour le calcul de $F_{X_{(j)}}$. En effet, $X_{(j)}$ prendra une valeur inférieure à un nombre y si et seulement si j des variables au moins prennent des valeurs inférieures ou égales à y . Comme le nombre des variables X_i à valeur inférieure ou égale à y est une variable binomiale de paramètres $(n, p = F(y))$, on peut écrire

$$\begin{aligned} F_{X_{(j)}}(y) &= P\{X_{(j)} \leq y\} = P\{\text{au moins } j \text{ des } X_i \text{ sont } \leq y\} \\ &= \sum_{k=j}^n \binom{n}{k} [F(y)]^k [1-F(y)]^{n-k} \end{aligned} \quad (6.22)$$

En marge de ce résultat général, on peut appliquer (6.21) et (6.22) au cas où F correspond à une distribution uniforme sur $(0, 1)$. On aura alors évidemment $F(x) = x$ lorsque $0 < x < 1$. Cela nous fait aboutir à une identité intéressante:

$$\sum_{k=j}^n \binom{n}{k} y^k (1-y)^{n-k} = \frac{n!}{(n-j)!(j-1)!} \int_0^y x^{j-1} (1-x)^{n-j} dx \quad 0 \leq y \leq 1 \quad (6.23)$$

6.5.5 Densité conjointe de deux statistiques d'ordre

On peut déterminer la densité conjointe de deux statistiques d'ordre $X_{(i)}$ et $X_{(j)}$ en invoquant le même genre d'argument que celui qui a permis d'établir (6.20). En supposant que l'on soit dans le cas où $i < j$, cette densité est

$$\begin{aligned} f_{X_{(i)}, X_{(j)}}(x_i, x_j) &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \\ &\times [F(x_i)]^{i-1} [F(x_j) - F(x_i)]^{j-i-1} [1-F(x_j)]^{n-j} f(x_i) f(x_j) \end{aligned} \quad (6.24)$$

pour tous les arguments satisfaisant $x_i < x_j$.

Exemple 6.25 Distribution de l'étendue d'un échantillon.

Soient n variables indépendantes et identiquement distribuées X_1, X_2, \dots, X_n . L'*étendue* de cet échantillon est la variable aléatoire $R = X_{(n)} - X_{(1)}$. Si l'on note f et F respectivement la densité et la fonction de répartition des variables X_i , la fonction de répartition de R peut être calculée à partir de (6.24) comme suit pour $a \geq 0$:

$$\begin{aligned} P\{R \leq a\} &= P\{X_{(n)} - X_{(1)} \leq a\} \\ &= \iint_{x_n - x_1 \leq a} f_{X_{(1)}, X_{(n)}}(x_1, x_n) dx_1 dx_n \\ &= \int_{-\infty}^{\infty} \int_{x_1}^{x_1+a} \frac{n!}{(n-2)!} [F(x_n) - F(x_1)]^{n-2} f(x_1) f(x_n) dx_n dx_1 \end{aligned}$$

En effectuant le changement de variable $y = F(x_n) - F(x_1)$, $dy = f(x_n) dx_n$, on trouve

$$\begin{aligned} \int_{x_1}^{x_1+a} [F(x_n) - F(x_1)]^{n-2} f(x_n) dx_n &= \int_0^{F(x_1+a)-F(x_1)} y^{n-2} dy \\ &= \frac{1}{n-1} [F(x_1+a) - F(x_1)]^{n-1} \end{aligned}$$

et donc

$$P\{R \leq a\} = n \int_{-\infty}^{\infty} [F(x_1+a) - F(x_1)]^{n-1} f(x_1) dx_1 \quad (6.25)$$

L'évaluation explicite de (6.25) n'est possible que dans quelques cas particuliers. Par exemple, celui où les X_i sont uniformément distribuées dans $(0, 1)$. On obtient alors en appliquant (6.25) et pour $0 < a < 1$:

$$\begin{aligned} P\{R < a\} &= n \int_0^1 [F(x_1+a) - F(x_1)]^{n-1} f(x_1) dx_1 \\ &= n \int_0^{1-a} a^{n-1} dx_1 + n \int_{1-a}^1 (1-x_1)^{n-1} dx_1 \\ &= n(1-a)a^{n-1} + a^n \end{aligned}$$

Par dérivation on peut calculer la densité de R , à savoir ici

$$f_R(a) = \begin{cases} n(n-1)a^{n-2}(1-a) & 0 \leq a \leq 1 \\ 0 & \text{sinon} \end{cases} \quad \blacksquare$$

6.6 CHANGEMENT DE VARIABLES MULTIDIMENSIONNELLES

6.6.1 Cas bidimensionnel

Considérons deux variables aléatoires X_1 et X_2 conjointement continues de densité f_{X_1, X_2} . Il arrive qu'on s'intéresse à la densité conjointe de Y_1 et Y_2 , deux fonctions de X_1 et X_2 . Supposons pour être plus précis que $Y_1 = g_1(X_1, X_2)$ et $Y_2 = g_2(X_1, X_2)$. Admettons encore que les fonctions g_1 et g_2 satisfassent les deux conditions suivantes:

- on peut résoudre par rapport à x_1 et x_2 le système d'équations $y_1 = g_1(x_1, x_2)$, $y_2 = g_2(x_1, x_2)$. Les solutions sont notées $x_1 = h_1(y_1, y_2)$ et $x_2 = h_2(y_1, y_2)$;
- les fonctions g_1 et g_2 sont continûment différentiables et de *Jacobien* partout non nul. Ceci revient à écrire, en utilisant la définition du Jacobien:

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} \equiv \frac{\partial g_1}{\partial x_1} \frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2} \frac{\partial g_2}{\partial x_1} \neq 0$$

pour tout couple (x_1, x_2) .

Sous ces conditions, on peut montrer que les variables Y_1 et Y_2 sont conjointement continues et de densité

$$f_{Y_1, Y_2}(y_1, y_2) = f_{X_1, X_2}(x_1, x_2) |J(x_1, x_2)|^{-1} \quad (6.26)$$

où il faut remplacer x_1 par $h_1(y_1, y_2)$ et x_2 par $h_2(y_1, y_2)$.

La démarche suivante mènerait à la démonstration de (6.26). D'abord calculer

$$p\{Y_1 \leq y_1, Y_2 \leq y_2\} = \iint_{\substack{(x_1, x_2): \\ g_1(x_1, x_2) \leq y_1 \\ g_2(x_1, x_2) \leq y_2}} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \quad (6.27)$$

La dérivation de (6.27) par rapport à y_1 puis y_2 donnerait alors la densité conjointe voulue. La preuve que cette dérivation donne bien le membre de droite de (6.26) est un exercice avancé d'analyse qui ne sera pas traité dans cet ouvrage.

6.6.2 Exemples de changement de variables bidimensionnelles

Exemple 6.26 Soient X_1 et X_2 deux variables aléatoires conjointement continues de densité f_{X_1, X_2} . Posons $Y_1 = X_1 + X_2$ et $Y_2 = X_1 - X_2$. Exprimer la densité conjointe de Y_1 et Y_2 en fonction de f_{X_1, X_2} .

SOLUTION. Posons $g_1(x_1, x_2) = x_1 + x_2$ et $g_2(x_1, x_2) = x_1 - x_2$. On a

$$J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

Comme la solution du système $y_1 = x_1 + x_2$, $y_2 = x_1 - x_2$ est $x_1 = (y_1 + y_2)/2$, $x_2 = (y_1 - y_2)/2$, l'application de (6.26) livre

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{1}{2} f_{X_1, X_2} \left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2} \right)$$

Dans le cas où X_1 et X_2 sont indépendantes et uniformes sur $(0, 1)$, ce dernier résultat devient

$$f_{Y_1, Y_2}(y_1, y_2) = \begin{cases} \frac{1}{2} & 0 \leq y_1 + y_2 \leq 2, 0 \leq y_1 - y_2 \leq 2 \\ 0 & \text{sinon} \end{cases}$$

Si, par contre, X_1 et X_2 sont indépendantes et exponentielles de paramètres λ_1 et λ_2 respectivement, on obtient

$$f_{Y_1, Y_2}(y_1, y_2) = \begin{cases} \frac{\lambda_1 \lambda_2}{2} \exp \left\{ -\lambda_1 \left(\frac{y_1 + y_2}{2} \right) - \lambda_2 \left(\frac{y_1 - y_2}{2} \right) \right\} & y_1 + y_2 \geq 0, y_1 - y_2 \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Si enfin X_1 et X_2 sont indépendantes et normales centrées réduites, on aura

$$\begin{aligned} f_{Y_1, Y_2}(y_1, y_2) &= \frac{1}{4\pi} e^{-[(y_1+y_2)^2/8 + (y_1-y_2)^2/8]} \\ &= \frac{1}{4\pi} e^{-(y_1^2+y_2^2)/4} \\ &= \frac{1}{\sqrt{4\pi}} e^{-y_1^2/4} \frac{1}{\sqrt{4\pi}} e^{-y_2^2/4} \end{aligned}$$

où l'on découvre un résultat secondaire intéressant: $X_1 + X_2$ est ici indépendante de $X_1 - X_2$. En fait, on peut montrer que pour deux variables aléatoires X_1 et X_2 indépendantes et de même fonction de répartition F , les variables $X_1 + X_2$ et $X_1 - X_2$ sont indépendantes si et seulement si F est une fonction de répartition de variable normale. ■

Exemple 6.27 On désigne par (X, Y) les coordonnées d'un point aléatoirement choisi dans le plan. On suppose que X et Y sont de distribution normale centrée et réduite. On voudrait connaître la distribution de R et θ , les coordonnées polaires du même point (voir figure 6.5).

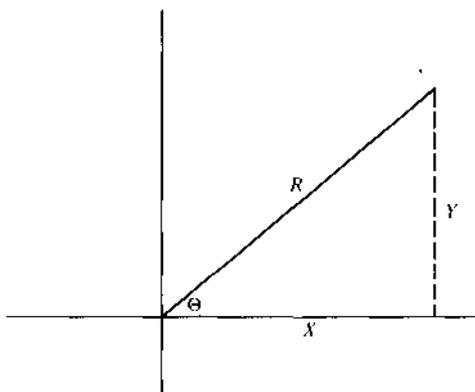


Figure 6.5 . = Point aléatoire. $(X, Y) = R, \Theta$

Si l'on écrit $r = g_1(x, y) = \sqrt{x^2 + y^2}$ et $\theta = g_2(x, y) = \text{Arc tg } y/x$, on peut calculer

$$\begin{aligned} \frac{\partial g_1}{\partial x} &= \frac{x}{\sqrt{x^2 + y^2}} & \frac{\partial g_1}{\partial y} &= \frac{y}{\sqrt{x^2 + y^2}} \\ \frac{\partial g_2}{\partial x} &= \frac{1}{1 + (y/x)^2} \left(\frac{-y}{x^2} \right) = \frac{-y}{x^2 + y^2} & \frac{\partial g_2}{\partial y} &= \frac{1}{x[1 + (y/x)^2]} = \frac{x}{x^2 + y^2} \end{aligned}$$

Ainsi

$$J(x, y) = \frac{x^2}{(x^2 + y^2)^{3/2}} + \frac{y^2}{(x^2 + y^2)^{3/2}} = \frac{1}{\sqrt{x^2 + y^2}} = \frac{1}{r}$$

Or la densité conjointe de X et Y est

$$f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

Par conséquent, la densité conjointe de R et θ sera

$$f(r, \theta) = \frac{1}{2\pi} r e^{-r^2/2} \quad 0 < \theta < 2\pi \quad 0 < r < \infty$$

Cette densité étant décomposable et constituée du produit des densités marginales de R et θ , il en résulte que ces deux variables sont indépendantes. θ est ici uniformément distribuée sur $(0, 2\pi)$, tandis que R suit une distribution de Rayleigh de densité

$$f(r) = r e^{-r^2/2} \quad 0 < r < \infty$$

A titre d'illustration, lors du tir sur une cible bidimensionnelle, il peut arriver que les erreurs verticale et horizontale soient de distribution normale centrée réduite. Dans ce cas, la distance entre le centre de la cible et l'impact du tir suit une distribution de Rayleigh.

Ces résultats sont remarquables. Il n'est en effet pas évident a priori que le vecteur des erreurs, dont les coordonnées aléatoires sont de distribution normale centrée réduite et indépendantes, soit d'orientation uniformément répartie, et ce, de plus, à n'importe quelle distance du centre.

On peut ainsi s'intéresser à la distribution conjointe de R^2 et θ : le Jacobien de la transformation $d = g_1(x, y) = x^2 + y^2$, $\theta = g_2(x, y) = \text{Arc tg}(y/x)$ étant

$$J = \left| \begin{array}{cc} 2x & 2y \\ -y & x \\ x^2 + y^2 & x^2 + y^2 \end{array} \right| = 2$$

on voit que

$$f_{R^2, \theta}(d, \theta) = \frac{1}{2} e^{-d/2} \frac{1}{2\pi} \quad 0 < d < \infty, 0 < \theta < 2\pi$$

on constate que R^2 et θ sont indépendantes, R^2 suivant une distribution exponentielle de paramètre $\frac{1}{2}$. Comme par définition $R^2 = X^2 + Y^2$ suit par ailleurs une loi χ^2 à deux degrés de liberté, on vérifie un résultat déjà présenté: la loi χ^2 à deux degrés de liberté et la loi exponentielle de paramètre $\frac{1}{2}$ ne font qu'une. ■

Exemple 6.28 Les résultats précédents permettent de réaliser un simulateur ou générateur de nombres aléatoires distribués normalement, et ce, à partir d'un générateur de nombres uniformément distribués.

Notons par U_1 et U_2 deux variables de distribution uniforme sur $(0, 1)$. Nous allons déterminer une transformation de U_1, U_2 qui donne deux variables normales centrées réduites X_1, X_2 en considérant d'abord les variables (R, θ) associées à (X_1, X_2) par un

passage aux coordonnées polaires. On vient de voir que si X_1, X_2 sont indépendantes et normales standard, alors $R^2 = X_1^2 + X_2^2$ et θ seront indépendantes, R^2 étant de plus de distribution exponentielle avec paramètre $\lambda = \frac{1}{2}$. Or, $-2 \ln U_1$ suit une telle distribution puisque, lorsque $x > 0$,

$$\begin{aligned} P\{-2 \ln U_1 < x\} &= P\left\{\ln U_1 > -\frac{x}{2}\right\} \\ &= P\{U_1 > e^{-x/2}\} \\ &= 1 - e^{-x/2} \end{aligned}$$

Par ailleurs, on peut utiliser pour simuler θ la variable $2\pi U_2$ qui suit une loi uniforme sur $(0, 2\pi)$. De ce fait, en posant

$$\begin{cases} R^2 = -2 \ln U_1 \\ \theta = 2\pi U_2 \end{cases}$$

R^2 peut être considéré comme le carré de la distance à l'origine et θ l'angle donnant l'orientation du point (X_1, X_2) . Comme $X_1 = R \cos \theta$ et $X_2 = R \sin \theta$, les deux variables seront bien indépendantes et normales centrées réduites. ■

Exemple 6.29 Soient X et Y deux variables aléatoires indépendantes suivant des lois gamma de paramètres respectifs (α, λ) et (β, λ) . On veut connaître la densité conjointe de $U = X + Y$ et $V = X/(X + Y)$.

SOLUTION. La densité conjointe de X et Y est

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \frac{\lambda e^{-\lambda y} (\lambda y)^{\beta-1}}{\Gamma(\beta)} \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda(x+y)} x^{\alpha-1} y^{\beta-1} \end{aligned}$$

Mais, si $g_1(x, y) = x + y$, $g_2(x, y) = x/(x + y)$, alors

$$\frac{\partial g_1}{\partial x} = \frac{\partial g_1}{\partial y} = 1 \quad \frac{\partial g_2}{\partial x} = \frac{y}{(x+y)^2} \quad \frac{\partial g_2}{\partial y} = -\frac{x}{(x+y)^2}$$

et ainsi

$$J(x, y) = \left| \begin{array}{cc} 1 & 1 \\ \frac{y}{(x+y)^2} & -\frac{x}{(x+y)^2} \end{array} \right| = -\frac{1}{x+y}$$

Comme la solution du système $u = x + y$, $v = x/(x + y)$ est $x = uv$, $y = u(1 - v)$, on peut écrire

$$f_{U,V}(u, v) = f_{X,Y}[uv, u(1-v)]u \\ = \frac{\lambda e^{-\lambda u} (\lambda u)^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)} \frac{v^{\alpha-1} (1-v)^{\beta-1} \Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

On constate que $X + Y$ et $X/(X + Y)$ sont indépendantes; de plus, $X + Y$ suit une loi gamma de paramètres $(\alpha + \beta, \lambda)$ tandis que $X/(X + Y)$ suit une loi bêta de paramètres (α, β) . On peut de plus en tirer que $B(\alpha, \beta)$, le facteur d'ajustement dans l'expression de la densité des variables bêta, est

$$B(\alpha, \beta) = \int_0^1 v^{\alpha-1} (1-v)^{\beta-1} dv \\ = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Ce dernier résultat est très intéressant et peut être illustré de la manière suivante: supposons que $n + m$ tâches doivent être exécutées et que la durée d'exécution de chacune soit une variable aléatoire exponentielle de paramètre λ et soit indépendante des autres durées. On dispose de deux travailleurs pour mener à bien ces tâches. Le premier de ces travailleurs exécutera les tâches 1, 2, ..., n , tandis que l'autre se chargera des m tâches restantes. Désignons par X et Y respectivement le temps de travail total de ces deux personnes. On peut alors affirmer – d'après ce qui précède ou ce qui a été établi au paragraphe 6.3.3 – que X et Y seront des variables gamma indépendantes de paramètres (n, λ) et (m, λ) respectivement. De plus, les résultats précédents indiquent qu'indépendamment de la durée totale d'exécution des $n + m$ tâches (donc indépendamment de $X + Y$), la part réalisée par le premier employé suit une distribution bêta de paramètres (n, m) . ■

6.6.3 Cas général

L'approche du cas général est similaire. On a ici n variables X_1, X_2, \dots, X_n de densité conjointe connue et l'on s'intéresse à la densité conjointe de Y_1, Y_2, \dots, Y_n où

$$Y_1 = g_1(X_1, \dots, X_n) \quad Y_2 = g_2(X_1, \dots, X_n), \dots \\ Y_n = g_n(X_1, \dots, X_n)$$

On admettra que ces fonctions g_i ont des dérivées partielles continues et que leur Jacobien $J(x_1, \dots, x_n)$ est partout non nul. Par $J(x_1, \dots, x_n)$, on entend

$$J(x_1, \dots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \dots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \dots & \frac{\partial g_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \dots & \frac{\partial g_n}{\partial x_n} \end{vmatrix}$$

On admettra encore que le système d'équations $y_1 = g_1(x_1, \dots, x_n)$, $y_2 = g_2(x_1, \dots, x_n), \dots$, $y_n = g_n(x_1, \dots, x_n)$ a une solution unique notée $x_1 = h_1(y_1, \dots, y_n), \dots, x_n = h_n(y_1, \dots, y_n)$. Si ces conditions sont réalisées, la densité conjointe des variables Y_i est

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = f_{X_1, \dots, X_n}(x_1, \dots, x_n) |J(x_1, \dots, x_n)|^{-1} \quad (6.28)$$

où x_i est à remplacer par $h_i(y_1, \dots, y_n)$, $i = 1, 2, \dots, n$.

Exemple 6.30 Les trois variables X_1 , X_2 , et X_3 sont normales centrées réduites et indépendantes. On cherche ici la densité conjointe de $Y_1 = X_1 + X_2 + X_3$, $Y_2 = X_1 - X_2$, $Y_3 = X_1 - X_3$.

SOLUTION. Le Jacobien de la transformation décrite est

$$J = \begin{vmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{vmatrix} = 3$$

La transformation réciproque donne

$$X_1 = \frac{Y_1 + Y_2 + Y_3}{3} \quad X_2 = \frac{Y_1 - 2Y_2 + Y_3}{3} \quad X_3 = \frac{Y_1 + Y_2 - 2Y_3}{3}$$

et par conséquent, d'après (6.28), la densité cherchée est

$$f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3) = \frac{1}{3} f_{X_1, X_2, X_3} \left(\frac{y_1 + y_2 + y_3}{3}, \frac{y_1 - 2y_2 + y_3}{3}, \frac{y_1 + y_2 - 2y_3}{3} \right)$$

Ainsi, comme

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = \frac{1}{(2\pi)^{3/2}} e^{-\sum_{i=1}^3 x_i^2/2}$$

on voit que

$$f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3) = \frac{1}{3(2\pi)^{3/2}} e^{-Q(y_1, y_2, y_3)/2}$$

où

$$\begin{aligned} Q(y_1, y_2, y_3) &= \left(\frac{y_1 + y_2 + y_3}{3} \right)^2 + \left(\frac{y_1 - 2y_2 + y_3}{3} \right)^2 + \left(\frac{y_1 + y_2 - 2y_3}{3} \right)^2 \\ &= \frac{y_1^2}{3} + \frac{2}{3} y_2^2 + \frac{2}{3} y_3^2 - \frac{2}{3} y_2 y_3. \end{aligned}$$

6.7 EXERCICES THÉORIQUES

6.7.1 Vérifier l'équation 6.2.

6.7.2 Supposons que le nombre d'événements qui se produisent durant un intervalle de temps donné soit une variable aléatoire de Poisson, de paramètre λ . Si chaque événement est classé comme étant un événement de type i associé à une probabilité $p_i, i = 1, \dots, n, \sum p_i = 1$, et ceci indépendamment des autres événements, montrer que les nombres d'événements de type i qui se produisent, $i = 1, \dots, n$, sont des variables aléatoires de Poisson, de paramètres respectifs $\lambda p_i, i = 1, \dots, n$ et qui de plus sont indépendantes.

6.7.3 Proposer une méthode utilisant le problème de l'aiguille de Buffon afin d'estimer π . Il est assez surprenant de noter que c'était une méthode commune autrefois pour estimer π .

6.7.4 Résoudre le problème de l'aiguille de Buffon quand $L > D$.

RÉPONSE. $\frac{2L}{\pi D} (1 - \sin \theta) + 2\theta/\pi$, où θ est tel que $\cos \theta = D/L$.

6.7.5 Si X et Y sont des variables aléatoires indépendantes continues et positives, exprimer la fonction de densité de

- $Z = X/Y$ et
- $Z = XY$

en termes de fonctions de densité de X et Y . Évaluer ces expressions dans le cas particulier où X et Y sont toutes deux des variables aléatoires exponentielles.

6.7.6 Montrer analytiquement (par un raisonnement d'induction) que $X_1 + \dots + X_n$ suit une distribution binomiale négative quand les $X_i, i = 1, \dots, n$ sont des variables aléatoires géométriques indépendantes et identiquement distribuées. Donner également un second argument qui vérifie ce qui précède, ceci sans calcul.

6.7.7

- a) Si X suit une distribution gamma de paramètres (t, λ) , quelle est la distribution de $cX, c > 0$?
- b) Montrer que

$$\frac{1}{2\lambda} \chi_{2n}^2$$

possède une distribution gamma de paramètres n, λ lorsque n est un entier positif et χ_{2n}^2 est une variable aléatoire chi-carré avec $2n$ degrés de liberté.

6.7.8 Soient X et Y des variables aléatoires continues indépendantes avec des fonctions taux de panne respectives $\lambda_X(t)$ et $\lambda_Y(t)$, et soit $W = \min(X, Y)$.

- a) Déterminer la fonction de répartition de W en fonction de celles de X et de Y .
 b) Montrer que $\lambda_w(t)$, la fonction taux de panne de W , est donnée par

$$\lambda_w(t) = \lambda_X(t) + \lambda_Y(t)$$

6.7.9 Soient X_1, \dots, X_n des variables aléatoires indépendantes exponentielles de paramètre commun λ . Déterminer la distribution de $\min(X_1, \dots, X_n)$.

6.7.10 Les durées de vie de batteries sont des variables aléatoires indépendantes exponentielles de même paramètre λ . Une torche électrique a besoin de 2 batteries pour fonctionner. Si l'on a une torche et n batteries de réserve, quelle est la distribution du temps de fonctionnement de la torche?

6.7.11 Soient X_1, X_2, X_3, X_4 et X_5 des variables aléatoires continues indépendantes identiquement distribuées de fonction de répartition F et de densité f et soit

$$I = P\{X_1 < X_2 > X_3 < X_4 > X_5\}$$

- a) Montrer que I ne dépend pas de F . Pour cela, exprimer I comme une intégrale à 5 dimensions et effectuer le changement de variables $u_i = F(x_i)$, $i = 1, \dots, 5$.
 b) Evaluer I .

6.7.12 Etablir le théorème 6.2.

Pour cela, l'établir d'abord pour $n = 2$, puis utiliser un raisonnement d'induction.

6.7.13 Dans l'exemple 6.22, nous avons calculé la densité conditionnelle de la probabilité d'obtenir un succès, qui était variable, ceci sachant qu'une suite de $n + m$ premières épreuves avait donné n succès. Est-ce que cette densité conditionnelle aurait changé si, parmi toutes les épreuves, nous avons spécifié celles représentant les n succès?

6.7.14 Supposer que X et Y sont des variables aléatoires géométriques indépendantes ayant le même paramètre p .

- a) Sans aucun calcul, que pensez-vous de la valeur de $P\{X = i | X + Y = n\}$?

Pour répondre, imaginer que vous lancez continuellement une pièce ayant pour probabilité p de tomber sur face. Si le second face apparaît au n -ième lancer, quelle est la fonction de distribution de l'apparition du premier face?

- b) Vérifier votre conjecture de la partie a).

6.7.15 Si X et Y sont des variables aléatoires binomiales indépendantes et de même paramètres n et p , montrer analytiquement que la distribution conditionnelle de X , étant donné que $X + Y = m$, est une loi hypergéométrique. Donner également un second argument qui permette d'obtenir le même résultat, mais sans calcul.

A titre d'indication, supposons que l'on jette $2n$ fois une pièce de monnaie. Soit X le nombre de piles dans les n premiers jets et Y le nombre de piles dans les n jets suivants. Montrer que, le total des piles étant fixé à m , le nombre de piles dans les n premiers jets a la même distribution que le nombre de boules blanches obtenues lorsqu'un échantillon de taille m est tiré parmi n boules blanches et n boules noires.

6.7.16 Considérer une expérience pouvant aboutir à trois résultats et où le résultat i apparaît avec une probabilité p_i , $i = 1, 2, 3$. Supposons que cette expérience soit effectuée n fois de façon indépendante et soit X_i , $i = 1, 2, 3$ le nombre de fois où le résultat i se produit. Déterminer la loi de probabilité conditionnelle de X_1 , étant donné que $X_2 = m$.

6.7.17 Soit X_1, X_2, X_3 trois variables aléatoires continues indépendantes et identiquement distribuées. Calculer:

- $P\{X_1 > X_2 | X_1 > X_3\}$;
- $P\{X_1 > X_2 | X_1 < X_3\}$;
- $P\{X_1 > X_2 | X_2 > X_3\}$;
- $P\{X_1 > X_2 | X_2 < X_3\}$.

6.7.18 Soit U une variable aléatoire uniformément distribuée sur l'intervalle $(0, 1)$. Calculer la distribution conditionnelle de U étant donné que:

- $U > a$;
- $U < a$

où $0 < a < 1$.

6.7.19 Supposons que W , le taux d'humidité de l'air un jour donné, soit une variable aléatoire gamma de paramètres (t, β) . Cela veut dire que sa densité est $f(w) = \beta e^{-\beta w} (\beta w)^{t-1} / \Gamma(t)$, $w > 0$. Supposons également qu'étant donné que $W = w$, le nombre d'accidents durant ce jour – appelons le N – suit une distribution de Poisson de moyenne w . Montrer que la distribution conditionnelle de W , étant donné que $N = n$, est la distribution bêta de paramètres $(t + n, \beta + 1)$.

6.7.20 Soit W une variable aléatoire gamma de paramètres (t, β) et supposons que conditionnellement à $W = w$, X_1, \dots, X_n sont des variables aléatoires exponentielles indépendantes de paramètre w . Montrer que la distribution conditionnelle de W , étant donné que $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$, est la distribution gamma de paramètres $(t + n, \beta + \sum_{i=1}^n x_i)$.

6.7.21 Un tableau rectangulaire de mn nombres arrangés en n lignes et m colonnes est dit contenir un *point de selle* s'il y a un nombre qui est à la fois le minimum de sa ligne et le maximum de sa colonne. Par exemple dans le tableau

1	3	2
0	-2	6
.5	12	3

le nombre 1 de la première ligne et de la première colonne est un point de selle. L'existence d'un point de selle revêt une importance dans la théorie des jeux. Considérons un tableau rectangulaire de nombres comme décrit précédemment et supposons que deux personnes A et B jouent au jeu suivant: A choisit un des nombres $1, 2, \dots, n$ et B un des nombres $1, 2, \dots, m$. Ces choix sont annoncés simultanément; si A choisit

i et B choisit j , alors A reçoit de B la somme spécifiée par le nombre situé à la i -ème ligne et j -ème colonne du tableau. Supposons maintenant que le tableau contienne un point de selle – disons le nombre se trouvant à la ligne r et à la colonne k – et appelons ce nombre x_{rk} . Si le joueur A choisit la ligne r , alors il peut être sûr de réaliser un gain au moins égal à x_{rk} (puisque x_{rk} est le nombre minimum de la ligne r). D'autre part, si le joueur B choisit la colonne k , alors il peut être sûr qu'il ne perdra pas plus que x_{rk} (puisque x_{rk} est le nombre maximum de la colonne k). Ainsi, comme A a une possibilité de jeu qui lui assure un gain de x_{rk} , et comme B a une possibilité de jeu lui garantissant qu'il ne perdra pas plus que x_{rk} , il semble raisonnable de considérer ces deux stratégies comme optimales et de déclarer que la valeur du jeu pour le joueur A est x_{rk} .

Si les nm nombres du tableau rectangulaire décrit ci-dessus sont tirés de manière indépendante d'une distribution continue quelconque, quelle est la probabilité que le tableau obtenu contienne un point de selle?

6.7.22 On dit que les variables aléatoires X et Y ont une distribution normale bivariée si leur fonction de densité conjointe est donnée par:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}$$

- Montrer que la densité conditionnelle de X , étant donné que $Y = y$, est la densité normale de paramètres

$$\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y) \quad \text{et} \quad \sigma_x^2(1 - \rho^2)$$

- Montrer que X et Y sont toutes deux des variables aléatoires normales de paramètres μ_x , σ_x^2 et μ_y , σ_y^2 respectivement.
- Montrer que X et Y sont indépendantes quand $\rho = 0$.

6.7.23 Soit $F(x)$ une fonction de répartition. Montrer que

- $F^n(x)$ et
- $1 - [1 - F(x)]^n$

sont aussi des fonctions de répartition quand n est un entier positif.

A titre d'indication, considérer n variables aléatoires indépendantes X_1, \dots, X_n ayant la même fonction de répartition F . Définir alors des variables aléatoires Y et Z en termes de X_i de telle sorte que $P\{Y \leq x\} = F^n(x)$ et $P\{Z \leq x\} = 1 - [1 - F(x)]^n$.

6.7.24 Montrer que si n personnes sont réparties au hasard le long d'une route de L km, alors la probabilité de ne jamais rencontrer deux personnes situées à une distance inférieure à D km est $[1 - (n-1)D/L]^n$, dans le cas où $D \leq L/(n-1)$. Qu'en est-il si $D > L/(n-1)$?

6.7.25 Etablir l'équation (6.20) en dérivant l'équation (6.22).

6.7.26 Montrer que la médiane d'un échantillon de taille $2n + 1$ provenant d'une distribution uniforme sur l'intervalle $(0, 1)$ a une distribution bêta de paramètres $(n + 1, n + 1)$.

6.7.27 Vérifier l'équation (6.24) qui donne la densité conjointe de $X_{(i)}$ et $X_{(j)}$.

6.7.28 Calculer la densité de l'étendue d'un échantillon de taille n provenant d'une distribution continue de fonction de densité f .

6.7.29 Soient $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ les valeurs ordonnées de n variables aléatoires uniformes sur l'intervalle $(0, 1)$. Prouver que pour $1 \leq k \leq n + 1$

$$P\{X_{(k)} - X_{(k-1)} > t\} = (1 - t)^n$$

où $X_0 \equiv 0, X_{n+1} \equiv 1$.

6.7.30 Soient X_1, \dots, X_n un ensemble de variables aléatoires continues indépendantes et identiquement distribuées selon une fonction de répartition F et soient $X_{(i)}$, $i = 1, \dots, n$ leurs valeurs ordonnées. Si X , indépendant des X_i , $i = 1, \dots, n$, a la même fonction de répartition F , déterminer

- $P\{X > X_{(n)}\}$,
- $P\{X > X_{(i)}\}$ et
- $P\{X_{(i)} < X < X_{(j)}\}$, $1 \leq i < j \leq n$.

6.7.31 Soient X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F et de densité f . La quantité $M \equiv [X_{(1)} + X_{(n)}]/2$, définie comme la moyenne de la plus petite et de la plus grande valeur, est appelée le milieu de l'étendue. Montrer que sa fonction de répartition est:

$$F_M(m) = n \int_{-\infty}^m [F(2m - x) - F(x)]^{n-1} f(x) dx$$

6.7.32 Soient X_1, \dots, X_n des variables aléatoires indépendantes et uniformes sur l'intervalle $(0, 1)$. Soit $R = X_{(n)} - X_{(1)}$ l'étendue et $M = [X_{(n)} + X_{(1)}]/2$ le milieu de l'étendue. Calculer la fonction de densité conjointe de R et M .

6.8 PROBLÈMES

6.8.1 On jette deux dés équilibrés. Trouver la loi de probabilité conjointe de X et Y dans les cas suivants:

- X est la plus grande des deux valeurs obtenues et Y en est la somme;
- X est la valeur obtenue avec le premier dé et Y est la plus grande des deux valeurs;
- X et Y sont respectivement la plus petite et la plus grande des deux valeurs obtenues.

6.8.2 Supposer que 3 balles sont tirées sans remise d'une urne contenant 5 balles blanches et 8 balles rouges. Soit X_i égal à 1 si la i -ème balle sélectionnée est blanche et égal à 0 sinon. Donner la loi de probabilité conjointe de

- a) X_1, X_2 ;
b) X_1, X_2, X_3 .

6.8.3 Dans le problème 6.8.2, supposer que les balles blanches sont numérotées et soit Y_i égal à 1 si la i -ème balle blanche est tirée et égal à 0 sinon. Trouver la loi de probabilité conjointe de

- a) Y_1, Y_2 ;
b) Y_1, Y_2, Y_3 .

6.8.4 Recommencer le problème 6.8.2 lorsque chaque balle tirée est remise dans l'urne avant le tirage suivant.

6.8.5 Recommencer le problème 6.8.3 lorsque chaque balle tirée est remise dans l'urne avant le tirage suivant.

6.8.6 On sait qu'il y a deux transistors défectueux dans un emballage en contenant 5. Les transistors sont testés, l'un après l'autre, jusqu'à ce que les deux éléments défectueux aient été identifiés. Soit N_1 le nombre de tests effectués pour trouver le premier transistor défectueux et soit N_2 le nombre de tests additionnels pour trouver le second transistor défectueux; établir la loi de probabilité conjointe de N_1 et N_2 .

6.8.7 On considère une suite d'épreuves de Bernoulli indépendantes avec une probabilité de succès p pour chacune d'entre elles. Soit X_1 le nombre d'échecs avant le premier succès et X_2 le nombre d'échecs entre le premier et le second succès. Trouver la loi de probabilité simultanée de X_1 et X_2 .

6.8.8 Soit la fonction de densité conjointe de X et Y donnée par:

$$f(x, y) = c(y^2 - x^2)e^{-y} \quad -y \leq x \leq y, 0 < y < \infty$$

- a) Trouver c .
b) Trouver les densités marginales de X et de Y .

6.8.9 Considérons la fonction de densité simultanée de X et Y donnée par:

$$f(x, y) = \frac{6}{7} \left(x^2 + \frac{xy}{2} \right) \quad 0 < x < 1, 0 < y < 2$$

- a) Vérifier que c'est bien là une fonction de densité conjointe.
b) Déterminer la fonction de densité de X .
c) Trouver $P\{X > Y\}$.
d) Trouver $P\{Y > \frac{1}{2} | X < \frac{1}{2}\}$.

6.8.10 La fonction de densité de X et Y est donnée par:

$$f(x, y) = e^{-(x+y)} \quad 0 \leq x < \infty, 0 \leq y < \infty$$

Trouver:

- a) $P\{X < Y\}$;
 b) $P\{X < a\}$.

6.8.11 Le propriétaire d'un magasin de télévision évalue que 45 % des clients entrant dans son magasin achètent un appareil de télévision ordinaire, 15% achètent un appareil de télévision couleur et 40% d'entre eux font juste du lèche-vitrine. Si cinq clients entrent dans son magasin un jour donné, quelle est la probabilité qu'il vende exactement 2 appareils ordinaires et 1 poste TV couleur ce jour-là?

6.8.12 Le nombre de personnes qui entrent dans un magasin durant une heure donnée est une variable aléatoire de Poisson de paramètre $\lambda = 10$. Déterminer la probabilité conditionnelle qu'au plus 3 hommes entrent dans ce magasin, étant donné que 10 femmes y sont entrées durant cette heure-là. Quelles hypothèses faites-vous?

6.8.13 Un homme et une femme se sont donnés rendez-vous à un endroit donné à 12 h 30 environ. Si l'homme arrive entre 12 h 15 et 12 h 45, et si la femme arrive indépendamment à une heure uniformément distribuée entre 12 h 00 et 13 h 00, trouver la probabilité que le premier arrivé n'attende pas plus de 5 minutes. Quelle est la probabilité que l'homme arrive le premier?

6.8.14 Une ambulance fait la navette à vitesse constante le long d'une route de longueur L . A un certain moment, un accident se produit en un point aléatoire qui est uniformément distribué sur la route (c'est-à-dire que la distance de ce point à une extrémité de la route servant de référence est uniformément distribuée sur l'intervalle $(0, L)$). En supposant que l'emplacement de l'ambulance, au moment de l'accident, est aussi uniformément distribué, calculer, en admettant les hypothèses d'indépendance nécessaires, la distribution de la distance de l'ambulance au point de l'accident.

6.8.15 Le vecteur aléatoire (X, Y) est distribué uniformément dans une région R du plan si, pour une constante c , sa densité conjointe est

$$f(x, y) = \begin{cases} c & \text{si } (x, y) \in R \\ 0 & \text{sinon} \end{cases}$$

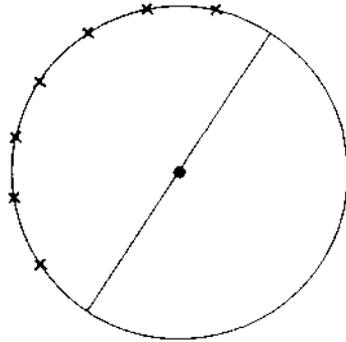
a) Montrer que $1/c = \text{aire de la région } R$.

Supposer que (X, Y) est uniformément distribué sur le carré de centre $(0, 0)$ et de longueur d'arête 2.

b) Montrer que X et Y sont indépendants, chacun étant distribué uniformément sur $(-1, 1)$.

c) Quelle est la probabilité que (X, Y) se trouve dans le cercle de rayon 1 centré à l'origine? Autrement dit, trouver $P\{X^2 + Y^2 \leq 1\}$.

6.8.16 Supposer que n points sont indépendamment choisis au hasard sur le périmètre d'un cercle et qu'on cherche la probabilité qu'ils se trouvent tous sur le même demi-cercle. (Autrement dit, on cherche la probabilité qu'il y ait une ligne passant par le centre du cercle de telle sorte que tous les points choisis soient du même côté de cette ligne.)



Soient P_1, \dots, P_n les n points choisis. Soit A l'événement «tous les points sont situés sur un demi-cercle», et soit A_i l'événement «tous les points sont dans le demi-cercle commençant au point P_i et partant dans le sens des aiguilles d'une montre à 180° , $i = 1, \dots, n$ ».

- Exprimer A en termes de A_i .
- Les A_i sont-ils mutuellement exclusifs?
- Trouver $P\{A\}$.

6.8.17 Trois points X_1, X_2, X_3 sont choisis au hasard sur une droite de longueur L . Quelle est la probabilité que X_2 se trouve entre X_1 et X_3 ?

6.8.18 Deux points sont choisis sur un segment de longueur L , de manière à ce qu'ils soient de part et d'autre du milieu du segment. En d'autres termes, les deux points X et Y sont des variables aléatoires indépendantes telles que X soit uniformément distribué sur $(0, L/2)$ et Y soit uniformément distribué sur $(L/2, L)$. Trouver la probabilité que la distance entre les deux points soit plus grande que $L/3$.

6.8.19 Dans 6.8.18, trouver la probabilité que les trois segments de droite, de 0 à X , de X à Y et de Y à L , puissent constituer les trois côtés d'un triangle (noter que trois segments de droite peuvent former un triangle si la longueur de chacun d'entre eux est inférieure à la somme des longueurs des deux autres).

6.8.20 Soit la densité conjointe de X et Y donnée par:

$$f(x, y) = \begin{cases} xe^{-(x+y)} & x > 0, y > 0 \\ 0 & \text{sinon} \end{cases}$$

X et Y sont-elles indépendantes? Qu'en est-il si f est donnée par:

$$f(x, y) = \begin{cases} 2 & 0 < x < y, 0 < y < 1 \\ 0 & \text{sinon} \end{cases}$$

6.8.21 Supposons que 10^6 personnes arrivent à une station-service à des temps qui sont des variables aléatoires indépendantes, chacune de ces variables étant uniformément

distribuée sur l'intervalle $(0, 10^6)$. Soit N le nombre de personnes qui arrivent pendant la première heure. Trouver une approximation pour $P\{N = i\}$.

6.8.22 Supposons que A, B, C sont des variables aléatoires indépendantes, chacune étant uniformément distribuée sur l'intervalle $(0, 1)$.

- Quelle est la fonction de répartition conjointe de A, B, C ?
- Quelle est la probabilité que toutes les racines de l'équation $Ax^2 + Bx + C = 0$ soient réelles?

6.8.23 Si X est uniformément distribuée sur l'intervalle $(0, 1)$ et Y exponentiellement distribuée avec un paramètre $\lambda = 1$, trouver la distribution de:

- $Z = X + Y$ et de
- $Z = X/Y$.

Supposer l'indépendance de X et de Y .

6.8.24 Si X_1 et X_2 sont des variables aléatoires exponentielles indépendantes avec paramètres respectifs λ_1 et λ_2 , trouver la distribution de $Z = X_1/X_2$. Calculer aussi $P\{X_1 < X_2\}$.

6.8.25 Quand un courant I (mesuré en ampères) traverse une résistance R (mesurée en ohms), la puissance dégagée est donnée par $W = RI^2$ (mesurée en watts). Supposons que I et R soient des variables aléatoires indépendantes de densité:

$$\begin{cases} f_I(x) = 6x(1-x) & 0 \leq x \leq 1 \\ f_R(x) = 2x & 0 \leq x \leq 1 \end{cases}$$

Déterminer la densité de W .

6.8.26 L'espérance du nombre d'erreurs typographiques sur une page d'un magazine est 2. Quelle est la probabilité qu'un article de 10 pages contienne a) 0 et b) 2 ou plus erreurs typographiques? Expliquer votre raisonnement!

6.8.27 Le nombre moyen d'accidents d'avion par mois dans le monde est 2.2. Quelle est la probabilité qu'il y ait

- plus de 2 accidents le mois prochain;
- plus de 4 accidents les deux prochains mois;
- plus de 5 accidents les trois prochains mois?

Expliquer votre raisonnement!

6.8.28 La recette hebdomadaire d'un restaurant est une variable aléatoire normale de moyenne \$2200 et d'écart-type \$230. Quelle est la probabilité que

- la recette totale des deux prochaines semaines dépasse \$5000;
- la recette hebdomadaire dépasse \$2000 lors d'au moins deux des trois prochaines semaines?

Quelle hypothèse d'indépendance avez-vous faite?

6.8.29 Dans le problème 6.8.2, calculer la loi de probabilité conditionnelle de X_1 sachant que

- a) $X_2 = 1$;
- b) $X_2 = 0$.

6.8.30 Dans le problème 6.8.4, calculer la loi de probabilité conditionnelle de X_1 sachant que

- a) $X_2 = 1$;
- b) $X_2 = 0$.

6.8.31 Dans le problème 6.8.3, calculer la loi de probabilité conditionnelle de Y_1 sachant que

- a) $Y_2 = 1$;
- b) $Y_2 = 0$.

6.8.32 Dans le problème 6.8.5, calculer la loi de probabilité conditionnelle de Y_1 sachant que

- a) $Y_2 = 1$;
- b) $Y_2 = 0$.

6.8.33 Choisissons un nombre X au hasard dans l'ensemble des nombres $\{1, 2, 3, 4, 5\}$. Puis choisissons au hasard un nombre du sous-ensemble $\{1, 2, \dots, X\}$. Appelons Y ce second nombre.

- a) Trouver la loi de probabilité simultanée de X et Y .
- b) Trouver la loi de probabilité conditionnelle de X , étant donné que $Y = i$. Le faire pour $i = 1, 2, 3, 4, 5$.
- c) X et Y sont-elles indépendantes? Pourquoi?

6.8.34 On jette deux dés. Soient X et Y respectivement la plus grande et la plus petite des valeurs obtenues. Calculer la loi de probabilité conditionnelle de Y , étant donné que $X = i$ pour $i = 1, 2, \dots, 6$. X et Y sont-elles indépendantes? Pourquoi?

6.8.35 La loi de probabilité conjointe de X et Y est donnée par:

$$\begin{array}{ll} p(1, 1) = \frac{1}{8} & p(1, 2) = \frac{1}{4} \\ p(2, 1) = \frac{1}{8} & p(2, 2) = \frac{1}{2} \end{array}$$

- a) Calculer la loi de probabilité conditionnelle de X étant donné que $Y = i, i = 1, 2$.
- b) X et Y sont-elles indépendantes?
- c) Calculer $P\{XY \leq 3\}$, $P\{X + Y > 2\}$, $P\{X/Y > 1\}$.

6.8.36 La densité conjointe de X et Y est donnée par

$$f(x, y) = xe^{-x(y+1)} \quad x > 0, y > 0$$

- a) Trouver la densité conditionnelle de X , étant donné que $Y = y$, et celle de Y , étant donné que $X = x$.
- b) Trouver la densité de $Z = XY$.

6.8.37 La densité conjointe de X et Y est:

$$f(x, y) = c(x^2 - y^2)e^{-x} \quad 0 \leq x < \infty, -x \leq y \leq x$$

Trouver la distribution conditionnelle de Y , étant donné que $X = x$.

6.8.38 Si X_1, X_2, X_3 sont des variables aléatoires indépendantes uniformément distribuées sur l'intervalle (a, b) , calculer la probabilité que la plus grande des trois soit plus élevée que la somme des deux autres.

6.8.39 Une machine complexe est effectivement opérationnelle tant que trois au moins de ses cinq moteurs fonctionnent. Si pour chaque moteur la durée de fonctionnement est une variable aléatoire indépendante, dont la densité est $f(x) = xe^{-x}, x > 0$, calculer la densité du temps de fonctionnement de la machine.

6.8.40 Si trois camions tombent en panne en des endroits aléatoirement distribués sur une route de longueur L , trouver la probabilité que chaque camion soit à une distance supérieure à d des deux autres lorsque $d \leq L/2$.

6.8.41 Considérons un échantillon de taille 5 issu d'une distribution uniforme sur l'intervalle $(0, 1)$. Calculer la probabilité que la médiane se trouve dans l'intervalle $(\frac{1}{4}, \frac{3}{4})$.

6.8.42 Si X_1, X_2, X_3, X_4, X_5 sont des variables aléatoires exponentielles indépendantes, identiquement distribuées et de paramètre λ , calculer:

- $P\{\min(X_1, \dots, X_5) \leq a\}$ et
- $P\{\max(X_1, \dots, X_5) \leq a\}$.

6.8.43 Déterminer la distribution de l'étendue d'un échantillon de taille 2 provenant d'une distribution dont la densité est $f(x) = 2x, 0 < x < 1$.

6.8.44 Soient X et Y les coordonnées d'un point choisi de façon uniforme dans un cercle de rayon 1 centré à l'origine, c'est-à-dire que leur densité conjointe est:

$$f(x, y) = \frac{1}{\pi} \quad x^2 + y^2 \leq 1$$

Trouver la densité simultanée des coordonnées polaires $R = (X^2 + Y^2)^{1/2}$ et $\theta = \text{Arc tg}(Y/X)$.

6.8.45 Si X et Y sont des variables aléatoires indépendantes toutes deux uniformément distribuées sur l'intervalle $(0, 1)$, trouver la densité conjointe de $R = \sqrt{X^2 + Y^2}$ et $\theta = \text{Arc tg}(Y/X)$.

6.8.46 Si U est uniforme sur l'intervalle $(0, 2\pi)$ et Z , indépendante de U , est exponentielle de paramètre 1, montrer directement (sans utiliser les résultats de l'exemple 6.27) que X et Y définies par:

$$\begin{cases} X = \sqrt{2Z} \cos U \\ Y = \sqrt{2Z} \sin U \end{cases}$$

sont des variables aléatoires normales centrées réduites.

6.8.47 Si X et Y ont pour densité simultanée

$$f(x, y) = \frac{1}{x^2 y^2} \quad x \geq 1, y \geq 1$$

- a) Calculer la densité conjointe de $U = XY$, $V = X/Y$.
 b) Quelles sont leurs densités marginales?

6.8.48 Si X et Y sont des variables aléatoires indépendantes de distribution uniforme sur l'intervalle $(0, 1)$, déterminer la densité conjointe de

- a) $U = X + Y$, $V = X/Y$;
 b) $U = X$, $V = X/Y$;
 c) $U = X + Y$, $V = X/(X + Y)$.

6.8.49 Refaire le problème 6.8.48 quand X et Y sont des variables aléatoires indépendantes exponentielles, chacune de paramètre $\lambda = 1$.

6.8.50 Si X_1 et X_2 sont des variables aléatoires indépendantes exponentielles, chacune de paramètre λ , trouver la densité simultanée de $Y_1 = X_1 + X_2$ et $Y_2 = e^{X_1}$.

6.8.51 Si X , Y et Z sont des variables aléatoires indépendantes et identiquement distribuées de densité $f(x) = e^{-x}$, $0 < x < \infty$, déterminer la distribution conjointe de $U = X + Y$, $V = X + Z$ et $W = Y + Z$.

6.8.52 Dans l'exemple 6.30, montrer que Y_2 et Y_3 ont une distribution normale bivariée.

6.8.53 Les âges de futurs parents contrôlés dans un hôpital sont approximativement distribués suivant une loi normale bivariée de paramètres $\mu_x = 28,4$, $\sigma_x = 6,8$, $\mu_y = 31,6$, $\sigma_y = 7,4$ et $\rho = 0,82$. (Les paramètres avec l'indice x se réfèrent à l'âge de la future mère et ceux indicés par y à l'âge du futur père.) En utilisant les résultats de l'exercice théorique 6.7.22, déterminer

- a) la proportion de femmes enceintes âgées de plus de 30 ans,
 b) la proportion de futurs pères âgés de 35 ans dont la femme a plus de 30 ans.

Propriétés de l'espérance

7.1 INTRODUCTION

7.1.1 Rappel

Dans ce chapitre, nous allons développer et utiliser des propriétés supplémentaires de l'espérance. Pour commencer, rappelons que l'espérance d'une variable aléatoire X est définie par

$$E[X] = \sum_x xp(x)$$

lorsque X est une variable aléatoire discrète de fonction de répartition $p(x)$ et par

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

lorsque X est une variable aléatoire continue de densité $f(x)$.

7.1.2 Bornes pour l'espérance

Puisque $E[X]$ est une moyenne pondérée des valeurs possibles de X , alors, si X est compris entre a et b , son espérance l'est aussi.

Si

$$P\{a \leq X \leq b\} = 1$$

alors

$$a \leq E[X] \leq b$$

Pour vérifier ce qui précède, on suppose que X est une variable aléatoire discrète pour laquelle $P\{a \leq X \leq b\} = 1$. Puisque ceci implique que $p(x) = 0$ pour tout x

n'appartenant pas à l'intervalle $[a, b]$, on voit que

$$\begin{aligned} E[X] &= \sum_{x:p(x)>0} xp(x) \\ &\geq \sum_{x:p(x)>0} ap(x) \\ &= a \sum_{x:p(x)>0} p(x) \\ &= a \end{aligned}$$

De la même manière, on peut montrer que $E[X] \leq b$ et le résultat est donc montré pour les variables aléatoires discrètes. Comme la preuve dans le cas continu est similaire, le résultat est démontré.

7.2 ESPÉRANCE D'UNE SOMME DE VARIABLES ALÉATOIRES

7.2.1 Espérance d'une fonction de deux variables

Le théorème suivant est un équivalent bidimensionnel des théorèmes 4.1 du chapitre 4 et 5.1 du chapitre 5, qui donnent les formules pour calculer l'espérance d'une fonction d'une variable aléatoire. Supposons que X et Y sont des variables aléatoires et que g est une fonction de deux variables. Nous avons alors le résultat suivant.

Théorème 7.1

Si X et Y sont discrètes et ont pour loi $p(x, y)$ alors

$$E[g(X, Y)] = \sum_y \sum_x g(x, y)p(x, y)$$

Si X et Y sont continues et ont pour densité conjointe $f(x, y)$ alors

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)dx dy$$

7.2.2 Exemple d'espérance d'une fonction de deux variables

Exemple 7.1 Un accident se produit en un point X uniformément distribué sur une route de longueur L . Au moment de l'accident, une ambulance se trouve en un point Y lui aussi uniformément distribué sur la route. En supposant que X et Y sont indépendants, trouver l'espérance de la distance entre l'ambulance et le lieu de l'accident.

SOLUTION. Nous devons calculer $E[|X - Y|]$. Puisque la densité conjointe de X et Y est

$$f(x, y) = \frac{1}{L^2}, \quad 0 < x < L, \quad 0 < y < L$$

le théorème 7.1 donne

$$E[|X - Y|] = \frac{1}{L^2} \int_0^L \int_0^L |x - y| dy dx$$

On a

$$\begin{aligned} \int_0^L |x - y| dy &= \int_0^x (x - y) dy + \int_x^L (y - x) dy \\ &= \frac{x^2}{2} + \frac{L^2}{2} - \frac{x^2}{2} - x(L - x) \\ &= \frac{L^2}{2} + x^2 - xL \end{aligned}$$

Donc

$$\begin{aligned} E[|X - Y|] &= \frac{1}{L^2} \int_0^L \left(\frac{L^2}{2} + x^2 - xL \right) dx \\ &= \frac{L}{3} \end{aligned}$$

■

7.2.3 Application au cas d'une somme de variables

Admettons que deux variables X et Y ont des espérances $E[X]$ et $E[Y]$ finies. On pose $g(X, Y) = X + Y$. En appliquant le théorème 7.1, on obtient, dans le cas continu,

$$\begin{aligned} E[X + Y] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dy dx + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= E[X] + E[Y] \end{aligned}$$

Ce résultat est valable de manière tout à fait générale; on peut écrire

$$E[X + Y] = E[X] + E[Y] \quad (7.1)$$

dès que $E[X]$ et $E[Y]$ sont finies.

Exemple 7.2 Supposer que pour des variables aléatoires X et Y ,

$$X \geq Y$$

Ceci signifie que pour tout résultat d'une expérience, la valeur de la variable X est supérieure ou égale à celle de la variable Y . Puisque l'inégalité précédente est équivalente à $X - Y \geq 0$, on a $E[X - Y] \geq 0$ ou, de manière équivalente,

$$E[X] \geq E[Y] \quad \blacksquare$$

En utilisant l'équation (7.1), une démonstration par induction permet d'établir sans difficulté que si $E[X_i]$ est finie pour tout $i = 1, \dots, n$, alors

$$E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n] \quad (7.2)$$

L'équation (7.2) est d'une grande utilité comme le montrent les exemples du paragraphe suivant.

7.2.4 Exemples de calculs d'espérances de sommes

Exemple 7.3 Moyenne d'un échantillon

Soient X_1, \dots, X_n des variables aléatoires indépendantes, identiquement distribuées selon une fonction de distribution F et d'espérance μ . Une telle séquence de variables constitue un échantillon de la distribution F . La quantité \bar{X} , définie par

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$$

est appelée la moyenne de l'échantillon. Calculer $E[\bar{X}]$.

SOLUTION.

$$\begin{aligned} E[\bar{X}] &= E\left[\sum_{i=1}^n \frac{X_i}{n}\right] \\ &= \frac{1}{n} E\left[\sum_{i=1}^n X_i\right] \\ &= \frac{1}{n} \sum_{i=1}^n E[X_i] \\ &= \mu \quad \text{puisque } E[X_i] = \mu \end{aligned}$$

L'espérance de la moyenne de l'échantillon est donc μ , la moyenne de la distribution. Lorsque la moyenne μ de la distribution est inconnue, la moyenne de l'échantillon est souvent utilisée en statistiques pour l'estimer. \blacksquare

Exemple 7.4 Inégalité de Boole

Soient A_1, \dots, A_n des événements et définissons les variables indicatrices X_i , $i = 1, \dots, n$ par

$$X_i = \begin{cases} 1 & \text{si } A_i \text{ se produit} \\ 0 & \text{sinon} \end{cases}$$

Soit $X = \sum_{i=1}^n X_i$.

Donc X représente le nombre d'événements A_i qui se produisent. Finalement, soit

$$Y = \begin{cases} 1 & \text{si } X \geq 1 \\ 0 & \text{sinon} \end{cases}$$

Donc Y est égal à 1 si au moins un des A_i se produit et égal à 0 sinon. Il est clair que

$$X \geq Y$$

Donc

$$E[X] \geq E[Y]$$

Mais puisque

$$E[X] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n P(A_i)$$

et

$$E[Y] = P\{\text{au moins un des } A_i \text{ se produit}\} = P\left\{\bigcup_{i=1}^n A_i\right\}$$

on obtient l'inégalité de Boole, soit

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$$

■

Les trois prochains exemples montrent comment on peut utiliser l'équation (7.2) pour calculer l'espérance des variables aléatoires binomiale, binomiale négative et hypergéométrique. Ces méthodes peuvent être comparées à celles présentées au chapitre 4.

Exemple 7.5 Espérance d'une variable aléatoire binomiale

Soit X une variable aléatoire binomiale de paramètres n et p . On se souvient qu'une telle variable X représente le nombre de succès lors de la réalisation de n épreuves indépendantes, chaque épreuve débouchant sur un succès avec la même probabilité p . On peut donc écrire

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

où

$$X_i = \begin{cases} 1 & \text{si la } i\text{-ième épreuve est un succès} \\ 0 & \text{si la } i\text{-ième épreuve est un échec} \end{cases}$$

Chaque variable X_i est une variable de Bernoulli ayant par conséquent pour espérance $E[X_i] = 1(p) + 0(1-p) = p$. D'où

$$E[X] = E[X_1] + E[X_2] + \dots + E[X_n] = np$$

■

Exemple 7.6 Espérance d'une variable aléatoire binomiale négative

On réalise une séquence d'épreuves indépendantes, chacune ayant la probabilité p d'aboutir à un succès. Le nombre d'épreuves à réaliser pour obtenir le r -ième succès est une variable aléatoire suivant une distribution binomiale négative. On cherche l'espérance de cette variable.

SOLUTION. Désignons par X cette variable. Elle peut-être représentée par

$$X = X_1 + X_2 + \dots + X_r$$

où X_1 est le nombre d'épreuves nécessaires à l'obtention du premier succès, X_2 le nombre supplémentaire d'épreuves pour obtenir un deuxième succès, X_3 celui des épreuves à ajouter pour avoir un troisième succès, etc. De manière générale, X_i représente le nombre d'épreuves supplémentaires nécessaires à partir du $(i-1)$ -ième succès pour obtenir le i -ième. Une courte réflexion permet de voir que chacune des variables X_i suit une loi géométrique de paramètre p . D'après les résultats obtenus dans l'exemple 4.31 du chapitre 4, $E[X_i] = 1/p$, $i = 1, 2, \dots, r$. Donc

$$E[X] = E[X_1] + E[X_2] + \dots + E[X_r] = \frac{r}{p}$$

■

Exemple 7.7 Espérance d'une variable aléatoire hypergéométrique

Si n boules sont tirées au hasard d'une urne contenant N boules dont m blanches, trouver l'espérance du nombre de boules blanches tirées.

SOLUTION. Soit X le nombre de boules blanches tirées et représentons X par

$$X = X_1 + X_2 + \dots + X_m$$

où

$$X_i = \begin{cases} 1 & \text{si la } i\text{-ième boule blanche a été tirée} \\ 0 & \text{si elle ne l'a pas été} \end{cases}$$

Or

$$\begin{aligned} E[X_i] &= P\{X_i = 1\} \\ &= P\{\text{la } i\text{-ième boule blanche a été tirée}\} \\ &= \frac{\binom{1}{1} \binom{N-1}{n-1}}{\binom{N}{n}} \\ &= \frac{n}{N} \end{aligned}$$

Et donc

$$E[X] = E[X_1] + \dots + E[X_m] = \frac{mn}{N}$$

On aurait pu obtenir ce résultat en utilisant une autre décomposition

$$X = Y_1 + \dots + Y_n$$

où

$$Y_i = \begin{cases} 1 & \text{si la } i\text{-ième boule tirée est blanche} \\ 0 & \text{sinon} \end{cases}$$

Or les chances pour chacune des N boules d'être tirée en i -ième position sont les mêmes. Par conséquent,

$$E[Y_i] = \frac{m}{N}$$

et par suite

$$E[X] = E[Y_1] + \dots + E[Y_n] = \frac{nm}{N}$$

Exemple 7.8 Espérance du nombre de rencontres

Chacun des N hommes d'une assemblée jette son chapeau au milieu de la pièce. On mélange les chapeaux et chacun en ramasse un au hasard. On veut savoir le nombre moyen de rencontres, c'est-à-dire d'hommes ayant récupéré leur propre chapeau.

SOLUTION. Désignons le nombre de rencontres par X . Le moyen le plus simple de calculer $E[X]$ consiste à écrire

$$X = X_1 + X_2 + \dots + X_N$$

où

$$X_i = \begin{cases} 1 & \text{si le } i\text{-ième homme ramasse son chapeau} \\ 0 & \text{s'il ramasse celui de quelqu'un d'autre} \end{cases}$$

Comme le i -ième homme a autant de chances de ramasser n'importe lequel des N chapeaux,

$$E[X_i] = P\{X_i = 1\} = \frac{1}{N}$$

pour tout $i = 1, 2, \dots, N$. Par conséquent

$$E[X] = E[X_1] + \dots + E[X_N] = \left(\frac{1}{N}\right)N = 1$$

On conclut qu'en moyenne un participant seulement aura ramassé son propre chapeau. ■

Exemple 7.9 Le problème ci-dessous fut posé pour la première fois et résolu par Daniel Bernoulli au 18^e siècle. On suppose qu'une urne contient $2N$ cartes, deux d'entre elles portant le numéro 1, deux autres le 2, deux autres le 3, etc. On tire m cartes au hasard. Quel est le nombre moyen de paires encore présentes dans l'urne après ce tirage? (Il est intéressant de savoir que Bernoulli a proposé ce modèle comme l'un de ceux permettant de déterminer combien de couples mariés il reste après la mort de m personnes dans un groupe composé exclusivement de couples au départ, au nombre de N).

SOLUTION. On définit pour $i = 1, 2, \dots, N$

$$X_i = \begin{cases} 1 & \text{si la } i\text{-ème paire est intacte} \\ 0 & \text{si elle a disparu, totalement ou en partie} \end{cases}$$

Or,

$$\begin{aligned} E[X_i] &= P\{X_i = 1\} \\ &= \frac{\binom{2N-2}{m}}{\binom{2N}{m}} \\ &= \frac{(2N-2)!}{m!(2N-2-m)!} \\ &= \frac{(2N-2)!}{(2N)!} \\ &= \frac{m!(2N-m)!}{m!(2N-m)!} \\ &= \frac{(2N-m)(2N-m-1)}{(2N)(2N-1)} \end{aligned}$$

aussi le résultat cherché est-il

$$\begin{aligned} E[X_1 + X_2 + \dots + X_N] &= E[X_1] + \dots + E[X_N] \\ &= \frac{(2N-m)(2N-m-1)}{2(2N-1)} \quad \blacksquare \end{aligned}$$

Exemple 7.10 Problème de collection de bons à compléter

Il existe N sortes de bons différents, aucune sorte n'étant plus rare que les autres. Le but du jeu est de rassembler une collection complète où les N sortes soient représentées.

- On se procure un lot de n bons. Il faut d'abord trouver le nombre moyen de sortes qui y seront représentées;
- on cherche encore le nombre moyen de bons à amasser pour obtenir une collection complète.

SOLUTION.

- a) Soit X le nombre de sortes différentes représentées dans le lot de taille n considéré. On calcule $E[X]$ en utilisant la décomposition

$$X = X_1 + \cdots + X_N$$

où

$$X_i = \begin{cases} 1 & \text{si la sorte } i \text{ est représentée dans le lot} \\ 0 & \text{si elle ne l'est pas} \end{cases}$$

Or

$$\begin{aligned} E[X_i] &= P\{X_i = 1\} \\ &= 1 - P\{\text{la sorte } i \text{ n'est pas représentée dans le lot}\} \\ &= 1 - \left(\frac{N-1}{N}\right)^n \end{aligned}$$

Donc

$$E[X] = E[X_1] + \cdots + E[X_N] = N \left[1 - \left(\frac{N-1}{N}\right)^n \right]$$

- b) Désignons par Y la variable qui compte le nombre de bons à amasser pour former une collection complète. On calcule $E[Y]$ en utilisant la même technique que celle appliquée au calcul de l'espérance d'une variable binomiale négative (exemple 7.6). En l'occurrence, on appelle Y_i , $i = 0, 1, \dots, N-1$ le nombre de bons à ajouter, une fois que i sortes sont représentées, pour faire apparaître une sorte de plus. On a bien sûr

$$Y = Y_0 + Y_1 + \cdots + Y_{N-1}$$

Lorsque i sortes sont déjà représentées, un nouveau bon sera d'une nouvelle sorte avec probabilité $(N-i)/N$. Par conséquent,

$$P\{Y_i = k\} = \frac{N-i}{N} \left(\frac{i}{N}\right)^{k-1} \quad k \geq 1$$

En d'autres termes, Y_i est une variable géométrique de paramètre $(N-i)/N$. D'après l'exemple 7.5, on sait déjà que

$$E[Y_i] = \frac{N}{N-i}$$

ce qui entraîne

$$E[Y] = 1 + \frac{N}{N-1} + \frac{N}{N-2} + \cdots + \frac{N}{1} = N \left[1 + \cdots + \frac{1}{N-1} + \frac{1}{N} \right] \quad \blacksquare$$

Exemple 7.11 Dix chasseurs guettent le passage d'un vol de canards. Lorsque les canards passent en groupe, les chasseurs font tous feu en même temps, mais chacun choisit sa cible au hasard, indépendamment des autres. On admet que chaque chasseur touche son canard avec la même probabilité p . Combien de canards survivront-ils au tir lorsque le vol se compose de 10 oiseaux?

SOLUTION. Disons que la variable X_i vaudra 1 si le i -ème canard survit et 0 sinon, $i = 1, 2, \dots, 10$. Le nombre moyen de canards épargnés sera

$$E[X_1 + \dots + X_{10}] = E[X_1] + \dots + E[X_{10}]$$

Pour le calcul de $E[X_i] = P\{X_i = 1\}$, on remarque que chacun des tireurs atteindra indépendamment des autres le i -ème canard avec la probabilité $p/10$. Aussi,

$$P\{X_i = 1\} = \left(1 - \frac{p}{10}\right)^{10}$$

Et donc

$$E[X] = 10 \left(1 - \frac{p}{10}\right)^{10} \quad \blacksquare$$

Exemple 7.12 Nombre moyen de chaînes

On considère les diverses permutations composées de n caractères «1» et de m caractères «0». On admet que l'une d'entre elles est tirée au hasard, si bien que chacune des $(n+m)!/(n!m!)$ permutations distinguables a la même probabilité de sortir. Toute succession ininterrompue de 1 sera appelée chaîne de 1. Si par exemple $n = 6$ et $m = 4$, et si l'ordre de tirage est 1, 1, 1, 0, 1, 1, 0, 0, 1, 0, nous serons en présence de trois chaînes de 1. On souhaite ici déterminer le nombre moyen de ces chaînes.

SOLUTION. Pour ce faire on peut poser

$$I_i = \begin{cases} 1 & \text{si une chaîne de 1 commence au } i\text{-ème caractère} \\ 0 & \text{sinon} \end{cases}$$

On peut donc écrire $C(1)$, le nombre de chaînes de 1, de la manière suivante:

$$C(1) = \sum_{i=1}^{n+m} I_i$$

et par conséquent

$$E[C(1)] = \sum_{i=1}^{n+m} E[I_i]$$

Or

$$\begin{aligned} E[I_1] &= P\{\text{le premier caractère est un 1}\} \\ &= \frac{n}{n+m} \end{aligned}$$

et pour $1 < i \leq n + m$

$$\begin{aligned} E[I_i] &= P\{0 \text{ occupe la } (i-1)\text{-ième position et } 1 \text{ la } i\text{-ème}\} \\ &= \frac{m}{n+m} \frac{n}{n+m-1} \end{aligned}$$

Donc

$$E[C(1)] = \frac{n}{n+m} + (n+m-1) \frac{nm}{(n+m)(n+m-1)}$$

De manière similaire, $E[C(0)]$, le nombre moyen de séquences de 0, sera

$$E[C(0)] = \frac{m}{n+m} + \frac{nm}{n+m}$$

et le nombre moyen de séquences de tous types sera

$$E[C(1) + C(0)] = 1 + \frac{2nm}{n+m} \quad \blacksquare$$

Exemple 7.13 On pose une à une les cartes d'un jeu ordinaire sur une table, ouvertes et côte à côte. On se demande combien il faut poser de cartes en moyenne pour obtenir

- le premier as;
- le premier pique.

SOLUTION. Les questions a) et b) ne sont que des cas particuliers du problème plus général suivant: une urne contient n boules blanches et m noires. On prélève ces boules une à une jusqu'à ce que la première boule blanche apparaisse. Si on désigne par X le nombre des boules alors prélevées, quelle est l'espérance de X ?

Pour résoudre la version générale du problème, on va rendre distinguables les boules noires de l'urne en les baptisant N_1, N_2, \dots, N_m . Posons encore

$$X_i = \begin{cases} 1 & \text{si } N_i \text{ est prélevée avant l'apparition de la première boule blanche} \\ 0 & \text{sinon.} \end{cases}$$

On voit facilement que

$$X = 1 + \sum_{i=1}^m X_i$$

Donc

$$E[X] = 1 + \sum_{i=1}^m P\{X_i = 1\}$$

Or, X_i vaudra 1 si la boule N_i est tirée avant toutes les boules blanches. Mais chacune de ces $n+1$ boules (à savoir les n blanches et la boule N_i) a la même probabilité d'être la première dans l'ordre des prélèvements, ce qui permet d'écrire

$$E[X_i] = P\{X_i = 1\} = \frac{1}{n+1}$$

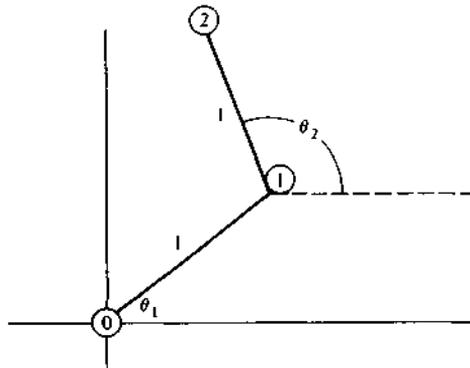
et ainsi

$$E[X] = 1 + \frac{m}{n+1}$$

La réponse à la question a) sera donc, avec $n = 4$ et $m = 48 \cdot \frac{53}{5}$, c'est-à-dire 10,6 cartes en moyenne. Pour b) on aura $n = 13$ et $m = 39$, ce qui donne un nombre moyen de $\frac{53}{14} = 3,79$ cartes avant d'obtenir le premier pique. ■

Exemple 7.14 Déplacement aléatoire dans le plan

On considère une particule située dans un plan et se déplaçant par sauts de longueur fixe mais orientés dans n'importe quelle direction. Plus précisément, on admettra que la longueur des sauts est égale à une unité, tandis que l'angle entre l'axe des abscisses et la direction prise à la suite d'un saut est une variable uniforme sur $(0, 2\pi)$ (voir figure 7.1). On cherche le carré de la distance entre la particule et sa position initiale après n sauts.



- ① = position initiale
- ① = position après le premier saut
- ② = position après le second saut

Figure 7.1

SOLUTION. Désignons les variations de coordonnées associées au i -ème saut par (X_i, Y_i) , $i = 1, 2, \dots, n$. On a les relations

$$\begin{cases} X_i = \cos \theta_i \\ Y_i = \sin \theta_i \end{cases}$$

où les θ_i , $i = 1, 2, \dots, n$ sont par hypothèse des variables uniformes sur $(0, 2\pi)$. La

position au bout de n sauts aura pour coordonnées $(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i)$. On voit donc que la grandeur D^2 cherchée, le carré de la distance de la particule à l'origine, est

$$\begin{aligned} D^2 &= \left(\sum_{i=1}^n X_i \right)^2 + \left(\sum_{i=1}^n Y_i \right)^2 \\ &= \sum_{i=1}^n (X_i^2 + Y_i^2) + \sum_{i \neq j} (X_i X_j + Y_i Y_j) \\ &= n + \sum_{i \neq j} (\cos \theta_i \cos \theta_j + \sin \theta_i \sin \theta_j) \end{aligned}$$

où l'on a utilisé la relation $\cos^2 \theta_i + \sin^2 \theta_i = 1$. En passant aux espérances, en utilisant l'indépendance de θ_i et θ_j lorsque $i \neq j$, ainsi que les relations

$$\begin{aligned} E[\cos \theta_i] &= \int_0^{2\pi} \cos u \, du = \sin 2\pi - \sin 0 = 0 \\ E[\sin \theta_i] &= \int_0^{2\pi} \sin u \, du = \cos 0 - \cos 2\pi = 0 \end{aligned}$$

on trouve

$$E[D^2] = n \quad \blacksquare$$

7.2.5 Espérance de la somme d'une infinité de variables

Lorsque l'on traite d'un ensemble infini de variables aléatoires X_i , $i \geq 1$, ayant toutes une espérance finie, il n'est pas certain que

$$E \left[\sum_{i=1}^{\infty} X_i \right] = \sum_{i=1}^{\infty} E[X_i] \quad (7.3)$$

Pour déterminer dans quel cas (7.3) est valable, on calcule son premier membre en s'appuyant sur la relation $\sum_{i=1}^{\infty} X_i = \lim_{n \rightarrow \infty} \sum_{i=1}^n X_i$

$$\begin{aligned} E \left[\sum_{i=1}^{\infty} X_i \right] &= E \left[\lim_{n \rightarrow \infty} \sum_{i=1}^n X_i \right] \\ &\stackrel{?}{=} \lim_{n \rightarrow \infty} E \left[\sum_{i=1}^n X_i \right] \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n E[X_i] \\ &= \sum_{i=1}^{\infty} E[X_i] \end{aligned} \quad (7.4)$$

Ces calculs – et par conséquent (7.3) – ne sont fondés que lorsque la permutation de l'espérance et de la prise de limite est correcte à la deuxième ligne du calcul de (7.4). Dans le cas général, cette permutation n'est pas justifiée. Cependant, on peut montrer que dans les deux cas particuliers suivants elle est néanmoins acceptable:

- a) les variables X_i sont toutes à valeurs non négatives, ce qui signifie que $P\{X_i \geq 0\} = 1$;
 b) $\sum_{i=1}^{\infty} E[|X_i|] < \infty$.

Exemple 7.15 Considérons une variable discrète X quelconque à valeurs positives ou nulles. On définit pour tout $i \geq 1$

$$X_i = \begin{cases} 1 & \text{si } X \geq i \\ 0 & \text{si } X < i \end{cases}$$

Alors

$$\begin{aligned} \sum_{i=1}^{\infty} X_i &= \sum_{i=1}^X X_i + \sum_{i=X+1}^{\infty} X_i \\ &= \sum_{i=1}^X 1 + \sum_{i=X+1}^{\infty} 0 \\ &= X \end{aligned}$$

De ce fait, les X_i étant toutes à valeurs non négatives,

$$\begin{aligned} E[X] &= \sum_{i=1}^{\infty} E[X_i] \\ &= \sum_{i=1}^{\infty} P\{X \geq i\} \end{aligned} \tag{7.5}$$

qui est une identité intéressante. ■

Exemple 7.16 On souhaite stocker en mémoire d'ordinateur une liste ordonnée comportant n éléments que nous appellerons $1, 2, \dots, n$. On sait que l'utilisateur de cette liste aura à consulter un élément de la liste par unité de temps. L'élément i sera consulté indépendamment du passé avec probabilité $P(i)$, $i \geq 1$ et $\sum_{i=1}^n P(i) = 1$. Ces probabilités étant supposées connues, quel est l'ordre de stockage qui minimisera le temps d'accès moyen à un élément appelé pour consultation?

SOLUTION. On numérote les éléments de telle manière que $P(1) \geq P(2) \geq \dots \geq P(n)$. Nous allons montrer que $1, 2, \dots, n$ est la permutation optimale. Pour ce faire, désignons par X la position de l'élément consulté. Pour toute permutation envisageable $O = i_1, i_2, \dots, i_n$

$$\begin{aligned}
 P_O\{X \geq k\} &= \sum_{j=k}^n P(i_j) \\
 &\geq \sum_{j=k}^n P(j) \\
 &= P_{1,2,\dots,n}\{X \geq k\}
 \end{aligned}$$

En sommant sur k et en utilisant (7.12), on obtient

$$E_O[X] \geq E_{1,2,\dots,n}[X]$$

ce qui montre bien que disposer les éléments dans l'ordre des probabilités décroissantes minimise effectivement le temps moyen d'accès à un élément demandé. ■

Exemple 7.17 Probabilité d'une réunion d'événements

Soient les événements A_1, \dots, A_n et les variables indicatrices $X_i, i = 1, \dots, n$ définies par

$$X_i = \begin{cases} 1 & \text{si } A_i \text{ a lieu} \\ 0 & \text{sinon} \end{cases}$$

Remarquons que

$$1 - \prod_{i=1}^n (1 - X_i) = \begin{cases} 1 & \text{si } \cup A_i \text{ a lieu} \\ 0 & \text{sinon} \end{cases}$$

Alors

$$E \left[1 - \prod_{i=1}^n (1 - X_i) \right] = P \left(\bigcup_{i=1}^n A_i \right)$$

Le développement du membre de gauche de cette expression donne

$$\begin{aligned}
 P \left(\bigcup_{i=1}^n A_i \right) &= E \left[\sum_{i=1}^n X_i - \sum_{i<j} \sum X_i X_j + \sum_{i<j<k} \sum X_i X_j X_k \right. \\
 &\quad \left. - \dots + (-1)^{n+1} X_1 \dots X_n \right] \tag{7.6}
 \end{aligned}$$

Mais, puisque

$$X_{i_1} X_{i_2} \dots X_{i_k} = \begin{cases} 1 & \text{si } A_{i_1} A_{i_2} \dots A_{i_k} \text{ a lieu} \\ 0 & \text{sinon} \end{cases}$$

on voit que

$$E[X_{i_1} \dots X_{i_k}] = P(A_{i_1} \dots A_{i_k})$$

Ainsi (7.6) n'établit rien d'autre que la formule bien connue pour la réunion d'événements

$$\begin{aligned}
 P(\cup A_i) &= \sum P(A_i) - \sum_{i<j} \sum P(A_i A_j) + \sum_{i<j<k} \sum P(A_i A_j A_k) \\
 &\quad - \dots + (-1)^{n+1} P(A_1 \dots A_n)
 \end{aligned}$$

■

Le dernier exemple de cette section montre d'une autre façon que l'introduction du hasard peut parfois être employée avantageusement.

Exemple 7.18 Un tournoi de n participants se déroule de la manière suivante. Chaque joueur rencontre chaque autre joueur une fois (il y a $\binom{n}{2}$ paires de joueurs donc $\binom{n}{2}$ matchs). A chaque fois, un concurrent gagne et l'autre perd. Supposer que les n joueurs sont initialement numérotés comme joueur 1, joueur 2, etc. La permutation i_1, i_2, \dots, i_n est une permutation hamiltonienne si i_1 bat i_2 , i_2 bat i_3 , ... et i_{n-1} bat i_n . Le problème est de déterminer le plus grand nombre possible de permutations hamiltoniennes.

Par exemple, supposer qu'il y a 3 joueurs. Il est alors facile de voir que si un des joueurs gagne deux fois, il n'y a qu'une seule permutation hamiltonienne (par exemple, si 1 gagne deux fois et si 2 bat 3 alors la seule hamiltonienne est 1, 2, 3); et si chaque joueur gagne une fois, il y a alors trois permutations hamiltoniennes (par exemple, si 1 bat 2, 2 bat 3 et 3 bat 1, alors 1,2,3, 2,3,1 et 3,1,2 sont hamiltoniennes). Donc, lorsque $n = 3$, le plus grand nombre possible de permutations hamiltoniennes est 3.

Bien que le problème des permutations hamiltoniennes n'implique pas les probabilités, nous allons introduire le hasard pour montrer que dans un tel tournoi de n joueurs, $n > 2$, il y a un résultat pour lequel le nombre de permutations hamiltoniennes est supérieur à $n!/2^{n-1}$.

Pour vérifier ceci, supposons que les résultats des $\binom{n}{2}$ matchs sont indépendants et que chacun des deux joueurs a autant de chance de gagner chaque match. Si X est le nombre d'hamiltoniennes à la fin du tournoi, alors X est une variable aléatoire dont l'ensemble des valeurs possibles est constitué de tous les nombres possibles de permutations hamiltoniennes qui peuvent résulter d'un tel tournoi à n participants. Puisqu'au moins une des valeurs possibles d'une variable aléatoire non constante doit dépasser sa moyenne, il doit au moins y avoir un résultat possible du tournoi qui possède plus de $E[X]$ permutations hamiltoniennes. Pour trouver $E[X]$, numérotons les $n!$ permutations et soit, pour $i = 1, \dots, n!$,

$$X_i = \begin{cases} 1 & \text{si la permutation } i \text{ est hamiltonienne} \\ 0 & \text{sinon} \end{cases}$$

On a

$$X = \sum_i X_i$$

donc

$$E[X] = \sum_i E[X_i]$$

Mais

$$\begin{aligned} E[X_i] &= P\{\text{la permutation } i \text{ est hamiltonienne}\} \\ &= \left(\frac{1}{2}\right)^{n-1} \end{aligned}$$

Cette égalité est vraie car la probabilité qu'une permutation i_1, \dots, i_n soit hamiltonienne est, par indépendance, la probabilité que i_1 batte i_2 multipliée par la probabilité que i_2 batte i_3 , etc. On obtient donc

$$E[X] = \frac{n!}{2^{n-1}}$$

Puisque, pour $n > 2$, X n'est pas une variable aléatoire constante, il existe au moins une valeur possible dépassant $n!/2^{n-1}$. ■

7.3 COVARIANCE, VARIANCE DE SOMMES, CORRÉLATION

7.3.1 Espérance d'un produit de variables indépendantes

A titre de préliminaire à cette section 7.3, nous présentons le théorème suivant, établissant que l'espérance du produit de deux variables indépendantes est égale au produit de leurs espérances.

Théorème 7.2

Soient X et Y deux variables aléatoires indépendantes et h et g deux fonctions. Alors

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

DÉMONSTRATION. Plaçons-nous dans le cas où X et Y sont conjointement continues de densité $f(\cdot, \cdot)$. On peut écrire

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f(x, y) \, dx \, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y) \, dx \, dy \\ &= \int_{-\infty}^{\infty} h(y)f_Y(y) \, dy \int_{-\infty}^{\infty} g(x)f_X(x) \, dx \\ &= E[h(Y)]E[g(X)] \end{aligned}$$

et la démonstration est similaire dans le cas discret. ■

De la même manière que l'espérance et la variance d'une unique variable aléatoire nous donnent des informations sur cette variable, la covariance entre deux variables nous donne des informations sur la relation entre ces deux variables.

7.3.2 Covariance

La *covariance* de deux variables aléatoires quelconques X et Y est notée $\text{Cov}(X, Y)$ et est définie par l'expression:

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Le développement du membre de droite donne

$$\begin{aligned} \text{Cov}(X, Y) &= E[XY - E[X]Y - XE[Y] + E[Y]E[X]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

On remarquera qu'en application du théorème 7.2, la covariance de deux variables indépendantes X et Y est nulle. La réciproque n'est cependant pas vraie. On peut donner comme contre-exemple simple le cas des deux variables X et Y suivantes; X est telle que

$$P\{X = 0\} = P\{X = 1\} = P\{X = -1\} = \frac{1}{3}$$

et on définit Y par rapport à X

$$Y = \begin{cases} 0 & \text{si } X \neq 0 \\ 1 & \text{si } X = 0 \end{cases}$$

Or XY est clairement nulle, donc $E[XY]$ aussi. Comme $E[X]$ l'est aussi, il reste

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0$$

Il est pourtant manifeste que X et Y ne sont pas indépendantes.

La proposition suivante énonce plusieurs propriétés de la covariance.

Théorème 7.3

- i) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- ii) $\text{Cov}(X, X) = \text{Var}(X)$
- iii) $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$
- iv) $\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$

DÉMONSTRATION. Les parties i) et ii) proviennent directement de la définition de la covariance et la partie iii) est laissée en exercice au lecteur. Pour démontrer iv), qui stipule que la covariance est additive (comme l'espérance), soit $\mu_i = E[X_i]$ et $\nu_j = E[Y_j]$. Alors

$$E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mu_i, \quad E\left[\sum_{j=1}^m Y_j\right] = \sum_{j=1}^m \nu_j$$

et

$$\begin{aligned} \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) &= E\left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i\right)\left(\sum_{j=1}^m Y_j - \sum_{j=1}^m \nu_j\right)\right] \\ &= E\left[\sum_{i=1}^n (X_i - \mu_i) \sum_{j=1}^m (Y_j - \nu_j)\right] \\ &= E\left[\sum_{i=1}^n \sum_{j=1}^m (X_i - \mu_i)(Y_j - \nu_j)\right] \\ &= \sum_{i=1}^n \sum_{j=1}^m E[(X_i - \mu_i)(Y_j - \nu_j)] \end{aligned}$$

où la dernière égalité provient du fait que l'espérance d'une somme de variables est égale à la somme des espérances. ■

7.3.3 Variance de sommes

Les parties ii) et iv) du théorème 7.3, en prenant $Y_j = X_j$, $j = 1, \dots, n$, entraînent que

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n X_i\right) &= \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j) \end{aligned}$$

Puisque chaque paire d'indice i, j , $i \neq j$, apparaît deux fois dans la double sommation, l'expression ci-dessus est équivalente à

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \quad (7.7)$$

Si X_1, \dots, X_n sont indépendantes deux à deux, c'est-à-dire si X_i et X_j sont indépendantes pour $i \neq j$, alors l'équation (7.7) se réduit à

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$$

L'exemple suivant illustre l'utilisation de l'équation (7.7).

7.3.4 Exemples de variance de sommes

Exemple 7.19 Soient X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées de moyenne μ et de variance σ^2 , et, comme dans l'exemple 7.3, soit $\bar{X} = \sum_{i=1}^n X_i / n$ la moyenne de l'échantillon. Les quantités $X_i - \bar{X}$, $i = 1, \dots, n$, sont

appelées déviations car elles sont égales aux différences entre une donnée individuelle et la moyenne de l'échantillon. Soit S^2 la somme des carrés des déviations

$$S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$$

La variable aléatoire $S^2 / (n - 1)$ est appelée la variance de l'échantillon. Trouver

a) $\text{Var}(\bar{X})$;

b) $E[S^2 / (n - 1)]$.

SOLUTION.

a)

$$\begin{aligned} \text{Var}(\bar{X}) &= \left(\frac{1}{n}\right)^2 \text{Var}\left(\sum_{i=1}^n X_i\right) \\ &= \left(\frac{1}{n}\right)^2 \sum_{i=1}^n \text{Var}(X_i) && \text{par indépendance} \\ &= \frac{\sigma^2}{n} \end{aligned}$$

par indépendance

b) Commençons par l'identité algébrique suivante.

$$\begin{aligned} S^2 &= \sum_{i=1}^n (X_i - \mu + \mu - \bar{X})^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 - 2(\bar{X} - \mu) \sum_{i=1}^n (X_i - \mu) \\ &= \sum_{i=1}^n (X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu)n(\bar{X} - \mu) \\ &= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \end{aligned}$$

En prenant les espérances de l'expression ci-dessus, on obtient

$$\begin{aligned} E[S^2] &= \sum_{i=1}^n E[(X_i - \mu)^2] - nE[(\bar{X} - \mu)^2] \\ &= n\sigma^2 - n\text{Var}(\bar{X}) \\ &= (n - 1)\sigma^2 \end{aligned}$$

où l'égalité finale provient de la partie a) et où celle d'avant provient du résultat de l'exemple 7.3 qui donne $E[\bar{X}] = \mu$. En divisant par $(n - 1)$, on voit que l'espérance de la variance de l'échantillon est la variance σ^2 de la distribution. ■

L'exemple suivant présente une autre méthode que celle utilisée dans le chapitre 4 pour obtenir la variance d'une variable aléatoire binomiale.

Exemple 7.20 Variance d'une variable aléatoire binomiale

On considère ici le cas d'une variable aléatoire binomiale de paramètres n et p .

SOLUTION. Une telle variable représentant le nombre de succès lors de n épreuves indépendantes de même probabilité de succès p , on peut écrire

$$X = X_1 + \cdots + X_n$$

où chaque X_i est une variable de Bernoulli indépendante des autres et telle que

$$X_i = \begin{cases} 1 & \text{si la } i\text{-ème épreuve est réussie} \\ 0 & \text{sinon} \end{cases}$$

D'après (7.7), on obtient

$$\text{Var}(X) = \text{Var}(X_1) + \cdots + \text{Var}(X_n)$$

Mais

$$\begin{aligned} \text{Var}(X_i) &= E[X_i^2] - (E[X_i])^2 \\ &= E[X_i] - (E[X_i])^2 \quad \text{puisque } X_i^2 = X_i \\ &= p - p^2 \end{aligned}$$

et donc

$$\text{Var}(X) = np(1 - p) \quad \blacksquare$$

Exemple 7.21 Variance d'un nombre de rencontres

On cherche la variance de X , la variable comptant le nombre d'hommes ramassant leur propre chapeau dans la situation de l'exemple 7.8.

SOLUTION. On utilise la même décomposition de X que dans l'exemple 7.8, à savoir

$$X = X_1 + \cdots + X_N$$

où

$$X_i = \begin{cases} 1 & \text{si le } i\text{-ème homme ramasse son chapeau} \\ 0 & \text{sinon} \end{cases}$$

D'après l'équation (7.7), on a

$$\text{Var}(X) = \sum_{i=1}^N \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \quad (7.8)$$

Comme $P\{X_i = 1\} = 1/N$, on peut écrire en s'inspirant de l'exemple précédent que

$$\text{Var}(X_i) = \frac{1}{N} \left(1 - \frac{1}{N} \right) = \frac{N-1}{N^2}$$

De plus

$$\text{Cov}(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$$

Or

$$X_i X_j = \begin{cases} 1 & \text{si le } i\text{-ème et le } j\text{-ème hommes récupèrent leurs propres chapeaux} \\ 0 & \text{sinon} \end{cases}$$

et par conséquent

$$\begin{aligned} E[X_i X_j] &= P\{X_i = 1, X_j = 1\} \\ &= P\{X_i = 1\}P\{X_j = 1 | X_i = 1\} \\ &= \frac{1}{N} \frac{1}{N-1} \\ \text{Cov}(X_i, X_j) &= \frac{1}{N(N-1)} - \left(\frac{1}{N}\right)^2 = \frac{1}{N^2(N-1)} \end{aligned}$$

ce qui d'après (7.8) donne

$$\begin{aligned} \text{Var}(X) &= \frac{N-1}{N} + 2 \binom{N}{2} \frac{1}{N^2(N-1)} \\ &= \frac{N-1}{N} + \frac{1}{N} \\ &= 1 \end{aligned}$$

On constate donc que la variance du nombre de rencontres aussi bien que son espérance valent 1. Ce résultat était d'une certaine manière prévisible puisqu'on a montré à la section 2.5 que lorsque N est grand, la probabilité d'obtenir i rencontres est approximativement $e^{-1}/i!$. Ceci revient à dire que pour N grand toujours le nombre de rencontres suit approximativement une loi de Poisson d'espérance 1. Comme on a l'égalité de l'espérance et de la variance pour une variable poissonnienne, le résultat obtenu est sans surprise. ■

Exemple 7.22 Echantillon de population finie

On considère un groupe comptant N individus; chacun d'entre eux s'est fait une opinion sur un certain sujet, opinion que l'on mesure au moyen d'un nombre réel v appelé intensité du sentiment de l'individu relativement au sujet. On note v_i l'intensité du sentiment de l'individu i , $i = 1, \dots, N$. Les quantités v_i , $i = 1, \dots, N$ sont inconnues et pour acquérir de l'information à leur sujet on choisit au hasard un sous-groupe de n personnes. Choisir au hasard signifie qu'on s'y prend de manière telle qu'aucun sous-groupe de taille n parmi les $\binom{N}{n}$ sélections possibles n'a une plus grande chance d'être tiré qu'un autre. On interroge alors ces n personnes et mesure les v_i correspondants. On forme la somme S de ces n valeurs. Que vaudront l'espérance et la variance de S ?

A titre d'application importante de ce problème on peut mentionner les élections à l'occasion desquelles chaque électeur est pour ou contre un certain candidat. On

prendra alors $v_i = 1$ si l'électeur est favorable au candidat, $v_i = 0$ sinon. La grandeur $\bar{v} = \sum_{i=1}^N v_i / N$ représente alors la proportion de la population soutenant ce candidat, autrement dit sa cote. Pour estimer \bar{v} on choisit au hasard un échantillon de n personnes que l'on interroge. La proportion S/n favorable au candidat est souvent utilisée pour estimer \bar{v} .

SOLUTION. On définit pour chaque électeur i , $i = 1, \dots, N$, une variable indicatrice I_i indiquant si cette personne appartient ou non à l'échantillon. Plus précisément,

$$I_i = \begin{cases} 1 & \text{si la personne } i \text{ est dans l'échantillon} \\ 0 & \text{sinon} \end{cases}$$

Dès lors, S peut être écrite

$$S = \sum_{i=1}^N v_i I_i$$

donc

$$\begin{aligned} E[S] &= \sum_{i=1}^N v_i E[I_i] \\ \text{Var}(S) &= \sum_{i=1}^N \text{Var}(v_i I_i) + 2 \sum_{i < j} \text{Cov}(v_i I_i, v_j I_j) \\ &= \sum_{i=1}^N v_i^2 \text{Var}(I_i) + 2 \sum_{i < j} v_i v_j \text{Cov}(I_i, I_j) \end{aligned}$$

Comme

$$E[I_i] = \frac{n}{N}$$

$$E[I_i I_j] = \frac{n}{N} \frac{n-1}{N-1}$$

on voit que

$$\begin{aligned} \text{Var}(I_i) &= \frac{n}{N} \left(1 - \frac{n}{N}\right) \\ \text{Cov}(I_i, I_j) &= \frac{n(n-1)}{N(N-1)} - \left(\frac{n}{N}\right)^2 \\ &= \frac{-n(N-n)}{N^2(N-1)} \end{aligned}$$

Alors

$$\begin{aligned} E[S] &= n \sum_{i=1}^N \frac{v_i}{N} = n\bar{v} \\ \text{Var}(S) &= \frac{n}{N} \left(\frac{N-n}{N}\right) \sum_{i=1}^N v_i^2 - \frac{2n(N-n)}{N^2(N-1)} \sum_{i < j} v_i v_j \end{aligned}$$

L'expression de $\text{Var}(S)$ peut être transformée en utilisant l'identité

$(v_1 + \dots + v_N)^2 = \sum_{i=1}^N v_i^2 + 2 \sum_{i < j}^N v_i v_j$, ce qui laisse après simplifications

$$\text{Var}(S) = \frac{n(N-n)}{N-1} \left(\frac{\sum_{i=1}^N v_i^2}{N} - \bar{v}^2 \right)$$

On peut considérer le problème sous un autre angle et supposer que le nombre de v égaux à 1 est Np . La variable S pouvant dans ce cas être considérée comme une variable hypergéométrique, son espérance et sa variance seront

$$E[S] = n\bar{v} = np \quad \text{puisque } \bar{v} = \frac{Np}{N} = p$$

$$\begin{aligned} \text{Var}(S) &= \frac{n(N-n)}{(N-1)} \left(\frac{Np}{N} - p^2 \right) \\ &= \frac{n(N-n)}{N-1} p(1-p) \end{aligned}$$

Si par ailleurs, plutôt qu'à S , on s'intéresse à S/n , la proportion dans l'échantillon des électeurs favorables à notre candidat, on aura

$$E\left[\frac{S}{n}\right] = p$$

$$\text{Var}\left(\frac{S}{n}\right) = \frac{(N-n)}{n(N-1)} p(1-p) \quad \blacksquare$$

7.3.5 Corrélation

La **corrélation** entre deux variables aléatoires X et Y est notée $\rho(X, Y)$ et est définie ainsi, pour autant que $\text{Var}(X) \text{Var}(Y)$ soit non nul:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}$$

On peut montrer que

$$-1 \leq \rho(X, Y) \leq 1 \quad (7.9)$$

Pour établir (7.9), admettons que X et Y aient des variances σ_X^2 et σ_Y^2 , respectivement.

On a

$$\begin{aligned} 0 &\leq \text{Var}\left(\frac{X}{\sigma_x} + \frac{Y}{\sigma_y}\right) \\ &= \frac{\text{Var}(X)}{\sigma_x^2} + \frac{\text{Var}(Y)}{\sigma_y^2} + \frac{2 \text{Cov}(X, Y)}{\sigma_x \sigma_y} \\ &= 2[1 + \rho(X, Y)] \end{aligned}$$

qui implique que

$$-1 \leq \rho(X, Y)$$

D'autre part

$$\begin{aligned} 0 &\leq \text{Var}\left(\frac{X}{\sigma_x} - \frac{Y}{\sigma_y}\right) \\ &= \frac{\text{Var}(X)}{\sigma_x^2} + \frac{\text{Var} Y}{(-\sigma_y)^2} - \frac{2 \text{Cov}(X, Y)}{\sigma_x \sigma_y} \\ &= 2[1 - \rho(X, Y)] \end{aligned}$$

entraîne

$$\rho(X, Y) \leq 1$$

ce qui finit d'établir (7.9).

On sait que $\text{Var}(Z) = 0$ entraîne que Z est constante avec probabilité 1 (ce résultat intuitif sera rigoureusement démontré au chapitre 8). La démonstration de (7.9) établit donc, compte tenu de cette implication, que $\rho(X, Y) = 1$ entraîne $Y = a + bX$ où $b = \sigma_y/\sigma_x > 0$ tandis que $\rho(X, Y) = -1$ entraîne $Y = a + bX$ où $b = -\sigma_y/\sigma_x < 0$. On laisse au lecteur le soin de prouver que la réciproque est également vraie: si $Y = a + bX$, alors $\rho(X, Y)$ vaudra 1 ou -1 selon que le signe de b est positif ou négatif.

Le coefficient de corrélation est une mesure du degré de linéarité entre X et Y . Les valeurs de ρ proches de 1 ou -1 indiquent une linéarité quasiment rigoureuse entre X et Y , tandis que des valeurs proches de 0 indiquent une absence de toute relation linéaire. Lorsque $\rho(X, Y)$ est positif, Y a tendance à augmenter si X en fait autant, tandis que pour $\rho(X, Y) < 0$, Y a tendance à diminuer si X augmente. Si $\rho(X, Y) = 0$, on dit que ces deux statistiques sont *non corrélées*.

7.3.6 Exemple de calcul de corrélation

Exemple 7.23 Soient I_A et I_B les variables indicatrices des événements A et B . Par définition

$$I_A = \begin{cases} 1 & \text{si } A \text{ survient} \\ 0 & \text{sinon} \end{cases}$$

$$I_B = \begin{cases} 1 & \text{si } B \text{ survient} \\ 0 & \text{sinon.} \end{cases}$$

Alors

$$\begin{aligned} E[I_A] &= P(A) \\ E[I_B] &= P(B) \\ E[I_A I_B] &= P(AB) \end{aligned}$$

et ainsi

$$\begin{aligned} \text{Cov}(I_A, I_B) &= P(AB) - P(A)P(B) \\ &= P(B)[P(A|B) - P(A)] \end{aligned}$$

On vient d'obtenir un résultat annoncé par une approche intuitive de la situation: les variables indicatrices de A et B sont positivement corrélées, non corrélées ou négativement corrélées selon que, respectivement, $P(A|B)$ est plus grande, égale ou inférieure à $P(A)$. ■

L'exemple suivant montre que la moyenne de l'échantillon et les déviations ne sont pas corrélées.

Exemple 7.24 Soient X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées de variance σ^2 . Montrer que

$$\text{Cov}(X_i - \bar{X}, \bar{X}) = 0$$

SOLUTION.

$$\begin{aligned} \text{Cov}(X_i - \bar{X}, \bar{X}) &= \text{Cov}(X_i, \bar{X}) - \text{Cov}(\bar{X}, \bar{X}) \\ &= \text{Cov}\left(X_i, \frac{1}{n} \sum_{j=1}^n X_j\right) - \text{Var}(\bar{X}) \\ &= \frac{1}{n} \sum_{j=1}^n \text{Cov}(X_i, X_j) - \frac{\sigma^2}{n} \\ &= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0 \end{aligned}$$

où l'avant-dernière égalité utilise le résultat de l'exemple 7.19 et où la dernière égalité provient du fait que

$$\text{Cov}(X_i, X_j) = \begin{cases} 0 & \text{si } j \neq i \text{ par indépendance} \\ \sigma^2 & \text{si } j = i \text{ puisque } \text{Var}(X_i) = \sigma^2 \end{cases}$$

par indépendance

Bien que \bar{X} et la déviation $X_i - \bar{X}$ ne soient pas corrélées, elles ne sont pas, en général, indépendantes. Cependant, dans le cas particulier où les X_i sont des variables aléatoires normales, \bar{X} n'est plus seulement indépendante d'une seule déviation, mais est indépendante de la séquence entière des déviations $X_j - \bar{X}$, $j = 1, \dots, n$. Ce

résultat sera établi dans la section 9 où nous montrerons aussi que, dans ce cas, la moyenne \bar{X} de l'échantillon et la variance de l'échantillon $S^2/(n-1)$ sont indépendantes et que S^2/σ^2 a une distribution chi-carrée avec $n-1$ degrés de liberté. (Voir l'exemple 7.19 pour la définition de S^2 .) ■

Exemple 7.25 On considère m épreuves indépendantes. Chacune peut donner r résultats de probabilités P_1, P_2, \dots, P_r avec $\sum_{i=1}^r P_i = 1$. On désigne par $N_i, i = 1, \dots, r$ le nombre parmi ces m épreuves de celles qui aboutissent au résultat i . N_1, N_2, \dots, N_r suit alors une distribution multinomiale

$$P\{N_1 = n_1, N_2 = n_2, \dots, N_r = n_r\} \\ = \frac{m!}{n_1! n_2! \dots n_r!} P_1^{n_1} P_2^{n_2} \dots P_r^{n_r} \quad \sum_{i=1}^r n_i = m$$

Lorsque $i \neq j$, il semble raisonnable qu'une grande valeur de N_i soit associée à de petites valeurs de N_j , aussi s'attend-on intuitivement à ce que ces deux variables soient négativement corrélées. Calculons leur covariance en utilisant le théorème 7.3 (iv) et la décomposition

$$N_i = \sum_{k=1}^m I_i(k) \quad \text{et} \quad N_j = \sum_{k=1}^m I_j(k)$$

où

$$I_i(k) = \begin{cases} 1 & \text{si la } k\text{-ième épreuve aboutit au résultat } i \\ 0 & \text{sinon} \end{cases}$$

$$I_j(k) = \begin{cases} 1 & \text{si la } k\text{-ième épreuve aboutit au résultat } j \\ 0 & \text{sinon} \end{cases}$$

Grâce au théorème 7.3 (iv) on peut écrire

$$\text{Cov}(N_i, N_j) = \sum_{l=1}^m \sum_{k=1}^m \text{Cov}(I_i(k), I_j(l))$$

Or, lorsque $k \neq l$

$$\text{Cov}(I_i(k), I_j(l)) = 0$$

puisque l'issue de l'épreuve k ne dépend pas de celle de l'épreuve l . D'autre part,

$$\text{Cov}(I_i(l), I_j(l)) = E[I_i(l)I_j(l)] - E[I_i(l)]E[I_j(l)] \\ = 0 - P_i P_j = -P_i P_j$$

où l'on a utilisé la relation $I_i(l)I_j(l) = 0$, puisque l'épreuve l ne peut donner les deux résultats i et j à la fois. On obtient donc

$$\text{Cov}(N_i, N_j) = -mP_iP_j$$

qui confirme notre conclusion intuitive que N_i et N_j sont négativement corrélées. ■

7.4 ESPÉRANCE CONDITIONNELLE

7.4.1 Définition du cas direct

On se souvient que, pour un couple de variables aléatoires discrètes, on avait défini la loi de probabilité conditionnelle de X , sachant que $Y = y$, pour autant que $P\{Y = y\} > 0$, par

$$p_{X|Y}(x|y) = P\{X = x | Y = y\} = \frac{p(x, y)}{p_Y(y)}$$

Il est dès lors naturel de vouloir définir dans le cas discret l'*espérance conditionnelle de X sous la condition $Y = y$* , pour autant que $p_Y(y) > 0$, par

$$\begin{aligned} E[X | Y = y] &= \sum_x xP\{X = x | Y = y\} \\ &= \sum_x xp_{X|Y}(x|y) \end{aligned}$$

Exemple 7.26 On considère deux variables aléatoires binomiales X et Y , indépendantes et de mêmes paramètres n et p . On souhaite calculer l'espérance conditionnelle de X sous la condition $X + Y = m$.

SOLUTION. Déterminons d'abord la loi de probabilité conditionnelle de X sous la condition $X + Y = m$. Pour $k \leq \min(n, m)$ on a

$$\begin{aligned} P\{X = k | X + Y = m\} &= \frac{P\{X = k, X + Y = m\}}{P\{X + Y = m\}} \\ &= \frac{P\{X = k, Y = m - k\}}{P\{X + Y = m\}} \\ &= \frac{P\{X = k\}P\{Y = m - k\}}{P\{X + Y = m\}} \\ &= \frac{\binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-m+k}}{\binom{2n}{m} p^m (1-p)^{2n-m}} \\ &= \frac{\binom{n}{k} \binom{n}{m-k}}{\binom{2n}{m}} \end{aligned}$$

où l'on a utilisé le fait que $X + Y$ est une variable aléatoire binomiale de paramètres $2n$ et p (voir l'exemple 6.17). On conclut que la distribution conditionnelle de X , sachant que $X + Y = m$, est hypergéométrique. D'après le résultat de l'exemple 7.7, on sait que

$$E[X|X + Y = m] = \frac{m}{2} \quad \blacksquare$$

7.4.2 Définition du cas continu

On se souvient également que pour un couple de variables X et Y continues de densité $f(\cdot, \cdot)$, la densité conditionnelle de X , sachant que $Y = y$, est définie – pour autant que $f_Y(y) > 0$ – par

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

Il est donc naturel de définir l'espérance conditionnelle de X , dans le cas continu et sous la condition $Y = y$, par

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$$

pour les valeurs de y telles que $f_Y(y) > 0$.

Exemple 7.27 Supposons que la densité conjointe de X et Y soit

$$f(x, y) = \frac{e^{-x/y} e^{-y}}{y} \quad 0 < x < \infty, 0 < y < \infty$$

On veut calculer $E[X|Y = y]$.

SOLUTION. Calculons d'abord la densité conditionnelle

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x, y)}{f_Y(y)} \\ &= \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx} \\ &= \frac{(1/y)e^{-x/y}e^{-y}}{\int_0^{\infty} (1/y)e^{-x/y}e^{-y} dx} \\ &= \frac{(1/y)e^{-x/y}}{\int_0^{\infty} (1/y)e^{-x/y} dx} \\ &= \frac{(1/y)e^{-x/y}}{-e^{-x/y}|_{x=0}^{x=\infty}} \\ &= \left(\frac{1}{y}\right) e^{-x/y} \end{aligned}$$

On constate que cette densité conditionnelle de X sous la condition $Y = y$ n'est autre que la densité exponentielle de paramètre et espérance y . Donc,

$$E[X | Y = y] = \int_0^{\infty} \frac{x}{y} e^{-x/y} dx = y \quad \blacksquare$$

7.4.3 Elargissement du point de vue

On sait que les probabilités conditionnelles satisfont toutes les propriétés des probabilités simples. De la même manière, les espérances conditionnelles ont toutes les propriétés des espérances ordinaires, en particulier

$$E[g(X) | Y = y] = \begin{cases} \sum g(x) p_{X|Y}(x|y) & \text{dans le cas discret} \\ \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y) dx & \text{dans le cas continu} \end{cases}$$

et

$$E\left[\sum_{i=1}^n X_i | Y = y\right] = \sum_{i=1}^n E[X_i | Y = y]$$

En fait, l'espérance conditionnelle, sachant que $Y = y$, peut être considérée comme une espérance ordinaire basée sur un espace de probabilité, réduit aux seuls événements pour lesquels $Y = y$ est vérifiée.

7.4.4 Théorème de calcul d'espérances par conditionnement

On décide de noter $E[X | Y]$ la composition des fonctions Y et $E[X | Y = y]$, cette dernière fonction faisant correspondre à un nombre y l'espérance conditionnelle de X , sachant que $Y = y$. On remarquera que la fonction composée obtenue est elle-même une variable aléatoire. Le théorème qui suit énonce une propriété fondamentale de l'espérance conditionnelle.

Théorème 7.4

$$E[X] = E[E[X | Y]] \quad (7.10)$$

Lorsque Y est une variable discrète, ce théorème signifie que

$$E[X] = \sum_y E[X | Y = y] P\{Y = y\} \quad (7.11)$$

tandis que lorsque Y est continue, le théorème entraîne

$$E[X] = \int_{-\infty}^{\infty} E[X | Y = y] f_Y(y) dy \quad (7.12)$$

DÉMONSTRATION. On suppose pour cette démonstration que X et Y sont deux variables discrètes. On doit établir (7.11). Or, le membre de droite de (7.11) peut être réécrit

$$\begin{aligned}
 \sum_y E[X|Y=y]P\{Y=y\} &= \sum_y \sum_x xP\{X=x|Y=y\}P\{Y=y\} \\
 &= \sum_y \sum_x x \frac{P\{X=x, Y=y\}}{P\{Y=y\}} P\{Y=y\} \\
 &= \sum_y \sum_x xP\{X=x, Y=y\} \\
 &= \sum_x x \sum_y P\{X=x, Y=y\} \\
 &= \sum_x xP\{X=x\} \\
 &= E[X]
 \end{aligned}$$

ce qui établit le résultat.

Pour mieux faire comprendre (7.11), on peut donner l'interprétation suivante: pour calculer $E[X]$, on construit une moyenne pondérée des espérances conditionnelles de X sous les diverses conditions $Y = y$, les masses de pondération étant les probabilités des conditions (cette construction a déjà été rencontrée; où?). Le résultat (7.11) est très utilisé, il permet souvent de calculer assez facilement une espérance après avoir conditionné la variable par une autre variable appropriée. C'est ce qu'illustrent les exemples du paragraphe suivant. ■

7.4.5 Exemples de calcul d'espérances par conditionnement

Exemple 7.28 Un mineur est prisonnier dans un puits d'où partent trois tunnels. Le premier de ces tunnels le mènerait à la sortie au bout de 3 heures de marche. Le second le ramènerait à son point de départ au bout de 5 heures de marche, ainsi que le troisième au bout de 7 heures. Si à chaque choix qu'il fait le mineur emprunte n'importe quel tunnel avec la même probabilité, quelle sera la durée moyenne de sa tentative de sortie?

SOLUTION. La variable X représentera la durée de la recherche de la sortie en heures; Y représente la première porte choisie. On a

$$\begin{aligned}
 E[X] &= E[X|Y=1]P\{Y=1\} + E[X|Y=2]P\{Y=2\} \\
 &\quad + E[X|Y=3]P\{Y=3\} \\
 &= \frac{1}{3}(E[X|Y=1] + E[X|Y=2] + E[X|Y=3])
 \end{aligned}$$

Or

$$\begin{aligned}
 E[X|Y=1] &= 3 \\
 E[X|Y=2] &= 5 + E[X] \\
 E[X|Y=3] &= 7 + E[X]
 \end{aligned} \tag{7.13}$$

Pour comprendre les équations (7.13), voyons le cas particulier $E[X|Y = 2]$: si le mineur emprunte le deuxième tunnel, il y passera 5 heures puis reviendra à son point de départ. Une fois là, le problème est exactement le même qu'à l'origine. A ce point, la durée moyenne de sa recherche est $E[X]$. C'est pourquoi $E[X|Y = 2] = 5 + E[X]$. Les arguments à la base des deux autres équations sont semblables. On a donc finalement:

$$E[X] = \frac{1}{3}(3 + 5 + E[X] + 7 + E[X])$$

ou

$$E[X] = 15 \quad \blacksquare$$

Exemple 7.29 Espérance de la somme d'un nombre aléatoire de variables aléatoires
Le nombre de clients se rendant à un grand magasin donné dans l'espace d'une journée est une variable aléatoire d'espérance 50. La somme dépensée par chacun des clients quotidiens du magasin est aussi une variable aléatoire d'espérance 8 francs. On admet que les dépenses d'un client ne dépendent ni de celles des autres clients ni du nombre total de clients pour la journée. Quelle est l'espérance du chiffre d'affaires quotidien du magasin?

SOLUTION. Le nombre de clients par jour sera noté N , tandis que le montant dépensé par le client i est X_i . Le chiffre d'affaires du magasin est donc $\sum_{i=1}^N X_i$. Or

$$E\left[\sum_{i=1}^N X_i\right] = E\left[E\left[\sum_{i=1}^N X_i | N\right]\right]$$

Mais

$$\begin{aligned} E\left[\sum_{i=1}^N X_i | N = n\right] &= E\left[\sum_{i=1}^n X_i | N = n\right] \\ &= E\left[\sum_{i=1}^n X_i\right] \quad \text{du fait de l'indépendance des } X_i \text{ et de } N \\ &= nE[X] \end{aligned}$$

où l'on a désigné par $E[X]$ l'espérance commune à tous les X_i . Dès lors la variable aléatoire notée $E\left[\sum_{i=1}^N X_i | N\right]$ a pour expression ici

$$E\left[\sum_{i=1}^N X_i | N\right] = NE[X]$$

ce qui entraîne

$$E\left[\sum_{i=1}^N X_i\right] = E[NE[X]] = E[N]E[X]$$

Dans notre exemple, le chiffre d'affaires moyen du magasin est donc de $50 \times 8 = 400$ francs par jour. \blacksquare

Exemple 7.30 Une urne contient b boules blanches et n boules noires. On retire les boules une à une jusqu'à ce que la première blanche apparaisse. Quel sera le nombre moyen des boules noires tirées?

SOLUTION. On a déjà résolu ce problème dans l'exemple 7.13. Nous allons voir ici une solution faisant appel à la technique du conditionnement. Désignons par X le nombre des boules noires qu'il faut tirer. On notera $E[X] = M_{b,n}$ pour rendre explicite la dépendance de cette espérance envers b et n . On va obtenir une expression de $M_{b,n}$ en conditionnant sur la couleur de la première boule retirée de l'urne. On définit pour cela Y ainsi:

$$Y = \begin{cases} 1 & \text{si la première boule tirée est blanche} \\ 0 & \text{si elle est noire.} \end{cases}$$

Conditionnons selon Y :

$$M_{b,n} = E[X] = E[X|Y=1]P\{Y=1\} + E[X|Y=0]P\{Y=0\}$$

Mais

$$\begin{aligned} E[X|Y=1] &= 0 \\ E[X|Y=0] &= 1 + M_{b,n-1} \end{aligned} \quad (7.14)$$

Pour se convaincre de la validité de (7.14), supposons que la première boule tirée soit noire. Après le premier tirage, la situation est exactement ce qu'elle aurait été au départ si nous avions eu b boules blanches et $n-1$ boules noires. D'où l'expression de droite dans (7.14).

On connaît par ailleurs $P\{Y=0\} = n/(b+n)$, ce qui permet d'écrire

$$M_{b,n} = \frac{n}{b+n}[1 + M_{b,n-1}]$$

Or $M_{b,0}$ est évidemment nulle, à partir de quoi on peut faire les calculs suivants:

$$\begin{aligned} M_{b,1} &= \frac{1}{b+1}[1 + M_{b,0}] = \frac{1}{b+1} \\ M_{b,2} &= \frac{2}{b+2}[1 + M_{b,1}] = \frac{2}{b+2} \left[1 + \frac{1}{b+1} \right] = \frac{2}{b+1} \\ M_{b,3} &= \frac{3}{b+3}[1 + M_{b,2}] = \frac{3}{b+3} \left[1 + \frac{2}{b+1} \right] = \frac{3}{b+1} \end{aligned}$$

Par induction on vérifierait facilement que

$$M_{b,n} = \frac{n}{b+1} \quad \blacksquare$$

7.4.6 Exemple de calcul de variance par conditionnement

On peut appliquer la technique précédente au calcul de variance également, comme le montre l'exemple suivant.

Exemple 7.31 Variance d'une variable géométrique

On réalise des épreuves indépendantes ayant toutes une même probabilité p de succès. On désigne par N le nombre d'essais à faire jusqu'à obtenir le premier succès. Quelle sera la variance de N ?

SOLUTION. Soit Y la variable valant 1 lorsque la première épreuve est un succès et 0 sinon. On sait que

$$\text{Var}(N) = E[N^2] - (E[N])^2$$

Pour calculer $E[N^2]$ on utilise la technique du conditionnement:

$$E[N^2] = E[E[N^2 | Y]]$$

Mais

$$\begin{aligned} E[N^2 | Y = 1] &= 1 \\ E[N^2 | Y = 0] &= E[(1 + N)^2] \end{aligned}$$

Ces deux équations sont fondées. Si en effet la première épreuve est un succès, N vaut évidemment 1, N^2 aussi. Si par contre cette épreuve est un échec, le nombre total d'épreuves jusqu'à obtenir un succès suivra la même distribution que la variable N augmentée de 1 (le premier échec). On peut donc écrire $E[N^2 | Y = 0] = E[(N + 1)^2]$, puis

$$\begin{aligned} E[N^2] &= E[N^2 | Y = 1]P\{Y = 1\} + E[N^2 | Y = 0]P\{Y = 0\} \\ &= p + (1 - p)E[(1 + N)^2] \\ &= 1 + (1 - p)E[2N + N^2] \end{aligned}$$

On a par ailleurs montré dans l'exemple 4.31 que $E[N] = 1/p$. Il reste donc

$$E[N^2] = 1 + \frac{2(1 - p)}{p} + (1 - p)E[N^2]$$

ou encore

$$E[N^2] = \frac{2 - p}{p^2}$$

Par conséquent

$$\begin{aligned} \text{Var}(N) &= E[N^2] - (E[N])^2 \\ &= \frac{2 - p}{p^2} - \left(\frac{1}{p}\right)^2 \\ &= \frac{1 - p}{p^2} \end{aligned}$$

■

Dans l'exemple suivant, nous allons déterminer l'espérance du nombre de variables aléatoires uniformes $(0, 1)$ qu'on doit additionner pour que leur somme dépasse 1. La réponse, étonnamment, est e .

Exemple 7.32 Soit U_1, U_2, \dots une séquence de variables aléatoires uniformes $(0, 1)$ indépendantes. Trouver $E[N]$ lorsque

$$N = \min \left\{ n : \sum_{i=1}^n U_i > 1 \right\}$$

SOLUTION. Nous allons résoudre ce problème en cherchant un résultat plus général. Pour $x \in [0, 1]$, soit

$$N(x) = \min \left\{ n : \sum_{i=1}^n U_i > x \right\}$$

et posons

$$m(x) = E[N(x)]$$

$N(x)$ est le nombre de variables aléatoires uniformes $(0, 1)$ qu'on doit additionner pour que leur somme dépasse x , et $m(x)$ est son espérance. Nous allons maintenant donner une équation pour $m(x)$ en conditionnant sur U_1 . Cela donne, à partir de l'équation (7.12),

$$m(x) = \int_0^1 E[N(x)|U_1 = y] dy \quad (7.15)$$

On a

$$E[N(x)|U_1 = y] = \begin{cases} 1 & \text{si } y > x \\ 1 + m(x-y) & \text{si } y \leq x \end{cases} \quad (7.16)$$

La formule précédente est évidente lorsque $y > x$ car si la première valeur uniforme est y , alors le nombre restant d'uniformes à additionner est le même que si nous commençons et que nous additionnons des uniformes jusqu'à ce que leur somme dépasse $x - y$. En substituant (7.16) dans (7.15), on a

$$\begin{aligned} m(x) &= 1 + \int_0^x m(x-y) dy \\ &= 1 + \int_0^x m(u) du \quad \text{en posant } u = x - y \end{aligned}$$

En dérivant cette équation, on a

$$m'(x) = m(x)$$

ou, de manière équivalente,

$$\frac{m'(x)}{m(x)} = 1$$

En intégrant ceci, on obtient

$$\ln[m(x)] = x + c$$

ou

$$m(x) = ke^x$$

Puisque $m(0) = 1$, on voit que $k = 1$, donc on obtient

$$m(x) = e^x$$

Donc $m(1)$, l'espérance du nombre de variables aléatoires uniformes $(0, 1)$ qu'on doit additionner pour que leur somme dépasse 1, est égale à e . ■

7.4.7 Calcul de probabilité par conditionnement

La technique utilisée pour le calcul d'espérances et ci-dessus de variances, consistant à conditionner selon une variable aléatoire appropriée, peut également servir au calcul de probabilités. Nous allons voir comment l'appliquer au calcul de la probabilité d'un événement E quelconque. On définit d'abord la variable indicatrice X de E :

$$X = \begin{cases} 1 & \text{si } E \text{ est réalisé} \\ 0 & \text{sinon} \end{cases}$$

En conséquence de la définition de X , on aura:

$$\begin{aligned} E[X] &= P(E) \\ E[X | Y = y] &= P(E | Y = y) \text{ pour toute variable } Y. \end{aligned}$$

En vertu de (7.10) et (7.11), on peut alors écrire, respectivement

$$\begin{aligned} P(E) &= \sum_y P(E | Y = y)P(Y = y) && \text{si } Y \text{ est discrète} \\ &= \int_{-\infty}^{\infty} P(E | Y = y)f_Y(y) dy && \text{si } Y \text{ est continue} \end{aligned} \quad (7.17)$$

On remarquera que si Y est une variable discrète pouvant prendre les valeurs y_1, y_2, \dots, y_n , et si on définit les événements F_i pour i allant de 1 à n par $F_i = \{Y = y_i\}$, (7.17) se réduit à l'équation déjà connue

$$P(E) = \sum_{i=1}^n P(E | F_i)P(F_i)$$

où F_1, \dots, F_n sont des événements mutuellement exclusifs dont l'union est l'ensemble fondamental.

Exemple 7.33 Soient deux variables aléatoires X et Y continues indépendantes et de densités f_X, f_Y respectivement. On souhaite calculer $P\{X < Y\}$.

SOLUTION. On conditionne selon Y , ce qui donne:

$$\begin{aligned} P\{X < Y\} &= \int_{-\infty}^{\infty} P\{X < Y \mid Y = y\} f_Y(y) dy \\ &= \int_{-\infty}^{\infty} P\{X < y \mid Y = y\} f_Y(y) dy \\ &= \int_{-\infty}^{\infty} P\{X < y\} f_Y(y) dy \quad \text{par hypothèse d'indépendance} \\ &= \int_{-\infty}^{\infty} F_X(y) f_Y(y) dy \end{aligned}$$

où

$$F_X(y) = \int_{-\infty}^y f_X(x) dx \quad \blacksquare$$

Exemple 7.34 Soient X et Y deux variables indépendantes et continues. On veut connaître la distribution de $X + Y$.

SOLUTION. En conditionnant selon Y , on obtient:

$$\begin{aligned} P\{X + Y < a\} &= \int_{-\infty}^{\infty} P\{X + Y < a \mid Y = y\} f_Y(y) dy \\ &= \int_{-\infty}^{\infty} P\{X + y < a \mid Y = y\} f_Y(y) dy \\ &= \int_{-\infty}^{\infty} P\{X < a - y\} f_Y(y) dy \\ &= \int_{-\infty}^{\infty} F_X(a - y) f_Y(y) dy \quad \blacksquare \end{aligned}$$

7.4.8 Variance conditionnelle

On peut définir la *variance conditionnelle* de X selon Y en s'inspirant de la démarche suivie pour l'espérance conditionnelle. On obtient en fin de compte la variable aléatoire:

$$\text{Var}(X \mid Y) \equiv E[(X - E(X \mid Y))^2 \mid Y]$$

Cette variable fait correspondre à tout événement l'espérance conditionnelle (sous condition $Y = y$, y étant la valeur associée à cet événement par Y) de la variable égale au carré de la différence entre X et son espérance conditionnelle par rapport à Y . En d'autres termes, cette définition s'appuie sur la définition ordinaire de la variance, à cela près que toutes les espérances sont prises conditionnellement selon Y .

Il existe une relation très utile entre $\text{Var}(X)$, la variance conditionnelle de X et la variable $\text{Var}(X|Y)$; cette relation sert parfois au calcul de $\text{Var}(X)$. Pour construire cette relation on notera d'abord que

$$\text{Var}(X|Y) = E[X^2|Y] - (E[X|Y])^2$$

d'après le même raisonnement que celui grâce auquel on établit $\text{Var}(X) = E[X^2] - (E[X])^2$. Donc

$$\begin{aligned} E[\text{Var}(X|Y)] &= E[E[X^2|Y]] - E[(E[X|Y])^2] \\ &= E[X^2] - E[(E[X|Y])^2] \end{aligned} \quad (7.18)$$

Mais comme $E[E[X|Y]] = E[X]$, on a par ailleurs

$$\text{Var}(E[X|Y]) = E[(E[X|Y])^2] - [E[X]]^2 \quad (7.19)$$

On en arrive ainsi à achever la démonstration du théorème suivant, par simple addition de (7.18) et (7.19).

Théorème 7.5 *Formule de la variance conditionnelle*

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y])$$

7.4.9 Exemples de calculs de variances par conditionnement

Exemple 7.35 On admet qu'à toute heure t , le nombre de personnes qui se sont rendues à une gare suit une loi de Poisson de paramètre croissant λt . Si le premier train s'arrêtant à cette gare repart à une heure aléatoire uniformément distribuée dans l'intervalle $(0, T)$ et indépendante de l'arrivée des passagers, quelles seront l'espérance et la variance du nombre de passagers qui pourront prendre le train?

SOLUTION. Pour tout $t \geq 0$ on convient d'appeler $N(t)$, le nombre de personnes attendant à la gare à l'heure t . L'heure d'arrivée du train sera désignée par Y . Ce qui nous intéresse est donc $N(Y)$. Le conditionnement selon Y donne:

$$\begin{aligned} E[N(Y)|Y = t] &= E[N(t)|Y = t] \\ &= E[N(t)] \text{ en vertu de l'indépendance de } Y \text{ et } N(t) \\ &= \lambda t \quad \text{du fait que } N(t) \text{ est poissonnienne de paramètre } \lambda t \end{aligned}$$

Donc

$$E[N(Y)|Y] = \lambda Y$$

Prenons l'espérance des deux membres. On aura

$$E[N(Y)] = \lambda E[Y] = \frac{\lambda T}{2}$$

Pour obtenir $\text{Var}(N(Y))$, utilisons la formule de la variance conditionnelle:

$$\begin{aligned} \text{Var}(N(Y) | Y = t) &= \text{Var}(N(t) | Y = t) \\ &= \text{Var}(N(t)) \quad \text{par hypothèse d'indépendance} \\ &= \lambda t \end{aligned}$$

et donc

$$\begin{aligned} \text{Var}(N(Y) | Y) &= \lambda Y \\ E[N(Y) | Y] &= \lambda Y \end{aligned}$$

Ce qui donne en utilisant la formule précitée:

$$\begin{aligned} \text{Var}(N(Y)) &= E[\lambda Y] + \text{Var}(\lambda Y) \\ &= \lambda \frac{T}{2} + \lambda^2 \frac{T^2}{12} \end{aligned}$$

où l'on a utilisé la formule $\text{Var}(Y) = T^2/12$. ■

Exemple 7.36 Variance de la somme d'un nombre aléatoire de variables aléatoires
Soit X_1, X_2, \dots une collection de variables aléatoires indépendantes et identiquement distribuées. Soit encore N , une variable aléatoire indépendante des $X_i, i \geq 1$, à valeurs entières non-négatives. On désire calculer $\text{Var}\left(\sum_{i=1}^N X_i\right)$.

SOLUTION. On conditionne selon N :

$$\begin{aligned} E\left[\sum_{i=1}^N X_i | N\right] &= NE[X] \\ \text{Var}\left(\sum_{i=1}^N X_i | N\right) &= N \text{Var}(X) \end{aligned}$$

Ces relations sont valables, car lorsque N est connue, $\sum_{i=1}^N X_i$ n'est autre que la somme d'un nombre fixe de variables aléatoires indépendantes. Dans ce cas, l'espérance et la variance d'une telle somme sont simplement la somme des espérances et des variances, respectivement. On applique maintenant la formule de variance conditionnelle pour obtenir

$$\text{Var}\left(\sum_{i=1}^N X_i\right) = E[N] \text{Var}(X) + (E[X])^2 \text{Var}(N) \quad \blacksquare$$

7.5 ESPÉRANCE CONDITIONNELLE ET PRÉDICTION

7.5.1 Meilleur prédicteur

Il arrive que l'on puisse observer la valeur d'une variable aléatoire X et qu'ensuite, on souhaite prédire la valeur d'une autre variable Y en se basant sur ce que l'on sait de X . Désignons par $g(X)$ le *prédicteur* de Y , au sens où si X prend x pour valeur alors $g(x)$ est la valeur prédite de Y . On souhaite évidemment choisir g de telle manière que la variable $g(X)$ soit aussi proche que possible de Y . Un critère de qualité en matière de prédiction est d'avoir minimisé $E[(Y - g(X))^2]$. Nous allons montrer ci-dessous que selon ce critère, le meilleur prédicteur de Y est $g(X) = E[Y|X]$.

Théorème 7.6

Pour toute fonction g

$$E[(Y - g(X))^2] \geq E[(Y - E[Y|X])^2]$$

DÉMONSTRATION.

$$\begin{aligned} E[(Y - g(X))^2|X] &= E[(Y - E[Y|X] + E[Y|X] - g(X))^2|X] \\ &= E[(Y - E[Y|X])^2|X] \\ &\quad + E[(E[Y|X] - g(X))^2|X] \\ &\quad + 2E[(Y - E[Y|X])(E[Y|X] - g(X))|X] \quad (7.20) \end{aligned}$$

A ce point, si X est connue, on peut considérer $E[Y|X] - g(X)$ comme une constante. Aussi a-t-on

$$\begin{aligned} &E[(Y - E[Y|X])(E[Y|X] - g(X))|X] \\ &= (E[Y|X] - g(X))E[Y - E[Y|X]|X] \\ &= (E[Y|X] - g(X))(E[Y|X] - E[Y|X]) \\ &= 0 \quad (7.21) \end{aligned}$$

On tire de (7.20) et (7.21)

$$E[(Y - g(X))^2|X] \geq E[(Y - E[Y|X])^2|X]$$

le pas final consistant à prendre l'espérance des deux membres de l'inégalité ci-dessus. ■

On peut donner un argument plus intuitif, quoique moins rigoureux, pour faire comprendre le théorème 7.6. Il n'est pas difficile de montrer que $E[(Y - c)^2]$ est minimale lorsque $c = E\{Y\}$ (voir l'exercice théorique 7.10.5). Si donc, nous n'avons aucune information, la meilleure prédiction de Y dans le sens de la minimisation du carré moyen de l'erreur, est de déclarer que Y prendra comme valeur son espérance. Si nous disposons d'une observation x de la variable X , le problème de prédiction se pose exactement dans les mêmes termes qu'auparavant, à cela près qu'il faut considérer toutes les probabilités et espérances comme conditionnelles, la condition étant $X = x$. En résumé, la meilleure prédiction dans ce cas est de déclarer que Y prendra pour valeur l'espérance conditionnelle de Y sous la condition $X = x$, ce qui est bien ce qu'établit le théorème 7.6.

7.5.2 Exemples de prédiction

Exemple 7.37 On suppose que le fils d'un homme de taille x (en centimètres) atteindra une taille de distribution normale autour de $x + 2$ et de variance 10. Quelle est la meilleure prédiction que l'on puisse donner sur la taille finale du fils d'un homme de 180 cm?

SOLUTION. Formellement le modèle admis peut être écrit

$$Y = X + 2 + e$$

où e est une variable normale, indépendante de X et de paramètres 0 et 10. Les variables X et Y représentent bien sûr respectivement la taille du père et du fils. La meilleure prédiction est donc $E[Y|X = 180]$ à savoir

$$\begin{aligned} E[Y|X = 180] &= E[X + 2 + e|X = 180] \\ &= 182 + E[e|X = 180] \\ &= 182 + E[e] \quad \text{par hypothèse d'indépendance} \\ &= 182 \end{aligned}$$

■

Exemple 7.38 On suppose qu'un signal d'intensité s lors de son émission en A est perçu en B avec une intensité aléatoire normalement distribuée, de paramètres $(s, 1)$. Or le signal émis en A est lui aussi une variable aléatoire S , normale de paramètres (μ, σ^2) . On observe en B un signal que l'on note R . Quelle est la meilleure estimation de l'intensité d'émission si l'on enregistre $R = r$?

SOLUTION. Commençons par calculer la densité conditionnelle de S , R étant donnée:

$$\begin{aligned} f_{S|R}(s|r) &= \frac{f_{S,R}(s,r)}{f_R(r)} \\ &= \frac{f_S(s)f_{R|S}(r|s)}{f_R(r)} \\ &= Ke^{-(s-\mu)^2/2\sigma^2} e^{-(r-s)^2/2} \end{aligned}$$

où K ne dépend pas de s . Or

$$\begin{aligned} \frac{(s-\mu)^2}{2\sigma^2} + \frac{(r-s)^2}{2} &= s^2 \left(\frac{1}{2\sigma^2} + \frac{1}{2} \right) - \left(\frac{\mu}{\sigma^2} + r \right) s + C_1 \\ &= \frac{1+\sigma^2}{2\sigma^2} \left[s^2 - 2 \left(\frac{\mu + r\sigma^2}{1+\sigma^2} \right) s \right] + C_1 \\ &= \frac{1+\sigma^2}{2\sigma^2} \left(s - \frac{(\mu + r\sigma^2)}{1+\sigma^2} \right)^2 + C_2 \end{aligned}$$

où de nouveau C_1 et C_2 ne dépendent pas de s . Ainsi

$$f_{S|R}(s|r) = C \exp \left\{ \frac{- \left[s - \frac{(\mu + r\sigma^2)}{1 + \sigma^2} \right]^2}{2 \left(\frac{\sigma^2}{1 + \sigma^2} \right)} \right\}$$

où C ne dépend toujours pas de s . On en déduit que la distribution conditionnelle de S , intensité du signal émis, sachant que l'intensité du signal reçu est r , est normale de paramètres

$$E[S|R = r] = \frac{\mu + r\sigma^2}{1 + \sigma^2}$$

$$\text{Var}(S|R = r) = \frac{\sigma^2}{1 + \sigma^2}$$

D'après le théorème 7.6, la meilleure estimation de signal émis est, dans le sens d'un carré moyen de l'erreur minimum et d'après l'observation de l'intensité r du signal reçu

$$E[S|R = r] = \frac{1}{1 + \sigma^2} \mu + \frac{\sigma^2}{1 + \sigma^2} r$$

Cette dernière décomposition de l'espérance conditionnelle de S est instructive dans la mesure où elle montre que la prédiction est une moyenne pondérée de μ , espérance a priori du signal émis, et de r , valeur enregistrée. Les poids attribués à μ et r sont dans la même proportion entre eux que 1 et σ^2 , respectivement variance conditionnelle du signal reçu lorsque l'intensité d'émission est s et variance du signal émis. ■

Exemple 7.39 Lors du traitement digital de données basées sur des mesures analogiques brutes X , ces mesures doivent être rendues discrètes avant tout traitement. Pour ce faire, on détermine de manière tout à fait générale un ensemble ordonné de bornes a_i , $i = 0, \pm 1, \pm 2, \dots$ telles que $\lim_{i \rightarrow +\infty} a_i = \infty$ et $\lim_{i \rightarrow -\infty} a_i = -\infty$. Une mesure

analogique X est alors rendue discrète selon l'intervalle (a_i, a_{i+1}) auquel elle appartient. Notons par y_i la valeur attribuée à la variance X rendue discrète lorsque $X \in (a_i, a_{i+1})$; en d'autres termes

$$Y = y_i \quad \text{si } a_i < X \leq a_{i+1}$$

La distribution de Y est donnée par

$$P\{Y = y_i\} = F_X(a_{i+1}) - F_X(a_i)$$

Le problème consiste alors à choisir les valeurs y_i , $i = 0, \pm 1, \pm 2, \dots$ de manière à minimiser $E[(X - Y)^2]$, l'espérance du carré de l'erreur entre donnée brute et approximation discrète.

a) On veut trouver les valeurs y_i optimales, $i = 0, \pm 1, \pm 2, \dots$;

- b) pour le quantificateur optimal Y ainsi trouvé, on veut montrer que $E[Y] = E[X]$, ce qui signifie que ce quantificateur conserve l'espérance des données;
 c) on souhaite encore montrer que $\text{Var}(Y) = \text{Var}(X) - E[(X - Y)^2]$.

SOLUTION.

a) Pour tout quantificateur Y on obtient, en conditionnant selon les valeurs de Y

$$E[(X - Y)^2] = \sum_i E[(X - y_i)^2 | a_i < X \leq a_{i+1}] P\{a_i < X \leq a_{i+1}\}$$

Posons maintenant

$$I = i \quad \text{si } a_i < X \leq a_{i+1}$$

Alors

$$E[(X - y_i)^2 | a_i < X \leq a_{i+1}] = E[(X - y_i)^2 | I = i]$$

et en vertu du théorème 7.7, cette quantité est minimale lorsque

$$\begin{aligned} y_i &= E[X | I = i] \\ &= E[X | a_i < X \leq a_{i+1}] \\ &= \int_{a_i}^{a_{i+1}} \frac{xf_X(x) dx}{F_X(a_{i+1}) - F_X(a_i)} \end{aligned}$$

Or, comme le quantificateur optimal est donné par $Y = E[X|I]$, on a

$$\begin{aligned} \text{b) } E[Y] &= E[X] \\ \text{c) } \text{Var}(X) &= E[\text{Var}(X|I)] + \text{Var}(E[X|I]) \\ &= E[E[(X - Y)^2 | I]] + \text{Var}(Y) \\ &= E[(X - Y)^2] + \text{Var}(Y) \end{aligned} \quad \blacksquare$$

7.5.3 Meilleur prédicteur linéaire

Il arrive que la distribution conjointe de X et Y ne soit pas connue, ou le soit mais que le calcul de $E[Y|X = x]$ soit pratiquement hors de portée. Dans ce cas, on peut cependant trouver le meilleur *prédicteur linéaire* de Y basé sur X si les espérances et variances de ces deux variables sont connues, ainsi que leur corrélation.

Il s'agit alors de déterminer deux constantes a et b telles que $E[(Y - (a + bX))^2]$ soit minimum. Or

$$\begin{aligned} E[(Y - (a + bX))^2] &= E[Y^2 - 2aY - 2bXY + a^2 + 2abX + b^2X^2] \\ &= E[Y^2] - 2aE[Y] - 2bE[XY] + a^2 \\ &\quad + 2abE[X] + b^2E[X^2] \end{aligned}$$

La dérivation partielle par rapport à a et b donne

$$\begin{aligned}\frac{\partial}{\partial a} E[(Y - a - bX)^2] &= -2E[Y] + 2a + 2bE[X] \\ \frac{\partial}{\partial b} E[(Y - a - bX)^2] &= -2E[XY] + 2aE[X] + 2bE[X^2]\end{aligned}\quad (7.22)$$

La résolution du système d'équations identifiant les dérivées partielles à 0, donne les solutions a et b

$$\begin{aligned}b &= \frac{E[XY] - E[X]E[Y]}{E[X^2] - (E[X])^2} = \frac{\text{Cov}(X, Y)}{\sigma_x^2} = \rho \frac{\sigma_y}{\sigma_x} \\ a &= E[Y] - bE[X] = E[Y] - \frac{\rho\sigma_y E[X]}{\sigma_x}\end{aligned}\quad (7.23)$$

où ρ désigne la corrélation entre X et Y , tandis que $\sigma_y^2 = \text{Var}(Y)$ et $\sigma_x^2 = \text{Var}(X)$. On vérifie aisément que ces valeurs donnent bien un minimum de $E[(Y - (a + bX))^2]$, déterminant ainsi le meilleur prédicteur linéaire de Y basé sur X au sens du carré moyen de l'erreur

$$\mu_y + \frac{\rho\sigma_y}{\sigma_x}(X - \mu_x)$$

où $\mu_y = E[Y]$ et $\mu_x = E[X]$.

Le carré moyen de l'erreur de ce prédicteur est

$$\begin{aligned}E\left[\left(Y - \mu_y - \rho \frac{\sigma_y}{\sigma_x}(X - \mu_x)\right)^2\right] \\ &= E[(Y - \mu_y)^2] + \rho^2 \frac{\sigma_y^2}{\sigma_x^2} E[(X - \mu_x)^2] - 2\rho \frac{\sigma_y}{\sigma_x} E[(Y - \mu_y)(X - \mu_x)] \\ &= \sigma_y^2 + \rho^2 \sigma_y^2 - 2\rho^2 \sigma_y^2 \\ &= \sigma_y^2(1 - \rho^2)\end{aligned}\quad (7.24)$$

On constate grâce à (7.24) que, lorsque ρ est voisin de $+1$ ou -1 , le carré moyen de l'erreur du meilleur prédicteur linéaire est presque nul.

Exemple 7.40 On va montrer qu'il existe une situation parmi d'autres dans laquelle l'espérance conditionnelle de Y selon X est linéaire en X , ce qui signifie que le meilleur prédicteur linéaire de Y basé sur X , est le meilleur tout court: celle où X et Y sont des variables conjointement normales.

SOLUTION. La densité conjointe de X et Y est

$$\begin{aligned}f(x, y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right.\right. \\ &\quad \left.\left.-\frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}\end{aligned}$$

On laisse au lecteur le soin de vérifier que la densité conditionnelle de Y , $X = x$ étant donné, est

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left(y - \mu_y - \frac{\rho\sigma_y}{\sigma_x}(x - \mu_x)\right)^2\right\}$$

On voit donc que la distribution conditionnelle associée est normale, d'espérance

$$E[Y|X = x] = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x)$$

et de variance $\sigma_y^2(1 - \rho^2)$, ce qui montre bien que le meilleur estimateur est linéaire. On retrouve d'ailleurs l'expression du meilleur estimateur linéaire. ■

7.6 FONCTIONS GÉNÉRATRICES DES MOMENTS

7.6.1 Définition et propriétés

On définit pour tout réel t , la *fonction génératrice des moments* M de la variable aléatoire X par

$$M(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} p(x) & \text{si } X \text{ est discrète, de loi de probabilité } p \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{si } X \text{ est continue, de densité } f \end{cases}$$

Cette fonction M est appelée fonction génératrice des moments du fait que tous les moments d'ordre n de X peuvent être calculés en dérivant n fois M puis en évaluant la dérivée en $t = 0$. Par exemple,

$$\begin{aligned} M'(t) &= \frac{d}{dt} E[e^{tX}] \\ &= E\left[\frac{d}{dt}(e^{tX})\right] \\ &= E[Xe^{tX}] \end{aligned} \tag{7.25}$$

où l'on a admis que la permutation des opérations de dérivation et de prise d'espérance était légitime. Plus précisément on a admis que

$$\frac{d}{dt} \left[\sum_x e^{tx} p(x) \right] = \sum_x \frac{d}{dt} [e^{tx} p(x)]$$

dans le cas discret et que

$$\frac{d}{dt} \left[\int e^{tx} f(x) dx \right] = \int \frac{d}{dt} [e^{tx} f(x)] dx$$

dans le cas continu. Cette supposition est presque toujours fondée et d'ailleurs est valable pour toutes les distributions considérées dans cet ouvrage. Ceci fait que lorsqu'on évalue M' obtenue dans (7.25) en $t = 0$ on trouve bien

$$M'(0) = E[X]$$

Et de même

$$\begin{aligned} M''(t) &= \frac{d}{dt} M'(t) \\ &= \frac{d}{dt} E[Xe^{tX}] \\ &= E\left[\frac{d}{dt}(Xe^{tX})\right] \\ &= E[X^2 e^{tX}] \end{aligned}$$

par conséquent

$$M''(0) = E[X^2]$$

L'expression générale de la n -ième dérivée de M est

$$M^n(t) = E[X^n e^{tX}] \quad n \geq 1$$

qui laisse

$$M^n(0) = E[X^n] \quad n \geq 1$$

7.6.2 Exemples d'applications de la fonction génératrice des moments

Les exemples qui suivent étudient la fonction génératrice des moments de plusieurs distributions courantes.

Exemple 7.41 Cas de la distribution binomiale de paramètres n et p

Soit X une variable suivant une telle distribution. On souhaite déterminer M et appliquer ses propriétés au calcul de $E[X]$ et $\text{Var}(X)$.

SOLUTION. On a

$$\begin{aligned} M(t) &= E[e^{tX}] \\ &= \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\ &= (pe^t + 1 - p)^n \end{aligned}$$

où l'on reconnaît lors de la dernière égalité, une application de la formule du binôme. En dérivant,

$$M'(t) = n(pe^t + 1 - p)^{n-1} pe^t$$

et par suite

$$E[X] = M'(0) = np$$

ce qui est bien en accord avec le résultat calculé dans l'exemple 7.5. Une seconde dérivation donne

$$M''(t) = n(n-1)(pe^t + 1 - p)^{n-2}(pe^t)^2 + n(pe^t + 1 - p)^{n-1} pe^t$$

puis

$$E[X^2] = M''(0) = n(n-1)p^2 + np$$

La variance de X est donc

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= n(n-1)p^2 + np - n^2p^2 \\ &= np(1-p) \end{aligned}$$

ce qui une fois encore confirme le calcul effectué dans l'exemple 7.20. ■

Exemple 7.42 Cas de la distribution de Poisson de paramètre λ

L'objectif est le même que dans l'exemple précédent, X désigne toujours la variable aléatoire étudiée.

SOLUTION. On a

$$\begin{aligned} M(t) &= E[e^{tX}] \\ &= \sum_{n=0}^{\infty} \frac{e^{tn} e^{-\lambda} \lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} \\ &= e^{-\lambda} e^{\lambda e^t} \\ &= \exp\{\lambda(e^t - 1)\} \end{aligned}$$

Deux dérivations donnent successivement

$$\begin{aligned} M'(t) &= \lambda e^t \exp\{\lambda(e^t - 1)\} \\ M''(t) &= (\lambda e^t)^2 \exp\{\lambda(e^t - 1)\} + \lambda e^t \exp\{\lambda(e^t - 1)\} \end{aligned}$$

et donc

$$\begin{aligned} E[X] &= M'(0) = \lambda \\ E[X^2] &= M''(0) = \lambda^2 + \lambda \\ \text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= \lambda \end{aligned}$$

On constate encore une fois que l'espérance et la variance d'une variable poissonnienne sont toutes deux égales à λ . ■

Exemple 7.43 Cas d'une distribution exponentielle de paramètre λ

SOLUTION. On a

$$\begin{aligned} M(t) &= E[e^{tX}] \\ &= \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{-(\lambda-t)x} dx \\ &= \frac{\lambda}{\lambda-t} \quad \text{pour } t < \lambda \end{aligned}$$

On remarque ici que M n'est définie que pour des valeurs de t inférieures à λ . Deux dérivations livrent

$$M'(t) = \frac{\lambda}{(\lambda-t)^2} \quad M''(t) = \frac{2\lambda}{(\lambda-t)^3}$$

Et donc

$$E[X] = M'(0) = \frac{1}{\lambda} \quad E[X^2] = M''(0) = \frac{2}{\lambda^2}$$

et la variance de X est

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= \frac{1}{\lambda^2} \end{aligned}$$

Exemple 7.44 Cas d'une distribution normale

SOLUTION. On déterminera d'abord la fonction génératrice des moments d'une variable normale standard. Désignons cette variable par Z . On a

$$\begin{aligned} M_Z(t) &= E[e^{tZ}] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x^2 - 2tx)}{2}\right\} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x-t)^2}{2} + \frac{t^2}{2}\right\} dx \end{aligned}$$

$$\begin{aligned}
 &= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2} dx \\
 &= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \quad \text{en substituant } y \text{ à } x - t \\
 &= e^{t^2/2}
 \end{aligned}$$

La fonction génératrice des moments d'une variable aléatoire Z normale centrée réduite est donc $M_Z(t) = e^{t^2/2}$. Pour obtenir la fonction génératrice des moments d'une variable normale quelconque, il faut se souvenir que $X = \mu + \sigma Z$ sera de distribution normale avec paramètres μ et σ^2 si Z est normale standard. On peut alors calculer la fonction génératrice des moments de X ainsi

$$\begin{aligned}
 M_X(t) &= E[e^{tX}] \\
 &= E[e^{t(\mu + \sigma Z)}] \\
 &= E[e^{t\mu} e^{t\sigma Z}] \\
 &= e^{t\mu} E[e^{t\sigma Z}] \\
 &= e^{t\mu} M_Z(t\sigma) \\
 &= e^{t\mu} e^{(t\sigma)^2/2} \\
 &= \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}
 \end{aligned}$$

On dérive pour obtenir

$$\begin{aligned}
 M'_X(t) &= (\mu + t\sigma^2) \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\} \\
 M''_X(t) &= (\mu + t\sigma^2)^2 \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\} + \sigma^2 \exp\left\{\frac{\sigma^2 t^2}{2} + \mu t\right\}
 \end{aligned}$$

donc

$$\begin{aligned}
 E[X] &= M'(0) = \mu \\
 E[X^2] &= M''(0) = \mu^2 + \sigma^2
 \end{aligned}$$

et finalement

$$\begin{aligned}
 \text{Var}(X) &= E[X^2] - E([X])^2 \\
 &= \sigma^2
 \end{aligned}$$

■

7.6.3 Tableaux de fonctions génératrices des moments

Les tableaux 7.2 et 7.3 donnent la fonction génératrice des moments de plusieurs distributions courantes.

Tableau 7.2

Lois (discrètes)	Loi de probabilité $p(x)$	Fonction génératrice des moments $M(t)$	Espérance	Variance
Binomiale de paramètres n, p $0 \leq p \leq 1$	$\binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, \dots, n$	$(pe^t + 1 - p)^n$	np	$np(1-p)$
Poisson de paramètre $\lambda > 0$	$e^{-\lambda} \frac{\lambda^x}{x!}$ $x = 0, 1, 2, \dots$	$\exp\{\lambda(e^t - 1)\}$	λ	λ
Géométrique de paramètre $0 \leq p \leq 1$	$p(1-p)^{x-1}$ $x = 1, 2, \dots$	$\frac{pe^t}{1 - (1-p)e^t}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Binomiale négative de paramètres r, p $0 \leq p \leq 1$	$\binom{n-1}{r-1} p^r (1-p)^{n-r}$ $n = r, r+1, \dots$	$\left[\frac{pe^t}{1 - (1-p)e^t} \right]^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

Tableau 7.3

Lois (continues)	Densité $f(x)$	Fonction génératrice des moments	
		Esérance	Variance
		$M(t)$	
Uniforme sur (a, b)	$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{sinon} \end{cases}$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$ $\frac{(b-a)^2}{12}$
Exponentielle de paramètre $\lambda > 0$	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$	$\frac{\lambda}{\lambda - t}$	$\frac{1}{\lambda}$ $\frac{1}{\lambda^2}$
Gamma de paramètres (s, λ) $\lambda > 0$	$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{s-1}}{\Gamma(s)} & x \geq 0 \\ 0 & x < 0 \end{cases}$	$\left(\frac{\lambda}{\lambda - t}\right)^s$	$\frac{s}{\lambda}$ $\frac{s}{\lambda^2}$
Normale de paramètres (μ, σ^2)	$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2} \quad -\infty < x < \infty$	$\exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$	μ σ^2

7.6.4 Autres propriétés des fonctions génératrices des moments

Théorème 7.7

La fonction génératrice des moments d'une somme de variables aléatoires indépendantes est égale au produit des fonctions génératrices des moments individuels de ces variables.

DÉMONSTRATION. Soient deux variables aléatoires indépendantes X et Y de fonctions génératrices des moments respectives M_X et M_Y . La fonction génératrice des moments de $X + Y$ est

$$\begin{aligned} M_{X+Y}(t) &= E[e^{t(X+Y)}] \\ &= E[e^{tX}e^{tY}] \\ &= E[e^{tX}]E[e^{tY}] \\ &= M_X(t)M_Y(t) \end{aligned}$$

où l'avant dernière égalité est une conséquence du théorème 7.2 et de l'indépendance de X et Y . ■

Une autre propriété très importante des fonctions génératrices est que leur donnée détermine de manière univoque la distribution des variables auxquelles elles correspondent. Plus précisément, si M_X et ses dérivées existent et sont finies dans un voisinage de 0, la distribution de X est entièrement déterminée par ces fonctions. Si par exemple $M_X(t) = (\frac{1}{2})^{10}(e^t + 1)^{10}$ on pourra affirmer, grâce au tableau 7.2 que X est une variable binomiale de paramètres 10 et $\frac{1}{2}$.

7.6.5 Exemples d'application

Exemple 7.45 On sait que la fonction génératrice des moments d'une variable X est $M(t) = e^{3(e^t-1)}$. Que vaudra $P\{X = 0\}$?

SOLUTION. Le tableau 7.2 indique que M est la fonction génératrice des moments d'une variable de Poisson de paramètre 3. En vertu de l'unicité de la distribution associée à une fonction génératrice, on conclut que X est une variable de Poisson de paramètre 3. Par conséquent, $P\{X = 0\} = e^{-3}$. ■

Exemple 7.46 Somme de variables binomiales indépendantes

Considérons deux variables binomiales X et Y indépendantes et de paramètres respectifs (n, p) et (m, p) . Quelle sera la distribution de $X + Y$?

SOLUTION. La fonction génératrice des moments de $X + Y$ est

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) = (pe^t + 1 - p)^n(pe^t + 1 - p)^m \\ &= (pe^t + 1 - p)^{m+n} \end{aligned}$$

On reconnaît ici dans $(pe^t + 1 - p)^{m+n}$ la fonction génératrice des moments d'une variable binomiale de paramètres $m + n$ et p . Cette distribution est donc celle de $X + Y$. ■

Exemple 7.47 Somme de variables aléatoires de Poisson

Si X et Y sont deux variables poissonniennes indépendantes de paramètres respectifs λ_1 et λ_2 , quelle est la distribution de leur somme?

SOLUTION.

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= \exp\{\lambda_1(e^t - 1)\} \exp\{\lambda_2(e^t - 1)\} \\ &= \exp\{(\lambda_1 + \lambda_2)(e^t - 1)\} \end{aligned}$$

Et par conséquent $X + Y$ est de distribution poissonnienne avec paramètre $\lambda_1 + \lambda_2$, ce qui confirme le résultat établi dans l'exemple 6.16. ■

Exemple 7.48 Somme de variables normales indépendantes

On veut montrer que si X et Y sont deux variables normales indépendantes de paramètres respectifs (μ_1, σ_1^2) et (μ_2, σ_2^2) alors $X + Y$ est normale d'espérance $\mu_1 + \mu_2$ et de variance $\sigma_1^2 + \sigma_2^2$.

SOLUTION.

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) M_Y(t) \\ &= \exp\left\{\frac{\sigma_1^2 t^2}{2} + \mu_1 t\right\} \exp\left\{\frac{\sigma_2^2 t^2}{2} + \mu_2 t\right\} \\ &= \exp\left\{\frac{(\sigma_1^2 + \sigma_2^2) t^2}{2} + (\mu_1 + \mu_2) t\right\} \end{aligned}$$

que l'on reconnaît être la fonction génératrice des moments d'une variable normale d'espérance $\mu_1 + \mu_2$ et de variance $\sigma_1^2 + \sigma_2^2$. Le résultat annoncé est donc acquis du fait de l'unicité de la distribution associée à une fonction génératrice des moments. ■

Exemple 7.49 Calcul de la fonction génératrice des moments d'une variable aléatoire chi-carrée à n degrés de liberté.

SOLUTION. On peut représenter une telle variable aléatoire par

$$Z_1^2 + \dots + Z_n^2$$

où Z_1, \dots, Z_n sont des variables aléatoires normales standard indépendantes. Soit $M(t)$ sa fonction génératrice des moments. D'après ce qui précède,

$$M(t) = \left(E\left[e^{tZ^2} \right] \right)^n$$

où Z est normale standard. On a

$$\begin{aligned}
 E\left[e^{tZ^2}\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} e^{-x^2/2} dx \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} dx \quad \text{où } \sigma^2 = (1-2t)^{-1} \\
 &= \sigma \\
 &= (1-2t)^{-1/2}
 \end{aligned}$$

où l'avant-dernière égalité utilise le fait que l'intégrale de la densité normale de moyenne 0 et de variance σ^2 vaut 1. Donc

$$M(t) = (1-2t)^{-n/2} \quad \blacksquare$$

Exemple 7.50 Fonction génératrice des moments d'une somme d'un nombre aléatoire de variables aléatoires. On considère une collection X_1, X_2, \dots de variables aléatoires indépendantes et identiquement distribuées. On considère également une variable N à valeurs entières positives indépendantes des $X_i, i = 1, 2, \dots$. On souhaite déterminer la fonction génératrice des moments puis l'espérance de

$$Y = \sum_{i=1}^N X_i$$

(Dans l'exemple 7.29 Y représentait la somme totale dépensée dans un magasin lorsqu'à la fois les dépenses de chaque client et le nombre de clients sont des variables aléatoires).

SOLUTION. Pour trouver la fonction génératrice des moments de Y , on conditionne tout d'abord selon N comme suit

$$\begin{aligned}
 E[e^{t\sum_{i=1}^N X_i} | N = n] &= E[e^{t\sum_{i=1}^n X_i} | N = n] \\
 &= E[e^{t\sum_{i=1}^n X_i}] \\
 &= (M_X(t))^n
 \end{aligned}$$

où

$$M_X(t) = E[e^{tX_i}]$$

Par conséquent

$$E[e^{tY} | N] = (M_X(t))^N$$

et donc

$$M_Y(t) = E[(M_X(t))^N]$$

Il suffit à ce stade de dériver pour obtenir les premiers moments de Y :

$$M'_Y(t) = E[N(M_X(t))^{N-1} M'_X(t)]$$

et par suite

$$\begin{aligned} E[Y] &= M'_Y(0) \\ &= E[N(M_X(0))^{N-1} M'_X(0)] \\ &= E[NEX] \\ &= E[N]E[X] \end{aligned} \quad (7.26)$$

où le résultat de l'exemple 7.29 se trouve confirmé. On a utilisé dans ce calcul la relation $M_X(0) = E[e^{0X}] = 1$. Ensuite,

$$M''_Y(t) = E[N(N-1)(M_X(t))^{N-2}(M'_X(t))^2 + N(M_X(t))^{N-1}M''_X(t)]$$

et donc

$$\begin{aligned} E[Y^2] &= M''_Y(0) \\ &= E[N(N-1)(E[X])^2 + NE[X^2]] \\ &= (E[X])^2(E[N^2] - E[N]) + E[N]E[X^2] \\ &= E[N](E[X^2] - (E[X])^2) + (E[X])^2E[N^2] \\ &= E[N] \text{Var}(X) + (E[X])^2E[N^2] \end{aligned} \quad (7.27)$$

De (7.26) et (7.27) on tire $\text{Var}(Y)$:

$$\begin{aligned} \text{Var}(Y) &= E[N] \text{Var}(X) + (E[X])^2(E[N^2] - (E[N])^2) \\ &= E[N] \text{Var}(X) + (E[X])^2 \text{Var}(N) \end{aligned} \quad \blacksquare$$

7.6.6 Fonction génératrice des moments de variables conjointes

Le concept de fonction génératrice des moments peut être étendu à des variables conjointes. On procède comme suit: pour toute collection de n variables aléatoires X_1, X_2, \dots, X_n leur *fonction génératrice des moments conjoints* M est définie pour toute collection de n arguments réels t_1, t_2, \dots, t_n par

$$M(t_1, \dots, t_n) = E[e^{t_1 X_1 + \dots + t_n X_n}]$$

Les fonctions génératrices des moments individuelles sont calculables à partir de M en posant tous les arguments sauf un égaux à zéro; en clair

$$M_{X_i}(t) = E[e^{tX_i}] = M(0, \dots, 0, t, 0, \dots, 0)$$

où t est le i -ème argument de M .

On peut démontrer que la donnée de M détermine de manière univoque la distribution conjointe de X_1, X_2, \dots, X_n . La démonstration est cependant trop difficile pour figurer dans cet ouvrage. Ce résultat permet d'en démontrer alors un autre.

Théorème 7.8

n variables aléatoires X_1, X_2, \dots, X_n seront indépendantes si et seulement si

$$M(t_1, \dots, t_n) = M_{X_1}(t_1) \cdots M_{X_n}(t_n) \quad (7.28)$$

DÉMONSTRATION. Supposons l'indépendance des n variables aléatoires, alors

$$\begin{aligned} M(t_1, \dots, t_n) &= E[e^{t_1 X_1 + \dots + t_n X_n}] \\ &= E[e^{t_1 X_1} \cdots e^{t_n X_n}] \\ &= E[e^{t_1 X_1}] \cdots E[e^{t_n X_n}] \quad \text{du fait de l'hypothèse précitée} \\ &= M_{X_1}(t_1) \cdots M_{X_n}(t_n) \end{aligned}$$

Dans l'autre sens, si (7.28) est vraie, la fonction génératrice des moments conjoints M est identique à celle de n variables aléatoires indépendantes, la i -ème desquelles aura la même distribution que X_i . Comme la fonction génératrice des moments conjoints détermine une distribution conjointe de manière unique, la distribution reconnue est bien celle des X_i conjointement. Ces variables sont donc indépendantes. ■

7.7 AUTRES PROPRIÉTÉS DES VARIABLES ALÉATOIRES NORMALES

7.7.1 Distribution normale multivariée

On considère un ensemble de n variables aléatoires indépendantes et normales standard notées Z_1, \dots, Z_n . On définit des variables aléatoires X_1, X_2, \dots, X_m grâce aux équations

$$\begin{aligned} X_1 &= a_{11}Z_1 + \dots + a_{1n}Z_n + \mu_1 \\ X_2 &= a_{21}Z_1 + \dots + a_{2n}Z_n + \mu_2 \\ &\vdots \\ X_i &= a_{i1}Z_1 + \dots + a_{in}Z_n + \mu_i \\ &\vdots \\ X_m &= a_{m1}Z_1 + \dots + a_{mn}Z_n + \mu_m \end{aligned}$$

où a_{ij} , $1 \leq i \leq m$, $1 \leq j \leq n$ et μ_i , $1 \leq i \leq m$ sont des constantes. Tout jeu de variables X_i , $1 \leq i \leq m$ ainsi définies est appelé groupe de **variables aléatoires normales multivariées**.

La somme de variables aléatoires normales indépendantes étant elle-même de distribution normale d'après le résultat de l'exemple 7.48, on remarquera que chaque X_i est une variable aléatoire normale d'espérance et de variance données par

$$\begin{aligned} E[X_i] &= \mu_i \\ \text{Var}(X_i) &= \sum_{j=1}^n a_{ij}^2 \end{aligned}$$

Considérons à présent la fonction génératrice conjointe des moments de X_1, \dots, X_m

$$M(t_1, \dots, t_m) = E\left[\exp\{t_1 X_1 + \dots + t_m X_m\}\right]$$

Comme $\sum_{i=1}^m t_i X_i$ est une combinaison linéaire des variables aléatoires normales indépendantes Z_1, \dots, Z_n , elle est aussi normalement distribuée. Sa moyenne et sa variance sont

$$E\left[\sum_{i=1}^m t_i X_i\right] = \sum_{i=1}^m t_i \mu_i$$

et

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^m t_i X_i\right) &= \text{Cov}\left(\sum_{i=1}^m t_i X_i, \sum_{j=1}^m t_j X_j\right) \\ &= \sum_{i=1}^m \sum_{j=1}^m t_i t_j \text{Cov}(X_i, X_j) \end{aligned}$$

Si Y est une variable aléatoire normale de moyenne μ et de variance σ^2 , alors

$$E[e^Y] = M_Y(t) \Big|_{t=1} = e^{\mu + \frac{\sigma^2}{2}}$$

Nous voyons donc que

$$M(t_1, \dots, t_m) = \exp\left\{\sum_{i=1}^m t_i \mu_i + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m t_i t_j \text{Cov}(X_i, X_j)\right\}$$

ce qui montre que la distribution conjointe de X_1, \dots, X_m est complètement déterminée par la connaissance des valeurs de $E[X_i]$ et $\text{Cov}(X_i, X_j)$, $i, j = 1, \dots, m$.

7.7.2 Distribution conjointe de la moyenne et de la variance d'un échantillon

Soient X_1, \dots, X_n des variables aléatoires normales indépendantes, chacune ayant pour moyenne μ et pour variance σ^2 . Soit $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$ la moyenne de l'échantillon.

Puisque la somme de variables aléatoires normales indépendantes est aussi une variable aléatoire normale, \bar{X} est donc une variable aléatoire normale d'espérance μ et de variance σ^2/n (voir les exemples 7.3 et 7.19).

Rappelons que selon l'exemple 7.24

$$\text{Cov}(\bar{X}, X_i - \bar{X}) = 0 \quad i = 1, \dots, n \quad (7.29)$$

De plus, puisque $\bar{X}, X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X}$ sont toutes des combinaisons linéaires des normales standard indépendantes $(X_i - \mu)/\sigma$, $i = 1, \dots, n$, alors

$\bar{X}, X_i - \bar{X}, i = 1, \dots, n$ possède une distribution conjointe normale multivariée. Si Y est une variable aléatoire normale de moyenne μ et de variance σ^2/n indépendante des $X_i, i = 1, \dots, n$ alors $Y, X_i - \bar{X}, i = 1, \dots, n$ possède aussi une distribution normale multivariée et, à cause de (7.29), a les mêmes espérances et covariances que les variables aléatoires $\bar{X}, X_i - \bar{X}, i = 1, \dots, n$. Mais comme une distribution normale multivariée est complètement déterminée par ses espérances et ses covariances, on peut conclure que $Y, X_i - \bar{X}, i = 1, \dots, n$ et $\bar{X}, X_i - \bar{X}, i = 1, \dots, n$ ont la même distribution conjointe, ce qui montre que \bar{X} est indépendant de la séquence des déviations $X_i - \bar{X}, i = 1, \dots, n$.

Puisque \bar{X} est indépendant de la séquence des déviations $X_i - \bar{X}, i = 1, \dots, n$, \bar{X} est aussi indépendant de la variance de l'échantillon $S^2/(n-1) \equiv \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$.

Comme nous savons déjà que \bar{X} est normal de moyenne μ et de variance σ^2/n , il reste seulement à déterminer la distribution de S^2 . Pour faire ceci, reprenons l'identité algébrique de l'exemple 7.19

$$\begin{aligned} S^2 &\equiv \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \end{aligned}$$

En divisant cette équation par σ^2 , on obtient

$$\frac{S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \quad (7.30)$$

Or

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2$$

est la somme des carrés de n variables aléatoires normales standard et est donc une variable aléatoire chi-carrée à n degrés de liberté. Donc, selon l'exemple 7.49, sa fonction génératrice des moments est $(1 - 2t)^{-n/2}$. Ainsi,

$$\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2$$

est le carré d'une normale standard et est donc une variable aléatoire chi-carrée à 1 degré de liberté dont la fonction génératrice des moments est $(1 - 2t)^{-1/2}$. Nous avons vu précédemment que les deux variables aléatoires du membre gauche de l'équation (7.30) sont indépendantes. Donc, comme la fonction génératrice des moments d'une somme de variables indépendantes est égale au produit de leurs fonctions génératrices des moments, on voit que

$$E \left[e^{tS^2/\sigma^2} \right] (1 - 2t)^{-1/2} = (1 - 2t)^{-n/2}$$

ou

$$E\left[e^{tS^2/\sigma^2}\right] = (1 - 2t)^{-(n-1)/2}$$

Mais comme $(1 - 2t)^{-(n-1)/2}$ est la fonction génératrice des moments d'une variable aléatoire chi-carrée à $n - 1$ degrés de liberté, on peut conclure que, puisque la fonction génératrice des moments détermine de manière unique la distribution de la variable aléatoire, c'est la distribution de S^2/σ^2 .

En résumé, nous avons le théorème suivant.

Théorème 7.9

Si X_1, \dots, X_n sont des variables aléatoires normales indépendantes et identiquement distribuées de moyenne μ et de variance σ^2 , la moyenne de l'échantillon \bar{X} et la variance de l'échantillon $S^2/(n-1)$ sont indépendantes. \bar{X} est une variable aléatoire normale de moyenne μ et de variance σ^2/n ; S^2/σ^2 est une variable aléatoire chi-carrée à $n - 1$ degrés de liberté.

7.8 DÉFINITION GÉNÉRALE DE L'ESPÉRANCE MATHÉMATIQUE

7.8.1 Cadre du problème

Nous n'avons jusqu'à présent défini d'espérance que pour des variables discrètes ou continues. Or il existe des variables qui n'appartiennent à aucune de ces deux catégories et pour lesquelles il est malgré tout intéressant d'essayer de définir une espérance. A titre d'exemple, considérons la variable aléatoire W construite ainsi: soient X une variable de Bernoulli de paramètre $p = \frac{1}{2}$ et Y une variable uniforme sur $[0, 1]$. On pose

$$W = \begin{cases} X & \text{lorsque } X = 1 \\ Y & \text{lorsque } X \neq 1 \end{cases}$$

La variable W n'est manifestement ni discrète (son ensemble fondamental $[0, 1]$ étant non dénombrable), ni continue (discontinuité de F_W en 1 puisque $P\{W = 1\} = \frac{1}{2}$).

Il s'avère utile pour définir l'espérance de variables aléatoires absolument quelconques de disposer d'un outil, à savoir l'intégrale de Stieltjes définie ci-après.

7.8.2 Intégrale de Stieltjes

Avant de définir cette intégrale, procédons à un rappel. On se souvient que pour toute fonction g , $\int_a^b g(x) dx$ est définie ainsi:

$$\int_a^b g(x) dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n g(x_i)(x_i - x_{i-1})$$

où la limite est prise sur toutes les divisions $a = x_0 < x_1 < x_2 < \dots < x_n = b$ avec $n \rightarrow \infty$ et $\max_{i=1, \dots, n} (x_i - x_{i-1}) \rightarrow 0$.

On définit maintenant pour toute fonction de répartition F , l'*intégrale de Stieltjes* d'une fonction à valeurs non négatives g sur l'intervalle $[a, b]$ par

$$\int_a^b g(x) dF(x) = \lim \sum_{i=1}^n g(x_i)[F(x_i) - F(x_{i-1})]$$

où comme précédemment, la limite est prise sur toutes les divisions $a = x_0 < x_1 < x_2 < \dots < x_n = b$ de $[a, b]$ avec $n \rightarrow \infty$ et $\max_{i=1, \dots, n} (x_i - x_{i-1}) \rightarrow 0$. On peut étendre cette définition à l'axe réel tout entier en posant

$$\int_{-\infty}^{\infty} g(x) dF(x) = \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_a^b g(x) dF(x)$$

Enfin, si g est une fonction à valeurs quelconques non nécessairement positives, on définit g^+ et g^-

$$g^+(x) = \begin{cases} g(x) & \text{lorsque } g(x) \geq 0 \\ 0 & \text{lorsque } g(x) < 0 \end{cases}$$

$$g^-(x) = \begin{cases} 0 & \text{lorsque } g(x) \geq 0 \\ -g(x) & \text{lorsque } g(x) < 0 \end{cases}$$

Les fonctions g^+ et g^- étant à valeurs non négatives et g s'écrivant $g = g^+ - g^-$, il est naturel de définir en tout généralité

$$\int_{-\infty}^{\infty} g(x) dF(x) = \int_{-\infty}^{\infty} g^+(x) dF(x) - \int_{-\infty}^{\infty} g^-(x) dF(x)$$

et nous dirons que $\int_{-\infty}^{\infty} g(x) dF(x)$ existe dès que $\int_{-\infty}^{\infty} g^+(x) dF(x)$ et $\int_{-\infty}^{\infty} g^-(x) dF(x)$ ne sont pas toutes deux égales à $+\infty$.

7.8.3 Espérance d'une variable aléatoire quelconque

Soit X , une variable aléatoire quelconque de fonction de répartition F . On définit l'espérance de X par

$$E[X] = \int_{-\infty}^{\infty} x dF(x) \quad (7.31)$$

On peut montrer que lorsque X est une variable discrète de loi de probabilité p ,

$$\int_{-\infty}^{\infty} x dF(x) = \sum_{x: p(x) > 0} xp(x)$$

tandis que si X est continue de densité f

$$\int_{-\infty}^{\infty} x dF(x) = \int_{-\infty}^{\infty} xf(x) dx$$

Le lecteur remarquera que (7.31) correspond intuitivement à la définition précédente de $E[X]$ comprise comme moyenne. Si en effet, on considère l'approximation suivante de $E[X]$, à savoir

$$\sum_{i=1}^n x_i [F(x_i) - F(x_{i-1})]$$

on remarque que $F(x_i) - F(x_{i-1})$ n'est autre que la probabilité que X soit dans l'intervalle $(x_{i-1}, x_i]$. L'approximation citée revient donc à multiplier une valeur approximative de X lorsque la variable prend une valeur dans $(x_{i-1}, x_i]$ par la probabilité que cet événement se produise. Il est prévisible que lorsque les intervalles utilisés deviennent de plus en plus fins le résultat sera une bonne moyenne pondérée.

L'intégrale de Stieltjes présente un intérêt plutôt sur le plan théorique où elle constitue un outil synthétique pour la définition et l'exploitation des propriétés de l'espérance. Son usage dispense en particulier de l'élaboration de démonstrations distinctes pour traiter les cas continu et discret. Comme par ailleurs, ses propriétés sont assez voisines de celles de l'intégrale ordinaire, on peut facilement adapter les démonstrations données dans ce chapitre pour leur donner une validité générale.

7.9 EXERCICES THÉORIQUES

7.9.1 Montrer que $E[(X - a)^2]$ est minimisée en $a = E[X]$.

7.9.2 Supposer que X est une variable aléatoire continue de densité f . Montrer que $E[|X - a|]$ est minimisée lorsque a est égal à la médiane de F .
Pour cela, écrire

$$E[|X - a|] = \int |x - a| f(x) dx$$

puis fractionner l'intégrale en deux régions où $x < a$ et $x > a$ et dériver.

7.9.3 Démontrer le théorème 7.1 lorsque

- a) X et Y ont une fonction de distribution conjointe;
- b) X et Y ont une densité conjointe et que $g(x, y) \geq 0$ pour tout x, y .

7.9.4 On considère une variable aléatoire X d'espérance μ finie et de variance σ^2 ainsi qu'une fonction g deux fois dérivable. Montrer que

$$E[g(X)] = g(\mu) + \frac{g''(\mu)}{2} \sigma^2$$

On pourra pour cela développer g en série de Taylor et négliger les termes au-delà du troisième.

7.9.5 Soient A_1, A_2, \dots, A_n des événements quelconques et soit $C_k = \{\text{au moins } k \text{ événements } A_i \text{ se produisent}\}$. Montrer que

$$\sum_{k=1}^n P(C_k) = \sum_{k=1}^n P(A_k)$$

Pour cela, soit X le nombre de A_i qui se produisent. Montrer que les deux membres de l'égalité ci-dessus sont égaux à $E[X]$.

7.9.6 Dans le texte, on a noté que

$$E\left[\sum_{i=1}^{\infty} X_i\right] = \sum_{i=1}^{\infty} E[X_i]$$

lorsque les X_i sont toutes des variables aléatoires non négatives. Puisqu'une intégrale est une limite de sommes, on peut s'attendre à ce que

$$E\left[\int_0^{\infty} X(t) dt\right] = \int_0^{\infty} E[X(t)] dt$$

tant que $X(t)$, $0 \leq t < \infty$, sont des variables aléatoires non négatives; et ce résultat est effectivement vrai. Utilisez-le pour donner une autre preuve du résultat suivant: pour une variable aléatoire non négative X ,

$$E[X] = \int_0^{\infty} P\{X > t\} dt$$

Pour cela, définir pour chaque t non négatif la variable aléatoire $X(t)$ par

$$X(t) = \begin{cases} 1 & \text{si } t < X \\ 0 & \text{si } t \geq X \end{cases}$$

Relier ensuite $\int_0^{\infty} X(t) dt$ à X .

7.9.7 On dit que X est plus grand en probabilité que Y , et on note $X \geq_{st} Y$ si, pour tout t ,

$$P\{X > t\} \geq P\{Y > t\}$$

Montrer que si $X \geq_{st} Y$, alors $E[X] \geq E[Y]$ lorsque

- X et Y sont des variables aléatoires non négatives;
- X et Y sont des variables aléatoires quelconques.

Pour cela, écrire X comme

$$X = X^+ - X^-$$

où

$$X^+ = \begin{cases} X & \text{si } X \geq 0 \\ 0 & \text{si } X < 0 \end{cases} \quad X^- = \begin{cases} 0 & \text{si } X \geq 0 \\ -X & \text{si } X < 0 \end{cases}$$

De la même manière, représenter Y comme $Y^+ - Y^-$. Utiliser alors la question a).

7.9.8 Montrer que X est plus grand en probabilité que Y si et seulement si

$$E[f(X)] \geq E[f(Y)]$$

pour toute fonction croissante f .

Pour cela, si $X \geq_{st} Y$, montrer que $E[f(X)] \geq E[f(Y)]$ en montrant que $f(X) \geq_{st} f(Y)$ puis utiliser l'exercice 7.9.7. Pour montrer que $E[f(X)] \geq E[f(Y)]$ pour toutes les fonctions croissantes f implique que $P\{X > t\} \geq P\{Y > t\}$, définir une fonction croissante f appropriée.

7.9.9 On lance n fois une pièce donnant face avec probabilité p . Calculer l'espérance du nombre de chaînes de faces dans le résultat dont la longueur soit 1, puis 2, puis k , $1 \leq k \leq n$.

7.9.10 Soient X_1, X_2, \dots, X_n des variables aléatoires positives indépendantes et identiquement distribuées. Calculer pour $k \leq n$

$$E \left[\frac{\sum_{i=1}^k X_i}{\sum_{i=1}^n X_i} \right]$$

7.9.11 On considère n épreuves indépendantes pouvant chacune aboutir à r résultats différents avec les probabilités P_1, P_2, \dots, P_r . Désignons par X , le nombre de résultats qui ne sont pas représentés au bout de n tirages. Calculer $E[X]$ puis montrer que parmi tous les vecteurs P_1, P_2, \dots, P_r , celui qui minimise $E[X]$ est celui pour lequel $P_i = 1/r$, $i = 1, 2, \dots, r$.

7.9.12 On effectue une série d'épreuves indépendantes. La i -ème épreuve débouche sur un succès avec probabilité P_i . Calculer l'espérance puis la variance du nombre des succès au cours des n épreuves. L'hypothèse d'indépendance est-elle significative pour chacun de ces deux calculs?

7.9.13 Soient X_1, \dots, X_n des variables aléatoires indépendantes, identiquement distribuées et continues. On dira qu'un pic est observé au temps j , $j \leq n$, si $X_j \geq X_i$ pour tout i tel que $1 \leq i \leq j$. Montrer que

$$a) E[\text{nombre de pics}] = \sum_{j=1}^n 1/j.$$

$$b) \text{Var}(\text{nombre de pics}) = \sum_{j=1}^n (j-1)/j^2.$$

7.9.14 Pour l'exemple 7.10 montrer que la variance du nombre de bons nécessaires pour obtenir une collection complète est égale à

$$\sum_{i=1}^{N-1} \frac{iN}{(N-i)^2} 2$$

Quand N est grand, on peut montrer que cette expression est approximativement égale à $N^2\pi^2/6$ (c'est-à-dire que leur rapport tend vers 1 lorsque $N \rightarrow \infty$).

7.9.15 On considère une série de n épreuves indépendantes. La i -ème donne un succès avec probabilité P_i .

- Calculer l'espérance du nombre de succès sur l'ensemble des n épreuves. On la notera μ ;
- sous la condition que μ reste constant, quel est le choix des probabilités P_1, \dots, P_n pour lequel la variance du nombre de succès devient maximale?
- A l'inverse, quel choix rendra-t-il cette variance minimale?

7.9.16 Supposons que des boules soient prélevées au hasard dans une urne contenant initialement n boules blanches et m boules noires. On a montré, dans l'exemple 7.13, que $E[X] = 1 + m/(n+1)$ où X est le nombre de tirages nécessaires pour obtenir une boule blanche.

- Calculer $\text{Var}(X)$;
- montrer que l'espérance du nombre de boules qu'il faut tirer pour amasser un total de k boules blanches est $k[1 + m/(n+1)]$.

Pour cela considérer des variables aléatoires Y_i , $i = 1, \dots, n+1$, qui représentent le nombre de boules noires retirées après les $(i-1)$ premières boules blanches et avant la i -ème boule blanche. Utiliser le fait que les Y_i , $i = 1, \dots, n+1$ sont identiquement distribuées.

7.9.17 On considère deux variables aléatoires indépendantes X_1 et X_2 de même espérance μ . On sait également que $\text{Var}(X_1) = \sigma_1^2$ et $\text{Var}(X_2) = \sigma_2^2$. La valeur de μ est inconnue et on se propose de l'estimer grâce à une moyenne pondérée de X_1 et X_2 . La forme générale de cette estimation sera donc $\lambda X_1 + (1 - \lambda) X_2$, où λ reste à choisir. Pour quel choix de λ , l'estimation aura-t-elle la variance la plus faible? Dire pourquoi cette valeur de λ est intéressante.

7.9.18 Dans l'exemple 7.25 nous avons montré que la covariance des variables aléatoires multinomiales N_i et N_j est égale à $-mP_iP_j$. Pour cela, nous avons exprimé N_i et N_j comme une somme de variables indicatrices. On peut aussi obtenir ce résultat en utilisant la formule

$$\text{Var}(N_i + N_j) = \text{Var}(N_i) + \text{Var}(N_j) + 2\text{Cov}(N_i, N_j)$$

- Quelle est la loi de $N_i + N_j$?
- Utiliser l'identité précédente pour montrer que $\text{Cov}(N_i, N_j) = -mP_iP_j$

7.9.19 Si X et Y sont identiquement distribuées, non nécessairement indépendantes, montrer que

$$\text{Cov}(X + Y, X - Y) = 0$$

7.9.20 Formule de la covariance conditionnelle

La covariance conditionnelle de X et Y , étant donné Z , est définie par

$$\text{Cov}(X, Y|Z) \equiv E[(X - E\{X|Z\})(Y - E\{Y|Z\})|Z]$$

a) Montrer que

$$\text{Cov}(X, Y|Z) = E[XY|Z] - E[X|Z]E[Y|Z]$$

b) Montrer la formule de la covariance conditionnelle

$$\text{Cov}(X, Y) = E\{\text{Cov}(X, Y|Z)\} + \text{Cov}(E[X|Z], E[Y|Z])$$

c) Poser $X = Y$ dans b) et énoncer la formule de la variance conditionnelle.

7.9.21 Soient $X_{(i)}$, $i = 1, \dots, n$ les statistiques d'ordre d'un ensemble de n variables aléatoires uniformes sur $(0, 1)$. On sait que la densité de $X_{(i)}$ est donnée par

$$f(x) = \frac{n!}{(i-1)!(n-i)!} x^{i-1}(1-x)^{n-i}, \quad 0 < x < 1$$

a) Calculer $\text{Var}(X_{(i)})$, $i = 1, \dots, n$.

b) Pour quelles valeurs de i $\text{Var}(X_{(i)})$ atteint-elle son

- minimum;
- maximum?

7.9.22 Montrer que si $Y = a + bX$, alors

$$\rho(X, Y) = \begin{cases} +1 & \text{si } b > 0 \\ -1 & \text{si } b < 0 \end{cases}$$

7.9.23 Soit Z , une variable aléatoire normale centrée réduite. On pose $Y = a + bZ + cZ^2$. Montrer que

$$\rho(Y, Z) = \frac{b}{\sqrt{b^2 + 2c^2}}$$

7.9.24 Prouver l'inégalité de Cauchy-Schwarz, à savoir:

$$(E[XY])^2 \leq E[X^2]E[Y^2]$$

On remarquera qu'à moins qu'il existe une constante t telle que $Y = -tX$, auquel cas, cette inégalité devient une égalité, on peut écrire que pour tout t ,

$$0 < E[(tX + Y)^2] = E[X^2]t^2 + 2E[XY]t + E[Y^2]$$

Aussi, les racines de l'équation quadratique

$$E[X^2]t^2 + 2E[XY]t + E[Y^2] = 0$$

doivent-elles être imaginaires, ce qui implique que le discriminant de cette équation quadratique doit être négatif.

7.9.25 Montrer que pour deux variables X et Y indépendantes

$$E[X | Y = y] = E[X] \quad \text{pour tout } y$$

aussi bien dans le cas continu que discret.

7.9.26 Montrer que

$$E[g(X)Y | X] = g(X)E[Y | X].$$

7.9.27 Montrer que lorsque $E[Y | X = x] = E[Y]$ pour tout x alors X et Y sont non corrélées, puis donner un contre-exemple établissant que la réciproque n'est pas vraie. On pourra pour cela montrer puis utiliser la relation $E[XY] = E[XE[Y | X]]$.

7.9.28 Montrer que $\text{Cov}(X, E[Y | X]) = \text{Cov}(X, Y)$.

7.9.29 On considère des variables aléatoires X_1, \dots, X_n indépendantes et identiquement distribuées. Calculer

$$E[X_1 | X_1 + \dots + X_n = x]$$

7.9.30 On considère l'exemple 7.25 qui traite d'une distribution multinomiale. Utiliser l'espérance conditionnelle pour le calcul de $E[N_i N_j]$ puis utiliser ce résultat pour vérifier la formule donnant $\text{Cov}(N_i, N_j)$ citée dans ce même exemple.

7.9.31 Une urne contient au départ n boules noires et b boules blanches. A chaque étape on ajoute r boules noires puis retire r boules au hasard prises parmi les $n + b + r$ boules présentes. Montrer que

$$E[\text{nombre des boules blanches à la fin de l'étape } t] = \left(\frac{b + n}{b + n + r} \right)^t n.$$

7.9.32 Démontrer l'équation (7.12).

7.9.33 Une pièce est caractérisée par une probabilité p de tomber sur face. On la lance indéfiniment. Calculer l'espérance du nombre de jets qu'il faudra jusqu'à ce qu'une chaîne de r résultats de type face consécutifs apparaisse.

On pourra conditionner sur la date d'apparition du premier pile, ce qui amène à l'équation

$$E[X] = (1 - p) \sum_{i=1} p^{i-1} (i + E[X]) + (1 - p) \sum_{i=r+1} p^{i-1} r$$

Après simplification, on peut la résoudre par rapport à $E[X]$.

7.9.34 Pour avoir une autre approche de l'exercice 7.9.33, appeler T_r le nombre de jets nécessaires pour obtenir une série de r faces consécutives.

a) Déterminer $E[T_r | T_{r-1}]$.

b) Déterminer $E[T_r]$ en fonction de $E[T_{r-1}]$.

c) Qu'est-ce que $E[T_1]$?

d) Qu'est-ce que $E[T_r]$?

7.9.35 a) Montrer que

$$E[X] = E[X|X < a]P\{X < a\} + E[X|X \geq a]P\{X \geq a\}$$

Pour cela, définir une variable aléatoire appropriée et calculer $E[X]$ en conditionnant sur cette variable.

b) Utiliser la question a) pour démontrer l'inégalité de Markov, qui affirme que si $P\{X \geq 0\} = 1$, alors pour $a > 0$

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

7.9.36 On tire une à une des boules dans une urne en contenant b blanches et n noires. On continue jusqu'à ce que toutes les boules restantes soient de la même couleur. Soit $M_{b,n}$ le nombre attendu de boules dans l'urne à la fin de l'expérience. Calculer une formule récursive donnant $M_{b,n}$ et la résoudre pour $b = 3$ et $n = 5$.

7.9.37 Une urne contient b boules blanches et n boules noires. Lorsqu'une boule est tirée, elle est remise dans l'urne s'il s'agit d'une boule blanche. S'il s'agit d'une noire, elle est remplacée par une boule blanche provenant d'une autre urne. Soit M_b l'espérance du nombre attendu de boules blanches restant dans l'urne après que l'opération ait été répétée k fois.

a) Etablir l'équation de récurrence:

$$M_{k+1} = \left(1 - \frac{1}{b+n}\right) M_k + 1$$

b) Utiliser cette équation pour prouver que

$$M_k = b + n - n \left(1 - \frac{1}{b+n}\right)^k$$

c) Quelle est la probabilité que la $(k+1)$ -ième boule tirée soit blanche?

7.9.38 Le meilleur prédicteur linéaire de Y basé sur X_1 et X_2 est $a + bX_1 + cX_2$, où a , b et c doivent être choisis de manière à minimiser

$$E[(Y - (a + bX_1 + cX_2))^2].$$

Déterminer a , b et c .

7.9.39 Le meilleur prédicteur quadratique de Y basé sur X est $a + bX + cX^2$, où a , b et c doivent être choisis de manière à minimiser

$$E[(Y - (a + bX + cX^2))^2].$$

Déterminer a , b et c .

7.9.40 On considère X et Y , deux variables aléatoires conjointement normales de densité

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right] \right\}$$

- a) Montrer que la loi conditionnelle de Y lorsque $X = x$ est normale d'espérance $\mu_y + \rho(\sigma_y/\sigma_x)(x - \mu_x)$ et de variance $\sigma_y^2(1 - \rho^2)$;
 b) montrer que $\text{Corr}(X, Y) = \rho$;
 c) montrer que X et Y sont indépendantes si et seulement si $\rho = 0$.

7.9.41 Soient X une variable aléatoire normale de paramètres $\mu = 0$ et $\sigma = 1$ et I , indépendante de X , telle que $P\{I = 1\} = \frac{1}{2} = P\{I = 0\}$. On définit Y ainsi

$$Y = \begin{cases} X & \text{si } I = 1 \\ -X & \text{si } I = 0 \end{cases}$$

En d'autres termes, Y a autant de chances d'être égale à X ou à $-X$.

- a) X et Y sont-elles indépendantes?
 b) I et Y sont-elles indépendantes?
 c) Montrer que Y est normale, d'espérance 0 et de variance 1.
 d) Montrer que $\text{Cov}(X, Y) = 0$.

7.9.42 Du théorème 7.6 et du fait que le meilleur prédicteur linéaire de Y basé sur X est $\mu_y + \rho(\sigma_y/\sigma_x)(X - \mu_x)$ il résulte que si

$$E[Y|X] = a + bX$$

alors

$$a = \mu_y - \rho \frac{\sigma_y}{\sigma_x} \mu_x \quad b = \rho \frac{\sigma_y}{\sigma_x}$$

(pourquoi?). Vérifier ceci par un calcul immédiat.

7.9.43 Montrer que pour toutes variables aléatoires X et Z

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

où

$$Y = E[X|Z]$$

7.9.44 On considère une population dont les individus sont capables de produire seuls une descendance identique à eux-mêmes. On admet que chaque individu a une probabilité $P_j, j \geq 0$, d'avoir engendré j descendants au terme de sa vie, ceci indépendamment de l'activité des autres individus. Le nombre initial d'individus dans la population est noté X_0 et appelé taille de la 0-ième génération. Tous les descendants de cette génération 0 sont de la première génération, de taille X_1 . De manière générale X_n représente la taille de la n -ième génération. L'espérance et la variance de la descendance immédiate d'un individu donné seront respectivement notées $\mu = \sum_{j=0}^{\infty} jP_j$ et $\sigma^2 = \sum_{j=0}^{\infty} (j - \mu)^2 P_j$. On admettra que $X_0 = 1$, c'est-à-dire que la population provient à l'origine d'un individu unique.

a) Montrer que

$$E[X_n] = \mu E[X_{n-1}]$$

b) En déduire que

$$E[X_n] = \mu^n$$

c) Montrer que

$$\text{Var}(X_n) = \sigma^2 \mu^{n-1} + \mu^2 \text{Var}(X_{n-1})$$

d) En déduire que

$$\text{Var}(X_n) = \begin{cases} \sigma^2 \mu^{n-1} \left(\frac{\mu^n - 1}{\mu - 1} \right) & \text{si } \mu \neq 1 \\ n\sigma^2 & \text{si } \mu = 1 \end{cases}$$

Ce que l'on vient d'étudier est appelé *processus de ramification*. Une question importante pour une population évoluant selon ces lois est de connaître la probabilité de son extinction. Notons π , cette probabilité sous l'hypothèse que la population descend d'un ancêtre initial unique; en clair,

$$\pi = P(\text{la population s'éteigne} | X_0 = 1)$$

e) Prouver que

$$\pi = \sum_{j=0}^{\infty} P_j \pi^j$$

On pourra pour cela conditionner sur le nombre de descendants immédiats de l'ancêtre originel.

7.9.45 Vérifier la formule donnant la fonction génératrice des moments d'une variable aléatoire uniforme, telle qu'elle est donnée dans le tableau 7.3. Vérifier également les formules donnant l'espérance et la variance par dérivation.

7.9.46 Pour une variable aléatoire normale standard Z , soit $\mu_n = E[Z^n]$. Montrer que

$$\mu_n = \begin{cases} 0 & \text{si } n \text{ est impair} \\ \frac{(2j)!}{2^j j!} & \text{si } n = 2j \end{cases}$$

Commencer par développer la fonction génératrice des moments de Z en série de Taylor autour de 0 pour obtenir

$$\begin{aligned} E[e^{tZ}] &= e^{t^2/2} \\ &= \sum_{j=0}^{\infty} \frac{(t^2/2)^j}{j!} \end{aligned}$$

7.9.47 Soit X une variable aléatoire normale de moyenne μ et de variance σ^2 . Utiliser les résultats de l'exercice 7.9.46 pour montrer que

$$E[X^n] = \sum_{j=0}^{\lfloor n/2 \rfloor} \frac{\binom{n}{2j} \mu^{n-2j} \sigma^{2j} (2j)!}{2^j j!}$$

Dans cette expression, $\lfloor n/2 \rfloor$ est le plus grand entier inférieur ou égal à $n/2$. Vérifier votre réponse pour $n=1$ et $n=2$.

7.9.48 Soit $Y = aX + b$, où a et b sont des constantes. Exprimer la fonction génératrice des moments de Y en fonction de celle de X .

7.9.49 Soit X une variable de fonction génératrice des moments M . On pose $\Psi(t) = (\ln M(t))$. Montrer que

$$\Psi''(t)|_{t=0} = \text{Var}(X)$$

7.9.50 A l'aide du tableau 7.3, déterminer la distribution de $\sum_{i=1}^n X_i$ lorsque X_1, \dots, X_n sont des variables exponentielles indépendantes et identiquement distribuées d'espérance commune $1/\lambda$.

7.9.51 Montrer comment on peut calculer $\text{Cov}(X, Y)$ à partir de la fonction génératrice des moments conjoints de X et Y .

7.9.52 Soient X_1, \dots, X_n des variables de distribution normale multivariée. Montrer qu'elles seront indépendantes si et seulement si

$$\text{Cov}(X_i, X_j) = 0 \quad \text{pour } i \neq j$$

7.9.53 Si Z est une variable aléatoire normale standard, que vaut $\text{Cov}(Z, Z^2)$?

7.10 PROBLÈMES

7.10.1 Si X et Y sont des variables aléatoires uniformes $(0, 1)$ indépendantes, montrer que

$$E[|X - Y|^\alpha] = \frac{2}{(\alpha+1)(\alpha+2)} \quad \text{pour } \alpha > 0$$

7.10.2 Soient X et Y des variables aléatoires indépendantes, ayant autant de chance l'une et l'autre de prendre n'importe quelle valeur parmi $1, 2, \dots, m$. Montrer que

$$E[|X - Y|] = \frac{(m-1)(m+1)}{3m}$$

7.10.3 Un hôpital est situé au centre d'un parc dont les côtés mesurent 3 km. Si un accident se produit dans le parc, l'hôpital envoie une ambulance. Le réseau de routes est rectangulaire, donc la distance de l'hôpital, dont les coordonnées sont $(0, 0)$, au point (x, y) est $|x| + |y|$. Si un accident se produit en un point uniformément distribué dans le parc, trouver l'espérance de la longueur du trajet de l'ambulance.

7.10.4 On lance 10 fois un dé. Calculer l'espérance de la somme des 10 jets.

7.10.5 N personnes arrivent séparément à un dîner professionnel. En arrivant, chaque personne regarde si elle a un ami parmi les gens déjà présents. Une personne s'assoit soit à la table d'un ami, soit à une table vide si aucun de ses amis n'est présent. En posant que chacune des $\binom{N}{2}$ paires de gens sont, indépendamment, amis avec une probabilité p , trouver l'espérance du nombre de tables occupées.

Poser X_i égal à 1 ou 0 selon si le i -ème arrivant s'assoit à une table inoccupée ou non.

7.10.6 On place n balles, numérotées de 1 à n , dans n urnes, aussi numérotées de 1 à n , de telle façon que la balle i ait autant de chance de se trouver dans les urnes $1, 2, \dots, i$. Trouver

- l'espérance du nombre d'urnes vides;
- la probabilité qu'aucune urne ne soit vide.

7.10.7 Considérons 3 épreuves, chacune ayant la même probabilité de succès. Soit X le nombre total de succès dans ces épreuves. Si $E[X] = 1.8$, quelle est

- la plus grande valeur possible de $P\{X = 3\}$;
- la plus petite valeur possible de $P\{X = 3\}$?

Dans les deux cas, construire un scénario probabiliste qui donne à $P\{X = 3\}$ la valeur fixée.

Pour la partie b), commencer par poser U , une variable uniforme $(0, 1)$ et définir alors les épreuves en termes de valeurs de U .

7.10.8 Considérer n lancers indépendants d'une pièce ayant une probabilité p de tomber sur face. On dit qu'il y a un changement lorsqu'une sortie est différente de celle qui la précède. Par exemple, si $n = 5$ et que les sorties sont *FFPFP*, il y a 3 changements. Trouver l'espérance du nombre de changements.

On pourra exprimer le nombre de changements comme la somme de $n - 1$ variables aléatoires de Bernoulli.

7.10.9 Un groupe de n hommes et m femmes est placé en ligne aléatoirement. Déterminer l'espérance du nombre d'hommes qui ont au moins une femme à leurs côtés.

On pourra définir une variable aléatoire indicatrice pour chaque homme.

7.10.10 Refaire le problème 7.10.9 lorsque le groupe est assis en rond autour d'une table.

7.10.11 Une urne contient m balles noires. A chaque étape, on enlève une balle noire et on met à sa place une nouvelle balle, qui est soit noire avec une probabilité p , soit blanche avec une probabilité $1 - p$. Trouver l'espérance du nombre d'étapes nécessaires pour qu'il n'y ait plus de balle noire dans l'urne.

NOTE. Le problème ci-dessus a une application possible dans la compréhension de la maladie du SIDA. Une partie du système immunitaire du corps consiste en une certaine classe de cellules appelées les T-cellules. Il y a deux types de T-cellules, appelées CD4 et CD8. Bien que le nombre total de T-cellules d'un malade du SIDA (au moins au début de sa maladie) soit le même que celui d'un individu en bonne santé, on a découvert récemment que la proportion des T-cellules CD4 et CD8 est différent. Environ 60% des T-cellules d'une personne en bonne santé sont du type CD4 alors que le pourcentage des T-cellules CD4 décroît continuellement chez les personnes atteintes du SIDA. Un modèle récent propose que le virus HIV (le virus qui cause le SIDA) attaque les cellules CD4 et que le mécanisme du corps qui remplace les T-cellules mortes ne remarque pas si la T-cellule morte était du type CD4 ou CD8. Il produit simplement une nouvelle T-cellule qui est CD4 avec une probabilité .6 et CD8 avec une probabilité .4. Cependant, bien que ce semble être une façon efficace de remplacer les T-cellules mortes lorsque chaque cellule morte a autant de chance d'être n'importe quelle T-cellule du corps (et qui a une probabilité .6 d'être CD4), cela a des conséquences dangereuses lorsqu'on fait face à un virus qui ne vise que les T-cellules CD4.

7.10.12 Cinq urnes contiennent des boules. On tire au hasard une boule de chaque urne. Celles-ci contiennent respectivement 1 boule blanche et 5 noires; 3 blanches et 3 noires; 6 blanches et 4 noires; 2 blanches et 6 noires; 3 blanches et 7 noires. Calculer l'espérance du nombre des boules blanches obtenues.

7.10.13 Soit Z une variable aléatoire normale standard. Pour une valeur x fixée, poser

$$X = \begin{cases} Z & \text{si } Z > x \\ 0 & \text{sinon} \end{cases}$$

Montrer que $E[X] = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

7.10.14 On bat parfaitement un jeu de n cartes, numérotées de 1 à n , afin que toutes les $n!$ configurations possibles soient équiprobables. Supposons que vous deviez hasarder n conjectures successives, où la i -ème conjecture devine la carte en position i . Soit N le nombre de conjectures correctes.

- Sans aucune information sur les conjectures précédentes, montrer que, quelque soit la stratégie adoptée, $E[N] = 1$.
- Si, après chaque conjecture, on vous montre la carte qui se trouve dans la position correspondante, quelle est d'après vous la meilleure stratégie? Montrer que selon cette stratégie

$$E[N] = 1/n + 1/(n-1) + \dots + 1 \\ \approx \int_1^n \frac{1}{x} dx = \ln n$$

- c) Supposons qu'après chaque conjecture on vous dise si vous avez raison ou tort. Dans ce cas, on peut montrer que la stratégie qui maximise $E[N]$ est celle qui consiste à annoncer à chaque fois la même carte jusqu'à ce que l'on vous dise que vous avez raison et à passer ensuite à une nouvelle carte. Selon cette stratégie, montrer que

$$E[N] = 1 + 1/2! + 1/3! + \dots + 1/n! \\ \approx e - 1$$

Pour cela, exprimer N comme une somme de variables aléatoires indicatrices (ou de Bernoulli).

7.10.15 On retourne l'une après l'autre les cartes d'un jeu ordinaire de 52 cartes. Si la première est un as, ou la seconde un deux, ou la troisième un trois, ..., ou la treizième un roi, ou la quatorzième un as, etc., nous disons qu'une rencontre a lieu. Remarquons que nous ne demandons pas que la $(13n + 1)$ -ième carte soit un as particulier pour considérer qu'il y a rencontre, mais seulement que ce soit un as. Calculer l'espérance du nombre de rencontres.

7.10.16 Une certaine région est habitée par r espèces d'insectes. Chaque insecte attrapé sera de l'espèce i avec une probabilité donnée par

$$P_i, i = 1, \dots, r \quad \sum_1^r P_i = 1$$

ceci indépendamment des insectes attrapés antérieurement.

- Calculer le nombre moyen d'insectes qui sont attrapés avant qu'un insecte de l'espèce 1 soit pris;
- calculer le nombre moyen d'espèces représentées parmi les captures jusqu'à ce qu'un insecte d'espèce 1 soit pris.

7.10.17 Une urne contient n boules, la i -ème étant de poids $W(i)$, $i = 1, \dots, n$. Les boules sont prélevées une à une et sans remise, d'une manière telle que la propriété suivante sera vérifiée: à chaque tirage, la probabilité qu'une boule donnée soit choisie sera égale au rapport de son poids et de la somme des poids restant dans l'urne. Supposons par exemple qu'à un moment donné, l'ensemble des boules restant dans l'urne soit i_1, \dots, i_r ; donc le prochain choix sera la boule i_j avec une probabilité

$$W(i_j) / \sum_{k=1}^r W(i_k), \quad j = 1, \dots, r.$$

Calculer l'espérance du nombre des boules qui seront prélevées avant que la boule n° 1 n'apparaisse.

7.10.18 On considère un groupe de 100 personnes. Calculer l'espérance du nombre de jours où 3 personnes exactement auront leur anniversaire. Calculer aussi l'espérance du nombre de jours anniversaire distincts.

7.10.19 Combien de fois vous attendez-vous à jeter un dé équilibré avant que chacune des six faces soit apparue au moins une fois?

7.10.20 Une urne n° 1 contient 5 boules blanches et 6 boules noires, alors que l'urne n° 2 en contient 8 blanches et 10 noires. Deux boules sont choisies au hasard de l'urne 1, puis introduites dans l'urne 2. Si 3 boules sont ensuite prélevées au hasard de l'urne 2, calculer l'espérance du nombre des boules blanches présentes parmi ces 3 boules.

Pour cela poser $X_i = 1$ si la i -ème boule blanche initialement dans l'urne 1 est l'une des trois boules tirées et poser $X_i = 0$ si ce n'est pas le cas. De façon analogue, poser $Y_i = 1$ si la i -ème boule blanche de l'urne 2 est une des 3 boules choisies et $Y_i = 0$ sinon. Le nombre de boules blanches présentes dans le triplet peut alors s'exprimer comme

$$\sum_1^5 X_i + \sum_1^8 Y_i.$$

7.10.21 Une bouteille contient initialement m grandes pilules et n petites pilules. Chaque jour un patient choisit au hasard une des pilules. S'il choisit une petite pilule, il l'avale. S'il en choisit une grande, il la coupe en deux; il remet une part (considérée à présent comme une petite pilule) dans la bouteille et mange l'autre.

a) Soit X le nombre de petites pilules dans la bouteille après que la dernière grande pilule a été choisie et que sa petite moitié a été remplacée. Trouver $E[X]$.

On pourra définir $n + m$ variables indicatrices, une pour chaque petite pilule présente initialement et une pour chacune des m petites pilules créées en coupant une grande. Utiliser ensuite l'argument de l'exemple 7.13.

b) Soit Y le jour où la dernière grande pilule est choisie. Trouver $E[Y]$.

On pourra pour cela chercher la relation entre X et Y .

7.10.22 Soient X_1, X_2, \dots une séquence de variables aléatoires continues indépendantes et identiquement distribuées. Soit $N \geq 2$ tel que

$$X_1 \geq X_2 \geq \dots \geq X_{N-1} < X_N$$

N est le point à partir duquel la séquence cesse de décroître. Montrer que $E[N] = e$.

Trouver d'abord $P\{N \geq n\}$.

7.10.23 Soient X_1, X_2, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées selon une loi uniforme $(0, 1)$.

a) Trouver $E[\max(X_1, \dots, X_n)]$.

b) Trouver $E[\min(X_1, \dots, X_n)]$.

7.10.24 Dans le problème 7.10.4, calculer la variance de la somme des dés.

7.10.25 Dans le problème 7.10.6, calculer la variance du nombre d'urnes vides.

7.10.26 Si $E[X] = 1$ et $\text{Var}(X) = 5$, trouver

- a) $E[(2 + X)^2]$;
 b) $\text{Var}(4 + 3X)$.

7.10.27 Si 10 couples mariés s'installent autour d'une table ronde, ceci au hasard, calculer l'espérance et la variance du nombre de femmes qui seront assises à côté de leur mari.

7.10.28 On retourne l'une après l'autre des cartes provenant d'un jeu ordinaire. Calculer l'espérance du nombre de cartes à retourner avant d'obtenir

- a) 2 as;
 b) 5 piques;
 c) les 13 cœurs.

7.10.29 Soit X le nombre de 1 et Y le nombre de 2 apparaissant lors de n jets d'un dé équilibré. Calculer $\text{Cov}(X, Y)$.

7.10.30 Un dé est jeté 2 fois. Soit X la somme des résultats et soit Y la différence entre le premier et le second résultat. Calculer $\text{Cov}(X, Y)$.

7.10.31 Les variables aléatoires X et Y ont une densité conjointe donnée par

$$f(x, y) = \begin{cases} 2e^{-2x}/x & 0 \leq x < \infty, 0 \leq y \leq x \\ 0 & \text{sinon} \end{cases}$$

Calculer $\text{Cov}(X, Y)$.

7.10.32 Soient X_1, \dots des variables aléatoires indépendantes de moyenne commune μ et de variance commune σ^2 et soit $Y_n = X_n + X_{n+1} + X_{n+2}$. Trouver, pour $j \geq 0$, $\text{Cov}(Y_n, Y_{n+j})$.

7.10.33 La densité conjointe de X et Y est donnée par

$$f(x, y) = \frac{1}{y} e^{-(y+x/y)}, \quad x > 0, y > 0$$

Calculer $E[X]$ et $E[Y]$ et montrer que $\text{Cov}(X, Y) = 1$.

7.10.34 Un étang contient 100 poissons parmi lesquels se trouvent 30 carpes. On capture 20 de ces poissons. Quelle est l'espérance puis la variance du nombre des carpes parmi ces 20 poissons? Quelles hypothèses faites-vous?

7.10.35 Un groupe de 20 personnes, composé de 10 hommes et de 10 femmes, est réparti aléatoirement en 10 couples. Calculer l'espérance et la variance du nombre de couples mixtes. Supposons maintenant que ce groupe de 20 personnes soit en fait composé de 10 couples mariés. Calculer l'espérance et la variance du nombre de couples mariés qui seront réunis par le hasard.

7.10.36 Soit X_1, X_2, \dots, X_n des variables aléatoires indépendantes ayant une distribution continue F inconnue; soit Y_1, Y_2, \dots, Y_m des variables aléatoires indépendantes ayant une distribution continue G inconnue. Ordonnons maintenant ces $n + m$ variables et posons

$$I_i = \begin{cases} 1 & \text{si la } i\text{-ème plus petite des } n + m \text{ variables est de} \\ & \text{l'échantillon des } X \\ 0 & \text{sinon} \end{cases}$$

La variable aléatoire $R = \sum_{i=1}^{n+m} iI_i$ est la somme des rangs de l'échantillon des X ; elle est la base d'une méthode statistique classique (appelée test de la somme des rangs de Wilcoxon) utilisée pour tester si les distributions F et G sont identiques. Ce test accepte l'hypothèse que $F = G$ si R n'est ni trop grand ni trop petit. En supposant que cette égalité est en fait vérifiée, calculer la moyenne et la variance de R (utiliser les résultats de l'exemple 7.22).

7.10.37 Il existe deux procédés différents pour fabriquer une certaine pièce; supposons que la qualité d'une pièce obtenue par le procédé i soit une variable aléatoire continue de distribution F_i , $i = 1, 2$. Supposons encore que n pièces soient issues du procédé 1 et m du procédé 2. Ordonnons les $n + m$ pièces par ordre de qualité et posons

$$X_j = \begin{cases} 1 & \text{si la } j\text{-ème meilleure pièce est obtenue grâce au pro-} \\ & \text{cédé 1} \\ 2 & \text{sinon} \end{cases}$$

Pour le vecteur $X = (X_1, X_2, \dots, X_{n+m})$ composés de n «1» et de m «2», soit R le nombre de chaînes de «1». Par exemple, si $n = 5$, $m = 2$ et $X = (1, 2, 1, 1, 1, 1, 2)$, alors $R = 2$. Si $F_1 = F_2$ (c'est-à-dire si les deux procédés produisent des articles dont la qualité varie selon la même répartition), quelles sont l'espérance et la variance de R ?

7.10.38 Si X_1, X_2, X_3, X_4 sont des variables aléatoires deux à deux non corrélées, chacune d'espérance nulle et de variance 1, calculer les corrélations de

- $X_1 + X_2$ et $X_2 + X_3$;
- $X_1 + X_2$ et $X_3 + X_4$.

7.10.39 Considérons le jeu de dés suivant, pratiqué dans certains casinos: deux joueurs, 1 et 2, jettent chacun à leur tour une paire de dés. Puis, la banque jette à son tour les dés, après quoi on détermine qui a gagné selon la règle suivante: le joueur i , $i = 1, 2$ gagne si la somme de ses deux dés donne un résultat strictement plus grand que celui de la banque. Posons pour $i = 1, 2$

$$I_i = \begin{cases} 1 & \text{si } i \text{ gagne} \\ 0 & \text{sinon} \end{cases}$$

et montrer que I_1 et I_2 sont positivement corrélées. Expliquer pourquoi on pouvait s'attendre à un tel résultat.

7.10.40 On jette plusieurs fois un dé équilibré. Soient X et Y le nombre de jets nécessaires pour obtenir un 6 et un 5, respectivement. Trouver a) $E[X]$, b) $E[X|Y = 1]$ et c) $E[X|Y = 5]$.

7.10.41 Une urne contient 4 boules blanches et 6 boules noires. On en tire successivement deux échantillons aléatoires de taille 3 et 5 respectivement, ceci sans remise. Soient X et Y le nombre de boules blanches dans chacun de ces échantillons; calculer $E[X|Y = i]$ pour $i = 1, 2, 3, 4$.

7.10.42 La densité conjointe de X et Y est donnée par

$$f(x, y) = \frac{e^{-x/y} e^{-y}}{y} \quad 0 < x < \infty; 0 < y < \infty$$

Calculer $E[X^2|Y = y]$.

7.10.43 La densité conjointe de X et Y est donnée par

$$f(x, y) = \frac{e^{-y}}{y} \quad 0 < x < y, 0 < y < \infty$$

Calculer $E[X^3|Y = y]$.

7.10.44 Un prisonnier est enfermé dans une cellule contenant 3 portes. La première ouvre un tunnel qui revient dans la cellule après une marche de 2 jours. La seconde porte donne sur un tunnel qui revient aussi à la cellule au bout d'un voyage de 4 jours. La troisième porte conduit à la liberté au bout d'un jour de marche. On suppose que le prisonnier choisit à chaque tentative les portes 1, 2 et 3 avec des probabilités respectives de 0,5, 0,3 et 0,2. Quelle est l'espérance du nombre de jours qu'il faudra au prisonnier pour retrouver sa liberté?

7.10.45 Dix chasseurs attendent que des canards s'envolent. Lorsqu'un vol de canards apparaît, les chasseurs tirent tous en même temps, chacun choisissant sa cible de façon aléatoire et indépendamment des autres. Si chaque chasseur atteint sa cible indépendamment de la réussite des autres avec une probabilité de 0,6, calculer le nombre moyen de canards qui seront touchés. On suppose que le nombre de canards dans un vol est une variable aléatoire de Poisson de paramètre 6.

7.10.46 Le nombre de personnes qui entrent dans un ascenseur au rez-de-chaussée est une variable aléatoire de Poisson d'espérance 10. Supposons qu'il y ait N étages, que la probabilité de sortir à tout étage est la même pour toutes les personnes et que de plus chaque personne agit indépendamment des autres. Calculer le nombre d'arrêts moyen que devra faire l'ascenseur pour débarquer tous les passagers.

7.10.47 On admet que le nombre moyen d'accidents dans une installation industrielle est de 5 par semaine. Supposons aussi que les nombres de travailleurs blessés dans chaque accident sont des variables aléatoires indépendantes ayant une même espérance égale à 2,5. Si le nombre de travailleurs blessés dans chaque accident est indépendant du nombre d'accidents qui se produisent, calculer le nombre moyen de travailleurs blessés au cours d'une semaine.

7.10.48 On jette continuellement une pièce ayant pour probabilité p de tomber sur face jusqu'à ce que pile et face soient apparues. Trouver

- l'espérance du nombre de jets;
- la probabilité que le dernier jet donne face.

7.10.49 Une personne jette continuellement une pièce jusqu'à ce que 3 faces consécutives apparaissent. En admettant que chaque jet donne indépendamment face avec une probabilité p , déterminer l'espérance du nombre de jets requis.

Pour cela, soit P le premier jet donnant pile, et soit $P = 0$ si tous les jets donnent face. Conditionner ensuite sur P .

7.10.50 Dans un jeu à $n + 1$ participants, chaque personne est, indépendamment, un gagnant avec une probabilité p . Les gagnants partagent un prix d'une unité (par exemple, si 4 personnes gagnent, chacune d'elles reçoit $1/4$ et s'il n'y a pas de gagnant, aucun des participants ne reçoit quelque chose.) Soit A un des joueurs et soit X la somme reçue par A .

- Calculer l'espérance du prix total partagé par les joueurs.
- Montrer que

$$E[X] = \frac{1 - (1 - p)^{n+1}}{n + 1}$$

- Calculer $E[X]$ en conditionnant sur le fait que A gagne ou non et conclure que

$$E[(1 + B)^{-1}] = \frac{1 - (1 - p)^{n+1}}{(n + 1)p}$$

lorsque B est une variable aléatoire binomiale de paramètres n et p .

7.10.51 Chacun des $m + 2$ joueurs paye 1 unité dans une cagnotte pour jouer au jeu suivant. Une pièce équilibrée est lancée successivement n fois, où n est un nombre impair, et les sorties successives sont notées. Chaque joueur écrit, avant les jets, une prédiction de sortie. Par exemple, si $n = 3$, un joueur peut écrire (F, F, P) , ce qui signifie qu'il prédit un face sur le premier jet, un face sur le deuxième et un pile sur le troisième. Après les jets de la pièce, les joueurs comptent leur nombre de prédictions correctes. Ainsi, si toutes les sorties montrent face, le joueur ayant écrit (F, F, P) a 2 prédictions correctes. La cagnotte de $m + 2$ unités est alors répartie équitablement entre les joueurs ayant le plus grand nombre de prédictions correctes.

Puisque chaque jet a autant de chance de donner pile ou face, m joueurs ont décidé de faire leurs prédictions d'une manière totalement aléatoire. En fait, ils lancent chacun leur propre pièce n fois et utilisent les résultats comme prédictions. Cependant, les 2 derniers joueurs ont formé un groupe et utilisent la stratégie suivante. L'un d'eux fait des prédictions de la même façon que les m autres joueurs, mais le second prédit exactement le contraire du premier. En fait, lorsque le membre du groupe prédisant au hasard prédit un F , l'autre membre prédit un P . Par exemple, si celui prédisant au hasard prédit (F, F, P) , l'autre prédit (P, P, F) .

- Argumenter le fait qu'exactement un des deux membres du groupe fait plus de $n/2$ prédictions correctes. (Rappelez-vous que n est impair.)

b) Soit X le nombre de joueurs parmi les m ne faisant pas partie du groupe qui ont plus de $n/2$ prédictions correctes. Quelle est la distribution de X ?

c) Avec X défini en b), argumenter le fait que

$$E[\text{le groupe a une récompense}] = (m+2)E\left[\frac{1}{X+1}\right]$$

d) Utiliser la partie c) du problème 7.10.50 pour conclure que

$$E[\text{le groupe a une récompense}] = \frac{2(m+2)}{m+1} \left[1 - \left(\frac{1}{2}\right)^{m+1} \right]$$

et le calculer explicitement pour $m = 1, 2$ et 3 .

Comme on peut montrer que $\frac{2(m+2)}{m+1} \left[1 - \left(\frac{1}{2}\right)^{m+1} \right] > 2$, il s'ensuit que la stratégie

du groupe donne toujours une espérance de profit positive.

7.10.52 Soit U_1, U_2, \dots une séquence de variables aléatoires uniformes $(0, 1)$ indépendantes. Dans l'exemple 7.32, on a montré que pour $0 \leq x \leq 1$, $E[N(x)] = e^x$ où

$$N(x) = \text{Minimum} \left\{ n : \sum_{i=1}^n U_i > x \right\}$$

Ce problème propose une autre approche pour établir ce résultat.

a) Montrer par induction sur n que pour $0 < x \leq 1$ et pour tout $n \geq 0$,

$$P\{N(x) \geq n+1\} = \frac{x^n}{n!}$$

Pour cela, conditionner d'abord sur U_1 puis utiliser l'hypothèse d'induction.

b) Utiliser la question a) pour conclure que

$$E[N(x)] = e^x$$

7.10.53 Une urne contient 30 balles dont 10 rouges et 8 bleues. On tire aléatoirement 12 balles de cette urne. Soit X le nombre de balles rouges tirées et Y le nombre de balles bleues tirées. Trouver $\text{Cov}(X, Y)$

a) en définissant des variables aléatoires indicatrices (c'est-à-dire de Bernoulli) appropriées X_i, Y_j telles que $X = \sum_{i=1}^{10} X_i$, $Y = \sum_{j=1}^8 Y_j$;

b) en conditionnant (soit sur X , soit sur Y) pour déterminer $E[XY]$.

7.10.54 Des ampoules de type i fonctionnent pendant une durée aléatoire de moyenne μ_i et d'écart-type σ_i , $i = 1, 2$. Une ampoule choisie au hasard dans une boîte d'ampoules est de type 1 avec une probabilité p et de type 2 avec une probabilité $1 - p$. Soit X la durée de vie de cette ampoule. Trouver

- a) $E[X]$;
b) $\text{Var}(X)$.

7.10.55 Dans l'exemple 7.28, calculer la variance du temps qu'il faut au mineur pour retrouver la sortie.

7.10.56 Les règles du jeu de dés nommé craps ont été définies au problème 2.9.14. Calculer l'espérance et la variance du nombre de jets de dés qu'il faut pour terminer une partie de craps.

7.10.57 Considérons un joueur qui, à chaque partie, gagne ou perd son pari avec des probabilités p et $(1 - p)$. Lorsque $p > \frac{1}{2}$ la stratégie suivante, connue sous le nom de stratégie de Kelly, est souvent utilisée. Elle consiste à toujours parier la fraction $2p - 1$ de sa fortune restante. Calculer l'espérance de la fortune au bout de n parties d'un joueur parti avec x unités et qui fait usage de la stratégie de Kelly.

7.10.58 Le nombre d'accidents touchant un individu lors d'une année donnée est une variable aléatoire de Poisson d'espérance λ . Supposons que cette espérance varie en fonction des personnes, valant ainsi 2 pour 60% de la population et 3 pour les 40% restants. On choisit une personne au hasard. Quelle est la probabilité qu'au cours d'une année, elle n'ait aucun accident? Qu'elle en ait 3? Quelle est la probabilité conditionnelle qu'elle ait 3 accidents dans l'année sachant qu'elle n'a pas eu d'accidents l'année précédente?

7.10.59 Refaire le problème 7.10.59 lorsque la proportion de la population caractérisée par un paramètre λ inférieur à x est égale à $1 - e^{-x}$.

7.10.60 Considérons une urne contenant un nombre élevé de pièces de monnaie et supposons que chacune des pièces ait une probabilité p de montrer face lorsqu'elle est jetée. Cependant, cette valeur de p varie pour chaque pièce. Supposons que la composition de l'urne est telle que si une pièce est choisie au hasard, alors le paramètre p qui la caractérise peut être vu comme une variable aléatoire uniformément distribuée sur $[0, 1]$. Si une pièce est choisie au hasard et jetée deux fois, calculer la probabilité que le premier jet donne face, puis que les deux jets donnent face.

7.10.61 Dans le problème 7.10.60, supposons que la pièce est lancée n fois. Soit X le nombre de faces obtenues. Montrer que

$$P\{X = i\} = \frac{1}{n+1}, \quad i = 0, 1, \dots, n$$

Utiliser le fait que

$$\int_0^1 x^{a-1} (1-x)^{b-1} dx = \frac{(a-1)!(b-1)!}{(a+b-1)!}$$

où a et b sont des entiers positifs.

7.10.62 Supposons que dans le problème 7.10.60, on continue à lancer la pièce jusqu'à ce que face apparaisse. Soit N le nombre de jets nécessaires. Trouver

- $P\{N \geq i\}, i \geq 0;$
- $P\{N = i\};$
- $E[N].$

7.10.63 Dans l'exemple 7.38, appelons S le signal émis et R le signal reçu.

- Calculer $E[R].$
- Calculer $\text{Var}(R).$
- R est-elle normalement distribuée?
- Calculer $\text{Cov}(R, S).$

7.10.64 Dans l'exemple 7.39, supposons que X est uniformément distribuée sur $(0, 1)$. Si les régions discrètes sont déterminées par $a_0 = 0, a_1 = \frac{1}{2}$ et $a_2 = 1$, déterminer le quantificateur optimal Y et calculer $E[(X - Y)^2].$

7.10.65 La fonction génératrice des moments de X est donnée par $M_X(t) = \exp\{2e^t - 2\}$ et celle de Y par $M_Y(t) = (\frac{1}{4})^{10} (3e^t + 1)^{10}$. Si X et Y sont indépendantes, que valent a) $P\{X + Y = 2\}$, b) $P\{XY = 0\}$ et c) $E[XY]$?

7.10.66 On jette deux dés. Soit X la valeur du premier dé et Y la somme des deux valeurs. Calculer la fonction génératrice des moments conjoints de X et Y .

7.10.67 La densité conjointe de X et Y est donnée par

$$f(x, y) = \frac{e^{-x} e^{-y/x}}{x} \quad 0 < x < \infty, 0 < y < \infty$$

- Calculer la fonction génératrice des moments conjoints de X et Y ;
- calculer les fonctions génératrices des moments individuels de X et Y .

7.10.68 Deux enveloppes, contenant chacune un chèque, sont placées devant vous. Vous devez choisir une enveloppe, l'ouvrir et regarder le montant du chèque. A ce moment-là, vous pouvez soit accepter le montant soit l'échanger contre le chèque de l'enveloppe fermée. Que devez-vous faire? Est-il possible de trouver une stratégie meilleure que d'accepter la première enveloppe?

Soient A et $B, A < B$, les montants (inconnus) des chèques, et remarquer que la stratégie qui sélectionne une enveloppe au hasard et accepte toujours son chèque a une espérance de $(A + B) / 2$. Considérer la stratégie suivante: soit $F(\cdot)$ une fonction de distribution strictement croissante (donc continue). Choisir une enveloppe au hasard et l'ouvrir. Si le chèque découvert a pour valeur x , alors accepter le avec probabilité $F(x)$ et changer le avec probabilité $1 - F(x)$.

- Montrer que si vous employez cette stratégie, l'espérance de gain est supérieure à $(A + B) / 2$.

On pourra conditionner sur le fait que la première enveloppe contienne soit A soit B . Considérer la stratégie qui fixe une valeur x puis accepte le premier chèque si sa valeur est plus grande que x et l'échange sinon.

- b) Montrer que, pour tout x , l'espérance du gain sous la x -stratégie est toujours au moins $(A + B) / 2$ et qu'elle est strictement plus grande que $(A + B) / 2$ si x se trouve entre A et B .
- c) Soit X une variable aléatoire continue sur toute la ligne et considérer la stratégie suivante: générer la valeur de X et si $X = x$, employer la x -stratégie de la question b). Montrer que l'espérance du gain sous cette stratégie est plus grande que $(A + B) / 2$.

CHAPITRE 8

Théorèmes limites

8.1 INTRODUCTION

Les théorèmes limites constituent les résultats théoriques les plus importants des probabilités. Parmi eux, les principaux sont répertoriés sous deux dénominations: *lois des grands nombres* d'une part, et *théorèmes centraux limites* d'autre part. On s'accorde généralement à les considérer comme des lois des grands nombres s'ils énoncent des conditions sous lesquelles la moyenne d'une suite de variables aléatoires converge (dans un sens à définir) vers leur espérance commune. Les théorèmes centraux limites par contre déterminent sous quelles hypothèses la somme d'un grand nombre de variables aléatoires est de distribution approximativement normale.

8.2 LOI FAIBLE DES GRANDS NOMBRES

8.2.1 Inégalité de Tchebychev

Nous allons commencer par établir l'inégalité appelée inégalité de Markov.

Théorème 8.1 *Inégalité de Markov*

Soit X une variable aléatoire à valeurs non négatives. Pour tout réel $a > 0$

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

DÉMONSTRATION. Pour $a > 0$, soit

$$I = \begin{cases} 1 & \text{si } X \geq a \\ 0 & \text{sinon} \end{cases}$$

et noter que, puisque $X \geq 0$,

$$I \leq \frac{X}{a}$$

En prenant les espérances de l'expression ci-dessus, on obtient

$$E[I] \leq \frac{E[X]}{a}$$

ce qui, comme $E[I] = P\{X \geq a\}$, prouve le résultat. ■

Le théorème suivant est alors un corollaire de l'inégalité de Markov.

Théorème 8.2 *Inégalité de Tchebychev*

Soit X une variable aléatoire d'espérance μ et de variance σ^2 finies. Pour tout réel $k > 0$

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

DÉMONSTRATION. On peut appliquer l'inégalité de Markov, avec $a = k^2$, à la variable $(X - \mu)^2$ puisque celle-ci est à valeurs non négatives. On obtient

$$P\{(X - \mu)^2 \geq k^2\} \leq \frac{E[(X - \mu)^2]}{k^2} \quad (8.1)$$

Mais comme $(X - \mu)^2 \geq k^2$ équivaut à $|X - \mu| \geq k$, (8.1) peut être réécrite

$$P\{|X - \mu| \geq k\} \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

ce qui achève la démonstration. ■

L'importance des inégalités de Markov et de Tchebychev réside en ce qu'elles permettent de borner la valeur de certaines probabilités là où seule l'espérance de la distribution est connue, plus éventuellement sa variance. Il est évident que, si la distribution elle-même est connue, on ne recourra pas à des bornes, puisque la valeur exacte de ces probabilités est calculable.

8.2.2 Exemples d'utilisation des inégalités de Markov et Tchebychev

Exemple 8.1 On suppose que le nombre de pièces sortant d'une usine donnée en l'espace d'une semaine est une variable aléatoire d'espérance 50.

- a) Peut-on estimer la probabilité que la production de la semaine prochaine dépasse 75 pièces?
- b) On sait de plus que la variance de la production hebdomadaire est de 25. Peut-on estimer la probabilité que la production de la semaine à venir soit comprise entre 40 et 60?

SOLUTION. Désignons par X le nombre de pièces produites en une semaine.

a) L'inégalité de Markov donne

$$P\{X > 75\} \leq \frac{E[X]}{75} = \frac{50}{75} = \frac{2}{3}$$

b) L'inégalité de Tchebychev donne

$$P\{|X - 50| \geq 10\} \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$

et donc

$$P\{|X - 50| < 10\} \geq 1 - \frac{1}{4} = \frac{3}{4}$$

La probabilité que la production de la semaine à venir se situe entre 40 et 60 pièces est donc au moins 0,75. ■

L'inégalité de Tchebychev étant valable pour n'importe quelle distribution de la variable X , il ne faut pas s'attendre à ce que la borne qu'elle fournit soit très proche de la probabilité exacte dans la majorité des cas. C'est ce que montre l'exemple suivant.

Exemple 8.2 Soit X une variable uniforme sur l'intervalle $(0, 10)$. On sait qu'alors $E[X] = 5$ et $\text{Var}(X) = \frac{25}{3}$, ce qui donne dans l'inégalité de Tchebychev

$$P\{|X - 5| > 4\} \leq \frac{25}{3(16)} \approx .52$$

alors que le résultat exact est

$$P\{|X - 5| > 4\} = .20$$

On voit bien que si l'inégalité de Tchebychev est fondée, la borne qu'elle fournit est ici loin d'être proche de la probabilité exacte.

A titre d'illustration supplémentaire, la borne calculée par l'inégalité de Tchebychev pour une variable X normale d'espérance μ et de variance σ^2 est, lorsque $k = 2\sigma$,

$$P\{|X - \mu| > 2\sigma\} \leq \frac{1}{4}$$

alors que la probabilité exacte est, elle,

$$P\{|X - \mu| > 2\sigma\} = P\left\{\left|\frac{X - \mu}{\sigma}\right| > 2\right\} = 2[1 - \Phi(2)] \approx .0456 \quad \blacksquare$$

On utilise souvent l'inégalité de Tchebychev comme outil théorique de démonstration, comme on le verra dans la preuve de la loi faible des grands nombres. Le théorème suivant permet déjà de s'en rendre compte.

Théorème 8.3

Soit X une variable aléatoire de variance nulle. Alors X est égale à son espérance avec probabilité 1.

$$P\{X = E\{X\}\} = 1$$

DÉMONSTRATION. En vertu de l'inégalité de Tchebychev, on peut écrire pour $n \geq 1$

$$P\left\{|X - \mu| > \frac{1}{n}\right\} = 0$$

Faisons tendre n vers l'infini. La propriété de continuité des probabilités donne alors

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} P\left\{|X - \mu| > \frac{1}{n}\right\} = P\left\{\lim_{n \rightarrow \infty} \left\{|X - \mu| > \frac{1}{n}\right\}\right\} \\ &= P\{X \neq \mu\} \end{aligned}$$

ce qui établit le résultat voulu. ■

8.2.3 Enoncé de la loi faible des grands nombres

Théorème 8.4 Loi faible des grands nombres

Soient X_1, X_2, \dots une suite de variables aléatoires indépendantes et identiquement distribuées. On suppose que toutes admettent la même espérance finie $E\{X_i\} = \mu$. Pour tout $\varepsilon > 0$

$$P\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > \varepsilon\right\} \rightarrow 0 \quad \text{lorsque } n \rightarrow \infty$$

DÉMONSTRATION. Nous démontrons le théorème en admettant une hypothèse supplémentaire, à savoir que les variables considérées ont une variance σ^2 commune et finie. Dans ce cas, comme

$$E\left[\frac{X_1 + \dots + X_n}{n}\right] = \mu \quad \text{et} \quad \text{Var}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

il résulte de l'inégalité de Tchebychev que

$$P\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| > \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}$$

ce qui établit le résultat. ■

La loi faible des grands nombres fut établie pour la première fois par Jacob Bernoulli pour le cas particulier où les X_i ne prennent pour valeur que 0 ou 1 (et sont donc des variables de Bernoulli). Son énoncé de ce théorème et la démonstration qu'il en donne figurent dans son ouvrage *Ars Conjectandi*, publié en 1713 par son neveu Nicolas Bernoulli, huit ans après sa mort. Il faut savoir que l'inégalité de Tchebychev n'étant pas connue à l'époque, Bernoulli dut développer une démonstration extrêmement ingénieuse pour établir le résultat. La version générale de la loi faible des grands nombres, telle que la présente le théorème 8.4, est attribuée au mathématicien russe Khintchine.

8.3 THÉORÈME CENTRAL LIMITE

8.3.1. Version restreinte

Le théorème central limite est l'un des plus remarquables résultats de la théorie des probabilités. En gros, il établit que la somme d'un grand nombre de variables aléatoires indépendantes suit une distribution approximativement normale. Il fournit donc non seulement une méthode simple pour le calcul approximatif de probabilités liées à des sommes de variables aléatoires, mais il explique également ce fait empirique remarquable que bien des phénomènes naturels admettent une distribution en forme de cloche, c'est-à-dire de type normal.

La clé de la démonstration du théorème central limite est le théorème suivant, donné sans démonstration.

Théorème 8.5

Soient Z_1, Z_2, \dots une suite de variables aléatoires dont les fonctions de répartition sont notées F_{Z_n} et les fonctions génératrices des moments M_{Z_n} , $n \geq 1$; soit aussi une variable aléatoire Z de fonction de répartition F_Z et de fonction génératrice des moments M_Z . Si $M_{Z_n}(t) \rightarrow M_Z(t)$ pour tout t , alors $F_{Z_n}(t) \rightarrow F_Z(t)$ pour toutes les valeurs de t pour lesquelles $F_Z(t)$ est continue.

Dans le cas particulier d'une variable Z normale standard, on sait que $M_Z(t) = e^{t^2/2}$. Du théorème 8.5, il résulte que si $M_{Z_n}(t) \rightarrow e^{t^2/2}$ lorsque $n \rightarrow \infty$, alors $F_{Z_n}(t) \rightarrow \Phi(t)$ lorsque $n \rightarrow \infty$.

Nous allons maintenant donner la version la plus simple du théorème central limite.

Théorème 8.6 *Théorème central limite*

Soient X_1, X_2, \dots une suite de variables aléatoires indépendantes et identiquement distribuées, d'espérance μ , de variance σ^2 . Alors la distribution de

$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

tend vers la distribution normale lorsque $n \rightarrow \infty$, ce qui veut dire que

$$P \left\{ \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq a \right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-x^2/2} dx \quad \text{quand } n \rightarrow \infty$$

DÉMONSTRATION. Admettons pour l'instant que $\mu = 0$ et $\sigma^2 = 1$. Nous allons démontrer le théorème en faisant l'hypothèse que la fonction génératrice des moments des X_i , notée $M(t)$, existe et est finie. La fonction génératrice des moments de X_i/\sqrt{n} sera

$$E \left[\exp \left\{ \frac{tX_i}{\sqrt{n}} \right\} \right] = M \left(\frac{t}{\sqrt{n}} \right)$$

et par conséquent celle de $\sum_{i=1}^n X_i/\sqrt{n}$ sera $\left[M \left(\frac{t}{\sqrt{n}} \right) \right]^n$.

Posons

$$L(t) = \ln M(t)$$

et remarquons que

$$L(0) = 0$$

$$L'(0) = \frac{M'(0)}{M(0)}$$

$$= \mu$$

$$= 0$$

$$L''(0) = \frac{M(0)M''(0) - [M'(0)]^2}{[M(0)]^2}$$

$$= E[X^2]$$

$$= 1$$

Or, pour démontrer le théorème, il nous faut établir que $[M(t/\sqrt{n})]^n \rightarrow e^{t^2/2}$ lorsque $n \rightarrow \infty$, ou, ce qui est équivalent, que $nL(t/\sqrt{n}) \rightarrow t^2/2$ lorsque $n \rightarrow \infty$. On peut écrire

$$\lim_{n \rightarrow \infty} \frac{L(t/\sqrt{n})}{n^{-1}} = \lim_{n \rightarrow \infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}} \quad \text{en vertu de la règle de l'Hôpital}$$

$$= \lim_{n \rightarrow \infty} \left[\frac{L'(t/\sqrt{n})t}{2n^{-1/2}} \right]$$

$$= \lim_{n \rightarrow \infty} \left[\frac{-L''(t/\sqrt{n})n^{-3/2}t^2}{-2n^{-3/2}} \right] \quad \text{en vertu de la même règle}$$

$$\begin{aligned}
 &= \lim_{n \rightarrow \infty} \left[L'' \left(\frac{t}{\sqrt{n}} \right) \frac{t^2}{2} \right] \\
 &= \frac{t^2}{2}
 \end{aligned}$$

Le théorème central limite est donc démontré pour le cas où $\mu = 0$ et $\sigma^2 = 1$. Dans le cas général, on considère les variables standardisées $X_i^* = (X_i - \mu)/\sigma$ auxquelles s'applique la démonstration ci-dessus, puisque $E[X_i^*] = 0$ et $\text{Var}(X_i^*) = 1$; le résultat est ainsi établi en toutes circonstances. ■

A titre de remarque, on peut ajouter la considération suivante: bien que le théorème 8.6 affirme que pour chaque a

$$P \left\{ \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq a \right\} \rightarrow \Phi(a)$$

on peut en fait démontrer que cette convergence est uniforme en a (on dit que $f_n(a) \rightarrow f(a)$ uniformément en a si, pour tout $\varepsilon > 0$, il existe un N tel que $|f_n(a) - f(a)| < \varepsilon$, pour tout a , dès que $n \geq N$).

La première version du théorème central limite fut établie par De Moivre aux alentours de 1733 pour le cas particulier des variables de Bernoulli de paramètre $p = 1/2$. Laplace en donna une extension à des variables de Bernoulli quelconques (une variable binomiale pouvant être comprise comme une somme de n variables de Bernoulli indépendantes et de même paramètre, cette démonstration de Laplace justifie l'approximation normale donnée aux variables binomiales, par exemple dans la section 5.4.7). Laplace énonça aussi la version plus générale donnée ici au théorème 8.6. Sa démonstration n'était cependant pas totalement rigoureuse, et il n'est d'ailleurs pas aisé de la compléter. Ce fut le mathématicien russe Liapounoff qui, le premier, donna une démonstration absolument rigoureuse du théorème central limite, et ce entre 1901 et 1902.

8.3.2 Applications du théorème central limite

Exemple 8.3 Un astronome souhaite mesurer la distance, en années-lumière, entre son observatoire et une étoile lointaine. Bien qu'il connaisse une technique de mesure, il sait aussi que chaque résultat ne constitue qu'une distance approchée, en raison des influences atmosphériques et d'autres causes d'erreur inévitables. Par conséquent, notre astronome prévoit de prendre plusieurs mesures et d'accepter leur moyenne comme estimation de la distance réelle. Il a des raisons de penser que les différentes valeurs mesurées sont des variables aléatoires indépendantes et identiquement distribuées d'espérance commune d (la vraie valeur) et de variance commune 4 (l'unité étant toujours l'année-lumière). Combien de mesures doit-il réaliser pour être raisonnablement sûr que l'erreur soit inférieure à une demi-année-lumière?

SOLUTION. Admettons que l'astronome fasse n observations que l'on désignera par X_1, X_2, \dots, X_n . Le théorème central limite établit que

$$Z_n = \frac{\sum_{i=1}^n X_i - nd}{2\sqrt{n}}$$

suit approximativement une distribution normale. Par conséquent

$$\begin{aligned} P\left\{-.5 \leq \frac{\sum_{i=1}^n X_i}{n} - d \leq .5\right\} &= P\left\{-.5 \frac{\sqrt{n}}{2} \leq Z_n \leq .5 \frac{\sqrt{n}}{2}\right\} \\ &\approx \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 \end{aligned}$$

Si donc notre astronome souhaite que la probabilité que l'erreur soit de moins d'une demi-année-lumière reste au-delà de 95 chances sur 100, il lui faudra prendre n^* mesures où n^* vérifie

$$2\Phi\left(\frac{\sqrt{n^*}}{4}\right) - 1 = .95 \quad \text{ou} \quad \Phi\left(\frac{\sqrt{n^*}}{4}\right) = .975$$

ce qui, d'après le tableau 5.4 du chapitre 5, équivaut à

$$\frac{\sqrt{n^*}}{4} = 1.96 \quad \text{ou} \quad n^* = (7.84)^2 = 61.47$$

Comme n^* est nécessairement entier, il faudra faire 62 mesures.

On notera cependant que cette évaluation repose sur l'hypothèse d'une qualité suffisante de l'approximation de Z_n par une variable normale lorsque $n = 62$. Bien qu'en général ce soit le cas pour une telle valeur de n , le moment à partir duquel n est assez grand pour qu'on puisse qualifier l'approximation de bonne dépend de la distribution des X_i . Si notre astronome craint que cette distribution ne soit défavorable et s'il ne souhaite pas prendre de risques, il peut avoir recours à l'inégalité de Tchebychev. On a

$$E\left[\sum_{i=1}^n \frac{X_i}{n}\right] = d \quad \text{Var}\left(\sum_{i=1}^n \frac{X_i}{n}\right) = \frac{4}{n}$$

et l'inégalité donne alors

$$P\left\{\left|\sum_{i=1}^n \frac{X_i}{n} - d\right| > .5\right\} \leq \frac{4}{n(.5)^2} = \frac{16}{n}$$

L'astronome devra donc faire $n = \frac{16}{0.05} = 320$ observations pour avoir 95 chances sur 100 de maintenir l'erreur en dessous d'une demi-année lumière. ■

Exemple 8.4 Le nombre d'inscriptions à un cours de psychologie est une variable aléatoire de Poisson d'espérance 100. Le professeur donnant ce cours a décidé que, si le nombre des inscriptions est au-delà de 120, il créera deux sections et donnera donc deux cours, tandis qu'en deçà une seule classe sera formée. Quelle est la probabilité que ce professeur ait à donner deux fois le cours?

SOLUTION. La solution exacte est $e^{-100} \sum_{i=120}^{\infty} (100)^i / i!$ et son évaluation numérique n'est pas aisée. On se souvient cependant qu'une variable poissonnienne de paramètre 100 peut être considérée comme la somme de 100 variables poissonniennes indépendantes de paramètre 1 chacune, ce qui donne l'occasion d'utiliser le théorème central limite pour obtenir une approximation de la réponse exacte. Soit X le nombre d'inscriptions. On a

$$\begin{aligned} P\{X \geq 120\} &= P\left\{\frac{X - 100}{\sqrt{100}} \geq \frac{120 - 100}{\sqrt{100}}\right\} \\ &\approx 1 - \Phi(2) \\ &\approx .0228 \end{aligned}$$

où l'on a utilisé le fait qu'espérance et variance des variables poissonniennes sont toujours égales. ■

Exemple 8.5 On lance 10 dés équilibrés. On cherche la probabilité que la somme des dix résultats soit comprise entre 30 et 40.

SOLUTION. On note X_i le résultat montré par le i -ième dé, $i = 1, 2, \dots, 10$. Comme $E[X_i] = \frac{7}{2}$ et $\text{Var}(X_i) = E[X_i^2] - (E[X_i])^2 = \frac{35}{12}$, on obtient par application du théorème central limite

$$\begin{aligned} P\left\{30 \leq \sum_{i=1}^{10} X_i \leq 40\right\} &= P\left\{\frac{30 - 35}{\sqrt{\frac{350}{12}}} \leq \frac{\sum_{i=1}^{10} X_i - 35}{\sqrt{\frac{350}{12}}} \leq \frac{40 - 35}{\sqrt{\frac{350}{12}}}\right\} \\ &\approx 2\Phi(\sqrt{6/7}) - 1 \\ &\approx .65 \end{aligned}$$

Exemple 8.6 Soient X_i , $i = 1, \dots, 10$ des variables aléatoires uniformes sur l'intervalle $(0, 1)$. On cherche à évaluer approximativement $P\left\{\sum_{i=1}^{10} X_i > 6\right\}$.

SOLUTION. Comme $E[X_i] = \frac{1}{2}$ et $\text{Var}(X_i) = \frac{1}{12}$, le théorème central limite livre

$$\begin{aligned} P\left\{\sum_{i=1}^{10} X_i > 6\right\} &= P\left\{\frac{\sum_{i=1}^{10} X_i - 5}{\sqrt{10(\frac{1}{12})}} > \frac{6 - 5}{\sqrt{10(\frac{1}{12})}}\right\} \\ &\approx 1 - \Phi(\sqrt{1.2}) \\ &\approx .16 \end{aligned}$$

ce qui signifie que 16 fois sur 100 seulement, en moyenne, la somme $\sum_{i=1}^{10} X_i$ sera supérieure à 6. ■

Dans l'exemple 8.7 nous allons utiliser le théorème central limite pour obtenir l'approximation de Stirling de $n!$.

Exemple 8.7 Approximation de Stirling de $n!$

Soient X_1, X_2, \dots des variables aléatoires de Poisson indépendantes de moyenne (et de variance) égale à 1. Soit $S_n = \sum_{i=1}^n X_i$ et noter que S_n a une moyenne et une variance égales à n .

Soit Z une variable aléatoire normale standard. On a

$$\begin{aligned}
 P\{S_n = n\} &= P\{n-1 < S_n \leq n\} && \text{puisque } S_n \text{ est un entier} \\
 &= P\{-1 < S_n - n \leq 0\} \\
 &= P\left\{\frac{-1}{\sqrt{n}} < \frac{S_n - n}{\sqrt{n}} \leq 0\right\} \\
 &\approx P\left\{\frac{-1}{\sqrt{n}} < Z \leq 0\right\} && \text{lorsque } n \text{ est grand,} \\
 &&& \text{par le théorème central limite} \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-1/\sqrt{n}}^0 e^{-x^2/2} dx \\
 &\approx \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n}} && \text{par le théorème fondamental} \\
 &&& \text{du calcul intégral}
 \end{aligned}$$

Mais la somme de variables aléatoires de Poisson indépendantes est aussi une variable aléatoire de Poisson, donc S_n est Poisson de moyenne n . Donc

$$P\{S_n = n\} = e^{-n} \frac{n^n}{n!}$$

Donc pour n grand

$$\frac{e^{-n} n^n}{n!} \approx \frac{1}{\sqrt{2\pi n}}$$

ou, de manière équivalente,

$$n! \approx n^{n+1/2} e^{-n} \sqrt{2\pi}$$

ce qui est l'approximation de Stirling de $n!$. ■

8.3.3 Versions plus générales du théorème central limite

Il a été possible de démontrer des versions du théorème central limite où les X_i sont encore indépendantes mais plus nécessairement identiquement distribuées. L'une de ces versions suit ici, qui n'est d'ailleurs pas la plus générale.

Théorème 8.7 *Théorème central limite relatif à des variables indépendantes seulement*
Soit X_1, X_2, \dots une suite de variables aléatoires d'espérances μ_i et de variances σ_i^2 , $i = 1, 2, \dots$. Si

a) les variables X_i sont uniformément bornées, ce qui signifie qu'il existe un réel M tel que $P\{|X_i| < M\} = 1$ pour tout i et

$$b) \sum_{i=1}^{\infty} \sigma_i^2 = \infty,$$

alors

$$P \left\{ \frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq a \right\} \rightarrow \Phi(a) \quad \text{quand } n \rightarrow \infty$$

8.3.4 Note historique : Pierre Simon, marquis de Laplace

Le théorème central limite a été à l'origine posé et démontré par le mathématicien français Pierre Simon, le marquis de Laplace, qui aboutit à ce théorème en observant que les erreurs de mesure (qui peuvent usuellement être considérées comme étant la somme d'un grand nombre de forces minuscules) tendaient à être normalement distribuées. Laplace, qui était aussi un célèbre astronome (il était en effet appelé «le Newton français»), fut un des premiers grands mathématiciens à contribuer aux probabilités et statistiques. Il rendit populaire l'utilisation des probabilités dans la vie de tous les jours. Il croyait fortement à leur importance, comme on le voit dans cette citation de son livre *Analytical Theory of Probability*. «On voit que la théorie des probabilités est à la base du sens commun réduit au calcul; elle nous fait apprécier avec exactitude ce que des esprits raisonnables sentent par une sorte d'instinct, souvent sans être capables d'expliquer pourquoi [...]. Il est remarquable que cette science, qui trouve ses origines dans l'étude des jeux de chance, puisse devenir l'objet le plus important de la connaissance humaine [...]. Les questions les plus importantes de la vie sont, pour une grande part, des problèmes de probabilités.»

L'application du théorème central limite, pour montrer que les erreurs de mesure sont approximativement normalement distribuées, est vue comme une importante contribution à la science. En effet, dans les 17^e et 18^e siècles, le théorème central limite était souvent appelé la «loi de fréquence des erreurs». La loi de fréquence des erreurs était considérée comme une avancée majeure par les scientifiques. Voyons les mots de Francis Galton (tirés de son livre *Natural Inheritance*, publié en 1889) à ce sujet. «Je ne connais rien d'aussi remarquable pour l'imagination que la merveilleuse forme que prend l'ordre cosmique dans la "Loi de Fréquence des erreurs". La Loi aurait été personnifiée et déifiée par les Grecs s'ils l'avaient connue. Elle règne avec

sérénité et dans un effacement complet parmi la confusion la plus sauvage. Plus la foule est énorme et plus l'anarchie apparente est grande, plus son influence est parfaite. C'est la loi suprême de la déraison.»

8.4 LOI FORTE DES GRANDS NOMBRES

8.4.1 Loi forte, version restreinte et version générale

La *loi forte des grands nombres* est sans doute le résultat le plus célèbre en théorie des probabilités. Il établit que la moyenne d'une suite de variables aléatoires identiquement distribuées tendra avec probabilité 1 vers l'espérance de cette distribution commune.

Théorème 8.8 *Loi forte des grands nombres*

Soit X_1, X_2, \dots une suite de variables indépendantes et identiquement distribuées, d'espérance commune finie μ . Alors, avec probabilité 1

$$\frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow \mu \quad \text{quand } n \rightarrow \infty^1$$

Il est inutile de démontrer ce théorème, car on en prouvera plus loin une version plus générale. On peut cependant ici présenter une application très importante de la loi forte des grands nombres. Supposons qu'on réalise une série d'épreuves indépendantes. Soit E un événement donné relatif à l'expérience ainsi répétée et $P(E)$ sa probabilité, constante au cours des tirages. On pose

$$X_i = \begin{cases} 1 & \text{si } E \text{ survient lors du } i\text{-ème tirage} \\ 0 & \text{sinon} \end{cases}$$

La loi forte des grands nombres établit qu'avec probabilité 1

$$\frac{X_1 + \dots + X_n}{n} \rightarrow E[X] = P(E) \quad (8.2)$$

Comme $X_1 + X_2 + \dots + X_n$ représente le nombre des occurrences de E au cours des n premiers tirages, (8.2) peut recevoir l'interprétation suivante: la fréquence relative limite d'apparition de l'événement E est $P(E)$ avec probabilité 1.

Bien que le théorème puisse être démontré sans cette hypothèse, notre preuve de la loi forte des grands nombres admet que les variables aléatoires X_i ont un quatrième moment fini, ce qui signifie que $E[X_i^4] = K < \infty$.

¹ En d'autres termes la loi forte des grands nombres signifie que

$$P \left\{ \lim_{n \rightarrow \infty} (X_1 + \dots + X_n)/n = \mu \right\} = 1$$

DÉMONSTRATION DE LA LOI FORTE DES GRANDS NOMBRES. Pour commencer, admettons que μ , la moyenne des X_i , est égale à 0. Soit $S_n = \sum_{i=1}^n X_i$ et considérons

$$E[S_n^4] = E[(X_1, \dots, X_n) (X_1, \dots, X_n) (X_1, \dots, X_n) (X_1, \dots, X_n)]$$

En développant le terme de droite, on obtient un résultat en termes de la forme

$$X_i^4, X_j^4, X_k^4, X_l^4, X_i^2 X_j^2, X_i^2 X_k^2, X_i^2 X_l^2, X_j^2 X_k^2, X_j^2 X_l^2, X_k^2 X_l^2 \text{ et } X_i X_j X_k X_l$$

où i, j, k, l , sont tous différents. Comme tous les X_i ont pour moyenne 0, on obtient par indépendance que

$$E[X_i^3 X_j] = E[X_i^3] E[X_j] = 0$$

$$E[X_i^2 X_j X_k] = E[X_i^2] E[X_j] E[X_k] = 0$$

$$E[X_i X_j X_k X_l] = 0$$

Pour une paire donnée i et j , il y a $\binom{4}{2} = 6$ termes dans le développement qui sont égaux à $X_i^2 X_j^2$. Cela entraîne donc, en développant le produit précédent et en prenant les espérances terme par terme, que

$$\begin{aligned} E[S_n^4] &= n E[X_i^4] + 6 \binom{n}{2} E[X_i^2 X_j^2] \\ &= n K + 3n(n-1) E[X_i^2] E[X_j^2] \end{aligned}$$

où nous avons utilisé encore une fois l'hypothèse d'indépendance. Maintenant, puisque

$$0 \leq \text{Var}(X_i^2) = E[X_i^4] - (E[X_i^2])^2$$

nous voyons que

$$(E[X_i^2])^2 \leq E[X_i^4] = K$$

D'où nous tirons, de la précédente expression, que

$$E[S_n^4] \leq n K + 3n(n-1) K$$

qui implique que

$$E\left[\frac{S_n^4}{n^4}\right] \leq \frac{K}{n^3} + \frac{3K}{n^2}$$

Donc, cela entraîne que

$$E \left[\sum_{n=1}^{\infty} \frac{S_n^4}{n^4} \right] = \sum_{n=1}^{\infty} E \left[\frac{S_n^4}{n^4} \right] < \infty$$

Mais ceci implique avec une probabilité égale à 1 que $\sum_{n=1}^{\infty} S_n^4/n^4 < \infty$. (S'il y a une probabilité positive que la somme soit infinie, alors son espérance est infinie.) Mais la convergence d'une série implique que son n -ième terme tend vers 0; et nous pouvons donc conclure, avec une probabilité égale à 1 que

$$\lim_{n \rightarrow \infty} \frac{S_n^4}{n^4} = 0$$

Mais si $S_n^4/n^4 = (S_n/n)^4$ tend vers 0, alors S_n/n doit aussi; nous avons donc prouvé avec une probabilité égale à 1 que

$$\frac{S_n}{n} \rightarrow 0 \quad \text{lorsque } n \rightarrow \infty$$

Lorsque μ , la moyenne des X_i , n'est pas égale à 0, on peut appliquer l'argument précédent aux variables aléatoires $X_i - \mu$ pour obtenir avec une probabilité égale à 1 que

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{(X_i - \mu)}{n} = 0$$

ou de manière équivalente,

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{X_i}{n} = \mu$$

ce qui prouve le résultat. ■

La loi forte des grands nombres fut démontrée pour la première fois et dans le cas particulier de variables de Bernoulli par le mathématicien français Borel. La version générale du théorème 8.8 est due au mathématicien russe A.N. Kolmogorov.

8.4.2 Comparaison de la loi faible et de la loi forte

Bien des gens ont du mal à saisir au début la différence entre loi faible et loi forte des grands nombres. La loi faible assure que pour toute grande valeur de n , disons n^* par exemple, $(X_1 + \dots + X_n)/n^*$ est probablement très voisine de μ . Elle n'assure pas cependant que $(X_1 + \dots + X_n)/n$ devra rester dans un voisinage étroit de μ pour toutes les valeurs de n supérieures à n^* . Elle laisse donc la porte ouverte à une situation où de larges écarts entre $(X_1 + \dots + X_n)/n$ et μ peuvent se produire pour une infinité d'événements, infinité dont la probabilité collective est très faible cependant. La loi forte exclut cette situation. Elle assure en particulier qu'avec probabilité 1 et pour toute valeur $\varepsilon > 0$,

$$\left| \sum_{i=1}^n \frac{X_i}{n} - \mu \right|$$

ne sera supérieure à ε qu'un nombre fini de fois.

8.5 AUTRES INÉGALITÉS

8.5.1 Inégalité unilatérale de Tchebychev

On rencontre parfois des situations dans lesquelles on voudrait obtenir une borne supérieure pour une probabilité de la forme $P\{X - \mu \geq a\}$, où a est un réel positif et où X est de distribution inconnue, mais d'espérance μ et de variance σ^2 qui sont, elles, connues. On se trouve ici évidemment dans un cas d'application de l'inégalité de Tchebychev puisque $X - \mu \geq a > 0$ entraîne $|X - \mu| \geq a$; on obtient

$$P\{X - \mu \geq a\} \leq P\{|X - \mu| \geq a\} \leq \frac{\sigma^2}{a^2} \quad \text{lorsque } a > 0$$

Le théorème suivant montre cependant qu'on peut faire mieux.

Théorème 8.9 Inégalité unilatérale de Tchebychev

Soit X une variable aléatoire d'espérance nulle et de variance σ^2 finie. On a alors pour tout réel $a > 0$

$$P\{X \geq a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

DÉMONSTRATION. Soit $b > 0$ et notons que $X \geq a$ est équivalent à $X + b \geq a + b$.

Donc

$$\begin{aligned} P\{X \geq a\} &= P\{X + b \geq a + b\} \\ &\leq P\{(X + b)^2 \geq (a + b)^2\} \end{aligned}$$

où l'inégalité ci-dessus est obtenue en remarquant que comme $a + b > 0$, $X + b \geq a + b$ implique que $(X + b)^2 \geq (a + b)^2$. En appliquant l'inégalité de Markov, on obtient que

$$P\{X \geq a\} \leq \frac{E[(X + b)^2]}{(a + b)^2} = \frac{\sigma^2 + b^2}{(a + b)^2}$$

En posant $b = \sigma^2 / a$ (dont on peut facilement voir qu'il s'agit de la valeur de b qui minimise $(\sigma^2 + b^2) / (a + b)^2$), on obtient le résultat désiré. ■

Exemple 8.8 Le nombre de pièces produites par une usine en une semaine est une variable aléatoire d'espérance 100 et de variance 400. On cherche une borne supérieure pour la probabilité que la production dépasse 120 pièces lors d'une semaine donnée.

SOLUTION. Appliquons l'inégalité unilatérale de Tchebychev

$$P\{X \geq 120\} = P\{X - 100 \geq 20\} \leq \frac{400}{400 + (20)^2} = \frac{1}{2}$$

où l'on voit que la probabilité de dépasser une production de 120 pièces en une semaine ne dépasse pas $\frac{1}{2}$.

Il était possible d'obtenir une borne grâce à l'inégalité de Markov, qui aurait donné

$$P\{X \geq 120\} \leq \frac{E(X)}{120} = \frac{5}{6}$$

c'est-à-dire une borne nettement moins bonne que la précédente. ■

De l'inégalité unilatérale de Tchebychev, on peut déduire le théorème suivant.

Théorème 8.10

Soit X une variable aléatoire, avec $E[X] = \mu$ et $Var(X) = \sigma^2$. On aura pour tout réel $a > 0$,

$$P\{X \geq \mu + a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

$$P\{X \leq \mu - a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

DÉMONSTRATION. Les variables $X - \mu$ et $\mu - X$ ont toutes deux une espérance nulle et une variance σ^2 . De ce fait

$$P\{X - \mu \geq a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

et

$$P\{\mu - X \leq a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

ce qui établit le résultat annoncé. ■

Exemple 8.9 On répartit au hasard en 100 couples un groupe initialement composé de 100 hommes et 100 femmes. On cherche une borne supérieure à la probabilité que moins de 30 des 100 couples formés soient mixtes.

SOLUTION. On numérote les hommes arbitrairement de 1 à 100 puis on pose

$$X_i = \begin{cases} 1 & \text{si l'homme } i \text{ est associé à une femme} \\ 0 & \text{sinon} \end{cases}$$

ceci pour $i = 1, 2, \dots, 100$. On peut exprimer le nombre de couples mixtes comme la somme

$$X = \sum_{i=1}^{100} X_i$$

Comme l'homme i a autant de chances d'être associé à chacune des 199 autres personnes, dont 100 sont des femmes, on a

$$E[X_i] = P\{X_i = 1\} = \frac{100}{199}$$

Par un raisonnement similaire, et si $i \neq j$, on calcule

$$\begin{aligned} E[X_i X_j] &= P\{X_i = 1, X_j = 1\} \\ &= P\{X_i = 1\} P\{X_j = 1 | X_i = 1\} = \frac{100}{199} \frac{99}{197} \end{aligned}$$

On peut se convaincre que $P\{X_j = 1 | X_i = 1\} = \frac{99}{197}$ en réalisant que, si l'homme i est associé à une femme, l'homme j peut être associé avec autant de chances à l'une quelconque des 197 personnes restantes, dont 99 sont des femmes. On peut maintenant calculer

$$\begin{aligned} E[X] &= \sum_{i=1}^{100} E[X_i] \\ &= (100) \frac{100}{199} \\ &\approx 50.25 \\ \text{Var}(X) &= \sum_{i=1}^{100} \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \\ &= 100 \frac{100}{199} \frac{99}{199} + 2 \binom{100}{2} \left[\frac{100}{199} \frac{99}{197} - \left(\frac{100}{199} \right)^2 \right] \\ &\approx 25.126 \end{aligned}$$

et appliquer l'inégalité de Tchebychev

$$P\{X \leq 30\} \leq P\{|X - 50.25| \geq 20.25\} \leq \frac{25.126}{(20.25)^2} \approx .061$$

L'événement «moins de 30 hommes seront associés chacun à une femme» a donc au plus 6 chances sur 100 de se produire. Cette borne peut cependant être améliorée par application de l'inégalité unilatérale de Tchebychev qui donne, elle,

$$\begin{aligned}
 P\{X \leq 30\} &= P\{X \leq 50.25 - 20.25\} \\
 &\leq \frac{25.126}{25.126 + (20.25)^2} \\
 &\approx .058
 \end{aligned}$$

8.5.2 Bornes de Chernoff

Lorsque la fonction génératrice des moments d'une variable aléatoire X est connue, on peut obtenir une borne plus efficace pour $P\{X \geq a\}$. Soit $M(t) = E[e^{tX}]$ la fonction génératrice des moments de la variable aléatoire X . Alors pour $t > 0$,

$$\begin{aligned}
 P\{X \geq a\} &= P\{e^{tX} \geq e^{ta}\} \\
 &\leq E[e^{tX}] e^{-ta}
 \end{aligned}$$

par l'inégalité de Markov

De la même façon, pour $t < 0$,

$$\begin{aligned}
 P\{X \leq a\} &= P\{e^{tX} \geq e^{ta}\} \\
 &\leq E[e^{tX}] e^{-ta}
 \end{aligned}$$

Ainsi, nous avons les inégalités suivantes, connues sous le nom des bornes de Chernoff.

Théorème 8.11 *Bornes de Chernoff*

$$\begin{aligned}
 P\{X \geq a\} &\leq e^{-ta} M(t) \quad \text{pour tout } t > 0 \\
 P\{X \leq a\} &\leq e^{-ta} M(t) \quad \text{pour tout } t < 0
 \end{aligned}$$

Puisque les bornes de Chernoff sont valables pour tout t dans les quadrants positifs et négatifs, on obtient la meilleure borne pour $P\{X \geq a\}$ en utilisant le t qui minimise $e^{-ta} M(t)$.

Exemple 8.10 Bornes de Chernoff pour une variable aléatoire normale standard.

Si Z est une variable aléatoire normale standard, sa fonction génératrice des moments est $M(t) = e^{t^2/2}$, donc la borne de Chernoff pour $P\{Z \geq a\}$ est donnée par

$$P\{Z \geq a\} \leq e^{-ta} e^{t^2/2} \quad \text{pour tout } t > 0.$$

La valeur de t , $t > 0$, qui minimise $e^{t^2/2 - ta}$ est la valeur qui minimise $t^2/2 - ta$, soit $t = a$. Ainsi, pour $a > 0$, on voit que

$$P\{Z \geq a\} \leq e^{-a^2/2}$$

De la même manière, on peut montrer que pour $a < 0$,

$$P\{Z \leq a\} \leq e^{-a^2/2} \quad \blacksquare$$

Exemple 8.11 Bornes de Chernoff pour une variable aléatoire de Poisson.

Si X est une variable aléatoire de Poisson de paramètre λ , sa fonction génératrice des moments est $M(t) = e^{\lambda(e^t - 1)}$. Ainsi la borne de Chernoff pour $P\{X \geq i\}$ est

$$P\{X \geq i\} \leq e^{\lambda(e^t - 1) - it} \quad , t > 0$$

Minimiser le membre de droite de cette expression revient à minimiser $\lambda(e^t - 1) - it$ et le calcul montre que la valeur minimale est atteinte quand $e^t = i/\lambda$. En stipulant que $i/\lambda > 1$, cette valeur de t qui donne le minimum est positive. Donc, en admettant que $i > \lambda$ et en posant $e^t = i/\lambda$ dans la borne de Chernoff, on obtient que

$$P\{X \geq i\} \leq e^{\lambda(i/\lambda - 1)} \left(\frac{\lambda}{i}\right)^i$$

ou, de manière équivalente,

$$P\{X \geq i\} \leq \frac{e^{-\lambda} (e\lambda)^i}{i^i} \quad \blacksquare$$

Exemple 8.12 Considérons un joueur qui, à chaque coup, a autant de chance, indépendamment du passé, de gagner ou de perdre 1 unité. Ceci signifie que si X_i représente les gains du joueur au i -ème coup, alors les X_i sont indépendants et

$$P\{X_i = 1\} = P\{X_i = -1\} = 1/2.$$

Soit $S_n = \sum_{i=1}^n X_i$ les gains du joueur après n coups. Nous allons utiliser la borne de Chernoff sur $P\{S_n \geq a\}$. Pour commencer, notons que la fonction génératrice des moments de X_i est

$$E[e^{tX}] = \frac{e^t + e^{-t}}{2}$$

Maintenant, en utilisant les développements de McLaurin de e^t et e^{-t} , on voit que

$$\begin{aligned} e^t + e^{-t} &= 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots + (1 - t + \frac{t^2}{2!} - \frac{t^3}{3!} + \dots) \\ &= 2 \left(1 + \frac{t^2}{2!} + \frac{t^4}{4!} + \dots \right) \\ &= 2 \sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!} \end{aligned}$$

$$\begin{aligned} &\leq 2 \sum_{n=0}^{\infty} \frac{(t^2/2)^n}{n!} \quad \text{puisque } (2n)! \geq n! 2^n \\ &= 2 e^{t^2/2} \end{aligned}$$

Donc $E[e^{tX}] \leq e^{t^2/2}$.

Puisque la fonction génératrice des moments de la somme de variables aléatoires indépendantes est le produit de leurs fonctions génératrices des moments, on a

$$\begin{aligned} E[e^{tS_n}] &= (E[e^{tX}])^n \\ &\leq e^{nt^2/2} \end{aligned}$$

En utilisant ce résultat avec la borne de Chernoff, on obtient

$$P\{S_n \geq a\} \leq e^{-ta} e^{nt^2/2}, \quad t > 0$$

La valeur de t qui minimise le membre de droite de cette expression est la valeur qui minimise $nt^2/2 - ta$, et cette valeur est $t = a/n$. En supposant que $a > 0$ (de telle sorte que le t qui minimise soit positif) et en posant $t = a/n$ dans l'inégalité précédente, on obtient

$$P\{S_n \geq a\} \leq e^{-a^2/2n}, \quad a > 0$$

Par exemple, cette inégalité donne

$$P\{S_{10} \geq 6\} \leq e^{-36/20} \approx .1653$$

alors que la probabilité exacte est

$$P\{S_{10} \geq 6\} = P\{\text{le joueur gagne au moins 8 fois sur les 10 premiers coups}\}$$

$$\begin{aligned} &= \frac{\binom{10}{8} + \binom{10}{9} + \binom{10}{10}}{2^{10}} \\ &= \frac{56}{1024} \approx .0547 \end{aligned} \quad \blacksquare$$

8.5.3 Inégalité de Jensen

L'inégalité que nous allons maintenant traiter porte sur des espérances plutôt que des probabilités. Nous avons auparavant besoin d'introduire la notion de *convexité de fonction*: une fonction réelle f deux fois différentiable est dite convexe si $f''(x) \geq 0$ pour tout x . Une fonction sera de même dite concave si $f''(x) \leq 0$ pour tout x . A

titre d'exemples, les fonctions suivantes sont convexes: $f(x) = x^2$, $f(x) = e^{ax}$, $f(x) = -x^{1/n}$ lorsque $x \geq 0$. Si f est convexe la fonction $g = -f$ est concave et vice versa.

Théorème 8.12 Inégalité de Jensen

Soit f une fonction convexe. Alors

$$E[f(X)] \geq f(E[X])$$

pour autant que ces espérances existent et soient finies.

DÉMONSTRATION. Ecrivons le développement en série de Taylor de f autour de $\mu = E[X]$:

$$f(x) = f(\mu) + f'(\mu)(x - \mu) + \frac{f''(\xi)(x - \mu)^2}{2}$$

où ξ est un réel compris entre x et μ . Comme $f''(\xi) \geq 0$, on aura

$$f(x) \geq f(\mu) + f'(\mu)(x - \mu)$$

et donc

$$f(X) \geq f(\mu) + f'(\mu)(X - \mu)$$

Il suffit de prendre l'espérance des deux membres pour obtenir

$$E[f(X)] \geq f(\mu) + f'(\mu)E[X - \mu] = f(\mu)$$

ce qui achève la démonstration. ■

Exemple 8.13 Un investisseur est confronté à une alternative. Soit il place tout son capital dans une affaire risquée rapportant une somme aléatoire X d'espérance m , soit il le place en titres sans risques qui rapporteront une somme m avec probabilité 1. On sait qu'il va chercher à prendre sa décision de manière à maximiser l'espérance de $u(R)$, où R est son bénéfice et u sa fonction de préférence. L'inégalité de Jensen nous montre que si u est une fonction concave, $E[u(X)] \leq u(m)$, ce qui rend le placement sûr préférable. Si par contre u est convexe, le placement risqué doit être choisi puisque $E[u(X)] \geq u(m)$. ■

8.6 BORNES POUR L'ERREUR DE PROBABILITÉ COMMISE EN APPROXIMANT UNE LOI BINOMIALE PAR UNE LOI DE POISSON

Dans cette section, nous allons établir des bornes pour savoir comment une somme de variables aléatoires de Bernoulli indépendantes est approximée par une variable aléatoire de Poisson ayant la même moyenne.

Supposons qu'on veuille approximer la somme de variables aléatoires de Bernoulli indépendantes de moyennes respectives p_1, p_2, \dots, p_n . En partant avec une séquence Y_1, \dots, Y_n de variables aléatoires de Poisson indépendantes, Y_i ayant pour moyenne p_i , nous allons construire une séquence de variables aléatoires indépendantes de Bernoulli X_1, \dots, X_n de paramètres p_1, \dots, p_n telle que

$$P\{X_i \neq Y_i\} \leq p_i^2 \quad \text{pour tout } i$$

En posant $X = \sum_{i=1}^n X_i$ et $Y = \sum_{i=1}^n Y_i$, le fait précédent nous permet de conclure que

$$P\{X \neq Y\} \leq \sum_{i=1}^n p_i^2$$

Finalement, nous allons montrer que cette inégalité implique que pour tout ensemble A de nombres réels,

$$|P\{X \in A\} - P\{Y \in A\}| \leq \sum_{i=1}^n p_i^2$$

Puisque X est la somme de variables aléatoires indépendantes de Bernoulli et que Y est une variable aléatoire de Poisson, l'inégalité précédente nous mènera à la borne voulue.

Pour montrer comment l'inégalité précédente se produit, soient Y_i , $i=1, \dots, n$, des variables aléatoires indépendantes de Poisson de moyennes respectives p_i . Soient à présent U_1, \dots, U_n des variables aléatoires indépendantes qui sont aussi indépendantes des Y_i et qui sont telles que

$$U_i = \begin{cases} 0 & \text{avec probabilité } (1-p_i)e^{p_i} \\ 1 & \text{avec probabilité } 1 - (1-p_i)e^{p_i} \end{cases}$$

Cette définition utilise implicitement l'inégalité $e^{-p} \geq 1-p$ en admettant que $(1-p_i)e^{p_i} \leq 1$.

Définissons à présent les variables aléatoires X_i , $i=1, \dots, n$ par

$$X_i = \begin{cases} 0 & \text{si } Y_i = U_i = 0 \\ 1 & \text{sinon} \end{cases}$$

Noter que

$$P\{X_i = 0\} = P\{Y_i = 0\}P\{U_i = 0\} = e^{-p_i}(1-p_i)e^{p_i} = 1-p_i$$

$$P\{X_i = 1\} = 1 - P\{X_i = 0\} = p_i$$

Si X_i est nul, alors Y_i doit être nul aussi (par définition de X_i). Donc on voit que

$$\begin{aligned}
 P\{X_i \neq Y_i\} &= P\{X_i = 1, Y_i \neq 1\} \\
 &= P\{Y_i = 0, X_i = 1\} + P\{Y_i > 1\} \\
 &= P\{Y_i = 0, U_i = 1\} + P\{Y_i > 1\} \\
 &= e^{-p_i} [1 - (1 - p_i)e^{p_i}] + 1 - e^{-p_i} - p_i e^{-p_i} \\
 &= p_i - p_i e^{-p_i} \\
 &\leq p_i^2 \quad (\text{puisque } 1 - e^{-p} \leq p)
 \end{aligned}$$

A présent, soit $X = \sum_{i=1}^n X_i$ et $Y = \sum_{i=1}^n Y_i$ et notons que X est la somme de variables aléatoires indépendantes de Bernoulli et que Y est une Poisson, d'espérances $E[Y] = E[X] = \sum_{i=1}^n p_i$.

Notons aussi que l'inégalité $X \neq Y$ implique que $X_i \neq Y_i$ pour un i , donc

$$\begin{aligned}
 P\{X \neq Y\} &\leq P\{X_i \neq Y_i \text{ pour un } i\} \\
 &\leq \sum_{i=1}^n P\{X_i \neq Y_i\} && (\text{inégalité de Boole}) \\
 &\leq \sum_{i=1}^n p_i^2
 \end{aligned}$$

Pour n'importe quel événement B , soit I_B la variable indicatrice de cet événement B , définie par

$$I_B = \begin{cases} 1 & \text{si } B \text{ se produit} \\ 0 & \text{sinon} \end{cases}$$

Notons que pour n'importe quel ensemble de nombres réels A ,

$$I_{\{X \in A\}} - I_{\{Y \in A\}} \leq I_{\{X \neq Y\}}$$

Cette inégalité provient du fait que, puisqu'une variable indicatrice vaut soit 0 soit 1, le membre de gauche ne vaut 1 que lorsque $I_{\{X \in A\}} = 1$ et $I_{\{Y \in A\}} = 0$.

Mais ceci implique que $X \in A$ et $Y \notin A$, ce qui signifie que $X \neq Y$, donc que le membre de droite est aussi égal à 1. En prenant les espérances de cette inégalité, on obtient

$$P\{X \in A\} - P\{Y \in A\} \leq P\{X \neq Y\}$$

En inversant X et Y , on obtient de la même façon que

$$P\{Y \in A\} - P\{X \in A\} \leq P\{X \neq Y\}$$

et nous pouvons conclure que

$$|P\{X \in A\} - P\{Y \in A\}| \leq P\{X \neq Y\}$$

Donc, on a démontré qu'avec $\lambda = \sum_{i=1}^n p_i$,

$$\left| P\left\{ \sum_{i=1}^n X_i \in A \right\} - \sum_{i \in A} \frac{e^{-\lambda} \lambda^i}{i!} \right| \leq \sum_{i=1}^n p_i^2$$

REMARQUE. Lorsque tous les p_i sont égaux à p , X est une variable aléatoire binomiale. Ainsi, cette inégalité montre que pour n'importe quel ensemble A d'entiers non négatifs,

$$\left| \sum_{i \in A} \binom{n}{i} p^i (1-p)^{n-i} - \sum_{i \in A} \frac{e^{-np} (np)^i}{i!} \right| \leq np^2$$

8.7 EXERCICES THÉORIQUES

8.7.1 Soit σ^2 la variance de X . La racine positive σ de la variance est appelée *écart-type*. Si X est de moyenne μ et d'écart-type σ , montrer que

$$P\{|X - \mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

8.7.2 Soit X une variable de moyenne μ et d'écart-type σ ; le rapport $r \equiv |\mu|/\sigma$ est appelé *rapport signal-bruit* de X . L'idée vient de ce que X peut être exprimée comme $X = \mu + (X - \mu)$ où μ représente le signal et $X - \mu$ le bruit, dont l'écart-type est σ . On définit $|(X - \mu)/\mu| \equiv D$ comme l'erreur relative de X par rapport à son signal (ou espérance) μ . Montrer que pour $\alpha > 0$

$$P\{D \leq \alpha\} \geq 1 - \frac{1}{r^2 \alpha^2}$$

8.7.3 Calculer le rapport signal-bruit, c'est-à-dire $|\mu|/\sigma$, où $\mu = E[X]$ et $\sigma^2 = \text{Var}(X)$ pour les variables aléatoires suivantes:

- Poisson de moyenne λ ;
- binomiale de paramètres n et p ;
- géométrique de moyenne $1/p$;
- uniforme sur l'intervalle (a, b) ;
- exponentielle de moyenne $1/\lambda$;
- normale de paramètres μ, σ^2 .

8.7.4 Soit $Z_n, n \geq 1$, une suite de variables aléatoires et c une constante telle que pour chaque $\varepsilon > 0, P\{|Z_n - c| > \varepsilon\} \rightarrow 0$ quand $n \rightarrow \infty$. Montrer que pour toute fonction

continue bornée g ,

$$E[g(Z_n)] \rightarrow g(c) \text{ quand } n \rightarrow \infty$$

8.7.5 Soit $f(x)$ une fonction continue définie pour $0 \leq x \leq 1$. Considérer les fonctions

$$B_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$$

(appelées *polynômes de Bernstein*) et prouver que

$$\lim_{n \rightarrow \infty} B_n(x) = f(x)$$

Considérer, à titre d'indication, des variables aléatoires indépendantes de Bernoulli X_1, X_2, \dots ayant pour espérance x . Montrer, puis utiliser le fait suivant (en faisant usage du résultat de 8.7.4):

$$B_n(x) = E\left[f\left(\frac{X_1 + \dots + X_n}{n}\right)\right]$$

Comme on peut montrer que la convergence de $B_n(x)$ vers $f(x)$ est uniforme en x , ce qui est établi ci-dessus donne une preuve probabiliste du fameux théorème d'analyse dit *de Weierstrass*, théorème qui énonce que toute fonction continue sur un intervalle fermé peut être approximée de façon arbitrairement proche par un polynôme.

8.7.6

- Soit X une variable aléatoire discrète pouvant prendre les valeurs $1, 2, \dots$. Si $P\{X = k\}$ est non croissante en $k = 1, 2, \dots$, prouver que

$$P\{X = k\} \leq 2 \frac{E[X]}{k^2}$$

- Soit X une variable aléatoire continue non négative de densité non croissante. Montrer que

$$f(x) \leq \frac{2E[X]}{x^2} \quad \text{pour tout } x > 0$$

8.7.7 Supposons qu'un dé non pipé soit jeté 100 fois. Soit X_i la valeur obtenue au i -ème jet. Calculer une approximation pour

$$P\left\{\prod_{i=1}^{100} X_i \leq a^{100}\right\} \quad 1 < a < 6$$

8.7.8 Expliquer pourquoi une variable aléatoire gamma de paramètres (t, λ) a une distribution approximativement normale quand t est grand.

8.7.9 Une pièce de monnaie équilibrée est jetée 1 000 fois. Si les 100 premiers jets donnent tous des piles, quelle proportion de piles peut-on s'attendre à obtenir lors des 900 derniers jets? Faites un commentaire sur l'énoncé «la loi forte des grands nombres n'est pas une anomalie dans la masse mais ne la compense pas».

8.7.10 Si X est une variable aléatoire de Poisson de moyenne λ , montrer que pour $i < \lambda$,

$$P\{X \leq i\} \leq \frac{e^{-\lambda} (e \lambda)^i}{i^i}$$

8.7.11 Soit X une variable aléatoire binomiale de paramètres n et p . Montrer que pour $i > np$:

a) le minimum de $e^{-ti} E[e^{tX}]$ est atteint lorsque t est tel que $e^t = \frac{iq}{(n-i)p}$ où $q = 1 - p$.

b) $P\{X \geq i\} \leq \frac{n^n}{i^i (n-i)^{n-i}} p^i (1-p)^{n-i}$

8.7.12 La borne de Chernoff d'une variable aléatoire normale standard Z donne que $P\{Z > a\} \leq e^{-a^2/2}$, $a > 0$. Montrer, en considérant la densité de Z , que la partie droite de l'inégalité peut être réduite par un facteur 2. En fait, montrer que

$$P\{Z > a\} \leq \frac{1}{2} e^{-a^2/2}, \quad a > 0$$

8.7.13 Si $E[X] < 0$ et $\theta \neq 0$ est tel que $E[e^{\theta X}] = 1$, montrer que $\theta > 0$.

8.8 PROBLÈMES

8.8.1 Soit X une variable aléatoire d'espérance et de variance toutes deux égales à 20. Que peut-on dire de $P\{0 \leq X \leq 40\}$?

8.8.2 Un professeur sait par expérience que la note de test d'un étudiant se présentant à un examen final est une variable aléatoire d'espérance 75.

a) Donner une borne supérieure à la probabilité que la note de test d'un étudiant dépasse 85;

Supposons maintenant que le professeur sache en plus que la variance de la note de test d'un étudiant est 25.

b) Que peut-on dire de la probabilité qu'un étudiant obtienne une note comprise entre 65 et 85?

c) Combien faudrait-il qu'il se présente d'étudiants à cet examen pour assurer, avec une probabilité d'au moins 0,9, que la moyenne de la classe soit de 75 plus ou moins 5? Ne pas utiliser le théorème central limite.

8.8.3 Utiliser le théorème central limite pour résoudre la partie c) du problème 8.7.2.

8.8.4 Soient X_1, \dots, X_{20} des variables aléatoires indépendantes de Poisson d'espérance 1.

- a) Utiliser l'inégalité de Markov pour obtenir une borne de $P\{\sum_{i=1}^{20} X_i > 15\}$.
 b) Utiliser le théorème central limite pour obtenir une approximation de $P\{\sum_{i=1}^{20} X_i > 15\}$.

8.8.5 On arrondit 50 nombres à l'entier le plus proche et on effectue la somme. Si les erreurs d'arrondi individuels sont distribuées uniformément sur $(-0,5, 0,5)$, quelle est la probabilité que la somme obtenue ait un écart de plus de 3 par rapport à la somme exacte?

8.8.6 On lance un dé jusqu'à ce que la somme totale des nombres obtenus dépasse 300. Quelle est la probabilité qu'il faille au moins 80 jets?

8.8.7 On a 100 ampoules dont les durées de vie sont des variables aléatoires indépendantes exponentielles de moyenne 5 heures. Si l'on allume une ampoule à la fois et qu'une ampoule grillée est instantanément remplacée par une neuve, qu'elle est la probabilité qu'il reste encore une ampoule intacte après 525 heures?

8.8.8 Si, dans le problème 8.8.7, le temps de remplacement d'une ampoule défectueuse suit une loi uniforme sur $(0, 0,5)$, quelle est la probabilité que toutes les ampoules soient grillées après 550 heures?

8.8.9 Soit X une variable aléatoire gamma de paramètres $(n, 1)$. Quelle doit être la valeur de n pour que

$$P\{|X/n - 1| > .01\} < .01$$

8.8.10 Des ingénieurs civils pensent que W , le poids (en milliers de livres) qu'une travée d'un pont peut supporter sans subir de dommage au niveau de sa structure, suit une loi normale, de moyenne 400 et d'écart-type 40. Supposons que le poids (également en milliers de livres) d'une voiture est une variable aléatoire normale de moyenne 3 et d'écart-type 0,3. Combien de voitures devraient se trouver sur cette travée pour que la probabilité de rupture soit supérieure à 0,1?

8.8.11 Plusieurs personnes pensent que la fluctuation journalière du prix de l'action d'une société donnée, cotée en bourse, est une variable aléatoire de moyenne 0 et de variance σ^2 . Cela veut dire que, si Y_n représente le prix de l'action lors du n -ième jour,

$$Y_n = Y_{n-1} + X_n \quad n \geq 1$$

où X_1, X_2, \dots sont des variables aléatoires indépendantes identiquement distribuées d'espérance 0 et de variance σ^2 . Supposons que le prix de l'action soit aujourd'hui de 100. Si $\sigma^2 = 1$, que peut-on dire de la probabilité que le prix de l'action excède 105 après 10 jours?

8.8.12 Nous avons 100 composants que nous allons employer les uns après les autres. Cela veut dire que le composant 1 sera d'abord utilisé, puis lorsqu'il tombera en panne, il sera remplacé par le composant 2, qui sera lui-même remplacé après

défaillance par le composant 3, et ainsi de suite. Si la durée de vie du composant i est distribuée de façon exponentielle avec espérance $10 + i/10$, $i = 1, \dots, 100$, estimer la probabilité que la durée de vie totale de l'ensemble des composants dépasse 1 200. Refaites le même exercice lorsque la distribution de la durée de vie des composants est uniforme sur l'intervalle $(0, 20 + i/5)$, $i = 1, \dots, 100$.

8.8.13 Les résultats d'un examen donné par un certain professeur ont une moyenne de 74 et une déviation standard de 14. Ce professeur a donné deux examens; l'un à une classe de 25 élèves et l'autre à une classe de 64 élèves.

- Approximer la probabilité que la moyenne du test dans la classe de 25 élèves dépasse 80.
- Refaire la question a) pour la classe de 64.
- Approximer la probabilité que la moyenne de la plus grande classe dépasse celle de l'autre de 2.2 points.
- Approximer la probabilité que la moyenne de la plus petite classe dépasse celle de l'autre de 2.2 points.

8.8.14 Un certain composant joue un rôle critique dans un système électrique et doit être remplacé immédiatement à chaque panne. Si la durée de vie moyenne de ce type de composant est de 100 heures et que sa déviation standard est 30 heures, combien de ces composants doit-on avoir en stock pour que la probabilité que le système marche continuellement les 2000 prochaines heures soit au moins de .95 ?

8.8.15 Une assurance a 10 000 automobilistes assurés. L'espérance annuelle d'indemnités demandées par un assuré est \$240 avec une déviation standard de \$800. Approximer la probabilité que les indemnités annuelles totales dépassent \$2.7 millions.

8.8.16 Refaire l'exemple 8.9 sous l'hypothèse que le nombre de couples est (approximativement) distribuée selon une loi normale. Cela vous semble-t-il être une supposition raisonnable?

8.8.17 Refaire la partie a) du problème 8.8.2 en sachant que la variance de la note de test de l'étudiant est 25.

8.8.18 Un lac contient 4 espèces de poissons. On suppose qu'un poisson capturé au hasard a autant de chances de représenter n'importe laquelle des espèces. Soit Y le nombre de poissons à prendre de façon à obtenir au moins un poisson de chaque type.

- Donner un intervalle (a, b) tel que $P\{a \leq Y \leq b\} \geq 0,90$.
- En utilisant l'inégalité de Tchebychev unilatérale, combien de poissons doit-on envisager de capturer pour être certain à au moins 90% d'obtenir au moins un poisson de chaque type?

8.8.19 Soit X une variable aléatoire non négative de moyenne 25. Que peut-on dire des espérances suivantes:

- $E[X^3]$;
- $E[\sqrt{X}]$;
- $E[\ln x]$;
- $E[e^{-X}]$?

8.8.20 Soit X une variable aléatoire non négative. Prouver que

$$E[X] \leq (E[X^2])^{1/2} \leq (E[X^3])^{1/3} \leq \dots$$

8.8.21 Est-ce que les résultats de l'exemple 8.13 auraient changé si l'homme d'affaires avait pu répartir son argent à raison d'une part α , $0 < \alpha < 1$, destinée au placement à haut risque et d'une autre part $1 - \alpha$ destinée à l'opération sans risque? Son gain après un tel investissement mixte serait alors $R = \alpha X + (1 - \alpha)m$.

8.8.22 Soit X une variable aléatoire de Poisson de moyenne 20.

- Utiliser l'inégalité de Markov pour obtenir une borne supérieure de $p = P\{X \geq 26\}$.
- Utiliser l'inégalité unilatérale de Chebyshev pour obtenir une borne supérieure de p .
- Approximer p en utilisant le théorème central limite.
- Déterminer p en faisant tourner un programme approprié.

Thèmes choisis de probabilité

9.1 PROCESSUS DE POISSON

9.1.1 Définition

Avant de définir le processus de Poisson, rappelons qu'une fonction f est dite $o(h)$ si $\lim_{h \rightarrow 0} f(h)/h = 0$. Ceci revient à dire que lorsque f est $o(h)$, si h prend de très petites valeurs, $f(h)$ reste très petit par rapport à h .

Supposons maintenant que certains événements se répartissent de manière aléatoire dans le temps. Désignons par $N(t)$ le nombre d'événements survenus dans l'intervalle $[0, t]$. On dit que le processus stochastique $\{N(t), t \geq 0\}$ est **un processus de Poisson de paramètre λ , $\lambda > 0$** , si

- a) $N(0) = 0$;
- b) les nombres d'événements survenant au cours d'intervalles disjoints sont des variables indépendantes;
- c) la distribution du nombre d'événements survenant pendant un intervalle donné ne dépend que de la longueur de cet intervalle, et pas de son origine en particulier;
- d) $P\{N(h) = 1\} = \lambda h + o(h)$;
- e) $P\{N(h) \geq 2\} = o(h)$.

La condition a) précise que le processus commence au temps 0. La condition b) est dite hypothèse d'*indépendance des incréments*. Son nom lui vient de ce qu'elle impose entre autres que le nombre d'occurrences d'événements entre les instants t et $t + s$ [à savoir $N(t + s) - N(t)$] soit indépendant du nombre des occurrences jusqu'au temps t [à savoir $N(t)$]. La condition c) appelée *hypothèse de stationnarité* établit que la distribution de $N(t + s) - N(t)$ est la même pour toutes les valeurs de t .

Nous avons donné au chapitre 4 une explication qui montrait que $N(t)$ suit une loi de Poisson de paramètre λt et qui est basée sur le fait que la loi de Poisson est la

forme limite de certaines distributions de variables binomiales. Ce même résultat va maintenant être démontré d'une autre manière dans les paragraphes 9.1.2 et 9.1.3.

9.1.2 Distributions des temps d'attente

Avant d'en venir à ces distributions, établissons le théorème suivant:

Théorème 9.1

Pour tout processus de Poisson de paramètre λ

$$P\{N(t) = 0\} = e^{-\lambda t}$$

DÉMONSTRATION. Posons $P_0(t) = P\{N(t) = 0\}$. On va faire apparaître une équation différentielle dont $P_0(t)$ sera solution:

$$\begin{aligned} P_0(t+h) &= P\{N(t+h) = 0\} \\ &= P\{N(t) = 0, N(t+h) - N(t) = 0\} \\ &= P\{N(t) = 0\}P\{N(t+h) - N(t) = 0\} \\ &= P_0(t)[1 - \lambda h + o(h)] \end{aligned}$$

où les deux dernières transformations sont légitimées par l'hypothèse b) d'une part et par le fait, d'autre part, que $P\{N(h) = 0\} = 1 - \lambda h + o(h)$ en vertu des hypothèses c) et d). Donc

$$\frac{P_0(t+h) - P_0(t)}{h} = -\lambda P_0(t) + \frac{o(h)}{h}$$

Faisons maintenant tendre h vers 0. On obtient

$$P_0'(t) = -\lambda P_0(t)$$

ou encore

$$\frac{P_0'(t)}{P_0(t)} = -\lambda$$

qui entraîne, en intégrant les deux membres

$$\ln P_0(t) = -\lambda t + c$$

ou

$$P_0(t) = Ke^{-\lambda t}$$

Comme $P_0(0) = P\{N(0) = 0\} = 1$ on peut conclure que

$$P_0(t) = e^{-\lambda t} \quad \blacksquare$$

Nous noterons désormais T_1 la date du premier événement. Pour $n > 1$ les variables aléatoires T_n représenteront le temps écoulé entre le $(n-1)$ -ième et le n -ième

événement. La suite $\{T_n, n = 1, 2, \dots\}$ est appelée suite des *intervalles d'attente*. Si par exemple $T_1 = 5$ et $T_2 = 10$, nous sommes dans le cas où le premier événement de notre processus de Poisson est arrivé au temps 5 et le second au temps 15. La distribution de ces intervalles d'attente est donnée par le théorème suivant:

Théorème 9.2

T_1, T_2, \dots sont des variables aléatoires indépendantes, distribuées exponentiellement et d'espérance commune $1/\lambda$.

DÉMONSTRATION. On remarquera d'abord que l'événement $\{T_1 > t\}$ est équivalent à l'événement «il ne survient aucun événement pendant l'intervalle $[0, t]$ ». Par conséquent

$$P\{T_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}$$

La variable T_1 suit donc une distribution exponentielle d'espérance $1/\lambda$. Mais par ailleurs

$$P\{T_2 > t\} = E[P\{T_2 > t | T_1\}]$$

Or

$$\begin{aligned} P\{T_2 > t | T_1 = s\} &= P\{\text{0 événement durant } (s, s + t) | T_1 = s\} \\ &= P\{\text{0 événement durant } (s, s + t)\} \\ &= e^{-\lambda t} \end{aligned}$$

où les deux dernières inégalités résultent des hypothèses b) et c). Ce calcul permet de conclure que T_2 est également une variable exponentielle d'espérance $1/\lambda$, indépendante de T_1 qui plus est. Il suffit alors de répéter ce calcul pour établir complètement le théorème. ■

Une autre variable qui va nous servir est S_n , date à laquelle se produit le n -ième événement. On l'appelle aussi *temps d'attente* pour le n -ième événement. On voit facilement que

$$S_n = \sum_{i=1}^n T_i \quad n \geq 1$$

ce qui permet de dire, en invoquant le théorème 9.2 et les conclusions de 5.2 que S_n suit une loi gamma de paramètres n et λ . Sa densité est donc

$$f_{S_n}(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!} \quad x \geq 0$$

9.1.3 Distribution du nombre d'occurrences au temps t

Nous sommes maintenant en mesure de démontrer que $N(t)$ est une variable aléatoire de Poisson d'espérance λt .

Théorème 9.3

Pour tout processus de Poisson de paramètre λ

$$P\{N(t) = n\} = \frac{e^{-\lambda t} (\lambda t)^n}{n!}$$

DÉMONSTRATION. On remarquera que la n -ième occurrence du processus de Poisson surviendra au temps t ou avant si et seulement si le nombre d'événements survenus au temps t est de n au moins. Ceci peut être écrit

$$N(t) \geq n \Leftrightarrow S_n \leq t$$

et donc

$$\begin{aligned} P\{N(t) = n\} &= P\{N(t) \geq n\} - P\{N(t) \geq n + 1\} \\ &= P\{S_n \leq t\} - P\{S_{n+1} \leq t\} \\ &= \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!} dx - \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^n}{n!} dx \end{aligned}$$

On utilise la formule d'intégration par parties $\int u dv = uv - \int v du$ avec $u = e^{-\lambda x}$, $dv = \lambda[(\lambda x)^{n-1}/(n-1)!] dx$, ce qui donne

$$\int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^{n-1}}{(n-1)!} dx = e^{-\lambda t} \frac{(\lambda t)^n}{n!} + \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^n}{n!} dx$$

qui achève la démonstration. ■

9.2 CHAÎNES DE MARKOV

9.2.1 Définitions

Considérons une suite X_0, X_1, \dots de variables aléatoires dont l'ensemble fondamental commun est $\{0, 1, \dots, M\}$. Ce modèle peut servir à représenter l'état d'un système au cours du temps, X_n désignant cet état au temps n . Dans ce cadre, nous dirons que le système se trouve dans l'état i au temps n si $X_n = i$. La suite des variables considérées est appelée *chaîne de Markov* si à partir de tout état i la probabilité P_{ij} de passer immédiatement après à l'état j est constante au cours du temps. Plus précisément encore la condition requise s'écrit, pour tout ensemble de réels $i_0, i_1, \dots, i_{n-1}, i, j$,

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} = P_{ij}$$

Les grandeurs P_{ij} , où $0 \leq i \leq M$ et $0 \leq j \leq M$, sont appelées *probabilités de transition* de la chaîne de Markov et vérifient (dire pourquoi?) les relations suivantes:

$$P_{ij} \geq 0 \quad \sum_{j=0}^M P_{ij} = 1 \quad i = 0, 1, \dots, M$$

Il est commode de disposer ces probabilités de transition P_{ij} dans un tableau carré comme suit:

$$\begin{pmatrix} P_{00} & P_{01} & \cdots & P_{0M} \\ P_{10} & P_{11} & \cdots & P_{1M} \\ \vdots & \vdots & \ddots & \vdots \\ P_{M0} & P_{M1} & \cdots & P_{MM} \end{pmatrix}$$

Un tel tableau est appelé matrice.

La connaissance de la *matrice de transition* ainsi créée et de la distribution de X_0 permet théoriquement de calculer toutes les probabilités désirées. La densité conjointe de X_0, \dots, X_n par exemple est calculable ainsi:

$$\begin{aligned} P\{X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} \\ &= P\{X_n = i_n | X_{n-1} = i_{n-1}, \dots, X_0 = i_0\} P\{X_{n-1} = i_{n-1}, \dots, X_0 = i_0\} \\ &= P_{i_{n-1}, i_n} P\{X_{n-1} = i_{n-1}, \dots, X_0 = i_0\} \end{aligned}$$

et la répétition de cet argument finit par montrer que la probabilité ci-dessus vaut

$$= P_{i_{n-1}, i_n} P_{i_{n-2}, i_{n-1}} \cdots P_{i_1, i_2} P_{i_0, i_1} P\{X_0 = i_0\}$$

9.2.2 Exemples de chaînes de Markov

Exemple 9.1 On admet que le fait qu'il ait plu ou non un jour donné est la seule considération à prendre en compte pour prévoir s'il pleuvra le lendemain. Plus précisément, s'il pleut aujourd'hui il pleuvra demain aussi avec probabilité α et s'il ne pleut pas aujourd'hui la probabilité qu'il pleuve demain est β .

On convient de dire que le système est dans l'état 0 s'il pleut et 1 s'il ne pleut pas. La situation peut être représentée par une chaîne de Markov à deux états dont la matrice de transition est

$$\begin{pmatrix} \alpha & 1 - \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

C'est-à-dire, $P_{00} = \alpha = 1 - P_{01}$, $P_{10} = \beta = 1 - P_{11}$. ■

Exemple 9.2 On considère un joueur qui à chaque tour de jeu ne peut rencontrer que deux résultats possibles; soit il gagne 1 avec probabilité p , soit il perd 1 avec probabilité $1 - p$. On suppose que le joueur cesse de miser dès que sa fortune atteint 0 ou M . On reconnaît dans la suite des avoirs de ce joueur une chaîne de Markov ayant pour probabilités de transition

$$\begin{aligned} P_{i, i+1} &= p = 1 - P_{i, i-1} & i = 1, \dots, M-1 \\ P_{00} &= P_{MM} = 1 \end{aligned} \quad \blacksquare$$

Exemple 9.3 Les physiciens P. et T. Ehrenfest ont étudié un modèle pour le mouvement de molécules dans lequel M molécules sont réparties parmi deux urnes. Au bout

d'une unité de temps une molécule est désignée au hasard, on la retire de son urne et la place dans l'autre. On désigne le nombre de molécules contenues dans la première urne après la n -ième opération par X_n . Dans ces conditions $\{X_0, X_1, \dots\}$ est une chaîne de Markov ayant pour probabilités de transition

$$P_{i,i+1} = \frac{M-i}{M} \quad 0 \leq i \leq M$$

$$P_{i,i-1} = \frac{i}{M} \quad 0 \leq i \leq M$$

$$P_{ij} = 0 \quad \text{si } |j-i| > 1$$

■

9.2.3 Matrices de transition d'ordre supérieur

Lorsqu'on considère une chaîne de Markov, P_{ij} représente la probabilité que le système passe de l'état i à l'état j en une transition. Mais on peut définir une matrice de transition d'ordre deux, composée des coefficients $P_{ij}^{(2)}$ donnant la probabilité que le système passe de l'état i à l'état j en l'espace de deux transitions:

$$P_{ij}^{(2)} = P\{X_{m+2} = j | X_m = i\}$$

On peut calculer les coefficients $P_{ij}^{(2)}$ à partir des valeurs P_{ij} comme suit:

$$\begin{aligned} P_{ij}^{(2)} &= P\{X_2 = j | X_0 = i\} \\ &= \sum_{k=0}^M P\{X_2 = j, X_1 = k | X_0 = i\} \\ &= \sum_{k=0}^M P\{X_2 = j | X_1 = k, X_0 = i\} P\{X_1 = k | X_0 = i\} \\ &= \sum_{k=0}^M P_{kj} P_{ik} \end{aligned}$$

On définit de manière plus générale la *matrice de transition d'ordre n* , notée $P_{ij}^{(n)}$, par

$$P_{ij}^{(n)} = P\{X_{n+m} = j | X_m = i\}$$

Le théorème qui suit montre comment on peut calculer les coefficients $P_{ij}^{(n)}$.

Théorème 9.4 Equations de Chapman-Kolmogorov

$$P_{ij}^{(n)} = \sum_{k=0}^M P_{ik}^{(r)} P_{kj}^{(n-r)} \quad \text{pour tout } 0 < r < n$$

DÉMONSTRATION.

$$\begin{aligned}
 P_{ij}^{(n)} &= P\{X_n = j | X_0 = i\} \\
 &= \sum_k P\{X_n = j, X_r = k | X_0 = i\} \\
 &= \sum_k P\{X_n = j | X_r = k, X_0 = i\} P\{X_r = k | X_0 = i\} \\
 &= \sum_k P_{kj}^{(n-r)} P_{ik}^{(r)} \quad \blacksquare
 \end{aligned}$$

Exemple 9.4 Marche aléatoire

Il s'agit, sous cette dénomination connue, d'un exemple de chaîne de Markov ayant un nombre infini dénombrable d'états. Une particule se déplace le long d'un axe dans un espace unidimensionnel. Après chaque unité de temps, la particule se déplace d'un cran vers la droite ou d'un cran vers la gauche avec probabilités respectives p et $1 - p$. Ceci revient à dire que la position de la particule est représentée par une chaîne de Markov ayant pour probabilités de transition

$$P_{i,i+1} = p = 1 - P_{i,i-1} \quad i = 0, \pm 1, \dots$$

Une particule située à l'abscisse i se retrouvera à l'abscisse j au bout de n transitions avec une probabilité égale à celle que $(n - i + j)/2$ de ces transitions se fassent vers la droite et $n - [(n - i + j)/2] = (n + i - j)/2$ se fassent vers la gauche. Comme chaque transition vers la droite se fera avec probabilité p indépendamment des transitions passées, on voit que la probabilité cherchée est de type binomial:

$$P_{ij}^{(n)} = \binom{n}{\frac{n-i+j}{2}} p^{(n-i+j)/2} (1-p)^{(n+i-j)/2}$$

où $\binom{n}{x}$ vaut 0 si x n'est pas un entier inférieur à n et non négatif. Ce résultat peut être décrit ainsi:

$$\begin{aligned}
 P_{i,i+2k}^{(2n)} &= \binom{2n}{n+k} p^{n+k} (1-p)^{n-k} \quad k = 0, \pm 1, \dots, \pm n \\
 P_{i,i+2k+1}^{(2n+1)} &= \binom{2n+1}{n+k+1} p^{n+k+1} (1-p)^{n-k} \\
 &\quad k = 0, \pm 1, \dots, \pm n, -(n+1) \quad \blacksquare
 \end{aligned}$$

On remarquera que bien que les $P_{ij}^{(n)}$ soient des probabilités conditionnelles, on peut les utiliser pour le calcul de probabilités non conditionnelles en conditionnant selon l'état initial. Par exemple,

$$\begin{aligned}
 P\{X_n = j\} &= \sum_i P\{X_n = j | X_0 = i\} P\{X_0 = i\} \\
 &= \sum_i P_{ij}^{(n)} P\{X_0 = i\}
 \end{aligned}$$

9.2.4 Chaînes ergodiques

Il apparaît que pour un grand nombre de chaînes de Markov, la suite des $P_{ij}^{(n)}$ converge lorsque n tend vers l'infini vers une limite Π_j qui ne dépend que de j . En d'autres termes, la probabilité de se trouver dans l'état j après n transitions lorsque n est grand, est approximativement Π_j , indépendamment de l'état de départ. On peut montrer qu'une condition suffisante pour qu'une chaîne de Markov possède cette propriété est qu'il existe un $n >$ tel que

$$P_{ij}^{(n)} > 0 \quad \text{pour tous les } i, j = 0, 1, \dots, M \quad (9.1)$$

Les chaînes de Markov satisfaisant (9.1) sont dites *ergodiques*. Comme d'après le théorème 9.4

$$P_{ij}^{(n+1)} = \sum_{k=0}^M P_{ik}^{(n)} P_{kj}$$

il résulte que lorsque n tend vers l'infini, les chaînes ergodiques vérifient

$$\Pi_j = \sum_{k=0}^M \Pi_k P_{kj} \quad (9.2)$$

De plus, $1 = \sum_{j=0}^M P_{ij}^{(n)}$. On obtient donc également, toujours si $n \rightarrow \infty$,

$$\sum_{j=0}^M \Pi_j = 1 \quad (9.3)$$

On peut en fait montrer que les Π_j pour $0 \leq j \leq M$, sont les seules solutions non négatives de (9.2) et (9.3). Ces résultats sont rassemblés dans le théorème 9.5, cité sans démonstration.

Théorème 9.5

Pour toute chaîne de Markov ergodique,

$$\Pi_j = \lim_{n \rightarrow \infty} P_{ij}^{(n)}$$

existe et les Π_j , $0 \leq j \leq M$, sont les seules solutions non négatives de

$$\Pi_j = \sum_{k=0}^M \Pi_k P_{kj}$$

$$\sum_{j=0}^M \Pi_j = 1.$$

Exemple 9.5 Reprenons la situation présentée dans l'exemple 9.1, où l'on admet qu'il pleuvra demain avec une probabilité α ou β selon qu'il pleuve aujourd'hui ou non. Le théorème 9.5 indique que les probabilités limites de pluie et d'absence de pluie, notées Π_0 et Π_1 , seront

$$\begin{aligned}\Pi_0 &= \alpha\Pi_0 + \beta\Pi_1 \\ \Pi_1 &= (1 - \alpha)\Pi_0 + (1 - \beta)\Pi_1 \\ \Pi_0 + \Pi_1 &= 1\end{aligned}$$

ce qui donne

$$\Pi_0 = \frac{\beta}{1 + \beta - \alpha} \quad \Pi_1 = \frac{1 - \alpha}{1 + \beta - \alpha}$$

Si par exemple $\alpha = 0,6$ et $\beta = 0,3$, la probabilité limite qu'il pleuve le n -ième jour, n tendant vers l'infini, est $\Pi_0 = \frac{3}{7}$. ■

La quantité Π_j représente aussi la proportion de fois, à long terme, où la chaîne de Markov reste dans l'état j , $j = 0, \dots, M$. Pour le comprendre intuitivement, notons par P_j la proportion de fois où la chaîne se trouve dans l'état j . (La loi forte des grands nombres permet de montrer que les proportions ainsi définies existent et sont constantes.) Or, comme la proportion de fois, à long terme, où la chaîne reste dans l'état k est P_k et que, de l'état k elle passe à l'état j avec probabilité P_{kj} , il s'ensuit que la proportion de fois où la chaîne de Markov entre dans l'état j à partir de l'état k est égale à $P_k P_{kj}$. En sommant sur k , on trouve que P_j , la proportion de fois où la chaîne de Markov passe à l'état j , satisfait

$$P_j = \sum_k P_k P_{kj}$$

Comme il est vrai aussi que

$$\sum_j P_j = 1$$

et que les Π_j , $j = 0, \dots, M$ sont l'unique solution des équations précédentes d'après le théorème 9.5, on conclut que $P_j = \Pi_j$, $j = 0, \dots, M$.

Exemple 9.6 Dans l'exemple 9.3, intéressons-nous à la proportion de fois où l'urne 1 contient j molécules, $j = 0, \dots, M$. On sait, d'après le théorème 9.4 que ces quantités ainsi définies constituent l'unique solution du système suivant:

$$\Pi_0 = \Pi_1 \times \frac{1}{M}$$

$$\Pi_j = \Pi_{j-1} \times \frac{M - j + 1}{M} + \Pi_{j+1} \times \frac{j + 1}{M}, \quad j = 1, \dots, M$$

$$\Pi_M = \Pi_{M-1} \times \frac{1}{M}$$

$$\sum_{j=0}^M \Pi_j = 1$$

Or, comme on peut le vérifier facilement, les quantités

$$\Pi_j = \binom{M}{j} \left(\frac{1}{2}\right)^M, \quad j = 0, \dots, M$$

satisfont les équations précédentes, elles représentent donc les proportions de fois où la chaîne de Markov reste dans chacun des états. (Le problème 9.6.11 indique comment on aurait pu deviner cette solution.) ■

9.3 SURPRISE, INCERTITUDE, ENTROPIE

9.3.1 Définition formelle de la surprise

Considérons E , l'un des événements pouvant survenir à la suite d'une expérience. A quel point serions-nous surpris d'apprendre que E a effectivement eu lieu? Il semble que la réponse à cette question doit dépendre de la probabilité de E : si par exemple l'expérience consiste à jeter deux dés, il ne nous semblerait pas spécialement étonnant d'apprendre que la somme des dés est paire (événement de probabilité $\frac{1}{2}$); nous serions par contre plus étonnés d'apprendre que cette somme vaut douze puisque la probabilité en est $\frac{1}{36}$.

Nous allons nous attacher dans ce qui suit à quantifier la notion de surprise. Pour commencer, admettons que la surprise que nous éprouverons consécutivement à la réalisation d'un événement E ne dépendra que de la probabilité de E . On notera $S(p)$ la surprise créée par la réalisation de tout événement de probabilité p . Nous allons essayer de déterminer la forme de S en nous fixant préalablement un ensemble de conditions intuitivement acceptables que S devra satisfaire. Les propriétés de S ne devront être déterminées que par ces conditions; tout au long de ce qui suit on considérera S définie pour toute valeur de p comprise entre 0 (exclusivement) et 1 (inclusivement): la fonction S ne sera pas définie pour des événements de probabilité nulle.

Notre première condition n'est que la traduction du fait intuitif qu'il n'y aura pas de surprise à apprendre qu'un événement certain est effectivement arrivé.

Axiome 9.6

$$S(1) = 0$$

Notre seconde condition sera que plus un événement est improbable, plus grande sera notre surprise.

Axiome 9.7

S est une fonction strictement décroissante de p , c'est-à-dire que si $p < q$ alors $S(p) > S(q)$.

La troisième condition est l'expression mathématique de l'opinion qu'une petite variation de p devrait n'entraîner qu'une faible variation de $S(p)$.

Axiome 9.8

S est une fonction continue.

Pour étayer la dernière condition, considérons deux événements indépendants E et F de probabilités respectives $P(E) = p$ et $P(F) = q$. Comme $P(EF) = pq$, la surprise

correspondant à l'apparition simultanée de E et F est $S(pq)$. Supposons maintenant que l'on apprenne d'abord que E est survenu, puis plus tard que F est survenu lui aussi. La surprise créée par la réalisation de E étant $S(p)$, on peut dire que $S(pq) - S(p)$ représente la surprise additionnelle créée par le fait que F survienne. Mais F étant indépendant de E , l'information sur E ne change pas la probabilité de F ; la surprise additionnelle doit donc être $S(q)$ simplement. Ce raisonnement suggère la dernière condition.

Axiome 9.9

$$S(pq) = S(p) + S(q) \quad 0 < p \leq 1, 0 < q \leq 1$$

Nous avons maintenant les éléments nécessaires pour énoncer le théorème 9.10 qui va nous livrer l'expression analytique de S .

Théorème 9.10

Si S satisfait les axiomes 9.6 à 9.9, alors

$$S(p) = -C \log_2 p$$

où C est un entier positif quelconque.

DÉMONSTRATION. De l'axiome 9.9 il résulte que

$$S(p^2) = S(p) + S(p) = 2S(p)$$

et par induction

$$S(p^m) = mS(p) \tag{9.4}$$

Ceci permet d'écrire que pour tout entier n

$$S(p) = S(p^{1/n} \cdots p^{1/n}) = nS(p^{1/n})$$

ce qui entraîne à son tour que

$$S(p^{1/n}) = \frac{1}{n} S(p) \tag{9.5}$$

De (9.4) et (9.5) on tire

$$\begin{aligned} S(p^{m/n}) &= mS(p^{1/n}) \\ &= \frac{m}{n} S(p) \end{aligned}$$

qui est équivalent à

$$S(p^x) = xS(p) \tag{9.6}$$

pour autant que x soit rationnel positif. En vertu de l'axiome 9.8 (continuité de S), la relation (9.6) reste vraie pour toute valeur non négative de x (le lecteur peut l'établir).

Posons alors, pour tout réel p tel que $0 < p \leq 1$, la relation $x = -\log_2 p$. Ceci équivaut à $p = (\frac{1}{2})^x$ et de (9.6) on tire finalement

$$S(p) = S((\frac{1}{2})^x) = xS(\frac{1}{2}) = -C \log_2 p$$

où $C = S(\frac{1}{2}) > S(1) = 0$ en vertu des axiomes 9.7 et 9.6. ■

Il est habituel de poser $C = 1$ et c'est ce que nous ferons. La surprise est alors mesurée en *bits* (abréviation des mots anglais «binary digits»).

9.3.2 Surprise, incertitude, information: synthèse

Considérons une variable aléatoire X qui prend ses valeurs dans l'ensemble $\{x_1, x_2, \dots, x_n\}$ avec les probabilités correspondantes p_1, p_2, \dots, p_n . La grandeur $-\log p_i$ représentant¹ la mesure de la surprise associée à l'événement $\{X = x_i\}$, la surprise moyenne créée lorsqu'on apprend quelle valeur X a prise est

$$H(X) = -\sum_{i=1}^n p_i \log p_i$$

En théorie de l'information cette quantité $H(X)$ est appelée *entropie* de la variable X (par convention, si l'un des p_i vaut 0 on déclare que $0 \log 0$ vaut 0). On peut montrer (et cela est laissé à titre d'exercice) que $H(X)$ est maximale lorsque tous les p_i sont égaux (est-ce intuitivement acceptable?).

La grandeur $H(X)$ représentant la surprise moyenne associée à la découverte de la valeur de X , on peut aussi l'assimiler à la quantité d'incertitude relative à cette variable aléatoire. La théorie de l'information considère $H(X)$ comme l'*information* liée à l'observation de X . On peut donc dire que la surprise moyenne causée par X , l'incertitude liée à X , l'information moyenne recelée par X , représentent en fait le même concept abordé par des points de vue un peu différents.

9.3.3 Cas de variables conjointes

Considérons deux variables aléatoires X et Y prenant leurs valeurs dans respectivement $\{x_1, x_2, \dots, x_n\}$ et $\{y_1, y_2, \dots, y_m\}$, leur loi de probabilité conjointe étant

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\}$$

On calcule grâce à cette loi conjointe l'incertitude $H(X, Y)$ attachée au vecteur aléatoire (X, Y) :

$$H(X, Y) = -\sum_i \sum_j p(x_i, y_j) \log p(x_i, y_j)$$

¹ Dans le reste de ce chapitre nous écrirons $\log x$ pour $\log_2 x$. On continuera d'écrire $\ln x$ pour désigner $\log_e x$.

Admettons maintenant que l'on ait observé que $Y = y_j$. La quantité restante d'incertitude sur X est dès lors

$$H_{Y=y_j}(X) = -\sum_i p(x_i | y_j) \log p(x_i | y_j)$$

où

$$p(x_i | y_j) = P\{X = x_i | Y = y_j\}$$

Aussi écrira-t-on comme suit l'incertitude moyenne associée à X après observation de Y :

$$H_Y(X) = \sum_j H_{Y=y_j}(X) p_Y(y_j)$$

où

$$p_Y(y_j) = P\{Y = y_j\}$$

Le théorème 9.11 se propose de mettre $H(X, Y)$, $H(Y)$ et $H_Y(X)$ en relation. Il établit que l'incertitude associée à X et Y est égale à celle associée à Y plus celle restant sur X une fois que Y a été observée.

Théorème 9.11

$$H(X, Y) = H(Y) + H_Y(X)$$

DÉMONSTRATION. On utilise l'identité $p(x_i, y_j) = p_Y(y_j)p(x_i | y_j)$ qui entraîne

$$\begin{aligned} H(X, Y) &= -\sum_i \sum_j p(x_i, y_j) \log p(x_i, y_j) \\ &= -\sum_i \sum_j p_Y(y_j) p(x_i | y_j) [\log p_Y(y_j) + \log p(x_i | y_j)] \\ &= -\sum_j p_Y(y_j) \log p_Y(y_j) \sum_i p(x_i | y_j) \\ &\quad - \sum_j p_Y(y_j) \sum_i p(x_i | y_j) \log p(x_i | y_j) \\ &= H(Y) + H_Y(X) \quad \blacksquare \end{aligned}$$

Un résultat fondamental de la théorie de l'information est que l'incertitude associée à une variable aléatoire X diminuera, en moyenne, si l'on peut observer une autre variable Y . Avant de démontrer cela on notera que pour tout réel positif x ,

$$\ln x \leq x - 1 \quad x > 0 \quad (9.7)$$

et l'égalité n'a lieu que lorsque $x = 1$. La démonstration est laissée en exercice.

Théorème 9.12

$$H_Y(X) \leq H(X)$$

et l'égalité a lieu si et seulement si X et Y sont indépendantes.

DÉMONSTRATION.

$$\begin{aligned} H_Y(X) - H(X) &= -\sum_i \sum_j p(x_i | y_j) \log [p(x_i | y_j)] p(y_j) \\ &\quad + \sum_i \sum_j p(x_i, y_j) \log p(x_i) \\ &= \sum_i \sum_j p(x_i, y_j) \log \left[\frac{p(x_i)}{p(x_i | y_j)} \right] \\ &\leq \log e \sum_i \sum_j p(x_i, y_j) \left[\frac{p(x_i)}{p(x_i | y_j)} - 1 \right] \quad \text{en application de (9.7)} \\ &= \log e \left[\sum_i \sum_j p(x_i) p(y_j) - \sum_i \sum_j p(x_i, y_j) \right] \\ &= \log e [1 - 1] \\ &= 0 \end{aligned}$$

9.4 THÉORIE DU CODAGE ET ENTROPIE

9.4.1 Efficacité d'un code

Imaginons qu'il faille transmettre la valeur prise par une variable discrète X d'un point A , où on l'observe, jusqu'à un point B , ceci grâce à un système de communication ne véhiculant que des 0 ou des 1. La première chose à faire est de coder les diverses valeurs que X peut prendre pour les transformer en chaînes de 0 et de 1. Pour éviter toute ambiguïté on exige normalement qu'il soit impossible de former un code simplement en ajoutant quelque chose à un autre.

A titre d'exemple supposons que X puisse prendre l'une des quatre valeurs x_1, x_2, x_3, x_4 . Un des codes auxquels on peut penser serait:

$$\begin{aligned} x_1 &\leftrightarrow 00 \\ x_2 &\leftrightarrow 01 \\ x_3 &\leftrightarrow 10 \\ x_4 &\leftrightarrow 11 \end{aligned} \tag{9.8}$$

Ceci signifie que lorsque $X = x_1$, le message envoyé en B est 00, qu'il est 01 lorsque $X = x_2$, etc. Mais on peut aussi coder ainsi:

$$\begin{aligned} x_1 &\leftrightarrow 0 \\ x_2 &\leftrightarrow 10 \\ x_3 &\leftrightarrow 110 \\ x_4 &\leftrightarrow 111 \end{aligned} \tag{9.9}$$

Par contre, le codage

$$\begin{aligned}x_1 &\leftrightarrow 0 \\x_2 &\leftrightarrow 1 \\x_3 &\leftrightarrow 00 \\x_4 &\leftrightarrow 01\end{aligned}$$

n'est pas admis car les codes correspondant à x_3 et x_4 sont des extensions du code de x_1 .

L'un des objectifs ordinairement assignés au codage est de minimiser le nombre moyen de bits (c'est-à-dire de chiffres binaires) qui seront nécessaires pour transmettre le message de A à B . Supposons par exemple que

$$\begin{aligned}P\{X = x_1\} &= \frac{1}{2} \\P\{X = x_2\} &= \frac{1}{4} \\P\{X = x_3\} &= \frac{1}{8} \\P\{X = x_4\} &= \frac{1}{8}\end{aligned}$$

Le code donné par (9.9) utilisera en moyenne $\frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{8}(3) + \frac{1}{8}(3) = 1,75$ bits; mais le code donné par (9.8) aura en moyenne besoin de 2 bits. Aussi dit-on, pour cette distribution de X , que ce dernier est moins efficace que l'autre.

9.4.2 Efficacité maximale d'un code sans bruit

Les considérations qui précèdent soulèvent la question suivante: pour une variable X donnée, quelle est l'efficacité maximale que puisse présenter un codage? La réponse est que pour tout codage il faudra transmettre un nombre de bits moyen au moins égal à l'entropie de X . Nous allons démontrer ce résultat, connu en théorie de l'information sous le nom de théorème du codage sans bruit, en deux phases.

Théorème 9.13

Soient X une variable aléatoire et $\{x_1, x_2, \dots, x_N\}$ l'ensemble des valeurs qu'elle peut prendre. Pour que des chaînes binaires de longueurs respectives n_1, \dots, n_N constituent un codage des valeurs de X (sans que certains codes ne constituent des extensions d'autres codes), il faut et il suffit que

$$\sum_{i=1}^N \left(\frac{1}{2}\right)^{n_i} \leq 1$$

DÉMONSTRATION. Soient N entiers positifs notés n_1, \dots, n_N . Désignons par w_j le nombre des n_i valant j , $j = 1, 2, \dots$. Il est évident que si nous voulons obtenir un codage exprimant pour tout $i = 1, \dots, N$ la valeur x_i en n_i bits, il est nécessaire que $w_1 \leq 2$. Comme de plus il est exclu qu'une chaîne binaire constitue une extension d'une autre, on aura $w_2 \leq 2^2 - 2w_1$ (ceci résulte du fait qu'au nombre 2^2 de chaînes de longueur 2 il faut soustraire le nombre $2w_1$ des séquences qui ne sont qu'une extension des chaînes de longueur 1 déjà utilisées). Le même raisonnement permet d'établir de manière générale que

$$w_n \leq 2^n - w_1 2^{n-1} - w_2 2^{n-2} - \dots - w_{n-1} 2 \quad (9.10)$$

pour $n = 1, \dots$. Le lecteur peut en fait se convaincre en réfléchissant tant soit peu que ces conditions sont également suffisantes pour qu'il existe un codage traduisant x_i en une chaîne de longueur n_i , $i = 1, \dots, N$. On peut réécrire (9.10):

$$w_n + w_{n-1} 2 + w_{n-2} 2^2 + \dots + w_1 2^{n-1} \leq 2^n \quad n = 1, \dots$$

Une division par 2^n livre une nouvelle expression de nos conditions nécessaires et suffisantes:

$$\sum_{j=1}^n w_j \left(\frac{1}{2}\right)^j \leq 1 \quad \text{pour tout } n \quad (9.11)$$

La suite des termes $\sum_{j=1}^n w_j \left(\frac{1}{2}\right)^j$ étant croissante et bornée par 1, il en résulte que (9.11) équivaut à:

$$\sum_{j=1}^{\infty} w_j \left(\frac{1}{2}\right)^j \leq 1$$

Le résultat est ainsi acquis puisque par définition des w_j , qui représentent le nombre des n_i de longueur j ,

$$\sum_{j=1}^{\infty} w_j \left(\frac{1}{2}\right)^j = \sum_{i=1}^N \left(\frac{1}{2}\right)^{n_i}$$

Nous sommes maintenant en mesure de démontrer le théorème principal.

Théorème 9.14 *Théorème du codage sans bruit*

Soit X une variable aléatoire pouvant prendre les valeurs x_1, \dots, x_N avec pour probabilités respectives $p(x_1), \dots, p(x_N)$. Tout codage de X traduisant x_i grâce à n_i bits utilisera un nombre moyen de bits supérieur ou égal à l'entropie de X :

$$\sum_{i=1}^N n_i p(x_i) \geq H(X) = - \sum_{i=1}^N p(x_i) \log p(x_i)$$

DÉMONSTRATION. POSONS $P_i = p(x_i)$, $q_i = 2^{-n_i} / \sum_{j=1}^N 2^{-n_j}$, $i = 1, \dots, N$. Alors

$$\begin{aligned} - \sum_{i=1}^N P_i \log \left(\frac{P_i}{q_i} \right) &= - \log e \sum_{i=1}^N P_i \ln \left(\frac{P_i}{q_i} \right) \\ &= \log e \sum_{i=1}^N P_i \ln \left(\frac{q_i}{P_i} \right) \\ &\leq \log e \sum_{i=1}^N P_i \left(\frac{q_i}{P_i} - 1 \right) \quad \text{en vertu de (9.7)} \\ &= 0 \quad \text{puisque } \sum_{i=1}^N P_i = \sum_{i=1}^N q_i = 1 \end{aligned}$$

Donc

$$\begin{aligned}
 -\sum_{i=1}^N P_i \log P_i &\leq -\sum_{i=1}^N P_i \log q_i \\
 &= \sum_{i=1}^N n_i P_i + \log \left(\sum_{j=1}^N 2^{-n_j} \right) \\
 &\leq \sum_{i=1}^N n_i P_i \quad \text{en vertu de théorème 9.13} \quad \blacksquare
 \end{aligned}$$

Exemple 9.7 On considère la variable aléatoire X de loi

$$p(x_1) = \frac{1}{2} \quad p(x_2) = \frac{1}{4} \quad p(x_3) = p(x_4) = \frac{1}{8}$$

On sait que

$$\begin{aligned}
 H(X) &= -\left[\frac{1}{2} \log \frac{1}{2} + \frac{1}{4} \log \frac{1}{4} + \frac{1}{4} \log \frac{1}{8} \right] \\
 &= \frac{1}{2} + \frac{2}{4} + \frac{3}{4} \\
 &= 1.75
 \end{aligned}$$

Il résulte alors du théorème 9.14 qu'il n'existe pas de codage plus efficace que

$$\begin{aligned}
 x_1 &\leftrightarrow 0 \\
 x_2 &\leftrightarrow 10 \\
 x_3 &\leftrightarrow 110 \\
 x_4 &\leftrightarrow 111
 \end{aligned} \quad \blacksquare$$

9.4.3 Borne pour codage non optimal

Dans la plupart des cas il n'existe pas de codage dont l'efficacité atteint la borne correspondant à $H(X)$. Il est cependant toujours possible de créer un codage tel que le nombre moyen de bits utilisé s'approche de $H(X)$ à moins d'une unité. Pour l'établir, il suffit de considérer les entiers n_i satisfaisant

$$-\log p(x_i) \leq n_i \leq -\log p(x_i) + 1$$

pour $i = 1, \dots, N$. La première inégalité entraîne

$$\sum_{i=1}^N 2^{-n_i} \leq \sum_{i=1}^N 2^{\log p(x_i)} = \sum_{i=1}^N p(x_i) = 1$$

et donc, en vertu du théorème 9.13, il existe un codage des valeurs x_i de X par des chaînes de longueur n_i , $i = 1, \dots, N$. La longueur moyenne de ces chaînes est

$$L = \sum_{i=1}^N n_i p(x_i)$$

satisfaisant

$$-\sum_{i=1}^N p(x_i) \log p(x_i) \leq L \leq -\sum_{i=1}^N p(x_i) \log p(x_i) + 1$$

ou enfin

$$H(X) \leq L \leq H(X) + 1$$

Exemple 9.8 On jette dix fois une pièce tombant sur face avec probabilité p . Les jets sont indépendants. On désire transmettre le résultat en un point B , alors que l'expérience est réalisée en un autre point A . Le résultat de cette expérience est donc un vecteur aléatoire $X = (X_1, \dots, X_{10})$ où $X_i = 1$ ou 0 selon que le jet correspondant a ou n'a pas donné face. Les résultats précédents établissent que le nombre moyen L de bits nécessaires à la transmission satisfera pour tout codage

$$H(X) \leq L$$

et qu'on aura pour un codage au moins

$$L \leq H(X) + 1$$

Or les X_i sont indépendantes et il résulte donc des théorèmes 9.11 et 9.12 que

$$\begin{aligned} H(X) &= H(X_1, \dots, X_n) = \sum_{i=1}^n H(X_i) \\ &= -10[p \log p + (1-p) \log (1-p)] \end{aligned}$$

Pour $p = \frac{1}{2}$ on trouve $H(X) = 10$. Dans ce cas on ne peut pas faire mieux que coder X grâce à ses composantes. Si par exemple les cinq premiers jets donnent face et les cinq derniers pile, il suffit de transmettre en B le message 1111100000.

Si par contre $p \neq \frac{1}{2}$ il sera souvent possible de faire mieux que ce dernier codage. Pour $p = \frac{1}{4}$ par exemple,

$$H(X) = -10\left(\frac{1}{4} \log \frac{1}{4} + \frac{3}{4} \log \frac{3}{4}\right) = 8.11$$

et il existe donc un codage dont la longueur moyenne ne dépassera pas 9,11.

Un codage simple qui serait ici plus efficace que le codage trivial consiste à décomposer d'abord le vecteur (X_1, \dots, X_{10}) en cinq paires de variables puis à convertir chaque paire comme suit:

$$\begin{aligned} X_i = 0, X_{i+1} = 0 &\leftrightarrow 0 \\ X_i = 0, X_{i+1} = 1 &\leftrightarrow 10 \\ X_i = 1, X_{i+1} = 0 &\leftrightarrow 110 \\ X_i = 1, X_{i+1} = 1 &\leftrightarrow 111 \end{aligned}$$

où $i = 1, 3, 5, 7, 9$. Le message total est constitué des messages partiels mis bout à bout. Si par exemple on observe $P P P F F P P P P F$ il faudra émettre 010110010. La longueur moyenne du message en bits est pour ce codage

$$5\left[1\left(\frac{3}{4}\right)^2 + 2\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) + 3\left(\frac{1}{4}\right)\left(\frac{3}{4}\right) + 3\left(\frac{1}{4}\right)^2\right] = \frac{135}{16} \approx 8.44 \quad \blacksquare$$

9.4.4 Codage en cas de bruit

Nous avons jusqu'ici supposé que le message émis en A était reçu sans erreur en B . Cependant, une erreur peut toujours résulter des perturbations aléatoires agissant sur le canal de communication. Il peut arriver qu'à cause d'une telle perturbation un message émis sous la forme 00101101 en A parvienne en B sous la forme 01101101.

Supposons qu'un bit émis en A soit correctement enregistré en B avec probabilité p , et ce indépendamment d'un bit à l'autre. Un tel système de communication est appelé *canal binaire symétrique*. Admettons encore que $p = 0,8$ et qu'on veuille transmettre un message constitué d'une grande quantité de bits de A à B . La transmission du message tel quel entraînera une probabilité d'erreur de 0,2 par bit, ce qui est plutôt fort. Un moyen de réduire ce taux d'erreur consiste à émettre trois fois consécutivement chaque bit du message puis à décoder en appliquant le système majoritaire. Cela revient en d'autres termes à utiliser le codage

<i>Encodage</i>	<i>Décodage</i>
$0 \rightarrow 000$	$\left. \begin{array}{l} 000 \\ 001 \\ 010 \\ 100 \end{array} \right\} \rightarrow 0$
$1 \rightarrow 111$	$\left. \begin{array}{l} 111 \\ 110 \\ 101 \\ 011 \end{array} \right\} \rightarrow 1$

On remarquera que tant qu'une erreur au plus se produit lors de la transmission, le décodage restitue la valeur correcte. La probabilité d'erreur par bit est donc ramenée à

$$(0.2)^3 + 3(0.2)^2(0.8) = 0.104$$

ce qui constitue une amélioration notable. Il est évident qu'on peut rendre le taux d'erreur aussi petit que l'on veut en répétant le bit à transmettre de nombreuses fois puis en décodant selon le système majoritaire. Le codage suivant, par exemple réduirait le taux d'erreur par bit à moins de 0,01:

Encodage	Décodage
$0 \rightarrow$ chaîne de 17 «0»	à la majorité
$1 \rightarrow$ chaîne de 17 «1»	

Le problème avec ce type de codage est que, tout en réduisant le taux d'erreur, on réduit aussi le nombre de bits significatifs par signal (voir tableau 9.1).

Tableau 9.1 Codage par répétition des bits

Probabilité d'erreur de transmission par bit	Débit des bits significatifs
0.20	1
0.10	0.33 (= $\frac{1}{3}$)
0.01	0.06 (= $\frac{1}{17}$)

A ce point de l'exposé, il apparaîtra peut-être au lecteur inévitable que l'abaissement de la probabilité d'erreur par bit vers 0 s'accompagne toujours d'un abaissement du débit effectif vers 0 également. Or, et c'est un résultat remarquable de la théorie de l'information, ce n'est pas le cas. Ce résultat, connu sous le nom de théorème du codage avec bruit, est dû à Claude Shannon. En voici l'énoncé.

Théorème 9.15 *Théorème du codage avec bruit*

Il existe un nombre C tel que pour tout $R < C$ et pour tout $\varepsilon > 0$ il y ait un système de codage et décodage transmettant au taux moyen de R bits par signal et avec une probabilité d'erreur par bit inférieure à ε . La plus grande valeur¹ que l'on puisse donner à C , notée C^ , est appelée capacité du canal et vaut, pour les canaux binaires symétriques*

$$C^* = 1 + p \log p + (1 - p) \log (1 - p)$$

9.5 EXERCICES THÉORIQUES ET PROBLÈMES

9.5.1 Des clients arrivent dans une banque à un rythme poissonien de taux λ . Supposons que deux clients arrivent durant la première heure. Quelle est la probabilité que

- les deux soient arrivés durant les 20 premières minutes?
- L'un au moins soit arrivé pendant les 20 premières minutes?

9.5.2 Sur une autoroute les voitures franchissent une ligne transversale à un rythme décrit par un processus poissonien de taux par minute $\lambda = 3$. Joe traverse l'autoroute sans regarder. Quelle est alors la probabilité qu'il ne soit pas blessé s'il met s secondes pour traverser? (on suppose que s'il se trouve sur la route alors qu'une voiture passe, il sera blessé). Prendre $s = 2, 5, 10, 20$.

9.5.3 Supposer, dans le scénario du problème 9.5.2, que Joe est assez agile pour échapper à une seule voiture, mais que s'il rencontre 2 ou plusieurs voitures en

¹ Voir au problème 9.6.18 une interprétation de C^* en termes d'entropie.

essayant de traverser la route, il sera blessé. Quelle est la probabilité qu'il ne soit pas blessé s'il met s secondes pour traverser? Prendre $s = 5, 10, 20, 30$.

9.5.4 On suppose que 3 boules blanches et 3 boules noires sont réparties dans deux urnes de façon que chacune de celles-ci contienne trois boules. On dira que le système est dans l'état i si la première urne contient i boules blanches, $i = 0, 1, 2, 3$. A chaque étape une boule est tirée de chaque urne, puis la boule tirée de la première urne est placée dans la deuxième et inversement. Soit X_n l'état du système après la n -ième étape; calculer les probabilités de transition de la chaîne de Markov $\{X_n, n \geq 0\}$.

9.5.5 Considérer l'exemple 9.1. S'il y a 50 chances sur 100 qu'il pleuve aujourd'hui, calculer la probabilité qu'il pleuve pendant 3 jours à partir d'aujourd'hui lorsque $\alpha = 0,7$ et $\beta = 0,3$.

9.5.6 Calculer les probabilités limites correspondant au modèle du problème 9.5.4.

9.5.7 Une matrice de probabilités de transition est dite doublement stochastique si

$$\sum_{i=0}^M P_{ij} = 1$$

pour tous les états $j = 0, 1, \dots, M$. Si une telle chaîne de Markov est ergodique, montrer que $\Pi_j = 1/(M + 1)$, $j = 0, 1, \dots, M$.

9.5.8 Un jour donné, Buffy est soit de bonne humeur (b), soit comme-ci comme-ça (c), soit mélancolique (m). Si elle est de bonne humeur aujourd'hui, elle sera b , c ou m demain avec probabilité 0,7, 0,2, 0,1 respectivement. Si elle est comme-ci comme-ça aujourd'hui, elle sera b , c ou m demain avec probabilité 0,4, 0,3, 0,3 respectivement. Si elle est mélancolique aujourd'hui, elle sera b , c ou m demain avec probabilité 0,2, 0,4, 0,4 respectivement. Quelle proportion de temps Buffy est-elle de bonne humeur?

9.5.9 Supposons que le fait qu'il pleuve ou non demain ne dépende que des conditions météorologiques des deux derniers jours. Plus précisément, supposons que s'il a plu hier et aujourd'hui, il pleuvra demain avec probabilité 0,8; s'il a plu aujourd'hui mais pas hier, il pleuvra demain avec probabilité 0,4 et s'il n'a plu ni hier ni aujourd'hui, il pleuvra demain avec probabilité 0,2. Quel est le pourcentage de jours pluvieux?

9.5.10 Un homme fait une promenade tous les matins. Lorsqu'il quitte la maison pour sa promenade, il emprunte avec équiprobabilité la porte de devant ou celle de derrière et de même, au retour, il rentre avec équiprobabilité par la porte de devant ou de derrière. Le promeneur possède 5 paires de chaussures de marche qu'il ôte, après la promenade, devant la porte qu'il emprunte pour rentrer. S'il ne trouve pas de chaussures devant la porte, en sortant de chez lui, il se promènera pieds nus. On veut déterminer le pourcentage de promenades qu'il effectue pieds nus.

- Représenter cette situation par une chaîne de Markov. Donner les états et les probabilités de transition.
- Déterminer le pourcentage de promenades qu'il effectue pieds nus.

9.5.11 Reprenez l'exemple 9.6.

- Vérifiez que la valeur proposée des Π_j satisfait les relations nécessaires.
- Pour une molécule quelconque, quelle est, à votre avis, la probabilité limite qu'elle se trouve dans l'urne 1.
- Pensez-vous que les événements «la molécule j est dans l'urne 1 après un long moment», $j \geq 1$, sont (à l'infini) indépendants?
- Expliquer comment on obtient ces probabilités limites?

9.5.12 Déterminer l'entropie de la somme obtenue lors du jet d'une paire de dés non pipés.

9.5.13 Soit X une variable pouvant prendre n valeurs différentes avec pour probabilités respectives P_1, \dots, P_n . Montrer que $H(X)$ est maximale lorsque $P_i = 1/n$, $i = 1, \dots, n$. Que vaut $H(X)$ dans ce cas?

9.5.14 Une paire de dés non pipés est lancée. Soit

$$X = \begin{cases} 1 & \text{si la somme est } 6 \\ 0 & \text{sinon} \end{cases}$$

et soit Y la valeur montrée par le premier dé. Calculer $H(Y)$, $H_Y(X)$ et $H(X, Y)$.

9.5.15 Une pièce ayant une probabilité $p = \frac{2}{3}$ de tomber sur face est lancée 6 fois. Calculer l'entropie du résultat de l'expérience.

9.5.16 Une variable aléatoire peut prendre les valeurs x_1, \dots, x_n avec probabilités respectives $p(x_i)$, $i = 1, \dots, n$. On essaie de déterminer la valeur de X en posant une série de questions, les seules réponses possibles étant oui ou non. Par exemple, on peut demander « X est-elle égale à x_1 ?» ou « X est-elle égale à x_1 ou x_2 ou x_3 ?», et ainsi de suite. Que pouvez-vous dire sur le nombre moyen de questions nécessaires pour déterminer la valeur de X ?

9.5.17 Montrer que pour toute variable aléatoire discrète X et toute fonction f on a

$$H(f(X)) \leq H(X)$$

9.5.18 Lors de la transmission d'un bit d'un point A à un point B , si on désigne par X la valeur du bit émis de A puis par Y la valeur reçue en B , alors $H(X) - H_Y(X)$ est appelé taux de transmission de l'information de A à B . Le taux maximal de transmission, en tant que fonction de $P\{X = 1\} = 1 - P\{X = 0\}$, est appelé *capacité* du canal. Montrer que pour un canal binaire symétrique avec $P\{Y = 1|X = 1\} = P\{Y = 0|X = 0\} = p$, la capacité du canal est atteinte par le taux de transmission de l'information quand $P\{X = 1\} = \frac{1}{2}$ et sa valeur est $1 + p \log p + (1 - p) \log (1 - p)$.

9.6 RÉFÉRENCES

Les références suivantes concernent les paragraphes 9.1 et 9.2

- Kemeny, J., L. Snell, and A. Knapp. *Denumerable Markov Chains*. New York: D. Van Nostrand Company, 1966.
- Parzen, E. *Stochastic Processes*. San Francisco: Holden-Day, Inc., 1962.
- Ross, S. M. *Introduction to Probability Models*, 3rd ed. New York: Academic Press, Inc., 1984.
- Ross, S. M. *Stochastic Processes*. New York: John Wiley & Sons, Inc., 1983.

Les références suivantes concernent les paragraphes 9.3 et 9.4

- Abramson, N. *Information Theory and Coding*. New York: McGraw-Hill Book Company, 1963.
- McElice, R. *Theory of Information and Coding*. Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1977.
- Peterson, W., and E. Weldon. *Error Correcting Codes*, 2nd ed. Cambridge, Mass.: The M.I.T. Press, 1972.

CHAPITRE 10

Simulation

10.1 INTRODUCTION

10.1.1 Exemple de situations à simuler

Comment peut-on déterminer la probabilité de gagner une partie de solitaire? (nous entendons par solitaire n'importe quelle version connue du jeu utilisant un paquet de 52 cartes normales et basée sur une stratégie fixe). Une méthode consiste à admettre l'hypothèse raisonnable d'équiprobabilité des $(52)!$ permutations possibles des cartes, puis à déterminer combien parmi celles-ci sont gagnantes. Il ne semble malheureusement pas aisé de mettre au point un critère systématique discriminant les permutations gagnantes: $(52)!$ est un nombre fort grand et il semble qu'il n'y ait pas d'autre moyen de savoir si une permutation mène à une réussite que de jouer la partie. Cette approche ne conviendra pas.

Il apparaît à ce point que la détermination de la probabilité de gagner échappe au traitement mathématique. Cependant les choses ne s'arrêtent pas là car les probabilités sont du domaine des sciences appliquées comme de celui des mathématiques; or dans toutes les sciences appliquées l'expérimentation est fort utile. Dans le cas de notre partie de solitaire, par exemple, l'expérimentation revient à exécuter un grand nombre de parties ou mieux encore à programmer un ordinateur pour qu'il le fasse. Après l'exécution de, disons, n parties on pourra poser

$$X_i = \begin{cases} 1 & \text{si le } i\text{-ème jeu est une victoire} \\ 0 & \text{sinon} \end{cases}$$

Les variables X_i , $i = 1, \dots, n$ seront alors des variables de Bernoulli pour lesquelles

$$E[X_i] = P\{\text{gagner une partie}\}$$

D'après la loi forte des grands nombres, nous saurons que

$$\sum_{i=1}^n \frac{X_i}{n} = \frac{\text{nombre de parties gagnées}}{\text{nombre de parties jouées}}$$

tendra avec probabilité 1 vers la probabilité de gagner une partie. Ou encore, on peut dire qu'après un grand nombre de parties on peut utiliser la proportion de parties gagnées sur le nombre total de parties pour obtenir une estimation de la probabilité cherchée. La méthode consistant à déterminer des probabilités de manière empirique à travers l'expérimentation est appelée *simulation*.

10.1.2 Procédé de simulation

Dans le but d'utiliser un ordinateur pour mener une étude de simulation, on doit pouvoir générer les valeurs d'une variable aléatoire uniforme sur $(0, 1)$; ces valeurs sont appelées des *nombres aléatoires*. Pour générer ces nombres, la plupart des ordinateurs possèdent une fonction prédéfinie appelée *générateur de nombres aléatoires*, qui produit une suite de nombres *pseudo-aléatoires*. C'est une suite de nombres qui, pratiquement, est semblable à un échantillon issu d'une distribution uniforme sur $(0, 1)$. La plupart des générateurs de nombres aléatoires procèdent en se basant sur une valeur initiale X_0 , appelée le *germe*, puis en calculant selon un processus récursif les valeurs suivantes à l'aide de nombres prédéterminés a , c et m , selon la formule

$$X_{n+1} = (aX_n + c) \text{ modulo } m \quad n \geq 0$$

Ceci signifie que $aX_n + c$ est divisé par m et que l'on assigne à X_{n+1} le reste de cette division. Chaque X_n peut donc prendre des valeurs dans $0, 1, \dots, m-1$ et c'est la quantité X_n/m qui est prise comme approximation d'une variable aléatoire uniforme sur $(0, 1)$. On peut montrer qu'avec des choix convenables de a , c et m , cette méthode produit une suite de nombres qui semblent provenir de l'observation de variables aléatoires indépendantes uniformes sur $(0, 1)$.

Pour simuler des variables de distribution quelconque, nous admettrons désormais que nous disposons déjà d'un simulateur de variable uniforme sur $(0, 1)$ et nous utiliserons le terme «nombres aléatoires» pour désigner les variables aléatoires indépendantes ayant cette distribution.

Dans l'exemple du solitaire nous aimerions programmer un ordinateur pour jouer la partie en partant avec un certain arrangement des cartes. Cependant, étant donné que la configuration initiale est l'une des $(52)!$ permutations possibles, et ceci avec la même probabilité, par hypothèse, il est également nécessaire de pouvoir générer une permutation aléatoire. L'algorithme suivant montre comment, seulement à partir de nombres aléatoires, on peut l'obtenir. L'algorithme commence par le choix aléatoire d'un des éléments et le place à la position n ; puis il prend un autre élément au hasard parmi ceux qui restent et le met en position $n-1$; et ainsi de suite. Le choix parmi les éléments restants se fait de manière efficace en maintenant ceux-ci dans une liste ordonnée et en sélectionnant au hasard une position dans cette liste.

Exemple 10.1 Génération d'une permutation aléatoire

Supposons que l'on s'intéresse à générer une permutation des entiers $1, 2, \dots, n$ telle que les $n!$ arrangements possibles soient équiprobables. On commencera avec n'importe quel arrangement et on obtiendra la permutation désirée après $n-1$ étapes où à chaque étape on intervertira les positions de deux des nombres de la permutation. On gardera en permanence trace de la permutation en notant $X(i)$, $i = 1, \dots, n$, le nombre qui se trouve actuellement en position i . L'algorithme procède comme suit:

1. Considérer une permutation arbitraire et noter $X(i)$ l'élément qui est placé dans la position i , $i = 1, \dots, n$. (Par exemple, on pourrait prendre $X(i) = i$, $i = 1, \dots, n$).
2. Générer une variable aléatoire N_n qui peut prendre les valeurs $1, 2, \dots, n$ de façon équiprobable.
3. Intervertir les valeurs de $X(N_n)$ et de $X(n)$. Désormais la valeur de $X(n)$ sera gardée fixe. [Par exemple, prenons $n = 4$ et au départ $X(i) = i$, $i = 1, 2, 3, 4$. Si $N_4 = 3$, alors la nouvelle permutation est $X(1) = 1$, $X(2) = 2$, $X(3) = 4$, $X(4) = 3$ et l'élément 3 sera maintenu à la position 4.]
4. Générer une variable aléatoire N_{n-1} qui vaudra $1, 2, \dots, n-1$ avec équiprobabilité.
5. Intervertir les valeurs de $X(N_{n-1})$ et de $X(n-1)$. [Si maintenant $N_3 = 1$, alors la nouvelle permutation est $X(1) = 4$, $X(2) = 2$, $X(3) = 1$ et $X(4) = 3$].
6. Générer une variable aléatoire N_{n-2} à valeurs dans $\{1, 2, \dots, n-2\}$ avec équiprobabilité.
7. Intervertir les valeurs de $X(N_{n-2})$ et de $X(n-2)$. [Si $N_2 = 1$ alors la nouvelle permutation est $X(1) = 2$, $X(2) = 4$, $X(3) = 1$, $X(4) = 3$ et c'est la permutation finale].
8. Générer N_{n-3} , et ainsi de suite. L'algorithme continue jusqu'à la génération de N_2 et, après l'intervention correspondante, la permutation résultante est la permutation cherchée.

Pour exécuter cet algorithme, il est nécessaire de pouvoir générer une variable aléatoire qui prend les valeurs $1, 2, \dots, k$ avec équiprobabilité. Pour le réaliser, noter U un nombre aléatoire – c'est-à-dire que U est uniformément distribué sur $(0, 1)$ – et remarquer que kU est uniforme sur $(0, k)$. Ainsi,

$$P\{i-1 < kU < i\} = \frac{1}{k} \quad i = 1, \dots, k$$

et en prenant $N_k = [kU] + 1$, où $[x]$ est la partie entière de x (c'est-à-dire le plus grand entier inférieur ou égal à x), alors N_k aura la distribution voulue.

L'algorithme peut se résumer de la manière suivante:

- Etape 1: Choisir $X(1), X(2), \dots, X(n)$, une des permutations de $1, 2, \dots, n$. (Par exemple, poser $X(i) = i$, $i = 1, \dots, n$).
- Etape 2: Poser $I = n$.
- Etape 3: Générer un nombre aléatoire U et poser $N = [IU] + 1$.
- Etape 4: Intervertir les valeurs de $X(N)$ et de $X(I)$.
- Etape 5: Diminuer I de 1 et si $I > 1$ aller à l'étape 3.
- Etape 6: $X(1), X(2), \dots, X(n)$ est la permutation aléatoire cherchée. ■

L'algorithme précédent, qui génère une permutation aléatoire, est extrêmement utile. Par exemple, supposons qu'un statisticien désire développer une expérience pour comparer les effets de m traitements différents sur un ensemble de n individus. Il décide de partager les individus en m groupes distincts d'effectifs respectifs n_1, n_2, \dots, n_m , où $\sum_{i=1}^m n_i = n$, et tels que les membres du groupe i reçoivent le traitement i . Pour éviter toute sorte de biais lors de l'affectation des individus aux groupes de traitement (l'interprétation des résultats de l'expérience pourrait être faussée s'il s'avérait que les «meilleurs» sujets étaient assignés au même groupe, par exemple), il est impératif que cette affectation d'un individu à un groupe donné soit faite de façon «aléatoire». Comment réaliser cette répartition?¹

Une procédure simple et efficace consiste à numéroter de façon arbitraire les individus $1, 2, \dots, n$ puis à générer une permutation aléatoire $X(1), \dots, X(n)$ de $1, 2, \dots, n$. Affecter alors les individus $X(1), X(2), \dots, X(n_1)$ au groupe 1, $X(n_1 + 1), \dots, X(n_1 + n_2)$ au groupe 2 et de façon générale le groupe j sera composé des individus numérotés $X(n_1 + n_2 + \dots + n_{j-1} + k)$, $k = 1, \dots, n_j$.

10.2 TECHNIQUES GÉNÉRALES POUR LA SIMULATION DE VARIABLES ALÉATOIRES CONTINUES

Dans cette section nous présenterons deux méthodes générales, basées sur l'utilisation des nombres aléatoires, pour simuler des variables aléatoires continues.

10.2.1 Méthode de la transformation inverse

Sur le théorème qui suit est basée une méthode générale pour la simulation de variables aléatoires continues, appelée *méthode de la transformation inverse*.

Théorème 10.1

Soit une variable uniforme U sur $(0, 1)$ et soit F une fonction de répartition continue quelconque. La variable aléatoire Y définie par

$$Y = F^{-1}(U)$$

est F pour fonction de répartition ($F^{-1}(x)$ désigne la valeur y pour laquelle $F(y) = x$).

DÉMONSTRATION.

$$\begin{aligned} F_Y(a) &= P\{Y \leq a\} \\ &= P\{F^{-1}(U) \leq a\} \end{aligned} \quad (10.1)$$

La fonction F étant monotone, on peut écrire que $F^{-1}(U) \leq a$ équivaut à $U \leq F(a)$. Placée dans (10.1) cette transformation donne

¹ Dans le cas $m = 2$, une autre technique pour la répartition aléatoire des individus est présentée dans l'exemple 6.12 du chapitre 6. La procédure générale est plus rapide mais nécessite plus de place en mémoire que celle de l'exemple 6.12.

$$F_Y(a) = P\{U \leq F(a)\} \\ = F(a) \quad \blacksquare$$

Selon le théorème 10.1, nous pouvons simuler une variable aléatoire X de fonction de répartition continue F en produisant un nombre aléatoire U et en posant $X = F^{-1}(U)$.

Exemple 10.2 Simulation d'une variable aléatoire exponentielle
Prenons $F(x) = 1 - e^{-x}$; alors $F^{-1}(y)$ est la valeur de x telle que

$$1 - e^{-x} = y$$

d'où

$$x = -\ln(1 - y)$$

Si U est uniforme sur $(0, 1)$ alors la variable

$$F^{-1}(U) = -\ln(1 - U)$$

sera de distribution exponentielle d'espérance 1. Comme $1 - U$ est également uniforme sur $(0, 1)$, il reste que $-\ln U$ est de distribution exponentielle d'espérance 1. Plus généralement, comme cX est exponentielle d'espérance c , lorsque X est exponentielle d'espérance 1, on conclut que $-c \ln U$ suit une loi exponentielle d'espérance c . ■

On peut utiliser les résultats de l'exemple 10.2 pour simuler une variable aléatoire gamma.

Exemple 10.3 Simulation d'une variable gamma (n, λ)

Pour obtenir une simulation d'une variable gamma de paramètres (n, λ) où n est entier, on utilise le fait que la somme de n variables exponentielles de paramètre λ et indépendantes suit la distribution voulue. Par conséquent, si U_1, \dots, U_n sont des variables aléatoires indépendantes uniformes sur $(0, 1)$ alors

$$X = -\sum_{i=1}^n \frac{1}{\lambda} \ln U_i = -\frac{1}{\lambda} \ln \left(\prod_{i=1}^n U_i \right)$$

suit la distribution demandée. ■

10.2.2 Méthode de rejet

Supposons que l'on dispose d'une méthode pour simuler une variable aléatoire de fonction de densité $g(x)$. Sur cette base, on peut simuler une variable aléatoire continue de fonction de densité $f(x)$. En effet, on simule d'abord Y ayant la densité g puis on accepte cette valeur générée avec une probabilité proportionnelle à $f(Y) / g(Y)$.

Plus précisément, soit c une constante telle que

$$\frac{f(y)}{g(y)} \leq c \quad \text{pour tout } y$$

Nous appliquons la procédure suivante pour simuler une variable aléatoire de fonction de densité f .

Méthode de rejet

Etape 1: On simule Y de fonction de densité g et on produit un nombre aléatoire U .

Etape 2: Si $U \leq f(Y) / cg(Y)$, on pose $X = Y$. Sinon on revient à l'étape 1.

La figure 10.1 représente graphiquement la méthode de rejet.

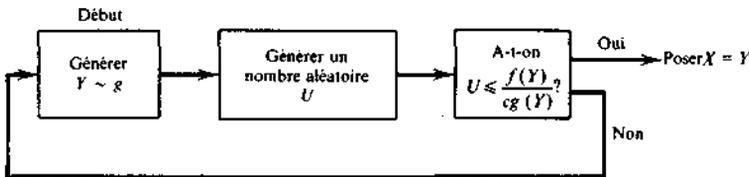


Figure 10.1 Méthode de rejet pour simuler une variable aléatoire X de fonction de densité f .

Le théorème suivant démontre la validité de la méthode de rejet.

Théorème 10.2

La variable aléatoire X générée par la méthode de rejet a la fonction de densité f .

DÉMONSTRATION. Soient X la valeur obtenue et N le nombre d'itérations nécessaire. Alors

$$\begin{aligned} P\{X \leq x\} &= P\{Y_N \leq x\} \\ &= P\left\{Y \leq x \mid U \leq \frac{f(Y)}{cg(Y)}\right\} \\ &= \frac{P\left\{Y \leq x, U \leq \frac{f(Y)}{cg(Y)}\right\}}{K} \end{aligned}$$

où $K = P\{U \leq f(Y)/cg(Y)\}$. Or la fonction de densité conjointe de Y et de U est, grâce à leur indépendance,

$$f(y, u) = g(y) \quad 0 < u < 1$$

En utilisant cette expression, on obtient

$$\begin{aligned}
 P\{X \leq x\} &= \frac{1}{K} \int \int_{\substack{r \leq x \\ 0 \leq u \leq f(y)/cg(y)}} g(y) \, du \, dy \\
 &= \frac{1}{K} \int_0^x \int_0^{f(y)/cg(y)} du \, g(y) \, dy \\
 &= \frac{1}{cK} \int_0^x f(y) \, dy
 \end{aligned} \tag{10.2}$$

En faisant tendre x vers l'infini, comme f est une densité, on a

$$1 = \frac{1}{cK} \int_0^{\infty} f(y) \, dy = \frac{1}{cK}$$

Grâce à l'équation (10.2), on obtient

$$P\{X \leq x\} = \int_0^x f(y) \, dy$$

Ce qui achève la démonstration. ■

On peut remarquer les points suivants.

- Lorsqu'on dit qu'on «accepte la valeur Y avec probabilité $f(Y)/cg(Y)$ », cela signifie que l'on génère un nombre aléatoire U et que l'on accepte Y si $U \leq f(Y)/cg(Y)$.
- Comme chaque itération produit, indépendamment d'une autre, un nombre qui sera accepté avec probabilité $P\{U \leq f(Y)/cg(Y)\} = K = 1/c$, le nombre d'itérations suit donc une loi géométrique de moyenne c .

Exemple 10.4 Simulation d'une variable aléatoire normale

Pour simuler une variable aléatoire normale standard Z (de moyenne 0 et de variance 1), il faut noter d'abord que la valeur absolue de Z a comme fonction de densité de probabilité

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2} \quad 0 < x < \infty \tag{10.3}$$

On commencera par simuler une variable de fonction de densité définie en (10.3) en utilisant la méthode de rejet où g est une fonction de densité exponentielle de moyenne 1 (dite standard), à savoir

$$g(x) = e^{-x} \quad 0 < x < \infty$$

Comme

$$\begin{aligned} \frac{f(x)}{g(x)} &= \sqrt{2/\pi} \exp \left\{ \frac{-(x^2 - 2x)}{2} \right\} \\ &= \sqrt{2/\pi} \exp \left\{ \frac{-(x^2 - 2x + 1)}{2} + \frac{1}{2} \right\} \\ &= \sqrt{2e/\pi} \exp \left\{ \frac{-(x - 1)^2}{2} \right\} \\ &\leq \sqrt{2e/\pi} \end{aligned} \quad (10.4)$$

en prenant $c = \sqrt{2e/\pi}$, grâce à l'équation (10.4) on a

$$\frac{f(x)}{cg(x)} = \exp \left\{ \frac{-(x - 1)^2}{2} \right\}$$

De ce fait, grâce à la méthode de rejet, on peut simuler la valeur absolue d'une variable aléatoire normale standard selon la procédure suivante:

- a) Générer deux variables aléatoires indépendantes Y et U , Y étant exponentielle standard et U uniforme sur $(0, 1)$.
- b) Si $U \leq \exp \{ -(Y - 1)^2/2 \}$ poser $X = Y$. Sinon revenir au point a).

Une fois que l'on a simulé une variable aléatoire X de fonction de densité donnée en (10.3), on peut générer une variable aléatoire normale standard Z où Z vaudra X ou $-X$ de façon équiprobable.

Dans l'étape b), la valeur Y est acceptée si $U \leq \exp \{ -(Y - 1)^2/2 \}$, ou de façon équivalente si $-\ln U \geq (Y - 1)^2/2$. Cependant, on a montré dans l'exemple 10.2 que $-\ln U$ est exponentiel standard, et par conséquent les étapes a) et b) sont équivalentes à

- a') Générer deux variables aléatoires indépendantes Y_1 et Y_2 exponentielles standard.
- b') Si $Y_2 \geq (Y_1 - 1)^2/2$, poser $X = Y_1$. Sinon revenir en a').

Supposons maintenant que dans cette procédure on accepte Y_1 - on sait donc que Y_2 est plus grand que $(Y_1 - 1)^2/2$. Que vaut cet écart? Pour répondre à cette question, rappelons que Y_2 est exponentiel de moyenne 1, par conséquent, étant donné qu'il dépasse une certaine valeur, le surplus entre Y_2 et $(Y_1 - 1)^2/2$ (c'est-à-dire «la durée de vie additionnelle» au-delà du temps $(Y_1 - 1)^2/2$) suit également une loi exponentielle de moyenne 1, en vertu de la propriété d'absence de mémoire. En d'autres termes, si l'on s'arrête à l'étape b'), on obtient non seulement X , la valeur absolue d'une normale standard, mais encore en calculant $Y_2 - (Y_1 - 1)^2/2$ une variable aléatoire exponentielle standard (indépendante de X).

Ce qui donne, en résumé, l'algorithme suivant pour simuler une variable exponentielle standard et une variable normale standard indépendantes.

Etape 1: Générer Y_1 , une variable aléatoire exponentielle standard.

Etape 2: Générer Y_2 , une variable aléatoire exponentielle standard.

Etape 3: Si $Y_2 - (Y_1 - 1)^2/2 > 0$ poser $Y = Y_2 - (Y_1 - 1)^2/2$ et aller à l'étape 4. Sinon aller à l'étape 1.

Etape 4: Générer un nombre aléatoire U et poser

$$Z = \begin{cases} Y_1 & \text{si } U \leq 1/2 \\ -Y_1 & \text{si } U > 1/2 \end{cases}$$

Les variables aléatoires Z et Y générées par l'algorithme précédent sont indépendantes; Z est normale de moyenne 0 et de variance 1 et Y est exponentielle standard. (Pour obtenir une variable normale de moyenne μ et de variance σ^2 , effectuer simplement la transformation $\mu + \sigma Z$).

Le lecteur remarquera que

- puisque $c = \sqrt{2e/\pi} \approx 1,32$, alors le nombre de fois où le point 2 est exécuté dans l'algorithme précédent doit être distribué selon une loi géométrique de moyenne 1,32.
- De plus, pour générer une suite de variables aléatoires normales standard, on peut utiliser la variable aléatoire exponentielle Y obtenue à l'étape 3 comme point de départ pour la génération de la variable normale suivante. Ainsi, en moyenne, on peut simuler une variable normale en générant 1,64 ($= 2 \times 1,32 - 1$) variables exponentielles et en calculant 1,32 termes au carré.

■

Exemple 10.5 Simulation de variables aléatoires normales par la méthode des coordonnées polaires

On a montré dans l'exemple 6.27 du chapitre 6 que si X et Y sont des variables aléatoires normales standard indépendantes, alors leurs coordonnées polaires $R = \sqrt{X^2 + Y^2}$, $\theta = \text{Arc tg}(Y/X)$ sont indépendantes. R^2 suit une loi exponentielle de moyenne 2 et θ est uniformément distribuée sur $(0, 2\pi)$. Ainsi, si U_1 et U_2 sont deux nombres aléatoires alors, grâce au résultat de l'exemple 10.2, on peut poser

$$\begin{cases} R = (-2 \ln U_1)^{1/2} \\ \theta = 2\pi U_2 \end{cases}$$

ce qui entraîne que

$$\begin{cases} X = R \cos \theta = (-2 \ln U_1)^{1/2} \cos(2\pi U_2) \\ Y = R \sin \theta = (-2 \ln U_1)^{1/2} \sin(2\pi U_2) \end{cases} \quad (10.5)$$

sont des variables aléatoires normales standard indépendantes.

L'approche précédente pour la génération de variables aléatoires normales standard est appelée l'approche de Box-Muller. Son efficacité souffre quelque peu de la nécessité de calculer les valeurs du sinus et du cosinus mentionnées. Il existe cependant un moyen pour réduire la consommation potentielle de temps de calcul. Pour commencer, remarquons que si U est uniforme sur $(0, 1)$ alors $2U$ est uniforme sur $(0, 2)$ et par conséquent $2U - 1$ est uniforme sur $(-1, 1)$. Si l'on a généré deux nombres aléatoires U_1 et U_2 et que l'on pose

$$\begin{cases} V_1 = 2U_1 - 1 \\ V_2 = 2U_2 - 1 \end{cases}$$

alors (V_1, V_2) est uniformément distribué à l'intérieur du carré d'aire 4 centré en $(0, 0)$ illustré sur la figure 10.2

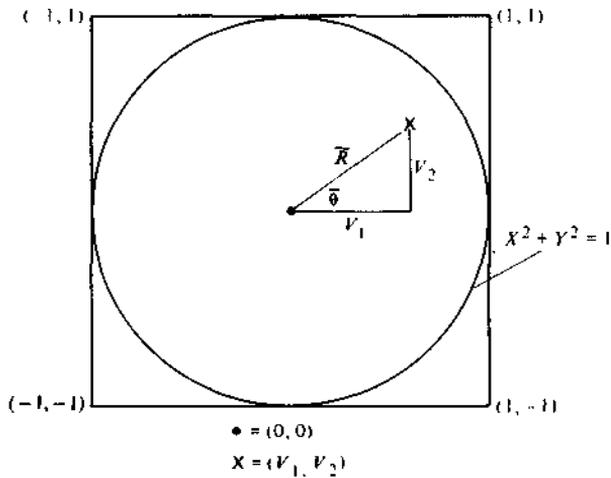


Figure 10.2

Supposons maintenant que nous générons une suite de couples (V_1, V_2) jusqu'à ce que nous obtenions un couple qui soit contenu dans un cercle de rayon 1 centré en $(0, 0)$ – c'est-à-dire jusqu'à ce que (V_1, V_2) soit tel que $V_1^2 + V_2^2 \leq 1$. Cela implique que ce couple (V_1, V_2) est uniformément distribué à l'intérieur du cercle. Soient \bar{R} et $\bar{\theta}$ leurs coordonnées polaires, il est facile de vérifier que \bar{R} et $\bar{\theta}$ sont indépendantes, avec \bar{R}^2 uniformément distribuée sur $(0, 1)$ et $\bar{\theta}$ uniformément distribuée sur $(0, 2\pi)$ – voir pour cela le problème 10.5.13.

Comme

$$\begin{cases} \sin\bar{\theta} = V_2/\bar{R} = \frac{V_2}{\sqrt{V_1^2 + V_2^2}} \\ \cos\bar{\theta} = V_1/\bar{R} = \frac{V_1}{\sqrt{V_1^2 + V_2^2}} \end{cases}$$

on déduit de l'équation (10.5) qu'on peut générer des variables aléatoires X et Y normales standard indépendantes en générant un autre nombre aléatoire U et en posant

$$\begin{cases} X = (-2 \ln U)^{1/2} V_1/\bar{R} \\ Y = (-2 \ln U)^{1/2} V_2/\bar{R} \end{cases}$$

Or, comme \bar{R}^2 , sachant que $V_1^2 + V_2^2 \leq 1$, a une distribution uniforme sur $(0, 1)$ et qu'il est indépendant de $\bar{\theta}$, on peut l'utiliser au lieu de générer un nouveau nombre aléatoire U ; de cette façon on montre que

$$\begin{cases} X = (-2 \ln \bar{R}^2)^{1/2} V_1 / \bar{R} = \sqrt{\frac{-2 \ln S}{S}} V_1 \\ Y = (-2 \ln \bar{R}^2)^{1/2} V_2 / \bar{R} = \sqrt{\frac{-2 \ln S}{S}} V_2 \end{cases}$$

sont des variables aléatoires normales standard indépendantes, où

$$S = \bar{R}^2 = V_1^2 + V_2^2$$

La procédure suivante résume la méthode pour générer un couple de variables aléatoires normales standard indépendantes.

Etape 1: Générer des nombres aléatoires U_1 et U_2 .

Etape 2: Poser $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$, $S = V_1^2 + V_2^2$.

Etape 3: Si $S > 1$ revenir à l'étape 1.

Etape 4: Produire les variables aléatoires normales standard indépendantes

$$X = \sqrt{\frac{-2 \ln S}{S}} V_1, Y = \sqrt{\frac{-2 \ln S}{S}} V_2$$

La méthode exposée ci-dessus s'appelle *la méthode des coordonnées polaires*. Comme la probabilité qu'un point, pris au hasard dans un carré, se trouve à l'intérieur du cercle est $\pi/4$, l'aire du cercle divisée par l'aire du carré, il s'ensuit que, en moyenne, la méthode des coordonnées polaires nécessite $4/\pi \approx 1,273$ itérations de l'étape 1. Par conséquent, en moyenne, on aura besoin de 2,546 nombres aléatoires, 1 logarithme, 1 racine carrée, 1 division et de 4,546 multiplications pour générer deux variables aléatoires normales standard indépendantes. ■

Exemple 10.6 Simulation d'une variable aléatoire Chi-carré

La loi Chi-carré à n degrés de liberté est la loi de $X_n^2 = Z_1^2 + \dots + Z_n^2$ où $Z_i, i = 1, \dots, n$, sont des variables aléatoires normales standard indépendantes. Comme il a été montré dans la section 6.3.4 du chapitre 6, $Z_1^2 + Z_2^2$ suit une loi exponentielle de paramètre $1/2$. Dans le cas où n est pair, disons $n = 2k$, X_{2k}^2 a une distribution gamma de paramètres $(k, 1/2)$. Par conséquent, $-2 \ln (\prod_{i=1}^k U_i)$ a une distribution Chi-carré avec $2k$ degrés de liberté. On peut simuler une variable aléatoire Chi-carré avec $2k + 1$ degrés de liberté en générant en premier lieu une variable aléatoire normale standard Z puis en ajoutant Z^2 à la variable Chi-carré précédente. En d'autres termes

$$X_{2k+1}^2 = Z^2 - 2 \ln \left(\prod_{i=1}^k U_i \right)$$

où Z, U_1, \dots, U_k sont indépendants, Z est normale standard et les autres variables sont uniformes sur $(0, 1)$. ■

10.3 SIMULATION DE VARIABLES ALÉATOIRES DISCRÈTES

10.3.1 Méthode de simulation

Toutes les méthodes générales pour simuler des variables aléatoires continues ont leur analogue dans le cas discret. Par exemple, supposons que l'on veuille simuler une variable aléatoire X dont la loi de probabilité est

$$P\{X = x_j\} = P_j, \quad j = 0, 1, \dots, \quad \sum_j P_j = 1$$

On peut utiliser la méthode discrète analogue à celle de la transformation inverse.

Pour simuler X tel que $P\{X = x_j\} = P_j$, on considère une variable U continue uniforme sur $(0, 1)$ puis l'on pose

$$X = \begin{cases} x_1 & \text{si } U < P_1 \\ x_2 & \text{si } P_1 < U < P_1 + P_2 \\ \vdots & \\ x_j & \text{si } \sum_1^{j-1} P_i < U < \sum_1^j P_i \\ \vdots & \end{cases}$$

Comme

$$P\{X = x_j\} = P \left\{ \sum_1^{j-1} P_i < U < \sum_1^j P_i \right\} = P_j$$

X aura la distribution voulue.

10.3.2 Exemples de simulation de variables discrètes

Exemple 10.7 Simulation d'une variable géométrique

Supposons que l'on effectue une suite d'épreuves indépendantes, ayant chacune une probabilité p , $0 < p < 1$, de succès, jusqu'à ce que l'on obtienne un succès. Si l'on désigne le nombre d'épreuves nécessaires par X , on aura

$$P\{X = i\} = (1 - p)^{i-1} p \quad i \geq 1$$

qui est clair si l'on remarque que $X = i$ si les $(i - 1)$ premières épreuves résultent en des échecs et la i -ème épreuve est un succès. La variable aléatoire X est appelée une variable aléatoire géométrique de paramètre p . Comme

$$\begin{aligned} \sum_{i=1}^{j-1} P\{X = i\} &= 1 - P\{X > j - 1\} \\ &= 1 - P\{j - 1 \text{ premières épreuves sont des échecs}\} \\ &= 1 - (1 - p)^{j-1} \quad j \geq 1 \end{aligned}$$

on peut simuler cette variable en générant un nombre aléatoire U et en donnant à X la valeur j telle que

$$1 - (1 - p)^{j-1} < U < 1 - (1 - p)^j$$

ou, de façon équivalente,

$$(1 - p)^j < 1 - U < (1 - p)^{j-1}$$

Comme $1 - U$ a la même distribution que U , on peut donc définir X par

$$\begin{aligned} X &= \min \{j : (1 - p)^j < U\} \\ &= \min \{j : j \ln(1 - p) < \ln U\} \\ &= \min \left\{ j : j > \frac{\ln U}{\ln(1 - p)} \right\} \end{aligned}$$

où l'inégalité change de signe puisque $\ln(1 - p)$ est négatif (en effet, $\ln(1 - p) < \ln 1 = 0$). En notant $[x]$ la partie entière de x (c'est-à-dire que $[x]$ est le plus grand entier inférieur ou égal à x), on peut écrire

$$X = 1 + \left\lceil \frac{\ln U}{\ln(1 - p)} \right\rceil \quad \blacksquare$$

Comme dans le cas continu, des techniques particulières de simulation ont été développées pour la plupart des distributions discrètes usuelles et certaines vont être présentées maintenant.

Exemple 10.8 Simulation d'une variable binomiale

Le meilleur moyen pour simuler une variable binomiale de paramètres (n, p) s'appuie sur le fait que celle-ci peut être exprimée comme la somme de n variables de Bernoulli indépendantes. Si U_1, \dots, U_n sont des variables uniformes sur $(0, 1)$ et indépendantes, on peut poser

$$X_i = \begin{cases} 1 & \text{si } U_i < p \\ 0 & \text{sinon} \end{cases}$$

et $X \equiv \sum_{i=1}^n X_i$ sera une variable aléatoire binomiale de paramètres n et p . ■

Exemple 10.9 Simulation d'une variable aléatoire poissonnienne

Pour simuler une variable aléatoire de Poisson de moyenne λ , on génère des variables aléatoires indépendantes uniformes sur $(0, 1)$, U_1, U_2, \dots jusqu'à

$$N = \min \left\{ n : \prod_{i=1}^n U_i < e^{-\lambda} \right\}$$

La variable aléatoire $X \equiv N - 1$ a la distribution voulue. En d'autres termes, si l'on continue de produire des nombres aléatoires jusqu'à ce que leur produit tombe au-dessous de $e^{-\lambda}$, alors le nombre de générations nécessaire moins 1 suit une loi de Poisson de moyenne λ .

Que $X \equiv N - 1$ soit effectivement poissonnienne de moyenne λ , on peut le voir facilement en remarquant que

$$X + 1 = \min \left\{ n : \prod_{i=1}^n U_i < e^{-\lambda} \right\}$$

est équivalent à

$$X = \max \left\{ n : \prod_{i=1}^n U_i \geq e^{-\lambda} \right\} \quad \text{où} \quad \prod_{i=1}^0 U_i = 1$$

ou, en prenant le logarithme, à

$$X = \max \left\{ n : \sum_{i=1}^n \ln U_i \geq -\lambda \right\}$$

ou encore à

$$X = \max \left\{ n : \sum_{i=1}^n -\ln U_i \leq \lambda \right\}$$

Cependant, $-\ln U_i$ est exponentielle de paramètre 1 et de ce fait X peut être considéré comme le nombre maximal de variables exponentielles de paramètre 1 dont la somme reste inférieure à λ . Nous avons vu que les temps d'attente entre deux événements successifs d'un processus de Poisson de moyenne 1 sont exponentiels de paramètre 1 et indépendants. Ainsi X est égal au nombre d'événements durant un intervalle de temps λ d'un processus de Poisson de paramètre 1; par conséquent X a une distribution de Poisson de moyenne λ . ■

10.4 TECHNIQUES DE LA RÉDUCTION DE LA VARIANCE

Soient X_1, \dots, X_n dont la distribution conjointe est donnée; supposons que l'on s'intéresse au calcul de

$$\theta \equiv E[g(X_1, \dots, X_n)]$$

où g est une fonction fixée. Le calcul analytique de cette expression s'avère souvent extrêmement difficile et, dans ce cas, on peut essayer de recourir à la simulation pour estimer θ . La procédure est la suivante: générer les variables $X_1^{(1)}, \dots, X_n^{(1)}$ ayant la même distribution conjointe que X_1, \dots, X_n et poser

$$Y_1 = g(X_1^{(1)}, \dots, X_n^{(1)})$$

Simuler alors un deuxième ensemble de variables aléatoires (indépendant du premier ensemble) $X_1^{(2)}, \dots, X_n^{(2)}$ avec toujours la même distribution que celle de X_1, \dots, X_n puis poser

$$Y_2 = g(X_1^{(2)}, \dots, X_n^{(2)})$$

On continue jusqu'à la génération du k -ième ensemble, où k est un nombre prédéfini, et jusqu'à la détermination de Y_1, \dots, Y_k . Or, Y_1, \dots, Y_k sont des variables aléatoires indépendantes et identiquement distribuées ayant chacune la même distribution que $g(X_1, \dots, X_n)$. Si l'on note alors par \bar{Y} la moyenne de ces k variables aléatoires, c'est-à-dire

$$\bar{Y} = \sum_{i=1}^k \frac{Y_i}{k}$$

on a

$$\begin{aligned} E[\bar{Y}] &= \theta \\ E[(\bar{Y} - \theta)^2] &= \text{Var}(\bar{Y}) \end{aligned}$$

On peut donc prendre \bar{Y} comme estimateur de θ . Comme l'espérance du carré de la différence entre \bar{Y} et θ est égale à la variance de \bar{Y} , il faudrait que cette quantité soit aussi petite que possible. [Dans la situation précédente, $\text{Var}(\bar{Y}) = \text{Var}(Y_i)/k$, que l'on ne connaît en général pas à l'avance mais qui doit être estimé à partir des valeurs générées Y_1, \dots, Y_n]. Nous allons maintenant présenter trois techniques pour réduire la variance de cet estimateur.

10.4.1 Utilisation de variables antithétiques

Dans la situation précédente, supposons que nous ayons généré deux variables Y_1 et Y_2 identiquement distribuées de moyenne θ . On a

$$\begin{aligned} \text{Var} \left(\frac{Y_1 + Y_2}{2} \right) &= \frac{1}{4} [\text{Var}(Y_1) + \text{Var}(Y_2) + 2 \text{Cov}(Y_1, Y_2)] \\ &= \frac{\text{Var}(Y_1)}{2} + \frac{\text{Cov}(Y_1, Y_2)}{2} \end{aligned}$$

Il serait alors avantageux (dans le sens où la variance serait réduite) d'avoir Y_1 et Y_2 non pas indépendants mais corrélés négativement. Comment s'arrange-t-on pour l'obtenir? Supposons que les variables X_1, \dots, X_n sont indépendantes et qu'en plus elles ont été générées grâce à la technique de la transformation inverse. En clair, X_i est simulée à partir de $F_i^{-1}(U_i)$ où U_i est un nombre aléatoire et F_i la distribution de X_i . Y_1 peut donc s'écrire comme

$$Y_1 = g(F_1^{-1}(U_1), \dots, F_n^{-1}(U_n))$$

Or, pour tout nombre aléatoire U , $1 - U$ est aussi uniforme sur $(0, 1)$ et est corrélé négativement avec U . Y_2 défini par

$$Y_2 = g(F_1^{-1}(1 - U_1), \dots, F_n^{-1}(1 - U_n))$$

aura donc la même distribution que Y_1 . Par conséquent, si Y_1 et Y_2 ont une corrélation négative, alors en générant Y_2 par cette méthode, on obtiendra une variance plus petite que s'il provenait d'un nouvel ensemble de nombres aléatoires. (De plus, on économise des opérations, car au lieu de générer n nombres aléatoires supplémentaires, il suffit de soustraire chacun des n nombres précédents de 1). Bien qu'en général on ne puisse pas être certain que Y_1 et Y_2 soient corrélés négativement, il se trouve que c'est souvent le cas et on peut effectivement montrer que c'est le cas si g est une fonction monotone.

10.4.2 Réduction de la variance par conditionnement

Rappelons d'abord la formule de la variance conditionnelle (voir la section 7.6.8 du chapitre 7)

$$\text{Var}(Y) = E[\text{Var}(Y|Z)] + \text{Var}(E[Y|Z])$$

Supposons que l'on s'intéresse à l'estimation de $E[g(X_1, \dots, X_n)]$ en simulant $\mathbf{X} = (X_1, \dots, X_n)$ et en calculant $Y = g(\mathbf{X})$. Si pour une variable aléatoire Z on arrive à calculer $E[Y|Z]$ alors, comme $\text{Var}(Y|Z) \geq 0$, on obtient grâce à la formule de la variance conditionnelle que

$$\text{Var}(E[Y|Z]) \leq \text{Var}(Y)$$

Ceci implique, en plus du fait que $E[E[Y|Z]] = E[Y]$, que $E[Y|Z]$ est un meilleur estimateur de $E[Y]$ que Y .

Exemple 10.10 Estimation de π

Soient U_1 et U_2 des nombres aléatoires et $V_i = 2U_i - 1$, $i = 1, 2$. On a vu dans l'exemple 10.5 que (V_1, V_2) est uniformément distribué dans le carré d'aire 4 centré en $(0, 0)$. La probabilité que ce point se trouve à l'intérieur du cercle inscrit de rayon 1 et centré en $(0, 0)$ – voir la figure 10.2 – est égale à $\pi/4$ (le rapport entre l'aire du cercle et celle du carré). Par conséquent, grâce à la simulation d'un grand nombre n de couples ainsi définis et à la définition de

$$I_j = \begin{cases} 1 & \text{si la } j\text{-ème paire tombe dans le cercle} \\ 0 & \text{sinon} \end{cases}$$

il s'ensuit que I_j , $j = 1, \dots, n$ sont des variables aléatoires indépendantes, identiquement distribuées et ayant la même moyenne $E[I_j] = \pi/4$. On a donc, en vertu de la loi forte des grands nombres, que

$$\frac{I_1 + \dots + I_n}{n} \longrightarrow \pi/4 \quad \text{lorsque } n \rightarrow \infty$$

De ce fait, en simulant un grand nombre de couples (V_1, V_2) et en multipliant la proportion de ceux qui se trouvent à l'intérieur du cercle par 4, on obtient une approximation précise de π .

L'estimateur mentionné ci-dessus peut encore être amélioré en recourant à l'espérance conditionnelle. Soit I la variable indicatrice définie précédemment pour le couple (V_1, V_2) . Au lieu de considérer la valeur observée de I , il est préférable de conditionner sur V_1 et d'utiliser

$$\begin{aligned} E[I|V_1] &= P\{V_1^2 + V_2^2 \leq 1 | V_1\} \\ &= P\{V_2^2 \leq 1 - V_1^2 | V_1\} \end{aligned}$$

Or

$$\begin{aligned} P\{V_2^2 \leq 1 - V_1^2 | V_1 = v\} &= P\{V_2^2 \leq 1 - v^2\} \\ &= P\{-\sqrt{1 - v^2} \leq V_2 \leq \sqrt{1 - v^2}\} \\ &= \sqrt{1 - v^2} \end{aligned}$$

et ainsi

$$E[I|V_1] = E[\sqrt{1 - V_1^2}]$$

De cette façon, pour estimer $\pi/4$, on obtient une amélioration si l'on utilise non pas la moyenne de I mais plutôt la valeur moyenne de $\sqrt{1 - V_1^2}$. Plus précisément, comme

$$E[\sqrt{1 - V_1^2}] = \int_{-1}^1 \frac{1}{2} \sqrt{1 - v^2} dv = \int_0^1 \sqrt{1 - u^2} du = E[\sqrt{1 - U^2}]$$

où U est uniforme sur $(0, 1)$, on peut générer n nombres aléatoires U et prendre la valeur moyenne de $\sqrt{1 - U^2}$ comme estimation de $\pi/4$. (Le problème 10.5.14 montre que cet estimateur a la même variance que la moyenne des n valeurs $\sqrt{1 - V_1^2}$.)

On peut même encore améliorer cet estimateur de π si l'on remarque que la fonction $g(u) = \sqrt{1 - u^2}$, $0 \leq u \leq 1$ est une fonction monotone décroissante et ainsi la méthode des variables antithétiques réduira donc la variance de l'estimateur de $E[\sqrt{1 - U^2}]$. En d'autres termes, plutôt que de générer n nombres aléatoires et de prendre la valeur moyenne de $\sqrt{1 - U^2}$ comme estimation de $\pi/4$, on peut obtenir un meilleur estimateur en générant seulement $n/2$ nombres aléatoires U et en prenant la moyenne de $\sqrt{1 - U^2} + \sqrt{1 - (1 - U)^2}$ divisée par 2 comme estimation de $\pi/4$.

La table suivante donne les estimations de π résultant de simulations basées sur les 3 estimateurs cités, en prenant $n = 10\,000$.

<i>Méthode</i>	<i>Estimation de π</i>
Utilisant la proportion des points aléatoires tombant dans le cercle	3,1612
Utilisant la valeur moyenne de $\sqrt{1 - U^2}$	3,128448
Utilisant la valeur moyenne de $\sqrt{1 - U^2} + \sqrt{1 - (1 - U)^2}$	3,139578

Une autre simulation, qui considère la dernière approche avec $n = 64\,000$, donne l'estimation 3,143288. ■

10.4.3 Variables de contrôle

Supposons de nouveau que l'on veuille estimer $E[g(\mathbf{X})]$, où $\mathbf{X} = (X_1, \dots, X_n)$, par la méthode de simulation. Mais maintenant nous supposons que pour une fonction f la valeur moyenne de $f(\mathbf{X})$ est connue – à savoir $E[f(\mathbf{X})] = \mu$. Pour toute constante a , on peut alors prendre

$$W = g(\mathbf{X}) + a[f(\mathbf{X}) - \mu]$$

comme estimateur de $E[g(\mathbf{X})]$. Or

$$\text{Var}(W) = \text{Var}[g(\mathbf{X})] + a^2 \text{Var}[f(\mathbf{X})] + 2a \text{Cov}[g(\mathbf{X}), f(\mathbf{X})] \quad (10.6)$$

Un simple calcul montre que cette expression est minimale lorsque

$$a = \frac{-\text{Cov}[f(\mathbf{X}), g(\mathbf{X})]}{\text{Var}[f(\mathbf{X})]} \quad (10.7)$$

et, pour cette valeur de a , elle vaut

$$\text{Var}(W) = \text{Var}[g(\mathbf{X})] - \frac{[\text{Cov}[f(\mathbf{X}), g(\mathbf{X})]]^2}{\text{Var}[f(\mathbf{X})]} \quad (10.8)$$

Malheureusement, $\text{Var}[f(\mathbf{X})]$ et $\text{Cov}[f(\mathbf{X}), g(\mathbf{X})]$ ne sont en général pas connues, aussi nous ne pouvons obtenir cette réduction de variance. Dans la pratique, une approche consiste à estimer ces valeurs et à espérer que la variable W obtenue ait effectivement une variance plus petite que celle de $g(\mathbf{X})$, alors qu'une seconde possibilité est de simuler les données pour estimer ces quantités.

10.5 PROBLÈMES

10.5.1 L'algorithme suivant génère une permutation aléatoire des éléments 1, 2, ..., n . Il est un peu plus rapide que celui présenté dans l'exemple 10.1 mais il est tel qu'aucune

position n'est fixée avant que l'algorithme s'arrête. Dans cet algorithme, $P(i)$ peut être considéré comme l'élément se trouvant en position i .

Étape 1: Poser $k = 1$.

Étape 2: Poser $P(1) = 1$.

Étape 3: Si $k = n$, s'arrêter. Sinon poser $k = k + 1$.

Étape 4: Générer un nombre aléatoire U et poser

$$P(k) = P([kU] + 1)$$

$$P([kU] + 1) = k.$$

Revenir à l'étape 3.

- a) Expliquer avec des phrases ce que fait l'algorithme.
 b) Montrer qu'à l'itération k – c'est-à-dire au moment où la valeur de $P(k)$ est initialisée – $P(1), P(2), \dots, P(k)$ est une permutation aléatoire de $1, 2, \dots, k$.
 Utiliser pour cela un raisonnement par induction et montrer que

$$\begin{aligned} P_k\{i_1, i_2, \dots, i_{j-1}, k, i_j, \dots, i_{k-2}, i\} \\ &= P_{k-1}\{i_1, i_2, \dots, i_{j-1}, i, i_j, \dots, i_{k-2}\} \frac{1}{k} \\ &= \frac{1}{k!} \text{ grâce à l'hypothèse d'induction} \end{aligned}$$

10.5.2 Développer une technique pour simuler une variable aléatoire ayant la fonction de densité de probabilité

$$f(x) = \begin{cases} e^{2x} & -\infty < x < 0 \\ e^{-2x} & 0 < x < \infty \end{cases}$$

10.5.3 Développer une technique pour simuler une variable aléatoire ayant la fonction de densité de probabilité

$$f(x) = \begin{cases} \frac{1}{2}(x-2) & 2 \leq x \leq 3 \\ \frac{1}{2}\left(2 - \frac{x}{3}\right) & 3 < x \leq 6 \\ 0 & \text{ailleurs} \end{cases}$$

10.5.4 Présenter une méthode pour simuler une variable aléatoire ayant la fonction de répartition

$$F(x) = \begin{cases} 0 & x \leq -3 \\ \frac{1}{2} + \frac{x}{6} & -3 < x < 0 \\ \frac{1}{2} + \frac{x^2}{32} & 0 < x \leq 4 \\ 1 & x > 4 \end{cases}$$

10.5.5 Utiliser la méthode de la transformation inverse pour présenter une approche pour la simulation d'une variable aléatoire distribuée selon la loi de Weibull

$$F(t) = 1 - e^{-at^\beta} \quad t \geq 0$$

10.5.6 Donner une méthode de simulation d'une variable aléatoire ayant la fonction taux de panne

- a) $\lambda(t) = c$
- b) $\lambda(t) = ct$
- c) $\lambda(t) = ct^2$
- d) $\lambda(t) = ct^3$

10.5.7 Soit F la fonction de répartition

$$F(x) = x^n \quad 0 < x < 1$$

- a) Donner une méthode pour simuler une variable aléatoire ayant F comme fonction de répartition, qui n'utilise qu'un seul nombre aléatoire.
- b) Soient U_1, U_2, \dots, U_n des nombres aléatoires indépendants. Montrer que

$$P\{\max(U_1, \dots, U_n) \leq x\} = x^n$$

- c) Utiliser la partie b) pour élaborer une seconde méthode de simulation d'une variable aléatoire de fonction de répartition F .

10.5.8 Supposer qu'il est relativement facile de simuler à partir de $F_i, i = 1, \dots, n$. Comment peut-on simuler à partir de

a)
$$F(x) = \prod_{i=1}^n F_i(x)$$

b)
$$F(x) = 1 - \prod_{i=1}^n [1 - F_i(x)]$$

10.5.9 Supposer que l'on a une méthode pour simuler des variables aléatoires ayant les distributions F_1 et F_2 . Expliquer comment on simule à partir de la distribution

$$F(x) = pF_1(x) + (1 - p)F_2(x) \quad 0 < p < 1$$

Donner une méthode pour simuler à partir de

$$F(x) = \begin{cases} \frac{1}{3}(1 - e^{-3x}) + \frac{2}{3}x & 0 < x \leq 1 \\ \frac{1}{3}(1 - e^{-3x}) + \frac{2}{3} & x > 1 \end{cases}$$

10.5.10 Dans l'exemple 10.4 nous avons simulé la valeur absolue d'une variable aléatoire normale standard selon la procédure de rejet appliquée sur des variables

exponentielles de paramètre 1. Ceci soulève la question suivante: pouvait-on obtenir un algorithme plus efficace à partir d'une densité exponentielle différente – c'est-à-dire d'une densité $g(x) = \lambda e^{-\lambda x}$, par exemple? Montrer que le nombre moyen d'itérations nécessaire dans la technique de rejet est minimal lorsque $\lambda = 1$.

10.5.11 Utiliser la méthode de rejet avec $g(x) = 1$, $0 < x < 1$, pour déterminer un algorithme de simulation d'une variable aléatoire ayant la fonction de densité de probabilité

$$f(x) = \begin{cases} 60x^3(1-x)^2 & 0 < x < 1 \\ 0 & \text{ailleurs} \end{cases}$$

10.5.12 Expliquer comment on peut utiliser des nombres aléatoires pour approximer $\int_0^1 k(x) dx$ où $k(x)$ est une fonction arbitraire. Pour cela, dire ce que représente $E[k(U)]$ si U est uniforme sur $(0, 1)$.

10.5.13 Si (X, Y) est distribué uniformément dans le cercle de rayon 1 centré à l'origine, sa densité conjointe est

$$f(x,y) = 1/\pi, \quad 0 \leq x^2 + y^2 \leq 1$$

Soient $R = (X^2 + Y^2)^{1/2}$ et $\theta = \text{Arc tg}(Y/X)$ leurs coordonnées polaires. Montrer que R et θ sont indépendants, avec R^2 uniforme sur $(0, 1)$ et θ uniforme sur $(0, 2\pi)$.

10.5.14 Dans l'exemple 10.10 nous avons montré que

$$E[(1 - V^2)^{1/2}] = E[(1 - U^2)^{1/2}] = \pi/4$$

où V est uniforme sur $(-1, 1)$ et U sur $(0, 1)$. Montrer que

$$\text{Var} [(1 - V^2)^{1/2}] = \text{Var} [(1 - U^2)^{1/2}]$$

et trouver leur valeur commune.

10.5.15

- Vérifier que (10.6) atteint son minimum lorsque a est donné par (10.7).
- Vérifier que la valeur minimale de (10.6) est donnée par (10.8).

10.5.16 Soit X une variable aléatoire définie sur $(0, 1)$ ayant la fonction de densité $f(x)$. Montrer que l'on peut estimer $\int_0^1 g(x) dx$ en simulant X et en prenant $g(X)/f(X)$ comme estimateur. Cette méthode, appelée *échantillonnage par importance*, essaie de choisir une fonction f de forme semblable à celle de g telle que $g(X)/f(X)$ ait une variance faible.

10.6 RÉFÉRENCE

Solutions à quelques problèmes choisis

Chapitre 1

- 1.8.1 a) 67 600 000 b) 19 656 000 1.8.2 24, 4 1.8.3 144, 4
1.8.4 2401 1.8.5 5184
1.8.6 720 a) 72 b) 144 c) 72 1.8.7 a) 120 b) 1260 c) 34 650
1.8.8 27 720
1.8.9 a) 40 320 b) 10 080 c) 1152 d) 2880 e) 384
1.8.10 a) 720 b) 72 c) 144 1.8.11 a) 720 b) 672 c) 384 d) 216 e) 576
1.8.12 a) 24 300 000 b) 17 100 720 1.8.13 190 1.8.14 2 598 960
1.8.15 600 1.8.16 a) 120 b) 110
1.8.17 a) 896 b) 1000 c) 910 1.8.18 a) 36 b) 26
1.8.19 35 1.8.20 18 1.8.21 48
1.8.25 27 720 1.8.26 210 1.8.27 165, 35
1.8.28 a) 65 536 b) 2520 1.8.29 a) 1287 b) 14 112
1.8.30 1 852 200 1.8.31 a) 12 600 b) 945
1.8.32 564 480 1.8.33 a) 220 b) 552

Chapitre 2

- 2.9.6 74 2.9.7 a) 0.4 b) 0.1 2.9.8 a) 0.49 b) 0.28
2.9.9 a) 0.5 b) 0.32 c) 149/198
2.9.10 a) 20 000 b) 12 000 c) 11 000 d) 10 000 e) 68 000
2.9.11 1.057 2.9.12 0.0769, 0.03116
2.9.13 a) 0.0020 b) 0.4226 c) 0.0475 d) 0.0211 e) 0.00024
2.9.15 9.10946×10^{-6} 2.9.16 0.048 2.9.18 0.4
2.9.19 0.492929 2.9.20 0.58333 2.9.21 0.2477, 0.2099

2.9.22 0.5	2.9.23 2/9, 1/9	2.9.25 70/323
2.9.28 0.8363	2.9.29 a) 0.0045 b) 0.0588	
2.9.30 a) 0.0833 b) 0.5	2.9.31 4	2.9.32 0.48
2.9.34 0.5177	2.9.41 1.0604×10^{-3}	2.9.42 0.4329
2.9.43 2.6084×10^{-6}	2.9.45 a) 0.09145 b) 0.4268	
2.9.46 4/7	2.9.47 12/35	2.9.48 0.0511
2.9.49 a) 0.2198 b) 0.0342		

Chapitre 3

3.7.1 1/3	3.7.3 0.339	3.7.5 6/91
3.7.6 1/2	3.7.7 2/3	3.7.8 1/2
3.7.9 7/11	3.7.10 a) 0.1818 b) 0.4073 c) 0.2845, 0.2532	
3.7.11 0.22	3.7.12 0.4697	3.7.13 0.9835
3.7.14 a) 0.0792 b) 0.264	3.7.15 a) 0.331 b) 0.383 c) 0.286 d) 48.62	
3.7.16 a) 44.3 b) 41.18	3.7.17 a) 0.4 b) 1/26	3.7.18 a) 0.496 b) 3/14 c) 9/62
3.7.19 a) 1/2 b) 1/8	3.7.21 35/768	3.7.22 4/9, 1/2
3.7.24 a) 1/3 b) 1/2	3.7.26 20/21, 40/41	3.7.27 7/12, 3/5
3.7.28 5/11	3.7.29 4/5	3.7.30 27/31
3.7.31 3/4	3.7.32 1/2	3.7.33 a) 1/3 b) 1/5 c) 1
3.7.34 12/37	3.7.35 46/185	
3.7.36 a) 3/13 b) 5/13 c) 5/52 d) 15/52		3.7.37 $b/(b+r+c)$
3.7.38 43/459	3.7.39 34.48	3.7.40 4/9
3.7.42 1/11	3.7.44 2/3	3.7.45 19/268
3.7.46 17.5, 38/165, 17/33	3.7.48 9	3.7.50 c) 2/3
3.7.51 a) 2/3 b) 1/3 c) 3/4	3.7.55 a) 9/13 b) 1/2	
3.7.59 9, 9, 18, 110, 4, 4, 8, 120 tous sur 128		3.7.60 1/9, 1/18
3.7.61 38/64, 13/64, 13/64	3.7.63 a) 1/16 b) 1/132 c) 5/16 d) 1/4 e) 31/32	
3.7.65 $P_1/(P_1 + P_2 - P_1P_2)$	3.7.68 3/10	3.7.69 0.5550
3.7.71 0.9530	3.7.73 a) 1/2 b) 3/5 c) 4/5	
3.7.74 9/19, 6/19, 4/19; 77/165, 53/165, 35/165		
3.7.79 97/142, 15/26, 33/102		

Chapitre 4

4.11.1 $P(4) = 6/91, P(2) = 8/91, P(1) = 32/91, P(0) = 1/91, P(-1) = 16/91, P(-2) = 28/91$	
4.11.4 1/2, 5/18, 5/36, 5/84, 5/252, 1/252, 0, 0, 0, 0	4.11.5 $n - 2i, i = 0, \dots, n$
4.11.6 $P(3) = P(-3) = 1/8, P(1) = P(-1) = 3/8$	
4.11.12 a) $P(4) = 1/16, P(3) = 1/8, P(2) = 1/16, P(0) = 1/2, P(-i) = P(i)$ b) $P(0) = 1$	
4.11.13 $P(0) = 0.28, P(500) = 0.27, P(1000) = 0.315, P(1500) = 0.09, P(2000) = 0.045$	

- 4.11.14 $P(0) = 1/2, P(1) = 1/6, P(2) = 1/12, P(3) = 1/20, P(4) = 1/5$
 4.11.17 a) $1/4, 1/6, 1/12$ b) $1/2$ 4.11.19 $1/2, 1/10, 1/5, 1/10, 1/10$
 4.11.20 a) 0.5918 b) non c) -0.108 4.11.21 $38.793, 36.25$
 4.11.23 3.5 4.11.24 $p = 11/18, \text{maximin} = 23/72$
 4.11.26 $11/2, 17/5$ 4.11.27 $A(p + 1/10)$ 4.11.28 $3/5$
 4.11.31 p^* 4.11.32 $110 - 100(0.9)^{10}$
 4.11.33 96 4.11.35 a) -0.0067 b) 1.093
 4.11.37 $92.5, 92.1875$ 4.11.39 $3/8$ 4.11.40 $11/243$
 4.11.42 $p \geq 1/2$ 4.11.45 3 4.11.50 a) $1/10$ b) $1/10$
 4.11.51 a) e^{-2} b) $1 - 3e^{-2}$ 4.11.53 a) $1 - e^{-6}$ b) $1 - e^{-219.18}$
 4.11.56 $365 \log(2)$ 4.11.57 a) 0.5768 b) 0.6070
 4.11.59 a) 0.3935 b) 0.3033 c) 0.0902 4.11.60 0.8886
 4.11.61 0.4082 4.11.63 a) 0.0821 b) 0.2424
 4.11.65 a) 0.3935 b) 0.2293 c) 0.3935 4.11.66 a) 0.1500 b) 0.1012
 4.11.68 5.8125 4.11.69 a) $32/243$ b) $4864/6561$ c) $160/729$ d) $160/729$
 4.11.73 $18(17)^{n-1}/(35)^n$ 4.11.76 $3/10, 5/6, 75/138$
 4.11.77 0.3439

Chapitre 5

- 5.9.2 $3.5e^{-5/2}$ 5.9.3 non, non 5.9.4 a) $1/2$
 5.9.5 $1 - (0.01)^{1/5}$ 5.9.6 a) 4 b) 0 c) ∞ 5.9.7 $3/5, 6/5$
 5.9.8 2 5.9.10 a) $2/3$ b) $2/3$ 5.9.11 $2/5$
 5.9.13 a) $2/3$ b) $1/3$ 5.9.16 $(0.9938)^{10}$ 5.9.18 22.66
 5.9.19 14.56 5.9.20 a) 0.0006 b) 0.75 c) 0.977
 5.9.22 a) 9.5 b) 0.0019 5.9.23 $0.9258, 0.1762$
 5.9.26 $0.0606, 0.0525$ 5.9.30 a) e^{-1} b) $e^{-1/2}$ 5.9.32 $e^{-1}, 1/3$
 5.9.36 $3/5$ 5.9.38 $1/y$

Chapitre 6

- 6.8.2 a) $14/39, 10/39, 10/39, 5/39$ b) $84, 70, 70, 70, 40, 40, 40, 15$ tous sur 429
 6.8.3 $15/26, 5/26, 5/26, 1/26$ 6.8.4 $25/169, 40/169, 40/169, 64/169$
 6.8.6 $p(i, j) = 1/10$ 6.8.7 $p(i, j) = p^2 (1 - p)^{i+j}$
 6.8.8 $1/8$ 6.8.9 b) $(12x^2 + 6x)/7$ c) $15/56$ d) 0.8625
 6.8.10 a) $1/2$ b) $1 - e^{-a}$ 6.8.11 0.1458 6.8.12 $(39.3)e^{-5}$
 6.8.13 $1/6, 1/2$ 6.8.15 $\pi/4$ 6.8.16 $n(1/2)^{n-1}$
 6.8.17 $1/3$ 6.8.18 $7/9$ 6.8.19 $1/2$
 6.8.21 $e^{-1/i!}$ 6.8.24 $\lambda_1^2 a / [(\lambda_1 a + \lambda_2)(\lambda_1 + \lambda_2)]$
 6.8.26 a) e^{-2} b) $1 - 3e^{-2}$ 6.8.28 0.0446 6.8.29 a) $1/3, 2/3$ b) $5/12, 7/12$
 6.8.30 $5/13, 8/13$ 6.8.31 a) $1/6, 5/6$ b) $1/4, 3/4$

6.8.33

$i \backslash j$	1	2	3	4	5
1	0.438	0.219	0.146	0.1095	0.0876
2	0	0.3896	0.2597	0.1948	0.1558
3	0	0	0.4255	0.3191	0.2553
4	0	0	0	0.5556	0.4444
5	0	0	0	0	1

$$P\{X = j | Y = i\}$$

$$6.8.36 \text{ a) } (y+1)^2 x e^{-x(y+1)}, x e^{-xy} \text{ b) } e^{-x}$$

$$6.8.37 \frac{1}{2} + 3y/(4x) - y^3/(4x^3)$$

$$6.8.41 \ 0.79297$$

$$6.8.44 \ r/\pi$$

$$6.8.42 \text{ a) } 1 - e^{-5\lambda a}$$

$$6.8.45 \ r$$

$$6.8.40 \ (1 - 2d/L)^3$$

$$\text{b) } (1 - e^{-\lambda a})^5$$

$$6.8.48 \ u/(v+1)^2$$

Chapitre 7

$$7.10.3 \ 3/2$$

$$7.10.7 \text{ a) } 0.6 \text{ b) } 0$$

$$7.10.11 \ n/(1-p)$$

$$7.10.18 \ 0.9301, 87.5757$$

$$7.10.23 \ n/(n+1), 1/(n+1)$$

$$7.10.27 \ 20/19, 360/361$$

$$7.10.29 \ -n/36$$

$$7.10.34 \ 6, 112/33$$

$$7.10.38 \ \text{a) } 1/2 \ \text{b) } 0$$

$$7.10.41 \ 9/5, 6/5, 3/5, 0$$

$$7.10.44 \ 12$$

$$7.10.53 \ -96/145$$

$$7.10.59 \ 1.2, 1/16, 2/81$$

$$7.10.63 \ \text{a) } \mu \ \text{b) } 1 + \sigma^2 \ \text{c) } \text{oui} \ \text{d) } \sigma^2$$

$$7.10.4 \ 35$$

$$7.10.8 \ 2(n-1)p(1-p)$$

$$7.10.12 \ 109/60$$

$$7.10.19 \ 14.7$$

$$7.10.24 \ 175/6$$

$$7.10.28 \ \text{a) } 21.2 \ \text{b) } 18.929 \ \text{c) } 49.214$$

$$7.10.30 \ 0$$

$$7.10.35 \ 100/19, 16 \ 200/6137, 10/19, 3240/6137$$

$$7.10.40 \ \text{a) } 6 \ \text{b) } 7 \ \text{c) } 5.8192$$

$$7.10.42 \ 2y^2$$

$$7.10.46 \ N[1 - e^{10/N}]$$

$$7.10.55 \ 218$$

$$7.10.60 \ 1/2, 1/3$$

$$7.10.5 \ [1 - (1-p)^N]/p$$

$$7.10.15 \ 4$$

$$7.10.20 \ 147/110$$

$$7.10.26 \ 14$$

$$7.10.31 \ 1/8$$

$$7.10.43 \ y^3/4$$

$$7.10.47 \ 12.5$$

$$7.10.57 \ x[1 + (2p-1)^2]^n$$

$$7.10.62 \ 1/i, [i(i+1)]^{-1}, \infty$$

Chapitre 8

$$8.8.1 \ p \geq 19/20$$

$$8.8.3 \ n \geq 3$$

$$8.8.5 \ 0.1416$$

$$8.8.8 \ 0.6932$$

$$8.8.2 \ \text{a) } 15/17 \ \text{b) } p \geq 3/4 \ \text{c) } n \geq 10$$

$$8.8.4 \ \text{a) } p \leq 4/3 \ \text{b) } 0.8438$$

$$8.8.6 \ 0.9431$$

$$8.8.9 \ 66.564$$

$$8.8.7 \ 0.3085$$

$$8.8.10 \ 117$$

8.8.11 $p \geq 0.6$

8.8.13 a) 0.0162 b) 0.0003 c) 0.2514 d) 0.2514

8.8.14 $n \geq 23$

8.8.15 0

8.8.17 $p \leq 0.2$

8.8.22 a) 0.769 b) 0.357 c) 0.1093 d) 0.112184

Chapitre 9

9.5.1 $1/9, 5/9$

9.5.3 0.0265, 0.0902, 0.2642, 0.4422

9.5.10 b) $1/6$

9.5.14 2.585, 0.5417, 3.1267

9.5.15 5.5098

Index

A

- Analyse combinatoire 1
- Antithétiques (variables) 439
- Aléatoire(s)
 - (marche) 300, 407
 - (sous-ensembles) 251-253, 428
 - (nombres) 426
- Approximation
 - binomiale d'une loi hypergéométrique 161
 - normale d'une loi binomiale 204
 - poissonienne d'une loi binomiale 145
- Associativité de la réunion et de l'intersection d'événements 27
- Axiomes des probabilités 28-29
- Axiomes de la théorie de l'information 410-411

B

- Banach (problème des allumettes de) 157
- Bayes (formule de) 68
 - formule de
 - généralisée 72
- Bernoulli
 - Daniel 296
 - Jacob 375
 - Jacques 86, 134, 142
 - Nicolas 142, 375
 - (épreuves de) 99

- (loi de) 134
- (espérance d'une variable de) 139
- Bertrand (paradoxe de) 194
- Bêta (loi) 219-220, 227, 265, 275
- Binôme (théorème du) 8
- Binomiale
 - (loi) 134, 260, 262, 278
 - approximation
 - d'une variable hypergéométrique 161
 - (approximation normale d'une variable) 204
 - (approximation poissonienne d'une variable) 145
 - (espérance d'une variable) 139, 293
 - (fonction génératrice des moments d'une variable) 334-335, 338
 - (calcul de la fonction de répartition) 142-143, 181
 - (simulation d'une variable) 437-438
 - (variance d'une variable) 139
 - loi - négative 156-157
 - loi - négative, relation avec la loi binomiale 167
 - loi - négative, relation avec la loi géométrique 157
- Bit 412
- Bivariée (distribution normale) 280, 332, 356
- Bonferroni (inégalité de) 50
- Boole (inégalité de) 50, 292-293

- Borel 384
 Bornes de Chernoff 388-390
 Box-Muller (approche de) 433
 Buffon (problème de l'aiguille de) 248-249, 277
- C**
- Canal binaire symétrique 419
 Capacité d'un canal de transmission 420
 Cauchy (loi de) 218-219
 Cauchy-Schwarz (inégalité de) 353
 Central limite (théorème) 371, 375
 Chaîne de Markov 404
 - ergodique 408
 - probabilités limites pour une - ergodique 408-409
 Chaînes de résultats identiques 43, 52, 92, 298
 Chapman-Kolmogorov (équations de) 406
 Chernoff (bornes de) 388-390
 Chi-carré (loi du) 217, 258, 277, 341, 346 (simulation d'une variable) 435
 Codage (théorie du) 414
 - théorème du - avec bruit 420
 - théorème du - sans bruit 416
 Coefficient
 - binomial 8
 - multinomial 10-11
 Combinaison 6
 Commutativité de la réunion et de l'intersection d'événements 27
 Conditionnelle (densité) 262 (distribution) 260-261, 263-264 (espérance) 316-319 (fonction de répartition) 263 (formule de la variance) 325-326, 440 (loi de probabilité) 260 (probabilité) 61 (utilisation pour la prédiction de l'espérance) 328 (variance) 325-326
 Conditionnellement, événements - indépendants 95
 Conjointe (densité) 239, 243 (fonction de répartition) 235, 244 (loi de probabilité) 236
 Conjointement, variables - continues 239, 244
 Conjoints (fonction génératrice des moments) 343-344
 Continues (variables aléatoires) 183 (variables conjointement) 239, 244
 Continuité, propriété de - des probabilités 44-45
 Convexité de fonction 390
 Convolution 255-256
 Corrélacion 312
 Corrélées (variables non) 313
 Covariance 305-306
 - formule de la - conditionnelle 352
- D**
- De Méré 81
 Demi-vie 253-255
 De Moivre 195, 207, 377
 - théorème limite de De Moivre-Laplace 204
 De Morgan (lois de) 27
 Densité 183
 - conditionnelle 262
 - conjointe ou simultanée 239, 244
 Dépendantes (variables aléatoires) 248
 Dépendants (événements) 75-76
 Déviation 308, 314
 Diagramme de Venn 26
 Discrètes (variables aléatoires) 123
 Distributions particulières: voir sous leur nom spécifique
 Distribution
 - conditionnelle 260-261, 263-264
 - marginale 238
 Distributivité de la réunion et de l'intersection d'événements 27
- E**
- Ecart-type 133, 394
 Echantillon tiré d'une population finie 310
 Echantillonnage par importance 445
 Ehrenfest (modèle des urnes d') 405, 409
 Ensemble fondamental 23
 Entropie 410
 - relation à la théorie du codage 414-420
 Epreuves 78
 - de Bernoulli 99
 Equations de Chapman-Kolmogorov 406
 Ergodique (chaîne de Markov) 408

- Eriang (loi d') 217
- Espérance mathématique 126, 187, 289, 347-348
- comme centre de gravité 128
 - conditionnelle 316-319
- définition générale
- d'une fonction de variable aléatoire 128-129, 188-189
 - d'une somme de variables aléatoires 290
 - de variable de Bernoulli 134
 - de variable binomiale 134, 293
 - de variable binomiale négative 158, 294
 - de variable exponentielle 208
 - de variable gamma 217
 - de variable géométrique 155
 - de variable hypergéométrique 161-162, 294
 - de variable normale 197-198
 - de variable poissonnienne 147
 - de variable uniforme 193
 - du nombre de rencontres 295
 - de la somme d'une quantité aléatoire de variables aléatoires 320
- (tables récapitulatives des) 338, 339
- Estimation du maximum de vraisemblance 160, 165
- Etendue d'un échantillon aléatoire 269
- Événement(s) 24
- complémentaires 25
 - conditionnellement indépendants 95
 - dépendants 75-76
 - élémentaire 35
 - indépendants 75-76, 77-78
- (intersection d') 25
- limite 45
 - mutuellement exclusifs 25
 - totalement indépendants 77-78
 - vide 25
- Exclusifs (événements mutuellement) 25
- Exponentielle 208
- (loi) 208
- (espérance de variable) 208
- (variance de variable) 208-209
- (fonction génératrice des moments de variable) 336
- (absence de mémoire d'une variable) 210
- variable -, relation avec la loi gamma 258
- relation avec la demi-vie 253-255
- loi - double 212
- (simulation d'une variable) 429
- F**
- Factorielle 3
- Fermat 82, 86
- Fonction de variable aléatoire 221
- Fonction génératrice des moments 333, 340
- conjoints 343-344
 - de variable binomiale 334-335, 338
 - de variable exponentielle 336
 - de variable normale 336-337, 339
 - de variable poissonnienne 335, 338
 - d'une quantité aléatoire de variables aléatoires 342-343
- (tables des) 338, 339
- Fonction de répartition 120
- conditionnelle 263
 - marginale 236,
 - simultanée ou conjointe 235, 244
 - simultanée de n variables 244
- Formule de Bayes 68
- généralisée 72
- Formule
- des probabilités totales 67
 - des probabilités totales généralisées 72
 - de Stirling 141
 - de la variance conditionnelle 325, 326, 440
- Fréquence relative et définition des probabilités 28-29
- G**
- Gamma (loi) 216, 257, 274
- relation avec la loi chi-carré 217, 258
 - relation avec la loi exponentielle 258
 - relation avec le processus de Poisson 216-217
- (simulation de variable) 429
- Gauss 207
- Générateur de nombres aléatoires 426
- Géométrique (loi) 154
- (simulation de variable) 436-437
- H**
- Huyghens 82, 86
- Hypergéométrique (loi) 159, 278, 316-317
- relation avec la loi binomiale 163
- I**
- Incertitude 412
- Indépendantes (variables aléatoires) 245, 250

- Indépendants**
 (événements) 75-78
 (événements conditionnellement) 95
- Inégalité**
 - de Bonferroni 50
 - de Boole 50, 52
 - de Cauchy-Schwarz 353
 - de Jensen 390
 - de Markov 371
 - de Tchebychev 372
- Information** 412
- Intégrale de Stieltjes** 347-348
- Intersection d'événements** 25
- Intervalles d'attente poissoniens** 403
- J**
- Jensen (inégalité de)** 390
- Jeux (théorie des)** 172, 279
- K**
- Khintchine** 375
- Kolmogorov** 384
- L**
- Laplace** 212, 377, 381
 (règle de succession de) 96, 101
- Laplacienne (distribution)** 212
- Legendre (théorème de)** 227
- Liapounoff** 377
- Limite**
 - d'événements 45
 - supérieure d'événements 52
 probabilité - d'une chaîne de Markov
 ergodique 408
- Log-normale (loi)** 227
- Lois particulières: voir sous leur nom
 spécifique**
- Lois de De Morgan** 27
- Loi des grands nombres** 31, 371
 (loi faible des) 374, 384
 (loi forte des) 31, 382, 384, 409, 426
- Loi de probabilité** 123
 - conditionnelle 260
 - simultanée ou conjointe 236
- Markov (chaîne de)** 404
 (inégalité de) 371
- Matrice de transition** 405-406
- Maximum de vraisemblance (estimation du)**
 160, 165
- Médiane d'un échantillon** 268, 280
 - d'une variable aléatoire 225, 349
- Mémoire (variable aléatoire sans)** 210
- Mendel** 136
- Méthode**
 - de rejet 429-430
 - de la transformation inverse 428
 - probabiliste 89
- Mode d'une variable aléatoire** 225
- Modèle des urnes d'Ehrenfest** 405
- Moments**
 - par rapport à l'origine 132
 (fonction génératrice des) 333, 340
 (fonction génératrice des) - conjoints
 343-344
- Montmort (problème de rencontre de)** 40, 51,
 94, 295, 309
- Moyenne d'échantillon** 292, 345-347
- Multinôme (théorème du)** 11
- Multinomiale (loi)** 244, 315
- Multinomiaux (coefficients)** 10-11
- N**
- Normale**
 (loi) 195
 approximation
 - de la loi binomiale 204
 loi - bivariée 280, 332, 356
 loi - multivariée 344
 (caractérisation de la loi) 249
 loi - standard ou centrée réduite 199
 (espérance de variable) 197-198
 (fonction génératrice des moments de
 variable) 336-337, 339
 (simulation de variable) 431, 433
 tables de la loi-standard 200
 (variance de variable) 197-198
- O**
- Ordre (statistiques d')** 265-266
- P Q**
- Paradoxe**
 - de Bertrand 194

- de St-Petersbourg 174
- Pareto (loi de) 163
- Partition 50
- Pascal 81-82
- Permutation 3
 - aléatoire 427, 442-443
- Pics 351
- Plus grand en probabilité 350
- Poisson 144
 - (loi de) 42, 144, 246, 259, 262, 277, 310
 - approximation de - d'une loi binomiale 145
 - (espérance de variable de) 147
 - (fonction génératrice des moments de variable de) 335, 338
 - (simulation de variable de) 437-438
 - (variance de variable de) 147
 - (calcul de la fonction de répartition de) 153, 181
- Poker 36
- Prédicteur 328, 355
 - linéaire 331, 355
- Prédiction (utilisation de l'espérance conditionnelle pour la) 328
- Principe fondamental de dénombrement 2
- Probabilité
 - subjective 48
 - de transition 404
 - de transition d'ordre n 406
- Problème
 - de l'aiguille de Buffon 248-249, 277
 - des anniversaires 38, 148-149
 - des allumettes de Banach 157
 - de collecte de coupons 118-120, 296-297, 351
 - de la durée du jeu 86
 - d'élection 100
 - du paradoxe 46
 - des points 81-82
 - de rencontre de Montmort 40, 51, 94, 295, 309
 - de la roue de la fortune 136
 - de la ruine du joueur 83, 172
- Processus
 - de Poisson 150-152, 401
 - de ramification 357
- Propriété de continuité des probabilités 44-45
- Pseudo-aléatoires (nombres) 426

R

- Ramification (processus de) 357
- Rapport signal-bruit 394

- Rayleigh (loi de) 214, 273
- Règle de multiplication 65
- Règle de succession de Laplace 96, 101
- Rejet (méthode de) voir simulation
- Réunion d'événements 24
- Riemann
 - (fonction zéta de) 163

S

- St-Petersbourg (paradoxe de) 174
- Shannon 420
- Simulation 425
 - méthode de rejet 429-430
 - méthode des coordonnées polaires 433, 435
- Simultanée
 - (densité) 239, 243
 - (fonction de répartition) 235, 244
 - (loi de probabilité) 236
- Somme de variables aléatoires
 - (espérance de la) 290
- Sous-ensemble aléatoirement choisi 251-253
- Statistique d'ordre 265-266
- Stieltjes (intégrale de) 347-348
- Stirling
 - (formule de) 141
 - (approximation de) 39, 380
- Subjective, approche - des probabilités 48
- Surprise 412

T

- Taux
 - de variable aléatoire exponentielle 214
 - de panne 213
 - de mortalité 215
 - de transmission 422
- Tchebychev
 - (inégalité de) 372
 - (inégalité unilatérale de) 385
- Temps d'attente poissonniens 402-403
- Théorème
 - du binôme 8
 - central limite 375
 - central limite étendu 381
 - du codage avec bruit 420
 - du codage sans bruit 416
 - de Legendre 227
 - limite de De Moivre-Laplace 204
 - de Weierstrass 395

Théorie

- du codage 414
- du codage en relation à l'entropie 414-420
- de l'hérédité 136
- de l'information 412
- des jeux 172, 279

Traitement de signaux 329-330**Transition**

- (matrice de) 405
- matrice de - d'ordre n 406
- (probabilité de) 404

U**Uniforme (loi) 191-192****V****Variable aléatoire 115**

- continue 183
- discrète 123

Variance

- comme moment d'inertie 133
- conditionnelle 325-326
- formule de la - conditionnelle 325-326
- d'une variable binomiale 139, 309
- d'une variable binomiale négative 158

- d'une variable exponentielle 208-209
- d'une variable géométrique 156, 322
- d'une variable hypergéométrique 162
- d'une variable normale 197-198
- d'une variable de Poisson 147
- d'une variable uniforme 193
- d'une variable 132, 190-191
- d'échantillon 308, 345-347
- du nombre de rencontres 309
- de la somme d'une quantité aléatoire de variables aléatoires 327
- d'une somme de variables aléatoires 307
- (tables récapitulatives de) 338, 339

Venn (diagramme de) 26**Von Neumann 317****W****Weibull (loi de) 218, 226, 444**

relation avec la loi exponentielle 226

Weierstrass (théorème de) 395**X Y Z****Zéta**

(loi) 163

fonction - de Riemann 163

Zipf (loi de) 163