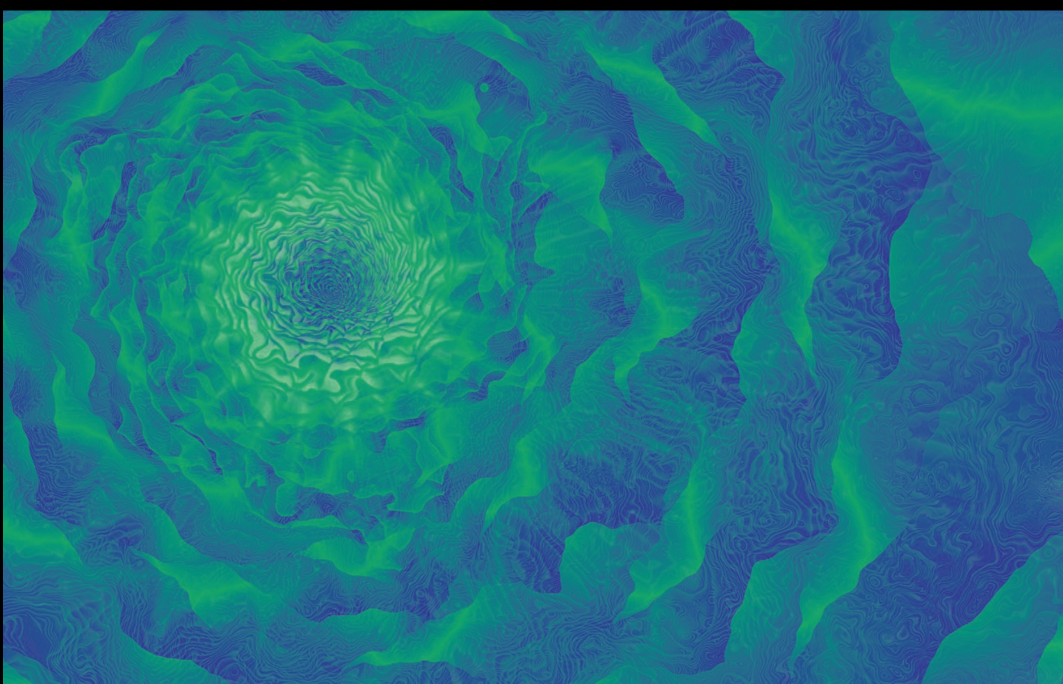
Enseignement des mathématiques

Introduction à la théorie des probabilités

Deuxième édition revue et augmentée

Robert C. Dalang et Daniel Conus



Presses polytechniques et universitaires romandes

Introduction à la théorie des probabilités

Enseignement des mathématiques

Introduction à la théorie des probabilités

Deuxième édition revue et augmentée

Robert C. Dalang et Daniel Conus

Presses polytechniques et universitaires romandes

Les auteurs et l'éditeur remercient l'Ecole polytechnique fédérale de Lausanne (EPFL)
pour le soutien apporté à la publication de cet ouvrage.

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

Calcul différentiel et intégral
Fonctions réelles d'une ou de plusieurs variables réelles
Jacques Douchet et Bruno Zwahlen

Cours d'analyse, 3 volumes
Srishti D. Chatterji

Algèbre linéaire
Robert C. Dalang et Amel Chaabouni
Aide-mémoire, exercices et applications

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

Analyse avancée pour ingénieurs
Bernard Dacorogna, Chiara Tanteri

Introduction à l'analyse numérique
Jacques Rappaz et Marco Picasso

Introduction à l'optimisation différentiable
Michel Bierlaire

Initiation aux probabilités
Sheldon M. Ross
Traduction de la neuvième édition américaine

Théorie des probabilités
Cours d'introduction avec application à la statistique mathématique
Charles-Edouard Pfister

Introduction à la statistique
Stéphan Morgenthaler

Calcul différentiel et intégral
Fonctions réelles d'une ou de plusieurs variables réelles
Jacques Douchet et Bruno Zwahlen

Recherche opérationnelle pour ingénieurs I et II
Jean-François Hêche, Thomas M. Liebling, Dominique de Werra

Introduction à l'optimisation différentiable
Michel Bierlaire

Algèbre linéaire
Aide-mémoire, exercices et applications
Robert C. Dalang et Amel Chaabouni

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

Analyse avancée pour ingénieurs
Bernard Dacorogna, Chiara Tanteri

Introduction à l'analyse numérique
Jacques Rappaz et Marco Picasso

La Fondation des Presses polytechniques et universitaires romandes (PPUR) publie principalement les travaux d'enseignement et de recherche de l'Ecole polytechnique fédérale de Lausanne (EPFL), des universités et des hautes écoles francophones.
PPUR, EPFL-Rolex Learning Center, CP 119, CH-1015 Lausanne,
ppur@epfl.ch, tél.: +41 21 693 21 30; fax: +41 21 693 40 27.

www.ppur.org

Deuxième édition revue et augmentée

ISBN 978-2-88915-148-6
CH – 1015 Lausanne

Imprimé en France

Avant-propos

Cet ouvrage est une première introduction à la théorie des probabilités. Il présente donc les concepts de base de la théorie, en particulier ceux liés aux espaces de probabilités finis et aux variables aléatoires discrètes et continues. Leurs fonctions de répartition et de densité, les principales lois de probabilités, la notion d'espérance et d'espérance conditionnelle sont introduites sans faire appel à la théorie de la mesure.

Il existe de nombreux et excellents ouvrages d'introduction au *calcul* des probabilités, notamment les livres de N. Bouleau [2], de R. Durrett [4], de G. Grimmett et D. Stirzaker [6], de S. Morgenthaler [7], de S. Ross [8], de M. Sanz-Solé [9] et de D. Stirzaker [10]. Ces ouvrages de base sont des sources d'inspiration et des sources d'exercices. Si notre ouvrage traite les mêmes thèmes principaux que ces références, il s'en distingue par le fait qu'il vise un public d'étudiants intéressés non seulement par le calcul des probabilités, mais aussi par les fondements mathématiques de cette théorie. Il suppose que l'étudiant a déjà suivi deux semestres d'analyse et d'algèbre linéaire et a acquis une bonne maîtrise de ces matières. En général, l'étudiant suit en parallèle un troisième semestre d'analyse et un cours de topologie générale. Notre objectif est donc de présenter la théorie des probabilités avec un niveau de rigueur comparable à ceux de ces derniers cours et de proposer un certain nombre d'exercices avec un contenu théorique substantiel. Nous n'introduisons cependant aucune des techniques de la théorie de la mesure, contrairement par exemple à M. Sanz-Solé [9], qui en fait un usage modéré.

Tout d'abord, nous insistons sur les notions d'espace de probabilité et de tribu, qui sont devenues plus importantes ces dernières années vu par exemple leur utilisation en mathématiques financières. Si nous n'abordons pas ces applications, nous présentons plusieurs situations où ces notions jouent un rôle important, par exemple dans la solution du problème de Monty-Hall (exercice 3.10). Dès le chapitre 1, nous proposons à l'étudiant quelques exercices concernant des événements à l'interprétation non élémentaire, ce qui permet par exemple de leur faire démontrer la première partie du lemme de Borel-Cantelli (la deuxième partie est traitée au chapitre 3).

Ensuite, dans le cadre de l'étude des variables aléatoires (absolument) continues et des vecteurs aléatoires continus, nous cherchons à donner des démonstrations rigoureuses des principaux résultats, même lorsque la plupart des autres ouvrages se contentent de renvoyer à la littérature plus avancée. Par exemple, dans le chapitre 4, nous proposons aux étudiants de montrer que toute fonction de répartition est une combinaison convexe d'une fonction de répartition discrète et d'une fonction de répartition continue. Cet effort de rigueur nécessite parfois d'introduire des hypothèses légèrement plus fortes que nécessaires, et certains compromis restent indispensables. Dans le chapitre 6, nous donnons une démonstration de la formule pour l'espérance d'une fonction d'un vecteur aléatoire (théorème 6.10), ce qui permet d'établir de manière satisfaisante la linéarité de l'espérance. Nous proposons aussi une démonstration élémentaire du fait que la fonction génératrice des moments d'une variable aléatoire détermine sa fonction de répartition. L'exemple le plus important de notre effort de rigueur est notre démonstration complète mais élémentaire du théorème limite central, sous l'hypothèse de l'existence du troisième moment. Malgré son caractère rigoureux, cette démonstration n'occupe qu'une heure de cours environ ! Afin d'être complets, nous proposons aussi une démonstration de la loi forte des grands nombres. Finalement, dans le chapitre 8, nous introduisons la notion d'espérance conditionnelle comme une projection orthogonale sur un sous-espace vectoriel de variables aléatoires, en prenant la liberté de ne pas parler de fonctions boréliennes.

Quelques sections sont marquées d'un astérisque. Elles contiennent des informations complémentaires et peuvent être omises en première lecture. Quelques exercices sont utilisés ultérieurement dans la partie théorique (exercices 1.7, 1.9, 5.9 et 6.29).

Le lecteur sera peut-être étonné du petit nombre d'exemples qui sont présentés dans le texte. Ceci est dû au fait que cet ouvrage est conçu comme un support de cours et est compensé par la présence de nombreux exercices. Sauf pour les exercices marqués d'une étoile (qui peuvent servir pour un travail personnel noté), un corrigé complet est donné à la fin du document. L'enseignant peut donc puiser dans les exercices les exemples qu'il souhaite présenter au cours et sélectionner chaque semaine les exercices qu'il recommande aux étudiants pour leur permettre de mieux assimiler la théorie. Les étudiants auront tout intérêt à tenter de résoudre les exercices sans se tourner vers le corrigé et de n'utiliser celui-ci que pour vérifier qu'ils ont bien obtenu la bonne solution. Le chapitre 9 contient trente exercices donnés lors de travaux écrits ou d'examens et permet à l'étudiant de se préparer pour ces épreuves.

La matière théorique, les exercices et les solutions présentés dans la première édition de cet ouvrage ont été affinés au cours de quatre années d'enseignement à la Section de mathématiques de l'Ecole Polytechnique Fédérale de Lausanne (EPFL). Dans cette deuxième édition, plusieurs nouveaux exercices ont été ajoutés et de nombreuses corrections ont été apportées. La transcription en LaTeX des chapitres 1 à 8 a été réalisée par Madame Erika Gindraux que nous remercions vivement.

R. C. Dalang et D. Conus

Janvier 2015

Table des matières

Conventions

Les conventions et notations suivantes sont utilisées dans cet ouvrage.

Ensembles de nombres

$\mathbb{N}$	$\{0, 1, 2, \ldots\}$, ensemble des entiers naturels
$\mathbb{N}^*$	$\{1, 2, \ldots\}$
$\mathbb{Z}$	$\{\ldots, -2, -1, 0, 1, 2, \ldots\}$, ensemble des entiers relatifs
$\mathbb{Z}_+$	$\mathbb{N}$
$\mathbb{Q}$	ensemble des nombres rationnels
$\mathbb{R}$	ensemble des nombres réels
$\mathbb{R}^*$	$\mathbb{R} \setminus \{0\}$
$]a, b[$	$\{x \in \mathbb{R} : a < x < b\}$
$[a, b]$	$\{x \in \mathbb{R} : a \leqslant x \leqslant b\}$
$]a, b]$	$\{x \in \mathbb{R} : a < x \leqslant b\}$
$\mathbb{R}_+$	$[0, +\infty[$, ensemble des nombres réels non négatifs
$\mathbb{R}_+^*$	$]0, +\infty[$, ensemble des nombres réels positifs

Notations ensemblistes

$\emptyset$	ensemble vide
$x \in E$	x appartient à l'ensemble E
$E \subset F$	E est un sous-ensemble de F
$E \cap F$	intersection de E et F
$E \cup F$	réunion de E et F
E^c	complémentaire de E
$E \setminus F$	$E \cap F^c$
$E \times F$	produit cartésien de E et F
$\mathcal{P}(E)$	ensemble des parties de l'ensemble E
$\sharp\,\Omega$ ou $\lvert\Omega\rvert$	cardinalité de Ω

Notations générales

$\det A$	déterminant de la matrice A
$\mathrm{Im}(g)$	ensemble image de la fonction g
$g^{-1}(E)$	préimage de l'ensemble E par la fonction g
$[x]$	partie entière de x, c.-à-d. $[x] = k$ si et seulement si $x \in [k, k+1[$ $(k \in \mathbb{Z})$

Chapitre 1

Espaces de probabilité

Dans ce chapitre, nous présentons d'abord le modèle mathématique qui sert de base à toute la théorie des probabilités, puis nous examinons de plus près les espaces de probabilité finis.

1.1 Evénements

Avant d'introduire cette première notion, examinons d'abord dans quels contextes il est naturel d'utiliser la théorie des probabilités.

Champ d'application de la théorie des probabilités

La théorie des probabilités vise à analyser tous les phénomènes ou expériences qui ont plusieurs résultats possibles, non connus à l'avance. Seul l'ensemble des résultats possibles, souvent noté Ω et appelé *ensemble fondamental* ou *univers*, est connu à l'avance. C'est uniquement à la fin de l'expérience que l'observateur saura quel en est le résultat $\omega \in \Omega$.

Exemple 1.1

a) Une expérience consiste à tirer une boule d'une urne contenant douze boules, numérotées de 1 à 12. L'ensemble fondamental est $\Omega = \{1, 2, \ldots, 12\}$.

b) Une expérience consiste à choisir un nombre réel au hasard. Dans ce cas, l'ensemble fondamendal est $\Omega = \mathbb{R}$.

c) Une expérience consiste à lancer une pièce de monnaie jusqu'à obtenir face. Dans ce cas, l'ensemble fondamendal est

$$\Omega = \{\mathrm{F}, \mathrm{PF}, \mathrm{PPF}, \mathrm{PPPF}, \ldots\}$$

où F désigne le résultat « face » et P désigne « pile ».

Fonction de probabilité

Dans l'exemple 1.1a), on peut s'intéresser par exemple aux propriétés suivantes : « la boule tirée porte un numéro pair » ou « la boule tirée porte un numéro inférieur ou égal à 5 ». La première propriété correspond à un sous-ensemble de Ω constitué de 6 boules, à savoir $G = \{2, 4, 6, 8, 10, 12\}$, alors que le second correspond au sous-ensemble $H = \{1, 2, 3, 4, 5\}$ formé de cinq boules. Dans l'exemple 1.1b), on peut s'intéresser à la propriété « le nombre choisi est positif », ce qui correspond au sous-ensemble $\mathbb{R}_+^*$.

Dans ce livre, on veut attribuer une probabilité à ces sous-ensembles, c'est-à-dire qu'on veut définir une fonction

$$P : \mathcal{P}(\Omega) \to \mathbb{R}\,,$$

où $\mathcal{P}(\Omega)$ est l'ensemble de tous les sous-ensembles de Ω. En général, il n'est pas nécessaire que le domaine de définition de la fonction P soit $\mathcal{P}(\Omega)$ tout entier, c'est-à-dire qu'on considérera

$$P : \mathcal{F} \to \mathbb{R}\,,$$

où $\mathcal{F} \subset \mathcal{P}(\Omega)$. On postulera cependant que le domaine de définition $\mathcal{F}$ est une *tribu*, notion que nous définissons maintenant.

Tribu

Définition 1.2

Une tribu *(ou σ-*algèbre*) sur Ω est une famille $\mathcal{F}$ de sous-ensembles de Ω qui possède les trois propriétés suivantes :*

a) $\Omega \in \mathcal{F}$;

b) si $F \in \mathcal{F}$, alors $F^c \in \mathcal{F}$, où F^c désigne le complémentaire de F ;

c) si $(F_n,\, n \geqslant 1)$ est une suite d'éléments de $\mathcal{F}$, alors $\bigcup_{n=1}^{\infty} F_n \in \mathcal{F}$.

REMARQUE 1.3 Si c) est remplacé par c′) : si $F_1 \in \mathcal{F}$ et $F_2 \in \mathcal{F}$, alors $F_1 \cup F_2 \in \mathcal{F}$, alors $\mathcal{F}$ est appelé une *algèbre* sur Ω.

Les éléments de la tribu $\mathcal{F}$ sont appelés des *événements.* Un singleton $\{\omega\}$ avec $\omega \in \Omega$ est appelé un *événement élémentaire.* L'ensemble Ω est appelé *l'événement certain.* L'ensemble vide $\emptyset$ est appelé *l'événement impossible.* L'exemple de base d'une tribu est $\mathcal{F} = \mathcal{P}(\Omega)$.

Si G est un événement et si, à la fin de l'expérience, on constate que $\omega \in G$, alors on dit que l'événement G est *réalisé.*

Evénements observables

Certaines expériences se déroulent en plusieurs étapes. Dans une phase intermédiaire d'une telle expérience, un événement G est dit *observable* si, pour tout $\omega \in \Omega$, on peut décider si $\omega \in G$ ou non. Divers exemples sont donnés dans les exercices 1.11 à 1.14.

1.2 Axiomes des probabilités

Dans l'exemple 1.1a), si la boule est tirée « au hasard », on aimerait dire que l'événement « la boule numéro 1 est tirée » a la même probabilité que les autres événements élémentaires, ou qu'il est plus probable de tirer une des 6 boules portant un numéro pair que la boule portant le numéro 1. Afin que les probabilités attribuées correspondent à ce qu'exige notre intuition, on postule qu'elles satisfont à certains axiomes, rassemblés dans la définition suivante.

Définition 1.4

a) Une fonction de probabilité *(ou* mesure *de probabilité) sur* $(\Omega, \mathcal{F})$ *est une application* $P : \mathcal{F} \to [0, 1]$ *qui vérifie les deux axiomes suivants :*

Axiome 1 *(événement certain).* $P(\Omega) = 1$.

Axiome 2 *(σ-additivité). Si* $(F_n,\ n \geqslant 1)$ *est une suite d'événements (dans* $\mathcal{F}$*) qui sont deux à deux disjoints (c'est-à-dire que* $F_n \cap F_m = \emptyset$ *si* $n \neq m$*), alors*

$$P\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} P(F_n)\,.$$

b) Un espace de probabilité *est un triplet* $(\Omega, \mathcal{F}, P)$*, où* Ω *est un ensemble non vide,* $\mathcal{F}$ *est une tribu sur* Ω *et* P *est une probabilité sur* $(\Omega, \mathcal{F})$.

Interprétation de la fonction de probabilité

L'interprétation la plus commune de la fonction de probabilité est *l'interprétation fréquentiste* : on imagine que l'expérience est répétée n fois et on note $n(G)$ le nombre de fois où l'événement G est réalisé. On souhaite que

$$P(G) = \lim_{n \to \infty} \frac{n(G)}{n}\,,$$

c'est-à-dire que la probabilité de G est la valeur limite de la proportion d'expériences où G est réalisé. Ainsi, dans l'exemple 1.1a), poser

$$P\big(\{1\}\big) = \frac{1}{12}$$

revient à dire que sur 12 000 réalisations de l'expérience, on tirera la boule numérotée 1 environ 1000 fois, ce qui est une hypothèse naturelle.

Une deuxième interprétation courante est celle de la *probabilité subjective* : on interprète la probabilité comme la mesure de la vraisemblance d'un fait. Ainsi, dire que « la probabilité que la Mona Lisa représente Léonard De Vinci est 60 % » signifie en particulier qu'on pense que cette affirmation est plus probable que l'affirmation contraire. Il faut cependant faire attention à ce que le choix des probabilités ainsi attribuées vérifie bien les deux axiomes de la définition.

Premières propriétés des probabilités

Proposition 1.5

a) Probabilité de l'événement impossible. *$\emptyset \in \mathcal{F}$ et $P(\emptyset) = 0$.*

b) Additivité finie. *Soient $m \geqslant 2$ et $G_1 \in \mathcal{F}, \ldots, G_m \in \mathcal{F}$ des événements deux à deux disjoints. Alors*

$$P(G_1 \cup \cdots \cup G_m) = P(G_1) + \cdots + P(G_m)\,.$$

c) Probabilité de l'événement complémentaire. *$P(G^c) = 1 - P(G)$.*

d) Monotonie des probabilités. *Soient $G, H \in \mathcal{F}$ tels que $G \subset H$. Alors $P(G) \leqslant P(H)$.*

e) Probabilité d'une réunion. *$P(G \cup H) = P(G) + P(H) - P(G \cap H)$.*

DÉMONSTRATION. a) Observons que $\Omega \in \mathcal{F}$ et $\emptyset = \Omega^c$, donc $\emptyset \in \mathcal{F}$. Soit $G_n = \emptyset$ si $n \geqslant 1$. Par σ-additivité (qui s'applique puisque les G_n sont deux à deux disjoints),

$$P\left(\bigcup_{n=1}^{\infty} G_n\right) = \sum_{n=1}^{\infty} P(G_n)\,.$$

Ainsi,

$$1 \geqslant P(\emptyset) = \sum_{n=1}^{\infty} P(\emptyset)\,,$$

donc $P(\emptyset) = 0$ puisque $P(\emptyset) \geqslant 0$.

b) Posons $G_n = \emptyset$ pour $n > m$, puis appliquons la σ-additivité à $(G_n,\, n \geqslant 1)$, pour obtenir

$$P\left(\bigcup_{n=1}^{\infty} G_n\right) = \sum_{n=1}^{\infty} P(G_n)$$

ou encore

$$\begin{aligned} P(G_1 \cup \cdots \cup G_m) &= P(G_1) + \cdots + P(G_m) + 0 + 0 + \cdots \\ &= P(G_1) + \cdots + P(G_m)\,. \end{aligned}$$

c) Remarquons que $\Omega = G \cup G^c$, donc $1 = P(\Omega) = P(G \cup G^c) = P(G) + P(G^c)$ d'après la propriété b).

d) D'après b), $P(H) = P\big(G \cup (H \cap G^c)\big) = P(G) + P(H \cap G^c) \geqslant P(G)$ puisque $P(H \cap G^c) \geqslant 0$.

e) Observons, à l'aide par exemple du *diagramme de Venn* de la figure 1.1, que $G = (G \cap H) \cup (G \cap H^c)$ et $G \cup H = (G \cap H^c) \cup H$, donc

$$\begin{aligned} P(G) + P(H) &= P(G \cap H) + P(G \cap H^c) + P(H) \\ &= P(G \cap H) + P(G \cup H)\,, \end{aligned}$$

d'où le résultat. $\square$

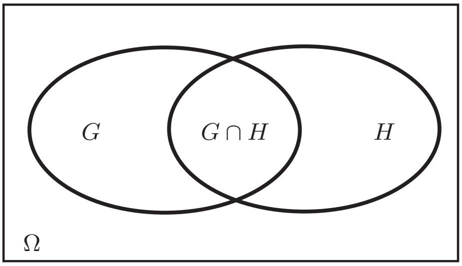


Figure 1.1 Représentation de deux événements par un diagramme de Venn.

Continuité de la probabilité

Une conséquence importante de l'axiome de σ-additivité des probabilités sont les deux propriétés suivantes de continuité le long de suites monotones d'événements.

Théorème 1.6

a) Soit $(G_n,\, n \in \mathbb{N})$ une suite d'événements tels que $G_n \subset G_{n+1}$, pour tout $n \in \mathbb{N}$. Alors

$$P\left(\bigcup_{n\in\mathbb{N}} G_n\right) = \lim_{n\to\infty} P(G_n)\,.$$

b) Soit $(G_n,\, n \in \mathbb{N})$ une suite d'événements tels que $G_n \supset G_{n+1}$, pour tout $n \in \mathbb{N}$. Alors

$$P\left(\bigcap_{n\in\mathbb{N}} G_n\right) = \lim_{n\to\infty} P(G_n)\,.$$

Pour la démonstration de ce théorème, voir l'exercice 1.7.

1.3 Espaces de probabilité finis

De nombreuses expériences ne comportent qu'un nombre fini de résultats possibles. On va donc étudier plus spécifiquement le cas où $\Omega = \{\omega_1, \ldots, \omega_r\}$ est un ensemble fini. Sauf mention du contraire, on prend aussi $\mathcal{F} = \mathcal{P}(\Omega)$.

Probabilités sur $(\Omega, \mathcal{F})$

De manière générale, il est intéressant de connaître toutes les fonctions de probabilité sur $\mathcal{F}$. Dans le cas où Ω est fini, nous allons décrire ces fonctions.

Posons $p_i = P\big(\{\omega_i\}\big)$ (pour simplifier la notation, on écrit plutôt $P\{\omega_i\}$). Alors $p_1, \dots, p_r$ sont des nombres réels tels que $p_i \geqslant 0$, $i = 1, \dots, r$ et $p_1 + \cdots + p_r = 1$.

La donnée des nombres p_i détermine la probabilité $P(G)$ de n'importe quel événement $G \in \mathcal{F}$. En effet, si $G = \{\omega_{i_1}, \dots, \omega_{i_k}\}$, alors d'après la propriété d'additivité finie,

$$P(G) = P\big(\{\omega_{i_1}\} \cup \cdots \cup \{\omega_{i_k}\}\big) = P\{\omega_{i_1}\} + \cdots + P\{\omega_{i_k}\} = p_{i_1} + \cdots + p_{i_k} \,.$$

Par conséquent,

$$P(G) = \sum_{j \colon \omega_j \in G} p_j \,. \tag{1.1}$$

Réciproquement, il est possible de se donner des nombres $p_1, \dots, p_r$ vérifiant $p_i \geqslant 0$, $i = 1, \dots, r$ et $p_1 + \cdots + p_r = 1$, de poser $P\{\omega_i\} = p_i$, puis de définir $P(G)$ par la formule (1.1) lorsque $G \neq \emptyset$ et de poser $P(\emptyset) = 0$. On a alors la proposition suivante.

Proposition 1.7 *La fonction* $P : \mathcal{F} \to [0,1]$ *ainsi définie est une probabilité sur* $(\Omega, \mathcal{F})$.

Démonstration. Nous allons vérifier les deux axiomes de la définition 1.4. Clairement,

$$P(\Omega) = \sum_{j \colon \omega_j \in \Omega} p_j = \sum_{j=1}^{r} p_j = 1 \,.$$

Quant à la σ-additivité, soit $(G_n,\, n \geqslant 1) \subset \mathcal{F}$ des événements deux à deux disjoints. Puisque Ω est un ensemble fini, seul un nombre fini de valeurs de n sont telles que $G_n \neq \emptyset$. Supposons que ce sont $G_1, \dots, G_k$. Alors

$$P\left(\bigcup_{n=1}^{\infty} G_n\right) = P(G_1 \cup \cdots \cup G_k) = \sum_{j \colon \omega_j \in G_1 \cup \cdots \cup G_k} p_j \,.$$

Puisque les G_j sont deux à deux disjoints, cette somme est égale à

$$\sum_{j \colon \omega_j \in G_1} p_j + \cdots + \sum_{j \colon \omega_j \in G_k} p_j = P(G_1) + \cdots + P(G_k) \,.$$

La σ-additivité est donc vérifiée. $\square$

1.4 Cas des événements élémentaires équiprobables

Un cas particulier important d'espace de probabilité fini est le cas où tous les événements élémentaires ont la même probabilité. Considérons Ω et $p_1, \dots, p_r$ comme dans la section 1.3. Dans cette section, nous allons donc faire l'hypothèse suivante.

Hypothèse. $p_1 = p_2 = \cdots = p_r$

La propriété $p_1 + \cdots + p_r = 1$ jointe à cette hypothèse, implique donc que $p_j = \frac{1}{r} = \frac{1}{\sharp\,\Omega}$, où $\sharp\,\Omega$ est le nombre d'éléments de Ω (parfois aussi notée $\operatorname{card}(\Omega)$ ou $|\Omega|$).

Il est maintenant facile de déterminer la probabilité d'un événement $G \in \mathcal{F}$. En effet,

$$P(G) = \sum_{j:\, \omega_j \in G} \frac{1}{r} = \frac{\sharp\,G}{r} = \frac{\sharp\,G}{\sharp\,\Omega}\,.$$

On retient souvent cette formule sous la forme

$$P(G) = \frac{\sharp\,G}{\sharp\,\Omega} = \frac{\text{nombre de cas « favorables » pour } G}{\text{nombre de cas « possibles »}}\,.$$

Le calcul de $P(G)$ se ramène donc au problème de compter (ou dénombrer) le nombre d'éléments de G et le nombre d'éléments de Ω.

Exemple 1.8

a) On lance une pièce de monnaie équitable deux fois de suite. Quelle est la probabilité que les deux résultats soient identiques ? Dans ce cas,

$$\Omega = \big\{(\mathrm{P},\mathrm{P}), (\mathrm{P},\mathrm{F}), (\mathrm{F},\mathrm{P}), (\mathrm{F},\mathrm{F})\big\}$$

et $\sharp\,\Omega = 4$. Soit G l'événement qui correspond à « les deux résultats sont identiques », c'est-à-dire $G = \big\{(\mathrm{P},\mathrm{P}), (\mathrm{F},\mathrm{F})\big\}$. Clairement, $\sharp\,G = 2$ et donc, en admettant que les quatre résultats élémentaires sont équiprobables,

$$P(G) = \frac{\sharp\,G}{\sharp\,\Omega} = \frac{2}{4} = \frac{1}{2}\,.$$

b) On tire 5 cartes d'un jeu de 52 cartes. Quelle est la probabilité d'obtenir exactement une paire ? Dans ce cas, la valeur de $\sharp\,\Omega$ ne peut être déterminée par énumération, car le nombre de mains de 5 cartes est trop grand. En fait, on verra plus loin que $\sharp\,\Omega = 2\,598\,960$ et que $\sharp$ « avoir une paire » $= 1\,098\,240$ donc $P(1 \text{ paire}) = 0{,}422569$.

Ce deuxième exemple est typique des situations que l'on rencontre pour calculer des probabilités dans le cadre des espaces de probabilités finis et montre la nécessité de disposer de techniques pour dénombrer (compter) le nombre d'éléments d'ensembles finis mais de très grande cardinalité. Ce sera l'objet du chapitre 2.

1.5 Complément : motivations à l'introduction des tribus*

Dans de nombreux exemples élémentaires, la tribu $\mathcal{F}$ de l'espace de probabilité $(\Omega, \mathcal{F}, P)$ est simplement $\mathcal{P}(\Omega)$. Cependant, il y a aussi de nombreuses situations où il est commode, voire nécessaire, de prendre $\mathcal{F} \subsetneq \mathcal{P}(\Omega)$. En voici quelques exemples.

Imaginons qu'une urne contient deux boules rouges R_1, R_2 et trois boules noires N_1, N_2, N_3, les boules de même couleur n'étant pas distinguables à l'œil nu. Dans l'interprétation fréquentiste de la notion de probabilité, l'expérimentateur qui tire successivement des boules avec remise ne pourra attribuer des probabilités qu'aux ensembles $F = \{\mathrm{R}_1, \mathrm{R}_2\}$ et $F^c = \{\mathrm{N}_1, \mathrm{N}_2, \mathrm{N}_3\}$ (ainsi qu'à $\emptyset$ et à $\Omega = \{\mathrm{R}_1, \mathrm{R}_2, \mathrm{N}_1, \mathrm{N}_2, \mathrm{N}_3\}$), puisqu'il ne saura jamais s'il a tiré R_1 plutôt que R_2, ni N_1 plutôt que N_3. Ainsi, il ne pourra définir la probabilité que sur la tribu $\mathcal{G} = \{\emptyset, F, F^c, \Omega\}$, qui apparaît donc de manière naturelle. Néanmoins, rien n'empêche l'observateur de décider que chaque singleton a la probabilité $\frac{1}{5}$ et ainsi de prolonger la définition de la probabilité à la tribu $\mathcal{P}(\Omega)$.

Dans l'exercice 2.22, on verra un autre exemple où l'observation de l'expérience ne permet pas de déterminer la probabilité de tous les sous-ensembles de Ω, même si, dans cet exemple, il y a un seul prolongement naturel de la probabilité à tous les sous-ensembles de Ω. Dans l'exercice 3.10, on sera confronté à une situation un peu différente : l'énoncé du problème ne permettra pas de déterminer la probabilité de tous les sous-ensembles et plusieurs prolongements de la probabilité seront raisonnables et conduiront d'ailleurs à des conclusions légèrement différentes !

Une deuxième raison qui motive l'introduction de la notion de tribu est de nature plus fondamentale. En effet, dans certains contextes et en particulier lorsque Ω est un ensemble infini non dénombrable, il n'est simplement pas possible de définir la probabilité de n'importe quel sous-ensemble de Ω tout en préservant les autres propriétés exigées pour une fonction de probabilité (sect. 4.6).

Finalement, certaines applications, en particulier en mathématiques financières, utilisent de manière naturelle la notion de tribu. Typiquement, il s'agit de distinguer les événements observables aujourd'hui (le fait que le prix de clôture d'une action dépasse 100 francs, par exemple) et les événements qui ne seront observables que demain (typiquement, le fait que le prix de clôture du lendemain dépasse 100 francs).

1.6 Exercices

1.1 Soit Ω un ensemble et soient E, F et G trois événements. Quel événement correspond à chacune des descriptions suivantes ?

a) E et F se réalisent, mais G ne se réalise pas ;

b) au moins l'un de ces événements se réalise ;

c) au plus deux des trois événements se réalisent ;

d) exactement un de ces événements est réalisé.

1.2 Soit Ω un ensemble et soient E, F et G trois événements. Quel événement correspond à chacune des descriptions suivantes ?

a) E est le seul événement qui se réalise ;
b) au moins deux de ces événements se réalisent ;
c) au plus l'un des trois événements se réalise ;
d) exactement deux de ces événements se réalisent.

1.3 (Tribu engendrée)

a) Soient Ω et I deux ensembles non vides. Pour $i \in I$, soit $\mathcal{F}_i$ une tribu sur Ω. Montrer que $\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i$ est une tribu sur Ω.
b) Soit Ω un ensemble non vide et soit $\mathcal{G} \subset \mathcal{P}(\Omega)$. Montrer qu'il existe une tribu sur Ω qui contient $\mathcal{G}$ et qui est contenue dans toute tribu sur Ω qui contient $\mathcal{G}$. Cette tribu, appelée *tribu engendrée par* $\mathcal{G}$, est notée $\sigma(\mathcal{G})$.

1.4 Soit Ω un ensemble non vide et soit $\mathcal{G} = \{A_1, \dots, A_n\}$ une partition de Ω (c'est-à-dire $A_i \subset \Omega$, $A_i \neq \emptyset$, $A_i \cap A_j = \emptyset$ si $i \neq j$ et $A_1 \cup \cdots \cup A_n = \Omega$). Déterminer $\sigma(\mathcal{G})$, définie dans l'exercice 1.3.

1.5 Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et soit $G \in \mathcal{F}$. Montrer que si $P(G) = 1$, alors pour tout $F \in \mathcal{F}$, $P(F \cap G) = P(F)$.

1.6 Soit Ω un ensemble non vide et soit $\mathcal{F}$ une tribu sur Ω.

a) Montrer que si $(F_n,\, n \geqslant 1)$ est une suite d'événements dans $\mathcal{F}$, alors $\bigcap_{n=1}^{\infty} F_n \in \mathcal{F}$.
Indication. On rappelle les *lois de De Morgan* :

$$(F \cup G)^c = F^c \cap G^c \qquad \text{et} \qquad (F \cap G)^c = F^c \cup G^c,$$

qui s'étendent aussi aux unions et intersections dénombrables.
b) Soit $(E_n,\, n \geqslant 1) \subset \mathcal{F}$. Trouver une suite $(F_n,\, n \geqslant 1) \subset \mathcal{F}$ d'événements deux à deux disjoints telle que pour tout $m \geqslant 1$,

$$\bigcup_{n=1}^{m} E_n = \bigcup_{n=1}^{m} F_n \qquad \text{et} \qquad \bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} F_n\,.$$

1.7* (Continuité de la probabilité) Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et soit $(E_n, n \geqslant 1)$ une suite d'événements dans $\mathcal{F}$. Etablir les propriétés suivantes.

a) Si $E_n \subset E_{n+1}$, pour tout $n \geqslant 1$, alors $P(\bigcup_{n=1}^{\infty} E_n) = \lim_{n\to\infty} P(E_n)$.
b) Si $E_n \supset E_{n+1}$, pour tout $n \geqslant 1$, alors $P(\bigcap_{n=1}^{\infty} E_n) = \lim_{n\to\infty} P(E_n)$.

Indication. Pour la partie a), on utilisera l'exercice 1.6b) et les axiomes de la définition 1.4. La partie b) sera déduite de a).

1.8* Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soient E, F et G des événements dans $\mathcal{F}$.

a) Montrer que $P(E \cap F^c) = P(E) - P(E \cap F)$.

b) Montrer que

$$P(E \cup F \cup G) = P(E) + P(F) + P(G) \\ - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G)\,.$$

1.9* Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et $(E_n, n \geqslant 1)$ des événements dans $\mathcal{F}$.

a) Etablir l'inégalité de Boole : $P(\bigcup_{n=1}^m E_n) \leqslant \sum_{n=1}^m P(E_n)$.

b) Montrer que $P(\bigcup_{n=1}^\infty E_n) \leqslant \sum_{n=1}^\infty P(E_n)$.

1.10* (Lemme de Borel-Cantelli, première partie) Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Etant donné une suite $(G_n)_{n\in\mathbb{N}}$ d'événements dans $\mathcal{F}$, on considère l'événement « il y a une infinité de valeurs de n pour lesquelles G_n se réalise » et on note cet événement $\limsup_{n\to\infty} G_n$.

a) Montrer que

$$\limsup_{n\to\infty} G_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} G_m\,.$$

b) Montrer que si $\sum_{n=1}^\infty P(G_n) < \infty$, alors

$$P\left(\limsup_{n\to\infty} G_n\right) = 0\,.$$

1.11 Une expérience consiste à lancer une pièce de monnaie deux fois de suite. On peut donc distinguer trois étapes de l'expérience : avant le premier lancer, puis juste après le premier lancer mais avant le deuxième, et enfin, après le deuxième lancer.

a) Quel est l'ensemble fondamental pour cette expérience ?

b) Montrer que l'événement $G = \{(\mathrm{P},\mathrm{P}), (\mathrm{F},\mathrm{P})\}$ n'est *pas* observable après le premier lancer mais avant le deuxième lancer.

c) Montrer que l'événement $H = \{(\mathrm{P},\mathrm{P}), (\mathrm{P},\mathrm{F})\}$ est observable après le premier lancer et avant le deuxième lancer.

1.12 Une boîte contient trois billes, une noire, une blanche et une rouge. On considère l'expérience qui consiste à tirer au hasard une bille de la boîte, à l'y remettre, puis à tirer une deuxième bille.

a) Quel est l'ensemble fondamental pour cette expérience ?

b) Quels sont les événements observables après le premier tirage mais avant le deuxième ?

c) Mêmes questions qu'en a) et b) dans le cas où l'on ne remet pas la première bille tirée avant de tirer la deuxième.

1.13 Une boîte contient deux boules rouges, une boule noire et une boule verte. On considère l'expérience qui consiste à tirer au hasard une boule de la boîte, à l'y remettre, puis à tirer une deuxième boule.

a) Quel est l'ensemble fondamental Ω_1 le plus petit possible qui permet de décrire cette expérience ? Quel est l'espace de probabilité correspondant à l'expérience ?

b) On souhaite décrire cette expérience à l'aide d'un ensemble fondamental où tous les singletons sont équiprobables. Donner un tel ensemble fondamental Ω_2.

c) Quelle tribu sur Ω_2 permet de modéliser le fait qu'il n'est pas possible de distinguer les deux boules rouges ?

1.14 Deux joueurs A et B lancent un dé à tour de rôle (en commençant par A). Le premier qui obtient 6 a gagné et le jeu s'arrête.

a) Donner un ensemble fondamental Ω pour cette expérience.

b) Quels sont les événements qui correspondent respectivement à « A gagne », « B gagne » et « ni A, ni B ne gagne » ?

1.15 On dispose d'un dé équitable.

a) Donner l'espace de probabilité $(\Omega, \mathcal{F}, P)$ si on lance le dé trois fois de suite.

b) Quelle est la probabilité que la somme des résultats vaille 3, vaille 4 et vaille 5 ?

Chapitre 2

Analyse combinatoire

Dans ce chapitre, nous allons étudier diverses notions et méthodes qui permettent de compter les nombres d'éléments d'ensembles finis de grande cardinalité, pour lesquels une simple énumération des éléments n'est pas réalisable. On appelle cet ensemble de techniques les *techniques de dénombrement* ou encore l'*analyse combinatoire*.

2.1 Principe de multiplication

Ce principe est à la base des techniques de dénombrement. Il permet de compter le nombre de résultats d'expériences qui se décomposent en une suite d'expériences partielles, réalisées l'une après l'autre.

Principe. Une expérience consiste en m expériences partielles, réalisées l'une après l'autre dans un ordre bien déterminé. On suppose que la première peut produire n_1 résultats, et pour chaque résultat de la première expérience partielle, il y a n_2 résultats possibles pour la deuxième expérience ; pour chaque résultat des deux premières expériences, il y en a n_3 pour la troisième ; etc. Alors le nombre total de résultats de l'expérience complète est le produit $n_1\, n_2 \cdots n_m$.

Exemple 2.1 Une expérience se déroule en trois étapes. La première comporte trois résultats possibles, la deuxième quatre et la troisième deux résultats possibles. Le principe de multiplication s'applique donc avec $n_1 = 3$, $n_2 = 4$ et $n_3 = 2$. Le nombre total de résultats possibles de cette expérience est égal à $3 \times 4 \times 2 = 24$.

2.2 Permutations

Une *permutation* de n objets distincts $a_1, \ldots, a_n$ est un arrangement ordonné, sans répétition, de ces n objets. On s'intéresse à déterminer le nombre de permutations de ces n objets.

Exemple 2.2 Supposons que $n = 3$ et que les trois objets sont a, b et c. Les permutations possibles sont *abc*, *acb*, *bac*, *bca*, *cab* et *cba*. Le nombre de permutations de ces trois objets est donc 6.

Pour compter les permutations de n objets distincts, on observe que fabriquer une permutation de n objets correspond à une expérience qui se déroule en n étapes. La première étape est de choisir le premier objet. Il y a donc $n_1 = n$ résultats possibles. Une fois le premier objet choisi, il faut choisir le deuxième objet parmi les $n-1$ restants, avec $n_2 = n-1$ résultats possibles. Puis il faut choisir le troisième objet parmi les $n-2$ objets restants, et ainsi de suite, jusqu'à l'étape n, où il ne reste que 1 résultat possible.

D'après le principe de multiplication, le nombre de permutations des n objets distincts $a_1, \ldots, a_n$ est donc $n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$. Ce nombre est aussi noté $n!$, ce qui se lit n *factoriel*. Par convention, on pose $0! = 1$.

Exemple 2.3 Les premières valeurs de la fonction $n \mapsto n!$ sont $0! = 1$, $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$. La valeur de $n!$ augmente très rapidement avec n. Ainsi, $10! = 3\,628\,800$ et $20! \simeq 2{,}43 \cdot 10^{18}$. On a la propriété asymptotique, appelée *formule de Stirling* :

$$n! \sim \sqrt{2\pi}\, n^n \sqrt{n}\, e^{-n},$$

où $a_n \sim b_n$ signifie que $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

Arrangements

Un *arrangement* est une permutation de k objets pris parmi n objets distincts ($k \leqslant n$). Les objets sont donc pris sans répétition et sont ordonnés. Nous désignerons par $A_{n,k}$ le nombre de ces arrangements.

Puisqu'il y a n choix possibles pour le premier objet, puis $n-1$ choix pour le deuxième objet, $n-2$ pour le troisième, etc., et finalement $n-(k-1)$ choix possibles pour l'objet numéro k, on déduit du principe de multiplication que

$$\begin{aligned} A_{n,k} &= n \cdot (n-1) \cdots (n-k+1) \\ &= n \cdot (n-1) \cdots (n-k+1) \cdot \frac{(n-k)(n-k-1)\cdots 2 \cdot 1}{(n-k)(n-k-1)\cdots 2 \cdot 1} \end{aligned}$$

et donc que

$$A_{n,k} = \frac{n!}{(n-k)!}\,.$$

Exemple 2.4 On s'intéresse à déterminer le nombre de mots de trois lettres distinctes qui peuvent être formés avec un alphabet de 26 lettres. Un tel mot est un arrangement de trois lettres choisies parmi les 26 lettres de l'alphabet. Par conséquent, le nombre de ces mots est $A_{26,3} = 26 \times 25 \times 24 = 15\,600$.

2.3 Combinaisons

Une *combinaison de k éléments pris parmi n éléments* (distincts) est un sous-ensemble à k éléments de cet ensembe de n éléments. Le nombre de ces sous-ensembles est noté

$$C_{n,k} \qquad \text{ou} \qquad \binom{n}{k}.$$

Le nombre $C_{n,k}$ est aussi parfois appelé un *coefficient binomial.*

Dans un sous-ensemble, les éléments ne sont pas ordonnés, au contraire d'un arrangement de k éléments pris parmi n. Par conséquent, à chaque sous-ensemble correspond $k!$ arrangements, ce qui implique que $A_{n,k} = C_{n,k} \cdot k!$ et donc

$$C_{n,k} = \frac{A_{n,k}}{k!} = \frac{n!}{k!(n-k)!},$$

ou encore

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Exemple 2.5 Dans un club avec 23 membres, on veut former un comité de 4 personnes. Combien y a-t-il de comités possibles ? Dans cette question, il est sous-entendu que toutes les personnes jouent le même rôle. Puisque chaque comité consiste en un sous-ensemble de 4 personnes, la réponse est

$$\binom{23}{4} = \frac{23!}{4!\,19!} = \frac{23 \times 22 \times 21 \times 20}{4 \times 3 \times 2 \times 1} = 8\,855.$$

Le binôme de Newton

La notion de combinaison permet de retrouver la formule du binôme de Newton, qui donne une expression pour $(c_1+c_2)^n$, où c_1 et c_2 sont des nombres réels et $n \in \mathbb{N}$, $n \geqslant 1$. En effet, par définition,

$$(c_1+c_2)^n = (c_1+c_2)\cdot(c_1+c_2)\cdots(c_1+c_2) = \sum_{i_1,\ldots,i_n \in \{1,2\}} c_{i_1} c_{i_2} \cdots c_{i_n}.$$

Le nombre de termes de la somme est 2^n. En effet, il y a deux possibilités pour chacun des n indices $i_1, \ldots, i_n$, soient $2 \times \cdots \times 2 = 2^n$ termes au total.

Chaque terme de la somme est de la forme $c_1^k\, c_2^{n-k}$. Le nombre de termes de cette forme se calcule en observant qu'un terme de ce type s'obtient en choisissant k indices parmi les n indices $i_1, \ldots, i_n$, et en leur attribuant la valeur 1, tout en attribuant la valeur 2 aux $n-k$ indices restants.

Le nombre de termes de cette forme est donc $C_{n,k}$. Par conséquent,

$$(c_1+c_2)^n = \sum_{k=0}^{n} \binom{n}{k} c_1^k\, c_2^{n-k}.$$

Propriétés des combinaisons

Proposition 2.6

a) $C_{n,k} = C_{n,n-k}$.

b) $C_{n,k} = C_{n-1,k-1} + C_{n-1,k}$.

Démonstration. a) Cette égalité découle simplement du fait que les deux membres sont égaux à

$$\frac{n!}{k!(n-k)!}\,.$$

b) Soit $\Omega = \{\omega_0, \omega_1, \ldots, \omega_{n-1}\}$. Cet ensemble possède n éléments. Le nombre $C_{n,k}$ est le nombre de sous-ensembles de Ω de cardinalité k. Cette famille de sous-ensembles se décompose en deux familles disjointes.

La première est formée des sous-ensembles qui contiennent l'élément ω_0. Ils s'obtiennent en choisissant $k-1$ éléments dans le sous-ensemble $\{\omega_1, \ldots, \omega_{n-1}\}$ et en adjoignant ω_0 au sous-ensemble ainsi obtenu. Il y a donc $C_{n-1,k-1}$ sous-ensembles de ce type.

La deuxième est formée des sous-ensembles de Ω qui sont de cardinalité k et qui ne contiennent pas ω_0. Ils s'obtiennent en choisissant k éléments dans $\{\omega_1, \ldots \omega_{n-1}\}$ et sont donc au nombre de $C_{n-1,k}$. D'où l'égalité

$$\binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k}\,. \qquad \square$$

Le triangle de Pascal

On va remplir une suite de lignes selon les règles suivantes. Sur la première ligne, on écrit le nombre 1, puis sur la deuxième ligne, on écrit ce nombre deux fois, avec un léger décalage comme dans la figure 2.1. Ensuite, on remplit successivement les lignes suivantes en commençant par écrire le nombre 1, puis en écrivant la somme des deux nombres qui se trouvent sur la ligne précédente, juste à gauche et à droite de la position courante.

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
\iddots & & & & & & \vdots & & & & & & \ddots
\end{array}$$

Figure 2.1 Le triangle de Pascal.

Sur la ligne n, on retrouve

$$\binom{n}{0},\ \binom{n}{1},\ \ldots,\ \binom{n}{n},$$

ce qui est une conséquence de la Proposition 2.6b).

2.4 Coefficients multinomiaux

On veut découper un ensemble de n objets distincts en une suite de r sous-ensembles de tailles respectives $n_1, n_2, \dots, n_r$, où $n_1 + n_2 + \cdots + n_r = n$. On s'intéresse à déterminer de combien de manières ce découpage peut se faire.

Il y a C_{n,n_1} choix pour le premier sous-ensemble, C_{n-n_1,n_2} choix pour le deuxième sous-ensemble, $C_{n-n_1-n_2,n_3}$ choix pour le troisième sous-ensemble, ..., $C_{n-n_1-\cdots-n_{r-1},n_r}$ choix pour le $r^{\text{ième}}$ sous-ensemble.

D'après le principe de multiplication, le nombre de manières est donc

$$\binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdots \binom{n-n_1-\cdots-n_{r-1}}{n_r}.$$

En exprimant les coefficients binomiaux et en simplifiant, on constate que ce produit est égal à

$$\frac{n!}{n_1!(n-n_1)!} \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \cdots \frac{(n-(n_1-\cdots-n_{r-1})!}{n_r!(n-n_1-\cdots-n_r)!} = \frac{n!}{n_1!\, n_2! \cdots n_r!}.$$

On note le nombre obtenu

$$\binom{n}{n_1\; n_2\; \cdots\; n_r}$$

et on l'appelle un *coefficient multinomial.* Sa valeur est

$$\binom{n}{n_1\; n_2\; \cdots\; n_r} = \frac{n!}{n_1!\, n_2! \cdots n_r!}.$$

Il représente le nombre de manières de découper un ensemble de n éléments en une suite de r sous-ensembles de tailles respectives $n_1, \dots, n_r$.

2.5 Tirages sans remise

Une urne contient m boules rouges et n boules noires. On tire un paquet de k boules (sans remise). Quelle est la probabilité d'obtenir r boules rouges et $k-r$ boules noires ($r \leqslant k \leqslant m+n$ et $k-r \leqslant n$) ?

L'ensemble Ω des résultats possibles est l'ensemble des sous-ensembles de k éléments pris parmi $m+n$. Soient G l'événement « on obtient r boules rouges et $k-r$ boules noires ». Alors

$$\sharp\,\Omega = \binom{m+n}{k}.$$

Comme l'événement G se réalise si l'on tire r boules rouges parmi les m disponibles et $k-r$ boules noires parmi les n disponibles, on déduit que

$$\sharp\,G = \binom{m}{r}\binom{n}{k-r}$$

et la probabilité cherchée est donc

$$P(G) = \frac{\binom{m}{r}\binom{n}{k-r}}{\binom{m+n}{k}} .$$

2.6 Quelques applications de l'analyse combinatoire

Les applications de l'analyse combinatoire sont très nombreuses. Nous allons donner trois exemples typiques.

Loteries

Quelle est la probabilité de gagner le gros lot dans une loterie ? Evidemment, cela dépend de la manière dont la loterie est conçue (numéros qui peuvent sortir, nombre de numéros à choisir, etc.), mais les mêmes principes s'appliquent à toutes les loteries.

EXEMPLE 2.7 Dans une loterie, on choisit six nombres parmi $1, \ldots, 54$. On gagne le gros lot si on a les six chiffres justes. Le deuxième prix est obtenu si on a cinq chiffres justes. Quelles sont les probabilités d'obtenir le gros lot et le deuxième prix (en achetant un seul billet) ?

Pour répondre à cette question, remarquons que les résultats possibles sont les sous-ensembles à six éléments de l'ensemble $\{1, \ldots, 54\}$, donc

$$\sharp\,\Omega = \binom{54}{6} .$$

Par conséquent,

$$P\{\text{« gros lot »}\} = \frac{1}{\binom{54}{6}} = \frac{1}{25\,827\,165} .$$

Le deuxième prix est obtenu si on choisit 5 nombres parmi les 6 gagnants et 1 parmi les 48 perdants. Donc

$$\sharp\,\text{deuxième prix} = \binom{6}{5}\binom{48}{1}$$

et

$$P\{\text{« deuxième prix »}\} = \frac{\binom{6}{5}\binom{48}{1}}{\binom{54}{6}} \simeq \frac{1}{89\,678} .$$

Mains de poker

L'analyse combinatoire permet de calculer la probabilité de recevoir une main donnée au jeu de poker. On rappelle que le poker se joue avec un jeu de 52 cartes et qu'une main comporte cinq cartes. L'ensemble des cartes est

$$\mathcal{C} = \{\mathrm{P}_1, \mathrm{P}_2, \dots, \mathrm{P}_{13},\ \mathrm{T}_1, \mathrm{T}_2, \dots, \mathrm{T}_{13},\ \mathrm{C}_1, \mathrm{C}_2, \dots, \mathrm{C}_{13},\ \mathrm{Q}_1, \mathrm{Q}_2, \dots, \mathrm{Q}_{13}\},$$

où on a noté P pour la couleur « pique », T pour « trèfle », C pour « cœur » et Q pour « carreau », et on a numéroté les cartes de 1 (pour l'as) à 10, puis 11 pour le valet, 12 pour la dame et 13 pour le roi. Il y a donc 13 cartes de chacune des quatre couleurs P, T, C et Q et 4 cartes de chacune des 13 valeurs 1 à 13.

a) *Probabilité d'une suite royale.* Dans une main de poker, quelle est la probabilité d'avoir une *suite royale* (c-à-d. cinq cartes consécutives, toutes de la même couleur) ?
Le nombre de mains possibles est

$$\sharp\,\Omega = \binom{52}{5} = 2\,598\,960\,.$$

Le nombre de suites royales se calcule comme suit. Il y a 10 choix possibles pour la valeur de la carte la plus élevée (5 à as). Ensuite, il y a 4 choix possibles pour la couleur de cette carte, et les autres cartes sont alors déterminées. Il y a donc $10 \times 4 = 40$ suites royales et donc

$$P\{\text{« suite royale »}\} = \frac{40}{\binom{52}{5}} \simeq 0{,}00001539\,.$$

b) *Probabilité d'obtenir deux paires.* On obtient deux paires lorsqu'on obtient deux couples de cartes de mêmes valeurs et une cinquième carte d'une troisième valeur. Il faut donc choisir deux valeurs distinctes parmi les treize valeurs possibles, puis pour les deux valeurs, deux cartes parmi les quatre de cette valeur, et finalement une carte parmi les $52 - 8 = 44$ cartes de valeurs distinctes de celles des deux paires. Le nombre de mains de deux paires est donc

$$\sharp\,\text{« deux paires »} = \binom{13}{2}\binom{4}{2}\binom{4}{2}\binom{44}{1}$$

et donc

$$P\{\text{« deux paires »}\} = \frac{\binom{13}{2}\binom{4}{2}^2 \cdot 44}{\binom{52}{5}} \simeq 0{,}047539\,.$$

Anniversaires

Il y a n personnes dans une classe. Quelle est la probabilité de l'événement G = « au moins deux personnes ont le même jour d'anniversaire » ?

Pour modéliser ce problème, on exclut le 29 février et on numérote les jours de l'année de 1 à 365. Si on énumère les personnes (par exemple par ordre alphabétique), alors l'ensemble des résultats possibles est $\Omega = \{1, \ldots, 365\}^n$, puisque chaque résultat est une suite de n nombres, chacun étant l'un des entiers entre 1 et 365. Par conséquent, $\sharp\,\Omega = 365^n$. Soit G^c l'événement

$$G^c = \text{« les } n \text{ anniversaires sont distincts »} .$$

Cet événement G^c se réalise quel que soit l'anniversaire de la première personne (365 choix possibles), pour autant que l'anniversaire de la deuxième personne tombe parmi les 364 autres jours de l'année, que l'anniversaire de la troisième personne tombe l'un des 363 autres jours de l'année, etc. Ainsi,

$$\sharp\, G^c = A_{365,n} = 365 \times 364 \times \cdots \times (365 - n + 1)$$

et donc

$$P(G^c) = \frac{A_{365,n}}{365^n} .$$

Dans le cas où $n = 45$, par exemple, on obtient

$$P(G^c) = \frac{A_{365,45}}{365^{45}} = \frac{365 \times 364 \times \cdots \times 321}{365 \times 365 \times \cdots \times 365} = 0{,}059 ,$$

donc $P(G) = 1 - 0{,}059 = 0{,}941$.

Si l'on pose la même question pour 23 personnes, alors on obtient $P(G) = \frac{A_{365,23}}{365^{23}} = 0{,}508$. A partir de 23 personnes, il y a plus d'une chance sur deux d'avoir au moins deux personnes avec le même anniversaire !

2.7 Exemples de modélisations*

Lorsqu'on est confronté à un problème de calcul des probabilités de nature combinatoire, le choix de l'ensemble fondamental Ω est souvent la première question à laquelle il faut répondre. Dans de nombreux cas, il est souhaitable de choisir Ω de sorte que tous les événements élémentaires soient équiprobables. Pour ce faire, il est parfois nécessaire de modifier artificiellement la donnée.

Exemple 2.8 Une expérience consiste à lancer deux pièces de monnaie indistinguables. Quelle est la probabilité d'obtenir deux faces ?

Une première possibilité serait de dire qu'il y a trois résultats possibles et donc de poser $\Omega = \big\{(\mathrm{P},\mathrm{P}), (\mathrm{P},\mathrm{F}), (\mathrm{F},\mathrm{F})\big\}$: cet ensemble contient les trois résultats possibles, vu que les deux pièces sont indistinguables. Cependant, il faut faire attention au fait que, dans ce cas,

$$P\big((\mathrm{P},\mathrm{F})\big) = \frac{1}{2} \neq \frac{1}{4} = P\big((\mathrm{P},\mathrm{P})\big) .$$

On préfère en général faire comme si on ajoutait un élément distinctif aux deux pièces, ou faire comme si on lançait les pièces l'une après l'autre, de sorte qu'il y ait quatre couples de résultats possibles, et on pose $\Omega = \big\{(\mathrm{P},\mathrm{P}),(\mathrm{P},\mathrm{F}),(\mathrm{F},\mathrm{P}),(\mathrm{F},\mathrm{F})\big\}$. Ainsi, tous les événements élémentaires ont la probabilité $\frac{1}{4}$.

On procède de même dans une expérience qui consiste à jeter deux dés indistinguables : on préfère que l'ensemble Ω soit l'ensemble des 36 couples plutôt que l'ensemble des 21 paires de résultats possibles.

Exemple 2.9 On dispose de 15 boules indistinguables, que l'on répartit dans deux urnes. Quelle est la probabilité que l'une des urnes reste vide ?

On peut considérer que les résultats de l'expérience sont des couples (m,n), où m, $n \in \mathbb{N}$ avec $m+n=15$. Mais, dans ce cas, on n'a pas d'emblée la valeur de $P\big\{(n,m)\big\}$.

Il est donc préférable d'admettre que les boules sont numérotées et placées les unes après les autres, et que chaque boule a la probabilité $\frac{1}{2}$ d'aller dans chacune des deux urnes. Ainsi, $\Omega = \{1,2\}^{15}$, et $\omega = (1,\dots,1)$ signifie que toutes les boules sont placées dans l'urne numéro 2. De cette manière, $P\{\omega\} = 2^{-15}$, pour tout $\omega \in \Omega$, et la probabilité que l'une des urnes reste vide est $2 \cdot 2^{-15} = 2^{-14}$.

D'autres exemples intéressants sont proposés notamment dans les exercices 2.21 et 2.22. Il y a bien entendu des exceptions au principe des événements élémentaires équiprobables. La plus simple est la situation du lancer répété d'une pièce de monnaie biaisée, qui sera étudiée au chapitre 3 dans le cadre de la loi binomiale.

2.8 Exercices

2.1 Les plaques minéralogiques aux USA sont formées de trois lettres, suivies de trois chiffres.

a) Quel est le nombre de plaques minéralogiques possibles ?

b) Quel est le nombre de plaques qui commencent par la lettre A ?

2.2 Au jeu Euromillions, le joueur doit choisir 5 nombres sur une première liste des nombres 1 à 50 et 2 nombres sur une deuxième liste des nombres 1 à 9. Il gagne le gros lot si tous ses nombres sont tirés. Quelle est la probabilité de gagner le gros lot à ce jeu ?

2.3 Quatre joueurs A, B, C, D reçoivent chacun une main de treize cartes (d'un jeu de 52 cartes). Combien y a-t-il de répartitions possibles ?

2.4 Dans la loterie de l'exemple 2.7, on obtient le troisième prix si on a quatre chiffres justes. Quelle est la probabilité d'obtenir ce prix ?

2.5 Dans une main de poker, quelle est la probabilité d'avoir une *suite* (cinq cartes consécutives, pas toutes de la même couleur) ?

2.6* On tire une main de 5 cartes d'un jeu de 52 cartes.

a) Quelle est la probabilité d'avoir un carré (quatre cartes de la même valeur et une autre carte) ?
b) Quelle est la probabilité d'avoir un brelan (trois cartes de la même valeur et deux autres cartes de valeurs différentes) ?
c) Quelle est la probabilité d'avoir une main pleine (trois cartes de mêmes valeurs et une paire) ?

2.7 Dans un groupe de 4 femmes et 6 hommes, on doit former un comité de 2 femmes et 2 hommes. Combien de comités sont possibles si :

a) deux des femmes refusent d'être ensemble dans le comité ;
b) il y a un couple marié dans le groupe et les deux époux ne peuvent être tous deux au comité ;
c) il faut désigner l'un des quatre membres comme président du comité ;
d) deux femmes et un homme sont frère et sœurs et au plus un membre de cette famille peut servir au comité ?

2.8 De combien de manières peut-on asseoir 8 personnes (A, B, C, D, E, F, G et H) en rang si :

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

2.9* De combien de manières peut-on placer 3 disques de musique classique, 2 disques de jazz et un de rock (tous d'interprètes différents) sur une étagère si :

a) aucune restriction n'est imposée ;
b) les disques de jazz doivent être rangés ensemble et les disques de classique aussi ;
c) les disques de classique doivent être rangés ensemble, mais les disques de jazz séparés l'un de l'autre.

2.10* Combien de « mots » différents peut-on former en permutant les lettres des mots suivants ?

a) vélos ;

b) papier ;
c) banane ;
d) minimum.

2.11 Démontrer que

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

lorsque $0 \leqslant r \leqslant \min(n,m)$, r, m, $n \in \mathbb{N}$.

2.12* a) Etablir l'égalité

$$\sum_{k=1}^{n} k\binom{n}{k} = n\,2^{n-1}$$

en calculant de deux manières différentes le nombre de manières de former une équipe avec un capitaine à partir d'un groupe de n personnes.

b) En procédant par analogie avec a), montrer que

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

c) Retrouver les résultats de a) et b) en appliquant le binôme de Newton à $(1+x)^n$ et en dérivant par rapport à x.

2.13* a) Soit $1 \leqslant i \leqslant n$ fixé. Etablir l'égalité

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

en calculant de deux manières différentes le nombres de comités avec i postes à responsabilités que l'on peut former à partir d'un groupe de n personnes.

b) Retrouver le résultat de a) en appliquant le binôme de Newton à $(1+x)^n$ et en dérivant par rapport à x.

2.14 (Formule d'inclusion-exclusion) Soient Ω un ensemble fini et P une probabilité sur $(\Omega, \mathcal{P}(\Omega))$. Soit $F_i \in \mathcal{P}(\Omega)$, $i = 1, \ldots, n$. Montrer que

$$\begin{aligned} P(F_1 \cup \cdots \cup F_n) = &\sum_{i=1}^{n} P(F_i) - \sum_{i_1<i_2} P(F_{i_1} \cap F_{i_2}) + \cdots \\ &+ (-1)^{k+1} \sum_{i_1<i_2<\ldots<i_k} P(F_{i_1} \cap \cdots \cap F_{i_k}) + \cdots \\ &+ (-1)^{n+1} P(F_1 \cap \cdots \cap F_n)\,. \end{aligned}$$

2.15 On tire 13 cartes d'un jeu de 52 cartes. Quelle est la probabilité qu'au moins une couleur manque ?

2.16 Démontrer que

$$\binom{n+\ell}{n} = \sum_{i=0}^{n} \binom{i+\ell-1}{i}.$$

2.17 Soit Ω un ensemble fini. Montrer que si $|\Omega| = n$, alors $\left|\mathcal{P}(\Omega)\right| = 2^n$.

2.18 a) Combien de manières y a-t-il de répartir n boules distinguables dans k urnes distinguables ?

b) Combien de manières y a-t-il de répartir n boules indistinguables dans k urnes distinguables, de sorte que chaque urne contienne au moins une boule. On admettra que $k \leqslant n$.

c) Même question qu'en b), mais sans l'exigence que chaque urne contienne au moins une boule.

2.19 a) Combien y a-t-il de solutions entières positives de l'équation $x_1 + \cdots + x_k = n$ $(k,\, n \in \mathbb{N}^*)$?

b) Combien y a-t-il de solutions entières non négatives de l'équation $x_1 + \cdots + x_k = n$ $(k,\, n \in \mathbb{N})$?

c) Combien y a-t-il de solutions entières non négatives de l'équation $x_1 + \cdots + x_k = n$ pour laquelle exactement r des inconnues sont nulles $(k,\, n,\, r \in \mathbb{N},\, r \leqslant k)$?

2.20* Un voyageur se déplace dans le plan Oxy en respectant les règles suivantes :

— il reste dans le quadrant non négatif ;

— il se déplace parallèlement soit à l'axe Ox, soit à l'axe Oy ;

— il s'éloigne toujours de l'origine ;

— il ne peut changer de direction que lorsque ses deux coordonnées sont entières.

a) Combien y a-t-il de chemins possibles d'extrémités $(0,0)$ et (m,n), où $(m,n) \in \mathbb{N}^2$?

b) Si $m \leqslant n$, combien y a-t-il de chemins possibles d'extrémités $(0,0)$ et (n,n) qui passent par (m,m) ?

2.21* Le métro M1 de Lausanne-Flon à Renens-CFF a 16 arrêts. A 7 de ces arrêts, dont Renens-CFF et Lausanne-Flon, il y a des correspondances avec des bus, un autre métro ou des trains CFF. A Lausanne-Flon, il y a 31 personnes plus le chauffeur dans le métro. Au cours du trajet, personne ne monte. A Renens-CFF, le chauffeur voit descendre 11 personnes.

a) De combien de manières les autres personnes ont-elles pu se répartir parmi les arrêts du métro ?

b) Qu'en est-il si le chauffeur a vu que sur les 31 personnes, il y a 16 hommes et 15 femmes et qu'à Renens-CFF, 5 hommes et 6 femmes sont descendus ?

c) Qu'en est-il si le chauffeur se souvient que 13 personnes sorties avant Renens-CFF lui ont demandé des correspondances ?

2.22* On tire sans remise des boules d'une urne qui contient r boules rouges, j boules jaunes, v boules vertes, b boules bleues et a boules blanches. Le jeu s'arrête dès qu'une couleur a été tirée deux fois. On s'intéresse à la probabilité que le jeu s'arrête après k tirages et qu'on a tiré deux boules rouges.

a) Définir un espace de probabilité qui permet de décrire cette expérience et qui est tel que tous les événements élémentaires sont équiprobables.

b) Pour $k = 2, 3, 4$, calculer cette probabilité.

2.23 On s'intéresse aux nombres formés de 4 chiffres.

a) Combien d'entre eux ont 4 chiffres identiques ?

b) Combien d'entre eux sont formés de 2 paires de deux chiffres ?

c) Combien d'entre eux ont 4 chiffres distincts ?

d) Combien d'entre eux ont 4 chiffres distincts, rangés dans l'ordre croissant ?

e) Que deviennent les réponses aux questions a), c) et d) si l'on remplace 4 par n ?

2.24 (Formule de Leibniz) Soit f et g deux fonctions de classe C^∞. Pour une fonction f, on note $f^{(n)}$ la n-ème dérivée de f. Montrer que pour tout $n \geqslant 1$,

$$(f \cdot g)^{(n)}(x) = \sum_{k=0}^{n} \binom{n}{k} f^{(n-k)}(x)\, g^{(k)}(x).$$

Chapitre 3

Probabilité conditionnelle et indépendance

Soient G et H deux événements. La probabilité que H se réalise est $P(H)$, mais si on sait que G s'est réalisé, alors cela va modifier notre opinion concernant cette probabilité. Le but de ce chapitre est d'examiner comment cette probabilité est modifiée et quelles conséquences il est possible d'en tirer.

3.1 Définition et premières propriétés

Si on se place dans l'interprétation fréquentiste décrite dans le chapitre 1, imaginons qu'on répète n fois une expérience et notons $n(G)$ le nombre de fois que l'événement G est réalisé. Parmi les $n(G)$ expériences où G est réalisé, l'événement H est aussi réalisé $n(H \cap G)$ fois. Par conséquent, la probabilité de H sachant que G est réalisé devrait être voisin, pour des grandes valeurs de n, de

$$\frac{n(H \cap G)}{n(G)} = \frac{\frac{n(H \cap G)}{n}}{\frac{n(G)}{n}} \simeq \frac{P(H \cap G)}{P(G)} \,.$$

Ces considérations motivent la définition suivante.

Définition 3.1

Soient G et H deux événements. Si $P(G) > 0$, on pose

$$P(H \mid G) = \frac{P(H \cap G)}{P(G)} \,.$$

L'expression $P(H \mid G)$ se lit « probabilité conditionnelle de H sachant que G est réalisé ».

Exemple 3.2 Une urne contient 6 boules rouges et 5 boules noires. On tire deux boules sans remise. Quelle est la probablité conditionnelle que la deuxième boule soit noire sachant que la première est rouge ?

Soient H l'événement « la deuxième boule est noire » et G l'événement « la première boule est rouge ». La réponse cherchée est

$$P(H \mid G) = \frac{P(H \cap G)}{P(G)} = \frac{\dfrac{6 \times 5}{11 \times 10}}{\dfrac{6 \times 10}{11 \times 10}} = \frac{5}{10} = \frac{1}{2}\,.$$

Il est aussi possible de raisonner comme suit. Si la première boule est rouge, il reste 10 boules dans l'urne, dont 5 noires, d'où $P(H \mid G) = \frac{5}{10} = \frac{1}{2}$.

Proposition 3.3

a) Soient G et H deux événements tels que $P(G) > 0$. Alors

$$P(H \cap G) = P(G) \cdot P(H \mid G)\,.$$

b) Règle de multiplication. *Soient $G_1, \ldots, G_k$ des événements tels que $P(G_1 \cap \cdots \cap G_k) > 0$. Alors*

$$\begin{aligned} &P(G_1 \cap G_2 \cap \cdots \cap G_k) \\ &\quad = P(G_1) \cdot P(G_2 \mid G_1) \cdot P(G_3 \mid G_1 \cap G_2) \cdots P(G_k \mid G_1 \cap \cdots \cap G_{k-1})\,. \end{aligned}$$

Démonstration. a) Cette propriété découle directement de la définition de la probabilité conditionnelle.

b) Le membre de droite vaut

$$\begin{aligned} &P(G_1) \cdot \frac{P(G_2 \cap G_1)}{P(G_1)} \cdot \frac{P(G_1 \cap G_1 \cap G_3)}{P(G_1 \cap G_2)} \cdots \frac{P(G_1 \cap \cdots \cap G_k)}{P(G_1 \cap \cdots \cap G_{k-1})} \\ &\quad = P(G_1 \cap \cdots \cap G_k)\,. \end{aligned}$$

□

3.2 Formule des probabilités totales

Proposition 3.4 *Soit $\{G_1, \ldots, G_n\}$ une partition de Ω (c'est-à-dire $G_i \neq \emptyset$, $G_i \cap G_j = \emptyset$ si $i \neq j$ et $G_1 \cup \cdots \cup G_n = \Omega$) en événements tels que $P(G_i) > 0$, $i = 1, \ldots, n$. Alors pour tout événement G,*

$$P(G) = \sum_{i=1}^{n} P(G \mid G_i) P(G_i)\,.$$

DÉMONSTRATION. On observe que

$$\begin{aligned} P(G) &= P(G \cap \Omega) = P\big(G \cap (G_1 \cup \cdots \cup G_n)\big) \\ &= P\big((G \cap G_1) \cup \cdots \cup (G \cap G_n)\big) . \end{aligned}$$

Puisque les événements $G \cap G_i$ sont disjoints, cette probabilité est égale à

$$P(G \cap G_1) + \cdots + P(G \cap G_n) = P(G \mid G_1)P(G_1) + \cdots + P(G \mid G_n)P(G_n) . \quad \square$$

EXEMPLE 3.5 Une urne contient n_1 boules rouges, n_2 boules noires et n_3 boules bleues. Soit $n = n_1 + n_2 + n_3$. On effectue deux tirages sans remise. Quelle est la probabilité de l'événement « la deuxième boule tirée est rouge » ?

Soit G cet événement et soit G_1 l'événement « la première boule tirée est rouge », G_2 l'événement « la première boule tirée est noire », G_3 l'événement « la première boule tirée est bleue ». Alors $\{G_1, G_2, G_3\}$ est une partition de Ω. Par conséquent,

$$\begin{aligned} P(G) &= P(G \mid G_1)P(G_1) + P(G \mid G_2)P(G_2) + P(G \mid G_3)P(G_3) \\ &= \frac{n_1 - 1}{n - 1} \cdot \frac{n_1}{n} + \frac{n_1}{n-1} \cdot \frac{n_2}{n} + \frac{n_1}{n-1} \cdot \frac{n_3}{n} = \frac{n_1(n_1 - 1 + n_2 + n_3)}{n(n-1)} = \frac{n_1}{n} . \end{aligned}$$

Il est intéressant d'observer que $\frac{n_1}{n}$ est aussi la probabilité que la première boule tirée soit rouge ! En fait, que les tirages soient effectués avec remise ou sans remise ne change pas la probabilité de l'événement « la deuxième boule tirée est rouge ». La même propriété est valable si on tire n boules et qu'on s'intéresse à la probabilité de l'événement « la $i^{\text{ième}}$ boule tirée est rouge » ($i \leqslant n$).

3.3 Formule de Bayes

Proposition 3.6 *Soient G et H deux événements. Alors*

$$P(G \mid H) = \frac{P(H \mid G)P(G)}{P(H \mid G)P(G) + P(H \mid G^c)P(G^c)} .$$

DÉMONSTRATION. On utilise d'abord la définition de la probabilité conditionnelle, puis on applique la formule des probabilités totales à la partition $\{G, G^c\}$, pour obtenir

$$P(G \mid H) = \frac{P(H \cap G)}{P(H)} = \frac{P(H \mid G)P(G)}{P(H \mid G)P(G) + P(H \mid G^c)P(G^c)} . \qquad \square$$

Exemple 3.7 *Test pour détecter une maladie.* Considérons un test pour détecter une certaine maladie. Selon les indications du producteur, si le test est administré à une personne qui est effectivement atteinte par cette maladie, alors le test indiquera « positif » dans 95 % des cas. En revanche, si le test est administré à une personne qui n'est pas atteinte par cette maladie, alors le test indiquera « positif » dans seulement 1 % des cas : les « faux positifs » sont donc relativement rares. Heureusement, seulement 0,1 % de la population est porteuse de cette maladie.

Si on vous administre ce test et que le résultat est positif, quelle est la probabilité que vous soyez effectivement malade ?

Pour répondre à cette question, appelons G l'événement « être malade » et H l'événement « le résultat du test est positif ». L'énoncé du problème exprime que

$$P(H \mid G) = 0{,}95\,, \qquad P(H \mid G^c) = 0{,}01 \quad \text{et} \quad P(G) = 0{,}001\,.$$

D'après la formule de Bayes,

$$\begin{aligned} P(G \mid H) &= \frac{P(H \mid G)P(G)}{P(H \mid G)P(G) + P(H \mid G^c)P(G^c)} \\ &= \frac{0{,}95 \times 0{,}001}{0{,}95 \times 0{,}001 + 0{,}01 \times 0{,}999} \\ &= 0{,}0868 \simeq 9\,\% \ ! \end{aligned}$$

Ainsi, heureusement pour le patient, la probabilité d'être réellement malade n'est que de 9 %, malgré le résultat positif du test.

Formule de Bayes généralisée

Proposition 3.8 *Soit $\{G_1, \ldots, G_n\}$ une partition de Ω et H un événement. On suppose que $P(H) > 0$ et $P(G_i) > 0$, pour $i = 1, \ldots, n$. Alors*

$$P(G_i \mid H) = \frac{P(H \mid G_i)P(G_i)}{\displaystyle\sum_{j=1}^{n} P(H \mid G_j)P(G_j)}\,.$$

Démonstration. En appliquant la définition de la probabilité conditionnelle, puis la formule des probabilités totales, on obtient

$$P(G_i \mid H) = \frac{P(G_i \cap H)}{P(H)} = \frac{P(H \mid G_i)P(G_i)}{\displaystyle\sum_{j=1}^{n} P(H \mid G_j)P(G_j)}\,. \qquad \square$$

Proposition 3.9 *Soit H un événement tel que $P(H) > 0$. Alors $G \mapsto P(G \mid H)$ est une probabilité sur $(\Omega, \mathcal{F})$.*

Démonstration. Il s'agit de vérifier les axiomes 1 et 2 de la définition 1.4a). Clairement, l'axiome 1 est vérifié car

$$P(\Omega \mid H) = \frac{P(\Omega \cap H)}{P(H)} = \frac{P(H)}{P(H)} = 1\,.$$

Quant à l'axiome 2, soit $(G_n,\, n \in \mathbb{N})$ une suite d'événements dans $\mathcal{F}$ qui sont deux à deux disjoints. Alors

$$P\left(\bigcup_{n=1}^{\infty} G_n \,\middle|\, H\right) = \frac{P\left(\left(\bigcup_{n=1}^{\infty} G_n\right) \cap H\right)}{P(H)}\,.$$

Or le numérateur est égal à

$$P\left(\bigcup_{n=1}^{\infty} (G_n \cap H)\right) = \sum_{n=1}^{\infty} P(G_n \cap H)\,,$$

puisque les événements $(G_n \cap H,\, n \in \mathbb{N})$ sont deux à deux disjoints. Ainsi,

$$P\left(\bigcup_{n=1}^{\infty} G_n \,\middle|\, H\right) = \frac{1}{P(H)} \sum_{n=1}^{\infty} P(G_n \cap H) = \sum_{n=1}^{\infty} P(G_n \mid H)$$

et l'axiome 2 est vérifié. □

3.4 Indépendance

La notion d'indépendance d'événements est une des notions centrales de la théorie des probabilités. Elle décrit la situation où le fait de savoir qu'un événement G est réalisé ne change pas notre opinion de la probabilité qu'un autre événement H se réalise. Cette notion intuitive d'indépendance sera formalisée et ses conséquences seront étudiées dans cette section.

Définition et premières propriétés

Définition 3.10

Deux événements G et H sont indépendants *si $P(G \cap H) = P(G)P(H)$.*

Remarque 3.11 Si $P(G) > 0$, la propriété d'indépendance est équivalente à $P(H \mid G) = P(H)$, ce qui correspond à la notion intuitive mentionnée ci-dessus.

Exemple 3.12 On lance une pièce équitable deux fois de suite. L'espace de probabilité est donc

$$\Omega = \big\{(\mathrm{P},\mathrm{P}),\ (\mathrm{P},\mathrm{F}),\ (\mathrm{F},\mathrm{P}),\ (\mathrm{F},\mathrm{F})\big\}\,,$$

et pour $\omega_i \in \Omega$, $P\{\omega_i\} = \frac{1}{4}$, $i = 1, \dots, 4$.

Soit $G_i =$ l'événement « la pièce i tombe sur pile », $i = 1, 2$. On va vérifier que G_1 et G_2 sont indépendants. En effet, $\sharp\, G_i = 2$, $i = 1, 2$, donc $P(G_1) = \frac{2}{4} = P(G_2)$. D'autre part, $P(G_1 \cap G_2) = P\big\{(\mathrm{P},\mathrm{P})\big\} = \frac{1}{4}$ et on a donc $P(G_1 \cap G_2) = P(G_1)P(G_2)$. Les événements G_1 et G_2 sont donc bien indépendants.

Proposition 3.13 *Si deux événements G et H sont indépendants, alors G et H^c le sont aussi.*

Démonstration. Observons que

$$\begin{aligned} P(G \cap H^c) &= P(G) - P(G \cap H) = P(G) - P(G)P(H) \\ &= P(G)\big(1 - P(H)\big) = P(G)P(H^c)\,. \end{aligned}$$

D'où l'indépendance de G et H^c. □

Indépendance de plusieurs événements

Définition 3.14

Des événements $G_1, \dots, G_n$ ($n \geqslant 2$) sont mutuellement indépendants *si pour tout $1 < k \leqslant n$ et pour toute suite $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n$,*

$$P(G_{i_1} \cap G_{i_2} \cap \cdots \cap G_{i_k}) = P(G_{i_1})P(G_{i_2}) \cdots P(G_{i_k})\,.$$

Remarque 3.15 Lorsqu'on cherche à vérifier l'indépendance mutuelle de n événements, il *ne suffit pas* de vérifier que pour $i \neq j$,

$$P(G_i \cap G_j) = P(G_i)P(G_j)\,.$$

Cette dernière propriété signifie que les événements G_i sont *deux à deux indépendants*, ce qui est une propriété *plus faible* que l'indépendance mutuelle (exercice 3.16).

Lorsqu'on dit simplement que $G_1, \dots, G_n$ sont *indépendants*, on sous-entend que ces événements sont mutuellement indépendants.

Les événements d'une suite $(G_n)_{n \in \mathbb{N}}$ sont dits *indépendants* si pour tout $n \in \mathbb{N}$, $G_1, \dots, G_n$ sont indépendants.

Voici une propriété qui est équivalente à l'indépendance mutuelle.

Proposition 3.16 *Les événements $G_1, \dots, G_n$ sont mutuellement indépendants si et seulement si pour tous $a_1, \dots, a_n \in \{0, 1\}$,*

$$P(G_1^{a_1} \cap \cdots \cap G_n^{a_n}) = P(G_1^{a_1}) \cdots P(G_n^{a_n})\,,$$

où $G_i^1 = G_i$ et $G_i^0 = G_i^c$, $i = 1, \dots, n$.

La démonstration de cette proposition est laissée en exercice (exercice 3.22).

Epreuves indépendantes

Définition 3.17

On appelle épreuves indépendantes *des expériences partielles identiques mutuellement indépendantes.*

Deux exemples d'épreuves indépendantes sont les résultats de n lancers équitables d'une pièce de monnaie, et les résultats de n jets équitables d'un dé.

Loi binomiale

On exécute n épreuves indépendantes, où chaque épreuve est un « succès » avec probabilité p et un « échec » avec probabilité $1-p$. Quelle est la probabilité d'avoir k succès, $k \in \{0, \ldots, n\}$?

Un résultat possible est d'obtenir d'abord les k succès et ensuite les $n-k$ échecs :

$$\underbrace{S \ldots S}_{k \text{ fois}} \quad \underbrace{E \ldots E}_{(n-k) \text{ fois}} .$$

La probabilité d'obtenir les résultats dans cet ordre est $p^k(1-p)^{n-k}$. Cependant, les k succès peuvent apparaître dans chacune des $C_{n,k}$ positions possibles. Pour chaque ensemble de k positions, la probabilité est la même que ci-dessus, à savoir $p^k(1-p)^{n-k}$. Par conséquent,

$$P\{\text{« } k \text{ succès »}\} = \binom{n}{k} p^k (1-p)^{n-k} .$$

Remarque 3.18 On peut observer que d'après le binôme de Newton,

$$\sum_{k=0}^{n} P\{\text{« } k \text{ succès »}\} = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+1-p)^n = 1 .$$

Loi multinomiale

On effectue n épreuves indépendantes : chaque épreuve a m résultats possibles, de probabilités respectives $p_1, \ldots, p_m$. Quelle est la probabilité d'obtenir n_1 résultats de type 1, n_2 résultats de type $2, \ldots, n_m$ résultats de type m, où $n_1 + \cdots + n_m = n$?

La probabilité que les n résultats soient (dans l'ordre)

$$\underbrace{R_1, \ldots, R_1}_{n_1 \text{ fois}}, \underbrace{R_2, \ldots, R_2}_{n_2 \text{ fois}}, \ \ldots \ , \underbrace{R_m, \ldots, R_m}_{n_m \text{ fois}}$$

est $p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$. Le nombre de permutations possibles de ces n résultats est

$$\frac{n!}{n_1!\, n_2! \ldots n_m!} = \binom{n}{n_1\, n_2 \, \cdots \, n_m} .$$

On en déduit que cette probabilité est

$$\binom{n}{n_1 \; n_2 \; \cdots \; n_m} p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m} .$$

Cette formule porte le nom de *loi multinomiale* de paramètres n, $p_1, \ldots, p_n$.

3.5 Exercices

3.1 Une compagnie d'assurances estime que les individus 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 en l'espace d'un an ; cette probabilité tombe à 0,2 pour les personnes à risque modéré. On suppose que 30 % de la population appartient à la classe à haut risque.

a) Quelle est la probabilité qu'un nouvel assuré soit victime d'un accident durant l'année qui suit la signature de son contrat ?

b) 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 ?

3.2 Un système émetteur/récepteur transmet des 0 et des 1. La probabilité de transmission correcte d'un 0 est 0,8, celle pour un 1 est 0,9. On sait que 45 % des symboles transmis sont des 0.

a) Quelle est la probabilité que le récepteur reçoive le symbole 0 ?

b) Si le récepteur reçoit le symbole 0, quelle est la probabilité que l'émetteur ait effectivement envoyé un 0 ?

3.3 Une urne contient 6 boules rouges et 5 boules noires. On tire trois boules sans remise. Trouver la probabilité que les trois boules tirées soient rouges.

3.4 On divise un jeu de 52 cartes en 4 mains de 13 cartes chacune. Quelle est la probabilité que chaque main contienne un as ?

3.5 On choisit deux cartes au hasard (sans remise) d'un jeu de 52 cartes. Sachant que la deuxième carte est un pique, quelle est la probabilité que la première carte soit un pique ?

3.6 On tire successivement sans remise deux boules d'une urne qui contient 8 boules rouges et 4 boules noires. Sachant que la deuxième boule est rouge, quelle est la probabilité que la première soit noire ?

3.7 Une boîte contient 3 boules rouges et 12 boules noires. On verse 5 de ces boules dans une première urne, 5 dans une deuxième urne et les 5 dernières dans une troisième urne. Quelle est la probabilité que les trois urnes contiennent chacune une boule rouge ?

3.8 Soient F et G deux événements tels que $P(F) \in]0,1[$ et $P(G) \in]0,1[$. Montrer que $P(F \mid G) > P(F)$ si et seulement si $P(G \mid F^c) < P(G \mid F)$ (autrement dit, la réalisation de G accroît la probabilité que F se réalise si et seulement si G est plus probable lorsque F se réalise que lorsque F ne se réalise pas).

3.9 Dans le jeu télévisé « A prendre ou à laisser » (simplifié), 21 boîtes numérotées et fermées contiennent divers prix, dont un seul gros lot. Au début du jeu, le joueur choisit l'une de ces boîtes. Ensuite, il désigne une des boîtes restantes, qui est ouverte. Ceci est répété jusqu'à ce qu'il reste une seule des 20 boîtes. Il se trouve qu'aucune des 19 boîtes ouvertes ne contenait le gros lot. A ce moment-là, on offre au joueur la possibilité d'échanger la boîte initialement choisie contre la boîte restante. A-t-il intérêt de le faire ?

3.10 (Problème de Monty-Hall) Dans un jeu télévisé, le joueur peut choisir l'une des trois portes 1, 2 et 3. Derrière l'une de ces portes, il y a une voiture, derrière les deux autres des chèvres. Une fois que le joueur a désigné une porte, disons la porte 1, le présentateur ouvre l'une des deux autres portes, disons la 2 ; derrière celle-ci se trouve une chèvre. Le joueur a maintenant la possibilité de modifier son choix (et prendre la porte 3) ou de conserver son choix (garder la porte 1). Que devrait-il faire ?

3.11 Une boîte contient trois cartes. La première a deux faces noires, la deuxième a deux faces rouges et la dernière a une face rouge et une face noire. On tire une carte au hasard et on observe que sa face supérieure est rouge. Quelle est la probabilité que l'autre face soit aussi rouge ?

3.12* 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.

a) Quelle fraction d'électeurs a participé à l'élection locale ?
b) 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 ?

3.13* Pour se rendre à son bureau, un employé prend tantôt sa voiture – et il arrive alors une fois sur deux en retard – tantôt le métro – et il arrive alors en retard une fois sur quatre seulement. Quand il arrive à l'heure, il décide toujours de conserver le même moyen de transport le jour suivant, tandis que chaque fois qu'il est en retard, il décide d'en changer. Soit p la probabilité que l'employé prenne sa voiture le premier jour.

a) Quelle est la probabilité que l'employé prenne sa voiture le $n^{\text{ième}}$ jour ?
b) Quelle est la probabilité qu'il arrive en retard la $n^{\text{ième}}$ fois qu'il se rend au bureau ?
c) Trouver la limite lorsque $n \to \infty$ des deux résultats en a) et b).

3.14* Un dé peint en rouge est découpé en 27 sous-dés identiques. On choisit un sous-dé au hasard et on le jette une fois.

a) Quelle est la probabilité d'obtenir rouge sur la face supérieure du dé jeté ?
b) Sachant que l'on a obtenu rouge, quelle est la probabilité que le sous-dé provienne d'un des coins du dé original ?

3.15* (Urne de Polya) Une urne contient r boules rouges et b boules blanches. Une boule est tirée et sa couleur est notée. On remet alors la boule tirée dans l'urne et on ajoute a boules de la même couleur. La même procédure est répétée $n-1$ fois et l'on fait ainsi n tirages. Soient R_j l'événement « la $j^{\text{ième}}$ boule tirée est rouge » et B_j l'événement « la $j^{\text{ième}}$ boule tirée est blanche. »

a) Calculer $P(R_1)$, $P(B_1)$, $P(R_2)$, $P(B_2)$.

b) En procédant par récurrence, calculer $P(R_j)$ et $P(B_j)$ pour $2 \leqslant j \leqslant n$.

c) Interpréter le cas $a = -1$.

3.16 On lance une pièce de monnaie équitable trois fois de suite. Soit G_1 l'événement « les lancers 2 et 3 donnent le même résultat », G_2 l'événement « les lancers 1 et 3 donnent le même résultat » et G_3 l'événement « les lancers 1 et 2 donnent le même résultat ». Montrer que ces événements sont deux à deux indépendants mais ne sont pas mutuellement indépendants.

3.17 On jette un dé huit fois de suite. Quelle est la probabilité d'obtenir 1 trois fois exactement ?

3.18 On lance une pièce de monnaie n fois de suite. Quelle est la probabilité d'obtenir au moins une fois pile ?

3.19 On lance un dé jusqu'à obtenir un 6. Quelle est la probabilité que l'expérience ne s'arrête jamais ?

3.20 Soit $N \in \mathbb{N}$. On dispose d'une urne avec N boules numérotées de 1 à N. Le jeu consiste à tirer une boule. On a un succès si on tire la boule portant le numéro 1. On joue de manière répétée à ce jeu. Les jeux sont indépendants. Quel est le nombre k de jeux nécessaires pour que la probabilité d'avoir au moins un succès soit supérieure ou égale à $\frac{1}{2}$?

3.21 Une urne contient 4 boules rouges, 6 boules vertes et 8 boules bleues. On effectue 9 tirages avec remise. Quelle est la probabilité d'obtenir trois fois une boule rouge, deux fois une boule verte et quatre fois une boule bleue ?

3.22 a) Démontrer la condition nécessaire de la proposition 3.16.
Indication. Montrer le résultat par une double récurrence sur n et sur $k := \big|\{i : a_i = 0\}\big|$.

b) Démontrer la condition suffisante de la proposition 3.16.
Indication. Montrer le résultat par récurrence descendante.

3.23 Soient $E_1, \ldots, E_n$ des événements indépendants ($n \geqslant 2$). Montrer que

$$P\left(\bigcup_{i=1}^{n} E_i\right) = 1 - \prod_{i=1}^{n}\big(1 - P(E_i)\big)\,.$$

3.24* On admet que le sexe de chaque enfant d'un couple est indépendant de celui des autres enfants de la famille et que la probabilité qu'un enfant est un garçon est 0,52. Calculer, pour un couple ayant 4 enfants, les probabilités des événements suivants :

a) tous les enfants sont du même sexe ;

b) les trois aînés sont des garçons, la dernière est une fille ;

c) il y a exactement 3 garçons ;

d) les deux aînés sont des garçons ;

e) il y a au moins une fille.

3.25 Soit $E = \{1, \ldots, n\}$ et supposons que A_1 et A_2 sont choisis indépendemment et uniformément dans $\mathcal{P}(E)$, l'ensemble des parties de E. Montrer que

$$P\{A_1 \subset A_2\} = \left(\frac{3}{4}\right)^n .$$

3.26* (Lemme de Borel-Cantelli, deuxième partie) On rappelle que la première partie du lemme de Borel-Cantelli est présentée dans l'exercice 1.10.

a) Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et soit $(G_n)_{n\in\mathbb{N}}$ une suite d'événements indépendants. On suppose que $\sum_{n=1}^{\infty} P(G_n) = +\infty$. Montrer que $P\left(\limsup_{n\to\infty} G_n\right) = 1$.

b) Une expérience consiste à lancer de manière répétée un dé équilibré. Montrer que si on prolonge indéfiniment l'expérience, alors il est certain que la valeur 6 sortira une infinité de fois.

c) On dispose d'une suite infinie d'urnes. L'urne numéro n contient n boules numérotées de 1 à n. Une seule boule est tirée de chacune des urnes (les tirages sont indépendants). Quelle est la probabilité que la suite des numéros tirés comporte une infinité de 1 ?

3.27 (Le « problème des points » du chevalier de Méré) Deux équipes se rencontrent dans un tournoi qui se joue en n jeux gagnants. On admet que chaque équipe gagne un jeu avec la probabilité $\frac{1}{2}$. Après que l'équipe A a gagné i jeux et l'équipe B j jeux ($i < n$ et $j < n$), le tournoi est définitivement interrompu. Quelle est la manière équitable de répartir le prix de x francs entre les deux équipes ?

Chapitre 4

Variables aléatoires

Dans de nombreux exemples, on associe à chaque résultat possible d'une expérience une grandeur numérique. La formalisation mathématique de cette situation mène à la notion de variable aléatoire, que nous allons maintenant introduire.

4.1 Définition et exemples

Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité.

Définition 4.1

Une variable aléatoire *(en abrégé* v.a.*) est une fonction $X : \Omega \to \mathbb{R}$ telle que pour tout $x \in \mathbb{R}$,*

$$\big\{\omega \in \Omega : X(\omega) \leqslant x\big\} \in \mathcal{F}\,. \tag{4.1}$$

Commentaires

Une autre notation pour $\big\{\omega \in \Omega : X(\omega) \leqslant x\big\}$ est $X^{-1}\big(]-\infty, x]\big)$, où $X^{-1}(I)$ est la notation usuelle pour l'ensemble des antécédents du sous-ensemble I de $\mathbb{R}$. La condition (4.1) n'est pas très différente de la condition qui assure la continuité d'une fonction (l'image réciproque d'un ouvert est un ouvert).

Pour simplifier l'écriture, on écrit souvent $\{X \leqslant x\}$ au lieu de $\{\omega \in \Omega : X(\omega) \leqslant x\}$ ou $X^{-1}\big(]-\infty, x]\big)$.

Puisqu'on va s'intéresser à la probabilité de $\{X \leqslant x\}$, il est naturel d'exiger que cet ensemble appartienne à $\mathcal{F}$, le domaine de définition de la mesure de probabilité P.

Dans le cas particulier où $\mathcal{F} = \mathcal{P}(\Omega)$, toute fonction $X : \Omega \to \mathbb{R}$ est une variable aléatoire.

EXEMPLE 4.2 *(Variable aléatoire indicatrice)* Soit $G \in \mathcal{F}$. On définit $1_G : \Omega \to \mathbb{R}$ par

$$1_G(\omega) = \begin{cases} 1 & \text{si } \omega \in G\,, \\ 0 & \text{sinon}\,. \end{cases}$$

La fonction 1_G « indique » par sa valeur si l'événement G se réalise ou pas. C'est bien une v.a., car

$$\{1_G \leqslant x\} = \begin{cases} \emptyset & \text{si } x < 0\,, \\ G^c & \text{si } 0 \leqslant x < 1\,, \\ \Omega & \text{si } x \geqslant 1\,. \end{cases}$$

EXEMPLE 4.3 On jette un dé n fois. Soit X le nombre de fois que le résultat est 6. Pour décrire cette situation, on pose $\Omega = \{1, \ldots, 6\}^n$, $\mathcal{F} = \mathcal{P}(\Omega)$, et on définit $P : \mathcal{F} \to [0,1]$ de telle sorte que tous les événements élémentaires sont équiprobables. Les éléments de Ω sont notés $\omega = (\omega_1, \ldots, \omega_n)$, où ω_i est le résultat du jet numéro i. La variable aléatoire X est la fonction définie sur Ω et à valeurs dans $\mathbb{R}$ définie par

$$X(\omega) = \sum_{i=1}^{n} 1_{\{6\}}(\omega_i)\,,$$

où $1_{\{6\}}$ est la fonction de $\{1, \ldots, 6\}$ dans $\mathbb{R}$ définie par

$$1_{\{6\}}(\omega_i) = \begin{cases} 1 & \text{si } \omega_i = 6\,, \\ 0 & \text{sinon}\,. \end{cases}$$

On peut s'intéresser à des probabilités comme

$$P\{X = k\} = P\big\{\omega \in \Omega : X(\omega) = k\big\}\,.$$

Or l'événement $\big\{\omega \in \Omega : X(\omega) = k\big\}$ n'est rien d'autre que l'événement « obtenir 6 exactement k fois » parmi n jets, et donc

$$P\{X = k\} = P\{\text{« obtenir 6 exactement } k \text{ fois »}\} = \binom{n}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}\,.$$

Définition 4.4

Une v.a. X qui ne prend qu'une seule valeur $x_0 \in \mathbb{R}$, c'est-à-dire telle que $P\{X = x_0\} = 1$, est appelée une v.a. déterministe.

4.2 Fonction de répartition d'une variable aléatoire

Dans de nombreuses situations, la probabilité que la variable aléatoire prenne ses valeurs dans un intervalle est une grandeur qui joue un rôle plus important que les propriétés ensemblistes de la fonction X. Les probabilités sont déterminées par une fonction qui joue un rôle important et qui est définie comme suit.

Définition 4.5

La fonction de répartition *d'une v.a. X est la fonction $F : \mathbb{R} \to \mathbb{R}$ définie par*
$$F(x) = P\{X \leqslant x\}\,.$$

Proposition 4.6 *Soit F la fonction de répartition d'une v.a. X. Alors :*

a) $x \leqslant y \Rightarrow F(x) \leqslant F(y)$.

b) $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$.

c) $\lim_{y\downarrow x,\, y\neq x} F(y) = F(x)$. En particulier, F est continue à droite.

Démonstration. a) Si $x \leqslant y$, alors $\{X \leqslant x\} \subset \{X \leqslant y\}$. En utilisant la proposition 1.5d), on constate que $F(x) = P\{X \leqslant x\} \leqslant P\{X \leqslant y\} = F(y)$.

b) Soient $(x_n,\, n \in \mathbb{N})$ tel que $x_n \geqslant x_{n+1}$ et $\lim_{n\to\infty} x_n = -\infty$. Soit $G_n = \{X \leqslant x_n\}$. Alors $G_n \supset G_{n+1}$ et $\bigcap_{n=1}^{\infty} G_n = \emptyset$. En effet, soit $\omega \in \Omega$ et soit $m \in \mathbb{N}$ tel que $x_m < X(\omega)$. Alors $\omega \notin G_m$, donc $\omega \notin \bigcap_{n=1}^{\infty} G_n$ puisque $\bigcap_{n=1}^{\infty} G_n \subset G_m$.

Par conséquent, en utilisant la continuité de la probabilité (théorème 1.6b), on obtient
$$0 = P(\emptyset) = P\left(\bigcap_{n=1}^{\infty} G_n\right) = \lim_{n\to\infty} P(G_n) = \lim_{n\to\infty} P\{X \leqslant x_n\} = \lim_{n\to\infty} F(x_n)\,.$$

Puisque ceci vaut pour toute suite (x_n) qui décroît vers $-\infty$, on en déduit que $\lim_{x\to-\infty} F(x) = 0$.

De la même manière, soient $(y_n,\ n \in \mathbb{N})$ tel que $y_n \leqslant y_{n+1}$ et $\lim_{n\to\infty} y_n = +\infty$. Soit $H_n = \{X \leqslant y_n\}$. Alors $H_n \subset H_{n+1}$ et $\bigcup_{n=1}^{\infty} H_n = \Omega$. En effet, si $\omega \in \Omega$, soit m tel que $y_m \geqslant X(\omega)$. Alors $\omega \in H_m \subset \bigcup_{n=1}^{\infty} H_n$. Par conséquent, en utilisant à nouveau la continuité de la probabilité (théorème 1.6a), nous obtenons
$$1 = P(\Omega) = P\left(\bigcup_{n=1}^{\infty} H_n\right) = \lim_{n\to\infty} P(H_n) = \lim_{n\to\infty} P\{X \leqslant y_n\} = \lim_{n\to\infty} F(y_n)\,.$$

Puisque ceci vaut pour toute suite (y_n) qui croît vers $+\infty$, on en déduit que $\lim_{y\to\infty} F(y) = 1$.

c) Soient $(y_n,\ n \in \mathbb{N})$ telle que $y_n \geqslant y_{n+1} \geqslant x$ et $\lim_{n\to\infty} y_n = x$. Soit $G_n = \{X \leqslant y_n\}$, $H = \{X \leqslant x\}$. Alors $G_n \supset G_{n+1} \supset H$ et $\bigcap_{n\in\mathbb{N}} G_n = H$. En effet, l'inclusion $\bigcap_{n\in\mathbb{N}} G_n \supset H$ découle du fait que $G_n \supset H$ pour tout n. Pour l'inclusion inverse, soit $\omega \in \bigcap_{n\in\mathbb{N}} G_n$. Alors $X(\omega) \leqslant y_n$, pour tout $n \in \mathbb{N}$. Par conséquent, $X(\omega) \leqslant \lim_{n\to\infty} y_n = x$, donc $\omega \in H$.

Ainsi, en utilisant une troisième fois la continuité de la probabilité, nous obtenons
$$\begin{aligned} F(x) = P\{X \leqslant x\} = P(H) &= P\left(\bigcap_{n=1}^{\infty} G_n\right) \\ &= \lim_{n\to\infty} P(G_n) = \lim_{n\to\infty} P\{X \leqslant y_n\} = \lim_{n\to\infty} F(y_n)\,. \end{aligned}$$
□

Remarque 4.7 Toute fonction $F : \mathbb{R} \to \mathbb{R}$ qui possède les trois propriétés de la proposition 4.6 est appelée une *fonction de répartition*.

Notation. On note

$$F(a-) = \lim_{y \uparrow a} F(y)\,.$$

Observons que si F est continue en a, alors $F(a) = F(a-)$. Sinon, $F(a-) < F(a)$.

Proposition 4.8 *Soient a et b deux nombres tels que $a \leqslant b$. Alors :*

a) $P\{X < a\} = F(a-)$.
b) $P\{a < X \leqslant b\} = F(b) - F(a)$.
c) $P\{a \leqslant X \leqslant b\} = F(b) - F(a-)$.
d) Si $a < b$, alors $P\{a < X < b\} = F(b-) - F(a)$.
e) $P\{X = a\} = F(a) - F(a-)$.

Démonstration. a) Soit $(a_n,\, n \in \mathbb{N})$ une suite de nombres tels que $a_n \leqslant a_{n+1} < a$ et $\lim_{n\to\infty} a_n = a$. Puisque $\{X \leqslant a_n\} \subset \{X \leqslant a_{n+1}\}$, on déduit de la continuité de la probabilité que

$$\begin{aligned} P\{X < a\} &= P\left(\bigcup_{n=1}^{\infty} \{X \leqslant a_n\}\right) = \lim_{n\to\infty} P\{X \leqslant a_n\} \\ &= \lim_{n\to\infty} F(a_n) = F(a-)\,. \end{aligned}$$

b) D'après la propriété d'additivité,

$$P\{X \leqslant b\} = P\big(\{X \leqslant a\} \cup \{a < X \leqslant b\}\big) = P\{X \leqslant a\} + P\{a < X \leqslant b\}\,.$$

Par conséquent, $P\{a < X \leqslant b\} = P\{X \leqslant b\} - P\{X \leqslant a\} = F(b) - F(a)$.

c) et d) Voir l'exercice 4.6.

e) Observons que

$$P\{X = a\} = P\{a \leqslant X \leqslant a\} = F(a) - F(a-)\,,$$

d'après c). □

Proposition 4.9 *Soit F la fonction de répartition d'une v.a. X. Alors l'ensemble des points de discontinuité de F est dénombrable.*

Démonstration. Voir l'exercice 4.7.

Corollaire 4.10 *Soit X une v.a. Alors $\big\{x \in \mathbb{R} : P\{X = x\} > 0\big\}$ est un ensemble dénombrable.*

Démonstration. L'ensemble dont il est question est égal à $\{x \in \mathbb{R} : F(x) > F(x-)\}$, où F est la fonction de répartion de X. C'est donc l'ensemble des discontinuités de F. L'énoncé est donc une conséquence de la proposition 4.9. $\square$

Variables aléatoires discrètes et continues

Soit F la fonction de répartition d'une v.a. X. Considérons l'ensemble

$$D_F = \{x \in \mathbb{R} : F(x) - F(x-) > 0\} .$$

Si $P\{X \in D_F\} = 1$, alors on dit que X est une *v.a. discrète* et que F est une fonction de répartition *discrète*.

Si $D_F = \emptyset$, on dit que X est une *v.a. continue* ; dans ce cas, la fonction de répartition F est une fonction continue. Ces deux types de v.a. et de fonctions de répartition jouent un rôle très important, comme le montre le théorème suivant.

Théorème 4.11

Toute fonction de répartition est une combinaison convexe d'une fonction de répartition discrète et d'une fonction de répartition continue.

Démonstration. Voir l'exercice 4.9. $\square$

Remarque 4.12 Pour vérifier qu'une v.a. X est discrète, il suffit de montrer qu'il existe un sous-ensemble dénombrable D de $\mathbb{R}$ tel que $P\{X \in D\} = 1$. Dans ce cas, on aura nécessairement $D_F \subset D$.

Vu le résultat du théorème 4.11, il est naturel d'étudier séparément les v.a. discrètes et les v.a. continues, ce que nous allons faire dans les sections suivantes.

4.3 Variables aléatoires discrètes

Dans cette section, nous allons étudier les principales propriétés des v.a. discrètes et donner une liste des v.a. discrètes les plus importantes.

Fonction de densité d'une variable aléatoire discrète

Si $D = \{x \in \mathbb{R} : F(x) - F(x-) > 0\}$ est un ensemble fini $D = \{a_1, \ldots, a_n\}$ avec $a_1 < \cdots < a_n$, alors

$$1 = P\{X \in D\} = \sum_{i=1}^{n} \big(F(a_i) - F(a_i-)\big) .$$

Dans ce cas, F est une fonction en escalier, constante sur chacun des intervalles $[a_i, a_{i+1}[$, avec un saut en chaque a_i. Il devient préférable de s'interesser aux positions des sauts et à leurs amplitudes, comme suit.

Définition 4.13

On appelle fonction de densité discrète *d'une v.a. discrète X la fonction $f : \mathbb{R} \to \mathbb{R}$ définie par $f(x) = P\{X = x\}$. On appelle aussi cette fonction la* fonction de fréquence *de la v.a. X.*

Dans le cas où $D = \{a_1, \ldots, a_n\}$ comme ci-dessus, la fonction de densité discrète de X est nulle sauf sur D. Pour $a_i \in D$, $f(a_i) = F(a_i) - F(a_i-)$, alors que pour $x \notin D$, $f(x) = F(x) - F(x-) = 0$. D'autre part,

$$\sum_{x \in D} f(x) = 1\,.$$

En effet,

$$\sum_{x \in D} f(x) = \sum_{x \in D} P\{X = x\} = P\{X \in D\} = 1\,,$$

puisque X est discrète.

Variables aléatoires discrètes : exemples de base

Nous présentons ci-dessous les types de v.a. que l'on rencontre le plus fréquemment.

Exemple 4.14

a) *Variable aléatoire de Bernoulli.* On appelle ainsi toute v.a. X telle que $P\big\{X \in \{0,1\}\big\} = 1$. Si $P\{X = 1\} = p$, on dit que X est une *v.a. de Bernoulli de paramètre* p, et on note $X : B(1,p)$.
Cas particulier. Soit $A \in \mathcal{F}$ telle que $P(A) = p$. Posons $X = 1_A$, où $1_A : \Omega \to \mathbb{R}$ est définie par

$$1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A\,, \\ 0 & \text{si } \omega \in A^c\,. \end{cases}$$

Alors X est $B(1,p)$.

b) *Variables aléatoires simples.* Soient $a_1, \ldots, a_n \in \mathbb{R}$, $A_1, \ldots, A_n \in \mathcal{F}$ tels que $(A_1, \ldots, A_n)$ est une partition de Ω. Alors

$$X = \sum_{i=1}^{n} a_i 1_{A_i}$$

est une v.a. discrète. En effet, $P\big\{X \in \{a_1, \ldots, a_n\}\big\} = 1$.

c) *Variable aléatoire géométrique.* On dit que X est une *v.a. géométrique* de paramètre p $(0 < p < 1)$ si pour tout $n = 1, 2, \ldots$,

$$P\{X = n\} = (1-p)^{n-1} p\,.$$

On observera que le membre de droite est la probabilité, dans une suite d'épreuves indépendantes ayant chacune p pour probabilité de succès, que le premier succès ait lieu à la $n^{\text{ième}}$ épreuve. Ainsi, si E_i est l'événement « avoir un échec lors de l'épreuve i » et S_i est l'événement « avoir un succès lors de l'épreuve i », alors $P\{X = n\} = P(E_1 \cap \cdots \cap E_{n-1} \cap S_n)$.

d) *Variable aléatoire binomiale.* On dit que X est une *v.a. binomiale* de paramètres n et p, et on note $X : B(n, p)$ (où $n \in \mathbb{N}^*$ et $0 \leqslant p \leqslant 1$), si, pour $k = 0, \dots, n$,

$$P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k} .$$

On observera que le membre de droite est la probabilité d'obtenir k succès parmi n épreuves indépendantes, toutes ayant p pour probabilité de succès.

e) *Variable aléatoire hypergéométrique.* On appelle ainsi toute v.a. X telle que pour $r = 0, \dots, \min(m, k)$ et $k - r \leqslant n$ (k, m, $n \in \mathbb{N}^*$),

$$P\{X = r\} = \frac{\binom{m}{r}\binom{n}{k-r}}{\binom{m+n}{k}} .$$

On observe que le membre de droite est la probabilité d'obtenir r boules rouges et $k-r$ boules noires lors de tirages sans remise de k boules d'une urne qui contient m boules rouges et n boules noires.

f) *Variable aléatoire de Poisson.* On dit que X est une v.a. de Poisson de paramètre λ ($\lambda \in \,]0, \infty[$) si pour $k \in \mathbb{N}$,

$$P\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!} .$$

Cette loi de probabilité apparaît typiquement lorsqu'on s'intéresse au nombre de succès dans une suite d'épreuves indépendantes où la probabilité de succès est très petite. Nous en verrons la raison dans l'exercice 4.15.

g) *Variable aléatoire binomiale négative.* Dans une suite d'épreuves indépendantes, chacune ayant p pour probabilité de succès, soit X le numéro de l'épreuve où a lieu le $r^{\text{ième}}$ succès. Une telle v.a. X est dite *binomiale négative* de paramètres (r, p) ($0 < p < 1$, $r \in \{1, 2, \dots\}$).
Pour $n = r, r+1, \dots$,

$$\begin{aligned} P\{X = n\} &= P\{\text{« } r-1 \text{ succès parmi les épreuves } 1, \dots, n-1 \\ &\qquad\qquad \text{et succès à l'épreuve } n \text{ »}\} \\ &= P\{\text{« } r-1 \text{ succès parmi les épreuves } 1, \dots, n-1\} \\ &\qquad\qquad \times P\{\text{succès à l'épreuve } n \text{ »}\} \\ &= \binom{n-1}{r-1} p^{r-1} (1-p)^{(n-1)-(r-1)} p = \binom{n-1}{r-1} p^r (1-p)^{n-r} . \end{aligned}$$

Remarque 4.15 Pour $\ell \in \mathbb{R}$ et $k \in \mathbb{N}$, on pose

$$\binom{\ell}{k} = \frac{\ell(\ell-1)\cdots(\ell-k+1)}{k!} .$$

Alors, pour n, r et X comme au point g) de l'exemple 4.14,

$$\begin{aligned}\binom{n-1}{r-1} = \binom{n-1}{n-r} &= \frac{(n-1)(n-2)\cdots(n-1-(n-r)+1)}{(n-r)!} \\ &= \frac{(n-1)(n-2)\cdots(r+1)r}{(n-r)!} \\ &= \frac{-r(-r-1)\cdots(-n+2)(-n+1)}{(n-r)!}(-1)^{n-r} \\ &= \binom{-r}{n-r}(-1)^{n-r}\,.\end{aligned}$$

Par conséquent,

$$P\{X=n\} = \binom{-r}{n-r}p^r(p-1)^{n-r}\,.$$

Avec ces notations, la fonction de densité discrète de la loi binomiale négative ressemble beaucoup à celle de la loi binomiale.

On posera aussi par convention que

$$\binom{m}{k} = 0\,, \qquad \text{si } k < 0\,.$$

Remarque 4.16 On dit parfois qu'une v.a. X « *suit une loi* » binomiale ou hypergéométrique, ou que « *la distribution de probabilité de* X » est binomiale ou hypergéométrique.

4.4 Variables aléatoires continues

La notion de v.a. continue a été introduite à la fin de la section 4.2. Nous en rappelons ici la définition.

Définition 4.17

Soient X une v.a. et F sa fonction de répartition. Si F est une fonction continue de $\mathbb{R}$ dans $\mathbb{R}$, alors on dit que X est une v.a. continue. *En particulier, X est une v.a. continue si $P\{X=x\} = F(x) - F(x-) = 0$, pour tout $x \in \mathbb{R}$.*

Commentaire

Il est important d'observer que la notion de « continuité d'une v.a. » n'a rien à voir avec la notion de continuité de la fonction $X : \Omega \to \mathbb{R}$. Cette dernière notion n'est d'ailleurs pas définie en général, car Ω n'est pas forcément un espace topologique. En revanche, la notion de continuité de la fonction de répartition F est la notion usuelle de continuité d'une fonction de $\mathbb{R}$ dans $\mathbb{R}$.

Fonction de densité

Définition 4.18

Soit F la fonction de répartion d'une v.a. continue. On dit que F (ou X) admet pour fonction de densité *la fonction $f : \mathbb{R} \to \mathbb{R}_+$ si, pour tous $a, b \in \mathbb{R}$,*

$$F(b) - F(a) = \int_a^b f(x)\, dx\,.$$

On remarquera que lorsque $a \leqslant b$, le membre de gauche n'est rien d'autre que $P\{a \leqslant \ X \leqslant b\}$. Par ailleurs, il est possible de remplacer dans cette probabilité les inégalités strictes par des inégalités larges, puisque F est continue.

Observations

Soit F une fonction de répartition qui admet f pour fonction de densité.

a) $\int_{-\infty}^{+\infty} f(x)\, dx = 1$ (car l'intégrale est égale à $F(+\infty) - F(-\infty) = 1 - 0$).

b) Les intégrales sont interprétées au sens de Riemann. Généralement, les fonctions de densité f que nous rencontrerons dans cet ouvrage seront continues ou tout au moins continues par intervalles.

c) Toute fonction $f : \mathbb{R} \to \mathbb{R}_+$ telle que $\int_{-\infty}^{+\infty} f(x)\, dx = 1$ est appelée une *fonction de densité.*

d) Si F est la fonction de répartition d'une v.a. X, alors

$$F(x) = P\{X \leqslant x\} = \int_{-\infty}^{x} f(u)\, du\,,$$

donc $f(x) = F'(x)$. Ainsi, la fonction de densité f de X est la dérivée de la fonction de répartition F de X.

e) Dans cet ouvrage, nous n'étudierons pas les fonctions de répartition continues qui n'admettent pas de fonction de densité. Lorsque nous parlerons d'une v.a. continue, il sera sous-entendu que sa fonction de répartition admet une fonction de densité.

Variables aléatoires continues : exemples de base

Nous présentons ici les principaux exemples de v.a. continues.

EXEMPLE 4.19

a) *Variable aléatoire uniforme sur* $[a, b]$ $(a < b)$. On appelle ainsi toute v.a. dont la fonction de densité est

$$f(x) = \frac{1}{b-a}\, 1_{[a,b]}(x)\,,$$

où

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

La fonction de répartition associée à cette fonction de densité est

$$F(x) = \begin{cases} 0 & \text{si } x \leqslant a\,, \\ \dfrac{x-a}{b-a} & \text{si } a \leqslant x \leqslant b\,, \\ 1 & \text{si } b \leqslant x\,. \end{cases}$$

b) *Variable aléatoire exponentielle de paramètre* λ ($\lambda > 0$). On appelle ainsi toute v.a. dont la fonction de densité est

$$f(x) = \begin{cases} \lambda\, e^{-\lambda x} & \text{si } x \geqslant 0\,, \\ 0 & \text{si } x < 0\,. \end{cases}$$

La fonction de répartition associée à cette fonction de densité est

$$F(x) = \begin{cases} 0 & \text{si } x \leqslant 0\,, \\ 1 - e^{-\lambda x} & \text{si } x \geqslant 0\,, \end{cases}$$

puisque

$$\int_0^x \lambda\, e^{-\lambda y}\, dy = -e^{-\lambda y}\Big|_0^x = 1 - e^{-\lambda x}\,.$$

Remarque 4.20 Observons que pour cette fonction de densité,

$$\int_{-\infty}^{+\infty} f(x)\, dx = \int_0^{+\infty} \lambda\, e^{-\lambda x}\, dx = -e^{-\lambda x}\Big|_0^{+\infty} = 1\,.$$

Une propriété importante des v.a. exponentielles est la propriété d'*absence de mémoire* (exercice 4.23).

Exemple 4.21 *Variable aléatoire normale standard.* On appelle ainsi toute v.a. dont la fonction de densité est

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\,, \qquad x \in \mathbb{R}\,.$$

Remarque 4.22 On peut vérifier que

$$\int_{-\infty}^{+\infty} f(x)\, dx = 1$$

(cf. exercice 4.24).

Exemple 4.23 *Variable aléatoire de loi gamma.* Soient $p > 0$ et $\lambda > 0$. On appelle v.a. de loi gamma toute v.a. dont la fonction de densité est

$$f(x) = \begin{cases} \dfrac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} & \text{si } x > 0\,, \\ 0 & \text{si } x \leqslant 0\,. \end{cases}$$

Rappelons que la fonction $\Gamma : \mathbb{R}_+ \to \mathbb{R}$ (dite *fonction gamma d'Euler*) est définie par

$$\Gamma(p) = \int_0^{+\infty} x^{p-1}\, e^{-x}\, dx\,.$$

Fonction de répartition d'une variable aléatoire normale standard

Par définition, pour $-\infty < x < +\infty$, $F(x) = \Phi(x)$, où

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy\,.$$

Il n'y a pas de formule élémentaire pour $\Phi(x)$. Les valeurs de $\Phi(x)$ sont indiquées dans la tabelle A.2 de l'appendice.

On lit par exemple dans la tabelle que

$$\Phi(0{,}43) = 0{,}6664\,, \qquad \Phi(1{,}65) = 0{,}9505\,.$$

Si l'on cherche la valeur de $\Phi(-1{,}23)$, elle n'est pas indiquée dans la tabelle mais on peut utiliser la tabelle et les propriétés de symétrie de la fonction de densité normale standard pour l'obtenir. En effet, puisque la fonction f est paire ($f(x) = f(-x)$, pour tout $x \in \mathbb{R}$),

$$\Phi(-x) = \int_{-\infty}^{-x} f(u)\, du = \int_{x}^{+\infty} f(u)\, du = 1 - \Phi(x)\,.$$

Par conséquent, $\Phi(-1{,}23) = 1 - \Phi(1{,}23)$. La valeur de $\Phi(1{,}23) = 0{,}8907$ se lit dans la tabelle et on obtient donc que $\Phi(-1{,}23) = 1 - 0{,}8907 = 0{,}1093$.

4.5 Fonctions d'une variable aléatoire continue

Etant donné une v.a. $X : \Omega \to \mathbb{R}$ et une fonction $g : \mathbb{R} \to \mathbb{R}$, il est naturel de former une nouvelle v.a. $Y = g(X) : \Omega \to \mathbb{R}$. Formellement, Y est la composée $Y = g \circ X$ de X avec g, et donc pour tout $\omega \in \Omega$, $Y(\omega) = g(X(\omega))$. Le plus souvent, la fonction g sera relativement régulière : continue à droite, continue, différentiable, etc. Dans ce chapitre, nous allons examiner plus particulièrement la situation où X est une v.a. continue et g est une fonction différentiable et allons faire le lien entre la densité de la v.a. X et celle de $Y = g(X)$. Le résultat principal est présenté dans la proposition suivante.

Proposition 4.24 *Soit X une v.a. continue avec fonction de densité $f_X : \mathbb{R} \to \mathbb{R}_+$. Soient $a, b, \alpha, \beta \in \mathbb{R} \cup \{-\infty, +\infty\}$ tels que $a < b$, $\alpha < \beta$ et $P\{a < X < b\} = 1$. Soient $g :]a, b[\to]\alpha, \beta[$ bijective telle que g et g^{-1} sont de classe C^1. Posons $Y = g(X)$. Alors Y est une v.a. continue dont la fonction de densité est*

$$f_Y(y) = \begin{cases} \dfrac{f_X\big(g^{-1}(y)\big)}{\big|g'\big(g^{-1}(y)\big)\big|} & \text{si } y \in]\alpha, \beta[\,, \\ 0 & \text{sinon}\,. \end{cases}$$

Démonstration. Soit F_Y la fonction de répartition de Y :

$$F_Y(y) = P\{Y \leqslant y\} = P\big\{g(X) \leqslant y\big\}.$$

Puisque g est bijective, elle est strictement monotone. Nous allons donc distinguer les cas où elle est strictement croissante ou strictement décroissante.

1^{er} *cas* : g est strictement croissante. Alors pour $y \in \,]\alpha, \beta[$,

$$F_Y(y) = P\big\{g(X) \leqslant y\big\} = P\big\{X \leqslant g^{-1}(y)\big\} = F_X\big(g^{-1}(y)\big).$$

Puisque $y \mapsto F_X\big(g^{-1}(y)\big)$ est la composée de deux fonctions continues, c'est une fonction continue et donc Y est une v.a. continue. De plus,

$$F_Y'(y) = F_X'\big(g^{-1}(y)\big) \cdot \big(g^{-1}\big)'(y) = f_X\big(g^{-1}(y)\big) \cdot \big(g^{-1}\big)'(y) = \frac{f_X\big(g^{-1}(y)\big)}{g'\big(g^{-1}(y)\big)}.$$

2^e *cas* : g est strictement décroissante. Dans ce cas,

$$F_Y(y) = P\big\{g(X) \leqslant y\big\} = P\big\{X \geqslant g^{-1}(y)\big\} = \int_{g^{-1}(y)}^{\infty} f_X(u)\,du$$

et donc

$$F_Y'(y) = -f_X\big(g^{-1}(y)\big) \cdot \big(g^{-1}\big)'(y) = -\frac{f_X\big(g^{-1}(y)\big)}{g'\big(g^{-1}(y)\big)} = \frac{f_X\big(g^{-1}(y)\big)}{\big|g'\big(g^{-1}(y)\big)\big|}$$

puisque g' est négative. □

Remarque 4.25 Il est parfois plus simple d'énoncer la proposition précédente en termes de la fonction $h(y) = g^{-1}(y)$. Alors $h(Y) = X$ et le lien entre les fonctions de densité est

$$f_Y(y) = f_X\big(h(y)\big)\big|h'(y)\big|.$$

Exemple 4.26 *Variable aléatoire normale (ou gaussienne)* $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$. Soit X une v.a. normale standard. Posons

$$Y = \mu + \sigma X.$$

On dit alors que Y est une v.a. normale (ou gaussienne) de paramètres μ et σ^2, ce que l'on note $N(\mu, \sigma^2)$ (on verra l'interprétation de ces deux paramètres dans l'exercice 6.5). La fonction de densité de Y est

$$f_Y(y) = \frac{f_X\big(g^{-1}(y)\big)}{g'\big(g^{-1}(y)\big)},$$

où

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \quad g(x) = \mu + \sigma x, \quad g'(x) = \sigma, \quad g^{-1}(y) = \frac{y-\mu}{\sigma}.$$

Par conséquent,

$$f_Y(y) = \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right)}{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right).$$

Remarque 4.27 Si Y est $N(\mu, \sigma^2)$, alors $\frac{Y-\mu}{\sigma}$ est normale standard, c'est-à-dire $N(0,1)$. Pour calculer $P\{Y \leqslant y\}$, on écrit

$$P\{Y \leqslant y\} = P\left\{\frac{Y-\mu}{\sigma} \leqslant \frac{y-\mu}{\sigma}\right\} = P\left\{X \leqslant \frac{y-\mu}{\sigma}\right\}.$$

Le calcul de cette probabilité se réduit donc au calcul d'une probabilité pour la v.a. normale standard X. Pour celle-ci, on utilise la tabelle A.2 de l'appendice.

4.6 Construction de variables aléatoires continues*

Le lecteur aura sans doute remarqué que dans les sections 4.4 et 4.5, nous n'avons plus beaucoup parlé de l'espace de probabilité $(\Omega, \mathcal{F}, P)$. En particulier, nous n'avons jamais précisé l'ensemble Ω sur lequel est défini une v.a. continue X, même dans le cas très simple d'une v.a. uniforme sur $[0,1]$.

D'une part, cette précision n'est pas nécessaire dans la suite de cet ouvrage, en particulier pour de nombreux calculs concernant des v.a. continues, mais d'autre part, cette omission cache une difficulté importante. En effet, la donnée d'une v.a. exige de préciser l'espace de probabilité $(\Omega, \mathcal{F}, P)$ et la fonction $X : \Omega \to \mathbb{R}$.

Dans le cas d'une v.a. uniforme sur $[0,1]$, un choix naturel pour l'ensemble fondamental est $\Omega = [0,1]$ et $X(\omega) = \omega$. Il reste à définir $\mathcal{F}$ et P. Or, pour $x \in [0,1]$, il est nécessaire que $P\big\{\omega \in \Omega : X(\omega) \leqslant x\big\} = x$, donc que

$$[0,x] \in \mathcal{F} \quad \text{et} \quad P\big([0,x]\big) = x\,, \qquad \text{pour tout } x \in [0,1]\,. \tag{4.2}$$

On pourrait imaginer prendre $\mathcal{F} = \mathcal{P}\big([0,1]\big)$. Cependant, on peut montrer qu'il n'y a aucune mesure de probabilité P sur $\mathcal{P}\big([0,1]\big)$ qui vérifie $P\big([0,x]\big) = x$ pour tout $x \in [0,1]$. Il est donc nécessaire de considérer une tribu $\mathcal{F}$ plus petite que $\mathcal{P}\big([0,1]\big)$.

La tribu la plus petite possible serait $\mathcal{F} = \sigma\big([0,x] : x \in [0,1]\big)$ (cf. l'exercice 1.3). Cette tribu est appelée la *tribu de Borel* sur $[0,1]$. Il s'agit alors de montrer qu'il existe une (et une seule) mesure de probabilité P sur la tribu de Borel telle que $P\big([0,x]\big) = x$, pour tout $x \in [0,1]$. Cette mesure est appelée la *mesure de Lebesgue*.

Ces considérations sont au départ de la *théorie de la mesure* et les démonstrations de ces résultats dépassent le cadre du présent ouvrage.

4.7 Exercices

4.1 On se place dans le cadre de l'exemple 4.3 avec $n = 2$ (le dé est jeté deux fois). Soit X le nombre de fois que l'on obtient 6 lors de deux lancers du dé. Quelle est la fonction de répartition de X ?

4.2 On jette deux dés équilibrés. Soit X le produit des deux résultats. Quelle est la fonction de densité de X ?

4.3 Soit $\Omega = \{1, \ldots, 6\}^n$, $\mathcal{F} = \mathcal{P}(\Omega)$ et P la probabilité définie par $P\{\omega\} = 6^{-n}$, pour tout $\omega \in \Omega$. Définissons $X : \Omega \to \mathbb{R}$ par

$$X(\omega) = \sum_{i=1}^{n} 1_{\{2,4,6\}}(\omega_i)\,, \qquad \text{où } \omega = (\omega_1, \ldots, \omega_n)\,.$$

Décrire la variable aléatoire X par une phrase et donner sa fonction de densité.

4.4 On jette deux dés équilibrés. Soit X le maximum des deux résultats. Quelle est la fonction de densité de X ?

4.5 Soient $\Omega = \{0, 1\}^n$, $\mathcal{F} = \mathcal{P}(\Omega)$ et P la probabilité définie par $P\{\omega\} = 2^{-n}$, pour tout $\omega \in \Omega$. Définissons $X : \Omega \to \mathbb{R}$ par

$$X(\omega) = \sum_{i=1}^{n} \omega_i\,, \qquad \text{où } \omega = (\omega_1, \ldots, \omega_n)\,.$$

Décrire la variable aléatoire X par une phrase et donner sa fonction de densité.

4.6 Démontrer les points c) et d) de la proposition 4.8.

4.7 Démontrer la proposition 4.9.

4.8 Soient F la fonction de répartition d'une variable aléatoire X et D l'ensemble des discontinuités de F. Supposons que $P\{X \in D\} > 0$ et posons $F_d(x) = P\{X \leqslant x \mid X \in D\}$, pour tout $x \in \mathbb{R}$. Montrer que F_d est une fonction de répartition.

4.9* Nous reprenons les notations de l'exercice 4.8.

a) Supposons que $0 < P\{X \in D\} < 1$. Montrer qu'il existe une fonction de répartition $F_c : \mathbb{R} \to \mathbb{R}$ continue et $\alpha \in \,]0, 1[$ tels que

$$F = \alpha F_d + (1 - \alpha) F_c\,.$$

b) Que peut-on dire dans le cas où $P\{X \in D\} \in \{0, 1\}$?

4.10 a) Soit X une v.a. géométrique de paramètre p. Montrer que

$$\sum_{n=1}^{\infty} P\{X = n\} = 1\,.$$

b) Soit Y une v.a. de Poisson de paramètre λ. Montrer que

$$\sum_{n=0}^{\infty} P\{Y = n\} = 1\,.$$

c) Soit Z une v.a. binomiale négative de paramètres (r, p). Montrer que

$$\sum_{n=r}^{\infty} P\{Z = n\} = 1\,.$$

Indication. $(1+x)^{-r} = \sum_{n=0}^{\infty} \binom{-r}{n} x^n$, pour $r \in \mathbb{R}$ et $|x| < 1$.

d) Soient U une variable aléatoire binomiale de paramètres n et p, et Z la variable aléatoire définie au point c). Expliquer pourquoi $P\{U < r\} = P\{Z > n\}$.

4.11* Une urne contient N boules numérotées de 1 à N. On en tire n au hasard ($n < N$) sans remise. Soit X le plus grand numéro tiré. Donner la fonction de répartition de X.

4.12* On lance une pièce de monnaie équitable n fois. Soit X le nombre de fois que le résultat est pile. Donner un espace de probabilité $(\Omega, \mathcal{F}, P)$ et une variable aléatoire $X : \Omega \to \mathbb{R}$ qui correspondent à cette situation.

4.13* Une urne contient N boules numérotées de 1 à N. On en tire n au hasard ($n < N$) sans remise. Soit X le plus petit numéro tiré. Donner la fonction de répartition de X.

4.14* On lance un dé n fois. Soit X le produit des résultats. Donner un espace de probabilité $(\Omega, \mathcal{F}, P)$ et une variable aléatoire $X : \Omega \to \mathbb{R}$ qui correspondent à cette situation.

4.15 (Approximation poissonienne d'une loi binomiale) Pour $n \in \mathbb{N}$, soient $p_n \in\,]0, 1[$ et X_n une v.a. binomiale de paramètres n et p_n. Soient $\lambda > 0$ et Y une v.a. de Poisson de paramètre λ. Supposons que $\lim_{n\to\infty} p_n = 0$ et $\lim_{n\to\infty} np_n = \lambda$. Montrer que pour tout $k \in \mathbb{N}$,

$$\lim_{n\to\infty} P\{X_n = k\} = P\{Y = k\}\,.$$

Indication. Vérifier d'abord que si $\lim_{n\to\infty} \lambda_n = \lambda$, alors $\lim_{n\to\infty} \left(1 - \frac{\lambda_n}{n}\right)^n = e^{-\lambda}$.

4.16* En Suisse, il y a eu 39 800 mariages en 2006. En expliquant vos hypothèses, estimer :

a) la probabilité que pour exactement vingt-cinq de ces couples, les deux époux sont nés le 15 février ;

b) pour exactement vingt-cinq de ces couples, les deux époux ont le même jour d'anniversaire.

4.17* Un jour donné, 10 000 voitures effectuent un trajet dans la ville A ; un cinquième de ces voitures sont grises. On admet que la probabilité qu'une voiture ait un accident ce jour-là est 0,002. En expliquant vos hypothèses, estimer :

a) la probabilité que 15 voitures aient un accident ce jour-là ;

b) la probabilité que 3 voitures grises aient un accident ce jour-là.

4.18 Soit X une v.a. exponentielle de paramètre $\lambda > 0$. Quelle est la fonction de densité de $Y = X^2$?

4.19 Soit X une variable aléatoire dont la fonction de densité est

$$f(x) = \begin{cases} c(4 - x^2) & \text{si } -2 < x < 2\,, \\ 0 & \text{sinon}\,. \end{cases}$$

a) Quelle est la valeur de c ?

b) Quelle est la fonction de répartition de X ?

4.20 Soit X une variable aléatoire continue dont la fonction de densité est

$$f(x) = \begin{cases} c\cos^2(x) & \text{si } 0 < x < \frac{\pi}{2}\,, \\ 0 & \text{sinon}\,. \end{cases}$$

a) Quelle est la valeur de c ?

b) Quelle est la fonction de répartition de X ?

4.21* a) Soit $f : \mathbb{R} \to \mathbb{R}$ définie par

$$f(x) = \begin{cases} \dfrac{c}{\sqrt{x}} & \text{si } x \geqslant 1\,, \\ 0 & \text{si } x < 1\,. \end{cases}$$

Est-ce qu'il existe une valeur de c pour laquelle f est une fonction de densité ?

b) Soit $F : \mathbb{R} \to \mathbb{R}$ définie par

$$F(x) = \begin{cases} e^{-1/x} & \text{si } x > 0\,, \\ 0 & \text{si } x \leqslant 0\,. \end{cases}$$

Est-ce que F est une fonction de répartition ? Si oui, quelle est sa fonction de densité ?

4.22* a) Soit $f : \mathbb{R} \to \mathbb{R}$ définie par :

$$f(x) = \begin{cases} \dfrac{c}{1 + x^2} & \text{si } x \geqslant 0\,, \\ 0 & \text{si } x < 0\,. \end{cases}$$

Est-ce qu'il existe une valeur de c pour laquelle f est une fonction de densité ?

b) Soit $F : \mathbb{R} \to \mathbb{R}$ définie par :

$$F(x) = \frac{1}{2}\left(1 + \frac{x}{\sqrt{1+x^2}}\right), \qquad x \in \mathbb{R}\,.$$

Est-ce que F est une fonction de répartition ? Si oui, quelle est sa fonction de densité ?

4.23 Soit X une variable aléatoire exponentielle de paramètre λ. Pour $x > 0$ et $y > 0$, montrer que $P\{X > x + y \mid X > y\} = P\{X > x\}$ (« propriété d'absence de mémoire »)

4.24 Montrer que

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx = 1\,.$$

Indication. Calculer le carré de l'intégrale et utiliser les coordonnées polaires dans le plan.

4.25 a) (Loi log-normale) Soit X une v.a. normale standard. Trouver la fonction de densité de la v.a. $Y = e^X$.

b) Soient X une v.a. normale standard et Z une v.a. qui est solution de l'équation $Z^3 + Z + 1 = X$. Trouver la fonction de densité de Z.

4.26 a) Soit X une v.a. exponentielle de paramètre $\lambda > 0$. Trouver la fonction de densité de la v.a. $Y = \ln(X)$.

b) Soient X une v.a. normale standard et Z une v.a. à valeurs dans $]-\frac{\pi}{2}, \frac{\pi}{2}]$ qui est solution de l'équation $Z + \tan(Z) = X$. Trouver la fonction de densité de Z.

4.27 a) Soit f la fonction de densité d'une v.a. continue X quelconque. Trouver la fonction de densité de la v.a. $Y = X^2$.

b) Soit X une v.a. normale standard. Montrer que $Y = X^2$ est une v.a. de loi gamma dont on identifiera les paramètres.

4.28 Soit X une v.a. $N(\mu = 10, \sigma^2 = 36)$. Calculer :

a) $P\{X < 13\}$;
b) $P\{4 < X < 16\}$;
c) $P\{X < 8\}$;
d) $P\{7 < X < 35\}$.

4.29* Soit X une v.a. $N(\mu = 4, \sigma^2 = 16)$. Calculer :

a) $P\{X < 6\}$;
b) $P\{2 < X < 8\}$;
c) $P\{X < 0\}$;
d) $P\{1 < X < 28\}$.

4.30* Soit F une fonction de répartition. Pour $y \in \,]0,1]$, on définit une fonction *inverse généralisée* de F en posant

$$F^{-1}(y) = \inf\{x \in \mathbb{R} : F(x) \geqslant y\}\,.$$

a) Soit U une v.a. uniforme sur $[0,1]$. Trouver la fonction de répartition de la v.a. $Y = F^{-1}(U)$.

b) Supposons que F est continue et soit X une v.a. dont la fonction de répartition est F. Montrer que la v.a. $Z = F(X)$ est uniforme sur $[0,1]$.

Chapitre 5

Vecteurs aléatoires

Dès que l'on s'intéresse à plusieurs v.a. différentes, il est naturel de considérer qu'elles sont les composantes d'un vecteur aléatoire, ce qui amène à la définition suivante.

Définition 5.1

Un vecteur aléatoire (à m composantes) est une application $X : \Omega \to \mathbb{R}^m$, que l'on écrit $X = (X_1, \ldots, X_m)$ et dont chaque composante $X_i : \Omega \to \mathbb{R}$ est une v.a.

5.1 Fonctions de répartition conjointe et marginale

Définition 5.2

La fonction de répartition conjointe *du vecteur aléatoire X est la fonction $F : \mathbb{R}^m \to \mathbb{R}$ définie par*

$$F(x_1, \ldots, x_m) = P\{X_1 \leqslant x_1, \ldots, X_m \leqslant x_m\} \qquad \left(= P\left(\bigcap_{i=1}^{m}\{X_i \leqslant x_i\}\right)\right).$$

Propriétés des fonctions de répartition conjointes

Proposition 5.3

a) *Pour $i = 1, \ldots, m$, la fonction partielle $x_i \mapsto F(x_1, \ldots, x_i, \ldots, x_m)$ est non décroissante (les autres composantes étant fixées).*

b) $\lim_{x_1 \to +\infty, \ldots, x_m \to +\infty} F(x_1, \ldots, x_m) = 1$ *et* $\lim_{x_1 \to -\infty, \ldots, x_m \to -\infty} F(x_1, \ldots, x_m) = 0$.

c) $\lim_{y_1 \searrow x_1, \ldots, y_m \searrow x_m} F(y_1, \ldots, y_m) = F(x_1, \ldots, x_m)$.

d) Cas où $m = 2$. *Soient $a_1 < b_1, a_2 < b_2$. Alors*

$$\begin{aligned} &P\big\{X_1 \in \left]a_1, b_1\right],\ X_2 \in \left]a_2, b_2\right]\big\} \\ &\quad = P\{a_1 < X_1 \leqslant b_1,\ a_2 < X_2 \leqslant b_2\} \\ &\quad = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2)\,. \end{aligned}$$

Démonstration. Les démonstrations des affirmations a)-c) sont similaires à celles des v.a. (qui correspondent au cas $m = 1$). En ce qui concerne d), remarquons que

$$P\{X_1 \leqslant b_1,\ X_2 \leqslant a_2\} + P\{X_1 \leqslant b_1,\ a_2 < X_2 \leqslant b_2\} = P\{X_1 \leqslant b_1,\ X_2 \leqslant b_2\}\,. \tag{5.1}$$

Le second membre est égal à $F(b_1, b_2)$. Par conséquent,

$$P\{X_1 \leqslant b_1,\ a_2 < X_2 \leqslant b_2\} = F(b_1, b_2) - F(b_1, a_2)\,. \tag{5.2}$$

En remplaçant b_1 par a_1, on déduit que

$$P\{X_1 \leqslant a_1,\ a_2 < X_2 \leqslant b_2\} = F(a_1, b_2) - F(a_1, a_2)\,. \tag{5.3}$$

Comme en (5.1), on constate que

$$\begin{aligned} &P\{X_1 \leqslant a_1,\ a_2 < X_2 \leqslant b_2\} + P\{a_1 < X_1 \leqslant b_1,\ a_2 < X_2 \leqslant b_2\} \\ &\qquad = P\{X_1 \leqslant b_1,\ a_2 < X_2 \leqslant b_2\}\,. \end{aligned}$$

En remplaçant le premier terme du membre de gauche par l'expression donnée en (5.3) et le membre de droite par l'expression donnée en (5.2), on obtient bien la propriété d). □

Définition 5.4
Soit $X = (X_1, \ldots, X_m)$ un vecteur aléatoire à valeurs dans $\mathbb{R}^m$. Les fonctions de répartition des v.a. X_i sont appelées les fonctions de répartition marginales *de X.*

Proposition 5.5 *Soient $X = (X_1, \ldots, X_m)$ un vecteur aléatoire à valeurs dans $\mathbb{R}^m$, F_X sa fonction de répartition et $F_{X_1}, \ldots, F_{X_m}$ ses fonctions de répartition marginales. Alors pour tout $x_i \in \mathbb{R}$,*

$$F_{X_i}(x_i) = F_X(+\infty, \ldots, +\infty, x_i, +\infty, \ldots, +\infty)\,.$$

Remarque 5.6 Par définition,

$$F_X(+\infty, \ldots, +\infty, x_i, +\infty, \ldots, +\infty) = \lim_{x_j \to +\infty,\, \forall j \neq i} F(x_1, \ldots, x_i, \ldots, x_m)\,.$$

Démonstration. Supposons, sans restreindre la généralité, que $i = 1$ et observons que

$$\begin{aligned} F_{X_1}(x_1) &= P\{X_1 \leqslant x_1\} = P\{X_1 \leqslant x_1,\ X_2 < +\infty, \ldots, X_m < +\infty\} \\ &= P\big\{(X_1, \ldots, X_m) \in\,]-\infty, x_1] \times\,]-\infty, +\infty[\, \times \cdots \times\,]-\infty, +\infty[\big\}\,. \end{aligned}$$

D'après la continuité de la probabilité (théorème 1.6a), le membre de droite est égal à

$$\begin{aligned} &\lim_{x_2 \to +\infty, \ldots, x_m \to +\infty} P\big\{(X_1, \ldots, X_m) \in\,]-\infty, x_1] \times\,]-\infty, x_2] \times \cdots \times\,]-\infty, x_m]\big\} \\ &\qquad = \lim_{x_2 \to +\infty, \ldots, x_m \to +\infty} F_X(x_1, x_2, \ldots, x_m) = F_X(x_1, +\infty, \ldots, +\infty)\,. \end{aligned}$$

La proposition est démontrée. □

5.2 Vecteurs aléatoires discrets

Définition 5.7

a) Un vecteur aléatoire discret *est un vecteur aléatoire* $X = (X_1, \ldots, X_m)$ *dont toutes les composantes* X_i *sont des v.a. discrètes.*

b) La fonction de densité simultanée *(ou* conjointe) *de* X *est la fonction* $f : \mathbb{R}^m \to \mathbb{R}_+$ *définie par*

$$\begin{aligned} f(x_1, \ldots, x_n) &= P\{X_1 = x_1, \ldots, X_m = x_n\} \\ &= P\big\{(X_1, \ldots, X_m) = (x_1, \ldots x_m)\big\}\,. \end{aligned}$$

c) La fonction de densité marginale *de* X_i *est la fonction* $f_{X_i} : \mathbb{R} \to \mathbb{R}_+$ *définie par*

$$\begin{aligned} f_{X_i}(x_i) &= P\{X_i = x_i\} \\ &= \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_m} f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m)\,. \end{aligned}$$

où les sommes portent sur les $(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m)$ *tels que* $f(x_1, \ldots, x_m) > 0$.

Exemple 5.8 Soit $X = (X_1, X_2)$ un vecteur aléatoire qui prend ses valeurs dans $\{5; 5{,}5; 6\} \times \{100, 125, 150, 175\}$, avec les probabilités suivantes :

$X_1 \setminus X_2$	100	125	150	175	$f_{X_1}(x_1)$
5	0,08	0,08	0,06	0	0,22
5,5	0,08	0,16	0,16	0,08	0,48
6	0	0,08	0,1	0,12	0,3
$f_{X_2}(x_2)$	0,16	0,32	0,32	0,2	1

Observons que la somme des probabilités contenues dans la partie centrale du tableau est égale à 1 et que $f_X(x_1, x_2)$ est égale à la valeur indiquée dans le tableau si $(x_1, x_2) \in \{5; 5{,}5; 6\} \times \{100, 125, 150, 175\}$ et $f_X(x_1, x_2) = 0$ sinon.

Exemple 5.9 *Vecteur aléatoire multinomial.* Une expérience a m résultats possibles, avec probabilités respectives $p_1, \ldots, p_m$, où $p_1 + \cdots + p_m = 1$. On réalise n épreuves indépendantes.

Pour $i = 1, \ldots, m$, soit X_i le nombre de fois qu'on obtient un résultat de type i. D'après la définition de la loi multinomiale (voir p. 33), la fonction de densité conjointe du vecteur aléatoire $X = (X_1, \ldots, X_m)$ est

$$f(n_1, \ldots, n_m) = \begin{cases} \dbinom{n}{n_1 \ \cdots \ n_m} p_1^{n_1} \cdots p_m^{n_m} & \text{si } n_1 + \cdots + n_m = n\,,\ n_i \in \mathbb{N}\,, \\ 0 & \text{sinon} \end{cases}$$

Remarquons que la loi marginale de X_i est $B(n, p_i)$.

5.3 Vecteurs aléatoires continus

Définition 5.10

On dit que X est un vecteur aléatoire continu *avec* fonction de densité conjointe $f : \mathbb{R}^m \to \mathbb{R}_+$ *si*

$$P\{X \in A\} = \int_A f(x_1, \ldots, x_m)\, dx_1 \cdots dx_m$$

pour tout $A \subset \mathbb{R}^m$ (pour lequel l'intégrale est bien définie au sens de Riemann).

Proposition 5.11

a) Soit X un vecteur aléatoire continu avec fonction de densité conjointe f et fonction de répartition conjointe F. Alors

$$F(x_1, \ldots, x_m) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} f(y_1, \ldots, y_m)\, dy_1 \cdots dy_m\,. \tag{5.4}$$

b) La fonction de densité conjointe f se calcule à partir de la fonction de répartition conjointe F par la formule

$$f(x_1, \ldots, x_m) = \frac{\partial^m}{\partial x_1 \cdots \partial x_m} F(x_1, \ldots, x_m)\,.$$

c) La fonction de densité marginale de X_i est

$$\begin{aligned} f_{X_i}(x_i) = \int_{-\infty}^{+\infty} dy_1 \cdots \int_{-\infty}^{+\infty} dy_{i-1} \int_{-\infty}^{+\infty} dy_{i+1} \cdots \int_{-\infty}^{+\infty} dy_m \\ \times f(y_1, \ldots, y_{i-1}, x_i, y_{i-1}, \ldots, y_m)\,. \end{aligned}$$

d) L'intégrale de la fonction de densité conjointe sur tout $\mathbb{R}^m$ est égale à 1 :

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_m)\, dx_1 \cdots dx_m = 1\,.$$

Démonstration. a) D'après la définition de la fonction de répartition,

$$\begin{aligned} F(x_1, \ldots, x_m) &= P\{X_1 \leqslant x_1, \ldots, X_m \leqslant x_m\} \\ &= P\big\{(X_1, \ldots, X_m) \in\,]-\infty, x_1] \times \cdots \times\,]-\infty, x_m]\big\} \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} f(y_1, \ldots, y_m)\, dy_1 \cdots dy_m\,. \end{aligned}$$

b) Il suffit de dériver successivement la formule (5.4) par rapport à chacune des variables $x_1, \ldots, x_n$ pour obtenir le résultat.

c) On utilise la définition de la fonction de densité pour écrire

$$
\begin{aligned}
&P\{X_i \leqslant x_i\} \\
&= P\Big\{(X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots X_m) \\
&\qquad\qquad \in \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{i-1 \text{ facteurs}} \times]-\infty, x_i] \times \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{m-i \text{ facteurs}}\Big\} \\
&= \int_{-\infty}^{+\infty} dy_1 \cdots \int_{-\infty}^{+\infty} dy_{i-1} \int_{-\infty}^{x_i} dy_i \int_{-\infty}^{+\infty} dy_{i+1} \cdots \int_{-\infty}^{+\infty} dy_m \, f(y_1, \ldots, y_m)\,.
\end{aligned}
$$

La fonction de densité marginale de X_i est $f_{X_i}(x_i) = \frac{d}{dx_i} P\{X_i \leqslant x_i\}$. En dérivant l'expression ci-dessus, on obtient le résultat désiré.

d) L'intégrale sur $\mathbb{R}^m$ de la fonction de densité est égale à $P\{X \in \mathbb{R}^m\} = 1$.

□

Exemples importants

Exemple 5.12 *Vecteur aléatoire uniforme sur $S \subset \mathbb{R}^2$.* Soit S un sous-ensemble de $\mathbb{R}^2$ d'aire a positive et finie. Un vecteur aléatoire $X = (X_1, X_2)$ est un *vecteur aléatoire uniforme* sur S s'il admet la fonction de densité

$$
f(x_1, x_2) = \frac{1}{a} 1_S(x_1, x_2) = \begin{cases} \dfrac{1}{a} & \text{si } (x_1, x_2) \in S\,, \\ 0 & \text{sinon}\,. \end{cases}
$$

Ce vecteur aléatoire est tel que

$$
P\{X \in A\} = \frac{1}{a} \int_A 1_S(x_1, x_2)\, dx_1\, dx_2 = \frac{\text{aire de } A \cap S}{\text{aire de } S}\,.
$$

Le calcul des probabilités liées à ce vecteur aléatoire se ramènera donc généralement à un calcul d'aire. Par exemple, dans le cas où $S = [0,1]^2$, on peut s'intéresser à

$$
P\{X_1 \leqslant X_2\} = P\big\{(X_1, X_2) \in T\big\}\,,
$$

où $T = \big\{(x_1, x_2) \in [0,1]^2 : x_1 \leqslant x_2\big\}$ (fig. 5.1).

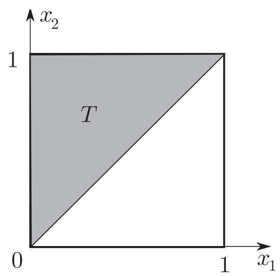


Figure 5.1 Ensemble T de l'exemple 5.12.

L'aire de S est $a = 1$ et l'aire de T est

$$\int_0^1 dx_1 \int_{x_1}^1 dx_2\, 1 = \int_0^1 dx_1\, (1 - x_1) = -\frac{(1-x_1)^2}{2}\Big|_0^1 = \frac{1}{2}\,.$$

Par conséquent, $P\{X_1 \leqslant X_2\} = \frac{1}{2}$.

La densité marginale de X_1 est

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_2 = \begin{cases} 0 & \text{si } x_1 \notin [0,1]\,, \\ \displaystyle\int_0^1 1\, dx_2 = 1 & \text{si } x_1 \in [0,1]\,. \end{cases}$$

Exemple 5.13

a) *Vecteur aléatoire normal standard dans* $\mathbb{R}^2$. On appelle ainsi tout vecteur aléatoire $X = (X_1, X_2)$ dont la fonction de densité est

$$f(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) = \frac{1}{2\pi} \exp\left(-\frac{\|x\|^2}{2}\right),$$

où $\|x\| = \sqrt{x_1^2 + x_2^2}$ et $x = (x_1, x_2)$.
La densité marginale de X_1 est alors

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{+\infty} dx_2\, f(x_1, x_2) = \int_{-\infty}^{+\infty} \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) dx_2 \\ &= \frac{1}{\sqrt{2\pi}}\, e^{-x_1^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x_2^2/2}\, dx_2 \\ &= \frac{1}{\sqrt{2\pi}}\, e^{-x_1^2/2}\,, \end{aligned}$$

et ceci est aussi la densité marginale de X_2. Ainsi, les composantes d'un vecteur aléatoire normal standard dans $\mathbb{R}^2$ sont des v.a. de loi $N(0,1)$.

b) *Vecteur aléatoire normal standard dans* $\mathbb{R}^m$. On appelle ainsi tout vecteur aléatoire $X = (X_1, \ldots, X_m)$ dont la fonction de densité est

$$f(x_1, \ldots, x_m) = \frac{1}{(2\pi)^{m/2}} \exp\left(-\frac{\|x\|^2}{2}\right),$$

où $\|x\| = \sqrt{x_1^2 + \cdots + x_m^2}$ et $x = (x_1, \ldots, x_m)$. La densité marginale de chaque X_i est $N(0,1)$. Ainsi, les composantes d'un vecteur aléatoire normal standard dans $\mathbb{R}^m$ sont des v.a. de loi $N(0,1)$.

5.4 Fonctions d'un vecteur aléatoire continu

Proposition 5.14 *Soit* $X = (X_1, \ldots, X_m)$ *un vecteur aléatoire continu avec fonction de densité conjointe* $f_X : \mathbb{R}^m \to \mathbb{R}$. *Soit* $U \subset \mathbb{R}^m$ *ouvert tel que*

$P\{X \in U\} = 1$. Soit $V \subset \mathbb{R}^m$ un autre sous-ensemble ouvert et $g : U \to V$ bijective :

$$g(x_1, \ldots, x_m) = \big(g_1(x_1, \ldots, x_m), \ldots, g_m(x_1, \ldots, x_m)\big)\,.$$

Posons $Y = g(X)$ et $h = g^{-1}$ (donc $h(Y) = X$). Si h est continûment différentiable, alors Y est un vecteur aléatoire continu dont la fonction de densité conjointe est

$$f_Y(y) = \begin{cases} f_X\big(h(y)\big) \cdot \big|J_h(y)\big| & \text{si } y \in V\,, \\ 0 & \text{sinon}\,, \end{cases}$$

où J_h est le jacobien de h. On rappelle que le jacobien *J_h de la fonction h est le nombre*

$$J_h(y) = \det \begin{pmatrix} \dfrac{\partial h_1}{\partial y_1} & \cdots & \dfrac{\partial h_m}{\partial y_1} \\ \vdots & & \vdots \\ \dfrac{\partial h_1}{\partial y_m} & \cdots & \dfrac{\partial h_m}{\partial y_m} \end{pmatrix}.$$

Démonstration. Soit $A \subset V$. Alors

$$\begin{aligned} P\{Y \in A\} &= P\{g(X) \in A\} = P\big\{X \in g^{-1}(A)\big\} \\ &= \int_{g^{-1}(A)} f_X(x_1, \ldots, x_n)\, dx_1 \cdots dx_m\,. \end{aligned}$$

Puisque h est bijective, le changement de variables $x = h(y)$ permet de voir que cette intégrale est égale à

$$\int_A f_X\big(h(y)\big)\,\big|J_h(y)\big|\,dy_1 \cdots dy_n\,.$$

Par conséquent, $f_Y(y) = f_X\big(h(y)\big)\,\big|J_h(y)\big|$ si $y \in V$. □

5.5 Vecteur aléatoire gaussien

Définition 5.15

Soient $X = (X_1, \ldots, X_m)$ un vecteur aléatoire normal standard dans $\mathbb{R}^m$, $b \in \mathbb{R}^m$ et A une matrice carrée inversible de type $m \times m$. Posons $Y = A \cdot X + b$. On dit que Y est un vecteur aléatoire gaussien *(ou* normal*) dans $\mathbb{R}^m$ et on note $Y : N(b, AA^T)$.*

Proposition 5.16 *Soit $Y : N(b, AA^T)$. Alors la fonction de densité conjointe de Y est*

$$f_Y(y) = \frac{1}{(2\pi)^{m/2}\,\big|\det(A)\big|}\,\exp\!\left(-\frac{1}{2}(y-b)^T\big(AA^T\big)^{-1}(y-b)\right), \qquad y \in \mathbb{R}^m\,.$$

Démonstration. On définit tout d'abord la fonction $g : \mathbb{R}^m \to \mathbb{R}^m$ par $g(x) = A \cdot x + b$, de sorte que g est inversible et $h(y) = g^{-1}(y) = A^{-1} \cdot (y - b)$. De plus, on vérifie par un calcul direct que $J_h(y) = \det(A^{-1}) = \frac{1}{\det(A)}$. D'après la proposition 5.14,

$$f_Y(y) = f_X\big(A^{-1}(y-b)\big) \cdot \big|J_h(y)\big| = \frac{1}{(2\pi)^{m/2}\,\big|\det(A)\big|} \exp\bigg(-\frac{1}{2}\big\|A^{-1}(y-b)\big\|^2\bigg).$$

Or, en se rappellant des propriétés du produit scalaire euclidien « $\cdot$ » et du produit matriciel, on constate que

$$\begin{aligned}\big\|A^{-1}(y-b)\big\|^2 &= \big(A^{-1}(y-b)\big) \cdot \big(A^{-1}(y-b)\big) = \big(A^{-1}(y-b)\big)^T \big(A^{-1}(y-b)\big)\\ &= (y-b)^T (A^{-1})^T A^{-1}(y-b)\\ &= (y-b)^T (AA^T)^{-1}(y-b)\,.\end{aligned}$$

Par conséquent, la densité du vecteur aléatoire Y est bien

$$f_Y(y) = \frac{1}{(2\pi)^{m/2}\big|\det(A)\big|} \exp\bigg(-\frac{1}{2}(y-b)^T \big(AA^T\big)^{-1}(y-b)\bigg), \quad y \in \mathbb{R}^m\,. \quad \square$$

Remarque 5.17 Les exercices 5.23, 6.30 et 8.8 traitent du vecteur aléatoire gaussien dans $\mathbb{R}^2$.

5.6 Variables aléatoires indépendantes

Définition 5.18

a) Soit $m \geqslant 2$. Les v.a. $X_1, \ldots, X_m$ sont indépendantes *si pour tous intervalles $A_1 \subset \mathbb{R}, \ldots, A_m \subset \mathbb{R}$,*

$$P\{X_1 \in A_1, \ldots, X_m \in A_m\} = \prod_{i=1}^{m} P\{X_i \in A_i\}\,.$$

b) Les v.a. $(X_n, n \in \mathbb{N}^*)$ sont indépendantes *si pour tout $m \geqslant 2$, les v.a. $X_1, \ldots, X_m$ sont indépendantes.*

Proposition 5.19 *Soit $X = (X_1, \ldots, X_m)$ un vecteur aléatoire discret ou continu (avec fonction de densité (resp. de répartition) conjointe f_X (resp. F_X)). Alors les trois propriétés suivantes sont équivalentes :*

a) les v.a. $X_1, \ldots, X_m$ sont indépendantes ;

b) pour tout $(x_1, \ldots x_m) \in \mathbb{R}^m$, $F_X(x_1, \ldots, x_m) = F_{X_1}(x_1) \cdots F_{X_m}(x_m)$;

c) pour tout $(x_1, \ldots x_m) \in \mathbb{R}^m$, $f_X(x_1, \ldots x_m) = f_{X_1}(x_1) \cdots f_{X_m}(x_m)$.

Démonstration. On va vérifier les trois implications a)⇒b), b)⇒c) et c)⇒a).

a)⇒b) On utilise la définition de F_X et la propriété a) pour écrire

$$\begin{aligned} F_X(x_1,\dots,x_m) &= P\{X_1 \leqslant x_1,\dots,X_m \leqslant x_m\} \\ &= P\big\{X_1 \in\,]-\infty,x_1],\,\dots,\,X_m \in\,]-\infty,x_m]\big\} \\ &= P\big\{X_1 \in\,]-\infty,x_1]\big\}\cdots P\big\{X_m \in\,]-\infty,x_m]\big\} \\ &= F_{X_1}(x_1)\cdots F_{X_m}(x_m)\,, \end{aligned}$$

ce qui établit b).

b)⇒c) Voir l'exercice 5.15.

c)⇒a) On se place d'abord dans le cas discret. Soient $A_1 \subset \mathbb{R},\dots,A_m \subset \mathbb{R}$ et, pour $i = 1,\dots,m$, soit D_i un ensemble dénombrable tel que $P\{X_i \in D_i\} = 1$. Alors

$$\begin{aligned} &P\{X_1 \in A_1,\dots,X_m \in A_m\} \\ &\qquad = \sum_{x_1\in A_1\cap D_1}\cdots\sum_{x_m\in A_m\cap D_m} P\{X_1 = x_1,\dots,X_m = x_m\} \\ &\qquad = \sum_{x_1\in A_1\cap D_1}\cdots\sum_{x_m\in A_m\cap D_m} f_X(x_1,\dots,x_m)\,. \end{aligned}$$

D'après la propriété c), ceci est égal à

$$\begin{aligned} &\sum_{x_1\in A_1\cap D_1}\cdots\sum_{x_m\in A_m\cap D_m} f_{X_1}(x_1)\cdots f_{X_m}(x_m) \\ &\qquad = \sum_{x_1\in A_1\cap D_1}\cdots\sum_{x_m\in A_m\cap D_m} P\{X_1 = x_1\}\cdots P\{X_m = x_m\} \\ &\qquad = \Bigg(\sum_{x_1\in A_1\cap D_1} P\{X_1 = x_1\}\Bigg)\cdots\Bigg(\sum_{x_m\in A_m\cap D_m} P\{X_m = x_m\}\Bigg) \\ &\qquad = P\{X_1 \in A_1\}\cdots P\{X_m \in A_m\}\,, \end{aligned}$$

ce qui établit bien a).

On se place maintenant dans le cas continu. Observons que

$$\begin{aligned} P\{X_1 \in A_1,\dots,X_m \in A_m\} &= P\big\{(X_1,\dots X_m) \in A_1\times\cdots\times A_m\big\} \\ &= \int_{A_1}\cdots\int_{A_m} f_X(x_1,\dots,x_m)\,dx_1\cdots dx_m\,. \end{aligned}$$

D'après la propriété c), ceci est égal à

$$\begin{aligned} &\int_{A_1}\cdots\int_{A_m} f_{X_1}(x_1)\cdots f_{X_m}(x_m)\,dx_1\cdots dx_m \\ &\qquad = \bigg(\int_{A_1} dx_1\, f_{X_1}(x_1)\bigg)\cdots\bigg(\int_{A_m} dx_m\, f_{X_m}(x_m)\bigg) \\ &\qquad = P\{X_1 \in A_1\}\cdots P\{X_m \in A_m\}\,. \end{aligned}$$

□

Fonctions de familles indépendantes de variables aléatoires

A plusieurs reprises, nous serons amenés à travailler avec des fonctions de familles indépendantes de v.a. Celles-ci restent indépendantes dans les conditions de la proposition suivante.

Proposition 5.20 *Soient $X_1, \dots, X_m$, $Y_1, \dots, Y_n$ des v.a. indépendantes, $g_1 : \mathbb{R}^m \to \mathbb{R}$, $g_2 : \mathbb{R}^n \to \mathbb{R}$. On pose $Z_1 = g_1(X_1, \dots, X_m)$ et $Z_2 = g_2(Y_1, \dots, Y_n)$. Alors Z_1 et Z_2 sont indépendantes.*

Démonstration. La démonstration de cette proposition fait l'objet de l'exercice 5.14.

Statistiques d'ordre

Définition 5.21

Soient $X_1, \dots, X_m$ des v.a. continues indépendantes et identiquement distribuées (ce que l'on notera i.i.d.*), toutes de densité f.* Les statistiques d'ordre *associées à $X_1, \dots, X_m$ sont les composantes du vecteur $(X_{(1)}, X_{(2)}, \dots, X_{(m)})$ définies par*

— *$X_{(1)}$ est la plus petite des valeurs des X_j ($= \min_{j=1,\dots,m} X_j$) ;*

— *$X_{(2)}$ est la seconde plus petite des valeurs des X_j ;*

— *$X_{(i)}$ est la $i^{\text{ième}}$ plus petite des valeurs des X_j ;*

— *$X_{(m)}$ est la valeur la plus grande des X_j ($= \max_{j=1,\dots,m} X_j$).*

En particulier, les inégalités $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(m)}$ sont vérifiées (et on peut même remplacer les inégalités larges par des inégalités strictes).

Proposition 5.22 *La densité du vecteur aléatoire $(X_{(1)}, \dots, X_{(m)})$ est*

$$f_{(X_{(1)},\dots,X_{(m)})}(x_1, \dots, x_m) = \begin{cases} m!\, f(x_1) \cdots f(x_m) & \text{si } x_1 < \cdots < x_m\,, \\ 0 & \text{sinon}\,. \end{cases}$$

Démonstration. Soient $x_1 < y_1 < x_2 < y_2 < \cdots < x_m < y_m$ des nombres et $I_j =]x_j, y_j]$ des intervalles, $j = 1, \dots, m$. Alors

$$P\{X_{(1)} \in I_1, \dots, X_{(m)} \in I_m\} = P\left(\bigcup \{X_{\pi(1)} \in I_1, \dots, X_{\pi(m)} \in I_m\}\right),$$

où la réunion porte sur toutes les permutations π de $\{1, \dots, m\}$. Puisque les intervalles I_j sont deux à deux disjoints, cette réunion porte sur des événements disjoints. L'additivité de la probabilité implique donc que le membre de droite est égal à

$$\sum P\{X_{\pi(1)} \in I_1, \dots, X_{\pi(m)} \in I_m\}\,,$$

où la somme porte à nouveau sur toutes les permutations π de $\{1, \dots, m\}$. Soit F la fonction de répartition de X_1. Vu l'indépendance des X_j et vu qu'il y a $m!$ permutations de $\{1, \dots, m\}$, cette somme est égale à

$$\begin{aligned}
\sum P\{X_{\pi(1)} \in I_1\} \cdots P\{X_{\pi(m)} \in I_m\} \\
&= m!\big(F(y_1) - F(x_1)\big) \cdots \big(F(y_m) - F(x_m)\big) \\
&= m! \left(\int_{x_1}^{y_1} du_1\, f(u_1)\right) \cdots \left(\int_{x_m}^{y_m} dy_m\, f(u_m)\right) \\
&= \int_{x_1}^{y_1} \cdots \int_{x_m}^{y_m} \big(m!\, f(u_1) \cdots f(u_m)\big)\, du_1 \cdots du_m\,.
\end{aligned}$$

D'où le résultat. □

5.7 Lois conditionnelles

Tout comme la probabilité conditionnelle permet de déterminer comment le fait de savoir qu'un événement s'est réalisé modifie notre opinion sur la probabilité d'un autre événement, la notion de loi conditionnelle décrit comment la connaissance de la valeur d'une v.a. affecte notre opinion concernant la loi d'une autre v.a. Nous allons étudier cette notion séparément dans les cas des vecteurs aléatoires discrets et continus.

Cas discret

Définition 5.23

Soit $X = (X_1, X_2)$ un vecteur aléatoire discret à valeurs dans $\mathbb{R}^2$. Soit $x_2 \in \mathbb{R}$ tel que $P\{X_2 = x_2\} > 0$. La fonction de densité conditionnelle de X_1 sachant $X_2 = x_2$ *est*

$$f_{X_1}(x_1 \mid X_2 = x_2) \overset{déf}{=} P\{X_1 = x_1 \mid X_2 = x_2\} = \frac{f_X(x_1, x_2)}{f_{X_2}(x_2)}\,.$$

La fonction de répartition conditionnelle de X_1 sachant $X_2 = x_2$ *est*

$$F_{X_1}(x_1 \mid X_2 = x_2) \overset{déf}{=} P\{X_1 \leqslant x_1 \mid X_2 = x_2\} = \sum_{y_1 \leqslant x_1} f_{X_1}(y_1 \mid X_2 = x_2)\,.$$

Remarque 5.24 Pour x_2 fixé, $x_1 \mapsto F_{X_1}(x_1 | X_2 = x_2)$ est bien une fonction de répartition. En effet, cette fonction est bien non décroissante, continue à droite et tend vers 0 (resp. 1) lorsque $x \to -\infty$ (resp. $x \to +\infty$).

Cas continu

Définition 5.25

Soit $X = (X_1, X_2)$ un vecteur aléatoire continu à valeurs dans $\mathbb{R}^2$, ayant f_X pour fonction de densité conjointe. Soit $x_2 \in \mathbb{R}$ tel que $f_{X_2}(x_2) > 0$. La fonction de densité conditionnelle de X_1 sachant $X_2 = x_2$ *est*

$$f_{X_1}(x_1 \mid X_2 = x_2) = \frac{f_X(x_1, x_2)}{f_{X_2}(x_2)}\,.$$

Remarque 5.26 Pour x_2 fixé avec $f_{X_2}(x_2) > 0$, la fonction $x_1 \mapsto f_{X_1}(x_1 \mid X_1 = x_2)$ est bien une fonction de densité. En effet, cette fonction est non négative et

$$\begin{aligned}\int_{-\infty}^{+\infty} f_{X_1}(x_1 \mid X_2 = x_2)\, dx_1 &= \int_{-\infty}^{+\infty} \frac{f_X(x_1, x_2)}{f_{X_2}(x_2)}\, dx_1 \\ &= \frac{1}{f_{X_2}(x_2)} \int_{-\infty}^{+\infty} f_X(x_1, x_2)\, dx_1 = \frac{f_{X_2}(x_2)}{f_{X_2}(x_2)} = 1\,.\end{aligned}$$

Définition 5.27

On définit la loi conditionnelle de X_1 sachant que $X_2 = x_2$ *par*

$$P\{X_1 \in A_1 \mid X_2 = x_2\} \stackrel{\textit{déf}}{=} \int_{A_1} f_{X_1}(x_1 \mid X_2 = x_2)\, dx_1\,,$$

et la fonction de répartition conditionnelle de X_1 sachant que $X_2 = x_2$ *par*

$$F_{X_1}(x_1 \mid X_2 = x_2) \stackrel{\textit{déf}}{=} \int_{-\infty}^{x_1} f_{X_1}(y_1 \mid X_2 = x_2)\, dy_1.$$

Exemple 5.28 Soit $X = (X_1, X_2)$ un vecteur aléatoire uniforme sur $B(0,1)$, c'est-à-dire un vecteur aléatoire dont la fonction de densité est (exemple 5.12)

$$f_X(x_1, x_2) = \begin{cases} \dfrac{1}{\pi} & \text{si } x_1^2 + x_2^2 < 1\,, \\ 0 & \text{sinon}\,. \end{cases}$$

Calculons $P\{X_1 \geqslant \frac{1}{2} \mid X_2 = \frac{1}{2}\}$.

On observe d'abord que la fonction de densité marginale de X_2 est

$$\begin{aligned}f_{X_2}(x_2) &= \int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_1 \\ &= \begin{cases} \displaystyle\int_{-\sqrt{1-x_2^2}}^{\sqrt{1-x_2^2}} \frac{1}{\pi}\, dx_1 = \frac{2}{\pi}\sqrt{1 - x_2^2} & \text{si } -1 < x_2 < 1\,, \\ 0 & \text{sinon}\,. \end{cases}\end{aligned}$$

Par conséquent, pour $-1 < x_2 < 1$:

$$f_{X_1}(x_1 \mid X_2 = x_2) = \begin{cases} \dfrac{1}{2\sqrt{1 - x_2^2}} & \text{si } |x_1| < \sqrt{1 - x_2^2}\,, \\ 0 & \text{sinon} \end{cases}$$

et

$$f_{X_1}\left(x_1 \,\middle|\, X_2 = \frac{1}{2}\right) = \begin{cases} \dfrac{1}{2\sqrt{\frac{3}{4}}} = \dfrac{1}{\sqrt{3}} & \text{si } |x_1| < \sqrt{\frac{3}{4}}\,, \\ 0 & \text{sinon}\,. \end{cases}$$

Donc

$$P\left\{X_1 \geqslant \frac{1}{2} \,\middle|\, X_2 = \frac{1}{2}\right\} = \int_{1/2}^{\sqrt{3/4}} \frac{1}{\sqrt{3}}\, dx_1 = \frac{1}{\sqrt{3}}\left(\frac{\sqrt{3}}{2} - \frac{1}{2}\right).$$

Cas mixte discret/continu

Dans certains cas, on est confronté à un vecteur aléatoire dont certaines composantes sont des v.a. discrètes et d'autres sont des v.a. continues. Par exemple, soit $X = (X_1, X_2)$ un vecteur aléatoire à valeurs dans $\mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$ où :

— X_1 est un vecteur aléatoire discret à valeurs dans $\mathbb{R}^{m_1}$,

— X_2 est un vecteur aléatoire continu à valeurs dans $\mathbb{R}^{m_2}$.

Définition 5.29

a) La fonction $f : \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \to \mathbb{R}$ est la fonction de densité (mixte) discrète/continue *de X si pour tout $x_1 \in \mathbb{R}^{m_1}$ et $A_2 \subset \mathbb{R}^{m_2}$*

$$P\{X_1 = x_1, X_2 \in A_2\} = \int_{A_2} f(x_1, x_2)\, dx_2\,.$$

b) Soit D un sous-ensemble dénombrable de $\mathbb{R}^{m_1}$ tel que $P\{X_1 \in D\} = 1$. Les fonctions de densités marginales *de X_1 et X_2 sont*

$$f_{X_1}(x_1) = \int_{\mathbb{R}^{m_2}} f(x_1, x_2)\, dx_2\,,$$
$$f_{X_2}(x_2) = \sum_{x_1 \in D} f(x_1, x_2)\,.$$

c) Les fonctions de densité conditionnelles *sont*

$$f_{X_1}(x_1 \mid X_2 = x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)}\,,$$
$$f_{X_2}(x_2 \mid X_1 = x_1) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)}\,.$$

Avec cette définition, on peut travailler avec la loi conditionnelle de X_1 sachant $X_2 = x_2$ de la même manière qu'avec toute loi discrète. De même, on peut travailler avec la loi conditionnelle de X_2 sachant $X_1 = x_1$ de la même manière qu'avec toute loi continue.

5.8 Exercices

5.1 On lance deux dés équilibrés. Soit X_1 (respectivement X_2) le plus petit (resp. le plus grand) des deux résultats.

a) Quelle est la fonction de densité conjointe de (X_1, X_2) ?

b) Quelles sont les fonctions de densité marginales de X_1 et X_2 ?

5.2 On tire successivement avec remise deux boules d'une urne contenant trois boules numérotées respectivement 2, 3 et 4. Soient X_1 la somme des numéros et X_2 le produit des numéros.

a) Quelle est la fonction de densité conjointe de (X_1, X_2) ?

b) Quelles sont les fonctions de densité marginales de X_1 et X_2 ?

5.3 Soit (X_1, X_2) un vecteur aléatoire continu ayant pour fonction de densité

$$f(x_1, x_2) = \begin{cases} c(x_1 + x_2) & \text{si } (x_1, x_2) \in [0,1]^2\,, \\ 0 & \text{sinon}\,. \end{cases}$$

a) Que vaut c ?

b) Calculer $P\{X_1 < X_2\}$.

c) Quelle est la fonction de densité marginale de X_1 ?

d) Que vaut $P\{X_1 = X_2\}$?

5.4 Soit (X_1, X_2) un vecteur aléatoire continu ayant pour fonction de densité

$$f(x_1, x_2) = \begin{cases} c(x_1 + x_1 x_2) & \text{si } (x_1, x_2) \in [0,1]^2\,, \\ 0 & \text{sinon}\,. \end{cases}$$

a) Que vaut c ?

b) Calculer $P\left\{X_2 > X_1 + \frac{1}{2}\right\}$.

c) Quelle est la fonction de densité marginale de X_2 ?

d) Que vaut $P\left\{X_2 = X_1 + \frac{1}{2}\right\}$?

5.5 a) Soit (X_1, X_2) un vecteur aléatoire uniforme sur $[0,1]^2$. Quelle est la fonction de répartition de $X_1 + X_2$? Quelle est sa fonction de densité ?

b) Plus généralement, soit $X = (X_1, X_2)$ un vecteur aléatoire continu dans $\mathbb{R}^2$ avec fonction de densité conjointe f_X. Montrer que la fonction de densité de $Y = X_1 + X_2$ est

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(x_1, y - x_1)\, dx_1\,.$$

5.6 Soit $X = (X_1, X_2)$ un vecteur aléatoire uniforme sur $[0,1]^2$. On pose

$$Y_1 = X_1 + X_2, \qquad \text{et} \qquad Y_2 = X_1 - X_2.$$

a) Trouver la fonction de densité conjointe de $Y = (Y_1, Y_2)$.

b) Montrer que Y est un vecteur aléatoire uniforme sur un domaine que l'on déterminera.

5.7 On note D le disque unité dans $\mathbb{R}^2$, i.e. $D = \{x \in \mathbb{R}^2 : \|x\| < 1\}$. Soit $X = (X_1, X_2)$ un vecteur aléatoire uniforme sur D. On note R et Θ les coordonnées polaires de X. Trouver la fonction de densité conjointe de $Y = (R, \Theta)$.

5.8* (Méthode de Box-Müller) Soit (X_1, X_2) un vecteur aléatoire uniforme sur $[0,1]^2$. Posons $Y_1 = \sqrt{-2\ln(X_1)}\cos(2\pi X_2)$ et $Y_2 = \sqrt{-2\ln(X_1)}\sin(2\pi X_2)$. Montrer que (Y_1, Y_2) est un vecteur aléatoire normal standard dans $\mathbb{R}^2$.

5.9* Soit $X = (X_1, \ldots, X_m)$ un vecteur aléatoire continu avec fonction de densité conjointe f_X. Soit $g : \mathbb{R}^m \to \mathbb{R}$ continue. On pose $Y = g(X)$ et on s'intéresse à la fonction de densité f_Y de Y. On admet que l'on dispose d'une paramétrisation des courbes de niveau de g, c'est-à-dire d'une fonction $\varphi : \mathbb{R}^m \to \mathbb{R}^m$ bijective et de classe C^1 telle que $g\big(\varphi(z_1, \ldots, z_m)\big) = z_m$.

a) Montrer que $\varphi^{-1}\big(g^{-1}\big(]-\infty, y]\big)\big) = \mathbb{R}^{m-1} \times\,]-\infty, y]$.

b) Montrer que la fonction de densité f_Y est donnée par la formule

$$f_Y(y) = \int_{-\infty}^{\infty} dz_1 \cdots \int_{-\infty}^{\infty} dz_{m-1} f_X\big(\varphi(z_1, \ldots, z_{m-1}, y)\big) \big|J_\varphi(z_1, \ldots, z_{m-1}, y)\big| \,,$$

où J_φ est la matrice jacobienne de φ.

c) Dans le cas où $m = 2$ et $g(x_1, x_2) = x_1 + x_2$, donner explicitement une paramétrisation φ des courbes de niveau de g et retrouver le résultat de l'exercice 5.5b).

5.10 Soient X_1 et X_2 les v.a. définies dans l'exercice 5.1. Est-ce que X_1 et X_2 sont indépendantes ?

5.11 Soient X_1 et X_2 les v.a. définies dans l'exercice 5.2. Est-ce que X_1 et X_2 sont indépendantes ?

5.12 Soient X_1 et X_2 deux v.a. indépendantes.

a) Exprimer la fonction de répartition de $X = \max(X_1, X_2)$ en fonction de celles de X_1 et X_2.

b) Même question, mais pour $Y = \min(X_1, X_2)$.

c) Dans le cas où X_1 et X_2 sont des v.a. continues (avec fonctions de densité f_{X_1} et f_{X_2}), exprimer les fonctions de densité de X et Y en fonction de celles de X_1 et X_2.

5.13 (Convolution de densités) Soient X_1 et X_2 deux v.a. indépendantes (avec fonctions de densité f_{X_1} et f_{X_2}). D'abord dans le cas discret, puis dans le cas continu, exprimer la fonction de densité de $X_1 + X_2$ à l'aide des fonctions de densité de X_1 et X_2.

5.14 Démontrer la proposition 5.20.

5.15* Soit $X = (X_1, \ldots, X_n)$ un vecteur aléatoire discret ou continu avec fonction de densité (resp. de répartition) conjointe f_X (resp. F_X). Montrer l'équivalence des trois propriétés suivantes en complétant la démonstration de la proposition 5.19.

a) Les v.a. $X_1, \ldots, X_n$ sont indépendantes.

b) Pour tout $(x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$F_X(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n) \,.$$

c) Pour tout $(x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$f_X(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)\,.$$

5.16 Soit (X_1, X_2) un vecteur aléatoire continu avec fonction de densité conjointe $f(x_1, x_2)$. On suppose qu'il existe deux fonctions g_1 et g_2 telles que, pour tous $x_1, x_2 \in \mathbb{R}$, $f(x_1, x_2) = g_1(x_1)g_2(x_2)$. Montrer que X_1 et X_2 sont indépendantes.

5.17 Soit $X = (X_1, X_2)$ un vecteur aléatoire normal standard dans $\mathbb{R}^2$. Montrer que X_1 et X_2 sont indépendantes et chacune de loi $N(0,1)$.

5.18 Soient $X_1, \ldots, X_n$ des v.a. continues i.i.d. avec fonction de densité f. Montrer que

$$P\{X_1 < X_2 < \cdots < X_n\} = \frac{1}{n!}\,.$$

5.19 Alain et Bernard se donnent rendez-vous entre 18 h et 19 h dans leur restaurant favori. Ils arrivent indépendamment l'un de l'autre, chacun à un instant uniformément réparti entre 18 h et 19 h. Le premier arrivé n'attendra pas l'autre plus de 15 minutes. Quelle est la probabilité qu'ils se rencontrent dans le restaurant ?

5.20 Pour se rendre à son travail, Jacques peut soit prendre les transports publics (un bus et un métro consécutivement), soit aller à pied. La ligne de bus a une fréquence de trois passages par heure, la ligne de métro de six passages par heure. Jacques arrive à l'arrêt de bus à 8 h 00 précises et il ne connaît pas l'horaire de passage exact des bus, ni des métros. Le trajet à pied (resp. en transports publics) dure 27 (resp. 12) minutes.

a) Quelle est la probabilité que prendre les transports publics lui fasse perdre du temps ?

b) Si les bus ont systématiquement deux minutes de retard, comment le résultat de a) change-t-il ?

5.21 (L'aiguille de Buffon) Des droites parallèles sont tracées sur le sol. Chaque droite est écartée de 1 unité des deux autres droites les plus proches. On laisse tomber une aiguille de longueur $L \leqslant 1$. Montrer que la probabilité que l'aiguille touche une des droites est $\frac{2L}{\pi}$.

5.22* a) Déterminer la loi de $X_1 + X_2$ lorsque X_1 et X_2 sont des v.a. de Poisson indépendantes de paramètres respectifs λ_1 et λ_2.

b) Déterminer la loi de $Y_1 + Y_2$ lorsque Y_1 et Y_2 sont des v.a. indépendantes respectivement $N(\mu_1, \sigma_1^2)$ et $N(\mu_2, \sigma_2^2)$.
Indication. Ecrire $Y_1 = \mu_1 + Z_1$, $Y_2 = \mu_2 + Z_2$, où Z_i est $N(0, \sigma_i^2)$ $(i = 1, 2)$.

5.23* Soit $X = \binom{X_1}{X_2}$ un vecteur aléatoire normal standard dans $\mathbb{R}^2$. On pose

$$\mu = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \qquad A = \begin{pmatrix}\sigma_1 & 0\\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2\end{pmatrix},$$

où $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1 > 0$, $\sigma_2 > 0$, $-1 < \rho < 1$, et $Y = \mu + AX$.

a) Montrer que la fonction de densité conjointe de Y est

$$f(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\frac{(y_1-\mu_1)(y_2-\mu_2)}{\sigma_1\sigma_2}\right]\right).$$

b) Montrer que Y_1 est une v.a. $N(\mu_1, \sigma_1^2)$ et que Y_2 est une v.a. $N(\mu_2, \sigma_2^2)$.

c) Montrer que si $\rho = 0$, alors Y_1 et Y_2 sont indépendantes.

5.24 Soient X_1 et X_2 deux v.a. indépendantes, où X_1 est de loi $B(n_1, p)$ et X_2 est de loi $B(n_2, p)$.

a) Quelle est la loi de $X_1 + X_2$?

b) Pour $n \in \{0, \dots, n_1 + n_2\}$, quelle est la loi conditionnelle de X_1 sachant $X_1 + X_2 = n$?

5.25 a) On reprend la donnée de l'exercice 5.1. Quelle est la fonction de densité conditionnelle $f_{X_2}(x_2 \mid X_1 = 2)$?

b) On reprend la donnée de l'exercice 5.3. Quelle est la fonction de densité conditionnelle de X_1 sachant $X_2 = \frac{1}{2}$? Calculer $P\{X_1 \leqslant x \mid X_2 = \frac{1}{2}\}$.

5.26 a) On reprend la donnée de l'exercice 5.2. Quelle est la fonction de densité conditionnelle $f_{X_2}(x_2 \mid X_1 = 6)$?

b) On reprend la donnée de l'exercice 5.4. Quelle est la fonction de densité conditionnelle de X_1 sachant $X_2 = \frac{1}{2}$? Calculer $P\{X_1 \leqslant x \mid X_2 = \frac{1}{2}\}$.

5.27 Soit X_1 une v.a. uniforme sur $[0, 1]$. Pour $0 < p < 1$, on suppose que si $X_1 = p$, alors X_2 est $B(1, p)$. Pour $x_2 \in \{0, 1\}$, trouver la fonction de densité conditionnelle de X_1 sachant $X_2 = x_2$.

5.28 Soit X_1 une v.a. uniforme sur $[0, 1]$. Pour $0 < p < 1$, on suppose que si $X_1 = p$, alors X_2 est une v.a. géométrique de paramètre p.

a) Trouver la fonction de densité conjointe $f_{X_1,X_2}(x_1, x_2)$.

b) Trouver la fonction de densité marginale de X_2.

Indication. Observer que $\displaystyle\int_0^1 (1-x)^{n-1} x\, dx = \frac{1}{n(n+1)}$.

c) Pour $x_2 \in \{1, 2, \dots\}$, trouver la fonction de densité conditionnelle de X_1 sachant $X_2 = x_2$.

Chapitre 6

Espérance mathématique

L'espérance mathématique est une des notions les plus importantes de la théorie des probabilités. Au niveau intuitif, lorsqu'on observe la valeur d'une v.a., cette valeur sera généralement située de part ou d'autre d'une valeur centrale qui est son espérance mathématique, que nous allons maintenant définir. Comme dans les chapitres précédents, nous distinguons le cas des v.a. discrètes de celui des v.a. continues.

6.1 Cas des variables aléatoires discrètes

Définition 6.1

*Soient X une v.a. discrète, f_X sa fonction de densité et $D = \big\{x \in \mathbb{R} : f_X(x) > 0\big\}$. Alors l'*espérance mathématique *(ou simplement l'*espérance*) de X est le nombre, noté $E[X]$, défini par*

$$E[X] = \sum_{x \in D} x P\{X = x\} = \sum_{x \in D} x\, f_X(x)\,.$$

Remarque 6.2 Si D n'est pas un ensemble fini, alors cette somme est une série avec une infinité de termes. Cette série converge (absolument) pour autant que

$$\sum_{x \in D} |x|\, P\{X = x\} < +\infty\,.$$

Sans cette convergence absolue, l'ordre dans lequel on effectuerait la somme aurait une importance. Ainsi, $E[X]$ est défini uniquement lorsque cette convergence absolue a lieu.

Exemple 6.3

a) Soit X le résultat du jet d'un dé. Alors $D = \{1, \dots, 6\}$ et $P\{X = x\} = \frac{1}{6}$ si $x \in D$. Par conséquent,

$$E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3{,}5\,.$$

b) *Variable aléatoire indicatrice.* Soit G un événement tel que $P(G) = p$. Posons $X = 1_G$, c'est-à-dire

$$X(\omega) = 1_G(\omega) = \begin{cases} 1 & \text{si } \omega \in G\,, \\ 0 & \text{sinon}\,. \end{cases}$$

Alors
$$E[X] = 1 \cdot P\{X = 1\} + 0 \cdot P\{X = 0\} = 1 \cdot p = p\,.$$

c) *Variable aléatoire géométrique de paramètre* $p \in \,]0,1]$. Rappelons que si X est une telle v.a. géométrique, alors $P\{X = k\} = (1-p)^{k-1}p$, où $k \in \mathbb{N}^*$. Posons $q = 1-p$. D'après la définition de l'espérance mathématique,

$$\begin{aligned} E[X] &= \sum_{k=1}^{\infty} kq^{k-1}p = p\sum_{k=1}^{\infty} kq^{k-1} = p\sum_{k=1}^{\infty} \frac{d}{dq}\,(q^k) = p\sum_{k=0}^{\infty} \frac{d}{dq}\,(q^k) \\ &= p\,\frac{d}{dq}\left(\sum_{k=0}^{\infty} q^k\right) = p\,\frac{d}{dq}\left(\frac{1}{1-q}\right) = p\,\frac{1}{(1-q)^2} = p\,\frac{1}{p^2} = \frac{1}{p}\,. \end{aligned}$$

Ainsi, l'espérance de X est $E[X] = \frac{1}{p}$: le temps moyen avant le premier succès dans une suite d'épreuves indépendantes est donc égal à l'inverse de la probabilité p de succès.

Espérance d'une fonction réelle d'un vecteur aléatoire discret

Théorème 6.4
Soient X un vecteur aléatoire discret à valeurs dans $\mathbb{R}^m$ et $D \subset \mathbb{R}^m$ un sous-ensemble dénombrable tel que $P\{X \in D\} = 1$. Soit $g : D \to \mathbb{R}$. Alors

$$E\big[g(X)\big] = \sum_{x \in D} g(x)P\{X = x\} = \sum_{x \in D} g(x) f_X(x)\,,$$

où f_X est la fonction de densité de X.

Remarque 6.5 Si D n'est pas fini, $E\big[g(X)\big]$ n'est définie que si

$$\sum_{x \in D} \big|g(x)\big| P\{X = x\} < \infty\,.$$

Démonstration. Posons $Y = g(X)$. Alors

$$P\{Y = y\} = P\big\{g(X) = y\big\}\,.$$

Or $\big\{g(X) = y\big\} = \big\{X \in g^{-1}(\{y\})\big\}$. Si $y \notin \mathrm{Im}(g)$, cet ensemble est vide et donc $P\{Y = y\} = 0$. Ainsi, l'ensemble des y tels que $P\{Y = y\} > 0$ est un sous-ensemble de $\mathrm{Im}(g) = g(D)$. C'est donc un ensemble dénombrable et Y est donc une v.a. discrète.

Par conséquent,

$$E[Y] = \sum_{y \in \operatorname{Im}(g)} yP\{Y = y\} = \sum_{y \in \operatorname{Im}(g)} yP\left\{X \in g^{-1}\bigl(\{y\}\bigr)\right\}.$$

Puisque $g^{-1}\bigl(\{y\}\bigr) \subset D$ est un ensemble dénombrable, le membre de droite s'écrit

$$\sum_{y \in \operatorname{Im}(g)} y \sum_{x \in g^{-1}(\{y\})} P\{X = x\} = \sum_{y \in \operatorname{Im}(g)} \sum_{x \in g^{-1}(\{y\})} yP\{X = x\}.$$

Au membre de droite, on peut remplacer y par $g(x)$, puisque $x \in g^{-1}\bigl(\{y\}\bigr)$, pour obtenir

$$\sum_{y \in \operatorname{Im}(g)} \sum_{x \in g^{-1}(\{y\})} g(x)P\{X = x\} = \sum_{x \in D} g(x)P\{X = x\}.$$

Le théorème est démontré. □

Corollaire 6.6 (Linéarité de l'espérance)

a) Soient X une v.a. discrète et $a, b \in \mathbb{R}$. Alors

$$E[aX + b] = aE[X] + b.$$

b) Soit (X_1, X_2) un vecteur aléatoire discret à valeurs dans $\mathbb{R}^2$. Alors

$$E[X_1 + X_2] = E[X_1] + E[X_2].$$

Démonstration. a) Soit $g : \mathbb{R} \to \mathbb{R}$ définie par $g(x) = ax + b$. D'après la proposition 6.4,

$$\begin{aligned}E[aX + b] = E\bigl[g(X)\bigr] &= \sum_x g(x)P\{X = x\} = \sum_x (ax + b)P\{X = x\} \\ &= \sum_x axP\{X = x\} + \sum_x bP\{X = x\} = aE[X] + b.\end{aligned}$$

b) Soit $h : \mathbb{R}^2 \to \mathbb{R}$ définie par $h(x_1, x_2) = x_1 + x_2$. D'après la proposition 6.4,

$$\begin{aligned}&E[X_1 + X_2] \\ &\quad = E\bigl[h(X_1, X_2)\bigr] = \sum_{(x_1, x_2)} h(x_1, x_2)P\{X_1 = x_1,\, X_2 = x_2\} \\ &\quad = \sum_{x_1}\sum_{x_2}(x_1 + x_2)P\{X_1 = x_1,\, X_2 = x_2\} \\ &\quad = \sum_{x_1}\sum_{x_2} x_1 P\{X_1 = x_1,\, X_2 = x_2\} + \sum_{x_1}\sum_{x_2} x_2 P\{X_1 = x_1,\, X_2 = x_2\} \\ &\quad = \sum_{x_1} x_1 \sum_{x_2} P\{X_1 = x_1,\, X_2 = x_2\} + \sum_{x_2} x_2 \sum_{x_1} P\{X_1 = x_1,\, X_2 = x_2\} \\ &\quad = \sum_{x_1} x_1 P\{X_1 = x_1\} + \sum_{x_2} x_2 P\{X_2 = x_2\} \\ &\quad = E[X_1] + E[X_2].\end{aligned}$$

□

6.2 Espérance d'une variable aléatoire continue

On va procéder par analogie avec le cas discret.

Définition 6.7

*Soit X une v.a. continue avec fonction de densité f_X. Alors l'*espérance *de X est le nombre*

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)\, dx \,.$$

Remarque 6.8 $E[X]$ est définie pour autant que $\int_{-\infty}^{+\infty} |x| f_X(x)\, dx < +\infty$.

Exemple 6.9

a) *Variable aléatoire uniforme sur* $[a, b]$. Rappelons que la fonction de densité d'une telle v.a. X est

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } a \leqslant x \leqslant b\,, \\ 0 & \text{sinon}\,. \end{cases}$$

L'espérance d'une telle v.a. est

$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} x f_X(x)\, dx = \int_a^b x\, \frac{1}{b-a}\, dx = \frac{1}{b-a}\, \frac{x^2}{2}\bigg|_a^b \\ &= \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}\,. \end{aligned}$$

b) *Variable aléatoire exponentielle de paramètre* λ. Rappelons que la fonction de densité d'une telle v.a. X est

$$f_X(x) = \begin{cases} \lambda\, e^{-\lambda x} & \text{si } x \geqslant 0\,, \\ 0 & \text{sinon}\,. \end{cases}$$

Son espérance est donc

$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} x f_X(x)\, dx = \int_0^{+\infty} x \lambda\, e^{-\lambda x}\, dx \\ &= -x\, e^{-\lambda x}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda x}\, dx = 0 + \frac{e^{-\lambda x}}{-\lambda}\bigg|_0^{+\infty} = \frac{1}{\lambda}\,. \end{aligned}$$

Espérance d'une fonction d'un vecteur aléatoire continu

Théorème 6.10

Soit X un vecteur aléatoire continu à valeurs dans $\mathbb{R}^m$ avec fonction de densité $f_X : \mathbb{R}^m \to \mathbb{R}_+$. Soit $g : \mathbb{R}^m \to \mathbb{R}$. Alors

$$E\big[g(X)\big] = \int_{\mathbb{R}^m} g(x) f_X(x)\, dx \,.$$

Remarque 6.11

a) $E\big[g(X)\big]$ est bien définie pour autant que

$$\int_{\mathbb{R}^m} \big|g(x)\big| f_X(x)\,dx < \infty\,.$$

b) Cette formule est analogue à celle du cas discret (proposition 6.4).

Démonstration. Nous ferons la démonstration dans deux cas.

1^{er} *cas* : $\mathrm{Im}(g)$ est un ensemble dénombrable. Dans ce cas, $Y = g(X)$ est une v.a. discrète puisque $P\big\{Y \in \mathrm{Im}(g)\big\} = 1$. Par conséquent,

$$\begin{aligned} E[Y] &= \sum_{y\in\mathrm{Im}(g)} yP\{Y = y\} = \sum_{y\in\mathrm{Im}(g)} yP\left\{X \in g^{-1}\big(\{y\}\big)\right\} \\ &= \sum_{y\in\mathrm{Im}(g)} y \int_{g^{-1}(\{y\})} f_X(x)\,dx = \sum_{y\in\mathrm{Im}(g)} \int_{g^{-1}(\{y\})} yf_X(x)\,dx\,. \end{aligned}$$

Puisque $g(x) = y$ lorsque $x \in g^{-1}\big(\{y\}\big)$, cette expression s'écrit aussi

$$\sum_{y\in\mathrm{Im}(g)} \int_{g^{-1}(\{y\})} g(x)f_X(x)\,dx = \int_{\mathbb{R}^m} g(x)f_X(x)\,dx\,.$$

Ceci démontre le théorème dans ce premier cas.

2^e *cas* : g vérifie les hypothèses de l'exercice 5.9. Sous les hypothèses de cet exercice, il existe $\varphi : \mathbb{R}^m \to \mathbb{R}^m$ bijective et de classe C^1 telle que $g(\varphi(z_1, \ldots, z_m)) = z_m$. Rappelons que cette hypothèse signifie en particulier que pour $y \in \mathbb{R}$ fixé, $(z_1, \ldots, z_{m-1}) \mapsto \varphi(z_1, \ldots, z_{m-1}, y)$ est une paramétrisation de la courbe de niveau C_y de g, qui est, par définition, $C_y = \big\{z \in \mathbb{R}^m : g(z) = y\big\}$. Posons $Y = g(X)$. L'exercice affirme que la fonction de densité de Y est

$$f_Y(y) = \int_{-\infty}^{+\infty} dz_1 \cdots \int_{-\infty}^{+\infty} dz_{m-1}\, f_X\big(\varphi(z_1, \ldots, z_{m-1}, y)\big)\, \big|J_\varphi(z_1, \ldots, z_{m-1}, y)\big|\,.$$

Donc

$$\begin{aligned} E[Y] = \int_{-\infty}^{+\infty} yf(y)\,dy = \int_{-\infty}^{+\infty} dy \int_{-\infty}^{+\infty} dz_1 \cdots \int_{-\infty}^{+\infty} dz_{m-1} \\ \times\, yf_X\big(\varphi(z_1, \ldots, z_{m-1}, y)\big)\, \big|J_\varphi(z_1, \ldots, z_{m-1}, y)\big|\,. \end{aligned}$$

Puisque $y = g\big(\varphi(z_1, \ldots, z_{m-1}, y)\big)$, ceci s'écrit

$$\begin{aligned} \int_{-\infty}^{+\infty} dy \int_{-\infty}^{+\infty} dz_1 \cdots \int_{-\infty}^{+\infty} dz_{m-1}\, g\big(\varphi(z_1, \ldots, z_{m-1}, y)\big) \\ \times\, f_X\big(\varphi(z_1, \ldots, z_{m-1}, y)\big)\big|J_\varphi(z_1, \ldots, z_{m-1}, y)\big|. \end{aligned}$$

Effectuons le changement de variables

$$(x_1, \ldots, x_m) = \varphi(z_1, \ldots, z_{m-1}, y)$$

pour constater que cette intégrale multiple est égale à

$$\int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_m \, g(x_1, \ldots, x_m) f_X(x_1, \ldots, x_m) = \int_{\mathbb{R}^m} g(x) f_X(x) \, dx \, .$$

Le théorème est démontré dans ce deuxième cas. □

Corollaire 6.12

a) Linéarité de l'espérance. *Soit X une v.a. continue avec fonction de densité f, $a, b \in \mathbb{R}$. Alors*

$$E[aX + b] = aE[X] + b \, .$$

b) Soit (X_1, X_2) un vecteur aléatoire continu dans $\mathbb{R}^2$ avec fonction de densité conjointe $f_{(X_1,X_2)}$. Alors

$$E[X_1 + X_2] = E[X_1] + E[X_2] \, .$$

Démonstration. Il suffit de reprendre la démonstration du cas discret (corollaire 6.6), de remplacer les sommes par des intégrales et les références à la proposition 6.4 par la proposition 6.10. □

Proposition 6.13 *Soient X_1 et X_2 deux v.a. indépendantes. Alors*

$$E[X_1 X_2] = E[X_1] E[X_2] \, .$$

De même, si $X_1, \ldots, X_n$ sont des v.a. indépendantes, alors

$$E[X_1 \cdots X_n] = E[X_1] \cdots E[X_n] \, .$$

Pour le démonstration de cette proposition, voir l'exercice 6.9.

Autres propriétés de l'espérance

a) Soit X une v.a. telle que $P\{X \geqslant 0\} = 1$. Alors $E[X] \geqslant 0$.

b) Soit (X_1, X_2) un vecteur aléatoire tel que $P\{X_1 \leqslant X_2\} = 1$. Alors $E[X_1] \leqslant E[X_2]$.

c) Soit X une v.a. Alors $\big|E[X]\big| \leqslant E\big[|X|\big]$.

La vérification de ces propriétés est laissée en exercice (cf. exercice 6.10).

6.3 Variance, covariance et corrélation

Définition 6.14

Soient X une v.a. telle que $E[X^2] < +\infty$, (X_1, X_2) un vecteur aléatoire tel que $E[X_i^2] < +\infty$, $i = 1, 2$.

a) La variance *de X est le nombre* $\mathrm{Var}(X) = E\left[\left(X - E[X]\right)^2\right]$*. Ce nombre représente l'écart quadratique moyen entre la valeur effective de X et son espérance.*

*b) L'*écart type *de X est le nombre $\sigma_X = \sqrt{\mathrm{Var}(X)}$.*

c) La covariance *entre X_1 et X_2 est le nombre*

$$\mathrm{Cov}(X_1, X_2) = E\left[\left(X_1 - E[X_1]\right)\left(X_2 - E[X_2]\right)\right] .$$

d) La corrélation *entre X_1 et X_2 est le nombre*

$$\mathrm{Corr}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)}\sqrt{\mathrm{Var}(X_2)}} .$$

Ce nombre est bien défini si $\mathrm{Var}(X_1) \neq 0 \neq \mathrm{Var}(X_2)$.

Proposition 6.15

a) $\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \mathrm{Cov}(X, X)$.

b) $\mathrm{Cov}(X_1, X_2) = E[X_1X_2] - E[X_1]E[X_2] = \mathrm{Cov}(X_2, X_1)$.

c) $\mathrm{Cov}(aX_1 + bX_2, Y) = a\ \mathrm{Cov}(X_1, Y) + b\ \mathrm{Cov}(X_2, Y)$.

d) $-1 \leqslant \mathrm{Corr}(X_1, X_2) \leqslant 1$.

e) $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$.

Démonstration. Pour les affirmations a) à d), voir l'exercice 6.11.

e) D'après la définition et la linéarité de l'espérance,

$$\begin{aligned}\mathrm{Var}(aX + b) &= E\left[\left(aX + b - E[aX + b]\right)^2\right] = E\left[\left(aX - aE[X]\right)^2\right] \\ &= a^2E\left[\left(X - E[X]\right)^2\right] = a^2\ \mathrm{Var}(X) .\end{aligned}$$

□

Remarque 6.16 Si dans e), on prend $a = 0$, alors on voit que $\mathrm{Var}(b) = 0$. Ainsi, une v.a. qui ne prend que la seule valeur b, c'est-à-dire une v.a. *déterministe,* est de variance nulle. La réciproque est aussi vraie (corollaire 6.18).

Théorème 6.17 Inégalité de Tchebychev

Pour tout $\varepsilon > 0$,

$$P\left\{\left|X - E[X]\right| \geqslant \varepsilon\right\} \leqslant \frac{\mathrm{Var}(X)}{\varepsilon^2} .$$

Démonstration. On utilise les propriétés de l'espérance mathématique pour écrire

$$P\left\{\left|X-E[X]\right| \geqslant \varepsilon\right\} = E[1_{\{|X-E[X]|\geqslant\varepsilon\}}]\,.$$

Sur l'événement $\left\{\left|X-E[X]\right| \geqslant \varepsilon\right\}$, l'inégalité $\frac{|X-E[X]|}{\varepsilon} \geqslant 1$ est vérifiée. Par conséquent,

$$1_{\{|X-E[X]|\geqslant\varepsilon\}} \leqslant \frac{\big(X-E[X]\big)^2}{\varepsilon^2}\,.$$

Vu la monotonie de l'espérance,

$$P\left\{\left|X-E[X]\right| \geqslant \varepsilon\right\} \leqslant E\left[\frac{\big(X-E[X]\big)^2}{\varepsilon^2}\right] = \frac{1}{\varepsilon^2}\operatorname{Var}(X)\,.$$

L'inégalité de Tchebychev est donc démontrée. □

Corollaire 6.18 *Si* $\operatorname{Var}(X)=0$, *alors* $P\big\{X=E[X]\big\}=1$.

Démonstration. Observons que

$$\big\{X \neq E[X]\big\} = \left\{\left|X-E[X]\right| > 0\right\} = \bigcup_{n\in\mathbb{N}^*}\left\{\left|X-E[X]\right| \geqslant \frac{1}{n}\right\}.$$

Or, pour tout $n \in \mathbb{N}^*$,

$$P\left\{\left|X-E[X]\right| \geqslant \frac{1}{n}\right\} \leqslant \frac{\operatorname{Var}(X)}{\left(\frac{1}{n}\right)^2} = 0\,.$$

Puisque les événements $\left\{\left|X-E[X]\right| \geqslant \frac{1}{n}\right\}$ sont emboîtés, la probabilité de leur réunion est égale, d'après la continuité de la probabilité, à

$$\lim_{n\to\infty} P\left\{\left|X-E[X]\right| > \frac{1}{n}\right\} = 0\,.$$

Ainsi, $P\big\{X \neq E[X]\big\} = 0$. Le corollaire est démontré. □

Proposition 6.19

a) $\operatorname{Var}(X_1 \pm X_2) = \operatorname{Var}(X_1) + \operatorname{Var}(X_2) \pm 2\operatorname{Cov}(X_1, X_2)$.

b) Si X_1 *et* X_2 *sont indépendantes, alors*

$$\operatorname{Var}(X_1 \pm X_2) = \operatorname{Var}(X_1) + \operatorname{Var}(X_2)$$

et pour $a, b \in \mathbb{R}$,

$$\operatorname{Var}(aX_1 + bX_2) = a^2\operatorname{Var}(X_1) + b^2\operatorname{Var}(X_2)\,.$$

c) Si $X_1, \dots, X_n$ sont des v.a. indépendantes, alors

$$\mathrm{Var}(X_1 \pm \cdots \pm X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)\,.$$

Démonstration. a) D'après les propriétés a) et c) de la proposition 6.15,

$$\begin{aligned}\mathrm{Var}(X_1 \pm X_2) &= \mathrm{Cov}(X_1 \pm X_2, X_1 \pm X_2)\\ &= \mathrm{Cov}(X_1, X_1 \pm X_2) \pm \mathrm{Cov}(X_2, X_1 \pm X_2)\\ &= \mathrm{Cov}(X_1, X_1) \pm \mathrm{Cov}(X_1, X_2) \pm \mathrm{Cov}(X_2, X_1) + \mathrm{Cov}(X_2, X_2)\\ &= \mathrm{Var}(X_1) \pm 2\,\mathrm{Cov}(X_1, X_2) + \mathrm{Var}(X_2)\,.\end{aligned}$$

b) Si X_1 et X_2 sont indépendantes, alors $E[X_1X_2] = E[X_1]E[X_2]$ d'après la proposition 6.13, et donc

$$\mathrm{Cov}(X_1, X_2) = E[X_1X_2] - E[X_1]E[X_2] = 0\,.$$

D'après a),

$$\mathrm{Var}(X_1 \pm X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)\,.$$

De même,

$$\mathrm{Var}(aX_1 + bX_2) = \mathrm{Var}(aX_1) + \mathrm{Var}(bX_2)\,.$$

D'après la propriété e) de la proposition 6.15, le membre de droite est égal à

$$a^2\,\mathrm{Var}(X_1) + b^2\,\mathrm{Var}(X_2)\,,$$

ce qui démontre b).

c) D'après la proposition 5.20, $X_1 \pm \cdots \pm X_{n-1}$ est indépendante de X_n. D'après la propriété b),

$$\mathrm{Var}(X_1 \pm \cdots \pm X_n) = \mathrm{Var}(X_1 \pm \cdots \pm X_{n-1}) + \mathrm{Var}(X_n)\,.$$

Un raisonnement par récurrence démontre donc c). □

Exemple 6.20 Soit X une v.a. de loi $B(n,p)$. Alors $X = X_1 + \cdots + X_n$, où X_i est la v.a. indicatrice d'un succès à l'épreuve i. La v.a. X_i est donc de loi $B(1,p)$. Les v.a. X_i sont indépendantes, $E[X_i] = p$ et $\mathrm{Var}(X_i) = p(1-p)$ (cf. exercice 6.3a). Par conséquent, $E[X] = np$ et $\mathrm{Var}(X) = np(1-p)$.

Remarque 6.21

a) Si $\mathrm{Cov}(X_1, X_2) = 0$, alors $\mathrm{Corr}(X_1, X_2) = 0$ et on dit que X_1 et X_2 sont *non corrélées.*
b) Si X_1 et X_2 sont indépendantes, alors elles sont non corrélées. La réciproque n'est pas vraie en général (cf. exercice 9.10).

Proposition 6.22 *Si $\big|\mathrm{Corr}(X_1, X_2)\big| = 1$, alors il existe deux nombres $a_1 \neq 0$ et $a_2 \in \mathbb{R}$ tels que $X_2 = a_1X_1 + a_2$ (c'est-à-dire $P\{X_2 = a_1X_1 + a_2\} = 1$).*

Démonstration. Observons que

$$\left|\mathrm{Corr}(X_1, X_2)\right| = 1 \iff \left|\mathrm{Cov}(X_1, X_2)\right| = \sqrt{\mathrm{Var}(X_1)}\sqrt{\mathrm{Var}(X_2)} \neq 0\,. \quad (6.1)$$

Soit $\lambda \in \mathbb{R}$. Alors

$$\begin{aligned}\mathrm{Var}(\lambda X_1 + X_2) &= \mathrm{Var}(\lambda X_1) + \mathrm{Var}(X_2) + 2\,\mathrm{Cov}(\lambda X_1, X_2) \\ &= \lambda^2\,\mathrm{Var}(X_1) + 2\lambda\,\mathrm{Cov}(X_1, X_2) + \mathrm{Var}(X_2)\,.\end{aligned}$$

D'après l'hypothèse $\left|\mathrm{Corr}(X_1, X_2)\right| = 1$ et (6.1), le discriminant réduit de ce polynôme du second degré en λ est

$$\Delta' = \big(\mathrm{Cov}(X_1, X_2)\big)^2 - \mathrm{Var}(X_1)\,\mathrm{Var}(X_2) = 0\,.$$

Il existe donc $\lambda_1 \in \mathbb{R}$ tel que $\mathrm{Var}(\lambda_1 X_1 + X_2) = 0$. D'après le corollaire 6.18,

$$P\big\{\lambda_1 X_1 + X_2 = E[\lambda_1 X_1 + X_2]\big\} = 1\,.$$

Il suffit pour conclure de noter que $a_1 = -\lambda_1 \neq 0$ puisque $\mathrm{Var}(X_2) \neq 0$ et de poser $a_2 = E[\lambda_1 X_1 + X_2]$. □

6.4 Moments d'une variable aléatoire

Définition 6.23
Soient $k \in \mathbb{N}$ et X une v.a. discrète ou continue avec fonction de densité f, telle que $E\big[|X|^k\big] < +\infty$. Le moment d'ordre k *de X est le nombre $E[X^k]$.*

Remarque 6.24

a) Le *premier moment* de X est $E[X]$, le *deuxième moment* de X est $E[X^2] = \mathrm{Var}(X) + (E[X])^2$.
b) Le moment d'ordre k de X n'est bien défini que si $E\big[|X|^k\big] < \infty$. Dans ce cas, pour $0 \leqslant \ell \leqslant k$, $E\big[|X|^\ell\big]$ est aussi fini, d'après l'exercice 6.18.

Fonction génératrice des moments

Définition 6.25
La fonction génératrice des moments *d'une v.a. X est la fonction $M_X : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ définie par*

$$M_X(t) = E[e^{tX}]\,.$$

Proposition 6.26 *S'il existe $t_0 > 0$ tel que $E[e^{tX}] < \infty$ pour $t \in\,]-t_0, t_0[$, alors*

$$M'_X(0) = E[X]\,, \quad M''_X(0) = E[X^2]\,, \quad \ldots\,, \quad M^{(k)}_X(0) = E[X^k]\,.$$

Démonstration. On vérifiera que l'hypothèse permet de permuter les opérations d'espérance et de dérivation dans le calcul suivant :

$$M'_X(t) = \frac{d}{dt}\, E[e^{tX}] = E\left[\frac{d}{dt}\, e^{tX}\right] = E[X\, e^{tX}]\,,$$

et donc $M'_X(0) = E[X]$.

De même,

$$\begin{aligned} M''_X(t) &= \frac{d}{dt}\, E[X\, e^{tX}] = E\left[\frac{d}{dt}\,(Xe^{tX})\right] \\ &= E[X^2\, e^{tX}]\,, \end{aligned}$$

donc $M''_X(0) = E[X^2]$. On procède ensuite par induction pour obtenir l'égalité $M_X^{(k)}(0) = E[X^k]$, pour tout $k \in \mathbb{N}$.

En ce qui concerne la possibilité de permuter espérance et dérivée, on va examiner séparément les cas des v.a. discrètes et continues.

Cas discret. Dans ce cas, soit D un ensemble dénombrable tel que $P\{X \in D\} = 1$. Alors

$$\frac{d}{dt}\, E[e^{tX}] = \frac{d}{dt}\left(\sum_{x\in D} e^{tx} f_X(x)\right)$$

et l'égalité

$$\frac{d}{dt}\left(\sum_{x\in D} e^{tx} f_X(x)\right) = \sum_{x\in D} \frac{d}{dt}\, e^{tx} f_X(x)$$

est vérifiée si D est fini ou, s'il existe $h > 0$ tel que $\sum_{x\in D} e^{h|x|} f_X(x) < \infty$. Cette dernière condition est une conséquence de l'hypothèse de la proposition et dans ce cas, l'égalité vaut pour $|t| < t_0$.

Cas continu. Dans ce cas,

$$\frac{d}{dt}\, E[e^{tX}] = \frac{d}{dt}\left(\int_{-\infty}^{+\infty} e^{tx} f_X(x)\, dx\right)$$

et, s'il existe $h > 0$ tel que $\int_{-\infty}^{+\infty} e^{h|x|} f_Y(x)\, dx < \infty$, alors l'égalité

$$\frac{d}{dt}\left(\int_{-\infty}^{+\infty} e^{tx} f_X(x)\, dx\right) = \int_{-\infty}^{+\infty} \frac{d}{dt}\,(e^{tx}) f_X(x)\, dx$$

est vraie pour $|t| < h$. Cette dernière condition est une conséquence de l'hypothèse de la proposition. □

Remarque 6.27 La tabelle A.1 de l'appendice donne la fonction génératrice des moments des lois principales.

Théorème 6.28

Soient X_1 et X_2 deux v.a. indépendantes. Alors

$$M_{X_1+X_2}(t) = M_{X_1}(t)M_{X_2}(t)\,.$$

Démonstration. Comme les v.a. X_1 et X_2 sont indépendantes, les v.a. e^{tX_1} et e^{tX_2} sont aussi indépendantes, d'après la proposition 5.20. En utilisant la proposition 6.13, on obtient

$$\begin{aligned} M_{X_1+X_2}(t) &= E[e^{t(X_1+X_2)}] = E[e^{tX_1}\, e^{tX_2}] \\ &= E[e^{tX_1}]E[e^{tX_2}] = M_{X_1}(t)M_{X_2}(t)\,. \end{aligned}$$

□

Théorème 6.29

Soient X_1 et X_2 deux v.a. telles que $M_{X_1}(t) = M_{X_2}(t) < +\infty$ pour $|t| < t_0$, où $t_0 > 0$. Alors X_1 et X_2 ont la même fonction de répartition (et donc la même loi).

Démonstration*. Nous allons démontrer ce théorème dans le cas où $M_{X_1}(t) = M_{X_2}(t) < +\infty$ pour tout $t \in \mathbb{R}$. Notons F_1 (resp. F_2) la fonction de répartition de la v.a. X_1 (resp. X_2). Nous supposerons également que la fonction de répartition F_1 est continue, admet f_1 pour fonction de densité et que la v.a. X_1 ne prend que des valeurs positives (la remarque 6.30 discute la situation où on ne fait pas ces hypothèses supplémentaires).

Par définition et d'après les hypothèses que nous venons de faire,

$$M_{X_1}(t) = E[e^{tX_1}] = \int_0^\infty e^{ty} f_1(y)\, dy\,.$$

Dans la démonstration de la proposition 6.26, nous avons vu que la fonction M_{X_1} est indéfiniment dérivable et admet pour dérivées

$$M_{X_1}^{(k)}(t) = \int_0^\infty y^k\, e^{ty} f_1(y)\, dy\,, \tag{6.2}$$

pour tout $k \in \mathbb{N}$. On considère alors la fonction $G : (\mathbb{R}_+)^2 \to \mathbb{R}$ définie par

$$G(t,x) = \sum_{k=0}^{[tx]} \frac{t^k}{k!}\, M_{X_1}^{(k)}(-t)\,. \tag{6.3}$$

Soit $x \in \mathbb{R}_+$ fixé. Nous allons montrer que $\lim_{t\to\infty} G(t,x) = F_1(x)$.

En remplaçant (6.2) dans (6.3), on obtient

$$\begin{aligned} G(t,x) &= \sum_{k=0}^{\lfloor tx \rfloor} \frac{t^k}{k!} \int_0^\infty y^k e^{-ty} f_1(y)\,dy \\ &= \int_0^\infty \left(\sum_{k=0}^{\lfloor tx \rfloor} \frac{(ty)^k}{k!} e^{-ty} \right) f_1(y)\,dy = \int_0^\infty \varphi_t(x,y) f_1(y)\,dy\,, \end{aligned}$$

où

$$\varphi_t(x,y) = \sum_{k=0}^{[tx]} \frac{(ty)^k}{k!} e^{-ty}\,.$$

Soit $Y_{t,y}$ une v.a. de loi de Poisson de paramètre ty. Nous pouvons écrire $\varphi_t(x,y) = P\{Y_{t,y} \leqslant tx\}$. Ainsi, dans le cas où $y > x$, par l'inégalité de Tchebychev (proposition 6.17),

$$\begin{aligned} 0 \leqslant \varphi_t(x,y) &= P\{Y_{t,y} \leqslant tx\} \\ &= P\big\{Y_{t,y} - ty \leqslant t(x-y)\big\} \leqslant P\big\{|Y_{t,y} - ty| \geqslant t(y-x)\big\} \\ &\leqslant \frac{\mathrm{Var}(Y_{t,y})}{t^2(y-x)^2} = \frac{y}{t(y-x)^2}\,. \end{aligned} \tag{6.4}$$

De même, dans le cas où $y < x$,

$$\begin{aligned} 0 \leqslant 1 - \varphi_t(x,y) &= P\{Y_{t,y} > tx\} \\ &= P\big\{Y_{t,y} - ty > t(x-y)\big\} \leqslant P\big\{|Y_{t,y} - ty| > t(x-y)\big\} \\ &\leqslant \frac{\mathrm{Var}(Y_{t,y})}{t^2(x-y)^2} = \frac{y}{t(x-y)^2}\,. \end{aligned} \tag{6.5}$$

Soit alors $\varepsilon > 0$ fixé. Comme F_1 est continue, choisissons $h > 0$ (dépendant de x et ε) suffisamment petit pour que $F_1(x+h) - F_1(x) < \frac{\varepsilon}{3}$ et $F_1(x) - F_1(x-h) < \frac{\varepsilon}{3}$. Choisissons ensuite $t > 0$ (dépendant de x, h et ε) suffisamment grand pour que $t > \frac{6x}{\varepsilon h^2}$. Alors,

$$\begin{aligned} \big|G(t,x) - F_1(x)\big| &= \left| \int_0^\infty \varphi_t(x,y) f_1(y)\,dy - \int_0^x f_1(y)\,dy \right| \\ &= \left| \int_0^{x-h} \big(\varphi_t(x,y) - 1\big) f_1(y)\,dy + \int_{x-h}^x \big(\varphi_t(x,y) - 1\big) f_1(y)\,dy \right. \\ &\qquad \left. + \int_x^{x+h} \varphi_t(x,y) f_1(y)\,dy + \int_{x+h}^\infty \varphi_t(x,y) f_1(y)\,dy \right| \\ &\leqslant \int_0^{x-h} \big|\varphi_t(x,y) - 1\big| f_1(y)\,dy + \int_{x-h}^x \big|\varphi_t(x,y) - 1\big| f_1(y)\,dy \\ &\qquad + \int_x^{x+h} \big|\varphi_t(x,y)\big| f_1(y)\,dy + \int_{x+h}^\infty \big|\varphi_t(x,y)\big| f_1(y)\,dy\,. \end{aligned} \tag{6.6}$$

D'après (6.5),

$$\begin{aligned}\int_0^{x-h} \bigl|\varphi_t(x,y)-1\bigr| f_1(y)\,dy &\leqslant \frac{1}{t}\int_0^{x-h} \frac{y}{(x-y)^2}\, f_1(y)\,dy\\ &\leqslant \frac{x-h}{th^2}\int_0^{x-h} f_1(y)\,dy\\ &\leqslant \frac{x-h}{th^2}\,. \end{aligned} \tag{6.7}$$

Un raisonnement analogue avec (6.4) montre que

$$\int_{x+h}^{\infty} \bigl|\varphi_t(x,y)\bigr| f_1(y)\,dy \leqslant \frac{x+h}{th^2}\,. \tag{6.8}$$

De plus, comme $\bigl|\varphi_t(x,y)-1\bigr| \leqslant 1$ et $\bigl|\varphi_t(x,y)\bigr| \leqslant 1$ pour tous x et y dans $\mathbb{R}_+$, nous avons

$$\int_{x-h}^{x} \bigl|\varphi_t(x,y)-1\bigr| f_1(y)\,dy \leqslant \int_{x-h}^{x} f_1(y)\,dy = F_1(x)-F_1(x-h) \tag{6.9}$$

et

$$\int_{x}^{x+h} \bigl|\varphi_t(x,y)\bigr| f_1(y)\,dy \leqslant \int_{x}^{x+h} f_1(y)\,dy = F_1(x+h)-F_1(x)\,. \tag{6.10}$$

En remplaçant (6.7)-(6.10) dans (6.6), on obtient

$$\begin{aligned}&\bigl|G(t,x)-F_1(x)\bigr|\\ &\quad\leqslant \frac{x-h}{th^2} + \bigl(F_1(x)-F_1(x-h)\bigr) + \bigl(F_1(x+h)-F_1(x)\bigr) + \frac{x+h}{th^2}\\ &\quad= \bigl(F_1(x)-F_1(x-h)\bigr) + \bigl(F_1(x+h)-F_1(x)\bigr) + \frac{2x}{th^2}\\ &\quad\leqslant \frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3} = \varepsilon\,.\end{aligned}$$

Ceci montre donc que $\lim_{t\to\infty} G(t,x) = F_1(x)$.

Comme la fonction G est univoquement déterminée par la fonction M_{X_1}, elle-même égale à M_{X_2}, nous obtenons

$$F_1(x) = \lim_{t\to\infty} G(t,x) = F_2(x)\,. \qquad \square$$

Remarque 6.30

a) Le raisonnement ci-dessus peut-être repris dans le cas où F_1 est une fonction de répartition discrète. Lorsque x est un point de continuité de F_1, il s'agit de reprendre le raisonnement et de remplacer les intégrales par des sommes. L'égalité de F_1 et F_2 aux points de discontinuité se déduit de la continuité à droite.

b) Le cas d'une v.a. non nécessairement positive exige un argument différent, voir [1, sect. 30].

c) La même référence traite le cas où $M_{X_1}(t) = M_{X_2}(t)$ pour $|t| < t_0$.

Exemple 6.31 (*Utilisation du théorème 6.29*)

a) Soient X_1 et X_2 deux v.a. indépendantes, où X_1 suit une loi de Poisson de paramètre λ_1 et X_2 une loi de Poisson de paramètre λ_2. On aimerait déterminer la loi de $X_1 + X_2$? (voir aussi l'exercice 5.22a).
On va calculer la fonction génératrice des moments de $X_1 + X_2$. D'après le théorème 6.28 et la tabelle A.1 de l'appendice,

$$M_{X_1+X_2}(t) = M_{X_1}(t)M_{X_2}(t) = e^{\lambda_1(e^t-1)}\, e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}\,.$$

On constate qu'il s'agit de la fonction génératrice des moments d'une v.a. de Poisson de paramètre $\lambda_1+\lambda_2$. D'après le théorème 6.29, on déduit que $X_1 + X_2$ est une v.a. de Poisson de paramètre $\lambda_1 + \lambda_2$.

b) Soient X_1 et X_2 deux v.a. indépendantes, où X_1 est $N(\mu_1, \sigma_1^2)$ et X_2 est $N(\mu_2, \sigma_2^2)$. Pour montrer que $X_1 + X_2$ est $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, on peut aussi utiliser le théorème 6.29 : cf. exercice 6.25b) (comparer avec l'exercice 5.22b)).

6.5 Exercices

6.1 Une roulette (à Las Vegas) comporte 38 cases, dont 18 cases noires, 18 cases rouges, une case 0 et une case 00. Si vous misez 1 dollar sur noir, vous recevez 2 dollars si la boule s'arrête dans une case noire et 0 sinon. Soit X votre gain net. Calculer $E[X]$.

6.2 A un jeu de la roue de la fortune, il y a 52 résultats possibles : un « 0 », un « 00 », deux « 20 », quatre « 10 », sept « 5 », quinze « 2 » et vingt-deux « 1 ». Si vous misez 1 franc sur un nombre, vous recevez ce montant (et le franc misé) si la roue s'arrête sur ce nombre. Si vous misez 1 franc sur « 0 » ou « 00 », vous recevez 40 francs (et le franc misé) si la roue s'arrête sur ce symbole.

a) Si vous misez simultanément 1 franc sur chacun des sept nombres ou symboles possibles, soit 7 francs au total, quel est votre espérance de gain ?

b) Si vous misez 1 franc sur un seul nombre ou symbole, à quelle mise correspondent la plus grande et la plus petite espérance de gain ?

6.3 a) Soit X une v.a. de Bernoulli $B(1, p)$. Calculer $E[X^2]$ et $\mathrm{Var}(X)$.

b) Soit Y une v.a. de loi Gamma de paramètres λ et s. Calculer $E[Y]$, $E[Y^2]$ et $\mathrm{Var}(Y)$.

6.4 a) Soit X une v.a. $B(n, p)$. Calculer $E[X]$.

b) Soit Y une v.a. uniforme sur $[a, b]$. Calculer $E[Y^2]$ et $\mathrm{Var}(Y)$.

6.5* a) Soit X une v.a. de Poisson de paramètre $\lambda > 0$. Calculer $E[X]$.

b) Soit Y une v.a. $N(\mu, \sigma^2)$. Calculer $E[Y]$, $E[Y^2]$ et $\mathrm{Var}[Y]$.
Indication. Traiter d'abord le cas où Y est $N(0,1)$.

6.6* a) Soit X une v.a. géométrique de paramètre p. Calculer $E[X^2]$ et $\mathrm{Var}(X)$.

b) Soit Y une v.a. exponentielle de paramètre $\lambda > 0$. Calculer $E[Y^2]$ et $\mathrm{Var}(Y)$.

6.7 Montrer que $F_X(x) = E\big[1_{\{X \leqslant x\}}\big] = E\big[1_{]-\infty,x]}(X)\big]$.

6.8 Soit X une v.a. telle que $P\{X \in \mathbb{N}\} = 1$. Montrer que

$$E[X] = \sum_{n=1}^{\infty} P\{X \geqslant n\}\,.$$

6.9 Montrer que si X_1 et X_2 sont indépendantes, alors $E[X_1X_2] = E[X_1]E[X_2]$ (proposition 6.13).

6.10* Soient X une v.a. discrète ou continue, (X_1, X_2) un vecteur aléatoire discret ou continu.

a) Montrer que si $P\{X \geqslant 0\} = 1$, alors $E[X] \geqslant 0$.

b) Montrer que si $P\{X_1 \leqslant X_2\} = 1$, alors $E[X_1] \leqslant E[X_2]$.

c) Montrer que $\big|E[X]\big| \leqslant E\big[|X|\big]$.

6.11 Vérifier les propriétés suivantes.

a) $\mathrm{Var}(X) = E[X^2] - \big(E[X]\big)^2 = \mathrm{Cov}(X, X)$.

b) $\mathrm{Cov}(X_1, X_2) = E[X_1X_2] - E[X_1]E[X_2]$.

c) $(X_1, X_2) \to \mathrm{Cov}(X_1, X_2)$ est une fonction bilinéaire, symétrique, non négative.

d) $-1 \leqslant \mathrm{Corr}(X_1, X_2) \leqslant 1$.

6.12 a) Soit X une v.a. binomiale avec $n = 4$ et $p = \frac{1}{2}$. Utiliser l'inégalité de Tchebychev pour estimer $P\big\{|X-2| \geqslant 2\big\}$ et comparer avec la valeur exacte.

b) Soit Z une v.a. $N(0,1)$. Même question qu'en a) avec $P\big\{|Z| \geqslant 3\big\}$.

6.13 On tire sans remise k boules d'une urne qui contient m boules rouges et n boules noires. On suppose que les boules rouges sont numérotées de 1 à m.

a) Soit G_i l'événement « la $i^{\text{ième}}$ boule rouge est tirée ». Calculer $P(G_i)$.

b) Posons $X = 1_{G_1} + \cdots + 1_{G_m}$. Expliquer pourquoi X est une v.a. hypergéométrique.

c) Calculer $E[X]$.

6.14* 50 couples mariés dînent ensemble. 20 personnes sont désignées au hasard et reçoivent un cadeau.

a) Quelle est la probabilité que les deux membres d'un couple donné ne reçoivent pas de cadeau ?
b) Quelle est l'espérance du nombre N de couples dont aucun des membres ne reçoit un cadeau ?
c) Quelle est la probabilité que chacun des deux membres de deux couples donnés ne reçoivent pas de cadeau ?
d) Quelle est la variance de N ?

6.15* Soient $F_1, \ldots, F_n$ des événements. En observant que $1_{F_1 \cup \cdots \cup F_n} = 1 - 1_{F_1^c} \cdots 1_{F_n^c}$ et en utilisant l'espérance des v.a. indicatrices, retrouver la formule d'inclusion-exclusion (cf. exercice 2.14).

6.16 Soit X une v.a. telle que $E[X^2] < \infty$. Trouver $a \in \mathbb{R}$ tel que $E\big[(X-a)^2\big]$ soit minimal.

6.17 (Inégalité de Jensen) Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction convexe. Montrer que $g\big(E[X]\big) \leqslant E\big[g(X)\big]$.

6.18 Soit X une variable aléatoire non négative. Montrer que

$$E[X] \leqslant E\big[X^2\big]^{1/2} \leqslant E\big[X^3\big]^{1/3} \leqslant \cdots$$

6.19 (Inégalité de Markov) Soit X une v.a. telle que $P\{X \geqslant 0\} = 1$ et $a > 0$.

a) Montrer que $P\{X \geqslant a\} \leqslant \frac{E[X]}{a}$.
b) Montrer que pour tout $p \geqslant 1$, $P\{X \geqslant a\} \leqslant \frac{E[X^p]}{a^p}$.

6.20 a) Soient X une v.a. continue (avec fonction de densité f) et $p > 0$ tels que $P\{X \geqslant 0\} = 1$ et $E[X^p] < \infty$. Montrer que

$$E[X^p] = \int_0^{+\infty} p x^{p-1} P\{X > x\}\, dx\,. \tag{6.11}$$

b) Montrer que l'égalité (6.11) est aussi vraie dans le cas où $P\{X \in \mathbb{N}\} = 1$ et que les deux autres hypothèses sur X sont satisfaites.

6.21* Calculer la fonction génératrice des moments de la v.a. X, puis en déduire $E[X]$ et $\mathrm{Var}(X)$ lorsque X est une v.a. :

a) de loi de Poisson de paramètre $\lambda > 0$,
b) de loi exponentielle de paramètre $\lambda > 0$.

6.22* Calculer la fonction génératrice des moments de la v.a. X, puis en déduire $E[X]$ et $\mathrm{Var}(X)$ lorsque X est une v.a. :

a) de loi binomiale de paramètres n et p,
b) de loi uniforme sur $[a, b]$.

6.23* Calculer la fonction génératrice des moments de la v.a. X, puis en déduire $E[X]$ et $\mathrm{Var}(X)$ lorsque X est une v.a. :

a) de loi géométrique de paramètre p,

b) de loi normale de paramètres μ et σ^2.

6.24* Calculer la fonction génératrice des moments de la v.a. X, puis en déduire $E[X]$ et $\mathrm{Var}(X)$ lorsque X est une v.a. :

a) de loi binomiale négative de paramètres r et p,

b) de loi Gamma de paramètres s et λ.

6.25* En utilisant la tabelle des fonctions génératrices des moments (appendice A.1), vérifier que si X_1 et X_2 sont deux v.a. indépendantes ayant les lois indiquées, alors $X_1 + X_2$ a la loi indiquée.

a) X_1 est $B(n_1, p)$, X_2 est $B(n_2, p)$, alors $X_1 + X_2$ est $B(n_1 + n_2, p)$.

b) X_1 est $N(\mu_1, \sigma_1^2)$, X_2 est $N(\mu_2, \sigma_2^2)$, alors X_1+X_2 est $N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

6.26* En utilisant la tabelle des fonctions génératrices des moments (appendice A.1), vérifier que si X_1 et X_2 sont deux v.a. indépendantes ayant les lois indiquées, alors $X_1 + X_2$ a la loi indiquée.

a) X_1 est Poisson(λ_1), X_2 est Poisson(λ_2), alors $X_1 + X_2$ est Poisson$(\lambda_1 + \lambda_2)$.

b) X_1 est Gamma(p_1, λ), X_2 est Gamma(p_2, λ), alors X_1+X_2 est Gamma(p_1+p_2, λ).

6.27 Soit X une v.a. On appelle *fonction génératrice des cumulants* la fonction $\kappa_X(t) = \ln M_X(t)$.

a) Montrer que $\kappa_X'(0) = E[X]$ et $\kappa_X''(0) = \mathrm{Var}(X)$.

b) Montrer que si X_1 et X_2 sont indépendantes, alors $\kappa_{X_1+X_2} = \kappa_{X_1} + \kappa_{X_2}$.

c) Dans le cas où X est une v.a. $N(\mu, \sigma^2)$, calculer $\kappa_X(t)$ et vérifier que $\kappa_X^{(n)}(0) = 0$ pour tout $n \geqslant 3$.

6.28 a) Soit X une v.a. $N(\mu, \sigma^2)$. Calculer $M_X(t)$.

b) Soit X une v.a. telle que $M_X(t) < \infty$ pour $|t| < t_0$, où $t_0 > 0$. Vérifier que

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E[X^k] .$$

c) Déduire la valeur de $E[X^k]$ lorsque X est $N(0, \sigma^2)$.

6.29* Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction bornée de classe C^3 dont les trois premières dérivées sont bornées. Soient X, Y, Z trois v.a. indépendantes telles que $E[Y] = E[Z]$ et $E[Y^2] = E[Z^2]$. Posons $C = \sup_{x\in\mathbb{R}} \left|g^{(3)}(x)\right|$. Montrer que

$$\left|E\big[g(X+Y)\big] - E\big[g(X+Z)\big]\right| \leqslant \frac{C}{3!} \left(E\big[|Y|^3\big] + E\big[|Z|^3\big]\right) .$$

Indication. Utiliser un dévelopement de Taylor d'ordre 2 de g avec une estimation du reste.

6.30* (Loi gaussienne bivariée) Soient $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1 > 0$, $\sigma_2 > 0$ et $\sigma_{1,2} \in \mathbb{R}$ tel que $|\sigma_{1,2}| < \sigma_1\sigma_2$. On pose

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \text{et} \qquad B = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix}.$$

Soit $Z = (Z_1, Z_2)$ un vecteur aléatoire continu dans $\mathbb{R}^2$ dont la fonction de densité est

$$f_Z(z) = \frac{1}{2\pi\sqrt{\det B}} \exp\left(-\frac{1}{2}(z-\mu)^T \cdot B^{-1} \cdot (z-\mu)\right), \qquad z \in \mathbb{R}^2.$$

On dit que Z est un vecteur aléatoire *gaussien bivarié* $N(\mu, B)$.

a) Montrer que $f_Z(z)$ a la même forme que la fonction de densité de l'exercice 5.23 et déterminer la valeur de ρ qui rend les deux densités identiques.

b) En déduire que

$$\mu = \begin{pmatrix} E[Z_1] \\ E[Z_2] \end{pmatrix} \qquad \text{et} \qquad B = \begin{pmatrix} \mathrm{Var}(Z_1) & \mathrm{Cov}(Z_1, Z_2) \\ \mathrm{Cov}(Z_1, Z_2) & \mathrm{Var}(Z_2) \end{pmatrix}.$$

c) En déduire aussi que si $\mathrm{Corr}(Z_1, Z_2) = 0$, alors Z_1 et Z_2 sont indépendantes. (*Attention* : cette propriété n'est valable que pour les composantes d'un vecteur aléatoire *gaussien.*)

6.31 Soit X une v.a. continue de densité f_X et soit $g : \mathbb{R} \to \mathbb{R}_+$ continue. Sans utiliser le théorème 6.10, montrer que

$$E\big[g(X)\big] = \int_0^\infty g(x) f_X(x)\, dx.$$

Indication. Utiliser l'exercice 6.20.

6.32 (Moments négatifs d'une v.a. positive) Soit X une v.a. continue avec fonction de densité f, telle que $P\{X > 0\} = 1$.

a) On suppose que $\int_0^\infty \frac{1}{x} f(x)\, dx < \infty$. Montrer que

$$\int_{-\infty}^0 M_X(t)\, dt = E\left[\frac{1}{X}\right].$$

b) Soit $n \geqslant 1$. On suppose que $\int_0^\infty x^{-n} f(x)\, dx < \infty$. Montrer que

$$\int_{-\infty}^0 dt_1 \int_{-\infty}^{t_1} dt_2 \cdots \int_{-\infty}^{t_{n-1}} dt_n\, M_X(t_n) = E\left[X^{-n}\right].$$

c) Soit X une v.a. de loi Gamma de paramètres p et λ, avec $p > 1$. Calculer $E[1/X]$.

Chapitre 7

Théorèmes limites

Dans ce chapitre, nous allons présenter les principaux théorèmes limites de la théorie des probabilités : les lois faible et forte des grands nombres et le théorème limite central. Ces théorèmes traitent du comportement asymptotique de la somme de v.a. *indépendantes et identiquement distribuées* (*i.i.d.*), ce qui correspond à la situation suivante.

Jeux répétés

Un joueur joue de manière répétée à un jeu de hasard. Soit X_i le bénéfice réalisé lors du $i^{\text{ième}}$ jeu, de sorte que

$$S_n = X_1 + \cdots + X_n$$

est le bénéfice cumulé après n jeux et $\frac{S_n}{n}$ est le bénéfice moyen par jeu, après n jeux. On s'intéresse au comportement asymtotique de S_n et de $\frac{S_n}{n}$.

7.1 Loi faible des grands nombres

Théorème 7.1

Soit $(X_n,\ n \in \mathbb{N}^*)$ *une suite de v.a. i.i.d., définies sur un espace de probabilité* $(\Omega, \mathcal{F}, P)$*, telles que* $\operatorname{Var}(X_1) = \sigma^2 < \infty$ *et* $E[X_1] = \mu$*. Pour* $n \in \mathbb{N}^*$*, posons* $S_n = X_1 + \cdots + X_n$*. Alors pour tout* $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - \mu\right| \geqslant \varepsilon\right\} = 0\,.$$

Démonstration. En utilisant la linéarité de l'espérance et les hypothèses, observons que

$$E\left[\frac{S_n}{n}\right] = \frac{1}{n}\,E[X_1 + \cdots + X_n] = \frac{1}{n}\big(E[X_1] + \cdots + E[X_n]\big) = \frac{1}{n}\,(n\mu) = \mu\,.$$

En utilisant l'indépendance de $X_1, \ldots, X_n$ et la proposition 6.19, on constate que

$$\begin{aligned}\operatorname{Var}\left(\frac{S_n}{n}\right) &= \frac{1}{n^2}\operatorname{Var}(X_1+\cdots+X_n) = \frac{1}{n^2}\bigl(\operatorname{Var}(X_1)+\cdots+\operatorname{Var}(X_n)\bigr)\\ &= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}\,.\end{aligned}$$

A l'aide de l'inégalité de Tchebychev, on constate que pour tout $\varepsilon > 0$,

$$\begin{aligned}0 \leqslant P\left\{\left|\frac{S_n}{n}-\mu\right| \geqslant \varepsilon\right\} &= P\left\{\left|\frac{S_n}{n}-E\left[\frac{S_n}{n}\right]\right| \geqslant \varepsilon\right\}\\ &\leqslant \frac{1}{\varepsilon^2}\operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{\varepsilon^2}\,\frac{\sigma^2}{n}\\ &\underset{n\to\infty}{\longrightarrow} 0\,.\end{aligned}$$

□

Remarque 7.2 Le théorème 7.1 signifie que pour de grandes valeurs de n, il est très probable que $\frac{S_n}{n}$ soit voisin de μ.

Exemple 7.3 Au jeu de la roulette (exercice 6.1), soit Y le gain d'un joueur qui mise 1 dollar sur noir. Puisque

$$P\{Y=+1\} = \frac{18}{38}\,, \qquad P\{Y=-1\} = \frac{20}{38}\,.$$

on voit que

$$E[Y] = 1\times\frac{18}{38} + (-1)\times\frac{20}{38} = -\frac{1}{19} \simeq -0{,}0526\,.$$

Imaginons qu'au cours de la soirée, il y a 1000 mises. Soit X_i le gain du casino lors de la $i^{\text{ième}}$ mise. Puisque le casino gagne ce que les joueurs perdent,

$$E[X_i] = \frac{1}{19} \simeq +0{,}0526\,.$$

Le gain total du casino lors de ces 1000 mises est alors

$$S_{1000} = X_1 + \cdots + X_{1000}\,.$$

Le gain moyen du casino à chaque mise est

$$\frac{X_1+\cdots+X_{1000}}{1000} = \frac{S_{1000}}{1000}\,.$$

Puisque $n = 1000$ est relativement grand, il est très probable, d'après le théorème 7.1, que $\frac{S_{1000}}{1000}$ soit voisin de 0,0526, donc $S_{1000} \simeq 52{,}6$ dollars. La conclusion de cet exemple est que les bénéfices du casino sont systématiques, même s'ils ne paraissent pas spectaculaires !

7.2 Théorème limite central

Ce théorème limite joue, comme l'indique son nom, un rôle central en théorie des probabilités. Il est aussi à la base d'une bonne partie de l'étude de la statistique. Il énonce la convergence de la loi d'une somme normalisée de v.a. i.i.d. vers la loi normale.

Théorème 7.4
Soit $(X_n,\ n \in \mathbb{N}^*)$ *une suite de v.a. i.i.d. définies sur un espace de probabilité* $(\Omega, \mathcal{F}, P)$ *telles que* $0 < \mathrm{Var}(X_1) < +\infty$*. Posons* $\mu = E[X_1]$ *et* $\sigma^2 = \mathrm{Var}(X_1)$*. Alors pour tout* $x \in \mathbb{R}$*,*

$$\lim_{n\to\infty} P\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} \leqslant x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du = \Phi(x)\,,$$

où Φ *est la fonction de répartition d'une v.a. de loi* $N(0,1)$*.*

Nous allons démontrer ce théorème sous une hypothèse un peu plus forte que celle indiquée dans l'énoncé.

Hypothèse A. $E\big[|X_1|^3\big] < +\infty$.

Nous utiliserons la notation suivante : $C_{\mathrm{b}}^3(\mathbb{R})$ désignera l'ensemble des fonctions $f : \mathbb{R} \to \mathbb{R}$ bornées, de classe C^3, dont les trois premières dérivées sont bornées.

Lemme 7.5 *Soit* $(\tilde{Y}_n,\ n \in \mathbb{N})$ *une suite de v.a. i.i.d. telles que* $E\big[|\tilde{Y}_1|^3\big] < \infty$*,* $E[\tilde{Y}_1^2] = 1$ *et* $E[\tilde{Y}_1] = 0$*. Soit* $\tilde{Z}$ *une v.a. de loi* $N(0,1)$*. Alors pour toute fonction* $g \in C_{\mathrm{b}}^3(\mathbb{R})$*,*

$$\lim_{n\to\infty} E\left[g\left(\frac{\tilde{Y}_1 + \cdots + \tilde{Y}_n}{\sqrt{n}}\right)\right] = E\big[g(\tilde{Z})\big]\,.$$

Démonstration. Fixons $g \in C_{\mathrm{b}}^3(\mathbb{R})$ et $n \in \mathbb{N}$. Soient $\tilde{Z}_1, \dots, \tilde{Z}_n$ des v.a. i.i.d. de loi $N(0,1)$. Posons $Y_i = n^{-1/2}\tilde{Y}_i$ et $Z_i = n^{-1/2}\tilde{Z}_i$. Puisque $Z_1 + \cdots + Z_n$ a la même loi que $\tilde{Z}$, il suffit de montrer que

$$\big|E\big[g(Y_1 + \cdots + Y_n)\big] - E\big[g(Z_1 + \cdots + Z_n)\big]\big| \leqslant \frac{C}{3!}\,\frac{E\big[|\tilde{Y}_1|^3\big] + E\big[|\tilde{Z}_1|^3\big]}{\sqrt{n}}\,, \tag{7.1}$$

où $C = \sup_{x\in\mathbb{R}}\big|g^{(3)}(x)\big|$. Posons

$$\begin{aligned} U_i &= Y_1 + \cdots + Y_{i-1} + Y_i + Z_{i+1} + \cdots + Z_n\,,\\ V_i &= Y_1 + \cdots + Y_{i-1} + \phantom{Y_i +{}} Z_{i+1} + \cdots + Z_n \end{aligned}$$

(noter l'absence du terme Y_i dans la définition de V_i), de sorte que

$$U_i = V_i + Y_i\,, \qquad U_{i-1} = V_i + Z_i$$

et la différence des espérances dans le membre de gauche de (7.1) est égale à

$$E\big[g(U_n)\big] - E\big[g(U_0)\big] = \sum_{i=1}^{n} \Big(E\big[g(U_i)\big] - E\big[g(U_{i-1})\big]\Big)$$
$$= \sum_{i=1}^{n} \Big(E\big[g(V_i + Y_i)\big] - E\big[g(V_i + Z_i)\big]\Big)\,.$$

D'après l'exercice 6.29, la valeur absolue de cette expression est majorée par

$$\sum_{i=1}^{n} \frac{C}{3!}\left(E\big[|Y_i|^3\big] + E\big[|Z_i|^3\big]\right) = n\frac{C}{3!}\,n^{-3/2}\left(E\big[|\tilde{Y}_1|^3\big] + E\big[|\tilde{Z}_1|^3\big]\right), \tag{7.2}$$

ce qui montre (7.1) et achève la démonstration du lemme 7.5. □

Proposition 7.6 *Soient Z une v.a. continue et $(Z_n,\, n \in \mathbb{N})$ une suite de v.a. On suppose que pour toute fonction $g \in C_{\mathrm{b}}^3(\mathbb{R})$, $\lim_{n\to\infty} E[g(Z_n)] = E[g(Z)]$. Alors pour tout $x \in \mathbb{R}$, $\lim_{n\to\infty} F_{Z_n}(x) = F_Z(x)$.*

DÉMONSTRATION. Soient $x \in \mathbb{R}$ et $k \in \mathbb{N}^*$ fixés. Construisons (fig. 7.1) deux fonctions $\tilde{g}_k$ et g_k dans $C_{\mathrm{b}}^3(\mathbb{R})$ telles que pour tout $z \in \mathbb{R}$,

$$1_{]-\infty,\, x-\frac{1}{k}]}(z) \leqslant \tilde{g}_k(z) \leqslant 1_{]-\infty,\, x]}(z) \leqslant g_k(z) \leqslant 1_{]-\infty,\, x+\frac{1}{k}]}(z)\,. \tag{7.3}$$

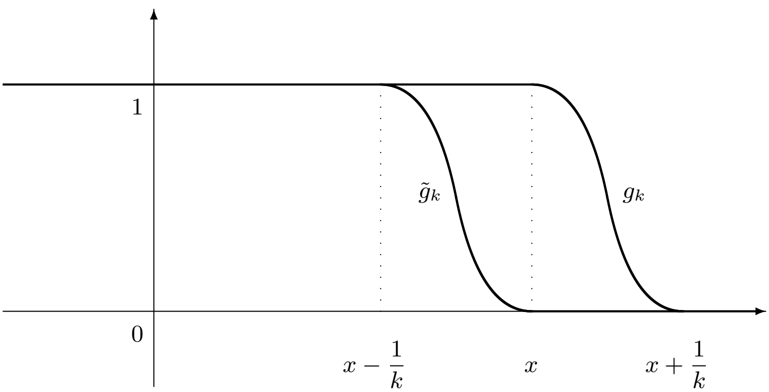


Figure 7.1 Esquisse des fonctions $\tilde{g}_k$ et g_k.

Observons que pour tout $n \in \mathbb{N}$,

$$F_{Z_n}(x) = P\{Z_n \leqslant x\} = E\big[1_{]-\infty,x]}(Z_n)\big] \leqslant E\big[g_k(Z_n)\big]\,,$$

et donc, d'après l'hypothèse,

$$\limsup_{n\to\infty} F_{Z_n}(x) \leqslant \lim_{n\to\infty} E\big[g_k(Z_n)\big] = E\big[g_k(Z)\big]$$
$$\leqslant E\big[1_{]-\infty,\, x+1/k]}(Z)\big] = F_Z\left(x + \frac{1}{k}\right). \tag{7.4}$$

De la même manière,

$$F_{Z_n}(x) = E\big[1_{]-\infty,x]}(Z_n)\big] \geqslant E\big[\tilde{g}_k(Z_n)\big]\,,$$

et donc

$$\begin{aligned}\liminf_{n\to\infty} F_{Z_n}(x) &\geqslant \lim_{n\to\infty} E\big[\tilde{g}_k(Z_n)\big] = E\big[\tilde{g}_k(Z)\big]\\ &\geqslant E\big[1_{]-\infty,\, x-1/k]}(Z)\big] = F_Z\left(x-\frac{1}{k}\right).\end{aligned} \tag{7.5}$$

Nous déduisons de (7.4) et (7.5) que pour tout $k \in \mathbb{N}^*$,

$$F_Z\left(x-\frac{1}{k}\right) \leqslant \liminf_{n\to\infty} F_{Z_n}(x) \leqslant \limsup_{n\to\infty} F_{Z_n}(x) \leqslant F_Z\left(x+\frac{1}{k}\right).$$

En faisant tendre k vers l'infini, nous concluons, F_Z étant continue, que $\liminf_{n\to\infty} F_{Z_n}(x)$ et $\limsup_{n\to\infty} F_{Z_n}(x)$ sont égales et que $\lim_{n\to\infty} F_{Z_n}(x) = F_Z(x)$. La proposition 7.6 est démontrée. □

Démonstration. Nous *démontrons le théorème 7.4 sous l'hypothèse A*. Posons $\tilde{Y}_n = \frac{X_n-\mu}{\sigma}$. Alors

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\tilde{Y}_1 + \cdots + \tilde{Y}_n}{\sqrt{n}}$$

et vu l'hypothèse A, $(\tilde{Y}_n,\, n \in \mathbb{N})$ vérifie les hypothèses du lemme 7.5. Soit Z une v.a. $N(0,1)$. D'après la conclusion du lemme 7.5, $Z_n = n^{-1/2}(\tilde{Y}_1+\cdots+\tilde{Y}_n)$ et Z vérifient les hypothèses de la proposition 7.6, donc aussi sa conclusion. Le théorème 7.4 est démontré. □

Remarque 7.7 Le lecteur souhaitant plus de détails concernant le théorème limite central peut se référer à [3]. On y trouve notamment la démonstration présentée ci-dessus, une démonstration complète sans l'hypothèse A, de même qu'un résultat concernant l'estimation de l'erreur commise lors d'une approximation par le théorème limite central.

7.3 Utilisation du théorème limite central

On cherche souvent à determiner une valeur numérique approximative pour des expressions du type

$$P\{S_n \in A\}\,.$$

Pour ce faire, on exprime d'abord cette probabilité comme somme ou différence de probabilités de la forme $P\{S_n \in I\}$, où I est un intervalle. Dans le cas où, par exemple, $I = \,]a,b]$ est un intervalle ouvert à gauche et fermé à droite, on a

$$P\{a < S_n \leqslant b\} = P\left\{\frac{a-n\mu}{\sigma\sqrt{n}} < \frac{S_n - n\mu}{\sigma\sqrt{n}} \leqslant \frac{b-n\mu}{\sigma\sqrt{n}}\right\}.$$

D'après le théorème limite central, le membre de droite est approximativement égal à

$$P\left\{\frac{a-n\mu}{\sigma\sqrt{n}} < Z \leqslant \frac{b-n\mu}{\sigma\sqrt{n}}\right\}$$

où Z est une v.a. $N(0,1)$. Cette probabilité s'écrit donc

$$\Phi\left(\frac{b-n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a-n\mu}{\sigma\sqrt{n}}\right).$$

La tabelle A.2 de l'appendice fournit ensuite la valeur numérique cherchée.

L'opération qui consiste à soustraite $n\mu$ et à diviser par $\sigma\sqrt{n}$ est appelée *normalisation* et on parle de *somme normalisée* pour se référer à l'expression

$$\frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Exemple 7.8

a) A la roulette, vous jouez 81 fois, pariant chaque fois 1 dollar sur noir. Quelle est la probabilité qu'à la fin, votre bénéfice soit positif ?
Pour répondre à cette question, soit X_i le bénéfice à la mise i, de sorte que (cf. exemple 7.3) :

$$E[X_i] = -\frac{1}{19} = \mu, \qquad P\{X_i = +1\} = \frac{18}{38}, \qquad P\{X_i = -1\} = \frac{20}{38}$$

et

$$\text{Var}(X_i) = E[X_i^2] - E[X_i]^2 = 1 - \left(-\frac{1}{19}\right)^2 = 0{,}9972 = \sigma^2.$$

Le bénéfice après 81 jeux est $S_{81} = X_1 + \cdots + X_{81}$. On cherche $P\{S_{81} \geqslant 0\}$. Cette probabilité est égale à

$$\begin{aligned} &P\left\{S_{81} - 81 \times \left(-\frac{1}{19}\right) \geqslant 0 - 81 \times \left(-\frac{1}{19}\right)\right\} \\ &\quad = P\left\{\frac{S_{81} - 81 \times \left(-\frac{1}{19}\right)}{\sqrt{0{,}9972}\sqrt{81}} \geqslant \frac{\frac{81}{19}}{\sqrt{0{,}9972}\sqrt{81}}\right\}. \end{aligned}$$

D'après le théorème limite central, cette probabilité est approximativement égale à

$$\begin{aligned} P\left\{Z \geqslant \frac{9}{19\sqrt{0{,}9972}}\right\} &= P\{Z \geqslant 0{,}47\} = 1 - P\{Z < 0{,}47\} \\ &= 1 - 0{,}6808 = 0{,}3192. \end{aligned}$$

b) On lance une pièce de monnaie 400 fois. Quelle est la probabilité d'avoir au moins 220 piles ? (On s'attend à 200 piles.)
Pour $i = 1, \ldots, 400$, soit $X_i = 1$ si le $i^{\text{ième}}$ jet tombe sur pile, $X_i = 0$ sinon. Alors X_i suit une loi de Bernoulli de paramètre $\frac{1}{2}$ et nous avons

$$E[X_i] = \frac{1}{2} \qquad \text{et} \qquad \operatorname{Var}(X_i) = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4}\,.$$

Posons $S_n = X_1 + \cdots + X_n$, de sorte que S_{400} représente le nombre de piles après 400 jets. Nous cherchons

$$\begin{aligned} P\{S_{400} \geqslant 220\} &= P\left\{\frac{S_{400} - 400 \times \frac{1}{2}}{\sqrt{\frac{1}{4}}\sqrt{400}} \geqslant \frac{220 - 400 \times \frac{1}{2}}{\sqrt{\frac{1}{4}}\sqrt{400}}\right\} \\ &= P\left\{\frac{S_{400} - 400 \times \frac{1}{2}}{\sqrt{\frac{1}{4}}\sqrt{400}} \geqslant 2\right\}. \end{aligned}$$

D'après le théorème limite central, cette probabilité est approximativement égale à

$$P\{Z \geqslant 2\} = 1 - P\{Z \leqslant 2\} = 1 - 0{,}9772 = 0{,}0228\,.$$

Remarque 7.9 Si la question était « quelle est la probabilité d'avoir au plus 219 piles ? », on aurait calculé

$$\begin{aligned} P\{S_{400} \leqslant 219\} &= P\left\{\frac{S_{400} - 200}{\sqrt{\frac{1}{4}}\sqrt{400}} \leqslant \frac{219 - 200}{\sqrt{\frac{1}{4}}\sqrt{400}}\right\} \simeq P\left\{Z \leqslant \frac{219 - 200}{10}\right\} \\ &= P\{Z \leqslant 1{,}9\} = 0{,}9713\,. \end{aligned}$$

La somme de ces deux probabilités devrait être égale à 1. Cependant, d'après nos approximations,

$$P\{S_{400} \leqslant 219\} + P\{S_{400} \geqslant 220\} \simeq 0{,}0228 + 0{,}9713 = 0{,}9941 < 1\,.$$

Pour pallier ce problème, lié au fait qu'on approxime une loi discrète par une loi continue, on utilise la correction de continuité, que nous introduisons ci-dessous.

Correction de continuité

En regardant de plus près l'exemple 7.8b) et la remarque ci-dessus, on constate que dans notre approximation de la somme des probabilités $P\{S_{400} \leqslant 219\} + P\{S_{400} \geqslant 220\}$, il manque la valeur de $P\{219 < 200 + 10Z < 220\}$. Ceci serait en principe une approximation de $P\{219 < S_{400} < 220\}$, qui est en fait égal à 0. On obtient une meilleure approximation de ces probabilités en répartissant la valeur de $P\{219 < 200 + 10Z < 220\}$ entre les deux termes $P\{S_{400} \leqslant 219\}$ et $P\{S_{400} \geqslant 220\}$, ce qui revient formellement à constater que puisque S_{400} prend des valeurs entières,

$$P\{S_{400} \leqslant 219\} = P\{S_{400} \leqslant 219{,}5\}$$

et

$$P\{S_{400} \geqslant 220\} = P\{S_{400} \geqslant 219{,}5\}.$$

On procède ensuite à la normalisation comme ci-dessus. On trouve que

$$\begin{aligned} P\{S_{400} \leqslant 219\} = P\{S_{400} \leqslant 219{,}5\} &= P\left\{\frac{S_{400} - 200}{10} \leqslant \frac{19{,}5}{10}\right\} \\ &\simeq P\left\{Z \leqslant \frac{19{,}5}{10}\right\} = 0{,}9744 \end{aligned}$$

et

$$\begin{aligned} P\{S_{400} \geqslant 220\} = P\{S_{400} \geqslant 219{,}5\} &= P\left\{\frac{S_{400} - 200}{10} \geqslant \frac{19{,}5}{10}\right\} \\ &\simeq P\left\{Z \geqslant \frac{19{,}5}{10}\right\} = 1 - 0{,}9744 = 0{,}0256\,. \end{aligned}$$

De manière générale, en vue d'approximer $P\{S_n \leqslant k\}$, on remplacera k par $k + \frac{1}{2}$ *avant* de procéder à la normalisation ; pour approximer $P\{S_n \geqslant k\}$ on remplacera k par $k - \frac{1}{2}$. Avec des inégalités strictes, on fait l'inverse : pour approximer $P\{S_n < k\}$ on remplacera k par $k - \frac{1}{2}$; pour approximer $P\{S_n > k\}$, on remplacera k par $k + \frac{1}{2}$.

7.4 Loi forte des grands nombres*

Théorème 7.10

Soit $(X_n, n \in \mathbb{N})$ *une suite de v.a. i.i.d., définies sur un espace de probabilité* $(\Omega, \mathcal{F}, P)$ *et telles que* $E\big[|X_1|\big] < +\infty$. *Pour* $n \in \mathbb{N}$, *posons* $S_n = X_1 + \cdots + X_n$. *Alors*

$$P\left\{\omega \in \Omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = E[X_1]\right\} = 1\,.$$

Nous allons démontrer ce théorème sous une hypothèse un peu plus forte que celle indiquée dans l'énoncé.

Hypothèse B. $E[X_1^4] < +\infty$.

Démonstration du théorème 7.10 sous l'hypothèse B

Posons $\mu = E[X_1]$, $Y_n = X_n - \mu$ et $\tilde{S}_n = Y_1 + \cdots + Y_n$. Il suffit de montrer que

$$P\left\{\omega \in \Omega : \lim_{n\to\infty} \frac{\tilde{S}_n(\omega)}{n} = 0\right\} = 1 .$$

D'après l'exercice 9.29, il suffit de montrer que pour tout $k \in \mathbb{N}^*$,

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{\tilde{S}_n}{n}\right| \geqslant \frac{1}{k}\right\} < +\infty . \tag{7.6}$$

D'après l'inégalité de Markov (exercice 6.19),

$$P\left\{\left|\frac{\tilde{S}_n}{n}\right| \geqslant \frac{1}{k}\right\} \leqslant \frac{E\left[\left(\frac{\tilde{S}_n}{n}\right)^4\right]}{\left(\frac{1}{k}\right)^4} = \frac{k^4}{n^4}\, E[\tilde{S}_n^4] . \tag{7.7}$$

Or

$$\begin{aligned} E[\tilde{S}_n^4] &= E\left[\left(\sum_{k=1}^{n} Y_k\right)^4\right] = E\left[\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\sum_{k_3=1}^{n}\sum_{k_4=1}^{n} Y_{k_1}Y_{k_2}Y_{k_3}Y_{k_4}\right] \\ &= \sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\sum_{k_3=1}^{n}\sum_{k_4=1}^{n} E[Y_{k_1}Y_{k_2}Y_{k_3}Y_{k_4}] . \end{aligned}$$

Vu que les Y_k sont indépendantes et $E[Y_k] = 0$, les seuls termes non nuls de cette quadruple somme sont ceux où $\{k_1, k_2, k_3, k_4\}$ est un singleton ou une paire. Puisqu'il y a n manières de choisir $(k_1, \ldots, k_4) \in \{1, \ldots, n\}^4$ pour obtenir un singleton et $C_{4,2}\, C_{n,2} = 3n(n-1)$ manières d'obtenir une paire,

$$\begin{aligned} E[\tilde{S}_n^4] &= n\, E[Y_1^4] + 3n(n-1)E[Y_1^2]E[Y_2^2] \\ &= n\, E[Y_1^4] + 3n(n-1)\big(E[Y_1^2]\big)^2 . \end{aligned}$$

En posant $c = 4\max\big(E[Y_1^4], \big(E[Y_1^2]\big)^2\big)$, on constate que pour tout $n \geqslant 1$,

$$E[\tilde{S}_n^4] \leqslant cn^2 . \tag{7.8}$$

Ainsi, d'après (7.7) et (7.8),

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{\tilde{S}_n}{n}\right| \geqslant \frac{1}{k}\right\} \leqslant ck^4 \sum_{n=1}^{\infty} \frac{1}{n^2} < +\infty$$

ce qui établit (7.6) et achève la démonstration. □

7.5 Exercices

7.1 On arrondit 36 nombres réels à l'entier le plus proche, puis on effectue leur somme. En admettant que les erreurs d'arrondi sont indépendantes et uniformément distribuées sur $]-0{,}5; 0{,}5]$, estimer, à l'aide du théorème central limite, la probabilité que la somme obtenue diffère de la somme effective des 36 nombres par plus de 5. (Observer que la différence maximale est 18.)

7.2 (Correction de continuité)

a) Soit X une v.a. binomiale de paramètres $n = 100$ et $p = \frac{1}{2}$. A l'aide de l'égalité $P\{X = 50\} = P\{49{,}5 < X < 50{,}5\}$ et du théorème central limite, estimer $P\{X = 50\}$.

b) Un dé est jeté 200 fois de suite. Estimer la probabilité que la somme des nombres obtenus se situe dans l'intervalle $[650, 750]$. Faire le calcul avec et sans la correction de continuité.

7.3* Un club de football organise une loterie. Pour une mise de 1 franc, cette loterie offre des gains de 1 franc et 2 francs avec probabilités respectives de $\frac{2}{5}$ et $\frac{1}{5}$. Pour 100 billets de loterie vendus, quelle est la probabilité que le club fasse un bénéfice supérieur à 10 francs ?

7.4* Un étudiant se présente à un examen. Il y a 100 questions auxquelles il faut répondre par *Vrai* ou *Faux*. L'examen est réussi si l'étudiant répond juste à 80 questions ou plus.

a) En admettant que la probabilité de répondre juste à une question est $p = \frac{3}{4}$ (indépendamment pour chaque question), estimer la probabilité que l'étudiant réussisse l'examen.

b) Une étudiante connaît les réponses à 50 des questions et joue à pile ou face pour répondre aux autres questions. Estimer la probabilité que l'étudiante réussisse l'examen. (Utiliser la correction de continuité.)

7.5 Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. i.i.d de loi uniforme sur $[0, 1]$. Pour $n \geqslant 1$, on pose $S_n = X_1 + \cdots + X_n$. Trouver le plus petit entier $n \in \mathbb{N}$ tel que $P\{\frac{S_n}{n} < 0{,}51\}$ soit plus grand ou égal à 0,84.

7.6 Une compagnie d'aviation a observé que 12 % des personnes qui réservent leur billet ne se présentent pas au départ. Elle admet que les désistements des passagers sont des événements indépendants. Pour un avion de 169 places, combien de billets est-ce que la compagnie peut vendre de sorte que, avec la probabilité 0,99, tous les passagers qui se présentent aient un siège dans l'avion ?

7.7* Lors d'une balade en montagne, Christian se trouve obligé de traverser une grotte. Il doit emporter des allumettes de façon à pouvoir s'éclairer lors de la traversée. Une allumette permet de s'éclairer en moyenne durant 30 secondes avec un écart type de 30 secondes. Sachant que la traversée de la grotte prend 18 minutes, combien Christian doit il prévoir d'allumettes pour être sûr à 95 % de pouvoir s'éclairer durant toute la traversée ?

7.8* Fabrice arrive devant un stade pour assister à un match de hockey sur glace 10 minutes avant le début de celui-ci.

a) Il y a 100 personnes dans la file devant lui. Sachant que le temps de service d'un client suit une loi exponentielle, combien de clients à la minute en moyenne le vendeur doit-il servir pour que Fabrice ait 3 chances sur 4 d'assister au début du match ?

b) Il ne reste malheureusement que 200 billets à vendre. Sachant qu'un client achète en moyenne 1,85 billet avec un écart type de 1,5 billet et que le vendeur sert en moyenne 9 clients à la minute, quelle est la probabilité que Fabrice assiste au début du match ?

7.9 Sous les hypothèses du théorème limite central, posons

$$Z_n = \frac{1}{\sigma\sqrt{n}}\left(S_n - n\mu\right).$$

Supposons qu'il existe $t_0 > 0$ tel que $E[e^{tX_1}] < \infty$ pour tout $t \in \,]-t_0, t_0[$. Montrer que pour tout $t \geqslant 0$, $\lim_{n\to\infty} M_{Z_n}(t) = M_Z(t)$, où Z est une v.a. $N(0,1)$.

Chapitre 8

Espérance et variance conditionnelles

L'espérance conditionnelle d'une v.a. Y, *sachant* une autre v.a. X est une troisième v.a., fonction de X, qui représente la meilleure estimation de Y lorsqu'on connaît la valeur de X (mais pas celle de Y !). Ce problème d'estimation est un cas particulier du problème de la meilleure approximation que l'on rencontre en algèbre linéaire et que nous rappelons ci-dessous.

8.1 Rappels d'algèbre linéaire

Proposition 8.1 *Soit V un espace vectoriel muni d'un produit scalaire $\langle\cdot\,,\cdot\rangle$ et de la norme $\|\cdot\|$ associée, W un sous-espace vectoriel de V et $v \in V$. S'il existe $w_0 \in W$ tel que pour tout $w \in W$,*

$$\langle v - w_0, w\rangle = 0\,, \tag{8.1}$$

alors w_0 est la projection orthogonale de v sur W et vérifie

$$\|v - w_0\| \leqslant \|v - w\|\,, \qquad \text{pour tout } w \in W\,.$$

Démonstration. Puisque $v = w_0 + (v - w_0)$, où $w_0 \in W$ et $v - w_0$ est orthogonal à W d'après (8.1), w_0 est bien la projection orthogonale de v sur W. D'autre part, pour tout $w \in W$,

$$\|v - w\|^2 = \|v - w_0 + w_0 - w\|^2.$$

Puisque $v - w_0$ est orthogonal à $w_0 - w$ d'après (8.1), le théorème de Pythagore affirme que le membre de droite est égal à

$$\|v - w_0\|^2 + \|w_0 - w\|^2 \geqslant \|v - w_0\|^2.$$

La proposition est démontrée. □

Remarque 8.2

a) Le vecteur w_0 est appelé la *meilleure approximation de v par un élément de W.*

b) D'après (8.1), le vecteur w_0 est caractérisé par la relation

$$\langle v, w\rangle = \langle w_0, w\rangle, \qquad \text{pour tout } w \in W\,. \tag{8.2}$$

8.2 Un espace vectoriel de variables aléatoires

Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Soit $L^2(\Omega)$ l'ensemble des v.a. $X : \Omega \to \mathbb{R}$ telles que $E[X^2] < \infty$.

Proposition 8.3 *$L^2(\Omega)$ est un espace vectoriel.*

Démonstration. Puisque $L^2(\Omega)$ est un sous-ensemble non vide de l'espace vectoriel de toutes les fonctions de $\Omega \to \mathbb{R}$, il suffit de vérifier la stabilité de $L^2(\Omega)$ par combinaison linéaire : si $X_1 \in L^2(\Omega)$, $X_2 \in L^2(\Omega)$ et $a_1, a_2 \in \mathbb{R}$, alors d'après l'inégalité de Cauchy-Schwarz usuelle,

$$\begin{aligned} E\big[(a_1X_1 + a_2X_2)^2\big] &\leqslant E\big[(a_1^2 + a_2^2)(X_1^2 + X_2^2)\big] \\ &= (a_1^2 + a_2^2)\big(E[X_1^2] + E[X_2^2]\big) < \infty\,. \end{aligned}$$

Observons que le vecteur nul dans $L^2(\Omega)$ est la v.a. constante égale à zéro. □

Semi-produit scalaire et semi-norme dans $L^2(\Omega)$

Considérons l'opération

$$(X, Y) \mapsto \langle X, Y\rangle \overset{\text{déf}}{=} E[XY]\,.$$

Remarquons que cette application possède les trois propriétés d'un semi-produit scalaire. En effet, elle est bilinéaire, puisque

$$\begin{aligned} \langle a_1X_1 + a_2X_2, Y\rangle &= E\big[(a_1X_1 + a_2X_2)Y\big] \\ &= a_1E[X_1Y] + a_2E[X_2Y] = a_1\langle X_1, Y\rangle + a_2\langle X_2, Y\rangle\,. \end{aligned}$$

Elle est symétrique, puisque $\langle X, Y\rangle = E[XY] = E[YX] = \langle Y, X\rangle$. Elle est positive, puisque $\langle X, X\rangle = E[X^2] \geqslant 0$. Elle n'est cependant pas définie-positive, car une v.a. qui n'est pas identiquement nulle peut avoir une norme nulle. C'est pour cela qu'on parle d'un *semi-produit scalaire* et non pas d'un produit scalaire.

Remarque 8.4 En fait, il manque très peu à l'opération $(X, Y) \mapsto \langle X, Y \rangle$ pour être définie-positive, puisque

$$\langle X, X \rangle = 0 \implies E[X^2] = 0 \implies \mathrm{Var}(X) + \big(E[X]\big)^2 = 0\,.$$

Puisque $\mathrm{Var}(X) \geqslant 0$, ceci implique que $\mathrm{Var}(X) = 0$ et $E[X] = 0$. D'après le corollaire 6.18, nous déduisons que $P\big\{X = E[X]\big\} = 1$ et $E[X] = 0$. Ainsi, $P\{X = 0\} = 1$.

La *semi-norme* associée au semi-produit scalaire que nous venons de définir est $\|X\|^2 = \langle X, X \rangle = E[X^2]$.

8.3 Estimation d'une variable aléatoire

D'après l'exercice 6.16, le nombre $a \in \mathbb{R}$ qui minimise l'expression $\|Y - a\|^2 = E\big[(Y-a)^2\big]$ est $a = E[Y]$. Ainsi, la meilleure estimation de Y par une v.a. constante est $E[Y]$. Etant donné une v.a. $Y \in L^2(\Omega)$ et une v.a. X, nous allons chercher la v.a. de la forme $\psi(X)$ qui approxime au mieux la v.a. Y, dans le sens où cette fonction $\psi : \mathbb{R} \to \mathbb{R}$ minimise l'expression

$$\big\|Y - \psi(X)\big\|^2 = E\left[\big(Y - \psi(X)\big)^2\right].$$

Espérance conditionnelle d'une variable aléatoire Y sachant une variable aléatoire X

Afin de pouvoir utiliser le problème de la meilleure approximation et sa solution rappelée dans la section 8.1, appelons W_X l'ensemble de toutes les v.a. de la forme $g(X)$, où $g : \mathbb{R} \to \mathbb{R}$ est une fonction telle que $g(X) \in L^2(\Omega)$.

Proposition 8.5 *W_X est un sous-espace vectoriel de $L^2(\Omega)$.*

Démonstration. Observons que $W_X \neq \emptyset$, puisque $1 \in W_X$. D'autre part, si $Z_1 = g_1(X)$, $Z_2 = g_2(X)$ et $a_1, a_2 \in \mathbb{R}$, alors $a_1 Z_1 + a_2 Z_2 = (a_1 g_1 + a_2 g_2)(X)$ appartient bien à W_X. $\square$

Remarque 8.6 Si $Z \in W_X$, alors les valeurs de Z sont entièrement déterminées par les valeurs de X : si $Z = g(X)$, alors pour tout $\omega \in \Omega$, $Z(\omega) = g(X(\omega))$.

Définition 8.7

Soit (X, Y) un vecteur aléatoire dans $\mathbb{R}^2$ tel que $Y \in L^2(\Omega)$. L'espérance conditionnelle de Y sachant X *est la v.a. Z dans W_X telle que $\|Y - Z\|^2 = E\big[(Y-Z)^2\big]$ est minimal (si une telle v.a. existe).*

Notation et interprétation. On note $\varphi(X) = E[Y \mid X]$ l'espérance conditionnelle de Y sachant X. Cette v.a. est la meilleure approximation (en moyenne quadratique) de Y par une v.a. qui est fonction de X. Ainsi, $E[Y \mid X] = \varphi(X)$, pour une certaine fonction $\varphi : \mathbb{R} \to \mathbb{R}$ à déterminer. Si on observe que $X(\omega) = x$, on estimera (ou on approximera) $Y(\omega)$ par $\varphi(x)$.

Caractérisation de $E[Y \mid X]$

Proposition 8.8 *La v.a. $E[Y \mid X]$ est déterminée par la relation*

$$E[Yg(X)] = E\big[E[Y \mid X]\, g(X)\big]\,, \tag{8.3}$$

pour toute fonction $g : \mathbb{R} \to \mathbb{R}$ telle que $g(X) \in L^2(\Omega)$.

Démonstration. C'est une conséquence du rappel d'algèbre linéaire, en particulier de (8.2). □

Pour chercher la fonction φ telle que $E[Y \mid X] = \varphi(X)$, on utilisera donc la propriété

$$E\big[Yg(X)\big] = E\big[\varphi(X)g(X)\big]\,, \tag{8.4}$$

pour toute fonction $g : \mathbb{R} \to \mathbb{R}$ telle que $g(X) \in L^2(\Omega)$, et en particulier pour les fonctions g qui sont continues et bornées.

Formule pour $E[Y \mid X]$ dans le cas discret

Nous allons faire le lien entre l'espérance conditionnelle de Y sachant X et la loi conditionnelle de Y sachant $X = x$ introduite dans la section 5.7.

Définition 8.9
Soit (X, Y) un vecteur aléatoire discret à valeurs dans $\mathbb{R}^2$.

a) Pour $x \in \mathbb{R}$ tel que $P\{X = x\} \neq 0$, posons

$$E[Y \mid X = x] \overset{\text{déf}}{=} \sum_y yP(Y = y \mid X = x) = \sum_y y\, f_Y(y \mid X = x)\,.$$

On observera que le nombre $E[Y \mid X = x]$ est l'espérance de la v.a. Y sous la loi conditionnelle de Y sachant $X = x$.

b) Plus généralement, pour $F \in \mathcal{F}$ tel que $P(F) \neq 0$, posons

$$E[Y \mid F] \overset{\text{déf}}{=} \sum_y yP(Y = y \mid F) = \frac{1}{P(F)} \sum_y yP\big(\{Y = y\} \cap F\big)\,.$$

Proposition 8.10 *Soit (X, Y) un vecteur aléatoire discret à valeurs dans $\mathbb{R}^2$. Pour $x \in \mathbb{R}$, posons*

$$\varphi(x) = \begin{cases} E[Y \mid X = x] & \text{si } P\{X = x\} \neq 0\,, \\ 0 & \text{sinon}\,. \end{cases}$$

Ceci définit une fonction $\varphi : \mathbb{R} \to \mathbb{R}$. L'espérance conditionnelle de Y sachant X est alors la v.a. $E[Y \mid X] = \varphi(X)$.

Démonstration. Soit $g : \mathbb{R} \to \mathbb{R}$ continue et bornée. Il suffit de vérifier l'égalité (8.4). Celle-ci s'écrit

$$\sum_x \sum_y y g(x) P\{X = x, Y = y\} = \sum_x \varphi(x) g(x) P\{X = x\},$$

ou encore

$$\sum_x g(x) \left(\sum_y y P\{X = x, Y = y\} \right) = \sum_x g(x)\, \varphi(x) P\{X = x\}.$$

Cette égalité vaudra pour tout $g : \mathbb{R} \to \mathbb{R}$ si

$$\sum_y y P\{X = x, Y = y\} = \varphi(x) P\{X = x\}.$$

Cette égalité est bien vérifiée puisque, pour $x \in \mathbb{R}$ tel que $P\{X = x\} \neq 0$,

$$\begin{aligned} \varphi(x) P\{X = x\} &= E[Y \mid X = x] P\{X = x\} \\ &= \sum_y y P\{X = x, Y = y\}. \end{aligned}$$

□

Formule pour $E[Y \mid X]$ dans le cas continu

Définition 8.11

Soit (X, Y) un vecteur aléatoire continu dans $\mathbb{R}^2$ avec fonction de densité conjointe $f_{(X,Y)}$. Pour $x \in \mathbb{R}$ tel que $f_X(x) \neq 0$, posons

$$E[Y \mid X = x] \overset{déf}{=} \int_{-\infty}^{+\infty} y\, f_Y(y \mid X = x)\, dy. \tag{8.5}$$

Ce nombre est l'espérance de Y calculée avec la densité conditionnelle de Y sachant $X = x$. Plus généralement, pour $h : \mathbb{R} \to \mathbb{R}$, nous posons

$$E[h(Y) \mid X = x] = \int_{-\infty}^{+\infty} h(y) f_Y(y \mid X = x)\, dy. \tag{8.6}$$

Proposition 8.12 *Soit (X, Y) un vecteur aléatoire continu dans $\mathbb{R}^2$ avec fonction de densité conjointe $f_{(X,Y)}$. Pour $x \in \mathbb{R}$, posons, à l'aide de (8.5),*

$$\varphi(x) = \begin{cases} E[Y \mid X = x] & \text{si } f_X(x) \neq 0, \\ 0 & \text{sinon}. \end{cases}$$

Ceci définit une fonction $\varphi : \mathbb{R} \to \mathbb{R}$. L'espérance conditionnelle de Y sachant X est alors la v.a. $E[Y \mid X] = \varphi(X)$. Plus généralement, pour $x \in \mathbb{R}$, posons

$$\psi(x) = \begin{cases} E[h(Y) \mid X = x] & \text{si } f_X(x) \neq 0, \\ 0 & \text{sinon}. \end{cases}$$

L'espérance conditionnelle de $h(Y)$ sachant X est alors la v.a. $E[h(Y) \mid X] = \psi(X)$.

Démonstration. Il suffit de traiter la deuxième affirmation car la première en découle en posant $h(y) = y$. Soit $g : \mathbb{R} \to \mathbb{R}$ continue et bornée. Il suffit de vérifier l'égalité

$$E\big[h(Y)g(X)\big] = E\big[\psi(X)g(X)\big]\,.$$

Celle-ci s'écrit

$$\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy\, h(y)g(x)f_{(X,Y)}(x,y) = \int_{-\infty}^{+\infty} \psi(x)g(x)f_X(x)\,dx\,,$$

ou encore

$$\int_{-\infty}^{+\infty} dx\, g(x)\left(\int_{-\infty}^{+\infty} dy\, h(y)f_{(X,Y)}(x,y)\right) = \int_{-\infty}^{+\infty} dx\, g(x)\psi(x)f_X(x)\,.$$

Cette égalité sera satisfaite pour tout g si

$$\int_{-\infty}^{+\infty} dy\, h(y)f_{(X,Y)}(x,y) = \psi(x)f_X(x)\,.$$

Cette égalité est bien vérifiée puisque, pour $x \in \mathbb{R}$ tel que $f_X(x) \neq 0$,

$$\begin{aligned}\psi(x)f_X(x) = E[h(Y) \mid X = x]\, f_X(x) = f_X(x) \int_{-\infty}^{+\infty} h(y)\frac{f_{(X,Y)}(x,y)}{f_X(x)}\,dy \\ = \int_{-\infty}^{+\infty} h(y)f_{(X,Y)}(x,y)\,dy\,.\end{aligned}$$

□

Formule pour $E[Y \mid X]$ dans le cas mixte discret / continu

Définition 8.13

Soit (X, Y) un vecteur aléatoire mixte dans $\mathbb{R}^2$ avec fonction de densité conjointe $f_{(X,Y)}$. Supposons que X soit une v.a. discrète et Y une v.a. continue. Rappelons que les probabilités conditionnelles dans le cas mixte sont introduites dans la définition 5.29.

a) Pour $x \in \mathbb{R}$ tel que $P\{X = x\} \neq 0$, posons

$$E[Y \mid X = x] \overset{déf}{=} \int_{-\infty}^{+\infty} y f_Y(y \mid X = x)\,dy\,. \tag{8.7}$$

Ce nombre est l'espérance de Y calculée avec la densité conditionnelle de Y sachant $X = x$.

b) Pour $y \in \mathbb{R}$ tel que $f_Y(y) \neq 0$, posons

$$E[X \mid Y = y] \overset{déf}{=} \sum_x x f_X(x \mid Y = y)\,. \tag{8.8}$$

Ce nombre est l'espérance de la v.a. X sous la loi conditionnelle de X sachant $Y = y$.

Proposition 8.14 *Soit (X, Y) un vecteur aléatoire mixte dans $\mathbb{R}^2$ avec fonction de densité conjointe $f_{(X,Y)}$. Supposons que X soit une v.a. discrète et Y une v.a. continue. Pour $x \in \mathbb{R}$, posons, à l'aide de (8.7),*

$$\varphi_1(x) = \begin{cases} E[Y \mid X = x] & \textit{si } P\{X = x\} \neq 0\,, \\ 0 & \textit{sinon}\,. \end{cases}$$

Ceci définit une fonction $\varphi_1 : \mathbb{R} \to \mathbb{R}$. De même, pour $y \in \mathbb{R}$, posons, à l'aide de (8.8),

$$\varphi_2(y) = \begin{cases} E[X \mid Y = y] & \textit{si } f_Y(y) \neq 0\,, \\ 0 & \textit{sinon}\,. \end{cases}$$

Ceci définit une fonction $\varphi_2 : \mathbb{R} \to \mathbb{R}$. L'espérance conditionnelle de Y sachant X est alors la v.a. $E[Y \mid X] = \varphi_1(X)$. L'espérance conditionnelle de X sachant Y est la v.a. $E[X \mid Y] = \varphi_2(Y)$.

Démonstration. Premièrement, l'égalité (8.4) s'écrit

$$\sum_x \int_{-\infty}^{+\infty} dy\, y g(x) f_{(X,Y)}(x, y) = \sum_x \varphi_1(x) g(x) P\{X = x\}$$

ou encore

$$\sum_x g(x) \left(\int_{-\infty}^{+\infty} dy\, y\ f_{(X,Y)}(x, y) \right) = \sum_x g(x)\, \varphi_1(x) P\{X = x\}\,.$$

Cette égalité sera satisfaite pour tout g si

$$\int_{-\infty}^{+\infty} dy\, y f_{(X,Y)}(x, y) = \varphi_1(x) P\{X = x\}\,.$$

Cette égalité est bien vérifiée puisque, pour $x \in \mathbb{R}$ tel que $P\{X = x\} \neq 0$,

$$\begin{aligned} \varphi_1(x) P\{X = x\} &= P\{X = x\} \int_{-\infty}^{+\infty} y\, f_Y(y \mid X = x)\, dy \\ &= P\{X = x\} \int_{-\infty}^{+\infty} y \frac{f_{(X,Y)}(x, y)}{P\{X = x\}}\, dy \\ &= \int_{-\infty}^{+\infty} y f_{X,Y}(x, y)\, dy\,. \end{aligned}$$

D'autre part, l'égalité (8.4) où on échange les rôles de X et Y, à savoir

$$E\big[Xg(Y)\big] = E\big[\varphi_2(Y) g(Y)\big]\,,$$

s'écrit

$$\sum_x \int_{-\infty}^{+\infty} dy\, x g(y) f_{(X,Y)}(x, y) = \int_{-\infty}^{+\infty} \varphi_2(y) g(y) f_Y(y)\, dy\,,$$

ou encore

$$\int_{-\infty}^{+\infty} dy\, g(y) \left(\sum_x x f_{(X,Y)}(x,y) \right) = \int_{-\infty}^{+\infty} dy\, g(y)\varphi_2(y) f_Y(y)\,.$$

Cette égalité sera satisfaite pour tout g si

$$\sum_x x f_{(X,Y)}(x,y) = \varphi_2(y) f_Y(y)\,.$$

Cette égalité est bien vérifiée puisque, pour $y \in \mathbb{R}$ tel que $f_Y(y) \neq 0$,

$$\begin{aligned}\varphi_2(y) f_Y(y) = f_Y(y) \sum_x x f_X(x \mid Y = y) &= f_Y(y) \sum_x x \frac{f_{(X,Y)}(x,y)}{f_Y(y)} \\ = \sum_x x\, f_{(X,Y)}(x,y)\,. \end{aligned}$$

□

Propriétés de l'espérance conditionnelle

Proposition 8.15

a) L'espérance de $E[Y \mid X]$ est $E[Y]$, c'est-à-dire $E\big[E[Y \mid X]\big] = E[Y]$.

b) Soit $h : \mathbb{R} \to \mathbb{R}$. Alors $E[h(X)Y \mid X] = h(X)E[Y \mid X]$.

c) $E[a_1 Y_1 + a_2 Y_2 \mid X] = a_1 E[Y_1 \mid X] + a_2 E[Y_2 \mid X]$.

d) Si X et Y sont deux v.a. indépendantes, alors $E[Y \mid X] = E[Y]$.

e) Soit $h : \mathbb{R}^2 \to \mathbb{R}$. Alors $E[h(X,Y) \mid X] = \psi(X)$, où $\psi(x) = E[h(x,Y) \mid X = x]$.

Démonstration. a) Il suffit de prendre $g \equiv 1$ dans l'égalité (8.3).

b) Observons que pour $g : \mathbb{R} \to \mathbb{R}$:

$$\begin{aligned} E\big[h(X)Y g(X)\big] &= E\big[Y h(X) g(X)\big] = E\big[Y (hg)(X)\big] \\ &= E\big[E[Y \mid X] h(X)\, g(X)\big] \end{aligned}$$

Par conséquent, $E\big[h(X)Y \mid X\big] = E[Y \mid X] h(X)$.

c) Cette propriété découle du fait que la projection orthogonale est une transformation linéaire.

d) et e) Voir l'exercice 8.5. □

Proposition 8.16 *Soit $F \in \mathcal{F}$ un événement tel que $P(F) > 0$.*

a) Soit X une v.a. discrète telle que $E[X^2] < \infty$. Alors

$$E[X \mid F] = \frac{1}{P(F)}\, E[X 1_F]\,.$$

b) Soit X une v.a. telle que $E[X^2] < \infty$. Sans exiger, comme dans la définition 8.9, que X soit une v.a. discrète, on pose

$$E[X \mid F] = \frac{1}{P(F)} E[X 1_F] \,.$$

Alors

$$E[X \mid 1_F] = E[X \mid F]\, 1_F + E[X \mid F^c]\, 1_{F^c} \,.$$

Pour la démonstration de cette proposition, voir l'exercice 8.7.

Formule des probabilités totales généralisée

Corollaire 8.17 *Soit $Y \in L^2(\Omega)$ une v.a discrète ou continue avec fonction de densité f.*

a) Soit X une v.a. discrète à valeurs dans un ensemble dénombrable D, tel que $P\{X = x\} > 0$, pour tout $x \in D$. Alors,

$$E[Y] = \sum_{x \in D} E[Y \mid X = x] P\{X = x\} \,.$$

b) Soit X une v.a. continue avec fonction de densité f_X. Alors

$$E[Y] = \int_{-\infty}^{+\infty} E[Y \mid X = x] f_X(x)\, dx$$

(avec la convention $E[Y \mid X = x] f_X(x) = 0$ si $f_X(x) = 0$).

Démonstration. a) D'après le point a) de la proposition 8.15, $E[Y] = E\big[E[Y \mid X]\big]$. D'après les propositions 8.10 et 8.14, $E[Y \mid X] = \varphi(X)$, où $\varphi(x) = E[Y \mid X = x]$. Ainsi, d'après le théorème 6.4,

$$\begin{aligned} E[Y] &= E\big[E[Y \mid X]\big] = E[\varphi(X)] \\ &= \sum_{x \in D} \varphi(x) P\{X = x\} = \sum_{x \in D} E[Y \mid X = x] P\{X = x\} \,. \end{aligned}$$

b) D'après le point a) de la proposition 8.15, $E[Y] = E\big[E[Y \mid X]\big]$. D'après les propositions 8.12 et 8.14, $E[Y \mid X] = \varphi(X)$, où $\varphi(x) = E[Y \mid X = x]$ si $f_X(x) > 0$, $\varphi(x) = 0$ sinon. Ainsi, d'après le théorème 6.4,

$$\begin{aligned} E[Y] &= E\big[E[Y \mid X]\big] = E[\varphi(X)] \\ &= \int_{-\infty}^{\infty} \varphi(x) f_X(x)\, dx \\ &= \int_{\{x : f_X(x) > 0\}} E[Y \mid X = x] f_X(x)\, dx \\ &= \int_{-\infty}^{+\infty} E[Y \mid X = x] f_X(x)\, dx \,, \end{aligned}$$

d'après la convention mentionnée dans l'énoncé. □

8.4 Variance conditionnelle

Imaginons qu'on observe X et qu'on estime Y par $E[Y \mid X]$. Il est alors naturel de s'intéresser à l'erreur quadratique moyenne de cette estimation, ce qui nous amène au concept de variance conditionnelle.

Définition 8.18
$\operatorname{Var}(Y \mid X) = E\left[\left(Y - E[Y \mid X]\right)^2 \mid X\right]$.

Proposition 8.19

a) $\operatorname{Var}(Y \mid X) = E[Y^2 \mid X] - (E[Y \mid X])^2$.

b) Posons $\operatorname{Var}(Y \mid X = x) \overset{déf}{=} E[Y^2 \mid X = x] - (E[Y \mid X = x])^2$ *et* $\psi(x) = \operatorname{Var}(Y \mid X = x)$ *(il s'agit donc de la variance de Y calculée avec la loi conditionnelle de Y sachant $X = x$). Alors* $\operatorname{Var}(Y \mid X) = \psi(X)$.

c) Si X et Y sont indépendantes, alors $\operatorname{Var}(Y \mid X) = \operatorname{Var}(Y)$.

DÉMONSTRATION. Voir l'exercice 8.6.

Proposition 8.20 $\operatorname{Var}(Y) = E\big[\operatorname{Var}(Y \mid X)\big] + \operatorname{Var}\big(E[Y \mid X]\big)$.

D'après la définition 8.18 et la proposition 8.15a), le premier terme du membre de droite est égal à $E[(Y - E[Y \mid X])^2]$. C'est donc l'erreur quadratique moyenne liée à l'estimation de Y par $E[Y \mid X]$. Quand au second terme, c'est la variance de l'estimation de Y.

DÉMONSTRATION. Observons que

$$\begin{aligned}\operatorname{Var}(Y) &= E\left[\left(Y - E[Y]\right)^2\right] = \left\|Y - E[Y]\right\|^2 \\ &= \left\|Y - E[Y \mid X] + E[Y \mid X] - E[Y]\right\|^2.\end{aligned}$$

D'après le théorème de Pythagore, le membre de droite est égal à

$$\left\|Y - E[Y \mid X]\right\|^2 + \left\|E[Y \mid X] - E[Y]\right\|^2.$$

Le premier terme de cette somme est égal à

$$E\big[\big(Y - E[Y \mid X]\big)^2\big] = E\big[\operatorname{Var}(Y \mid X)\big]$$

d'après la définition 8.18 et la proposition 8.15. Puisque $E[Y] = E\big[E[Y \mid X]\big]$, le deuxième terme de cette somme est

$$E\big[\big(E[Y \mid X] - E\big[E[Y \mid X]\big]\big)^2\big] = \operatorname{Var}\big(E[Y \mid X]\big).$$

La proposition est démontrée. $\square$

Exemple 8.21 Un casino a constaté que le nombre N de personnes qui jouent à la roulette un soir donné est une v.a. de Poisson de paramètre $\lambda > 0$. Soit X_i le bénéfice ou la perte du casino lié aux paris du joueur i. On admet que les v.a. X_i sont i.i.d., d'espérance μ et de variance σ^2, et que N et la suite des X_i sont mutuellement indépendantes. S'il y a n joueurs, le bénéfice du casino est $S_n = X_1 + \cdots + X_n$. Avant de connaître le nombre N de joueurs, le bénéfice pour la soirée est $S_N = X_1 + \cdots + X_N$.

a) Trouver $E[S_N \mid N]$ et $\mathrm{Var}(S_N \mid N)$. Calculons tout d'abord $E[S_n \mid N = n]$. Comme S_n et N sont indépendantes, la propriété d) de la proposition 8.15 affirme que $E[S_n \mid N = n] = E[S_n]$ et la linéarité de l'espérance nous dit que $E[S_n] = n\mu$. Par conséquent,

$$E[S_n \mid N = n] = n\mu\,.$$

Nous devons encore calculer $E[S_N \mid N]$. Par définition de l'espérance conditionnelle, nous cherchons une v.a. $Z = \varphi(N)$ telle que $E\big[S_N g(N)\big] = E\big[Zg(N)\big]$ pour toute fonction g continue et bornée. Par les propriétés de l'espérance (noter qu'il est possible de démontrer que la somme et l'espérance ci-dessous peuvent être permutées, mais nous ne présentons pas les détails ici),

$$\begin{aligned} E\big[S_N\, g(N)\big] &= E\left[S_N\, g(N) \sum_{n=0}^{\infty} 1_{\{N=n\}}\right] = \sum_{n=0}^{\infty} E\big[S_N\, g(N) 1_{\{N=n\}}\big] \\ &= \sum_{n=0}^{\infty} E\big[S_n\, g(n) 1_{\{N=n\}}\big] = \sum_{n=0}^{\infty} g(n) E[S_n 1_{\{N=n\}}]. \end{aligned}$$

Par la proposition 8.16,

$$\begin{aligned} E\big[S_N\, g(N)\big] &= \sum_{n=0}^{\infty} g(n) E[S_n \mid N = n] P\{N = n\} = \sum_{n=0}^{\infty} n\mu\, g(n) P\{N = n\} \\ &= \sum_{n=0}^{\infty} E\big[n\mu\, g(n) 1_{\{N=n\}}\big] = \sum_{n=0}^{\infty} E\big[N\mu\, g(N) 1_{\{N=n\}}\big] \\ &= E\left[N\mu\, g(N) \sum_{n=0}^{\infty} 1_{\{N=n\}}\right] = E\big[N\mu\, g(N)\big]\,. \end{aligned}$$

Ainsi, $Z = \varphi(N) = N\mu$ vérifie la définition de l'espérance conditionnelle et $E[S_N \mid N] = N\mu$. De plus, par l'indépendance de S_n et N, le point c) de la proposition 8.19 et le point c) de la proposition 6.19, nous avons $\mathrm{Var}(S_n \mid N = n) = \mathrm{Var}(S_n) = n\,\mathrm{Var}(X_1) = n\sigma^2$. Ainsi, un raisonnement analogue à celui effectué ci-dessus avec $(S_N - N\mu)^2$ en lieu et place de S_N montre que $\mathrm{Var}(S_N \mid N) = N\sigma^2$.

b) Trouver $E[S_N]$ et $\mathrm{Var}(S_N)$. D'après la propriété a) de la proposition 8.15 et le calcul qui vient d'être fait,

$$E[S_N] = E\big[E[S_N \mid N]\big] = E[N\mu] = \mu E[N] = \mu\lambda\,.$$

D'autre part, d'après la proposition 8.20 et la partie a),

$$\begin{aligned}\operatorname{Var}(S_N) &= E\big[\operatorname{Var}(S_N \mid N)\big] + \operatorname{Var}\big(E[S_N \mid N]\big) = E[N\sigma^2] + \operatorname{Var}(N\mu)\\ &= \sigma^2 E[N] + \mu^2 \operatorname{Var}(N) = \lambda(\sigma^2+\mu^2)\,.\end{aligned}$$

8.5 Exercices

8.1 Soit (X,Y) un vecteur aléatoire discret dans $\mathbb{R}^2$ dont la fonction de densité conjointe est

$$P\{X=x, Y=y\} = \binom{n}{x}\binom{n-x}{y} p^x q^y (1-p-q)^{n-x-y}\,,$$

pour $x, y \in \{0,\dots,n\}$ avec $x+y \leqslant n$ (loi multinomiale).

a) Déterminer la fonction de densité marginale de X.

b) Montrer que les fonctions de densité conditionnelles $P\{Y=y \mid X=x\}$ et $P\{X=x \mid Y=y\}$ sont des densités binomiales dont on déterminera les paramètres.

c) Calculer $E[Y \mid X=x]$ et $E[X \mid Y=y]$.

d) Trouver $E[Y \mid X]$ et $E[X \mid Y]$.

e) Calculer $P\big\{E[X \mid Y]=0\big\}$.

8.2 Soient X et Y deux v.a. indépendantes, où X est binomiale (n,p) et Y est binomiale (m,p). Posons $Z=X+Y$.

a) Trouver $P\{Y=y, Z=z\}$ et $P\{Y=y \mid Z=z\}$.

b) En déduire $E[Y \mid Z=z]$ et $E[Y \mid Z]$.

c) Calculer $P\big\{E[Y \mid Z]=0\big\}$.

8.3 Soient X_1 et X_2 deux v.a. indépendantes, où X_1 et X_2 sont des v.a. de Poisson de paramètres respectifs λ_1 et λ_2. Posons $Z=X_1+X_2$.

a) Trouver $P\{X_2=x_2, Z=z\}$ et $P\{X_2=x_2 \mid Z=z\}$.

b) En déduire $E[X_2 \mid Z=z]$ et $E[X_2 \mid Z]$.

c) Calculer $P\big\{E[X_2 \mid Z] \leqslant \frac{\lambda_2}{\lambda_1+\lambda_2}\big\}$.

8.4 Soit (X,Y) un vecteur aléatoire continu dont la fonction de densité conjointe est

$$f_{(X,Y)}(x,y) = \begin{cases} 10\,x^2 y & \text{si } 0<y<x<1\,,\\ 0 & \text{sinon}\,.\end{cases}$$

a) Déterminer $f_Y(y \mid X=x)$.

b) En déduire $E[Y \mid X=x]$ et $\operatorname{Var}(Y \mid X=x)$, puis $E[Y \mid X]$ et $\operatorname{Var}(Y \mid X)$.

8.5* a) Soient X et Y deux v.a. indépendantes. Montrer que $E[Y \mid X] = E[Y]$.

b) Soit (X, Y) un vecteur aléatoire (discret ou continu) dans $\mathbb{R}^2$ et $h : \mathbb{R}^2 \to \mathbb{R}$. Montrer que

$$E\big[h(X,Y) \mid X\big] = \psi(X)\,,$$

où $\psi : \mathbb{R} \to \mathbb{R}$ est définie par $\psi(x) = E\big[h(x,Y) \mid X = x\big]$.

8.6 Démontrer la proposition 8.19.

8.7 Démontrer la proposition 8.16.

8.8* Soit $Z = (Z_1, Z_2)$ un vecteur aléatoire continu dont la fonction de densité est donnée dans l'exercice 6.30, dont on utilise les notations.

a) Montrer que la fonction de densité conditionnelle $f_{Z_2}(z_2 \mid Z_1 = z_1)$ est celle d'une v.a.

$$N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}\,(z_1 - \mu_1), \sigma_2^2(1-\rho^2)\right)\,,$$

où $\rho = \mathrm{Corr}(Z_1, Z_2)$.

b) Déduire de a) que $E[Z_2 \mid Z_1] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}\,(Z_1 - \mu_1)$ et $\mathrm{Var}(Z_2 \mid Z_1) = \sigma_2^2(1-\rho^2)$.

8.9 Une poule pond X œufs, où X est une v.a. de Poisson de paramètre λ. Chaque œuf pondu éclot avec probabilité $p \in]0,1[$, indépendamment des autres. Soit Y le nombre de poussins obtenus.

a) Quelle est la loi conditionnelle de Y sachant $X = x$?

b) Déterminer $E[Y|X]$ et en déduire $E[Y]$ et $E[XY]$.

c) Calculer $\mathrm{Var}(Y \mid X)$ et $\mathrm{Var}(Y)$.

d) Quelle est la fonction de densité conjointe du vecteur aléatoire (X, Y) ?

8.10 Soit (Θ, X) un vecteur aléatoire, où Θ est une v.a. uniforme sur $[0,1]$ et la loi conditionnelle de X sachant $\Theta = \theta$ est binomiale $B(n, \theta)$ $(n \in \mathbb{N}^*)$.

a) Calculer $E[X \mid \Theta]$ et $\mathrm{Var}(X \mid \Theta)$.

b) Calculer $E[X]$ et $\mathrm{Var}(X)$.

Chapitre 9

Exercices de révision

Ce chapitre propose des exercices de révision regroupant toutes les notions traitées dans le cadre de cet ouvrage. De difficultés variées, ils constituent de bons exemples de questions pouvant apparaître dans un examen.

9.1* Un ordinateur est équipé d'un logiciel antivirus. Au moment d'allumer l'ordinateur, le logiciel informe l'utilisateur de la présence ou non d'un virus. La probabilité que l'ordinateur soit infecté par un virus est 0,2. Si l'ordinateur est infecté, la probabilité que le logiciel détecte le virus est 0,9. Si l'ordinateur n'est pas infecté, la probabilité que le logiciel déclare qu'un virus est présent est 0,02.

a) Si le logiciel signale la présence d'un virus, quelle est la probabilité que l'ordinateur soit infecté ?

b) Quelle est la probabilité que l'ordinateur soit infecté par un virus et que le logiciel ne le détecte pas ?

9.2* On tire une main de 4 cartes d'un jeu de 52 cartes.

a) Combien y a-t-il de résultats possibles ?

b) Quelle est la probabilité d'obtenir deux piques et deux trèfles ?

c) Sachant qu'on a obtenu deux piques et deux trèfles, quelle est la probabilité d'obtenir quatre valeurs différentes ?

9.3* On tire successivement deux boules avec remise d'une urne qui contient trois boules numérotées 1, 2 et 3. Soient X_1 le produit des numéros obtenus et X_2 la somme des numéros obtenus.

a) Donner un espace de probabilité $(\Omega, \mathcal{F}, P)$ qui correspond à cette expérience aléatoire.

b) Définir deux v.a. $X_1 : \Omega \to \mathbb{R}$ et $X_2 : \Omega \to \mathbb{R}$ qui correspondent à la description ci-dessus.

c) Donner la fonction de densité conjointe de X_1 et X_2 ainsi que leurs fonctions de densité marginales.

d) Est-ce que X_1 et X_2 sont indépendantes ? Justifier votre réponse.

9.4* Si $x \in \mathbb{R}$, on note $[x]$ la partie entière de x (c'est-à-dire $[x] \in \mathbb{Z}$ et $[x] \leqslant x < [x]+1$). Soit X une v.a. exponentielle de paramètre $\lambda > 0$. Posons $Y = [X]$.

a) Quelle est la fonction de densité de Y ?

b) Soit Z une v.a. telle que $Z^5 + Z + 1 = X$. Quelle est la fonction de densité de Z ?

9.5* Soient X_1, X_2 deux v.a. indépendantes et uniformes sur $[0,1]$. Calculer

$$P\left\{X_2 - X_1 < \frac{1}{4}\,,\, X_1 + X_2 < \frac{1}{2}\right\}.$$

9.6* Pour $n \in \mathbb{N}^*$, soient $X_1, \ldots, X_n$ des v.a. indépendantes ayant toutes F pour fonction de répartition. On pose $M_n = \max(X_1, \ldots, X_n)$.

a) Montrer que $P\{M_n \leqslant y\} = (F(y))^n$.

b) On suppose de plus que les X_i sont des v.a. uniformes sur $[0,1]$. Montrer que pour tout $x > 0$,

$$\lim_{n\to\infty} P\big\{n(1 - M_n) < x\big\} = 1 - e^{-x}\,.$$

9.7* Soient $n \geqslant 2$ et $X_1, \ldots, X_n$ des v.a. indépendantes et de loi uniforme sur $[0,1]$. Posons $Y = \min(X_1, \ldots, X_n)$ et $Z = \max(X_1, \ldots, X_n)$.

a) Trouver la fonction de densité f_Y de Y.

b) Pour $0 < y < z < 1$, calculer $P\{y < Y \leqslant Z \leqslant z\}$.

c) Trouver la fonction de répartition conjointe de (Y, Z) et sa fonction de densité conjointe.

9.8* On dispose de n jetons numérotés de 1 à n et de n urnes, aussi numérotées de 1 à n. On place indépendamment les jetons dans les urnes, de sorte que le jeton numéro i a probabilité $\frac{1}{i}$ de se retrouver dans chacune des urnes 1 à i.

a) Pour $i = 1, \ldots, n$, quelle est la probabilité que l'urne i reste vide ?

b) Quelle est l'espérance du nombre d'urnes vides ?

c) Quelle est la probabilité qu'aucune urne ne reste vide ?

9.9* Le temps nécessaire pour corriger une série d'exercices est de 5 minutes en moyenne, avec un écart type de 2 minutes. S'il y a 100 séries à corriger, quel est le laps de temps minimal tel que la probabilité que les 100 séries puissent être corrigées en ce laps de temps soit supérieure à 0,99 ? On indiquera les hypothèses utilisées pour effectuer le calcul.

9.10* Soient X une v.a. de loi $N(0,1)$ et Y une v.a. telle que $P\{Y = +1\} = P\{Y = -1\} = \frac{1}{2}$. On suppose que X et Y sont indépendantes et on pose $Z = XY$.

a) Est-ce que Y et Z sont indépendantes ?
b) Montrer que Z est de loi $N(0,1)$.
c) Montrer que $\text{Cov}(X,Z) = 0$.
d) Est-ce que X et Z sont indépendantes ?

9.11* Soient X_1, X_2, X_3 des v.a. i.i.d dont la fonction de densité est

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

Posons $U_1 = X_1 + X_2$, $U_2 = X_1 + X_3$, $U_3 = X_2 + X_3$. Quelle est la fonction de densité conjointe de (U_1, U_2, U_3) ?

9.12* Une expérience consiste à lancer successivement 5 dés équilibrés.

a) Quel est le nombre de résultats possibles de l'expérience ?
b) Quelle est la probabilité que le résultat soit « 1 paire » (c'est-à-dire deux résultats identiques et trois autres chiffres distincts) ?
c) Quelle est la probabilité que le résultat soit « 2 paires » (c'est-à-dire que 3 chiffres distincts apparaissent sur les cinq dés, dont deux chiffres apparaissent deux fois) ?

9.13* Trois pièces de monnaie sont sur une table. L'une est biaisée et tombe sur pile avec probabilité $p > \frac{1}{2}$. Les deux autres sont équilibrées. Vous en choisissez une au hasard. En dix lancers, vous obtenez 8 piles et 2 faces. Quelle est la probabilité d'avoir ramassé la pièce biaisée ?

9.14* Soient X et Y deux v.a. continues indépendantes ayant respectivement pour fonction de densité f_X et f_Y.

a) Exprimer $P\{X \leqslant x, X + Y \leqslant u\}$ à l'aide de f_X et f_Y.
b) Exprimer la fonction de densité conjointe du vecteur aléatoire $(X, X+Y)$ à l'aide de f_X et f_Y.
c) Donner la fonction de densité conditionnelle de X sachant $X + Y = u$ ($u \in \mathbb{R}$).
d) Calculer explicitement la fonction de densité conditionnelle en c) lorsque X et Y suivent des lois exponentielles de paramètre $\lambda = 1$ et $u > 0$.
e) Dans le cadre de d), quelle est la loi conditionnelle de X sachant $X+Y = u$ ($u > 0$) ?

9.15* Dans une ville, lors d'une élection, 20 % des habitants préfèrent le candidat A. On choisit 100 habitants au hasard et on s'intéresse à la probabilité que 18 à 22 des personnes choisies préfèrent le candidat A.

a) Exprimer la valeur exacte de cette probabilité (sans la calculer numériquement).
b) En utilisant la correction de continuité, donner une approximation de cette probabilité.

9.16* Soit $n > 1$ et soient $X_1, \dots, X_n$ des v.a. continues i.i.d., toutes ayant f pour fonction de densité et F pour fonction de répartition.

a) Pour $1 < m \leqslant n$, calculer $P\big\{X_m > \max(X_1, \dots, X_{m-1})\big\}$.

b) On observe successivement les v.a. $X_1, \dots, X_n$. On dit qu'un record a lieu à l'instant m si $m > 1$ et $X_m > \max(X_1, \dots, X_{m-1})$. Quelle est l'espérance du nombre N de records ?

9.17* Une urne contient quinze boules blanches numérotées de 1 à 15 et quinze boules noires aussi numérotées de 1 à 15. On tire quatre boules sans remise.

a) Quelle est la probabilité que tous les numéros obtenus soient distincts ?

b) Quelle est la probabilité d'obtenir une seule paire de boules portant le même numéro ?

c) Quelle est la probabilité d'obtenir deux paires de boules avec les mêmes numéros ?

9.18* On suppose que la fonction de densité conjointe du vecteur aléatoire (X, Y) est

$$f(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\,\left(x^2 - 2\rho xy + y^2\right)\right).$$

a) Soit $Z = \frac{Y-\rho X}{\sqrt{1-\rho^2}}$. Montrer que la densité conjointe du vecteur aléatoire (X, Z) est celle d'un vecteur aléatoire normal standard dans $\mathbb{R}^2$.

b) Soit $E = (\mathbb{R}_+^*)^2$. Montrer que

$$P\big\{(X, Y) \in E\big\} = \frac{1}{4} + \frac{1}{2\pi}\,\operatorname{arctg}\left(\frac{\rho}{\sqrt{1-\rho^2}}\right).$$

9.19* Une expérience consiste à lancer successivement 5 dés équilibrés.

a) Quel est le nombre de résultats possibles ?

b) Quelle est la probabilité que le résultat soit « 1 brelan » (c'est-à-dire trois chiffres identiques et deux autres chiffres distincts) ?

c) Quelle est la probabilité que le résultat soit « 2 paires » (c'est-à-dire que 3 chiffres distincts apparaissent sur les cinq dés, dont deux chiffres apparaissent deux fois).

9.20* 20 % des personnes qui partent en vacances d'hiver partent dans un pays chaud, les autres vont aux sports d'hiver. La probabilité d'attraper un coup de soleil dans un pays chaud est $\frac{4}{10}$, alors que cette probabilité est de $6/10$ aux sports d'hiver.

a) Quelle est la probabilité qu'une personne qui part en vacances d'hiver revienne sans coup de soleil ?

b) Une personne revient sans coup de soleil. Quelle est la probabilité qu'elle soit partie dans un pays chaud ?

9.21* Soit X une v.a. $N(\mu = 1, \sigma^2 = 4)$. On définit une v.a. Y par

$$Y = \begin{cases} -2 & \text{si } X < 0\,, \\ 0 & \text{si } 0 \leqslant X \leqslant 2\,, \\ 2 & \text{si } X > 2\,. \end{cases}$$

a) Quelle est la fonction de densité de Y ?

b) Soit Z une v.a. telle que $Z^7 + Z^3 + Z = X$. Quelle est la fonction de densité de Z ?

9.22* Soient $(\Omega, \mathcal{F}, P)$ un espace de probabilité, $(G_n, n \in \mathbb{N})$ une suite d'événements. L'événement « sauf pour un nombre fini de valeurs de n, tous les G_n se réalisent » est noté $\liminf_{n\to\infty} G_n$.

a) Montrer que

$$\liminf_{n\to\infty} G_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} G_m\,.$$

b) Montrer que

$$P\left(\liminf_{n\to\infty} G_n\right) \leqslant \liminf_{n\to\infty} P(G_n)\,.$$

Rappel : si $(a_n, n \in \mathbb{N})$ est une suite de nombres, alors $\liminf\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} \inf\limits_{m\geqslant n} a_m$.

9.23* On effectue une suite d'épreuves indépendantes. Chacune de ces épreuves a trois résultats possibles R_1, R_2 et R_3, de probabilités respectives p_1, p_2 et p_3, où $p_i > 0$ et $p_1 + p_2 + p_3 = 1$.

a) Soient $0 < n_1 < n_2$ deux entiers. Quelle est la probabilité que le résultat R_1 apparaisse pour la première fois lors de l'épreuve numéro n_1 et que le résultat R_2 apparaisse pour la première fois lors de l'épreuve numéro n_2 ?

b) A l'aide de a) , montrer que la probabilité que le résultat R_1 apparaisse avant le résultat R_2 est

$$\frac{p_1}{p_1 + p_2}\,.$$

9.24* On dispose de 6 urnes numérotées de 1 à 6. On dépose successivement au hasard 15 boules dans ces urnes.

a) Quel est l'ensemble Ω des résultats possibles ?

b) Quelle est la probabilité que l'urne 1 reste vide ?

c) Quelle est la probabilité que les urnes 1, 2, 3 et 4 contiennent chacune trois boules, l'urne 5 deux boules et l'urne 6 une boule ?

d) Quelle est la probabilité que quatre urnes contiennent chacune 3 boules, une urne contienne 2 boules et une urne contienne 1 boule ?

9.25* Soient X_1, X_2 et X_3 trois v.a. indépendantes et uniformes sur $[0,1]$. Posons $Y = \min(X_1, X_2, X_3)$ et $Z = \max(X_1, X_2, X_3)$.

a) Trouver les fonctions de densité marginales f_Y et f_Z des v.a. Y et Z.

b) Trouver la fonction de densité conjointe $f_{(Y,Z)}$ du vecteur aléatoire (Y, Z).

c) Trouver $E[Y \mid Z]$.

9.26* Un ouvrier enfonce un clou de 0,25 cm en moyenne par coup de marteau avec un écart type de 0,1 cm. Combien de coups de marteau sont-ils nécessaires pour que la probabilité d'enfoncer entièrement un clou de 9 cm de long soit supérieure à 0,9 ? On utilisera le théorème limite central en expliquant les hypothèses qui justifient son utilisation.

9.27* Soient X_1 et X_2 deux v.a. indépendantes, suivant une loi gamma respectivement de paramètres (s, λ) et (t, λ). Posons

$$Y_1 = X_1 + X_2 \qquad \text{et} \qquad Y_2 = \frac{X_1}{X_1 + X_2}\,.$$

Trouver la fonction de densité conjointe de (Y_1, Y_2).

9.28* Une urne contient N_i boules de couleur i, $i = 1, 2, 3$. On fait des tirages *sans remise*. Montrer que la probabilité d'obtenir une boule de couleur 1 avant une boule de couleur 2 est $\frac{N_1}{N_1+N_2}$.

9.29 Soient $(\Omega, \mathcal{F}, P)$ un espace de probabilité, $(X_n)_{n\in\mathbb{N}}$ et X des v.a. Soit

$$G = \Big\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big\}\,.$$

a) Expliquer pourquoi

$$G = \bigcap_{k\in\mathbb{N}^*} \bigcup_{N\in\mathbb{N}} \bigcap_{n>N} \left\{|X_n - X| < \frac{1}{k}\right\}.$$

b) On suppose que pour tout $k \in \mathbb{N}^*$, $\sum_{n=1}^{\infty} P\big\{|X_n - X| \geqslant \frac{1}{k}\big\} < \infty$. Montrer que $P(G^c) = 0$ et donc $P(G) = 1$.

9.30* Soit (X, Y) un vecteur aléatoire dans $\mathbb{R}^2$ dont la fonction de densité conjointe est

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\,(x^2 - 2\rho xy + y^2)\right).$$

a) Soient Z_1 et Z_2 telles que $X = \frac{Z_1+Z_2}{2}$ et $Y = \frac{Z_1-Z_2}{2}$. Trouver la fonction de densité conjointe du vecteur aléatoire (Z_1, Z_2).

b) Sans calcul d'intégrale, donner la fonction de densité marginale de Z_1, celle de Z_2, et indiquer la loi de ces deux v.a.

c) Est-ce que Z_1 et Z_2 sont indépendantes ?

d) A l'aide des propriétés de l'espérance conditionnelle, déterminer $E[X \mid Z_1]$.

Chapitre 10

Une brève histoire de la théorie des probabilités

Pratiquement toutes les activités humaines sont affectées par le hasard : un accident, le temps qu'il fait, une maladie qu'on attrape au mauvais moment, un mouvement inattendu de la bourse, sont des exemples courants. Dans ce chapitre, nous examinons la naissance et le développement de la théorie des probabilités depuis l'Antiquité jusqu'à nos jours.

L'Antiquité

Au contraire de la géométrie et de l'étude des nombres, il n'y a pas eu d'étude mathématique des probabilités pendant l'Antiquité. Les anciens utilisent le hasard dans le divertissement (jeux de hasard) et dans quelques autres circonstances. Par exemple, le hasard permet de prendre des décisions dans des situations où aucun argument rationnel n'est utile pour faire un choix. Ce rôle est notamment présent dans le sport, pour décider par exemple dans quel ordre des équipes vont se rencontrer, mais aussi dans quelques situations plus dramatiques (1). On se rend compte déjà à cette époque que l'utilisation de méthodes de décision basées sur le hasard a aussi l'avantage d'éviter d'engager la responsabilité d'un décideur !

De la Renaissance au dix-neuvième siècle

Une des premières références connues à des calculs de probabilités est un calcul élémentaire mentionné dans un commentaire, publié à Venise en 1477, sur la *Divine Comédie* de Dante. Quelques autres références aux probabilités apparaissent dans un court traité de Jérôme Cardan (1501-1576), publié en 1663, longtemps après le décès de l'auteur. Galilée (1564-1642) fait aussi quelques calculs concernant des jeux de dés.

Pierre-Simon de Laplace (1749-1827) attribue l'origine de la théorie des probabilités à Blaise Pascal (1623-1662) et Pierre de Fermat (1608 ?-1665),

(1). Une illustration de cette utilisation apparaît dans la chanson française « Il était un petit navire », qui raconte l'histoire d'un bateau à la dérive dans lequel les vivres viennent à manquer. La chanson raconte que les marins « tirent à la courte paille » pour décider « qui sera mangé » !

pour leur résolution en 1654 environ du « problème des points », qui leur avait été posé par Antoine Gambaud, dit le Chevalier de Méré (exercice 3.27). Un peu plus tard, en 1657, Christiaan Huygens (1629-1695) publie un traité qui contient des calculs de probabilités combinatoires. Ces calculs sont repris et développés dans le livre *Ars Conjectandi* de Jakob Bernoulli (1654-1705), publié en 1713. Ce livre contient des calculs liés aux permutations et au binôme de Newton, une étude de divers problèmes combinatoires ainsi que la loi faible des grands nombres.

Pierre Rémond de Montmort (1678-1719) publie en 1708 l'*Essay d'analyse sur les Jeux de Hasard,* qui mentionne aussi le binôme de Newton, divers problèmes de jeux de cartes, et le « problème des points ». Un autre traité particulièrement influent est celui du mathématicien français Abraham De Moivre (1667-1754), qui publie en 1718 (et en anglais !) *The Doctrine of Chances.* Ce traité analyse 26 problèmes de probabilités combinatoires. La « formule de Stirling » (exemple 2.3) y apparaît (le nom provient de James Stirling (1692-1770)), ainsi que la formule d'inclusion/exclusion, les probabilités conditionnelles et l'approximation d'une loi binomiale par la loi gaussienne, c'est-à-dire une toute première version d'un théorème limite central.

Nikolaus Bernoulli (1687-1759) publie en 1711 sa thèse de doctorat, qui est sans doute le premier écrit où apparaît la loi uniforme.

Daniel Bernoulli (1700-1782) étudie des problèmes d'assurance et montre l'intérêt du principe de répartition des risques. Il montre aussi en 1732 que le hasard n'explique pas les inclinaisons des orbites des planètes. Il utilise le calcul différentiel dans ses calculs de probabilités et s'intéresse au problème de l'accumulation des erreurs de mesures. Un de ses mémoires discute le « problème de Saint-Pétersbourg », qui concerne la valeur à attribuer à une récompense aléatoire dont l'espérance est infinie. Il étudie aussi l'approximation de la loi binomiale par la loi gaussienne, même si son approche n'est pas complètement rigoureuse.

Le mathématicien suisse Leonhard Euler (1707-1783) fait quelques contributions à la théorie des probabilités, principalement en liaison avec des questions d'assurance (annuités et mortalité).

Jean d'Alembert (1717-1783) étudie le problème de Saint-Pétersbourg et énonce la règle maintenant usuelle du calcul de l'espérance d'une variable aléatoire discrète. Il considère cependant qu'une très petite probabilité correspond à une probabilité nulle !

Thomas Bayes (1702-1761) publie en 1764, à titre posthume, un cas particulier de son théorème. La formule générale sera énoncée plus tard par Laplace.

Joseph Louis Lagrange (1736-1813) publie vers 1770 un mémoire qui concerne 10 problèmes de probabilités, dont le problème de la durée d'un jeu de hasard et des problèmes concernant des variables aléatoires de loi continue.

Le naturaliste Georges Buffon (1707-1788) publie en 1777 son *Essai d'Arithmétique Morale,* où il parle notamment de problèmes qui font le lien entre les probabilités et la géométrie, et en particulier, du fameux problème de « l'aiguille de Buffon » (exercice 5.21).

Un tournant dans la théorie des probabilités est la publication en 1812 du livre *Théorie analytique des probabilités* par Laplace. Laplace poursuit l'étude de D. Bernoulli concernant les inclinaisons des orbites des planètes. Il calcule l'intégrale de la fonction de densité gaussienne. Il utilise le calcul des probabilités pour expliquer pourquoi les orbites des comètes sont des paraboles, plutôt que des hyperboles ou des ellipses. Il développe la théorie des fonctions génératrices de variables aléatoires (qui apparaît déjà chez de Moivre). Il complète les travaux de D. Bernoulli, Stirling et de Moivre sur le théorème limite central et présente la méthode des moindres carrés. Il s'intéresse en particulier aux *incertitudes* dans les résultats de *mesures* : en vue de déterminer les orbites des planètes, il se demande comment une accumulation de petites erreurs inconnues affecte le résultat final.

Carl Friedrich Gauss (1777-1855) travaille aussi sur la méthode des moindres carrés : ses calculs permettent de prédire l'orbite de l'astéroïde Cérès (en 1801). En 1829, Gauss montre que si les erreurs sont réparties selon la loi normale, aussi appelée loi de Gauss, alors la méthode des moindres carrés est optimale.

John Venn (1834-1923), mathématicien et logicien, introduit le diagramme de Venn en 1881. Il cherche à représenter les liens entre différentes propositions logiques.

La première moitié du vingtième siècle

Un des résultats les plus importants du début du vingtième siècle est le théorème limite central, que nous avons présenté dans le chapitre 7. Ce théorème est l'aboutissement d'un long développement. La fonction de densité de la loi normale apparaît déjà dans les travaux de A. de Moivre en 1738, comme nous l'avons dit ci-dessus. En particulier, de Moivre, qui avait établi la formule dite « de Stirling » en 1733, a montré comment la loi binomiale peut être approchée par la loi normale. En 1810, Laplace démontre le théorème limite central pour les variables aléatoires indépendantes et bornées. Même si des grands mathématiciens français comme Joseph Bertrand (1822-1900) et Henri Poincaré (1854-1912) s'intéressent au theorème limite central, la première preuve rigoureuse sous des hypothèses plus faibles que celles de Laplace, mais qui s'appuie sur ses idées, est donnée en 1901 par le mathématicien russe Aleksandr M. Liapounov (1857-1918). La preuve que nous présentons dans le présent ouvrage est due à Jarl W. Lindeberg (1876-1932) et a été publiée en 1920. Cette démonstration a l'avantage d'être élémentaire et de se généraliser à toutes sortes d'autres contextes. Relevons enfin que c'est George Pólya (1887-1985) qui, en 1920, a qualifié ce théorème de « central ».

Les neuf premiers chapitres de ce livre contiennent les points essentiels développés au court des 300 années de recherche mathématique sur les probabilités qui sont résumées ci-dessus. Les développements que nous allons mentionner maintenant font partie d'ouvrages plus avancés.

Outre les travaux liés au théorème limite central, la théorie des probabilités a connu un essor important pendant tout le vingtième siècle. Mentionnons d'abord les travaux du grand mathématicien et probabiliste américain Norbert Wiener (1894-1964), qui étudie, dans un contexte militaire, comment prévoir les trajectoires d'avions et de missiles. Norbert Wiener donne aussi une construction mathématique du mouvement brownien, dont certaines propriétés avaient été établies dès 1900 par Louis Bachelier (1870-1946), et en 1905 par Albert Einstein (1879-1955). On considère cependant que la théorie moderne des probabilités démarre avec les travaux de Andreï N. Kolmogorov (1903-1987), qui publie en 1933 une axiomatisation de la théorie, qui demeure la base contemporaine. Elle avait été précédée de tentatives infructueuses par d'autres mathématiciens, comme Richard von Mises (1883-1953) en 1919. Outre l'axiomatique, Kolmogorov jette les bases modernes des notions de mesure de probabilité et d'espérance conditionnelle et étudie les processus de Markov.

En France, Paul Lévy (1886-1971), mathématicien pur, généralise les conditions d'application du théorème limite central et caractérise les lois dites « stables », qui jouent un rôle analogue à celui de la loi normale et qu'il étudie à l'aide des fonctions caractéristiques (autre nom de la transformée de Fourier). Il publie en 1937 un livre influent intitulé *Théorie de l'addition des variables aléatoires,* consacré à l'étude du comportement des sommes de variables aléatoires indépendantes. Il améliore divers résultats de Wiener sur le mouvement brownien et entreprend une étude de certains processus de Markov à plusieurs paramètres (champs aléatoires). Le mathématicien russe Alexandre Khintchine (1878-1959) introduit en 1934 la notion de processus stationnaire, qui joue un rôle fondamental aujourd'hui dans la théorie des communications et le traitement du signal. Ces travaux sont complétés par ceux de Herman Wold (1908-1992) en 1938.

Le mathématicien suédois Harald Cramér (1893-1985), motivé par des questions d'assurances, s'intéresse dès 1937 à un autre aspect des théorèmes limites. Il détermine la probabilité qu'une somme de variables aléatoires indépendantes et identiquement distribuées donne un résultat très éloigné de la valeur espérée : on parle de la probabilité d'un « événement rare » ou d'une « grande déviation ». Ce résultat donnera lieu à des travaux importants, que nous mentionnons ci-dessous.

La deuxième moitié du vingtième siècle

Durant cette période, la théorie des probabilités évolue de plus en plus vers l'étude des processus stochastiques, c'est-à-dire des phénomènes aléatoires qui évoluent dans le temps et dans l'espace. Cette étude s'appuie sur les travaux de Wiener, Lévy, Kolmogorov et Khintchine mentionnés ci-dessus et les généralise.

William Feller (1906-1970), d'origine croate mais établi aux Etats-Unis, obtient des versions plus générales du théorème limite central et des résultats très raffinés sur les processus de diffusion. Il publie en deux volumes (1950 et 1970) un livre intitulé *An Introduction to Probability Theory and Its Applications,* qui rassemble les principaux résultats de la théorie et qui est encore régulièrement utilisé aujourd'hui.

Le mathématicien américain Joseph L. Doob (1910-2004) développe la théorie des martingales, la théorie des processus de Markov et la théorie probabiliste du potentiel. Son livre *Stochastic Processes,* publié en 1953, a été l'un des livres de probabilités les plus influents du vingtième siècle. Ces développements utilisent de manière centrale les fondements introduits par Kolmogorov et en particulier, la notion d'espérance conditionnelle, sous une forme plus générale que celle que nous avons présentée dans le chapitre 8.

Eugene B. Dynkin (1924-2014), d'origine russe, travaille dans le prolongement des travaux de Kolmogorov. Il développe une théorie générale des processus de Markov et montre les liens entre les propriétés de ces processus et les solutions de certaines équations aux dérivées partielles. Ces travaux sont présentés dans un traité en deux volumes intitulé *Markov Processes*, publié en 1963.

Dans la continuation des travaux de Cramér, le mathématicien d'origine indienne S. R. Srinivasa Varadhan (né en 1940) formule, à partir de 1966, une théorie unifiée concernant les événements rares, dont l'élément central est un « principe des grandes déviations ». Ce principe très général permet de décrire le comportement qualitatif d'un système stochastique qui dévie de manière importante de son comportement moyen. Dans de nombreux cas, cette description utilise la notion d'entropie relative de deux mesures de probabilité, qui étend la notion d'entropie d'une loi de probabilité discrète, introduite par Claude E. Shannon (1916-2001) en 1948 dans le cadre de la théorie des communications. Ces outils jouent aujourd'hui un rôle important dans la modélisation des communications par Internet et l'étude du comportement de ce réseau. Varadhan reçoit en 2007 le Prix Abel pour ses contributions importantes à la théorie des probabilités, qui font appel à l'analyse non linéaire, aux équations aux dérivées partielles et à l'analyse fonctionnelle.

Le mathématicien français Paul André Meyer (1934-2003), fondateur de « l'Ecole de probabilités de Strasbourg », développe la théorie générale des processus stochastiques à temps continu, en particulier, la théorie des martingales, des processus de Markov et de l'intégrale stochastique. Avec Claude Dellacherie (né en 1943), il entreprend de publier ces développements théoriques dans le traité *Probabilités et Potentiel* en cinq volumes, dont le premier volume paraît en 1975 et le dernier en 1992. Cette théorie difficile et abstraite trouve sa pleine justification notamment à travers les applications qu'elles permettront dès les années 1990, en particulier en mathématiques financières.

Bien que Louis Bachelier avait déjà utilisé le mouvement brownien dans des modèles de marchés financiers en 1900, ce n'est qu'en 1969 que Robert C. Merton (né en 1944) publie un premier article sur l'optimisation des stratégies de placement dans les marchés boursiers, et plusieurs autres articles sur ce thème dans les années qui suivent. En 1973, Fischer Black (1938-1995) et Myron Scholes (né en 1941) publient un article sur le calcul du prix des options dans les marchés boursiers. Ces travaux seront le point de départ de la théorie moderne des marchés financiers, qui se développe avec vigueur dès les années 1980. Cette théorie utilise de manière centrale le calcul stochastique et la théorie des martingales, et donne un élan nouveau aux développements théoriques abstraits qui ont précédé (même la notion élémentaire de tribu, ou σ-algèbre, y retrouve une importance accrue). En 1997, Merton et Scholes reçoivent le Prix Nobel d'économie pour leurs travaux.

Une partie importante de ces avancées de nature « appliquée » s'appuie sur les travaux du mathématicien japonais Kiyoshi Itô (1915-2008). A partir des années 1940, ce mathématicien pur développe la théorie des diffusions aléatoires, de l'intégrale stochastique et des équations différentielles stochastiques. Il démontre le fameux « lemme d'Itô », qui joue dans le calcul stochastique un rôle analogue à celui du théorème fondamental du calcul différentiel et intégral en analyse. Le lemme d'Itô intervient dans la plupart des travaux qui concernent les diffusions aléatoires, et en particulier, dans les calculs requis par la théorie des marchés financiers. Pendant le Congrès International des Mathématiciens à Madrid en 2006, Itô reçoit le prix Gauss, décerné pour la première fois, qui récompense un mathématicien dont les travaux ont eu un impact important dans d'autres disciplines scientifiques : on admet aujourd'hui qu'il y a plus de non mathématiciens qui connaissent le lemme d'Itô que de mathématiciens !

Pendant ce même Congrès International des Mathématiciens de Madrid, le mathématicien français Wendelin Werner (né en 1968) reçoit la première médaille Fields jamais attribuée pour des travaux en théorie des probabilités. Les travaux de Werner s'appuient aussi sur l'analyse complexe et ont conduit à des applications en physique théorique. Deux autres lauréats de la médaille Fields en 2006 ont utilisé la théorie des probabilités dans leurs travaux : Andrei Okounkov (né en 1969) d'origine russe, qui a établit des liens surprenants entre la combinatoire et la géométrie algébrique, en utilisant la théorie des matrices aléatoires, et le mathématicien américain Terence Tao, âgé alors de 31 ans, qui a obtenu notamment des résultats difficiles de théorie des nombres à l'aide de la théorie des probabilités. Deux autres médailles Fields ont été attribuées pour des travaux en théorie des probabilités, à Stanislav Smirnov (né en 1970) en 2010 et à Martin Hairer (né en 1975) en 2014.

A l'aube du vingt-et-unième siècle, la théorie des probabilités est devenue une branche importante des mathématiques, liée de près aux autres thèmes importants des mathématiques (statistique, analyse, analyse numérique, théorie des nombres, géométrie, combinatoire, etc.), dont les résultats sont utilisés dans pratiquement toutes les disciplines scientifiques (physique, biologie, télécommunications, économie, finance, etc). La recherche en théorie des probabilités est promise à un bel avenir.

Sources

H. Cramér. Mathematical Probability and Statistical Inference. *International Statistical Review* 49 (1981), pp. 309-317.

A. Hald. *A History of Probability Theory and Statistics and the Applications Before 1750.* Wiley, New York (1990).

International Mathematical Union. Site `http://www.mathunion.org/`

L. Le Cam. The Central Limit Theorem around 1935. *Statistical Science* 1-1 (1986), pp. 78-96.

I. Todhunter. *A History of the Mathematical Theory of Probability.* Chelsea Publ. Co. (1865).

Wikipedia. Site `http://fr.wikipedia.org/wiki/Accueil`

Chapitre 11

Corrigés des exercices

11.1 Chapitre 1

$\boxed{\textbf{1.1}}$ a) $E \cap F \cap G^c$.

b) $E \cup F \cup G$, ou encore $\left(E^c \cap F^c \cap G^c\right)^c$.

c) $\left(E \cap F \cap G\right)^c$.

d) $(E \cap F^c \cap G^c) \cup (E^c \cap F \cap G^c) \cup (E^c \cap F^c \cap G)$.

$\boxed{\textbf{1.2}}$ a) $E \cap F^c \cap G^c$.

b) $(E \cap F) \cup (E \cap G) \cup (F \cap G)$, ou encore, si l'on veut une réunion d'événements disjoints :

$$(E \cap F \cap G^c) \cup (E \cap F^c \cap G) \cup (E^c \cap F \cap G) \cup (E \cap F \cap G)\,.$$

c) $(E^c \cap F^c) \cup (E^c \cap G^c) \cup (F^c \cap G^c)$, ou encore si l'on veut une réunion d'événements disjoints :

$$(E \cap F^c \cap G^c) \cup (E^c \cap F \cap G^c) \cup (E^c \cap F^c \cap G) \cup (E^c \cap F^c \cap G^c)\,.$$

d) $(E \cap F \cap G^c) \cup (E \cap F^c \cap G) \cup (E^c \cap F \cap G)$.

$\boxed{\textbf{1.3}}$ a) Nous devons vérifier que $\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i$ satisfait les trois propriétés de la définition 1.2. D'abord, l'ensemble Ω appartient à chaque $\mathcal{F}_i$ puisque $\mathcal{F}_i$ est une tribu. Ainsi, $\Omega \in \mathcal{F}$. Deuxièmement, soit $F \in \mathcal{F}$. Alors pour tout $i \in I$, $F \in \mathcal{F}_i$. Comme $\mathcal{F}_i$ est une tribu, $F^c \in \mathcal{F}_i$ pour tout $i \in I$ et donc $F^c \in \mathcal{F}$. Enfin, soit $(F_n)_{n \geqslant 1}$ une suite d'éléments de $\mathcal{F}$. Alors, pour tout $i \in I$, $(F_n)_{n \geqslant 1} \subset \mathcal{F}_i$. Puisque $\mathcal{F}_i$ est une tribu, $\bigcup_{n=1}^{\infty} F_n \in \mathcal{F}_i$, pour tout $i \in I$ et donc $\bigcup_{n=1}^{\infty} F_n \in \mathcal{F}$. Ainsi, $\mathcal{F}$ est bien une tribu.

b) Notons $\mathcal{T}$ l'ensemble de toutes les tribus sur Ω contenant $\mathcal{G}$. Cet ensemble est non vide. En effet, $\mathcal{P}(\Omega)$ est une tribu et contient $\mathcal{G}$. Posons alors $\sigma(\mathcal{G}) = \bigcap_{\mathcal{F} \in \mathcal{T}} \mathcal{F}$. Il s'agit bien d'une tribu par le point a). Cette tribu contient $\mathcal{G}$ puisque chaque tribu $\mathcal{F}$ apparaissant dans cette intersection contient $\mathcal{G}$. De plus, par construction, elle est contenue dans toute tribu contenant $\mathcal{G}$. Ainsi, $\sigma(\mathcal{G})$ est bien la tribu cherchée.

1.4 Soit $\mathcal{F}$ la tribu définie par

$$\mathcal{F} = \left\{ A_1^{i_1} \cup \cdots \cup A_n^{i_n} : i_1, \ldots, i_n \in \{0,1\} \right\},$$

où $A_j^1 = A_j$ et $A_j^0 = \emptyset$. Nous allons montrer que $\mathcal{F} = \sigma(\mathcal{G})$. Montrons tout d'abord que $\mathcal{F}$ est bien une tribu. Premièrement, $\Omega = A_1 \cup \cdots \cup A_n = A_1^1 \cup \cdots \cup A_n^1 \in \mathcal{F}$. Deuxièmement, si $F = A_1^{i_1} \cup \cdots \cup A_n^{i_n} \in \mathcal{F}$, nous remarquons que $F^c = A_1^{1-i_1} \cup \cdots \cup A_n^{1-i_n} \in \mathcal{F}$ (ces deux ensembles sont disjoints et leur union est Ω, puisque $\mathcal{G}$ est une partition de Ω). Finalement, si nous considérons une suite $(F_m = A_1^{i_1^m} \cup \cdots \cup A_n^{i_n^m})_{m \geqslant 1} \subset \mathcal{F}$, nous remarquons que $\bigcup_{m=1}^{\infty} F_m = A_1^{j_1} \cup \cdots \cup A_n^{j_n} \in \mathcal{F}$, où $j_k = \sup_{m \geqslant 1} i_k^m$. Ainsi, $\mathcal{F}$ est bien une tribu.

Vérifions alors que $\mathcal{G} \subset \mathcal{F}$. Soit $A_k \in \mathcal{G}$, $A_k = A_1^{i_1} \cup \cdots \cup A_n^{i_n} \in \mathcal{F}$ avec $i_j = 0$ pour tout $j \neq k$ et $i_k = 1$. Ainsi, $\mathcal{G} \subset \mathcal{F}$.

Il nous reste à vérifier que $\mathcal{F}$ est contenue dans toute tribu contenant $\mathcal{G}$. Ainsi, soit $\mathcal{H}$ une tribu contenant $\mathcal{G}$. La tribu $\mathcal{H}$ contient A_j pour tout $j = 1, \ldots, n$. De plus, $\emptyset = \Omega^c \in \mathcal{H}$. Ainsi, $A_1^{i_1} \cup \cdots \cup A_n^{i_n} \in \mathcal{H}$ pour tout choix de $i_1, \ldots, i_n$. Donc, $\mathcal{F} \subset \mathcal{H}$.

Nous avons alors montré que $\mathcal{F} = \sigma(\mathcal{G})$.

1.5 Soit $F \in \mathcal{F}$. Comme $F \cap G \subset F$, nous avons $P(F \cap G) \leqslant P(F)$. D'autre part, $P(F \cup G) \leqslant 1$, et donc

$$P(F \cap G) = P(F) + P(G) - P(F \cup G) \geqslant P(F) + 1 - 1 = P(F).$$

La double inégalité montre que $P(F \cap G) = P(F)$.

1.6 a) Nous devons montrer que l'événement $\bigcap_{n=1}^{\infty} F_n \in \mathcal{F}$. Pour cela, nous pouvons réécrire cet événement sous la forme suivante :

$$\bigcap_{n=1}^{\infty} F_n = \left(\bigcup_{n=1}^{\infty} F_n^c \right)^c.$$

Comme $F_n \in \mathcal{F}$ pour tout $n \geqslant 1$, nous avons $F_n^c \in \mathcal{F}$ pour tout $n \geqslant 1$. Ainsi, $\bigcup_{n=1}^{\infty} F_n^c \in \mathcal{F}$. Le complémentaire de $\bigcup_{n=1}^{\infty} F_n^c$ est donc également dans $\mathcal{F}$. Ainsi, $\bigcap_{n=1}^{\infty} F_n \in \mathcal{F}$.

b) Posons tout d'abord $F_1 = E_1$, ce qui donne le résultat pour $m = 1$. Posons alors pour $n \geqslant 2$,

$$F_n = E_n \cap \left(\bigcup_{i=1}^{n-1} E_i \right)^c.$$

Le résultat étant vérifié pour $m = 1$, supposons le vrai pour $m - 1$ et montrons par récurrence que celui-ci est vrai pour m :

$$\begin{aligned}\bigcup_{n=1}^{m} F_n = F_m \cup \left(\bigcup_{n=1}^{m-1} F_n\right) &= \left(E_m \cap \left(\bigcup_{i=1}^{m-1} E_i\right)^c\right) \cup \left(\bigcup_{n=1}^{m-1} E_n\right) \\ &= \left(\bigcup_{n=1}^{m} E_n\right) \cap \left(\left(\bigcup_{i=1}^{m-1} E_i\right)^c \cup \left(\bigcup_{n=1}^{m-1} E_n\right)\right) \\ &= \left(\bigcup_{n=1}^{m} E_n\right) \cap \Omega = \bigcup_{n=1}^{m} E_n\,.\end{aligned}$$

De plus, par construction,

$$F_n = E_n \cap \left(\bigcup_{i=1}^{n-1} E_i\right)^c = E_n \cap \left(\bigcup_{i=1}^{n-1} F_i\right)^c = E_n \cap F_1^c \cap \cdots \cap F_{n-1}^c\,.$$

Ainsi, $F_n \cap F_j = \emptyset$, pour tout $j < n$. Par le point a), $F_n = E_n \cap \left(\bigcup_{i=1}^{n-1} E_i\right)^c \in \mathcal{F}$. Ceci établit la première conclusion. La deuxième découle des équivalences

$$\begin{aligned}\omega \in \bigcup_{n=1}^{\infty} E_n &\iff \exists m \geqslant 1 : \omega \in \bigcup_{n=1}^{m} E_n \\ &\iff \exists m \geqslant 1 : \omega \in \bigcup_{n=1}^{m} F_n \iff \omega \in \bigcup_{n=1}^{\infty} F_n\,.\end{aligned}$$

1.11 a) L'ensemble fondamental de cette expérience est

$$\Omega = \big\{(\mathrm{P},\mathrm{P}), (\mathrm{P},\mathrm{F}), (\mathrm{F},\mathrm{P}), (\mathrm{F},\mathrm{F})\big\}\,.$$

b) L'événement G contient les deux résultats de l'expérience où le *deuxième* lancer est pile. Il correspond donc à la propriété « le deuxième lancer tombe pile (mais rien n'est dit sur le résultat du lancer 1) ». Il n'est donc pas observable avant le deuxième lancer.

c) L'événement H contient les deux résultats de l'expérience où le *premier* lancer est pile. Il correspond donc à la propriété « le premier lancer est un pile (mais rien n'est dit sur le résultat du lancer 2) ». Il est donc observable après le premier lancer.
On remarquera qu'une tribu intéressante sur Ω est donc $\mathcal{F} = \{\emptyset, H, H^c, \Omega\}$. C'est la tribu des événements qui sont observables après le premier lancer de la pièce.

1.12 a) Considérons l'ensemble $J = \{\mathrm{N}, \mathrm{B}, \mathrm{R}\}$, où N, B et R représentent respectivement les billes noire, blanche et rouge. L'ensemble J correspond à l'ensemble fondamental lorsque l'on effectue un seul tirage. Ainsi, l'ensemble fondamental correspondant à deux tirages avec remise est

$$\begin{aligned}\Omega &= J \times J \\ &= \big\{(\mathrm{N},\mathrm{N}), (\mathrm{N},\mathrm{B}), (\mathrm{N},\mathrm{R}), (\mathrm{B},\mathrm{N}), (\mathrm{B},\mathrm{B}), (\mathrm{B},\mathrm{R}), (\mathrm{R},\mathrm{N}), (\mathrm{R},\mathrm{B}), (\mathrm{R},\mathrm{R})\big\}\,.\end{aligned}$$

b) Les événements observables après le premier tirage sont les événements pour lesquels on peut décider si un certain $\omega \in \Omega$ y appartient ou non en ne connaissant que le premier résultat. Ainsi, ils ne doivent pas dépendre du second. Il s'agit donc des événements de la tribu suivante :

$$\mathcal{F} = \sigma\big(\{\mathrm{N}\} \times J\,,\ \{\mathrm{B}\} \times J\,,\ \{\mathrm{R}\} \times J\big)\,.$$

Remarquons que les trois événements nécessaires à la définition de $\mathcal{F}$ forment une partition de l'ensemble fondamental Ω. Nous pouvons donc appliquer le résultat de l'exercice 1.4 pour identifier $\mathcal{F}$:

$$\begin{aligned}\mathcal{F} = \big\{\emptyset\,,\ \{\mathrm{N}\} \times J\,,\ \{\mathrm{B}\} \times J\,,\ \{\mathrm{R}\} \times J\,,\ \{\mathrm{N},\mathrm{B}\} \times J\,,& \\ \{\mathrm{N},\mathrm{R}\} \times J\,,\ \{\mathrm{B},\mathrm{R}\} \times J\,,\ \Omega\big\}\,.&\end{aligned}$$

c) Dans le cas où l'on ne remet pas la bille tirée dans l'urne avant le deuxième tirage, l'ensemble fondamental est

$$\Omega = \big\{(\mathrm{N},\mathrm{B}), (\mathrm{N},\mathrm{R}), (\mathrm{B},\mathrm{N}), (\mathrm{B},\mathrm{R}), (\mathrm{R},\mathrm{N}), (\mathrm{R},\mathrm{B})\big\}\,.$$

De la même manière qu'au point b), la tribu des événements observables après le premier tirage est donnée par

$$\mathcal{F} = \sigma\big(\{\mathrm{N}\} \times \{\mathrm{R},\mathrm{B}\}\,,\ \{\mathrm{B}\} \times \{\mathrm{N},\mathrm{R}\}\,,\ \{\mathrm{R}\} \times \{\mathrm{N},\mathrm{B}\}\big)\,.$$

Le résultat de l'exercice 1.4 s'applique à nouveau et nous trouvons

$$\begin{aligned}\mathcal{F} = \Big\{\emptyset\,,\ &\big\{(\mathrm{N},\mathrm{B}), (\mathrm{N},\mathrm{R})\big\}\,,\ \big\{(\mathrm{B},\mathrm{N}), (\mathrm{B},\mathrm{R})\big\}\,,\ \big\{(\mathrm{R},\mathrm{N}), (\mathrm{R},\mathrm{B})\big\}\,,\\ &\big\{(\mathrm{N},\mathrm{B}), (\mathrm{N},\mathrm{R}), (\mathrm{B},\mathrm{N}), (\mathrm{B},\mathrm{R})\big\}\,,\ \big\{(\mathrm{N},\mathrm{B}), (\mathrm{N},\mathrm{R}), (\mathrm{R},\mathrm{N}), (\mathrm{R},\mathrm{B})\big\}\,,\\ &\big\{(\mathrm{B},\mathrm{N}), (\mathrm{B},\mathrm{R}), (\mathrm{R},\mathrm{N}), (\mathrm{R},\mathrm{B})\big\}\,,\ \Omega\Big\}\,.\end{aligned}$$

1.13 a) Considérons l'ensemble $B = \{\mathrm{R},\mathrm{V},\mathrm{N}\}$, où R, V et N représentent respectivement une boule rouge, verte et bleue. L'ensemble B correspond à l'ensemble fondamental lorsque l'on effectue un seul tirage. Ainsi, l'ensemble fondamental correspondant à deux tirages avec remise est

$$\begin{aligned}\Omega_1 &= B \times B\\ &= \big\{(\mathrm{R},\mathrm{R}), (\mathrm{R},\mathrm{V}), (\mathrm{R},\mathrm{N}), (\mathrm{V},\mathrm{R}), (\mathrm{V},\mathrm{V}), (\mathrm{V},\mathrm{N}), (\mathrm{N},\mathrm{R}), (\mathrm{N},\mathrm{V}), (\mathrm{N},\mathrm{N})\big\}\,.\end{aligned}$$

On munit Ω_1 de la tribu $\mathcal{P}(\Omega_1)$. Il s'agit encore de trouver la fonction de probabilité qui correspond à notre notion intuitive de « tirer au hasard ». Posons $\alpha = P\big\{(\mathrm{V},\mathrm{V})\big\}$. Alors

$$P\big\{(x,y)\big\} = \alpha\,,$$

pour $x, y \in \{\mathrm{V},\mathrm{N}\}$, car il n'y a qu'une seule boule noire et une seule boule verte dans la boîte. De plus, pour $x, y \in \{\mathrm{V},\mathrm{N}\}$,

$$P\big\{(\mathrm{R},y)\big\} = P\big\{(y,\mathrm{R})\big\} = 2\alpha\,,$$

car il y a deux boules rouges dans la boîte. Finalement, pour la même raison, pour $y \in \{\mathrm{V}, \mathrm{N}\}$,

$$P\big\{(\mathrm{R}, \mathrm{R})\big\} = 2P\big\{(\mathrm{R}, y)\big\} = 4\alpha\,.$$

Puisqu'il faut que $\sum_{\omega\in\Omega} P\{\omega\} = 1$, nous déduisons que $1 = 4\alpha + 4 \times 2\alpha + 4 \times \alpha = 16\,\alpha$ et donc $\alpha = \frac{1}{16}$. Nous constatons que les singletons ne sont pas équiprobables.

b) Si nous souhaitons construire un ensemble fondamental dans lequel tous les événements élémentaires sont équiprobables, nous devons distinguer chacune des boules présentes dans l'urne au moment du tirage. Ainsi, considérons qu'il y a une boule rouge R_1 et une autre R_2. L'ensemble fondamental pour un seul tirage devient donc $B = \{\mathrm{R}_1, \mathrm{R}_2, \mathrm{N}, \mathrm{V}\}$ et nous avons $\Omega_2 = B \times B$. De plus, nous avons $P\{\omega\} = \frac{1}{16}$ pour tout $\omega \in \Omega_2$.

c) Les deux boules rouges étant indistinguables au sortir de la boîte, les éléments de la forme (R_1, x) et (R_2, x) doivent toujours apparaître par paire dans les événements observables. Il en est de même pour les éléments de la forme (x, R_1) et (x, R_2). Nous pouvons représenter l'ensemble fondamental Ω_2 du point b) comme dans la figure 11.1.

$(\mathrm{R}_1, \mathrm{R}_1)$	$(\mathrm{R}_1, \mathrm{R}_2)$	$(\mathrm{R}_1, \mathrm{V})$	$(\mathrm{R}_1, \mathrm{N})$
$(\mathrm{R}_2, \mathrm{R}_1)$	$(\mathrm{R}_2, \mathrm{R}_2)$	$(\mathrm{R}_2, \mathrm{V})$	$(\mathrm{R}_2, \mathrm{N})$
$(\mathrm{V}, \mathrm{R}_1)$	$(\mathrm{V}, \mathrm{R}_2)$	(V, V)	(V, N)
$(\mathrm{N}, \mathrm{R}_1)$	$(\mathrm{N}, \mathrm{R}_2)$	(N, V)	(N, N)

<table>
<tr><td rowspan="2">(R₁, R₁) (R₁, R₂)
(R₂, R₁) (R₂, R₂)</td><td>(R₁, V)
(R₂, V)</td><td>(R₁, N)
(R₂, N)</td></tr>
<tr></tr>
<tr><td>(V, R₁) (V, R₂)</td><td>(V, V)</td><td>(V, N)</td></tr>
<tr><td>(N, R₁) (N, R₂)</td><td>(N, V)</td><td>(N, N)</td></tr>
</table>

Figure 11.1 A gauche, l'ensemble Ω_1, et à droite, un découpage de cet ensemble en une partition de 9 sous-sensembles.

Nous pouvons découper cet ensemble sous la forme de la partition présentée dans la partie droite de la figure 11.1. En tenant compte du fait que les deux boules rouges sont indistinguables, la tribu des événements observables est la tribu $\mathcal{G}$ engendrée par cette partition (exercice 1.4). Il est à noter que cette tribu n'est pas $\mathcal{P}(\Omega_2)$. On peut remarquer que si nous nous plaçons dans l'interprétation fréquentiste de la probabilité et que nous cherchons à déterminer la fonction de probabilité en répétant l'expérience un grand nombre de fois, alors cette méthode permettra uniquement de déterminer la probabilité des événements de $\mathcal{G}$ et non pas ceux de $\mathcal{P}(\Omega_2)$.

$\boxed{\textbf{1.14}}$ a) Une réalisation de l'expérience peut être représentée par une suite $(x_1, \ldots, x_n)$ de nombres de 1 à 6, de taille n quelconque, dont le dernier terme est $x_n = 6$ et dont chacun des termes précédents vérifie $x_i \neq 6$ pour tout $i < n$. La valeur de x_i représente le résultat du

$i^{\text{ième}}$ lancer. Ajoutons à toutes ces issues possibles un élément Δ correspondant au résultat « le 6 ne sort jamais ». Nous pouvons alors écrire

$$\Omega = \{6\} \cup \left(\bigcup_{n=2}^{\infty} \left(\{1,2,3,4,5\}^{n-1} \times \{6\} \right) \right) \cup \{\Delta\}\,.$$

b) L'événement « A gagne » est réalisé s'il obtient 6 le premier. Ainsi, cet événement se réalise si la suite correspondante se termine après un nombre impair de lancers, c'est-à-dire

$$\big\{\text{A gagne}\big\} = \{6\} \cup \left(\bigcup_{k=1}^{\infty} \left(\{1,2,3,4,5\}^{2k} \times \{6\} \right) \right).$$

De la même manière, le joueur B gagne si le jeu se termine après un nombre pair de lancers, c'est-à-dire

$$\big\{\text{B gagne}\big\} = \bigcup_{k=1}^{\infty} \left(\{1,2,3,4,5\}^{2k-1} \times \{6\} \right)\,.$$

Finalement, aucun des deux ne gagne si le 6 ne sort jamais et donc

$$\big\{\text{ni A ni B ne gagne}\big\} = \{\Delta\}\,.$$

1.15 a) Soit l'ensemble $D = \{1,2,3,4,5,6\}$ correspondant aux résultats du tirage du dé. L'ensemble fondamental correspondant à deux tirages successifs du dé est

$$\Omega = D \times D \times D = D^3 = \big\{(1,1,1),(1,1,2),\ldots,(6,6,5),(6,6,6)\big\}\,.$$

Le premier nombre du couple représente le résultat du premier jet, le deuxième nombre le deuxième jet et le troisième nombre le troisième jet. Comme chacun des singletons de Ω est observable à l'issue de l'expérience, la tribu associée à cette expérience est $\mathcal{F} = \mathcal{P}(\Omega)$. Puisque le dé est équilibré, tous les singletons de Ω sont équiprobables. Comme $|\Omega| = 6^3 = 216$, la probabilité de chaque singleton est $\frac{1}{216}$. La fonction de probabilité est donc déterminée par

$$P\left(\big\{(d_1,d_2,d_3)\big\}\right) = \frac{1}{216}\,,$$

pour tous $d_1, d_2, d_3 \in D$.

b) Notons A l'événement « la somme des résultats est 3 ». Il y a une seule manière d'obtenir une somme de 3 en trois jets de dé, à savoir d'obtenir trois fois 1. D'où

$$P(A) = P\left(\big\{(1,1,1)\big\}\right) = \frac{1}{216}\,.$$

Soit B l'événement « la somme des résultats est 4 ». Pour obtenir une somme de 4 en trois jets de dé, il est nécessaire d'obtenir deux fois 1 et

une fois 2. Il y a donc 3 situations où l'on obtient une somme de 4 et donc
$$P(B) = \frac{3}{216} = \frac{1}{72}\,.$$
Notons finalement C l'événement « la somme des résultats est 5 ». Dénombrons alors les situations où l'on obtient 5 avec trois lancers de dé. Observons qu'il y a deux manières différentes d'obtenir 5 comme somme de trois nombres entiers, à savoir $1+1+3$ ou $1+2+2$. Pour chacun de ces deux cas, il y a trois manières de jeter les dés correspondantes. Ainsi, il y a en tout 6 ($= 2 \cdot 3$) situations équiprobables conduisant à l'obtention d'une somme de 5. D'où
$$P(C) = \frac{6}{216} = \frac{1}{36}\,.$$

11.2 Chapitre 2

$\boxed{\textbf{2.1}}$ a) On peut assimiler la création d'une plaque minéralogique à une expérience qui se réalise en six étapes, à savoir choisir la première lettre, puis la deuxième et ensuite la troisième, puis choisir successivement les trois chiffres. Cela permet d'appliquer le principe de multiplication et de trouver que le nombre de plaques possibles est $26 \times 26 \times 26 \times 10 \times 10 \times 10 = 17\,576\,000$.

b) De la même manière, on vérifie que le nombre de plaques qui commencent par la lettre A est $1 \times 26 \times 26 \times 10^3 = 676\,000$.

$\boxed{\textbf{2.2}}$ Soit Ω l'ensemble des manières de remplir le billet. Alors
$$\sharp\,\Omega = \binom{50}{5}\binom{9}{2} = 76\,275\,360\,.$$
La probabilité de gagner le gros lot est donc $\frac{1}{76\,275\,360} \simeq 1{,}3 \cdot 10^{-8}$.

$\boxed{\textbf{2.3}}$ Puisqu'il s'agit de découper l'ensemble de 52 cartes en une suite de quatre sous-ensembles de cardinalité 13, la réponse est
$$\binom{52}{13\ 13\ 13\ 13} = \frac{52!}{(13!)^4} \simeq 5{,}36 \cdot 10^{28}\,.$$
On remarquera que si chaque joueur a reçu ses treize cartes et si A et B échangent leurs mains, alors cela compte comme une répartition différente.

$\boxed{\textbf{2.4}}$ Comme dans l'exemple 2.7, on constate que
$$\sharp\,\text{troisième prix} = \binom{6}{4}\binom{48}{2}$$
$$P\{\text{troisième prix}\} = \frac{\binom{6}{4}\binom{48}{2}}{\binom{54}{6}} \simeq \frac{1}{1\,526}\,.$$

$\boxed{\mathbf{2.5}}$ Le nombre de suites possibles se calcule comme suit. Il y a 10 choix possibles pour la valeur de la carte la plus élevée (5 à as). Ensuite, il y a 4 choix possibles pour la première carte, 4 choix pour la deuxième carte, 4 choix pour la troisième carte, 4 choix pour la quatrième carte et aussi 4 choix pour la cinquième carte. Il y a donc $10 \cdot 4^5$ suites, y compris les suites royales (où toutes les cartes sont de la même couleur). Les suites royales sont au nombre de $10 \cdot 4$. La probabilité d'une suite est donc

$$\frac{10 \cdot (4^5 - 4)}{\binom{52}{5}} \simeq 0{,}003925\,.$$

$\boxed{\mathbf{2.7}}$ a) Choisissons tout d'abord les 2 hommes devant faire partie du comité. Nous avons $\binom{6}{2}$ manières de faire ce choix. Pour chacun de ces choix, nous devons choisir les femmes faisant partie du comité. Nous avons $\binom{4}{2}$ manières de faire ce choix au total, dont nous devons soustraire le choix des 2 femmes ne souhaitant pas être ensemble. Nous obtenons alors, par le principe de multiplication, $\binom{6}{2} \cdot \left(\binom{4}{2} - 1\right)$ $(= 15 \cdot 5 = 75)$ comités possibles satisfaisant le critère demandé.

b) Nous avons au total $\binom{6}{2}\binom{4}{2}$ comités possibles. Parmi ces comités, il y en a $\binom{5}{1}\binom{3}{1}$ dont le couple marié fait partie. En effet, nous devons choisir le deuxième homme parmi les 5 restants et la deuxième femme parmi les 3 restantes. Ainsi, le nombre de comités où les époux ne siègent pas ensemble est de $\binom{6}{2}\binom{4}{2} - \binom{5}{1}\binom{3}{1}$ $(= 90 - 15 = 75)$.

c) Nous avons $\binom{6}{2}\binom{4}{2}$ comités possibles sans nomination de président. Pour chacun de ceux-ci, il y a 4 présidents possibles. Ainsi, il y a $4 \cdot \binom{6}{2}\binom{4}{2}$ $(= 360)$ comités possibles avec un président.

d) Au plus un des trois membres de la famille peut servir au comité. Il y a donc deux possibilités : soit l'un d'entre eux siège, soit aucun. Supposons que l'une des deux sœurs siège dans ce comité. Nous avons alors $\binom{2}{1}\binom{5}{2}(= 20)$ manières de compléter ce comité (nous ne pouvons prendre ni sa sœur, ni son frère). Si l'autre sœur siège, la situation est parfaitement symétrique. Si le frère est au comité, nous avons $\binom{2}{2}\binom{5}{1}$ $(= 5)$ manières de le compléter (nous ne pouvons prendre aucune de ses sœurs). Enfin, il y a $\binom{2}{2}\binom{5}{2}$ $(= 10)$ comités où aucun membre de cette famille ne siège. Ainsi, il y a en tout $20 + 20 + 5 + 10 = 55$ comités valides.

$\boxed{\mathbf{2.8}}$ a) Il s'agit de dénombrer les permutations possibles de 8 personnes distinctes. Ainsi, il y a $8!$ $(= 40\,320)$ répartitions possibles des 8 personnes sur un rang.

b) Les personnes A et B souhaitent être assises l'une à côté de l'autre. Nous allons considérer que les personnes A et B forment un groupe que nous allons traiter comme s'il s'agissait d'une seule personne. Ceci va nous permettre d'attribuer une place à chaque personne et une place globale au groupe dans le rang, puisque A et B doivent être de toute manière l'un à côté de l'autre. Il y a donc 7 entités (6 personnes et le groupe

AB) à répartir sur un rang. Nous savons qu'il y a 7! manières de le faire. Pour chacune de ces répartitions, A et B peuvent être disposés de deux manières différentes sur les deux sièges réservés. Ainsi, il y a $2 \cdot 7!$ ($= 10\,080$) répartitions possibles des 8 personnes avec A et B ensemble.

c) Tout d'abord, remarquons qu'il y a deux types de répartitions possibles, à savoir

$$\text{f h f h f h f h} \quad \text{ou} \quad \text{h f h f h f h f},$$

où f représente une femme et h un homme. Ainsi, pour chacune des répartitions ci-dessus, il s'agit de commencer par déterminer le nombre de répartitions possibles des 4 femmes sur les sièges qui leur sont attribués. Il y en a 4!, le nombre de permutations de 4 objets. Pour chacune des répartitions possibles des femmes, il y a également 4! manières de répartir les 4 hommes. Ainsi, en appliquant le principe de multiplication, nous constatons qu'il y a $2 \cdot 4! \cdot 4!$ ($= 1\,152$) répartitions des 4 femmes et des 4 hommes en alternance.

d) De la même manière qu'au point b), les 5 hommes forment un groupe que nous allons considérer comme une seule entité. Il y a donc 4 entités (3 personnes restantes et 1 groupe) à répartir, donc 4! manières de le faire. Pour chacune de ces répartitions, les 5 hommes du groupe peuvent se répartir de diverses manières sur les sièges qui leur sont attribués. Le nombre de répartitions possibles des 5 hommes est le nombre de permutations de 5 objets distincts, à savoir 5!. Ainsi, il y a $4! \cdot 5!$ ($= 2\,880$) répartitions possibles des 8 personnes avec les 5 hommes ensemble.

e) Nous considérons que chaque couple marié forme un groupe assimilé à une entité. Il y a donc 4 entités à répartir en rang, donc 4! manières différentes de le faire. Pour chacune de ces répartitions, chacun des 4 couples à deux manières différentes de s'asseoir sur les deux places qui lui correspondent. Ainsi, il y a globalement $4! \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 4! \cdot 2^4$ ($= 384$) répartitions possibles des 4 couples.

2.11 Afin de démontrer que

$$\binom{n+m}{r} = \binom{n}{0}\binom{m}{r} + \binom{n}{1}\binom{m}{r-1} + \cdots + \binom{n}{r}\binom{m}{0}, \tag{11.1}$$

nous allons considérer un ensemble de personnes composé de m femmes et de n hommes. Soit alors $0 \leqslant r \leqslant \min(m, n)$. Il s'agit de compter le nombre de groupes de r personnes que l'on peut composer à partir des $m + n$ que l'on a à disposition. D'une part, ce nombre est égal à $\binom{m+n}{r}$. D'autre part, nous pouvons compter le nombre de groupes de r personnes composés de k femmes (et donc de $r - k$ hommes), avec $k \leqslant r$. Nous en avons $\binom{m}{k}\binom{n}{r-k}$. En additionnant le nombre de tels groupes pour toutes les valeurs possibles de k, nous obtenons le nombre total de groupes de r personnes, à savoir

$$\sum_{k=0}^{r} \binom{n}{k}\binom{m}{r-k} = \binom{n}{0}\binom{m}{r} + \binom{n}{1}\binom{m}{r-1} + \cdots + \binom{n}{r}\binom{m}{0}.$$

Les expressions des deux membres de (11.1) sont solutions d'un même problème de dénombrement et sont donc égales. Le résultat est démontré.

2.14 L'espace de probabilité est fini, donc la probabilité d'un ensemble $F \in \mathcal{P}(\Omega)$ est $P(F) = \sum_{\omega \in F} P\{\omega\}$. Considérons alors un singleton $\{\omega\}$, avec $\omega \in \Omega$ et regardons son influence sur la formule à démontrer.

Tout d'abord, nous constatons que si $\omega \notin F_1 \cup \cdots \cup F_n$, il ne contribue ni à la probabilité du membre de gauche, ni à celles du membre de droite, car il n'appartient à aucun des ensembles considérés.

Soit alors $\omega \in F_1 \cup \cdots \cup F_n$ et supposons que ω appartienne à exactement m ensembles F_i parmi les n possibles. La probabilité $P\{\omega\}$ apparaît alors une fois dans le membre de gauche de l'égalité à démontrer. De plus, ω appartient à $\binom{m}{k}$ ensembles de type $F_{i_1} \cap \cdots \cap F_{i_k}$, c'est-à-dire le nombre de possibilités de choisir k ensembles de type F_i parmi les m auxquels ω appartient. Ainsi, la probabilité $P\{\omega\}$ apparaît $\binom{m}{1} - \binom{m}{2} + \cdots + (-1)^{m+1}\binom{m}{m}$ fois dans le membre de droite de l'équation à démontrer.

Pour que l'équation soit vérifiée, il suffit donc de montrer que la probabilité de chaque singleton $\{\omega\}$ contribue de la même manière dans le membre de gauche que dans celui de droite. C'est-à-dire que, pour tout $0 < m \leqslant n$,

$$1 = \binom{m}{1} - \binom{m}{2} + \cdots + (-1)^{m+1}\binom{m}{m}. \tag{11.2}$$

Or

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

pour tout $0 < m \leqslant n$, ce qui vérifie l'équation (11.2) et donc le résultat cherché.

Notons qu'il est également possible de démontrer la formule d'inclusion-exclusion par récurrence sur n. L'exercice 6.15 fournit une troisième démonstration de cette formule.

2.15 Supposons que les quatre couleurs du jeu de 52 cartes sont numérotées de 1 à 4. Notons alors E_i l'événement « la $i^{\text{ième}}$ couleur manque dans la main de 13 cartes » $(i = 1, \ldots, 4)$. Nous cherchons $P(E_1 \cup E_2 \cup E_3 \cup E_4)$. Par la formule d'inclusion-exclusion, nous avons

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3 \cup E_4) \\ &= \sum_{i=1}^{4} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) \\ &\quad + \sum_{i_1 < i_2 < i_3} P(E_{i_1} \cap E_{i_2} \cap E_{i_3}) - P(E_1 \cap E_2 \cap E_3 \cap E_4) \\ &= 4P(E_1) - \binom{4}{2} P(E_1 \cap E_2) \\ &\quad + \binom{4}{3} P(E_1 \cap E_2 \cap E_3) - P(E_1 \cap E_2 \cap E_3 \cap E_4)\,, \end{aligned} \tag{11.3}$$

car les événements dans chacune des sommes sont équiprobables. Or

$$P(E_1) = \frac{\binom{39}{13}}{\binom{52}{13}},$$

car les 13 cartes doivent être choisies parmi les 39 cartes qui ne sont pas de la couleur 1. De même,

$$P(E_1 \cap E_2) = \frac{\binom{26}{13}}{\binom{52}{13}},$$

car les 13 cartes doivent être choisies parmi les 26 qui ne sont ni de la couleur 1, ni de la couleur 2. Finalement,

$$P(E_1 \cap E_2 \cap E_3) = \frac{1}{\binom{52}{13}},$$

car toutes les cartes doivent être de la couleur 4 et $P(E_1 \cap E_2 \cap E_3 \cap E_4) = 0$, car il n'est pas possible que toutes les couleurs manquent. En remplaçant dans (11.3), nous obtenons $P(E_1 \cup E_2 \cup E_3 \cup E_4) \simeq 0{,}0511$.

$\boxed{\textbf{2.16}}$ Nous allons démontrer le résultat par récurrence sur n. Dans le cas où $n = 0$, nous avons, pour tout $\ell \geqslant 1$,

$$\binom{\ell}{0} = 1 = \binom{\ell - 1}{0} = \sum_{i=0}^{0} \binom{i + \ell - 1}{i}.$$

Soit $n \geqslant 0$ et supposons que le résultat soit vérifié pour n et pour tout $\ell \geqslant 1$. Vérifions le pour $n + 1$. Nous avons, par la proposition 2.6b) et par hypothèse de récurrence,

$$\begin{aligned}\binom{n+1+\ell}{n+1} &= \binom{n+\ell}{n+1} + \binom{n+\ell}{n} \\ &= \binom{n+\ell}{n+1} + \sum_{i=0}^{n} \binom{i+\ell-1}{i} = \sum_{i=0}^{n+1} \binom{i+\ell-1}{i}.\end{aligned}$$

Le résultat est démontré.

2.17 Le cardinal $|\mathcal{P}(\Omega)|$ de l'ensemble des parties de Ω correspond au nombre total de sous-ensembles de Ω. Il s'agit ainsi de la somme du nombre de sous-ensembles à k éléments ($k \leqslant n$) sur toutes les valeurs possibles de k. Nous savons que le nombre de sous-ensembles à k éléments d'un ensemble à n éléments est $\binom{n}{k}$. Ainsi,

$$|\mathcal{P}(\Omega)| = \sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \binom{n}{k} 1^k 1^{n-k} = (1+1)^n = 2^n ,$$

et le résultat est démontré.

Nous pouvons également calculer le nombre de sous-ensembles de Ω en identifiant $A \subset \Omega$ à un vecteur v_A de taille n dont les composantes peuvent prendre la valeur 0 ou 1. Chaque composante correspond à un élément $\omega \in \Omega$. Dans v_A, une composante prend la valeur 1 si $\omega \in A$ et 0 sinon. Le nombre de sous-ensembles de Ω est donc égal au nombre de vecteurs de taille n à valeurs dans $\{0, 1\}$. Celui-ci est égal à 2^n. En effet, il y a deux choix pour chacune des n composantes.

2.18 a) Admettons que les boules sont numérotées de 1 à n et les urnes de 1 à k. Considérons la boule numéro 1. Nous avons k possibilités pour la placer dans une urne. Ceci est valable pour chacune des n boules. Il y a donc globalement k^n possibilités de répartir les n boules dans les k urnes.

b) Les boules étant ici indistinguables, mais les urnes restant distinguables, c'est le nombre de boules que nous plaçons dans chaque urne qui devient déterminant. Notons x_i le nombre de boules que nous plaçons dans l'urne numéro i. Nous devons déterminer le nombre de solutions entières positives (c.-à-d. avec $x_i \in \mathbb{N}^*$) de l'équation

$$x_1 + \cdots + x_k = n .$$

En effet, il y a k urnes, le nombre total de boules est n et nous devons en placer au moins une dans chaque urne. Pour déterminer ce nombre de répartitions possibles, imaginons que les n boules sont disposées l'une à côté de l'autre en rang. Nous devons découper ce rang en k groupes non vides. Ceci revient à distribuer $k-1$ séparations dans les espaces entre les boules, sans en placer deux au même endroit. Dans l'ensemble des $n-1$ espaces disponibles, nous devons choisir un sous-ensemble de cardinalité $k-1$ où nous déposerons les séparations. Il y a donc $\binom{n-1}{k-1}$ manières de répartir les n boules indistinguables dans les k urnes distinguables, sans qu'aucune urne ne reste vide.

c) Lorsque nous n'avons plus de restrictions concernant le nombre d'urnes vides, il s'agit de déterminer le nombre de solutions entières non négatives (c.-à-d. avec $x_i \in \mathbb{N}$) de l'équation

$$x_1 + \cdots + x_k = n .$$

En ajoutant 1 à chacune des composantes du membre de gauche (on pose $y_i = x_i + 1$), cela revient à déterminer le nombre de solutions entières positives de l'équation

$$y_1 + \cdots + y_k = n + k\,,$$

qui est égal, par le point b), à $\binom{n+k-1}{k-1} = \binom{n+k-1}{n}$.

2.19 a) Supposons que nous avons à disposition n boules indistinguables (correspondant aux n unités) que nous devons disposer dans k urnes numérotées (correspondant aux variables $x_1, \ldots, x_k$). Le nombre de solutions de l'équation est identique au nombre de répartitions des n boules dans les k urnes, sachant qu'aucune urne ne peut être vide. Ainsi, la réponse est la même que celle de l'exercice 2.18b) : il y a $\binom{n-1}{k-1}$ solutions positives à notre équation.

b) En ajoutant 1 à chacune des composantes du membre de gauche (on pose $y_i = x_i + 1$), cela revient à déterminer le nombre de solutions entières positives de l'équation

$$y_1 + \cdots + y_k = n + k\,,$$

qui est égal, par le point a), à $\binom{n+k-1}{k-1} = \binom{n+k-1}{n}$. Demander des solutions non négatives en lieu et place de solutions positives revient à ne plus imposer de restrictions concernant le nombre d'urnes vides dans la modélisation du point a) et la réponse est la même qu'à l'exercice 2.18c).

c) Supposons tout d'abord que les r inconnues nulles sont $x_{k-r+1}, \ldots, x_k$. Le problème revient à trouver le nombre de solutions positives de l'équation $x_1 + \cdots + x_{k-r} = n$. Par le point a), nous en avons $\binom{n-1}{k-r-1}$. Ce nombre de solutions reste le même pour chacun des $\binom{k}{r}$ choix des r inconnues nulles parmi les k inconnues possibles. Ainsi, nous avons $\binom{n-1}{k-r-1}\binom{k}{r}$ solutions non négatives à l'équation avec exactement r inconnues nulles.

2.23 a) Pour chaque chiffre, sauf le zéro, nous pouvons construire un nombre à 4 chiffres. Il y a donc 9 possibilités.

b) Deux cas peuvent se présenter. Premièrement, le nombre de possibilités de construire un nombre à 4 chiffres contenant 2 paires de deux chiffres différents de 0 est de $\binom{9}{2}\binom{4}{2}$. En effet, nous choisissons deux chiffres parmi 9, puis deux places parmi les quatre possibles pour placer le premier des deux chiffres. Deuxièmement, si nous avons une des deux paires formée de 0, nous avons $9 \cdot \binom{3}{2}$ possibilités. En effet, il y a 9 manières de choisir la seconde paire et deux places parmi trois seulement pour le 0 (si le zéro est placé en tête, ce n'est plus un nombre à 4 chiffres). Finalement, il y a donc $\binom{9}{2}\binom{4}{2} + 9 \cdot \binom{3}{2} = 243$ nombres de 4 chiffres formés de 2 paires distinctes.

c) De façon à former un nombre à 4 chiffres distincts, nous avons tout d'abord 9 choix pour le chiffre initial. En effet, celui-ci ne peut pas être 0. Ensuite, pour chacun de ces choix, il y a 9 choix pour le deuxième chiffre

(8 chiffres non nuls restants et 0). Puis, 8 choix pour le troisième chiffre et 7 choix pour le quatrième. Ainsi, par le principe de multiplication, il y a $9 \cdot 9 \cdot 8 \cdot 7 = 4536$ nombres de 4 chiffres distincts.

d) Dans le cas où les nombres doivent apparaître dans l'ordre croissant, il y a $\binom{9}{4}$ possibilités de choisir 4 chiffres distincts (0 non compris), puis une seule manière de les ordonner. Le zéro ne peut effectivement pas apparaître en première place dans ces conditions. Ainsi, il y a $\binom{9}{4}$ nombres de 4 chiffres distincts ordonnés.

e) Au point a), il y a 9 possibilités, quelle que soit la valeur de n. Au point c), pour $1 \leqslant n \leqslant 10$, en suivant un raisonnement analogue au cas $n = 4$, nous avons 9 choix pour le premier chiffre, puis 9 choix pour le deuxième, 8 choix pour le troisième, et ainsi de suite, jusqu'à $(10 - n + 1)$ choix pour le n-ème chiffre. Ainsi, nous aurons $\frac{9 \cdot 9!}{(10-n)!}$ nombres de n chiffres distincts. Pour $n > 10$, il n'y en a aucun. Au point d), en suivant un raisonnement analogue à celui du cas $n = 4$, nous remarquons qu'il y aura $\binom{9}{n}$ nombres de n chiffres distincts ordonnés pour $1 \leqslant n \leqslant 9$. Il n'y en a aucun pour $n \geqslant 10$.

2.24 On procède par récurrence. Le cas $n = 1$ correspond à la formule usuelle pour la dérivée d'un produit, *i.e.* $(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x)$. On suppose maintenant que la formule est vraie pour un $n \geqslant 1$. Alors, par hypothèse de récurrence et utilisant la linéarité de la dérivée et la formule de dérivée d'un produit, nous avons

$$\begin{aligned}
(f \cdot g)^{(n+1)}(x) &= \frac{d}{dx}(f \cdot g)^{(n)}(x) \\
&= \frac{d}{dx}\left(\sum_{k=0}^{n} \binom{n}{k} f^{(n-k)}(x) g^{(k)}(x)\right) \\
&= \sum_{k=0}^{n} \binom{n}{k} \frac{d}{dx}\left(f^{(n-k)}(x) g^{(k)}(x)\right) \\
&= \sum_{k=0}^{n} \binom{n}{k} \left(f^{(n-k+1)}(x) g^{(k)}(x) + f^{(n-k)}(x) g^{(k+1)}(x)\right) \\
&= \sum_{k=0}^{n} \binom{n}{k} f^{(n-k+1)}(x) g^{(k)}(x) + \sum_{k=0}^{n} \binom{n}{k} f^{(n-k)}(x) g^{(k+1)}(x) \\
&= \sum_{k=0}^{n} \binom{n}{k} f^{(n-k+1)}(x) g^{(k)}(x) + \sum_{k=1}^{n+1} \binom{n}{k-1} f^{(n-k+1)}(x) g^{(k)}(x) \\
&= f^{(n+1)}(x) g(x) + \sum_{k=1}^{n} \left(\binom{n}{k} + \binom{n}{k-1}\right) f^{(n-k+1)}(x) g^{(k)}(x) \\
&\qquad + f(x) g^{(n+1)}(x) \\
&= \sum_{k=0}^{n+1} \binom{n+1}{k} f^{(n+1-k)}(x) g^{(k)}(x),
\end{aligned}$$

par la proposition 2.6 b). Ainsi, la formule est vraie pour $n+1$ et le résultat est démontré.

11.3 Chapitre 3

$\boxed{\textbf{3.1}}$ Notons A l'événement « le signataire a un accident durant l'année » et notons R l'événement « le signataire fait partie de la classe à haut risque ».

a) Nous voulons calculer $P(A)$. D'après la formule des probabilités totales,

$$P(A) = P(A \mid R)P(R) + P(A \mid R^c)P(R^c) = (0{,}4 \times 0{,}3) + (0{,}2 \times 0{,}7) = 0{,}26\,,$$

car R^c représente l'événement « le signataire fait partie de la classe à risque modéré ».

b) Nous devons calculer $P(R \mid A)$. Par la définition de probabilité conditionnelle (ou la formule de Bayes),

$$P(R \mid A) = \frac{P(A \mid R)P(R)}{P(A)} = \frac{0{,}4 \times 0{,}3}{0{,}26} \simeq 0{,}46\,,$$

par le résultat du point a).

Remarque* Dans cette résolution, on n'a pas précisé l'espace de probabilité, ce qui peut se faire comme suit.

Soit Ω_1 l'ensemble des individus que l'assurance a pris en compte. On assimile l'arrivée d'un nouvel assuré (signataire) au choix d'un élément $\omega_1 \in \Omega_1$. Soit $\Omega_2 = \{a, n\}$ où a signifie « avoir un accident dans l'année » et n siginifie « ne pas avoir d'accident dans l'année. » L'ensemble fondamental est alors $\Omega = \Omega_1 \times \Omega_2$.

Soit $R_1 \subset \Omega_1$ l'ensemble des individus à haut risque. Alors l'événement R ci-dessus est $R = R_1 \times \Omega_2$ et l'événement A ci-dessus est $A = \Omega_1 \times \{a\}$. Les données de l'énoncé du problème déterminent la fonction de probabilité P sur la tribu $\mathcal{F}$ engendrée par la partition de Ω en les quatre sous-ensembles

$$R \cap A = R_1 \times \{a\}, \qquad R \cap A^c = R_1 \times \{n\},$$

$$R^c \cap A = R_1^c \times \{a\}, \qquad R^c \cap A^c = R_1^c \times \{n\}.$$

L'énoncé sous-entend que $P(\{\omega_1\} \times \Omega_2) = \frac{1}{\sharp\Omega_1}$. Il n'impose pas que pour tout $\omega_1 \in R_1$, $P(\Omega_1 \times \{a\} \mid \{\omega_1\} \times \Omega_2) = 0,4$, mais uniquement que $P(A \mid R) = 0,4$.

3.2 Notons R_i l'événement « le récepteur reçoit un i » et E_i l'événement « l'émetteur émet un i » ($i \in \{0,1\}$).

a) Nous voulons calculer $P(R_0)$. D'après la formule des probabilités totales,

$$\begin{aligned} P(R_0) &= P(R_0 \mid E_0)P(E_0) + P(R_0 \mid E_1)P(E_1) \\ &= (0{,}8 \times 0{,}45) + (0{,}1 \times 0{,}55) = 0{,}415\,, \end{aligned}$$

puisque $E_0 = E_1^c$.

b) Nous devons calculer $P(E_0 \mid R_0)$. Par la définition de probabilité conditionnelle (ou la formule de Bayes),

$$P(E_0 \mid R_0) = \frac{P(R_0 \mid E_0)P(E_0)}{P(R_0)} = \frac{0{,}8 \times 0{,}45}{0{,}415} \simeq 0{,}867\,,$$

par le résultat du point a).

3.3 Soit G_i l'événement « la $i^{\text{ième}}$ boule tirée est rouge ». Il s'agit de calculer

$$\begin{aligned} P(G_1 \cap G_2 \cap G_3) &= P(G_1)P(G_2 \mid G_1)P(G_3 \mid G_1 \cap G_2) \\ &= \frac{6}{11} \cdot \frac{5}{10} \cdot \frac{4}{9} = \frac{4}{33} \simeq 0,12\,. \end{aligned}$$

On peut remarquer que cette probabilité est égale à

$$\frac{A_{6,3}}{A_{11,3}}\,.$$

3.4 Notons F_i l'événement « la $i^{\text{ième}}$ main contient exactement un as » pour $i = 1, \ldots, 4$. Nous cherchons $P(F_1 \cap F_2 \cap F_3 \cap F_4)$. Par la règle de multiplication des probabilités conditionnelles, on a

$$P(F_1 \cap F_2 \cap F_3 \cap F_4) = P(F_4 \mid F_1 \cap F_2 \cap F_3)P(F_3 \mid F_1 \cap F_2)P(F_2 \mid F_1)P(F_1)\,. \tag{11.4}$$

Or

$$P(F_1) = \frac{\binom{4}{1}\binom{48}{12}}{\binom{52}{13}}\,.$$

En effet, il s'agit de choisir un as parmi les 4 possibles et les douze autres cartes parmi les 48 restantes. De la même manière,

$$P(F_2 \mid F_1) = \frac{\binom{3}{1}\binom{36}{12}}{\binom{39}{13}}\,.$$

En effet, pour chaque répartition de la première main contenant un as, il nous reste effectivement 39 cartes à distribuer parmi lesquelles 3 as. De manière analogue,

$$P(F_3 \mid F_1 \cap F_2) = \frac{\binom{2}{1}\binom{24}{12}}{\binom{26}{13}}.$$

De plus, $P(F_4 \mid F_1 \cap F_2 \cap F_3) = 1$. En effet, si les trois premières mains ont chacune exactement un as, alors la dernière en a obligatoirement un aussi. Par (11.4), nous obtenons alors

$$\begin{aligned} P(F_1 \cap F_2 \cap F_3 \cap F_4) &= \frac{\binom{4}{1}\binom{48}{12}}{\binom{52}{13}} \cdot \frac{\binom{3}{1}\binom{36}{12}}{\binom{39}{13}} \cdot \frac{\binom{2}{1}\binom{24}{12}}{\binom{26}{13}} \cdot 1 \\ &= \frac{13^3}{17 \cdot 25 \cdot 49} \simeq 0{,}105\,. \end{aligned}$$

3.5 Soit A_j l'événement « la $j^{\text{ième}}$ carte est un pique » pour $j = 1, 2$. Nous souhaitons calculer $P(A_1 \mid A_2)$. Nous avons $P(A_1) = \frac{13}{52} = \frac{1}{4}$. Si la première carte tirée est un pique, il reste 12 piques sur 51 cartes. Par contre, si la première carte n'est pas un pique, il en reste 13 sur 51 cartes. Ainsi, $P(A_2 \mid A_1) = \frac{12}{51} = \frac{4}{17}$ et $P(A_2 \mid A_1^c) = \frac{13}{51}$. D'après la formule de Bayes,

$$\begin{aligned} P(A_1 \mid A_2) &= \frac{P(A_2 \mid A_1)P(A_1)}{P(A_2 \mid A_1)P(A_1) + P(A_2 \mid A_1^c)P(A_1^c)} \\ &= \frac{\frac{4}{17} \cdot \frac{1}{4}}{\frac{4}{17} \cdot \frac{1}{4} + \frac{13}{51} \cdot \frac{3}{4}} = \frac{4}{17} \simeq 0{,}24\,. \end{aligned}$$

Remarquons que $P(A_1 \mid A_2) = P(A_2 \mid A_1)$, bien que les tirages soient effectués *sans remise.*

3.6 Soit R_j l'événement « la $j^{\text{ième}}$ boule est rouge » et soit N_j l'événement « la $j^{\text{ième}}$ boule est noire » ($j = 1, 2$). Nous souhaitons calculer $P(N_1 \mid R_2)$. Nous avons $P(N_1) = \frac{4}{12} = \frac{1}{3}$. Si la première boule tirée est rouge, il reste 7 rouges sur 11 boules. Par contre, si la première boule est noire, il en reste 8 sur 11 boules. Ainsi, nous avons $P(R_2 \mid R_1) = \frac{7}{11}$ et $P(R_2 \mid N_1) = \frac{8}{11}$. D'après la formule de Bayes,

$$\begin{aligned} P(N_1 \mid R_2) &= \frac{P(R_2 \mid N_1)P(N_1)}{P(R_2 \mid N_1)P(N_1) + P(R_2 \mid R_1)P(R_1)} \\ &= \frac{\frac{8}{11} \cdot \frac{1}{3}}{\frac{8}{11} \cdot \frac{1}{3} + \frac{7}{11} \cdot \frac{2}{3}} = \frac{4}{11} \simeq 0{,}36\,. \end{aligned}$$

3.7 Notons F_i l'événement « la $i^{\text{ième}}$ urne contient exactement une boule rouge » pour $i = 1, 2, 3$. Nous cherchons $P(F_1 \cap F_2 \cap F_3)$. Par la règle de multiplication des probabilités conditionnelles, nous avons

$$P(F_1 \cap F_2 \cap F_3) = P(F_3 \mid F_1 \cap F_2)P(F_2 \mid F_1)P(F_1)\,. \tag{11.5}$$

Or nous avons

$$P(F_1) = \frac{\binom{3}{1}\binom{12}{4}}{\binom{15}{5}}\,.$$

En effet, il s'agit de choisir une boule rouge parmi les 3 possibles et les quatre autres boules parmi les 12 restantes. Nous avons également

$$P(F_2 \mid F_1) = \frac{\binom{2}{1}\binom{8}{4}}{\binom{10}{5}}\,.$$

En effet, pour chaque répartition des boules pour laquelle la première urne contient exactement une rouge, il nous reste 10 boules à distribuer parmi lesquelles 2 rouges. Il s'agit donc de choisir une boule rouge parmi les 2 restantes et 4 boules noires parmi les 8 restantes. Finalement, $P(F_3 \mid F_1 \cap F_2) = 1$. En effet, si les deux premières urnes ont chacune exactement une boule rouge, la dernière en a obligatoirement une aussi. Par (11.5), nous obtenons

$$P(F_1 \cap F_2 \cap F_3) = 1 \cdot \frac{\binom{2}{1}\binom{8}{4}}{\binom{10}{5}} \cdot \frac{\binom{3}{1}\binom{12}{4}}{\binom{15}{5}} = \frac{5^2}{7 \cdot 13} = \frac{25}{91} \simeq 0{,}27\,.$$

3.8 D'après la formule de Bayes,

$$P(F \mid G) = \frac{P(G \mid F)P(F)}{P(G \mid F)P(F) + P(G \mid F^c)P(F^c)}\,.$$

Ainsi,

$$\begin{aligned} P(F \mid G) > P(F) &\iff \frac{P(G \mid F)P(F)}{P(G \mid F)P(F) + P(G \mid F^c)P(F^c)} > P(F) \\ &\iff P(G \mid F) > P(G \mid F)P(F) + P(G \mid F^c)\big(1 - P(F)\big) \\ &\iff P(G \mid F)\big(1 - P(F)\big) > P(G \mid F^c)\big(1 - P(F)\big) \\ &\iff P(G \mid F) > P(G \mid F^c)\,, \end{aligned}$$

et le résultat est démontré.

$\boxed{\textbf{3.9}}$ Considérons qu'au début du jeu, nous avons un gros lot et 20 lots de moindre importance répartis aléatoirement dans les boîtes. Nous pouvons donc considérer comme espace de probabilité $\Omega = \{1, \dots, 21\}$, où $\omega \in \Omega$ représente le fait que le gros lot se trouve dans la boîte numéro ω. Nous munissons Ω de la tribu $\mathcal{F} = \mathcal{P}(\Omega)$ et de la fonction de probabilité donnée par $P\{\omega\} = \frac{1}{21}$ pour tout $\omega \in \Omega$. Sans limiter la généralité, nous pouvons supposer que le joueur choisit la boîte numéro 1. Toujours sans limiter la généralité, nous pouvons supposer qu'il fait ouvrir les boîtes au présentateur dans l'ordre numérique. Nous savons alors que le gros lot ne se trouve pas dans les boîtes 2 à 20. Autrement dit, nous savons que l'événement $\{1, 21\}$ se réalise. Calculons alors la probabilité que le gros lot se trouve dans la boîte du candidat sous cette condition. Nous voulons donc calculer $P(\{1\} \mid \{1, 21\})$ et nous avons

$$P\big(\{1\} \mid \{1,21\}\big) = \frac{P\big(\{1\} \cap \{1,21\}\big)}{P\{1,21\}} = \frac{P\{1\}}{P\{1,21\}} = \frac{\frac{1}{21}}{\frac{2}{21}} = \frac{1}{2}\,.$$

De même, $P(\{21\} \mid \{1, 21\}) = \frac{1}{2}$. Ainsi, le joueur n'a pas plus intérêt à changer de boîte qu'à conserver la sienne : les deux choix sont indifférents.

$\boxed{\textbf{3.10}}$ Nous allons étudier la probabilité que le candidat gagne la voiture en fonction de la stratégie utilisée. Un raisonnement intuitif pourrait être de dire qu'après que le présentateur a ouvert l'une des portes, la voiture a autant de chances de se trouver derrière l'une ou l'autre des portes restantes et que le candidat a une probabilité $\frac{1}{2}$ de gagner la voiture quel que soit son choix. Ce raisonnement n'est pas correct !

Définissons un modèle mathématique pour ce problème. Le joueur désigne une porte. On attribue alors le numéro I à la porte désignée par le candidat et parmi les deux portes restantes, le numéro II à la porte de gauche et le numéro III à celle de droite. La numérotation est ainsi indépendante de la porte désignée par le candidat.

A partir de là, les observations possibles, une fois le jeu terminé, sont présentées dans la figure 11.2.

Situation	Porte I	Porte II	Porte III	Le présentateur ouvre
ω_1	Chèvre	Chèvre	Voiture	la porte II
ω_2	Chèvre	Voiture	Chèvre	la porte III
ω_3	Voiture	Chèvre	Chèvre	la porte II
ω_4	Voiture	Chèvre	Chèvre	la porte III

Figure 11.2 Observations possibles à la fin du jeu de Monty-Hall.

Ainsi, l'ensemble fondamental est $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, les ω_i ayant le sens indiqué dans la figure 11.2. Une des difficultés du problème est que $P\{\omega_3\}$ et $P\{\omega_4\}$ ne sont pas des données du problème. En revanche, $P\{\omega_1\} = P\{\omega_2\} = P\{\omega_3, \omega_4\} = \frac{1}{3}$ reflète bien la description du problème. On considère en effet que le candidat choisit une porte sans information sur la position de la voiture.

Soit $\mathcal{G} = \sigma(\{\omega_1\}, \{\omega_2\}, \{\omega_3, \omega_4\})$. La description du problème précise donc la fonction de probabilité P sur la tribu $\mathcal{G}$ et non pas sur $\mathcal{P}(\Omega)$, : ceci est un exemple d'un espace de probabilité fini pour lequel la fonction de probabilité n'est pas définie sur $\mathcal{P}(\Omega)$.

Remarquons que l'événement « le présentateur désigne la porte II » est $\{\omega_1, \omega_3\}$ et que l'événement « le présentateur désigne la porte III » est $\{\omega_2, \omega_4\}$. D'autre part, l'événement « le joueur gagne la voiture en changeant de porte » est $\{\omega_1, \omega_2\}$. Puisque $\{\omega_1, \omega_3\} \notin \mathcal{G}$, nous allons prolonger la définition de P à $\mathcal{P}(\Omega)$.

Soit $0 \leqslant \alpha \leqslant 1$. Puisque $P\{\omega_3\} + P\{\omega_4\} = P\{\omega_3, \omega_4\} = \frac{1}{3}$, posons $P\{\omega_3\} = \frac{\alpha}{3}$ et $P\{\omega_4\} = \frac{1-\alpha}{3}$. Ainsi, $P\big(\{\omega_3\} \mid \{\omega_3, \omega_4\}\big) = \alpha$ et $P\big(\{\omega_4\} \mid \{\omega_3, \omega_4\}\big) = 1 - \alpha$. La valeur α représente donc la probabilité que le présentateur choisisse la porte II lorsqu'il a les deux possibilités qui lui sont offertes.

Si le présentateur désigne la porte II (resp. III), on s'intéresse à $P\big(\{\omega_1, \omega_2\} \mid \{\omega_1, \omega_3\}\big)$, qui est la probabilité de gagner la voiture si on change de porte, sachant que le présentateur a désigné II (resp. à $P\big(\{\omega_1, \omega_2\} \mid \{\omega_2, \omega_4\}\big)$). Or

$$\begin{aligned} P\big(\{\omega_1, \omega_2\} \mid \{\omega_1, \omega_3\}\big) &= \frac{P\{\omega_1\}}{P\{\omega_1, \omega_3\}} = \frac{P\{\omega_1\}}{P\{\omega_1\} + P\{\omega_3\}} \\ &= \frac{P\{\omega_1\}}{P\{\omega_1\} + P\big(\{\omega_3\} \mid \{\omega_3, \omega_4\}\big) P\{\omega_3, \omega_4\}} \\ &= \frac{\dfrac{1}{3}}{\dfrac{1}{3} + \alpha \cdot \dfrac{1}{3}} = \frac{1}{\alpha + 1}. \end{aligned}$$

On obtient de la même manière $P\big(\{\omega_1, \omega_2\} \mid \{\omega_2, \omega_4\}\big) = \frac{1}{2-\alpha}$. Remarquons que $\frac{1}{\alpha+1} \geqslant \frac{1}{2}$ et que $\frac{1}{2-\alpha} \geqslant \frac{1}{2}$ pour tout $0 \leqslant \alpha \leqslant 1$. Ainsi, quelle que soit la porte désignée par le présentateur, la probabilité conditionnelle de gagner en changeant de porte est toujours supérieure ou égale à $\frac{1}{2}$ qui est toujours supérieur ou égal à la probabilité conditionnelle de gagner sans changer de porte.

Dans le cas symétrique ($\alpha = \frac{1}{2}$), la probabilité de gagner en changeant de porte est $\frac{2}{3}$. Dans le cas extrême où $\alpha = 1$, il n'y a pas d'avantage à changer de porte si le présentateur désigne la porte II, mais il faudra absolument changer s'il désigne la porte III.

Remarque. Une raison plus intuitive qui justifie de changer de porte est la suivante. Dans tous les cas, on préférerait pouvoir choisir d'ouvrir les deux portes autres que celle désignée, plutôt que de garder sa porte. Comme le présentateur vous montre une chèvre derrière l'une de ces deux portes, changer de porte revient au même que choisir les deux autres portes.

Quelle que soit la valeur du paramètre α ci-dessus, la probabilité *a priori* de gagner avec la stratégie « changer de porte » est $\frac{2}{3} = P\{\omega_1, \omega_2\}$ et la probabilité de gagner avec la stratégie « ne pas changer de porte » est $\frac{1}{3} = P\{\omega_3, \omega_4\}$. Cependant, le joueur n'est pas obligé de choisir la stratégie à l'avance et la question porte sur le choix à prendre dans une phase intermédiaire du jeu, où le paramètre α joue un rôle.

3.11 Dans cet exercice, l'expérience consiste à tirer une carte au hasard et également la face de celle-ci que l'on observe. Ainsi, l'espace de probabilité est donné par $\Omega = \Omega_1 \times \Omega_2$, où $\Omega_1 = \{1, 2, 3\}$ représente le numéro de la carte tirée et $\Omega_2 = \{a, b\}$ représente la face observée pour une carte donnée. La tribu est $\mathcal{F} = \mathcal{P}(\Omega)$. Nous tirons la carte, de même que la face observée, uniformément et donc $P\{\omega\} = \frac{1}{3 \cdot 2} = \frac{1}{6}$, pour tout $\omega \in \Omega$. On admet que la couleur de la face observée est donnée en fonction de $\omega = (\omega_1, \omega_2)$ dans la figure 11.3.

ω_1	ω_2	Couleur observée	Autre couleur
1	a	noire	noire
1	b	noire	noire
2	a	rouge	rouge
2	b	rouge	rouge
3	a	noire	rouge
3	b	rouge	noire

Figure 11.3 Couleur de la face observée en fonction de ω.

Ainsi, soit R_1 l'événement « la face observée est rouge ». Nous avons $R_1 = \{(2, a), (2, b), (3, b)\}$ et $P(R_1) = \frac{1}{2}$. Soit R_2 l'événement « l'autre face est rouge ». Nous avons $R_2 = \{(2, a), (2, b), (3, a)\}$ et $P(R_2) = \frac{1}{2}$ également. Dans ce problème, nous souhaitons trouver $P(R_2 \mid R_1)$. Par la définition de probabilité conditionnelle, nous avons

$$P(R_2 \mid R_1) = \frac{P(R_2 \cap R_1)}{P(R_1)} = 2 \cdot P\{(2, a), (2, b)\} = 2 \cdot \frac{2}{6} = \frac{2}{3}.$$

3.16 L'ensemble fondamental pour cette expérience est

$$\begin{aligned}\Omega = \{&(\mathrm{P}, \mathrm{P}, \mathrm{P}), (\mathrm{P}, \mathrm{P}, \mathrm{F}), (\mathrm{P}, \mathrm{F}, \mathrm{P}), (\mathrm{P}, \mathrm{F}, \mathrm{F}), (\mathrm{F}, \mathrm{P}, \mathrm{P}),\\ &(\mathrm{F}, \mathrm{P}, \mathrm{F}), (\mathrm{F}, \mathrm{F}, \mathrm{P}), (\mathrm{F}, \mathrm{F}, \mathrm{F})\}.\end{aligned}$$

Tous les événements élémentaires sont équiprobables et donc $P\{\omega\} = \frac{1}{8}$, pour tout $\omega \in \Omega$. D'autre part, en comptant les cas favorables à chaque événement, on constate que

$$\begin{aligned} P(G_1) &= \frac{4}{8} = \frac{1}{2} = P(G_2) = P(G_3) \\ P(G_1 \cap G_2) &= \frac{2}{8} = \frac{1}{4} = P(G_1)P(G_2)\,. \end{aligned}$$

De même, on trouve que $P(G_1 \cap G_3) = P(G_1)P(G_3)$ et que $P(G_2 \cap G_3) = P(G_2)P(G_3)$. Les événements G_1, G_2 et G_3 sont bien deux à deux indépendants. En revanche,

$$P(G_1 \cap G_2 \cap G_3) = \frac{2}{8} = \frac{1}{4} \neq P(G_1) \cdot P(G_2) \cdot P(G_3) = \frac{1}{8},$$

donc G_1, G_2 et G_3 ne sont *pas* mutuellement indépendants. En fait, il est facile de remarquer que si G_1 et G_2 se réalisent, alors G_3 se réalise aussi, ce qui explique cette constatation.

$\boxed{\mathbf{3.17}}$ Soit G_i l'événement « le jet i donne 1 ». Alors $G_1, \ldots, G_8$ sont mutuellement indépendants et $P(G_i) = \frac{1}{6}$, $P(G_i^c) = \frac{5}{6}$. Une manière d'obtenir trois fois 1 est

$$x_1 \; 1 \; x_3 \; x_4 \; 1 \; 1 \; x_7 \; x_8\,, \qquad \text{avec } x_1, x_3, x_4, x_7, x_8 \neq 1\,.$$

La probabilité de cette réalisation de 1 trois fois est, d'après la propriété d'indépendance :

$$P(G_1^c) \cdot P(G_2) \cdot P(G_3^c) \cdot P(G_4^c) \cdot P(G_5) \cdot P(G_6) \cdot P(G_7^c) \cdot P(G_8^c) = \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^5.$$

Puisque les trois 1 peuvent apparaître en n'importe laquelle des $C_{8,3}$ positions possibles,

$$P\{\text{obtenir 1 trois fois}\} = \binom{8}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^5.$$

$\boxed{\mathbf{3.18}}$ Les n lancers de la pièce correspondent à des épreuves indépendantes, où un « succès » est le fait d'obtenir un pile. La probabilité de succès est donc $p = \frac{1}{2}$. Ainsi,

$$\begin{aligned} P\{\text{obtenir au moins une fois pile}\} &= 1 - P\{\text{obtenir aucun pile}\} \\ &= 1 - \binom{n}{0} p^0 (1-p)^{n-0} \\ &= 1 - (1-p)^n = 1 - \left(\frac{1}{2}\right)^n. \end{aligned}$$

3.19 Soit G_i l'événement « le jet i donne un 6 ». On cherche $P(\cap_{i=1}^{\infty} G_i^c)$. Soit $H_n = G_1^c \cap \cdots \cap G_n^c$, ce qui correspond à « aucun 6 parmi les n premiers jets ». Alors $H_n \supset H_{n+1}$ et $\cap_{i=1}^{n} H_i = \cap_{i=1}^{n} G_i^c$. De plus, $P(H_n) = P(G_1^c) \cdots P(G_n^c) = (\frac{5}{6})^n$. Par conséquent, en utilisant la continuité de la probabilité (théorème 1.6), on obtient

$$P\left(\bigcap_{i=1}^{\infty} G_i^c\right) = P\left(\bigcap_{i=1}^{\infty} H_i\right) = \lim_{n\to\infty} P(H_n) = \lim_{n\to\infty} \left(\frac{5}{6}\right)^n = 0\,.$$

Ainsi, il est certain qu'un pile apparaîtra au-moins une fois. Pour un résultat plus précis, voir l'exercice 3.26.

3.20 Soit F_k l'événement « on obtient au moins un succès » parmi les k jeux. Alors $P(F_k^c) = (1 - \frac{1}{N})^k$, donc $P(F_k) = 1 - (1 - \frac{1}{N})^k$. Ainsi,

$$\begin{aligned} P(F_k) \geqslant \frac{1}{2} &\iff \left(1 - \frac{1}{N}\right)^k \leqslant \frac{1}{2} \iff k \ln\left(1 - \frac{1}{N}\right) \leqslant \ln\frac{1}{2} \\ &\iff \ln 2 \leqslant k \ln\left(\frac{1}{1 - \frac{1}{N}}\right) \iff k \geqslant \frac{\ln 2}{\ln\left(\frac{1}{1-\frac{1}{N}}\right)}\,. \end{aligned}$$

Lorsque N est grand,

$$k \geqslant \frac{\ln 2}{\ln\left(1 + \frac{1}{N}\right)} \simeq \frac{\ln 2}{\frac{1}{N}} \simeq \frac{2}{3} N \qquad (\text{puisque } \ln 2 \simeq 0{,}6931)\,.$$

3.21 Une manière possible d'obtenir ce résultat est d'obtenir la séquence R R R V V B B B B, où R signifie rouge, V signifie verte et B signifie bleue (dans l'ordre indiqué). La probabilité d'obtenir cette séquence est

$$\left(\frac{4}{18}\right)^3 \cdot \left(\frac{6}{18}\right)^2 \cdot \left(\frac{8}{18}\right)^4\,.$$

Mais on peut aussi permuter les neuf éléments de la séquence, ce qui donne

$$\frac{9!}{3!\,2!\,4!} = \begin{pmatrix} 9 \\ 3\ 2\ 4 \end{pmatrix}$$

séquences différentes. La probabilité cherchée est donc

$$\begin{pmatrix} 9 \\ 3\ 2\ 4 \end{pmatrix} \left(\frac{4}{18}\right)^3 \left(\frac{6}{18}\right)^2 \left(\frac{8}{18}\right)^4.$$

3.22 a) Nous allons procéder par récurrence sur n. Dans le cas où $n = 2$, le résultat est une conséquence de la définition 3.10 et de la proposition 3.13. Admettons que le résultat est vrai dans le cas où il y a $n-1$ événements ($n \geqslant 3$). Nous devons montrer que

$$P(G_1^{a_1} \cap \cdots \cap G_n^{a_n}) = P(G_1^{a_1}) \cdots P(G_n^{a_n}),$$

pour tous $a_1, \ldots, a_n \in \{0, 1\}$. Nous allons démontrer ceci par récurrence sur le nombre k de complémentaires d'événements qui apparaissent dans l'intersection considérée, c'est-à-dire $k = \big|\{i : a_i = 0\}\big|$. Pour $k = 0$, le résultat est vérifié par définition de l'indépendance mutuelle. Admettons alors que le résultat est vrai pour $k-1$ et montrons-le pour k. Supposons donc qu'il y a k complémentaires qui apparaissent dans l'intersection. Par commutativité de l'intersection, supposons, sans limiter la généralité, que nous nous intéressons à $P(G_1^c \cap \cdots \cap G_k^c \cap G_{k+1} \cap \cdots \cap G_n)$. Nous avons

$$\begin{aligned}
&P(G_1^c \cap \cdots \cap G_k^c \cap G_{k+1} \cap \cdots \cap G_n)\\
&\quad = P(G_2^c \cap \cdots \cap G_k^c \cap G_{k+1} \cap \cdots \cap G_n)\\
&\qquad - P(G_1 \cap G_2^c \cap \cdots \cap G_k^c \cap G_{k+1} \cap \cdots \cap G_n)\\
&\quad = P(G_2^c) \cdots P(G_k^c) P(G_{k+1}) \cdots P(G_n)\\
&\qquad - P(G_1) P(G_2^c) \cdots P(G_k^c) P(G_{k+1}) \cdots P(G_n)\\
&\quad = \big(1 - P(G_1)\big) P(G_2^c) \cdots P(G_k^c) P(G_{k+1}) \cdots P(G_n)\\
&\quad = P(G_1^c) P(G_2^c) \cdots P(G_k^c) P(G_{k+1}) \cdots P(G_n).
\end{aligned}$$

Le premier terme du deuxième membre de l'équation a été traité par l'hypothèse de récurrence sur n, le deuxième par l'hypothèse de récurrence sur k. Ainsi, le résultat est démontré pour toute valeur de $k \leqslant n$, donc pour le cas général de n événements. Il est donc démontré pour toute valeur de n par récurrence.

b) Montrons tout d'abord que si $P(G_1^{a_1} \cap \cdots \cap G_n^{a_n}) = P(G_1^{a_1}) \cdots P(G_n^{a_n})$ pour tous $a_1, \ldots, a_n \in \{0, 1\}$, alors pour tout $1 \leqslant k \leqslant n$, toute suite $1 \leqslant i_1 < \cdots < i_k \leqslant n$ et tous $b_1, \ldots, b_k \in \{0, 1\}$, $P(G_{i_1}^{b_1} \cap \cdots \cap G_{i_k}^{b_k}) = P(G_{i_1}^{b_1}) \cdots P(G_{i_k}^{b_k})$. Ce résultat est clair dans le cas où $k = n$. Supposons par récurrence descendante que celui-ci est vrai pour une suite de taille $k+1$ et montrons-le pour une suite de taille k. Sans limiter la généralité, nous pouvons supposer que $i_j = j$ pour $j = 1, \ldots, k$. Comme G_{k+1}^1 et G_{k+1}^0 sont disjoints et $G_{k+1}^1 \cup G_{k+1}^0 = \Omega$, nous avons

$$\begin{aligned}
&P(G_1^{b_1} \cap \cdots \cap G_k^{b_k})\\
&\quad = P\big(G_1^{b_1} \cap \cdots \cap G_k^{b_k} \cap (G_{k+1}^1 \cup G_{k+1}^0)\big)\\
&\quad = P\big((G_1^{b_1} \cap \cdots \cap G_k^{b_k} \cap G_{k+1}^1) \cup (G_1^{b_1} \cap \cdots \cap G_k^{b_k} \cap G_{k+1}^0)\big)\\
&\quad = P(G_1^{b_1} \cap \cdots \cap G_k^{b_k} \cap G_{k+1}^1) + P(G_1^{b_1} \cap \cdots \cap G_k^{b_k} \cap G_{k+1}^0)\\
&\quad = P(G_1^{b_1}) \cdots P(G_k^{b_k}) P(G_{k+1}^1) + P(G_1^{b_1}) \cdots P(G_k^{b_k}) P(G_{k+1}^0)\\
&\quad = P(G_1^{b_1}) \cdots P(G_k^{b_k}),
\end{aligned}$$

par hypothèse de récurrence et car $P(G^1_{k+1}) + P(G^0_{k+1}) = 1$.
Pour montrer que les événements $G_1, \ldots, G_n$ sont mutuellement indépendants, il suffit de prendre $b_j = 1$ pour tout $j = 1, \ldots, k$ dans le résultat démontré ci-dessus.

3.23 Nous avons, par la loi de De Morgan et la proposition 3.16,

$$P\left(\bigcup_{i=1}^n E_i\right) = P\left(\left(\bigcap_{i=1}^n E_i^c\right)^c\right) = 1 - P\left(\bigcap_{i=1}^n E_i^c\right)$$
$$= 1 - \prod_{i=1}^n P(E_i^c) = 1 - \prod_{i=1}^n \big(1 - P(E_i)\big).$$

3.25 Notons $\mathcal{P}_j(E)$ la collection des sous-ensembles de E de cardinalité j, pour $0 \leqslant j \leqslant n$. Nous avons, en découpant $\{A_1 \subset A_2\}$ selon la valeur de A_2,

$$P\{A_1 \subset A_2\} = \sum_{G \in \mathcal{P}(E)} P\{A_1 \subset A_2, A_2 = G\}$$
$$= \sum_{j=0}^n \sum_{G \in \mathcal{P}_j(E)} P\{A_1 \subset G, A_2 = G\}.$$

Par l'indépendance de A_1 et A_2, nous avons

$$P\{A_1 \subset A_2\} = \sum_{j=0}^n \sum_{G \in \mathcal{P}_j(E)} P\{A_1 \subset G\} P\{A_2 = G\} = \sum_{j=0}^n \sum_{G \in \mathcal{P}_j(E)} \frac{2^{|G|}}{2^n} \frac{1}{2^n}.$$

Il y a en effet 2^n choix possibles pour chacun des ensembles A_1 et A_2, parmi lesquels $2^{|G|}$ appartiennent à $\mathcal{P}(G)$ et sont donc sous-ensembles de G. Il vient alors

$$P\{A_1 \subset A_2\} = \frac{1}{4^n} \sum_{j=0}^n \sum_{G \in \mathcal{P}_j(E)} 2^j = \frac{1}{4^n} \sum_{j=0}^n 2^j \big|\mathcal{P}_j(E)\big| = \frac{1}{4^n} \sum_{j=0}^n \binom{n}{j} 2^j$$
$$= \frac{1}{4^n} \sum_{j=0}^n \binom{n}{j} 2^j 1^{n-j} = \frac{1}{4^n} (2+1)^n = \left(\frac{3}{4}\right)^n.$$

3.27 Soit p la probabilité conditionnelle que l'équipe A gagne le tournoi, sachant que A a déjà gagné i jeux et que B a déjà gagné j jeux. On admet que la répartition du prix est équitable si A reçoit px francs et B reçoit $(1-p)x$ francs : il s'agit donc de calculer p. L'équipe A doit encore gagner $k = n - i$ jeux pour remporter le tournoi et B doit encore gagner $\ell = n - j$ jeux pour remporter le tournoi. Vu que les jeux sont indépendants, p est fonction de (k, ℓ) uniquement. L'équipe A gagnera en exactement m jeux supplémentaires $(k \leqslant m \leqslant k + \ell - 1)$ si A gagne le dernier jeu et $k - 1$ des $m - 1$ premiers jeux supplémentaires. Ceci peut se produire de $C_{m-1,k-1}$ manières. Chaque manière a une probabilité 2^{-m} de se réaliser, donc

$$p = \sum_{m=k}^{k+\ell-1} 2^{-m}\binom{m-1}{k-1}.$$

Cette formule a été obtenue en 1654 par Pascal et Fermat. D'après Durrett [4], divers résultats incorrects avaient été donnés auparavant : $\frac{i}{i+j}$ par Pacioli en 1494, $\frac{n+i-j}{2n}$ par Tartaglia en 1556, et $\frac{2n-1+i-j}{4n-2}$ par Forestani en 1603.

11.4 Chapitre 4

$\boxed{\mathbf{4.1}}$ On constate que

$$\begin{aligned}
P\{X=0\} &= \binom{2}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^2 \simeq 0{,}694\,,\\
P\{X=1\} &= \binom{2}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^1 \simeq 0{,}278\,,\\
P\{X=2\} &= \binom{2}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^0 \simeq 0{,}028\,,
\end{aligned}$$

Ainsi,

$$F(x) = P\{X \leqslant x\} = \begin{cases} 0 & \text{si } x < 0\,,\\ 0{,}694 & \text{si } 0 \leqslant x < 1\,,\\ 0{,}694 + 0{,}278 = 0{,}972 & \text{si } 1 \leqslant x < 2\,,\\ 1 & \text{si } x \geqslant 2\,. \end{cases}$$

$\boxed{\mathbf{4.2}}$ L'ensemble fondamental est $\Omega = \{1,\dots,6\}^2$, muni de la tribu $\mathcal{F} = \mathcal{P}(\Omega)$. Les deux dés étant équilibrés et indépendants, nous définissons la fonction de probabilité par $P\{\omega\} = \frac{1}{36}$, pour tout $\omega \in \Omega$. La variable aléatoire X est alors donnée par $X(\omega) = \omega_1\omega_2$, où $\omega = (\omega_1,\omega_2)$. L'ensemble des valeurs possibles pour la variable aléatoire X est donc donné par $S = \{x_1x_2 : x_1, x_2 \in \{1,\dots,6\}\}$. En comptant les ω qui donnent lieu à chacune de ces valeurs, nous obtenons pour X la fonction de densité suivante :

$$P\{X = x\} = \begin{cases} \dfrac{1}{36} & \text{si } x \in \{1,\, 9,\, 16,\, 25,\, 36\}\,,\\[2mm] \dfrac{1}{18} & \text{si } x \in \{2,\, 3,\, 5,\, 8,\, 10,\, 15,\, 18,\, 20,\, 24,\, 30\}\,,\\[2mm] \dfrac{1}{12} & \text{si } x = 4\,,\\[2mm] \dfrac{1}{9} & \text{si } x \in \{6, 12\}\,,\\[2mm] 0 & \text{sinon}\,. \end{cases}$$

4.3 L'espace de probabilité $(\Omega, \mathcal{F}, P)$ modélise le réalisation de n jets indépendants d'un dé équilibré. Pour chacun de ces jets, la variable aléatoire X est augmentée d'une unité si le résultat est pair. Ainsi, la variable aléatoire X représente le nombre de résultats pairs lors de n jets indépendants d'un dé équilibré. Cette variable aléatoire suit donc une loi binomiale de paramètres n et $p = \frac{1}{2}$. Il y a en effet une chance sur deux d'obtenir un nombre pair lors du lancer d'un dé équilibré. Sa fonction de densité est donc $f(x) = \binom{n}{x}\frac{1}{2^n}$, si $x \in \{0, \ldots, n\}$, $f(x) = 0$ sinon.

4.4 L'ensemble fondamental est $\Omega = \{1, \ldots, 6\}^2$, muni de la tribu $\mathcal{F} = \mathcal{P}(\Omega)$. Les deux dés étant équilibrés et indépendants, nous définissons la fonction de probabilité par $P\{\omega\} = \frac{1}{36}$, pour tout $\omega \in \Omega$. La variable aléatoire X est alors donnée par $X(\omega) = \max(\omega_1, \omega_2)$, où $\omega = (\omega_1, \omega_2)$. L'ensemble des valeurs possibles pour la variable aléatoire X est donc donné par $S = \{1, \ldots, 6\}$. En comptant les ω qui donnent lieu à chacune de ces valeurs, nous obtenons pour X la fonction de densité suivante :

x	1	2	3	4	5	6
$P\{X = x\}$	$\frac{1}{36}$	$\frac{1}{12}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{1}{4}$	$\frac{11}{36}$

4.5 L'espace de probabilité $(\Omega, \mathcal{F}, P)$ modélise le réalisation de n jets indépendants d'une pièce de monnaie. Admettons que 1 représente pile et 0 représente face. Pour chacun de ces jets, la variable aléatoire X est augmentée d'une unité si le résultat est pile. Ainsi, la variable aléatoire X représente le nombre de piles lors de n jets indépendants d'une pièce de monnaie. Cette variable aléatoire suit donc une loi binomiale de paramètres n et $p = \frac{1}{2}$. Sa fonction de densité est donc $f(x) = \binom{n}{x}\frac{1}{2^n}$, si $x \in \{0, \ldots, n\}$, $f(x) = 0$ sinon.

4.6 D'après l'additivité de la probabilité, la définition de F et le point a) de la proposition 4.8,

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

Si $a < b$, alors $\{X \leqslant a\} \subset \{X < b\}$ et donc

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

4.7 Notons D l'ensemble des discontinuités de la fonction F et D_n l'ensemble des discontinuités de F d'amplitude supérieure ou égale à $\frac{1}{n}$, c'est-à- dire

$$D = \big\{x \in \mathbb{R} : F(x) - F(x-) > 0\big\} \quad \text{et} \quad D_n = \left\{x \in \mathbb{R} : F(x) - F(x-) > \frac{1}{n}\right\}.$$

Alors $D = \bigcup_{n \in \mathbb{N}} D_n$. Montrons alors par l'absurde que $|D_n| < n$. Supposons donc que $|D_n| \geqslant n$. Choisissons alors $x_1 < \cdots < x_n \in D_n$. D'une part, par la monotonie de F,

$$\begin{aligned}
&\sum_{i=1}^{n}(F(x_i)-F(x_i-))\\
&\quad = F(x_n)+\big(-F(x_n-)+F(x_{n-1})\big)+\cdots+\big(-F(x_2-)+F(x_1)\big)-F(x_1-)\\
&\quad \leqslant F(x_n)-F(x_1-)\leqslant F(+\infty)-F(-\infty)=1\,,
\end{aligned}$$

et, d'autre part,

$$\sum_{i=1}^{n}\big(F(x_i)-F(x_i-)\big) > \sum_{i=1}^{n}\frac{1}{n}=\frac{n}{n}=1\,.$$

Nous obtenons une contradiction, ce qui montre que $|D_n|<n$ pour tout n. Ainsi, l'ensemble $D=\bigcup_{n\in\mathbb{N}}D_n$ est une réunion dénombrable d'ensembles finis et il est donc dénombrable.

$\boxed{\textbf{4.8}}$ Vérifions tout d'abord que la fonction F_d est non décroissante. En effet, soit $x<y$. Alors

$$\begin{aligned}
F_d(y)-F_d(x) &= P\{X\leqslant y \mid X\in D\}-P\{X\leqslant x\mid X\in D\}\\
&= P\{x<X\leqslant y\mid X\in D\}\geqslant 0\,,
\end{aligned}$$

par définition de la probabilité.

Calculons alors $\lim_{x\to-\infty}F_d(x)$. D'après la définition de la probabilité conditionnelle et le fait que $]-\infty,x]\cap D\subset\,]-\infty,x]$,

$$\begin{aligned}
0\leqslant F_d(x) &= P\{X\leqslant x\mid X\in D\}\\
&= \frac{P\{X\leqslant x, X\in D\}}{P\{X\in D\}}=\frac{P\{X\in\,]-\infty,x]\cap D\}}{P\{X\in D\}}\\
&\leqslant \frac{P\{X\in\,]-\infty,x]\}}{P\{X\in D\}}=\frac{F(x)}{P\{X\in D\}}\;\longrightarrow\;0\,,
\end{aligned}$$

lorsque $x\to-\infty$, car F est une fonction de répartition. Ainsi, $\lim_{x\to-\infty}F_d(x) = 0$.

Calculons $\lim_{x\to+\infty}F_d(x)$. Considérons pour cela une suite réelle croissante $(x_n)_{n\in\mathbb{N}}$ telle que $\lim_{n\to\infty}x_n=+\infty$. Alors

$$\lim_{n\to\infty}F_d(x_n)=\lim_{n\to\infty}\frac{P\big\{X\in\,]-\infty,x_n]\cap D\big\}}{P\{X\in D\}}=\lim_{n\to\infty}\frac{P(E_n)}{P\{X\in D\}}\,,$$

avec $E_n=\big\{X\in\,]-\infty,x_n]\cap D\big\}$. La suite d'événements $(E_n)_{n\in\mathbb{N}}$ est une suite croissante et nous savons que $\lim_{n\to\infty}P(E_n)=P(\bigcup_{n=1}^{\infty}E_n)$, par la continuité de la probabilité. Or $\bigcup_{n=1}^{\infty}E_n=\{X\in D\}$. Ainsi,

$$\lim_{n\to\infty}F_d(x_n)=\frac{P\left(\displaystyle\bigcup_{n=1}^{\infty}E_n\right)}{P\{X\in D\}}=\frac{P\{X\in D\}}{P\{X\in D\}}=1\,.$$

Finalement, montrons la continuité à droite de F_d. On souhaite montrer que $\lim_{h\downarrow 0} F_d(x+h) = F_d(x)$. En utilisant la non-décroissance de F_d et le fait que $]x, x+h] \cap D \subset]x, x+h]$, nous observons que

$$\begin{aligned}0 \leqslant F_d(x+h) - F_d(x) &= \frac{P\big\{X \in]x, x+h] \cap D\big\}}{P\{X \in D\}} \\ &\leqslant \frac{P\big\{X \in]x, x+h]\big\}}{P\{X \in D\}} = \frac{F(x+h) - F(x)}{P\{X \in D\}} \underset{h\downarrow 0}{\longrightarrow} 0\,,\end{aligned}$$

par la continuité à droite de F. Ainsi, F_d est bien une fonction de répartition.

$\boxed{\textbf{4.10}}$ a) La variable aléatoire X suit une loi géométrique de paramètre p, donc $P\{X = n\} = p(1-p)^{n-1}$. Ainsi,

$$\begin{aligned}\sum_{n=1}^{\infty} P\{X = n\} &= \sum_{n=1}^{\infty} p(1-p)^{n-1} = p\sum_{n=1}^{\infty}(1-p)^{n-1} = p\sum_{n=0}^{\infty}(1-p)^{n} \\ &= p\frac{1}{1-(1-p)} = 1\,.\end{aligned}$$

b) La variable aléatoire Y suit une loi de Poisson de paramètre λ, donc $P\{Y = n\} = e^{-\lambda}\frac{\lambda^n}{n!}$. Ainsi,

$$\sum_{n=0}^{\infty} P\{Y = n\} = \sum_{n=0}^{\infty} e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!} = e^{-\lambda}e^{\lambda} = 1\,.$$

c) La variable aléatoire Z suit une loi binomiale négative de paramètres (r, p), donc $P\{Z = n\} = \binom{-r}{n-r}p^r(p-1)^{n-r}$. Ainsi, en utilisant l'indication,

$$\begin{aligned}\sum_{n=r}^{\infty} P\{Z = n\} &= \sum_{n=r}^{\infty}\binom{-r}{n-r}p^r(p-1)^{n-r} = p^r\sum_{n=r}^{\infty}\binom{-r}{n-r}(p-1)^{n-r} \\ &= p^r\sum_{n=0}^{\infty}\binom{-r}{n}(p-1)^{n} = p^r(1+p-1)^{-r} = p^r p^{-r} = 1\,.\end{aligned}$$

d) Supposons que nous effectuons une suite d'épreuves indépendantes, chacune d'elle pouvant être un succès avec la probabilité p ou un échec avec la probabilité $1-p$. La variable aléatoire U compte alors le nombre de succès parmi les n premières épreuves. La variable aléatoire Z compte, quant à elle, le nombre d'épreuves nécessaires pour obtenir r succès. Considérons l'événement $\{U < r\}$. Celui-ci représente le fait qu'il y ait moins de r succès parmi les n premières épreuves. Dans ce cas-là, il sera nécessaire de procéder à plus de n épreuves pour obtenir r succès. Réciproquement, s'il faut procéder à plus de n épreuves pour obtenir r succès, il y a nécessairement moins de r succès parmi les n premières. Ainsi, $\{U < r\} = \{Z > n\}$ et donc $P\{U < r\} = P\{Z > n\}$.

4.15 Vérifions tout d'abord que

$$\lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^n = e^{-\lambda}$$

lorsque $\lambda_n \to \lambda$. Comme $\lambda_n \to \lambda$, nous avons $\lim_{n\to\infty}\frac{\lambda_n}{n} = 0$. Ainsi, il existe N tel que, pour $n \geqslant N$, $\left|\frac{\lambda_n}{n}\right| < 1$. Nous pouvons alors écrire

$$\begin{aligned}\lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^n &= \lim_{n\to\infty}\exp\left(\log\left(\left(1-\frac{\lambda_n}{n}\right)^n\right)\right) = \lim_{n\to\infty}\exp\left(n\log\left(1-\frac{\lambda_n}{n}\right)\right)\\ &= \lim_{n\to\infty}\exp\left(\lambda_n\frac{\log\left(1-\frac{\lambda_n}{n}\right)}{\frac{\lambda_n}{n}}\right)\\ &= \exp\left(\lambda\left(\lim_{x\to 0}\frac{\log(1-x)}{x}\right)\right) = \exp\left(\lambda\left(\lim_{x\to 0}\frac{-1}{1-x}\right)\right) = e^{-\lambda},\end{aligned}$$

où nous avons utilisé la règle de Bernoulli-L'Hospital pour calculer $\lim_{x\to 0}\frac{\log(1-x)}{x}$. Calculons maintenant $P\{X_n = k\}$. En posant $\lambda_n = np_n$, nous avons

$$P\{X_n = k\} = \binom{n}{k} p_n^k(1-p_n)^{n-k} = \frac{n!}{(n-k)!}\,\frac{1}{k!}\left(\frac{\lambda_n}{n}\right)^k\left(1-\frac{\lambda_n}{n}\right)^n\left(1-\frac{\lambda_n}{n}\right)^{-k}.$$

En utilisant

$$\lim_{n\to\infty}\frac{n!}{(n-k)!\,n^k} = \lim_{n\to\infty}\frac{n(n-1)\cdots(n-k+1)}{n^k} = 1\,,$$

$\lim_{n\to\infty}\lambda_n = \lambda$, $\lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^{-k} = 1$ et l'indication, nous obtenons

$$\lim_{n\to\infty} P\{X_n = k\} = \frac{\lambda^k}{k!}\,e^{-\lambda} = P\{Y = k\}\,.$$

4.18 Soit $g :]0,\infty[\to]0,\infty[$ définie par $g(x) = x^2$. Alors g est bijective. Vu que

$$f_X(x) = \begin{cases}\lambda\, e^{-\lambda x} & \text{si } x > 0\,,\\ 0 & \text{si } x \leqslant 0\,,\end{cases}$$

on déduit de la proposition 4.24 que

$$f_Y(y) = \frac{f_X\big(g^{-1}(y)\big)}{\left|g'\big(g^{-1}(y)\big)\right|}\,.$$

Or

$$g^{-1}(y) = \sqrt{y}\,,\qquad g'(x) = 2x\,,\qquad g'\big(g^{-1}(y)\big) = 2\sqrt{y}\,.$$

Par conséquent,

$$f_Y(y) = \begin{cases} \dfrac{\lambda\, e^{-\lambda\sqrt{y}}}{2\sqrt{y}} & \text{si } y > 0\,, \\ 0 & \text{si } y \leqslant 0\,. \end{cases}$$

4.19 a) Il s'agit de déterminer la constante c telle que $\int_{-\infty}^{+\infty} f(x)\,dx = 1$. Pour que f soit positive et non identiquement nulle, il faut $c > 0$. D'autre part,

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x)\,dx &= c\int_{-2}^{2}(4-x^2)\,dx = c\left(4x - \frac{x^3}{3}\right)\Big|_{-2}^{2} \\ &= c\left(\left(8-\frac{8}{3}\right) - \left(-8+\frac{8}{3}\right)\right) = \frac{32}{3}\,c\,. \end{aligned}$$

Nous devons donc poser $c = \frac{3}{32}$.

b) La fonction de répartition $F(x)$ de la variable aléatoire X est donnée par $F(x) = \int_{-\infty}^{x} f(u)\,du$. Lorsque $x \leqslant -2$, cette intégrale est nulle. De plus, lorsque $x \geqslant 2$, nous avons $\int_{-\infty}^{x} f(u)\,du = \int_{-\infty}^{+\infty} f(u)\,du = 1$, par le point a). Il reste donc à déterminer $F(x)$, pour $-2 < x < 2$. Dans ce cas,

$$\begin{aligned} \int_{-\infty}^{x} f(u)\,du &= \frac{3}{32}\int_{-2}^{x}(4-u^2)\,du = \frac{3}{32}\left(4u - \frac{u^3}{3}\right)\Big|_{-2}^{x} \\ &= \frac{3}{32}\left(\left(4x-\frac{x^3}{3}\right) - \left(-8+\frac{8}{3}\right)\right) = \frac{3}{32}\left(4x - \frac{x^3}{3} + \frac{16}{3}\right) \\ &= \frac{1}{32}\left(16 + 12\,x - x^3\right). \end{aligned}$$

Nous obtenons donc

$$F(x) = \begin{cases} 0 & \text{si } x \leqslant -2\,, \\ \dfrac{1}{32}\,(16 + 12\,x - x^3) & \text{si } -2 < x < 2\,, \\ 1 & \text{si } x \geqslant 2\,. \end{cases}$$

4.20 a) Il s'agit de déterminer la constante c telle que $\int_{-\infty}^{+\infty} f(x)\,dx = 1$. Pour que f soit positive et non identiquement nulle, il faut $c > 0$. D'autre part,

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x)\,dx &= c\int_{0}^{\pi/2}\cos^2(x)\,dx = c\int_{0}^{\pi/2}\frac{1+\cos(2x)}{2}\,dx \\ &= c\left(\frac{x}{2} + \frac{\sin(2x)}{4}\right)\Big|_{0}^{\pi/2} = \frac{\pi c}{4}\,. \end{aligned}$$

Nous devons donc poser $c = \frac{4}{\pi}$.

b) La fonction de répartition $F(x)$ de la variable aléatoire X est donnée par $F(x) = \int_{-\infty}^{x} f(u)\,du$. Lorsque $x \leqslant 0$, cette intégrale est nulle. De plus, lorsque $x \geqslant \frac{\pi}{2}$, nous avons $\int_{-\infty}^{x} f(u)\,du = \int_{-\infty}^{+\infty} f(u)\,du = 1$, par le point a). Il reste donc à déterminer $F(x)$, pour $0 < x < \frac{\pi}{2}$. Dans ce cas,

$$\begin{aligned}\int_{-\infty}^{x} f(u)\,du &= \frac{4}{\pi}\int_{0}^{x} \cos^2(u)\,du = \frac{4}{\pi}\left(\frac{u}{2} + \frac{\sin(2u)}{4}\right)\Bigg|_0^x \\ &= \frac{2}{\pi}\left(x + \frac{\sin(2x)}{2}\right).\end{aligned}$$

Nous obtenons donc

$$F(x) = \begin{cases} 0 & \text{si } x \leqslant 0\,, \\ \dfrac{2}{\pi}\left(x + \dfrac{\sin(2x)}{2}\right) & \text{si } 0 < x < \frac{\pi}{2}\,, \\ 1 & \text{si } x \geqslant \frac{\pi}{2}\,. \end{cases}$$

$\boxed{\mathbf{4.23}}$ Soit X une variable aléatoire exponentielle de paramètre λ. Alors $P\{X \leqslant x\} = 1 - e^{-\lambda x}$, pour $x > 0$. Ainsi, $P\{X > x\} = 1 - P\{X \leqslant x\} = e^{-\lambda x}$, pour $x > 0$. Par la définition de probabilité conditionnelle et le fait que $\{X > x + y\} \subset \{X > y\}$, pour $x > 0$ et $y > 0$, nous avons

$$\begin{aligned}P\{X > x + y \mid X > y\} &= \frac{P\{X > x + y, X > y\}}{P\{X > y\}} = \frac{P\{X > x + y\}}{P\{X > y\}} \\ &= \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} = e^{-\lambda x} = P\{X > x\}\,.\end{aligned}$$

$\boxed{\mathbf{4.24}}$ Il s'agit de montrer que l'intégrale de la densité de la loi normale vaut bien 1, c'est-à-dire que $I = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\,dx = 1$. Pour cela, nous allons d'abord évaluer le carré de cette intégrale, ce qui permet de se ramener à une intégrale dans $\mathbb{R}^2$ et de procéder à un changement de coordonnées polaires ($x = r\cos\theta, y = r\sin\theta$). En effet,

$$\begin{aligned}I^2 &= \left(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\,dx\right)\left(\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\,dy\right) \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{\infty} e^{-(1/2)(x^2+y^2)}\,dx\,dy = \frac{1}{2\pi}\int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2/2}\, r\,dr\,d\theta \\ &= \frac{1}{2\pi}\left(\int_{0}^{2\pi} d\theta\right)\left(\int_{0}^{\infty} r\,e^{-r^2/2}\,dr\right) = \left(-e^{-r^2/2}\right)\Bigg|_0^{\infty} = 1\,.\end{aligned}$$

D'où $I = 1$, puisque I est positive.

4.25 a) Observons que $Y = g(X)$, avec $g : \mathbb{R} \to \mathbb{R}_+^*$ définie par $g(x) = e^x$. La fonction g est bijective, g et g^{-1} sont de classe C^1 et $P\{X \in \mathbb{R}\} = 1$. Nous pouvons donc appliquer la proposition 4.24 :

$$f_Y(y) = \frac{f_X\big(g^{-1}(y)\big)}{\big|g'\big(g^{-1}(y)\big)\big|}\,.$$

Comme X est une v.a. normale standard, $f_X(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$. Ainsi, pour $y > 0$,

$$f_Y(y) = \frac{\dfrac{1}{\sqrt{2\pi}}\, e^{-(\log y)^2/2}}{|e^{\log(y)}|} = \frac{1}{\sqrt{2\pi}\, y}\, e^{-(\log y)^2/2}\,,$$

et $f_Y(y) = 0$ si $y \leqslant 0$.

b) Vérifions tout d'abord que la v.a. Z est bien définie. Observons que $h(Z) = X$, avec $h : \mathbb{R} \to \mathbb{R}$ définie par $h(z) = z^3 + z + 1$. Cette fonction est bijective, h et h^{-1} sont de classe C^1 et $P\{X \in \mathbb{R}\} = 1$. En effet, $h'(z) = 3z^2 + 1 > 0$ pour tout $z \in \mathbb{R}$, donc h est strictement croissante. De plus, $\lim_{x\to+\infty} h(z) = +\infty$ et $\lim_{x\to-\infty} h(z) = -\infty$, ce qui assure la bijectivité de h. La v.a. Z est donc bien définie, elle est donnée par $Z = h^{-1}(X)$.
D'après la remarque 4.25, la densité f_Z de la v.a. Z est $f_Z(z) = f_X\big(h(z)\big)\big|h'(z)\big|$. Comme X est une v.a. normale standard,

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}\,(3z^2 + 1)\, e^{-(z^3+z+1)^2/2}\,,$$

pour tout $z \in \mathbb{R}$.

4.26 a) Observons que $Y = g(X)$, avec $g : \mathbb{R}_+^* \to \mathbb{R}$ définie par $g(x) = \ln(x)$. La fonction g est bijective, g et g^{-1} sont de classe C^1 sur $]0, \infty[$ et $P\{X \in]0, \infty[\} = 1$. Nous pouvons donc appliquer la proposition 4.24 :

$$f_Y(y) = \frac{f_X\big(g^{-1}(y)\big)}{\big|g'\big(g^{-1}(y)\big)\big|}\,.$$

Comme X est une v.a. exponentielle de paramètre λ, $f_X(x) = \lambda e^{-\lambda x}$, pour $x \geqslant 0$ et $f_X(x) = 0$ sinon. Ainsi, pour $y \in \mathbb{R}$,

$$f_Y(y) = \frac{\lambda\, e^{-\lambda e^y}}{|e^{-y}|} = \lambda\, e^{y - \lambda e^y}\,.$$

b) Vérifions tout d'abord que la v.a. Z est bien définie. Observons que $h(Z) = X$, avec $h :]-\frac{\pi}{2}, \frac{\pi}{2}[\to \mathbb{R}$ définie par $h(z) = z + \tan(z)$. Cette fonction est bijective. En effet, $h'(z) = 1 + \frac{1}{\cos^2(z)} > 0$ pour tout $z \in]-\frac{\pi}{2}, \frac{\pi}{2}[$, donc h est strictement croissante, h et h^{-1} sont de classe C^1 et $P\{Z \in]-\frac{\pi}{2}, \frac{\pi}{2}[\} = 1$. De plus, $\lim_{x\to\pi/2} h(z) = +\infty$ et

$\lim_{x\to-\pi/2} h(z) = -\infty$, ce qui assure la bijectivité de h. La v.a. Z est donc bien définie, elle est donnée par $Z = h^{-1}(X)$.
D'après la remarque 4.25, la densité f_Z de la v.a. Z est

$$f_Z(z) = f_X\big(h(z)\big)\left|h'(z)\right| .$$

Comme X est une v.a. normale standard,

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}\left(1 + \frac{1}{\cos^2(z)}\right) e^{-(z+\tan(z))^2/2} ,$$

pour tout $z \in \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$.

4.27 a) Observons que $Y = g(X)$, avec $g : \mathbb{R} \to \mathbb{R}$ définie par $g(x) = x^2$. La fonction g n'est pas bijective de $\mathbb{R}$ dans $\mathbb{R}$. Nous ne pouvons donc pas appliquer la proposition 4.24. Déterminons alors la fonction de répartition F_Y de la v.a. Y directement à partir de la définition. Pour $y \leqslant 0$, $F_Y(y) = P\{Y \leqslant y\} = P\{X^2 \leqslant y\} = 0$. Pour $y > 0$,

$$\begin{aligned} F_Y(y) &= P\{Y \leqslant y\} = P\{X^2 \leqslant y\} = P\{-\sqrt{y} \leqslant X \leqslant \sqrt{y}\} \\ &= F(\sqrt{y}) - F(-\sqrt{y}) , \end{aligned}$$

où F est la fonction de répartition de la v.a. X. Nous trouvons alors la densité f_Y de la v.a. Y par dérivation. Pour $y < 0$, $f_Y(y) = 0$. Pour $y > 0$,

$$f_Y(y) = F'_Y(y) = \frac{1}{2\sqrt{y}}\, f(\sqrt{y}) + \frac{1}{2\sqrt{y}}\, f(-\sqrt{y}) = \frac{1}{2\sqrt{y}}\big(f(\sqrt{y}) + f(-\sqrt{y})\big) .$$

Finalement,

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}\big(f(\sqrt{y}) + f(-\sqrt{y})\big) & \text{si } y > 0 , \\ 0 & \text{si } y < 0 . \end{cases}$$

b) Nous allons appliquer le résultat du point a) au cas où X est une v.a. normale standard, c'est-à-dire avec $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Dans ce cas, pour $y > 0$,

$$f_Y(y) = \frac{1}{2\sqrt{y}}\, \frac{1}{\sqrt{2\pi}}\left(e^{-y/2} + e^{-y/2}\right) = \frac{1}{\sqrt{2\pi}}\, y^{-1/2}\, e^{-y/2} ,$$

ce qui correspond à la densité d'une v.a. de loi gamma avec paramètres $p = \frac{1}{2}$ et $\lambda = \frac{1}{2}$. En effet, $y^{-1/2}\, e^{-y/2}$ apparaît dans la densité de la loi gamma et la constante est forcément la bonne puisqu'elle est déterminée par la propriété $\int_0^{+\infty} f_Y(y)\, dy = 1$. En particulier, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

$\boxed{\textbf{4.28}}$ Lorsque X est une v.a. $N(10,36)$, $\frac{X-10}{6}$ est une v.a. normale standard. En notant Z une v.a. normale standard et Φ sa fonction de répartition, nous avons les résultats suivants.

a) $$P\{X<13\} = P\left\{\frac{X-10}{6} < \frac{13-10}{6}\right\} = P\left\{Z < \frac{1}{2}\right\}$$
$$= \Phi\left(\frac{1}{2}\right) \simeq 0{,}6915\,.$$

b) $$P\{4<X<16\} = P\left\{\frac{4-10}{6} < \frac{X-10}{6} < \frac{16-10}{6}\right\} = P\{-1<Z<1\}$$
$$= \Phi(1) - \Phi(-1) = \Phi(1) - \big(1-\Phi(1)\big)$$
$$= 2\Phi(1) - 1 \simeq 2\cdot 0{,}8413 - 1 = 0{,}6826\,.$$

c) $$P\{X<8\} = P\left\{\frac{X-10}{6} < \frac{8-10}{6}\right\} = P\left\{Z < -\frac{1}{3}\right\}$$
$$= \Phi\left(-\frac{1}{3}\right) = 1 - \Phi\left(\frac{1}{3}\right) \simeq 1 - 0{,}6293 = 0{,}3707\,.$$

d) $$P\{7<X<35\} = P\left\{\frac{7-10}{6} < \frac{X-10}{6} < \frac{35-10}{6}\right\}$$
$$= P\left\{-\frac{1}{2} < Z < \frac{25}{6}\right\} = \Phi\left(\frac{25}{6}\right) - \left(1-\Phi\left(\frac{1}{2}\right)\right)$$
$$\simeq 1 - 1 + 0{,}6915 = 0{,}6915\,,$$

car $\Phi\left(\frac{25}{6}\right) \simeq \Phi(+\infty) = 1$.

11.5 Chapitre 5

$\boxed{\textbf{5.1}}$ a) Remarquons d'abord que l'espace de probabilité est $(\Omega,\mathcal{F},P)$, où $\Omega = \{1,\ldots,6\}^2$, $\mathcal{F} = \mathcal{P}(\Omega)$ et la fonction de probabilité P est déterminée par $P\{\omega\} = \frac{1}{36}$, pour tout $\omega \in \Omega$. Le vecteur aléatoire (X_1,X_2) est défini par $\big(X_1(\omega), X_2(\omega)\big) = (\min_{i=1,2}\omega_i,\ \max_{i=1,2}\omega_i)$, où $\omega = (\omega_1,\omega_2)$. L'ensemble des valeurs possibles pour les v.a. X_1 et X_2 est $\{1,\ldots,6\}$. Par définition, nous avons toujours $X_1 \leqslant X_2$. Nous devons calculer $f(x_1,x_2) = P\{X_1 = x_1, X_2 = x_2\}$. Lorsque $x_1 = x_2$, les deux résultats des dés doivent être identiques et égaux à x_1. Il n'y a qu'un seul élément de Ω qui réalise cette condition et donc $f(x_1,x_2) = \frac{1}{36}$. Dans le cas où $x_1 < x_2$, l'un des dés doit donner x_1, l'autre x_2. Il y a deux éléments de Ω qui réalisent cette condition, à savoir (x_1,x_2) et (x_2,x_1). Nous obtenons $f(x_1,x_2) = \frac{1}{18}$. Ainsi, pour $x_1, x_2 \in \{1,\ldots,6\}$,

$$f(x_1,x_2) = \begin{cases} \dfrac{1}{36} & \text{si } x_1 = x_2\,, \\ \dfrac{1}{18} & \text{si } x_1 < x_2\,, \\ 0 & \text{sinon}\,. \end{cases}$$

b) Afin de déterminer la densité marginale de X_1, il s'agit d'additionner les probabilités de tous les événements pour lesquels X_1 prend la valeur voulue (c.-à-d. $f_{X_1}(x_1) = \sum_{i=1}^{6} f(x_1, i)$). Les résultats obtenus aux points a) et b) sont résumés dans le tableau suivant. Les valeurs centrales indiquent $f(x_1, x_2)$. La fonction f_{X_1} (resp. f_{X_2}) est présentée dans la dernière colonne (resp. ligne).

$x_1 \mid x_2$	1	2	3	4	5	6	$f_{X_1}(x_1)$
1	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{11}{36}$
2	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{9}{36}$
3	0	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{7}{36}$
4	0	0	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{5}{36}$
5	0	0	0	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{3}{36}$
6	0	0	0	0	0	$\frac{1}{36}$	$\frac{1}{36}$
$f_{X_2}(x_2)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	

5.2 a) Remarquons d'abord que l'espace de probabilité est $(\Omega, \mathcal{F}, P)$, où $\Omega = \{2, 3, 4\}^2$, $\mathcal{F} = \mathcal{P}(\Omega)$ et la fonction de probabilité P est déterminée par $P\{\omega\} = \frac{1}{9}$, pour tout $\omega \in \Omega$. Le vecteur aléatoire (X_1, X_2) est défini par $(X_1(\omega), X_2(\omega)) = (\omega_1 + \omega_2,\, \omega_1\omega_2)$, où $\omega = (\omega_1, \omega_2)$.
L'ensemble des valeurs possibles pour la v.a. X_1 est $\{4, 5, 6, 7, 8\}$ et l'ensemble des valeurs possibles pour X_2 est $\{4, 6, 8, 9, 12, 16\}$. Nous remarquons que les couples de valeurs possibles sont en correspondance biunivoque avec les couples de valeurs obtenues lors du tirage, à la symétrie du tirage près. Ainsi, les valeurs possibles $(4, 4), (6, 9)$ et $(8, 16)$ pour (X_1, X_2) correspondent respectivement au tirage de $(2, 2), (3, 3)$ et $(4, 4)$. Elles ont donc probabilité $\frac{1}{9}$. Les valeurs possibles $(5, 6), (6, 8)$ et $(7, 12)$ pour (X_1, X_2) correspondent respectivement au tirage de $(2, 3), (2, 4)$ et $(3, 4)$ (à la symétrie des tirages près). Elles ont donc probabilité $\frac{2}{9}$. Les autres couples de valeurs ne sont pas possibles et ont donc tous probabilité 0.

b) Afin de déterminer la densité marginale de X_1, il s'agit d'additionner les probabilités de tous les événements pour lesquels X_1 prend la valeur voulue. Les résultats obtenus aux points a) et b) sont résumés dans le tableau suivant. Les valeurs centrales indiquent $f(x_1, x_2)$. La fonction f_{X_1} (resp. f_{X_2}) est présentée dans la dernière colonne (resp. ligne).

$x_1 \mid x_2$	4	6	8	9	12	16	$f_{X_1}(x_1)$
4	$\frac{1}{9}$	0	0	0	0	0	$\frac{1}{9}$
5	0	$\frac{2}{9}$	0	0	0	0	$\frac{2}{9}$
6	0	0	$\frac{2}{9}$	$\frac{1}{9}$	0	0	$\frac{3}{9}$
7	0	0	0	0	$\frac{2}{9}$	0	$\frac{2}{9}$
8	0	0	0	0	0	$\frac{1}{9}$	$\frac{1}{9}$
$f_{X_2}(x_2)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	

$\boxed{\mathbf{5.3}}$ a) Nous devons choisir c de sorte que

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x_1,x_2)\,dx_1\,dx_2 = 1\,.$$

Or

$$\begin{aligned}
&\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x_1,x_2)\;dx_1\,dx_2 \\
&\quad = c\int_0^1\int_0^1 (x_1+x_2)\,dx_1\,dx_2 = c\int_0^1 \left[\frac{x_1^2}{2}+x_1x_2\right]_{x_1=0}^{x_1=1} dx_2 \\
&\quad = c\int_0^1 \left(\frac{1}{2}+x_2\right) dx_2 = c\left[\frac{x_2}{2}+\frac{x_2^2}{2}\right]_0^1 = c\left(\frac{1}{2}+\frac{1}{2}\right) = c\,.
\end{aligned}$$

Ainsi, $c=1$.

b) Observons que

$$\begin{aligned}
&P\{X_1<X_2\} \\
&\quad = \iint_{\{(x_1,x_2)\,:\,x_1<x_2\}} f(x_1,x_2)\,dx_1\,dx_2 = \int_0^1\int_0^{x_2}(x_1+x_2)\,dx_1\,dx_2 \\
&\quad = \int_0^1 \left[\frac{x_1^2}{2}+x_1x_2\right]_{x_1=0}^{x_1=x_2} dx_2 = \frac{3}{2}\int_0^1 x_2^2\,dx_2 = \frac{3}{2}\left[\frac{x_2^3}{3}\right]_0^1 = \frac{1}{2}\,.
\end{aligned}$$

c) Pour $x_1\notin[0,1]$, $f_{X_1}(x_1)=0$. Pour $x_1\in[0,1]$,

$$\begin{aligned}
f_{X_1}(x_1) &= \int_{-\infty}^{+\infty} f(x_1,x_2)\,dx_2 \\
&= \int_0^1 (x_1+x_2)\,dx_2 = \left[x_1x_2+\frac{x_2^2}{2}\right]_0^1 = \frac{1}{2}+x_1\,.
\end{aligned}$$

d) Nous pouvons écrire $P\{X_1 = X_2\}$ sous la forme

$$P\{X_1 = X_2\} = \iint_{\{(x_1,x_2)\,:\,x_1=x_2\}} f(x_1,x_2)\,dx_1\,dx_2$$
$$= \int_{-\infty}^{\infty}\int_{x_2}^{x_2} f(x_1,x_2)\,dx_1\,dx_2 = 0\,,$$

pour toute fonction de densité f. Ainsi, $P\{X_1 = X_2\} = 0$ pour tout vecteur aléatoire continu.

5.4 a) Nous devons choisir c de sorte que

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x_1,x_2)\,dx_1\,dx_2 = 1\,.$$

Or

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x_1,x_2)\,dx_1\,dx_2$$
$$= c\int_0^1\int_0^1 x_1(1+x_2)\,dx_1\,dx_2 = c\int_0^1 \left[\frac{x_1^2}{2}\,(1+x_2)\right]_{x_1=0}^{x_1=1} dx_2$$
$$= \frac{c}{2}\int_0^1 (1+x_2)\,dx_2 = \frac{c}{2}\left[x_2 + \frac{x_2^2}{2}\right]_0^1 = \frac{3c}{4}\,.$$

Ainsi, $c = \frac{4}{3}$.

b) Observons que

$$P\left\{X_2 > X_1 + \frac{1}{2}\right\} = \iint_{\{(x_1,x_2)\,:\,x_2>x_1+(1/2)\}} f(x_1,x_2)\,dx_1\,dx_2$$
$$= \frac{4}{3}\int_0^{1/2}\int_{x_1+(1/2)}^{1} x_1(1+x_2)\,dx_2\,dx_1$$
$$= \frac{4}{3}\int_0^{1/2} x_1\left[x_2 + \frac{x_2^2}{2}\right]_{x_2=x_1+(1/2)}^{x_2=1} dx_1$$
$$= \frac{4}{3}\int_0^{1/2}\left(\frac{7}{8}\,x_1 - \frac{3}{2}\,x_1^2 - \frac{1}{2}\,x_1^3\right) dx_1$$
$$= \frac{4}{3}\left[\frac{7}{16}\,x_1^2 - \frac{1}{2}\,x_1^3 - \frac{1}{8}\,x_1^4\right]_0^{1/2} = \frac{4}{3}\,\frac{5}{128} = \frac{5}{96}\,.$$

c) Pour $x_2 \notin [0,1]$, $f_{X_2}(x_2) = 0$. Pour $x_2 \in [0,1]$,

$$f_{X_2}(x_2) = \int_{-\infty}^{+\infty} f(x_1,x_2)\,dx_1 = \frac{4}{3}\int_0^1 x_1(1+x_2)\,dx_1$$
$$= \frac{4}{3}\,(1+x_2)\left[\frac{x_1^2}{2}\right]_0^1 = \frac{2}{3}\,(1+x_2)\,.$$

d) Nous pouvons écrire $P\{X_2 = X_1 + \frac{1}{2}\}$ sous la forme

$$P\left\{X_2 = X_1 + \frac{1}{2}\right\} = \iint\limits_{\{(x_1,x_2)\,:\,x_2=x_1+(1/2)\}} f(x_1,x_2)\,dx_1\,dx_2$$
$$= \int_{-\infty}^{\infty}\int_{x_1+(1/2)}^{x_1+(1/2)} f(x_1,x_2)\,dx_1\,dx_2 = 0\,,$$

pour toute fonction de densité f. Ainsi, $P\left\{X_2 = X_1 + \frac{1}{2}\right\} = 0$.

5.5 a) Soit (X_1, X_2) un vecteur aléatoire uniforme sur $U = [a,b] \times [c,d]$. Pour $A \subset U$,

$$P\big\{(X_1,X_2) \in A\big\} = \iint\limits_A \frac{1}{(b-a)(d-c)}\,dx_1\,dx_2$$
$$= \frac{\text{Aire}(A)}{(b-a)(d-c)} = \frac{\text{Aire}(A)}{\text{Aire}(U)}\,.$$

Calculons alors la fonction de répartition F de $X_1 + X_2$ dans le cas où $U = [0,1]^2$. Pour $x < 0$, nous avons $F(x) = P\{X_1 + X_2 \leqslant x\} = 0$. Pour $x \geqslant 2$, nous avons $F(x) = 1$. Dans le cas où $x \in [0,2[$, notons $A_x = \big\{(x_1,x_2) \in [0,1]^2 : x_1 + x_2 \leqslant x\big\}$ (fig. 11.4).

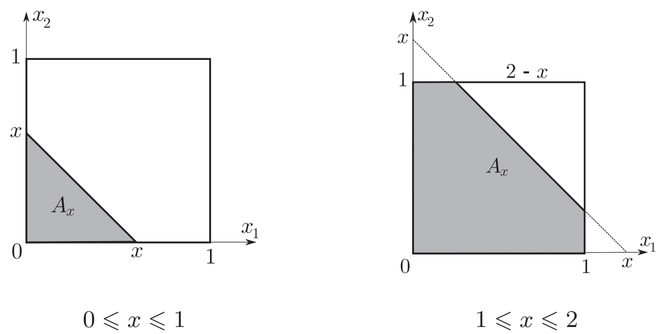


Figure 11.4 Ensemble A_x de l'exercice 5.5 selon la valeur de x.

Alors

$$F(x) = P\{X_1 + X_2 \leqslant x\} = P\big\{(X_1,X_2) \in A_x\big\} = \text{Aire}(A_x)\,.$$

Il s'agit alors de distinguer deux cas. Premièrement, lorsque $0 \leqslant x < 1$, l'ensemble A_x est un triangle rectangle isocèle de côté x. Son aire est donc égale à $\frac{x^2}{2}$. D'autre part, lorsque $1 \leqslant x < 2$, nous avons $A_x = [0,1]^2 \setminus \big\{(x_1,x_2) \in [0,1]^2 : x_1 + x_2 > x\big\}$. Cet ensemble que l'on retire à $[0,1]^2$ est un triangle rectangle isocèle de côté $2 - x$, donc

$F(x) = \text{Aire}(A_x) = 1 - \frac{(2-x)^2}{2}$. Ainsi,

$$F(x) = \begin{cases} 0 & \text{si } x < 0\,, \\ \dfrac{x^2}{2} & \text{si } 0 \leqslant x < 1\,, \\ 1 - \dfrac{(2-x)^2}{2} & \text{si } 1 \leqslant x < 2\,, \\ 1 & \text{si } x \geqslant 2\,. \end{cases}$$

Notons qu'il est également possible, en distinguant les divers cas, de calculer directement $F(x)$ par

$$F(x) = \iint\limits_{\{(x_1,x_2):x_1+x_2 \leqslant x\}} f(x_1, x_2)\, dx_1\, dx_2\,.$$

En dérivant F, nous obtenons

$$f(x) = \begin{cases} 0 & \text{si } x < 0\,, \\ x & \text{si } 0 \leqslant x < 1\,, \\ 2 - x & \text{si } 1 \leqslant x < 2\,, \\ 0 & \text{si } x \geqslant 2\,. \end{cases}$$

b) Calculons tout d'abord la fonction de répartition F_Y de $Y = X_1 + X_2$:

$$\begin{aligned} F_Y(y) &= P\{X_1 + X_2 \leqslant y\} \\ &= \iint\limits_{\{(x_1,x_2)\,:\,x_1+x_2 \leqslant y\}} f(x_1, x_2)\, dx_1\, dx_2 = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{y-x_1} dx_2\, f(x_1, x_2)\,. \end{aligned}$$

Ainsi,

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy} \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{y-x_1} dx_2\, f(x_1, x_2) \\ &= \int_{-\infty}^{\infty} dx_1 \frac{d}{dy} \int_{-\infty}^{y-x_1} dx_2\, f(x_1, x_2) = \int_{-\infty}^{\infty} f(x_1, y - x_1)\, dx_1\,. \end{aligned}$$

Le résultat est démontré. Notons que nous pouvons également calculer la densité conjointe du vecteur aléatoire $(X_1, X_1 + X_2)$ à l'aide de la proposition 5.14, puis calculer la densité marginale de la deuxième composante de celui-ci.

5.6 a) Observons que $Y = g(X)$ où la fonction $g : \mathbb{R}^2 \to \mathbb{R}^2$ est donnée par

$$g(x_1, x_2) = (x_1 + x_2, x_1 - x_2).$$

La fonction g est bijective et admet pour inverse la fonction $h : \mathbb{R}^2 \to \mathbb{R}^2$ donnée par

$$h(y_1, y_2) = \left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right).$$

La fonction h est continûment différentiable et admet pour jacobien

$$J_h(y_1, y_2) = \det \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}.$$

Notons que $P(X \in \mathbb{R}^2) = 1$ et que

$$f_X(x_1, x_2) = \begin{cases} 1 & \text{si } (x_1, x_2) \in [0,1]^2, \\ 0 & \text{sinon.} \end{cases}$$

Ainsi, par la proposition 5.14, nous avons

$$\begin{aligned} f_Y(y_1, y_2) &= f_X(h(y_1, y_2)) \, |J_h(y_1, y_2)| \\ &= \frac{1}{2} f_X\left(\frac{y_1 + y_2}{2}, \frac{y_1 - y_2}{2}\right) = \begin{cases} \frac{1}{2} & \text{si } (y_1, y_2) \in V, \\ 0 & \text{sinon,} \end{cases} \end{aligned}$$

où $V = g([0,1]^2)$.

b) On constate que $(y_1, y_2) \in V$ si et seulement si $\left(\frac{y_1+y_2}{2}, \frac{y_1-y_2}{2}\right) \in [0,1]^2$. Pour satisfaire cette condition, nous devons avoir $0 \leqslant y_1 + y_2 \leqslant 2$ et $0 \leqslant y_1 - y_2 \leqslant 2$. L'ensemble V est illustré dans la figure 11.5. Par le résultat de la partie a), on constate que la fonction de densité conjointe de Y est constante sur V et nulle en dehors. De plus, la valeur de la constante est $\frac{1}{2}$, qui est l'inverse de l'aire du domaine V. Ainsi, on constate que Y est uniforme sur V.

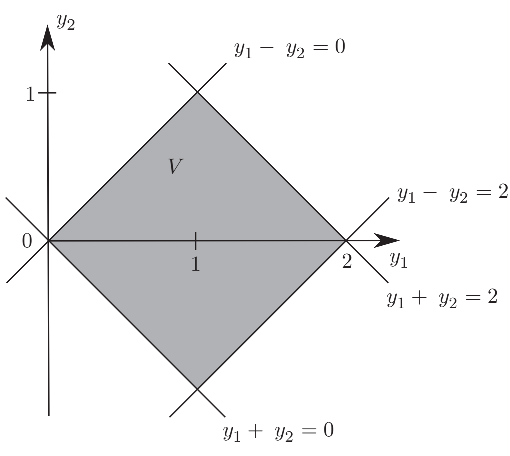


Figure 11.5 Ensemble V du problème 5.6

$\boxed{\textbf{5.7}}$ Observons que $Y = g(X)$, où la fonction $g : D \to \mathbb{R}^2$ est donnée par $g(x_1, x_2) = (g_1(x_1, x_2), g_2(x_1, x_2))$, avec

$$r = g_1(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$$

et

$$\theta = g_2(x_1,x_2) = \begin{cases} \arctan(\frac{x_2}{x_1}) & \text{si } x_1 > 0, \\ \frac{\pi}{2} & \text{si } x_1 = 0 \text{ et } x_2 \geqslant 0, \\ -\frac{\pi}{2} & \text{si } x_1 = 0 \text{ et } x_2 < 0, \\ \arctan(\frac{x_2}{x_1}) + \pi & \text{si } x_1 < 0 \text{ et } x_2 \geqslant 0, \\ \arctan(\frac{x_2}{x_1}) - \pi & \text{si } x_1 < 0 \text{ et } x_2 < 0. \end{cases}$$

Notons que θ est l'angle, défini sur $]-\pi,\pi]$, entre l'axe horizontal et le segment reliant l'origine avec le point (x_1,x_2). On note donc que la fonction g a pour image le domaine $[0,1[\times]-\pi,\pi]$. Afin de travailler sur un domaine ouvert, nous allons restreindre la fonction g au domaine $U = D \setminus \{x \in D : x_1 \leqslant 0, x_2 = 0\}$. On note que $P\{X \in U\} = 1$ étant donné que nous n'excluons de D qu'un segment d'aire nulle. Ainsi, l'image de la restriction de g à U devient $V =]0,1[\times]-\pi,\pi[$. La fonction $g : U \to V$ est bijective et admet pour inverse la fonction $h : V \to U$ donnée par

$$h(r,\theta) = (r\cos(\theta), r\sin(\theta)) .$$

La fonction h est continûment différentiable et admet pour jacobien

$$J_h(r,\theta) = \det \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -r\sin(\theta) & r\cos(\theta) \end{pmatrix} = r.$$

Sachant que

$$f_X(x_1,x_2) = \begin{cases} \frac{1}{\pi} & \text{si } (x_1,x_2) \in D, \\ 0 & \text{sinon,} \end{cases}$$

par la proposition 5.14, nous obtenons

$$\begin{aligned} f_Y(r,\theta) &= f_X(h(r,\theta))\,|J_h(r,\theta)| = r f_X\,(r\cos(\theta), r\sin(\theta)) \\ &= \begin{cases} \frac{r}{\pi} & \text{si } (y_1,y_2) \in V, \\ 0 & \text{sinon.} \end{cases} \end{aligned}$$

$\boxed{\mathbf{5.10}}$ Les deux v.a. ne sont pas indépendantes. En effet, en examinant la solution de l'exercice 5.1b), on constate que $f(x_1,x_2) \neq f_{X_1}(x_1)f_{X_2}(x_2)$. Par exemple, $f(6,1) = 0 \neq \frac{1}{36^2} = f_{X_1}(6)f_{X_2}(1)$.

$\boxed{\mathbf{5.11}}$ Les deux v.a. ne sont pas indépendantes. En effet, en examinant la solution de l'exercice 5.2b), on constate que $f(x_1,x_2) \neq f_{X_1}(x_1)f_{X_2}(x_2)$. Par exemple, $f(5,4) = 0 \neq \frac{2}{81} = f_{X_1}(5)f_{X_2}(4)$.

$\boxed{\mathbf{5.12}}$ a) Par l'indépendance de X_1 et X_2,

$$\begin{aligned} F_X(x) = P\{X \leqslant x\} &= P\big\{\max\{X_1,X_2\} \leqslant x\big\} = P\{X_1 \leqslant x, X_2 \leqslant x\} \\ &= P\{X_1 \leqslant x\}P\{X_2 \leqslant x\} = F_{X_1}(x)F_{X_2}(x) . \end{aligned}$$

b) Par indépendance de X_1 et X_2, nous avons

$$\begin{aligned}F_Y(y) &= P\{X \leqslant y\} = P\big\{\min\{X_1, X_2\} \leqslant y\big\} = 1 - P\big\{\min\{X_1, X_2\} \geqslant y\big\}\\ &= 1 - P\{X_1 \geqslant y, X_2 \geqslant y\} = 1 - P\{X_1 \geqslant y\}P\{X_2 \geqslant y\}\\ &= 1 - \big(1 - F_{X_1}(y)\big)\big(1 - F_{X_2}(y)\big)\,.\end{aligned}$$

c) Dans le cas où les v.a. sont continues, nous savons que la fonction de densité s'obtient en dérivant la fonction de répartition. Ainsi,

$$f_X(x) = \frac{d}{dx}\,F_X(x) = \frac{d}{dx}\,\big(F_{X_1}(x)F_{X_2}(x)\big) = f_{X_1}(x)F_{X_2}(x)+F_{X_1}(x)f_{X_2}(x)\,.$$

De façon analogue,

$$\begin{aligned}f_Y(y) = \frac{d}{dy}\,F_Y(y) &= \frac{d}{dy}\,\Big(1 - \big(1 - F_{X_1}(y)\big)\big(1 - F_{X_2}(y)\big)\Big)\\ &= f_{X_1}(y)\big(1 - F_{X_2}(y)\big) + \big(1 - F_{X_1}(y)\big)f_{X_2}(y)\,.\end{aligned}$$

$\boxed{\mathbf{5.13}}$ Traitons tout d'abord le cas discret. Nous cherchons à calculer $f_{X_1+X_2}(x)$, pour un x donné. Soit D_2 l'ensemble dénombrable des valeurs possibles de la v.a. X_2. Alors

$$\begin{aligned}f_{X_1+X_2}(x) = P\{X_1 + X_2 = x\} &= \sum_{x_2 \in D_2} P\{X_1 + X_2 = x, X_2 = x_2\}\\ &= \sum_{x_2 \in D_2} P\{X_1 = x - x_2, X_2 = x_2\}\\ &= \sum_{x_2 \in D_2} P\{X_1 = x - x_2\}P\{X_2 = x_2\}\\ &= \sum_{x_2 \in D_2} f_{X_1}(x - x_2)f_{X_2}(x_2)\,,\end{aligned}$$

par indépendance de X_1 et X_2. Notons aussi que

$$f_{X_1+X_2}(x) = \sum_{x_1 \in D_1} f_{X_1}(x_1)f_{X_2}(x - x_1)\,,$$

où D_1 est l'ensemble des valeurs possibles de X_1.

Traitons maintenant le cas continu. Comme X_1 et X_2 sont indépendantes, nous savons que $f(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$ pour tout $x_1, x_2 \in \mathbb{R}$. Par le résultat du point b) de l'exercice 5.5, nous avons donc

$$f_{X_1+X_2}(x) = \int_{-\infty}^{\infty} f(x_1, x - x_1)\,dx_1 = \int_{-\infty}^{\infty} f_{X_1}(x_1)f_{X_2}(x - x_1)\,dx_1\,.$$

Comme dans le cas discret, nous avons également

$$f_{X_1+X_2}(x) = \int_{-\infty}^{\infty} f_{X_1}(x - x_2)f_{X_2}(x_2)\,dx_2\,.$$

$\boxed{\mathbf{5.14}}$ Nous traitons explicitement uniquement le cas où les v.a. sont continues. Les autres cas se traitent de façon analogue.

Soient $A_1, A_2 \subset \mathbb{R}$. Nous souhaitons montrer que

$$P\{Z_1 \in A_1,\, Z_2 \in A_2\} = P\{Z_1 \in A_1\}P\{Z_2 \in A_2\}\,.$$

Or

$$\begin{aligned}
P\{Z_1 \in A_1, Z_2 \in A_2\} &= P\big\{g_1(X_1,\dots,X_m) \in A_1,\, g_2(Y_1,\dots,Y_n) \in A_2\big\}\\
&= P\big\{(X_1,\dots,X_m) \in g_1^{-1}(A_1),\ (Y_1,\dots,Y_n) \in g_2^{-1}(A_2)\big\}\\
&= P\big\{(X_1,\dots,X_m,Y_1,\dots,Y_n) \in g_1^{-1}(A_1) \times g_2^{-1}(A_2)\big\}\\
&= \underset{g_1^{-1}(A_1)}{\int\cdots\int} dx_1\cdots dx_m \underset{g_2^{-1}(A_2)}{\int\cdots\int} dy_1\cdots dy_n\\
&\qquad \times f_{X_1}(x_1)\cdots f_{X_m}(x_m) f_{Y_1}(y_1)\cdots f_{Y_n}(y_n)\\
&= \left(\underset{g_1^{-1}(A_1)}{\int\cdots\int} dx_1\cdots dx_m\, f_{X_1}(x_1)\cdots f_{X_m}(x_m)\right)\\
&\qquad \times \left(\underset{g_2^{-1}(A_2)}{\int\cdots\int} dy_1\cdots dy_n\, f_{Y_1}(y_1)\cdots f_{Y_n}(y_n)\right)\\
&= P\big\{(X_1,\dots,X_m) \in g_1^{-1}(A_1)\big\}\, P\big\{(Y_1,\dots,Y_n) \in g_2^{-1}(A_2)\big\}\\
&= P\big\{g_1(X_1,\dots,X_m) \in A_1\big\}\, P\big\{g_2(Y_1,\dots,Y_n) \in A_2\big\}\\
&= P\{Z_1 \in A_1\}P\{Z_2 \in A_2\}\,.
\end{aligned}$$

Ainsi, Z_1 et Z_2 sont indépendantes et le résultat est démontré.

$\boxed{\mathbf{5.16}}$ La fonction de densité conjointe de (X_1, X_2) s'écrit

$$f(x_1, x_2) = g_1(x_1)g_2(x_2)\,.$$

Ainsi,

$$\begin{aligned}
f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f(x_1,x_2)\,dx_2 = \int_{-\infty}^{\infty} g_1(x_1)\,g_2(x_2)\,dx_2\\
&= g_1(x_1)\int_{-\infty}^{\infty} g_2(x_2)\,dx_2 = c_2\,g_1(x_1)\,,
\end{aligned}$$

où $c_2 = \int_{-\infty}^{\infty} g_2(x_2)\,dx_2$. Un raisonnement symétrique montre que $f_{X_2}(x_2) = c_1\,g_2(x_2)$, où $c_1 = \int_{-\infty}^{\infty} g_1(x_1)\,dx_1$. Ainsi,

$$f_{X_1}(x_1)\,f_{X_2}(x_2) = c_1c_2 g_1(x_1)\,g_2(x_2) = f(x_1,x_2)\,.$$

En effet,

$$\begin{aligned}
c_1 c_2 &= \left(\int_{-\infty}^{\infty} g_1(x_1)\,dx_1\right) \cdot \left(\int_{-\infty}^{\infty} g_2(x_2)\,dx_2\right) = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, g_1(x_1)\, g_2(x_2) \\
&= \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2\, f(x_1, x_2) = 1\,.
\end{aligned}$$

Les v.a. X_1 et X_2 sont donc indépendantes par la proposition 5.19.

$\boxed{\mathbf{5.17}}$ Nous avons vu au point a) de l'exemple 5.13 que la densité marginale des composantes d'un vecteur aléatoire gaussien dans $\mathbb{R}^2$ est bien celle d'une loi $N(0,1)$. Ainsi,

$$\begin{aligned}
f(x_1, x_2) &= \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_1^2}{2}\right) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_2^2}{2}\right) = f_{X_1}(x_1)\, f_{X_2}(x_2)\,.
\end{aligned}$$

Par la proposition 5.19, le résultat est démontré.

$\boxed{\mathbf{5.18}}$ Notons F la fonction de répartition commune des v.a. $X_1, \ldots, X_n$. La fonction de densité conjointe de $X_1, \ldots, X_n$ est donc $f(x_1) \cdots f(x_n)$. Calculons alors la probabilité cherchée :

$$\begin{aligned}
&P\{X_1 < \cdots < X_n\} \\
&= \underset{\{x_1 < \cdots < x_n\}}{\int \cdots \int} f(x_1) \cdots f(x_n)\, dx_1 \cdots dx_n \\
&= \int_{-\infty}^{+\infty} dx_n\, f(x_n) \int_{-\infty}^{x_n} dx_{n-1}\, f(x_{n-1}) \cdots \int_{-\infty}^{x_3} dx_2\, f(x_2) \int_{-\infty}^{x_2} dx_1\, f(x_1) \\
&= \int_{-\infty}^{+\infty} dx_n\, f(x_n) \int_{-\infty}^{x_n} dx_{n-1}\, f(x_{n-1}) \cdots \int_{-\infty}^{x_3} dx_2\, f(x_2) F(x_2) \\
&= \int_{-\infty}^{+\infty} dx_n\, f(x_n) \int_{-\infty}^{x_n} dx_{n-1}\, f(x_{n-1}) \cdots \int_{-\infty}^{x_4} dx_3\, f(x_3) \left[\frac{F^2(x_2)}{2}\right]_{-\infty}^{x_3} \\
&= \int_{-\infty}^{+\infty} dx_n\, f(x_n) \int_{-\infty}^{x_n} dx_{n-1}\, f(x_{n-1}) \cdots \int_{-\infty}^{x_4} dx_3\, f(x_3) \frac{F^2(x_3)}{2} \\
&\;\;\vdots \\
&= \int_{-\infty}^{+\infty} dx_n\, f(x_n) \frac{F^{n-1}(x_n)}{(n-1)!} = \left[\frac{F^n(x_n)}{n!}\right]_{-\infty}^{+\infty} = \frac{1}{n!}\,,
\end{aligned}$$

et le résultat est démontré. Signalons encore qu'il est possible de démontrer ce résultat en considérant les v.a. $Y_i = F(X_i)$ $(i = 1, \ldots, n)$ qui sont de loi uniforme sur $[0,1]$, par le point b) de l'exercice 4.30. Le résultat dans le cas de lois uniformes se démontre par calcul de la même intégrale multiple que ci-dessus mais avec $f(x_i)$ remplacé par 1.

5.19 Soit X_1 (resp. X_2) le nombre de minutes après 18 h auxquelles Alain (resp. Bernard) arrive dans le restaurant. Les v.a. X_1 et X_2 sont indépendantes et uniformes sur l'intervalle $[0,60]$. Ainsi, le vecteur aléatoire (X_1, X_2) est uniforme sur $[0,60]^2$. Nous cherchons à calculer la probabilité p que les deux personnes arrivent à moins de 15 minutes l'une de l'autre, autrement dit $p = P\big(\{X_1 \leqslant X_2 \leqslant X_1 + 15\} \cup \{X_2 < X_1 \leqslant X_2 + 15\}\big)$. Ces deux événements sont disjoints et, par la symétrie du problème, leurs probabilités sont égales. Ainsi, $p = 2P\{X_1 \leqslant X_2 \leqslant X_1 + 15\} = 2P\{(X_1, X_2) \in A\}$, avec $A = \big\{(x_1, x_2) \in [0,60]^2 : x_1 \leqslant x_2 \leqslant x_1 + 15\big\}$ (fig. 11.6). Par l'exemple 5.12,

$$p = 2 \cdot \frac{\mathrm{Aire}(A)}{\mathrm{Aire}\big([0,60]^2\big)} = \frac{\mathrm{Aire}(A)}{1\,800}.$$

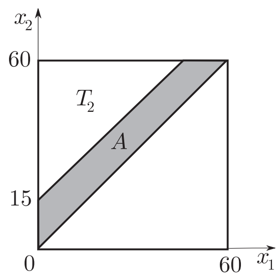


Figure 11.6 Ensemble A de l'exercice 5.19.

Calculons donc Aire(A). L'ensemble A est un trapèze formé du triangle $T_1 = \big\{(x_1, x_2) \in [0,60]^2 : x_1 \leqslant x_2\big\}$ privé du triangle $T_2 = \big\{(x_1, x_2) \in [0,60]^2 : x_2 > x_1 + 15\big\}$ (fig. 11.6). Nous avons donc $\mathrm{Aire}(A) = \mathrm{Aire}(T_1) - \mathrm{Aire}(T_2)$. Le triangle T_1 (resp. T_2) est un triangle rectangle isocèle de côté 60 (resp. 45). Nous trouvons alors $\mathrm{Aire}(A) = \frac{60^2}{2} - \frac{45^2}{2} = \frac{1\,575}{2}$. Finalement, $p = \frac{1\,575}{2 \cdot 1\,800} = \frac{7}{16} = 0{,}4375$.

5.20 a) Soit X_1 (resp. X_2) le nombre de minutes que Jacques va attendre le bus (resp. le métro). Comme Jacques ne connaît pas l'horaire exact de passage des bus et métros, on considère que les v.a. X_1 et X_2 sont indépendantes et uniformes sur l'intervalle $[0,20]$ (resp. $[0,10]$). Ainsi, le vecteur aléatoire (X_1, X_2) est uniforme sur $[0,20] \times [0,10]$. Nous cherchons à calculer la probabilité p que le temps de trajet total en transports publics soit plus grand que 27, autrement dit $p = P\{X_1 + X_2 + 12 > 27\} = P\{X_1 + X_2 > 15\}$. Ainsi, $p = P\big\{(X_1, X_2) \in A\big\}$, avec $A = \big\{(x_1, x_2) \in [0,20] \times [0,10] : x_1 + x_2 > 15\big\}$ (fig. 11.7). Par l'exemple 5.12,

$$p = \frac{\mathrm{Aire}(A)}{\mathrm{Aire}\big([0,20] \times [0,10]\big)} = \frac{\mathrm{Aire}(A)}{200}.$$

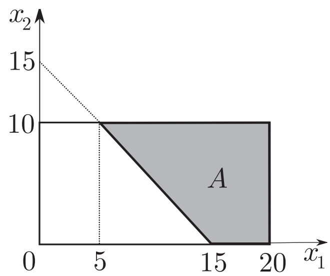


Figure 11.7 Ensemble A de l'exercice 5.20.

Calculons donc Aire(A). L'ensemble A est un trapèze dont la petite base est 5, la grande base 15 et la hauteur 10, donc Aire$(A) = \frac{(5+15)\cdot 10}{2} = 100$. Finalement, $p = \frac{100}{200} = \frac{1}{2}$.

b) Si nous savons que les bus ont deux minutes de retard systématique, cela ne change rien au problème ci-dessus. Si Jacques ne connaît pas l'horaire de passage exact, l'hypothèse d'uniformité du passage des bus reste inchangée.

$\boxed{\textbf{5.21}}$ Soient X la distance entre le centre de l'aiguille et la droite la plus proche et $\Theta \in [0, \pi[$ l'angle entre la moitié supérieure de l'aiguille et les droites parallèles. On suppose que toutes les positions finales de l'aiguille sont équiprobables, c'est-à-dire que la fonction de densité conjointe du vecteur aléatoire (X, Θ) est

$$f(x,\theta) = \begin{cases} \dfrac{2}{\pi} & \text{si } x \in \left[0, \frac{1}{2}\right] \text{ et } \theta \in [0, \pi[\,, \\ 0 & \text{sinon}\,. \end{cases}$$

A l'aide du dessin de la figure 11.8, on constate que l'aiguille touche une droite si et seulement si $\frac{L}{2}\sin(\Theta) \geqslant X$.

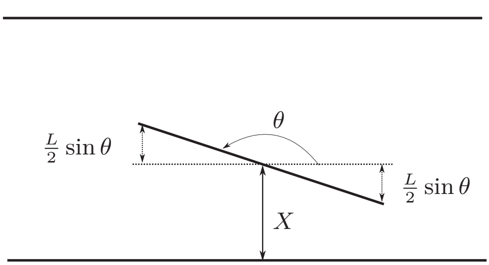


Figure 11.8 Position de l'aiguille dans l'exercice 5.21.

Ainsi, la probabilité cherchée est

$$\begin{aligned}\int_0^{\pi} d\theta \int_0^{\frac{L}{2}\sin\theta} dx\, f(x,\theta) &= \frac{2}{\pi}\int_0^{\pi} d\theta \int_0^{\frac{L}{2}\sin\theta} dx \\ &= \frac{L}{\pi}\int_0^{\pi} d\theta \sin\theta = \frac{2L}{\pi}.\end{aligned}$$

5.24 a) Supposons que X_1 est le nombre de succès parmi n_1 épreuves indépendantes, chacune ayant p pour probabilité de succès, et que X_2 est le nombre de succès parmi n_2 épreuves indépendantes, chacune ayant la même probabilité p de succès. On imagine qu'on réalise d'abord n_1 épreuves, puis n_2 autres épreuves : ainsi, X_1+X_2 est le nombre de succès parmi n_1+n_2 épreuves indépendantes, chacune ayant p pour probabilité de succès. Par conséquent, X_1+X_2 est de loi $B(n_1+n_2,p)$.

b) Calculons la fonction de densité conditionelle

$$f_{X_1}(x_1 \mid X_1+X_2=n) = \frac{P\{X_1=x_1,\, X_1+X_2=n\}}{P\{X_1+X_2=n\}}.$$

Le membre de droite est égal à

$$\frac{P\{X_1=x_1,\, X_2=n-x_1\}}{P\{X_1+X_2=n\}}.$$

Puisque X_1 est X_2 sont indépendantes, ceci devient

$$\frac{P\{X_1=x_1\}P\{X_2=n-x_1\}}{P\{X_1+X_2=n\}}.$$

En utilisant la définition de la loi binomiale et la réponse à la question a), on obtient

$$\begin{aligned}&\frac{\binom{n_1}{x_1} p^{x_1}(1-p)^{n_1-x_1}\binom{n_2}{n-x_1} p^{n-x_1}(1-p)^{n_2-(n-x_1)}}{\binom{n_1+n_2}{n} p^n(1-p)^{n_1+n_2-n}} \\ &= \frac{\binom{n_1}{x_1}\binom{n_2}{n-x_1}}{\binom{n_1+n_2}{n}}.\end{aligned}$$

Ainsi, la loi conditionnelle de X_1 sachant que $X_1+X_2=n$ est la loi d'une v.a. hypergéométrique X (rappellons que cette loi est celle que l'on obtient lors d'un tirage sans remise de n boules d'une urne contenant n_1 boules rouges et n_2 boules noires, où X est le nombre de boules rouges obtenues).

5.25 a) Dans cette question, (X_1, X_2) est un vecteur aléatoire discret. Dans ce cas, la densité conditionnelle de X_2 sachant $X_1 = 2$ est donnée par $f_{X_2}(x_2 \mid X_1 = 2) = \frac{f(2,x_2)}{f_{X_1}(2)}$. La valeur de la densité marginale de X_1 au point $x_1 = 2$ a été calculée à l'exercice 5.1 : $f_{X_1}(2) = \frac{1}{4}$. Ainsi, $f_{X_2}(x_2 \mid X_1 = 2) = 4f(2, x_2)$. En reprenant les résultats du tableau de la loi conjointe de X_1 et X_2, nous avons

x_2	1	2	3	4	5	6
$f_{X_2}(x_2 \mid X_1 = 2)$	0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$

b) Nous sommes cette fois-ci dans le cas d'un vecteur aléatoire continu. La densité conditionnelle de X_1 sachant $X_2 = \frac{1}{2}$ est donnée par $f_{X_1}(x_1 \mid X_2 = \frac{1}{2}) = \frac{f(x_1,1/2)}{f_{X_2}(1/2)}$. La densité marginale de X_2 est la même que celle de X_1 par symétrie. Celle-ci a été calculée à l'exercice 5.3 et vaut $f_{X_2}(x_2) = \frac{1}{2} + x_2$, pour $0 \leqslant x_2 \leqslant 1$. Lorsque $x_1 < 0$ ou $x_1 > 1$, $f_{X_1}(x_1 \mid X_2 = \frac{1}{2}) = 0$. Pour $0 \leqslant x_1 \leqslant 1$, nous obtenons

$$f_{X_1}\left(x_1 \,\middle|\, X_2 = \frac{1}{2}\right) = \frac{f\left(x_1, \frac{1}{2}\right)}{f_{X_2}\left(\frac{1}{2}\right)} = \frac{x_1 + \frac{1}{2}}{\frac{1}{2} + \frac{1}{2}} = \frac{1}{2} + x_1 \,.$$

Nous devons alors calculer $P\{X_1 \leqslant x \mid X_2 = \frac{1}{2}\}$. La fonction de densité conditionnelle étant elle-même une fonction de densité, nous obtenons la fonction de répartition conditionnelle en intégrant $f_{X_1}(\cdot \mid X_2 = \frac{1}{2})$. Pour $x < 0$, nous avons immédiatement $P\{X_1 \leqslant x \mid X_2 = \frac{1}{2}\} = 0$ et, pour $x > 1$, $P\{X_1 \leqslant x \mid X_2 = \frac{1}{2}\} = 1$. Pour $0 \leqslant x \leqslant 1$,

$$\begin{aligned} P\left\{X_1 \leqslant x \,\middle|\, X_2 = \frac{1}{2}\right\} &= \int_{-\infty}^{x} f_{X_1}\left(x_1 \,\middle|\, X_2 = \frac{1}{2}\right) dx_1 = \int_0^x \left(\frac{1}{2} + x_1\right) dx_1 \\ &= \left[\frac{x_1}{2} + \frac{x_1^2}{2}\right]_{x_1=0}^{x_1=x} = \frac{1}{2}\, x(1+x) \,. \end{aligned}$$

5.26 a) Dans cette question, (X_1, X_2) est un vecteur aléatoire discret. Dans ce cas, la densité conditionnelle de X_2 sachant $X_1 = 6$ est donnée par $f_{X_2}(x_2 \mid X_1 = 6) = \frac{f(6,x_2)}{f_{X_1}(6)}$. La valeur de la densité marginale de X_1 au point $x_1 = 6$ a été calculée à l'exercice 5.2 : $f_{X_1}(6) = \frac{1}{3}$. Ainsi, $f_{X_2}(x_2 \mid X_1 = 6) = 3f(6, x_2)$. En reprenant les résultats du tableau de la loi conjointe de X_1 et X_2, nous avons

x_2	8	9
$f_{X_2}(x_2 \mid X_1 = 6)$	$\frac{2}{3}$	$\frac{1}{3}$

et $f_{X_2}(x_2 \mid X_2 = 0) = 0$ si $x_2 \notin \{8, 9\}$.

b) Nous sommes cette fois-ci dans le cas d'un vecteur aléatoire continu. La densité conditionnelle de X_1 sachant $X_2 = \frac{1}{2}$ est donnée par $f_{X_1}(x_1 \mid X_2 = \frac{1}{2}) = \frac{f(x_1,1/2)}{f_{X_2}(1/2)}$. La densité marginale de X_2 a été calculée à l'exercice 5.4 et vaut $f_{X_2}(x_2) = \frac{2}{3}(1+x_2)$, pour $0 \leqslant x_2 \leqslant 1$. Lorsque $x_1 < 0$ ou $x_1 > 1$, $f_{X_1}(x_1|X_2 = \frac{1}{2}) = 0$. Pour $0 \leqslant x_1 \leqslant 1$, nous obtenons

$$f_{X_1}\left(x_1 \,\middle|\, X_2 = \frac{1}{2}\right) = \frac{f\left(x_1, \dfrac{1}{2}\right)}{f_{X_2}\left(\dfrac{1}{2}\right)} = \frac{2x_1}{\dfrac{2}{3}\,\dfrac{3}{2}} = 2x_1\,.$$

Nous devons alors calculer $P\{X_1 \leqslant x \mid X_2 = \frac{1}{2}\}$. La fonction de densité conditionnelle étant elle-même une fonction de densité, nous obtenons la fonction de répartition conditionnelle en intégrant $f_{X_1}(\cdot \mid X_2 = \frac{1}{2})$. Pour $x < 0$, nous avons immédiatement $P\{X_1 \leqslant x \mid X_2 = \frac{1}{2}\} = 0$ et, pour $x > 1$, $P\{X_1 \leqslant x \mid X_2 = \frac{1}{2}\} = 1$. Pour $0 \leqslant x \leqslant 1$,

$$\begin{aligned} P\left\{X_1 \leqslant x \,\middle|\, X_2 = \frac{1}{2}\right\} &= \int_{-\infty}^{x} f_{X_1}\left(x_1 \,\middle|\, X_2 = \frac{1}{2}\right) dx_1 \\ &= \int_0^x 2x_1\, dx_1 = \left[x_1^2\right]_{x_1=0}^{x_1=x} = x^2\,. \end{aligned}$$

5.27 Nous savons que la v.a. X_1 suit une loi uniforme sur $[0,1]$, ainsi $f_{X_1}(x_1) = 1_{[0,1]}(x_1)$. De plus, nous savons que si $X_1 = p$, alors X_2 suit une loi de Bernoulli de paramètre p. Ainsi, nous avons

$$f_{X_2}(x_2 \mid X_1 = p) = \begin{cases} p & \text{si } x_2 = 1\,, \\ 1-p & \text{si } x_2 = 0\,, \\ 0 & \text{si } x_2 \notin \{0,1\}\,. \end{cases}$$

Par la définition de la densité conditionnelle,

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= f_{X_2}(x_2 \mid X_1 = x_1) f_{X_1}(x_1) \\ &= \begin{cases} x_1 & \text{si } x_2 = 1\,,\ x_1 \in [0,1]\,, \\ 1-x_1 & \text{si } x_2 = 0\,,\ x_1 \in [0,1]\,, \\ 0 & \text{sinon}\,. \end{cases} \end{aligned}$$

Par conséquent,

$$f_{X_2}(x_2) = \int_{-\infty}^{+\infty} f_{X_1,X_2}(x_1,x_2)\, dx_1 = \int_0^1 f_{X_1,X_2}(x_1,x_2)\, dx_1$$

et donc $f_{X_2}(1) = \int_0^1 x_1\, dx_1 = \frac{1}{2}$ et $f_{X_2}(0) = \int_0^1 (1-x_1) dx_1 = \frac{1}{2}$. Finalement, comme $f_{X_2}(x_2) = 0$ pour $x_2 \notin \{0,1\}$, nous obtenons que $f_{X_2}(x_2) = \frac{1}{2}$, pour

$x_2 \in \{0,1\}$. Nous pouvons alors calculer la densité conditionnelle de X_1 sachant $X_2 = x_2$: pour $x_2 \in \{0,1\}$,

$$\begin{aligned} f_{X_1}(x_1 \mid X_2 = x_2) &= \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)} = 2f_{X_1,X_2}(x_1,x_2) \\ &= \begin{cases} 2x_1 & \text{si } x_1 \in [0,1]\,,\ x_2 = 1\,, \\ 2(1-x_1) & \text{si } x_1 \in [0,1]\,,\ x_2 = 0\,, \\ 0 & \text{si } x_1 \notin [0,1]\,. \end{cases} \end{aligned}$$

5.28 a) Puisque la v.a. X_1 suit une loi uniforme sur $[0,1]$, $f_{X_1}(x_1) = 1_{[0,1]}(x_1)$. De plus, nous savons que si $X_1 = p$, alors X_2 suit une loi géométrique de paramètre p. Ainsi,

$$f_{X_2}(x_2 \mid X_1 = p) = \begin{cases} p(1-p)^{x_2-1} & \text{si } x_2 \in \{1,2,\ldots\}\,, \\ 0 & \text{sinon}\,. \end{cases}$$

Par la définition de la densité conditionnelle, la densité conjointe est donnée par

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= f_{X_2}(x_2 \mid X_1 = x_1) f_{X_1}(x_1) \\ &= \begin{cases} x_1(1-x_1)^{x_2-1} & \text{si } x_2 \in \{1,2,\ldots\}\,,\ x_1 \in [0,1]\,, \\ 0 & \text{sinon}\,. \end{cases} \end{aligned}$$

b) Nous avons

$$f_{X_2}(x_2) = \int_{-\infty}^{+\infty} f_{X_1,X_2}(x_1,x_2)\,dx_1 = \int_0^1 f_{X_1,X_2}(x_1,x_2)\,dx_1\,.$$

Pour $x_2 \in \{1,2,\ldots\}$, nous avons

$$f_{X_2}(x_2) = \int_0^1 x_1(1-x_1)^{x_2-1}\,dx_1 = \frac{1}{x_2(x_2+1)}$$

d'après l'indication. Finalement, $f_{X_2}(x_2) = 0$ pour $x_2 \notin \{1,2,\ldots\}$.

c) Nous pouvons alors calculer la densité conditionnelle de X_1 sachant $X_2 = x_2$: pour $x_2 \in \{1,2,\ldots\}$,

$$\begin{aligned} f_{X_1}(x_1 \mid X_2 = x_2) &= \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)} \\ &= \begin{cases} x_1(1-x_1)^{x_2-1}x_2(x_2+1) & \text{si } x_1 \in [0,1]\,, \\ 0 & \text{sinon}\,. \end{cases} \end{aligned}$$

11.6 Chapitre 6

6.1 Nous avons

$$P\{X=+1\}=\frac{18}{38}\,,\qquad\qquad P\{X=-1\}=\frac{20}{38}\,.$$

Par conséquent,

$$E[X]=(+1)\times\frac{18}{38}+(-1)\times\frac{20}{38}=-\frac{2}{38}\simeq -0{,}0526\,.$$

Ceci signifie que vous perdez *en moyenne* 5,26 cents par jeu.

6.2 Soit Ω l'ensemble $\{1,2,5,10,20,\text{« 0 »},\text{« 00 »}\}$. Notons ω le résultat de la roue. La probabilité P est donnée par

ω	1	2	5	10	20	0	00
$P\{\omega\}$	$\frac{22}{52}$	$\frac{15}{52}$	$\frac{7}{52}$	$\frac{4}{52}$	$\frac{2}{52}$	$\frac{1}{52}$	$\frac{1}{52}$

a) Soit H la v.a. donnant le gain net du joueur lorsqu'il parie 1 franc sur chacun des nombres ou symboles possibles. Les valeurs possibles de H sont

ω	1	2	5	10	20	0	00
$H(\omega)$	-5	-4	-1	4	14	34	34

et la fonction de densité de H est

x	-5	-4	-1	4	14	34
$P\{H=x\}$	$\frac{22}{52}$	$\frac{15}{52}$	$\frac{7}{52}$	$\frac{4}{52}$	$\frac{2}{52}$	$\frac{2}{52}$

Ainsi, son espérance est donnée par

$$\begin{aligned}E[H]&=(-5)\cdot\frac{22}{52}+(-4)\cdot\frac{15}{52}+(-1)\cdot\frac{7}{52}+4\cdot\frac{4}{52}+14\cdot\frac{2}{52}+34\cdot\frac{2}{52}\\&=-\frac{65}{52}=-1{,}25\,.\end{aligned}$$

b) Pour $m\in\{1,2,5,10,20,\text{« 0 »},\text{« 00 »}\}$, soit H_m le bénéfice net lorsque le joueur mise 1 franc sur le nombre ou symbole m. Alors H_m ne prend que deux valeurs et sa fonction de densité est

x	-1	m
$P\{H_m=x\}$	$1-p_m$	p_m

, si $m\in\{1,2,5,10,20\}$,

x	-1	40
$P\{H_m=x\}$	$\frac{51}{52}$	$\frac{1}{52}$

, si $m\in\{0,00\}$,

où $p_m = P\{\omega = m\}$. Ainsi, $E[H_m] = mp_m + (-1)(1 - p_m)$, ce qui est explicité dans le tableau suivant :

m	1	2	5	10	20	0	00
$E[H_m]$	$-\frac{8}{52}$	$-\frac{7}{52}$	$-\frac{10}{52}$	$-\frac{8}{52}$	$-\frac{10}{52}$	$-\frac{11}{52}$	$-\frac{11}{52}$

Ainsi, miser sur « 0 » ou « 00 » offre la plus faible espérance de gain et miser sur « 2 » offre la plus forte. Notons que les espérances de gain sont toutes négatives et donc à l'avantage du casino.

6.3 a) La v.a. X prend la valeur 1 avec probabilité p et 0 avec probabilité $1 - p$. Rappelons que $E[X] = p$. Nous avons donc

$$E[X^2] = 1 \cdot p + 0 \cdot (1 - p) = p\,.$$

De plus, par le point a) de la proposition 6.15, nous avons

$$\mathrm{Var}(X) = E[X^2] - \big(E[X]\big)^2 = p - p^2 = p(1 - p)\,.$$

b) La densité de la v.a. Y est donnée par

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

Ainsi,

$$E[Y] = \int_0^\infty x\,\frac{\lambda\,e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)}\,dx = \frac{s}{\lambda}\int_0^\infty \frac{\lambda\,e^{-\lambda x}(\lambda x)^{s}}{\Gamma(s+1)}\,dx = \frac{s}{\lambda}\,.$$

De plus,

$$\begin{aligned} E[Y^2] &= \int_0^\infty x^2\,\frac{\lambda\,e^{-\lambda x}(\lambda x)^{s-1}}{\Gamma(s)}\,dx \\ &= \frac{s(s+1)}{\lambda^2}\int_0^\infty \frac{\lambda\,e^{-\lambda x}(\lambda x)^{s+1}}{\Gamma(s+2)}\,dx = \frac{s(s+1)}{\lambda^2}\,. \end{aligned}$$

Par le point a) de la proposition 6.15,

$$\mathrm{Var}(Y) = E[Y^2] - E[Y]^2 = \frac{s(s+1)}{\lambda^2} - \frac{s^2}{\lambda^2} = \frac{s}{\lambda^2}\,.$$

$\boxed{\mathbf{6.4}}$ a) La v.a. X suit une loi binomiale de paramètres n et p. Sa fonction de densité est donc donnée par $f(k) = P\{X = k\} = \binom{n}{k}p^k(1-p)^{n-k}$, pour $k \in \{0, \dots, n\}$. Ainsi, son espérance est donnée par

$$\begin{aligned} E[X] &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n} k\frac{n!}{k!(n-k)!}\,p^k(1-p)^{n-k} \\ &= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}\,p^k(1-p)^{n-k} \\ &= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!}\,p^{k-1}(1-p)^{n-k} \\ &= np\sum_{j=0}^{n-1} \binom{n-1}{j}p^j(1-p)^{n-1-j} = np\,. \end{aligned}$$

Ainsi, l'espérance d'une loi binomiale de paramètres n et p est np. Cela correspond à multiplier le nombre d'expériences aléatoires indépendantes effectuées par la probabilité de succès de chacune.

b) Rappelons que $E[X] = \frac{b+a}{2}$ et que la densité d'une variable aléatoire uniforme sur $[a, b]$ est donnée par

$$f(x) = \frac{1}{b-a}\,1_{[a,b]}(x)\,.$$

Ainsi,

$$E[X^2] = \int_a^b x^2\frac{1}{b-a}\,dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{b-a}\,\frac{b^3-a^3}{3} = \frac{b^2+ba+a^2}{3}\,.$$

Par le point a) de la proposition 6.15,

$$\begin{aligned} \operatorname{Var}(X) = E[X^2] - \big(E[X]\big)^2 &= \frac{b^2+ab+a^2}{3} - \frac{(b+a)^2}{4} \\ &= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}\,. \end{aligned}$$

$\boxed{\mathbf{6.7}}$ La v.a. $1_{\{X\leqslant x\}}$ ne prend que deux valeurs, à savoir 0 et 1. Il s'agit donc d'une variable aléatoire de Bernoulli et nous avons, par le point b) de l'exemple 6.3,

$$E[1_{\{X\leqslant x\}}] = P\{1_{\{X\leqslant x\}} = 1\} = P\{X \leqslant x\} = F_X(x)\,.$$

D'autre part, la v.a. $1_{]-\infty,x]}(X)$ est également une v.a. de Bernoulli et donc

$$E\big[1_{]-\infty,x]}(X)\big] = P\big\{1_{]-\infty,x]}(X) = 1\big\} = P\big\{X \in\,]-\infty, x]\big\} = F_X(x)\,.$$

$\boxed{\mathbf{6.8}}$ Par hypothèse, la v.a. X ne prend que des valeurs dans $\mathbb{N}$. Ainsi, nous avons

$$\sum_{n=1}^{\infty} P\{X \geqslant n\} = \sum_{n=1}^{\infty}\sum_{m=n}^{\infty} P\{X = m\} = \sum_{m=1}^{\infty}\sum_{n=1}^{m} P\{X = m\}$$
$$= \sum_{m=1}^{\infty} m\, P\{X = m\} = E[X]\,,$$

et le résultat est démontré.

$\boxed{\mathbf{6.9}}$ Nous allons distinguer les cas discret et continu. Considérons tout d'abors le cas où (X_1, X_2) est un vecteur aléatoire discret. Si X_1 et X_2 sont indépendantes, alors $P\{X_1 = x_1, X_2 = x_2\} = P\{X_1 = x_1\}P\{X_2 = x_2\}$. Notons D_1 et D_2 les ensembles dénombrables des valeurs possibles de X_1 et X_2, respectivement. Nous avons

$$\begin{aligned}
E[X_1X_2] &= \sum_{x_1\in D_1}\sum_{x_2\in D_2} x_1x_2P\{X_1 = x_1, X_2 = x_2\} \\
&= \sum_{x_1\in D_1}\sum_{x_2\in D_2} x_1x_2P\{X_1 = x_1\}P\{X_2 = x_2\} \\
&= \left(\sum_{x_1\in D_1} x_1\, P\{X_1 = x_1\}\right)\left(\sum_{x_2\in D_2} x_2\, P\{X_2 = x_2\}\right) = E[X_1]E[X_2]\,.
\end{aligned}$$

Considérons maintenant le cas d'un vecteur aléatoire continu. Si X_1 et X_2 sont indépendantes, alors $f(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$. Ainsi, d'après le théorème 6.10,

$$\begin{aligned}
E[X_1X_2] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1x_2f(x_1, x_2)\, dx_1dx_2 \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1x_2f_{X_1}(x_1)f_{X_2}(x_2)\, dx_1\, dx_2 \\
&= \left(\int_{-\infty}^{+\infty} x_1f_{X_1}(x_1)\, dx_1\right)\left(\int_{-\infty}^{+\infty} x_2f_{X_2}(x_2)\, dx_2\right) = E[X_1]E[X_2]\,.
\end{aligned}$$

Le résultat est démontré.

$\boxed{\mathbf{6.11}}$ a) Nous constatons aisément à l'aide des définitions que $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$. De plus, par définition de la variance et la linéarité de l'espérance,

$$\begin{aligned}
\mathrm{Var}(X) &= E\left[\big(X - E[X]\big)^2\right] = E\left[X^2 - 2XE[X] + \big(E[X]\big)^2\right] \\
&= E[X^2] - 2E[X]E[X] + \big(E[X]\big)^2 = E[X^2] - \big(E[X]\big)^2\,.
\end{aligned}$$

Notons que ce résultat est important, car il permet de déduire facilement la variance d'une v.a. à partir de ces deux premiers moments.

b) Par définition de la covariance et la linéarité de l'espérance, nous avons

$$\begin{aligned}\operatorname{Cov}(X_1, X_2) &= E\left[\big(X_1 - E[X_1]\big)\big(X_2 - E[X_2]\big)\right] \\ &= E\big[X_1X_2 - E[X_1]X_2 - X_1E[X_2] + E[X_1]E[X_2]\big] \\ &= E[X_1X_2] - E[X_1]E[X_2] - E[X_1]E[X_2] + E[X_1]E[X_2] \\ &= E[X_1X_2] - E[X_1]E[X_2]\,.\end{aligned}$$

De même que pour la variance, cette formule est importante, car elle simplifie souvent le calcul des covariances.

c) La symétrie de la covariance se déduit aisément de la définition. Montrons alors la linéarité par rapport à la première variable. Soient X_1, Y_1 et X_2 des v.a. et $\alpha, \beta \in \mathbb{R}$. D'après le point b),

$$\begin{aligned}&\operatorname{Cov}(\alpha X_1 + \beta Y_1, X_2) \\ &= E\big[(\alpha X_1 + \beta Y_1)X_2\big] - E[\alpha X_1 + \beta Y_1]E[X_2] \\ &= \alpha E[X_1X_2] + \beta E[Y_1X_2] - \alpha E[X_1]E[X_2] - \beta E[Y_1]E[X_2] \\ &= \alpha \operatorname{Cov}(X_1, X_2) + \beta \operatorname{Cov}(Y_1, X_2)\,.\end{aligned}$$

Par la symétrie de la covariance, celle-ci est également linéaire par rapport à la deuxième variable, elle est donc bilinéaire. Il reste à montrer qu'elle est bien non négative. Soit X une v.a. Par le point a) et la propriété a) p. 80,

$$\operatorname{Cov}(X, X) = \operatorname{Var}(X) = E\left[\big(X - E[X]\big)^2\right] \geqslant 0\,,$$

et le résultat est démontré.

d) Par le point c), nous remarquons que la fonction $(X_1, X_2) \to \operatorname{Cov}(X_1, X_2)$ satisfait les propriétés d'un semi-produit scalaire. Munissons l'espace vectoriel des v.a. à variance finie du semi-produit scalaire $\langle X_1, X_2\rangle = \operatorname{Cov}(X_1, X_2)$. La semi-norme associée à ce produit scalaire est

$$\|X\|^2 = \langle X, X\rangle = \operatorname{Cov}(X, X) = \operatorname{Var}(X)\,.$$

Ainsi, en appliquant l'inégalité de Cauchy-Schwarz, nous obtenons

$$\big|\operatorname{Cov}(X_1, X_2)\big| = \big|\langle X_1, X_2\rangle\big| \leqslant \|X_1\|\,\|X_2\| = \sqrt{\operatorname{Var}(X_1)}\sqrt{\operatorname{Var}(X_2)}\,.$$

Ainsi,

$$\big|\operatorname{Corr}(X_1, X_2)\big| = \frac{\big|\operatorname{Cov}(X_1, X_2)\big|}{\sqrt{\operatorname{Var}(X_1)}\sqrt{\operatorname{Var}(X_2)}} \leqslant 1\,,$$

et le résultat est démontré.

6.12 a) Rappelons tout d'abord l'inégalité de Tchebychev. Pour une variable aléatoire X et $\epsilon > 0$,

$$P\left\{\big|X - E[X]\big| \geqslant \epsilon\right\} \leqslant \frac{\operatorname{Var}(X)}{\epsilon^2}\,.$$

Dans le cas qui nous intéresse, X est $B(n=4, p=\frac{1}{2})$. Ainsi, $E[X] = np = 2$ et $\mathrm{Var}(X) = np(1-p) = 1$. Finalement

$$P\{|X-2| \geqslant 2\} \leqslant \frac{\mathrm{Var}(X)}{4} = \frac{1}{4}\,.$$

La valeur exacte de cette probabilité est

$$P\{|X-2| \geqslant 2\} = P\{X=0\} + P\{X=4\} = 2\binom{4}{0}\left(\frac{1}{2}\right)^4 = \frac{1}{8}\,.$$

L'inégalité est bien vérifiée et l'approximation est relativement satisfaisante.

b) Maintenant, Z est $N(0,1)$, donc $E[Z] = 0$ et $\mathrm{Var}(Z) = 1$. L'inégalité de Tchebychev affirme que

$$P\{|Z| \geqslant 3\} \leqslant \frac{\mathrm{Var}(Z)}{9} = \frac{1}{9}\,.$$

La valeur exacte est

$$\begin{aligned} P\{|Z| \geqslant 3\} &= 1 - P\{-3 \leqslant Z \leqslant 3\} \\ &= 1 - \big(\Phi(3) - \Phi(-3)\big) = 1 - \big(2\Phi(3) - 1\big) \\ &= 2\big(1 - \Phi(3)\big) = 2(1 - 0{,}9987) = 0{,}0026\,, \end{aligned}$$

où Φ est la fonction de répartition de la loi $N(0,1)$. Remarquons que dans ce cas-ci, l'approximation est nettement moins bonne.

6.13 a) Supposons que toutes les boules sont distinguables. Nous nous intéressons aux boules tirées sans tenir compte de l'ordre, de sorte que nous avons $\binom{n+m}{k}$ tirages possibles. Pour que l'événement G_i soit réalisé, nous devons tirer la boule rouge numéro i. Il reste alors $k-1$ boules à choisir parmi les $n+m-1$ restantes. Nous avons donc $\binom{n+m-1}{k-1}$ cas favorables de sorte que

$$P(G_i) = \frac{\dbinom{n+m-1}{k-1}}{\dbinom{n+m}{k}} = \frac{k}{n+m}\,.$$

Remarquons que cette probabilité est indépendante de la valeur de i.

b) Les événements $G_1, \ldots, G_m$ sont équiprobables. (*Attention.* Ils ne sont pas indépendants. En effet, pour s'en convaincre, prenons le cas $k=1$. Si nous tirons la boule numéro 1, nous ne pouvons pas tirer la boule numéro 2 et donc G_1 et G_2 ne sont pas indépendants. La v.a. X ne suit donc pas une loi binomiale.) La v.a. X représente le nombre de boules rouges tirées. En effet, pour que $X = r$, exactement r parmi les m boules rouges doivent être tirées. Les $k-r$ boules restantes doivent ensuite être choisies parmi les n boules noires. Nous constatons ainsi que X suit une loi hypergéométrique de paramètres m, n et k.

c) Par linéarité de l'espérance, nous constatons que

$$E[X] = E[1_{G_1}+\cdots+1_{G_m}] = P(G_1)+\cdots+P(G_m) = m\frac{k}{n+m} = \frac{mk}{n+m}\,.$$

Ainsi, l'espérance d'une v.a. hypergéométrique de paramètres m, n et k est donnée par $E[X] = \frac{mk}{n+m}$.

6.16 Soit $\varphi(a) = E\big[(X-a)^2\big]$. Alors

$$\varphi(a) = E\big[(X-a)^2\big] = E[X^2 - 2aX + a^2] = E[X^2] - 2aE[X] + a^2\,.$$

La fonction φ est un polynôme de degré 2 en a dont le terme quadratique est à coefficient positif. Ainsi, son minimum est atteint en $a = \frac{2E[X]}{2} = E[X]$. La valeur de a qui minimise la fonction φ correspond donc à l'espérance de X. De plus, la valeur minimale réalisée est $\varphi\big(E[X]\big) = E\big[\big(X - E[X]\big)^2\big] = \mathrm{Var}(X)$.

6.17 Remarquons tout d'abord que pour une fonction convexe g, avec $x_0 \in \mathbb{R}$ fixé, il existe une constante c telle que $g(x) \geqslant g(x_0)+c(x-x_0)$, pour tout $x \in \mathbb{R}$. Dans le cas où la fonction g est dérivable, $c = g'(x_0)$. Dans le cas contraire, nous savons que g est continue et que les dérivées à gauche $g'(x_0-)$ et à droite $g'(x_0+)$ existent. Toute valeur de $c \in \big[g'(x_0-), g'(x_0+)\big]$ convient.

Ainsi, en appliquant l'inégalité ci-dessus à X en fixant $x_0 = E[X]$, nous obtenons $g(X) \geqslant g\big(E[X]\big)+c\big(X-E[X]\big)$. En prenant l'espérance à gauche et à droite de l'inégalité (qui est préservée par la propriété b) p. 80), nous obtenons

$$\begin{aligned} E\big[g(X)\big] &\geqslant E\left[g\big(E[X]\big) + c\big(X - E[X]\big)\right] \\ &= g\big(E[X]\big) + cE\big[X - E[X]\big] = g\big(E[X]\big)\,, \end{aligned}$$

et le résultat est démontré.

6.18 Soit $n \geqslant 1$. Nous voulons montrer que $E[X^n]^{1/n} \leqslant E[X^{n+1}]^{1/(n+1)}$. Pour cela, considérons la fonction $g : \mathbb{R} \to \mathbb{R}$ définie par $g(x) = x^{\frac{n+1}{n}}$ si $x \geqslant 0$ et $g(x) = 0$ si $x \leqslant 0$. Comme $\frac{n+1}{n} \geqslant 1$ pour tout $n \geqslant 1$, la fonction g est croissante et convexe sur $\mathbb{R}$. Ainsi, par l'inégalité de Jensen (exercice 6.17), $E\big[X^n\big]^{(n+1)/n} \leqslant E[X^{n+1}]$. Finalement $E[X^n]^{1/n} \leqslant E\big[X^{n+1}\big]^{1/(n+1)}$. Le résultat étant valable pour tout $n \geqslant 1$, le résultat est démontré.

6.19 a) Nous avons $X\,1_{\{X \geqslant a\}} \geqslant a\,1_{\{X\geqslant a\}}$. Ainsi, par la monotonie de l'espérance,

$$\begin{aligned} E[X] &= E[X\,1_{\{X\geqslant a\}}] + E[X\,1_{\{X<a\}}] \\ &\geqslant aP\{X \geqslant a\} + E[X\,1_{\{X<a\}}] \geqslant aP\{X \geqslant a\}\,, \end{aligned}$$

car X est une v.a. non négative. Finalement, comme $a > 0$,

$$P\{X \geqslant a\} \leqslant \frac{E[X]}{a}\,.$$

b) En appliquant le résultat du point a) à la v.a. X^p ($p \geqslant 1$), nous avons

$$P\{X \geqslant a\} = P\{X^p \geqslant a^p\} \leqslant \frac{E[X^p]}{a^p}\,.$$

$\boxed{\textbf{6.20}}$ a) Notons F la fonction de répartition de X. Nous allons appliquer la formule d'intégration par parties au membre de droite de l'égalité à démontrer en considérant $u'(x) = px^{p-1}$ et $v(x) = P\{X > x\} = 1 - F(x)$. Nous avons alors $u(x) = x^p$, $v'(x) = -f(x)$ et nous obtenons

$$\begin{aligned}\int_0^{+\infty} px^{p-1} P\{X > x\}\,dx &= \Big[x^p P\{X > x\}\Big]_0^{+\infty} + \int_0^{+\infty} x^p f(x)\,dx \\ &= \Big[x^p P\{X > x\}\Big]_0^{+\infty} + E[X^p] = E[X^p]\,.\end{aligned}$$

En effet, le premier terme ci-dessus est nul. En $x = 0$, c'est immédiat. Lorsque $x \to +\infty$, nous avons

$$0 \leqslant x^p P\{X > x\} = x^p \int_x^{+\infty} f(y)\,dy \leqslant \int_x^{+\infty} y^p f(y)\,dy \to 0\,,$$

lorsque $x \to \infty$ car $E[X^p] = \int_0^{+\infty} y^p f(y)\,dy < \infty$ par hypothèse. Le résultat est donc démontré.

b) Considérons maintenant une v.a. X telle que $P\{X \in \mathbb{N}\} = 1$. Ainsi, la fonction $x \mapsto P\{X > x\}$ est constante et vaut $P\{X \geqslant k\}$ sur tout intervalle de la forme $[k-1, k[$, $k \in \mathbb{N}^*$. Ainsi, nous pouvons réécrire le membre de droite de l'égalité à démontrer sous la forme suivante :

$$\begin{aligned}&\int_0^{\infty} px^{p-1} P\{X > x\}\,dx \\ &= \sum_{k=1}^{\infty} \int_{k-1}^{k} px^{p-1} P\{X > x\}\,dx = \sum_{k=1}^{\infty} P\{X \geqslant k\} \int_{k-1}^{k} px^{p-1}\,dx \\ &= \sum_{k=1}^{\infty} P\{X \geqslant k\}\big(k^p - (k-1)^p\big) = \sum_{k=1}^{\infty}\big(k^p - (k-1)^p\big) \sum_{j=k}^{\infty} P\{X = j\} \\ &= \sum_{j=1}^{\infty} P\{X = j\} \sum_{k=1}^{j}\big(k^p - (k-1)^p\big) = \sum_{j=1}^{\infty} j^p P\{X = j\} = E[X^p]\,,\end{aligned}$$

et le résultat est démontré.

$\boxed{\textbf{6.27}}$ a) Rappelons que $M_X(0) = 1$, $M_X'(0) = E[X]$ et $M_X''(0) = E[X^2]$. A partir de la définition de $\kappa_X(t)$, nous avons

$$\kappa_X'(0) = \big(\ln M_X(t)\big)'\Big|_{t=0} = \frac{M_X'(t)}{M_X(t)}\bigg|_{t=0} = \frac{M_X'(0)}{M_X(0)} = E[X]\,.$$

Par suite,

$$\kappa''_X(0) = \left(\frac{M'_X(t)}{M_X(t)}\right)'\Bigg|_{t=0} = \frac{M''_X(t)M_X(t) - \big(M'_X(t)\big)^2}{M_X(t)^2}\Bigg|_{t=0}$$
$$= \frac{M''_X(0)M_X(0) - \big(M'_X(0)\big)^2}{M_X(0)^2} = E[X^2] - \big(E[X]\big)^2 = \mathrm{Var}(X)\,,$$

par le point a) de la proposition 6.15.

b) Dans le cas où X_1 et X_2 sont indépendantes, nous savons par le théorème 6.28 que $M_{X_1+X_2}(t) = M_{X_1}(t)M_{X_2}(t)$. Ainsi,

$$\kappa_{X_1+X_2}(t) = \ln M_{X_1+X_2}(t) = \ln\big(M_{X_1}(t)M_{X_2}(t)\big)$$
$$= \ln M_{X_1}(t) + \ln M_{X_2}(t) = \kappa_{X_1}(t) + \kappa_{X_2}(t)\,.$$

c) Nous savons que si X suit une loi $N(\mu, \sigma^2)$, $M_X(t) = \exp\Big(\mu t + \frac{\sigma^2 t^2}{2}\Big)$. Ainsi, $\kappa_X(t) = \mu t + \frac{\sigma^2 t^2}{2}$ est un polynôme de degré 2 en t. Donc, il est clair que $\kappa_X^{(n)}(t) \equiv 0$ pour tout $n \geqslant 3$.

6.28 a) Traitons tout d'abord le cas d'une v.a. Y de loi $N(0,1)$. Dans ce cas,

$$M_Y(t) = E[e^{tY}] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{ty}\, e^{-y^2/2}\, dy$$
$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-(y^2-2ty+t^2)/2}\, e^{t^2/2}\, dy$$
$$= \frac{1}{\sqrt{2\pi}}\, e^{t^2/2}\int_{-\infty}^{+\infty} e^{-(y-t)^2/2}\, dy = e^{t^2/2}\,.$$

Ainsi, en posant $X = \mu + \sigma Y$, avec Y de loi $N(0,1)$, nous obtenons

$$M_X(t) = E[e^{tX}] = E[e^{\mu t + \sigma t Y}]$$
$$= e^{\mu t}E[e^{\sigma t Y}] = e^{\mu t}e^{\sigma^2 t^2/2} = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

b) Rappelons tout d'abord que pour tout $x \in \mathbb{R}$, $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. Considérons $t \in \mathbb{R}$ tel que $|t| < t_0$. Comme $M_X(t) < \infty$, nous pouvons écrire

$$M_X(t) = E[e^{tX}] = E\left[\sum_{k=0}^{\infty} \frac{t^k}{k!}\, X^k\right] = \sum_{k=0}^{\infty} \frac{t^k}{k!}\, E[X^k] \qquad (11.6)$$

(l'hypothèse permet de permuter somme et espérance).

c) Considérons la fonction génératrice des moments de la loi $N(0,\sigma^2)$ et développons-la sous forme de série. Nous avons

$$M_X(t) = e^{\sigma^2 t^2/2} = \sum_{k=0}^{\infty} \frac{\sigma^{2k} t^{2k}}{2^k k!} = \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!} \frac{\sigma^{2k}(2k)!}{2^k k!}.$$

Ainsi, en identifiant la série ci-dessus et (11.6) terme à terme, nous obtenons que $E[X^{2k+1}] = 0$ et $E[X^{2k}] = \frac{\sigma^{2k}(2k)!}{2^k k!}$, pour tout $k \in \mathbb{N}$.

6.31 D'après l'exercice 6.20 dans le cas où $p = 1$,

$$\begin{aligned} E\big[g(X)\big] &= \int_0^\infty P\big\{g(X) \geqslant u\big\}\, du = \int_0^\infty du \int_{\{x:g(x)\geqslant u\}} dx f(x) \\ &= \int_0^\infty f(x)\, dx \int_0^{g(x)} du = \int_0^\infty f(x) g(x)\, dx \end{aligned}$$

et le résultat est démontré (il est possible de permuter les deux intégrales car la fonction de densité f est à valeurs positives).

6.32 a) Par définition de la fonction génératrice des moments et le fait que $P\{X > 0\} = 1$, nous avons :

$$\begin{aligned} \int_{-\infty}^0 dt\, M_X(t) &= \int_{-\infty}^0 dt \int_0^\infty dx\, e^{tx} f(x) = \int_0^\infty dx\, f(x) \int_{-\infty}^0 dt\, e^{tx} \\ &= \int_0^\infty dx\, f(x) \left[\frac{e^{tx}}{x}\right]_{t=-\infty}^{t=0} = \int_0^\infty dx\, \frac{1}{x} f(x) = E\left[\frac{1}{X}\right], \end{aligned}$$

d'après le théorème 6.10, et le résultat est démontré (il est possible de permuter les deux intégrales car l'intégrand est toujours positif).

b) On prouve le résultat plus général suivant : pour tout $n \geqslant 1$, pour tout $t_0 \leqslant 0$,

$$\int_{-\infty}^{t_0} dt_1 \int_{-\infty}^{t_1} dt_2 \cdots \int_{-\infty}^{t_{n-1}} dt_n\, M_X(t_n) = E\left[e^{t_0 X} X^{-n}\right].$$

On procède par récurrence. Le résultat pour $n = 1$ est prouvé de manière similaire à la partie (a). En effet, nous avons

$$\begin{aligned} \int_{-\infty}^{t_0} dt\, M_X(t) &= \int_{-\infty}^{t_0} dt \int_0^\infty dx\, e^{tx} f(x) = \int_0^\infty dx\, f(x) \int_{-\infty}^{t_0} dt\, e^{tx} \\ &= \int_0^\infty dx\, f(x) \left[\frac{e^{tx}}{x}\right]_{t=-\infty}^{t=t_0} = \int_0^\infty dx\, \frac{e^{t_0 x}}{x} f(x) = E\left[\frac{e^{t_0 X}}{X}\right]. \end{aligned}$$

D'après l'hypothèse et l'exercice 6.18, ce moment est fini car $e^{t_0 x} \leqslant 1$ pour $t_0 \leqslant 0$. Supposons maintenant que le résultat est vrai pour n et prouvons le pour $n+1$. Par l'hypothèse de récurrence, nous avons :

$$\begin{aligned}
\int_{-\infty}^{t_0} dt_1 \int_{-\infty}^{t_1} dt_2 \cdots \int_{-\infty}^{t_n} dt_{n+1}\, M_X(t_{n+1}) &= \int_{-\infty}^{t_0} dt_1\, E\left[e^{t_1 X} X^{-n}\right] \\
&= \int_{-\infty}^{t_0} dt_1 \int_0^{\infty} dx\, e^{t_1 x} x^{-n} f(x) \\
&= \int_0^{\infty} dx\, x^{-n} f(x) \int_{-\infty}^{t_0} dt_1\, e^{t_1 x} \\
&= \int_0^{\infty} dx\, e^{t_0 x} x^{-(n+1)} f(x) \\
&= E\left[e^{t_0 X} X^{-(n+1)}\right]
\end{aligned}$$

(à nouveau, il est possible de permuter les deux intégrales car l'intégrand est positif). Le résultat général est donc prouvé. Pour obtenir la conclusion demandée, il suffit de poser $t_0 = 0$.

c) En utilisant la densité de la loi Gamma de paramètres p et λ, nous avons

$$\int_0^{\infty} \frac{1}{x} f(x)\, dx = \int_0^{\infty} \frac{\lambda^p e^{-\lambda x} x^{p-2}}{\Gamma(p)}\, dx < \infty,$$

car $p > 1$. Ainsi, par la partie a) et comme $p > 1$, nous avons

$$\begin{aligned}
E\left[\frac{1}{X}\right] &= \int_{-\infty}^{0} M_X(t)\, dt = \lambda^p \int_{-\infty}^{0} (\lambda - t)^{-p}\, dt \\
&= \frac{\lambda^p}{-p+1} \left[-(\lambda - t)^{-p+1}\right]_{-\infty}^{0} = \frac{\lambda}{p-1}.
\end{aligned}$$

Remarque. Notons que la partie c) ci-dessus illustre le fait que $E[1/X] \neq 1/E[X]$.

11.7 Chapitre 7

7.1 Pour i allant de 1 à 36, notons X_i la v.a. représentant la différence entre le $i^{\text{ième}}$ nombre réel et sa valeur arrondie. Les v.a. X_i sont i.i.d. et suivent chacune une loi uniforme sur $]-0{,}5; 0{,}5[$. De plus, $E[X_i] = 0$ et $\mathrm{Var}(X_i) = \frac{1}{12}$, pour tout i. La différence entre la somme des 36 nombres réels et la somme des valeurs arrondies est représentée par $X = \sum_{i=1}^{36} X_i$. Nous cherchons à calculer

$P\big\{|X| > 5\big\}$. En utilisant le théorème limite central, nous obtenons

$$\begin{aligned}
P\big\{|X| > 5\big\} &= 1 - P\{-5 < X < 5\} \\
&= 1 - P\left\{\frac{-5 - 36 \cdot 0}{\dfrac{1}{\sqrt{12}} \cdot \sqrt{36}} < \frac{X - 36 \cdot 0}{\dfrac{1}{\sqrt{12}} \cdot \sqrt{36}} < \frac{5 - 36 \cdot 0}{\dfrac{1}{\sqrt{12}} \cdot \sqrt{36}}\right\} \\
&\simeq 1 - P\left\{-\frac{5}{\sqrt{3}} < Z < \frac{5}{\sqrt{3}}\right\} = 2 - 2\,\Phi\left(\frac{5}{\sqrt{3}}\right) \\
&= 2 - 2\,\Phi\,(2{,}89) = 2 \times (1 - 0{,}9981) = 0{,}0038\,,
\end{aligned}$$

où Z est une v.a. $N(0,1)$ et Φ est sa fonction de répartition. Ainsi, nous remarquons qu'avec une très forte probabilité, nous pouvons affirmer que la somme des valeurs arrondies diffère de la somme réelle de moins de 5.

7.2 a) La v.a. X peut s'écrire $X = \sum_{i=1}^{100} X_i$, où les X_i sont des v.a. i.i.d suivant chacune une loi de Bernoulli de paramètre $p = \frac{1}{2}$. Nous avons $E[X_i] = p = \frac{1}{2}$ et $\mathrm{Var}(X_i) = p(1-p) = \frac{1}{4}$. Ainsi, en appliquant le théorème limite central, nous obtenons

$$\begin{aligned}
&P\{49{,}5 < X < 50{,}5\} \\
&= P\left\{\frac{49{,}5 - 100 \cdot \dfrac{1}{2}}{\sqrt{\dfrac{1}{4} \cdot 100}} \leqslant \frac{X - 100 \cdot \dfrac{1}{2}}{\sqrt{\dfrac{1}{4} \cdot 100}} \leqslant \frac{50{,}5 - 100 \cdot \frac{1}{2}}{\sqrt{\frac{1}{4} \cdot 100}}\right\} \\
&\simeq P\{-0{,}1 < Z < 0{,}1\} = 2\,\Phi(0{,}1) - 1 = 2 \times 0{,}5398 - 1 = 0{,}0796\,,
\end{aligned}$$

où Z est une v.a. $N(0,1)$ et Φ est sa fonction de répartition. Nous pouvons alors estimer $P\{X = 50\}$ par 0,0796.

b) La somme Y des résultats obtenus peut s'écrire $Y = \sum_{i=1}^{200} Y_i$, où les Y_i sont des v.a. i.i.d suivant chacune une loi représentant le résultat du jet d'un dé (c.-à-d. elles sont uniformes sur $\{1, \ldots, 6\}$). Nous avons $E[Y_i] = \frac{7}{2}$ et $\mathrm{Var}(Y_i) = \frac{35}{12}$. Nous cherchons à calculer $P\{650 \leqslant Y \leqslant 750\}$.
Calculons tout d'abord la probabilité cherchée sans la correction de continuité. Nous obtenons, par le théorème limite central,

$$\begin{aligned}
P\{650 \leqslant Y \leqslant 750\} &= P\left\{\frac{650 - 200 \cdot \dfrac{7}{2}}{\sqrt{\dfrac{35}{12}} \cdot \sqrt{200}} \leqslant \frac{Y - 200 \cdot \dfrac{7}{2}}{\sqrt{\dfrac{35}{12}} \cdot \sqrt{200}} \leqslant \frac{750 - 200 \cdot \dfrac{7}{2}}{\sqrt{\dfrac{35}{12}} \cdot \sqrt{200}}\right\} \\
&\simeq P\left\{-\frac{1}{7}\,\sqrt{2}\sqrt{3}\sqrt{5}\sqrt{7} \leqslant Z \leqslant \frac{1}{7}\,\sqrt{2}\sqrt{3}\sqrt{5}\sqrt{7}\right\} \\
&= 2\,\Phi(2{,}07) - 1 = 2 \times 0{,}9808 - 1 = 0{,}9616\,,
\end{aligned}$$

où Z est une v.a. $N(0,1)$ et Φ est sa fonction de répartition.

Lorsque nous considérons la correction de continuité, nous utilisons l'égalité $P\{650 \leqslant Y \leqslant 750\} = P\{649{,}5 \leqslant Y \leqslant 750{,}5\}$. Nous obtenons

$$\begin{aligned}
&P\{649{,}5 \leqslant Y \leqslant 750{,}5\}\\
&= P\left\{\frac{649{,}5 - 200\cdot\frac{7}{2}}{\sqrt{\frac{35}{12}}\cdot\sqrt{200}} \leqslant \frac{Y - 200\cdot\frac{7}{2}}{\sqrt{\frac{35}{12}}\cdot\sqrt{200}} \leqslant \frac{750{,}5 - 200\cdot\frac{7}{2}}{\sqrt{\frac{35}{12}}\cdot\sqrt{200}}\right\}\\
&\simeq P\left\{-\frac{1{,}01}{7}\sqrt{2}\sqrt{3}\sqrt{5}\sqrt{7} \leqslant Z \leqslant \frac{1{,}01}{7}\sqrt{2}\sqrt{3}\sqrt{5}\sqrt{7}\right\}\\
&= 2\,\Phi(2{,}09) - 1 = 2\times 0{,}9817 - 1 = 0{,}9634\,,
\end{aligned}$$

où Z est une v.a. $N(0,1)$ et Φ est sa fonction de répartition. Ainsi, nous constatons que lorsque l'on lance 200 fois un dé, le résultat se trouve proche de 700 avec une grande probabilité.

7.5 Les v.a. X_i suivent chacune une loi uniforme sur $[0,1]$. Ainsi, $E[X_1] = \frac{1}{2}$ et $\mathrm{Var}(X_1) = \frac{(1-0)^2}{12} = \frac{1}{12}$. Par le théorème limite central, nous pouvons écrire

$$\begin{aligned}
P\left\{\frac{S_n}{n} < 0{,}51\right\} = P\{S_n < 0{,}51\,n\} &= P\left\{\frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}} < \frac{0{,}51\,n - \frac{n}{2}}{\sqrt{\frac{n}{12}}}\right\}\\
&\simeq P\left\{Z < \frac{\frac{n}{100}}{\sqrt{\frac{n}{12}}}\right\} = \Phi\left(\frac{\sqrt{12n}}{100}\right),
\end{aligned}$$

où Z est une v.a. $N(0,1)$ et Φ est sa fonction de répartition. Cette probabilité doit être supérieure à 0,84 pour satisfaire la condition souhaitée. D'après la tabelle A.2 de l'appendice, ce sera le cas si

$$\frac{\sqrt{12\,n}}{100} > 1\,.$$

Cette condition est équivalente à $\sqrt{n} > \frac{100}{\sqrt{12}}$ ou encore $n > \frac{10\,000}{12} = \frac{2\,500}{3} = 833{,}3$. Le plus petit entier n tel que la probabilité cherchée est plus grande ou égale à 0,84 est $n = 834$.

7.6 Notons n le nombre de billets que la compagnie peut se permettre de vendre et X le nombre de passagers se présentant au départ de l'avion. Nous pouvons écrire $X = \sum_{i=1}^{n} X_i$, où $X_i = 1_{\{\text{le } i^{\text{ième}} \text{ passager se présente au départ}\}}$. Nous avons donc $E[X_i] = P\{\text{le } i^{\text{ième}} \text{ passager se présente}\} = 0{,}88$ et $\mathrm{Var}(X_i) =$

0,12 × 0,88. Nous devons trouver n de sorte que $P\{X \leqslant 169\} \geqslant 0{,}99$. En utilisant le théorème limite central, nous pouvons écrire

$$\begin{aligned}P\{X \leqslant 169\} &= P\left\{\frac{X - 0{,}88\,n}{\sqrt{0{,}12 \cdot 0{,}88}\sqrt{n}} \leqslant \frac{169 - 0{,}88\,n}{\sqrt{0{,}12 \cdot 0{,}88}\sqrt{n}}\right\} \\ &\simeq P\left\{Z \leqslant \frac{169 - 0{,}88\,n}{\sqrt{0{,}12 \cdot 0{,}88}\sqrt{n}}\right\},\end{aligned}$$

où Z est une v.a. $N(0,1)$. Cette probabilité doit être supérieure à 0,99 pour satisfaire la condition souhaitée. D'après la tabelle A.2 de l'appendice, ce sera le cas si

$$\frac{169 - 0{,}88\,n}{\sqrt{0{,}12 \cdot 0{,}88}\sqrt{n}} \geqslant 2{,}33\,.$$

Cette condition est équivalente à

$$169 - 0{,}88\,n \geqslant 2{,}33\,\sqrt{0{,}12 \cdot 0{,}88}\sqrt{n}$$

ou encore à

$$0{,}88\,n + 0{,}757\,\sqrt{n} - 169 \leqslant 0\,.$$

En posant $n = x^2$ et en calculant les zéros du polynôme $0{,}88\,x^2 + 0{,}757\,x - 169$, nous obtenons la condition

$$0 \leqslant \sqrt{n} \leqslant \frac{-0{,}757 + \sqrt{0{,}757^2 + 4 \cdot 0{,}88 \cdot 169}}{2 \cdot 0{,}88} = 13{,}43$$

ou encore $0 \leqslant n \leqslant 180{,}5$. La compagnie d'aviation peut donc vendre jusqu'à 180 billets pour son vol de 169 places en étant sûre à 99 % que chacun aura un siège dans l'avion.

$\boxed{\textbf{7.9}}$ Traitons tout d'abord le cas où $\mu = 0$. Dans ce cas, $Z_n = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n X_i$. Ainsi, pour $t \geqslant 0$,

$$M_{Z_n}(t) = E[e^{tZ_n}] = \prod_{i=1}^n E\big[e^{\frac{tX_i}{\sigma\sqrt{n}}}\big] = \prod_{i=1}^n M_{X_i}\left(\frac{t}{\sigma\sqrt{n}}\right) = \left(M_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n.$$

Or, pour $t \in\,]-t_0, t_0[$, en effectuant un développement limité autour de 0,

$$\begin{aligned}M_{X_1}(t) &= M_{X_1}(0) + M'_{X_1}(0)t + M''_{X_1}(0)\frac{t^2}{2} + o(t^2) \\ &= 1 + E[X_1]t + E[X_1^2]\frac{t^2}{2} + o(t^2) = 1 + \frac{\sigma^2 t^2}{2} + o(t^2)\,.\end{aligned}$$

Soit $t \geqslant 0$ fixé et soit n suffisament grand pour que $\frac{|t|}{\sigma\sqrt{n}} < t_0$. Alors,

$$M_{Z_n}(t) = \left(1 + \frac{t^2}{2n} + o\left(\frac{t^2}{\sigma^2 n}\right)\right)^n.$$

Selon l'indication de l'exercice 4.15,

$$\lim_{n\to\infty}\left(1+\frac{\lambda_n}{n}\right)^n = \exp\left(\lim_{n\to\infty}\lambda_n\right),$$

et donc

$$\lim_{n\to 0} M_{Z_n}(t) = \lim_{n\to\infty}\left(1+\frac{t^2}{2n}+o\left(\frac{t^2}{\sigma^2 n}\right)\right)^n = e^{\frac{t^2}{2}} = M_Z(t)\,,$$

avec Z une v.a. de loi $N(0,1)$.

Dans le cas où $\mu \neq 0$, on pose $Y_i = X_i - \mu$ pour tout $i \geqslant 0$. Les v.a. Y_i vérifient les hypothèses de l'énoncé. De plus, $E[Y_i] = 0$ et $Z_n = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n Y_i$. Ainsi, par la première partie de l'exercice, nous avons bien

$$\lim_{n\to\infty} M_{Z_n}(t) = M_Z(t)\,,$$

pour tout $t \geqslant 0$.

11.8 Chapitre 8

8.1 a) La densité marginale de la v.a. X est

$$\begin{aligned} P\{X=x\} &= \sum_{y=0}^{n-x} P\{X=x, Y=y\} \\ &= \binom{n}{x} p^x \sum_{y=0}^{n-x}\binom{n-x}{y} q^y (1-p-q)^{n-x-y} \\ &= \binom{n}{x} p^x (1-p)^{n-x}\,, \end{aligned}$$

pour $x = 0, \ldots, n$.

b) Nous en déduisons que, pour $x \in \{0, \ldots, n\}$ fixé,

$$\begin{aligned} P\{Y=y \mid X=x\} &= \frac{P\{X=x, Y=y\}}{P\{X=x\}} \\ &= \frac{\binom{n}{x}\binom{n-x}{y} p^x q^y (1-p-q)^{n-x-y}}{\binom{n}{x} p^x (1-p)^{n-x}} \\ &= \binom{n-x}{y}\left(\frac{q}{1-p}\right)^y \left(1-\frac{q}{1-p}\right)^{n-x-y}, \end{aligned}$$

pour $y = 0, \ldots, n-x$. Ainsi, la densité conditionnelle de Y sachant $X = x$ est une densité binomiale de paramètres $n-x$ et $\frac{q}{1-p}$. Remarquons que $\binom{n}{x}\binom{n-x}{y} = \binom{n}{y}\binom{n-y}{x}$. Ainsi, le problème est symétrique en x et y et un raisonnement analogue nous montre que $P\{X = x \mid Y = y\}$ est une densité binomiale de paramètres $n-y$ et $\frac{p}{1-q}$.

c) D'après le résultat du point b), nous pouvons utiliser la formule pour l'espérance d'une v.a. binomiale, ce qui donne

$$E[Y \mid X = x] = (n-x)\frac{q}{1-p} \qquad \text{et} \qquad E[X \mid Y = y] = (n-y)\frac{p}{1-q}\,.$$

d) Au point c), nous avons calculé $\varphi(x) = E[Y \mid X = x]$. Nous en déduisons que $E[Y|X] = \varphi(X) = \frac{q}{1-p}(n-X)$. Un raisonnement analogue nous donne $E[X \mid Y] = \frac{p}{1-q}(n-Y)$.

e) Par le point d), nous savons que $E[X \mid Y] = \frac{p}{1-q}(n-Y)$. De plus, par le point a), Y suit une loi binomiale de paramètres n et q. Ainsi,

$$P\big\{E[X \mid Y] = 0\big\} = P\left\{\frac{p}{1-q}\,(n-Y) = 0\right\} = P\{Y = n\} = q^n\,.$$

8.2 a) La v.a. $Z = X + Y$ suit une loi binomiale de paramètres $n+m$ et p. Calculons la densité conjointe de (Y, Z) :

$$\begin{aligned} P\{Y = y, Z = z\} &= P\{Y = y,\, X + Y = z\} = P\{Y = y,\, X = z - y\} \\ &= P\{Y = y\}P\{X = z - y\} \\ &= \binom{m}{y} p^y (1-p)^{m-y} \binom{n}{z-y} p^{z-y}(1-p)^{n-(z-y)}\,. \end{aligned}$$

Donc,

$$\begin{aligned} &P\{Y = y \mid Z = z\} \\ &= \frac{\binom{m}{y} p^y (1-p)^{m-y} \binom{n}{z-y} p^{z-y}(1-p)^{n-(z-y)}}{\binom{n+m}{z} p^z (1-p)^{n+m-z}} = \frac{\binom{m}{y}\binom{n}{z-y}}{\binom{n+m}{z}}\,. \end{aligned}$$

Ainsi, la densité conditionnelle $P\{Y = y \mid Z = z\}$ est la densité d'une v.a. hypergéométrique de paramètres m, n et z.

b) En utilisant le résultat de l'exercice 6.13 sur l'espérance d'une v.a. hypergéométrique, nous obtenons $\varphi(z) = E[Y \mid Z = z] = \frac{mz}{m+n}$, d'où $E[Y \mid Z] = \varphi(Z) = \frac{mZ}{m+n}$.

c) Par le point b), nous savons que $E[Y \mid Z] = \frac{mZ}{m+n}$. De plus, Z suit une loi binomiale de paramètres $n+m$ et p. Ainsi,

$$P\big\{E[Y \mid Z] = 0\big\} = P\left\{\frac{mZ}{m+n} = 0\right\} = P\{Z = 0\} = (1-p)^{n+m}\,.$$

8.3 a) La v.a. $Z = X_1 + X_2$ suit une loi de Poisson de paramètre $\lambda_1 + \lambda_2$. Calculons la densité conjointe de (X_2, Z).

$$\begin{aligned} P\{X_2 = x_2, Z = z\} &= P\{X_2 = x_2,\ X_1 + X_2 = z\} \\ &= P\{X_2 = x_2,\ X_1 = z - x_2\} \\ &= P\{X_2 = x_2\} P\{X_1 = z - x_2\} \\ &= e^{-\lambda_2} \frac{\lambda_2^{x_2}}{x_2!}\, e^{-\lambda_1} \frac{\lambda_1^{z-x_2}}{(z-x_2)!} \end{aligned}$$

Donc,

$$\begin{aligned} &P\{X_2 = x_2 \mid Z = z\} \\ &= \frac{e^{-\lambda_2} \frac{\lambda_2^{x_2}}{x_2!}\, e^{-\lambda_1} \frac{\lambda_1^{z-x_2}}{(z-x_2)!}}{e^{-\lambda_1-\lambda_2} \frac{(\lambda_1+\lambda_2)^z}{z!}} = \binom{z}{x_2} \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{x_2} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{z-x_2} . \end{aligned}$$

Ainsi, la densité conditionnelle $P\{X_2 = x_2 \mid Z = z\}$ est la densité d'une v.a. binomiale de paramètres z et $\frac{\lambda_2}{\lambda_1+\lambda_2}$.

b) En utilisant l'espérance d'une v.a. binomiale, nous obtenons $\varphi(z) = E[X_2 \mid Z = z] = \frac{\lambda_2 z}{\lambda_1+\lambda_2}$, d'où $E[X_2 \mid Z] = \varphi(Z) = \frac{\lambda_2 Z}{\lambda_1+\lambda_2}$.

c) Par le point b), nous savons que $E[X_2 \mid Z] = \frac{\lambda_2 Z}{\lambda_1+\lambda_2}$. De plus, Z suit une loi de Poisson de paramètre $\lambda_1 + \lambda_2$. Ainsi,

$$\begin{aligned} P\left\{E[X_2 \mid Z] \leqslant \frac{\lambda_2}{\lambda_1+\lambda_2}\right\} &= P\left\{\frac{\lambda_2 Z}{\lambda_1+\lambda_2} \leqslant \frac{\lambda_2}{\lambda_1+\lambda_2}\right\} \\ &= P\{Z \leqslant 1\} = (1 + \lambda_1 + \lambda_2)\, e^{-(\lambda_1+\lambda_2)} . \end{aligned}$$

8.4 a) Calculons la densité marginale de X :

$$f_X(x) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)\, dy = 10\, x^2 \int_0^x y\, dy = 10\, x^2 \left[\frac{y^2}{2}\right]_0^x = 5x^4 ,$$

pour $0 < x < 1$, $f_X(x) = 0$ sinon. Par conséquent, pour $0 < x < 1$, la densité conditionnelle de Y sachant $X = x$ est

$$f_Y(y \mid X = x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)} = \frac{10\, x^2 y}{5x^4} = \frac{2y}{x^2} ,$$

pour $0 < y < x$ et $f_Y(y \mid X = x) = 0$ sinon.

b) Pour $0 < x < 1$,

$$\begin{aligned} \varphi(x) = E[Y \mid X = x] &= \int_{-\infty}^{+\infty} y f_Y(y \mid X = x)\, dy = \frac{2}{x^2} \int_0^x y^2\, dy \\ &= \frac{2}{x^2} \left[\frac{y^3}{3}\right]_0^x = \frac{2}{3}\, x , \end{aligned}$$

et

$$E[Y^2 \mid X = x] = \int_{-\infty}^{+\infty} y^2 f_Y(y \mid X = x)\, dy = \frac{2}{x^2}\int_0^x y^3\, dy$$
$$= \frac{2}{x^2}\left[\frac{y^4}{4}\right]_0^x = \frac{1}{2}\,x^2 .$$

Ainsi,

$$\operatorname{Var}(Y \mid X = x) = E[Y^2 \mid X = x] + \big(E[Y \mid X = x]\big)^2$$
$$= \frac{x^2}{2} - \frac{4x^2}{9} = \frac{x^2}{18} .$$

A l'aide des résultats du point b), nous obtenons finalement $E[Y \mid X] = \varphi(X) = \frac{2}{3}X$ et $\operatorname{Var}(Y \mid X) = \frac{X^2}{18}$.

8.6 a) Par définition de la variance conditionnelle,

$$\operatorname{Var}(Y \mid X) = E\left[\big(Y - E[Y \mid X]\big)^2 \,\Big|\, X\right]$$
$$= E[Y^2 \mid X] + E\big[E[Y \mid X]^2 \mid X\big] - 2E\big[YE[Y \mid X] \mid X\big] .$$

Par le point b) de la proposition 8.15,

$$\operatorname{Var}(Y \mid X) = E[Y^2 \mid X] + \big(E[Y \mid X]\big)^2 - 2\big(E[Y \mid X]\big)^2$$
$$= E[Y^2 \mid X] - \big(E[Y \mid X]\big)^2 .$$

b) D'après les propositions 8.10, 8.12 et 8.14, pour toute v.a Y, $E[Y \mid X] = \varphi(X)$, où $\varphi(x) = E[Y \mid X = x]$. Ainsi, par le point a), $\operatorname{Var}(Y \mid X) = \psi(X)$, avec

$$\psi(x) = E[Y^2 \mid X = x] - \big(E[Y \mid X = x]\big)^2 .$$

c) Le point d) de la proposition 8.15 montre que si X et Y sont indépendantes, alors $E[Y \mid X] = E[Y]$. Comme Y^2 est indépendante de X si Y l'est, nous avons, par le point a),

$$\operatorname{Var}(Y \mid X) = E[Y^2 \mid X] - \big(E[Y \mid X]\big)^2 = E[Y^2] - \big(E[Y]\big)^2 = \operatorname{Var}(Y) ,$$

par le point a) de la proposition 6.15.

8.7 a) La v.a. 1_F est discrète et ne prend que les valeurs 0 et 1. Ainsi, dans le cas où X est discrète, d'après le théorème 6.4, nous avons

$$E[X1_F] = \sum_x \sum_{y=0}^{1} xyP\{X = x, 1_F = y\} = \sum_x x\, P\{X = x, 1_F = 1\}$$
$$= \sum_x x\, P\big(\{X = x\} \cap F\big) = P(F)E[X \mid F] ,$$

par le point b) de la définition 8.9. Le résultat est démontré.

b) Par définition de l'espérance conditionnelle, nous devons montrer que pour toute fonction g bornée, la v.a. $Z = E[X \mid F]\, 1_F + E[X \mid F^c]\, 1_{F^c}$ est une fonction de 1_F et vérifie $E\big[Xg(1_F)\big] = E\big[Zg(1_F)\big]$. Les termes $E[X \mid F]$ et $E[X \mid F^c]$ étant des nombres et sachant que $1_{F^c} = 1 - 1_F$, la v.a. Z est bien une fonction de 1_F. De plus, notons que pour toute fonction g bornée, nous avons

$$g(1_F) = g(0) + \big(g(1) - g(0)\big)\, 1_F\,.$$

Ainsi, par linéarité de l'espérance, il est suffisant de vérifier le résultat pour $g \equiv 1$ et pour $g(x) = x$. Dans le premier cas, par définition de $E[X \mid F]$, nous avons

$$\begin{aligned}E[Z] &= E\big[E[X \mid F]\, 1_F + E[X \mid F^c]\, 1_{F^c}\big]\\ &= E[X \mid F]P(F) + E[X \mid F^c]P(F^c) = E[X\, 1_F] + E[X\, 1_{F^c}] = E[X]\,.\end{aligned}$$

D'autre part, comme $1_F^2 = 1_F$ et $1_F 1_{F^c} = 0$,

$$E[Z1_F] = E\big[E[X \mid F]\, 1_F + E[X \mid F^c]\, 1_F 1_{F^c}\big] = E[X \mid F]P(F) = E[X\, 1_F]\,,$$

et le résultat est démontré.

$\boxed{\textbf{8.9}}$ a) Sachant que $X = x$, le nombre d'œufs éclos correspond au nombre de succès (l'œuf éclot) dans une suite d'expériences indépendantes avec probabilité de succès p. Ainsi, conditionnellement à $X = x$, Y suit une loi binomiale de paramètres x et p.

b) Tout d'abord, en utilisant l'espérance de la loi binomiale, nous observons que $E[Y \mid X = x] = px$. Ainsi, $E[Y \mid X] = pX$. On rappelle que l'espérance et la variance d'une v.a. de Poisson de paramètre λ sont données par $E[X] = \lambda$ et $\mathrm{Var}(X) = \lambda$. Par la partie a) de la proposition 8.15, nous avons

$$E[Y] = E\left[E[Y \mid X]\right] = E[pX] = pE[X] = p\lambda.$$

De plus, en utilisant les parties a) et b) de la proposition 8.15, nous avons

$$\begin{aligned}E[XY] &= E[E[XY \mid X]] = E[XE[Y \mid X]] = E[pX^2] = pE[X^2]\\ &= p\left(\mathrm{Var}(X) + E[X]^2\right) = p\left(\lambda + \lambda^2\right).\end{aligned}$$

c) En utilisant les propriétés de la loi binomiale et la proposition 8.19b), nous avons $\mathrm{Var}(Y \mid X = x) = xp(1-p)$. Nous déduisons que $\mathrm{Var}(Y \mid X) = p(1-p)X$. Par la proposition 8.20 et la partie b), nous avons

$$\begin{aligned}\mathrm{Var}(Y) &= E[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(E[Y \mid X]) = p(1-p)E[X] + \mathrm{Var}(pX)\\ &= p(1-p)E[X] + p^2\,\mathrm{Var}(X) = (p(1-p) + p^2)\lambda = p\lambda.\end{aligned}$$

d) La fonction de densité conjointe de (X, Y) est donnée par

$$f_{(X,Y)}(x, y) = f_X(x) f_Y(y \mid X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \cdot \binom{x}{y} p^y (1-p)^{x-y},$$

pour $x, y \in \mathbb{N}$ avec $y \leqslant x$ et $f_{(X,Y)}(x, y) = 0$ sinon.

$\boxed{\textbf{8.10}}$ a) Etant donné que la loi conditionnelle de X sachant $\Theta = \theta$ est $B(n, \theta)$, nous savons, par les résultats sur l'espérance et la variance d'une v.a. binomiale, que :

$$E[X \mid \Theta = \theta] = n\theta \qquad \text{et} \qquad \operatorname{Var}(X \mid \Theta = \theta) = n\theta(1-\theta)$$

(voir en particulier le point b) de la proposition 8.19). Ainsi,

$$E[X \mid \Theta] = n\Theta \qquad \text{et} \qquad \operatorname{Var}(X \mid \Theta) = n\Theta(1-\Theta).$$

b) Sachant que Θ est une v.a. uniforme sur $[0,1]$, nous avons

$$E[\Theta] = \frac{1}{2} \qquad \text{et} \qquad \operatorname{Var}(\Theta) = \frac{1}{12}.$$

Par la proposition 8.15, nous avons

$$E[X] = E[E[X \mid \Theta]] = E[n\Theta] = nE[\Theta] = \frac{n}{2}.$$

De plus, par la proposition 8.20, nous avons

$$\begin{aligned}
\operatorname{Var}(X) &= \operatorname{Var}(E[X \mid \Theta]) + E[\operatorname{Var}(X \mid \Theta)] = \operatorname{Var}(n\Theta) + E[n\Theta(1-\Theta)] \\
&= n^2 \operatorname{Var}(\Theta) + nE[\Theta] - nE[\Theta^2] = (n^2 - n)\operatorname{Var}(\Theta) + n(E[\Theta] - E[\Theta]^2) \\
&= \frac{n^2 - n}{12} + \frac{n}{4} = \frac{n(n+2)}{12}.
\end{aligned}$$

11.9 Chapitre 9

$\boxed{\textbf{9.29}}$ a) Nous allons montrer que $G = \bigcap_{k\in\mathbb{N}^*} \bigcup_{N\in\mathbb{N}} \bigcap_{n>N} \{|X_n - X| < \frac{1}{k}\}$ en montrant la double inclusion. Premièrement, soit $\omega \in G$. Alors, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$. Par définition de la limite, pour tout $k \in \mathbb{N}^*$, il existe $N \in \mathbb{N}$ tel que pour tout $n > N$, $\big|X_n(\omega) - X(\omega)\big| < \frac{1}{k}$. Donc

$$\omega \in \bigcap_{k\in\mathbb{N}^*} \bigcup_{N\in\mathbb{N}} \bigcap_{n>N} \left\{|X_n - X| < \frac{1}{k}\right\}.$$

Réciproquement, soit $\omega \in \bigcap_{k\in\mathbb{N}^*} \bigcup_{N\in\mathbb{N}} \bigcap_{n>N} \left\{|X_n - X| < \frac{1}{k}\right\}$. Alors, pour tout $\epsilon > 0$, il existe $k \in \mathbb{N}^*$ tel que $\frac{1}{k} < \epsilon$. Ainsi, par hypothèse, il existe $N \in \mathbb{N}$ tel que, pour tout $n > N$, $\big|X_n(\omega) - X(\omega)\big| < \frac{1}{k} < \epsilon$. Par définition de la limite, nous avons bien $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ et $\omega \in G$.

b) Nous avons $G^c = \bigcup_{k\in\mathbb{N}^*} \bigcap_{N\in\mathbb{N}} \bigcup_{n>N} \left\{|X_n - X| \geqslant \frac{1}{k}\right\}$. On pose alors

$$A_{n,k} = \left\{|X_n - X| \geqslant \frac{1}{k}\right\}, \qquad B_{N,k} = \bigcup_{n>N} A_{n,k} \quad \text{et} \quad C_k = \bigcap_{N\in\mathbb{N}} B_{N,k}\,.$$

Ainsi, $G^c = \bigcup_{k\in\mathbb{N}^*} C_k$. Tout d'abord, nous constatons que $C_k \subset C_{k+1}$ et donc, par la continuité de la probabilité, $P(G^c) = \lim_{k\to\infty} P(C_k)$. De même, nous constatons que $B_{N+1,k} \subset B_{N,k}$ et donc $P(C_k) = \lim_{N\to\infty} P(B_{N,k})$. D'autre part,

$$P(B_{N,k}) = P\left(\bigcup_{n>N} A_{n,k}\right) \leqslant \sum_{n=N+1}^{\infty} P(A_{n,k})\,.$$

Finalement,

$$0 \leqslant P(G^c) \leqslant \lim_{k\to\infty}\lim_{N\to\infty}\sum_{n=N+1}^{\infty} P(A_{n,k}) = 0\,,$$

car $\sum_{n=0}^{\infty} P(A_{n,k}) < +\infty$ pour tout $k \in \mathbb{N}^*$. Donc, $P(G^c) = 0$ et $P(G) = 1$.

Appendice

A.1 Principales lois de probabilité

Table A.1 Tabelle des principales v.a.

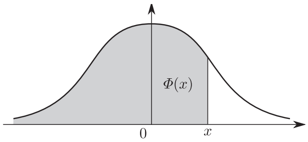
Variables aléatoires discrètes

Loi de la v.a X	$f_X(k)$	$E[X]$	$\mathrm{Var}(X)$	$M_X(t)$
Binomiale (n, p)	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	$(p\,e^t + 1 - p)^n$
Poisson (λ)	$e^{-\lambda} \dfrac{\lambda^k}{k!}$	λ	λ	$\exp\big(\lambda(e^t - 1)\big)$
Géométrique (p)	$p(1-p)^{k-1}$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$	$\dfrac{p\,e^t}{1-(1-p)e^t}$
Binomiale négative (r, p)	$\dbinom{k-1}{r-1} p^r (1-p)^{k-r}$	$\dfrac{r}{p}$	$\dfrac{r(1-p)}{p^2}$	$\left(\dfrac{p\,e^t}{1-(1-p)\,e^t}\right)^r$

Variables aléatoires continues

Loi de la v.a X	$f_X(x)$, $x \in \mathbb{R}$	$E[X]$	$\mathrm{Var}(X)$	$M_X(t)$
Uniforme sur $[a, b]$	$\dfrac{1}{b-a}\, 1_{[a,b]}(x)$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Exponentielle (λ)	$\lambda e^{-\lambda x}\, 1_{[0,\infty[}(x)$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	$\dfrac{\lambda}{\lambda - t}$
Normale (μ, σ^2)	$\dfrac{1}{\sqrt{2\pi}\sigma} \exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ^2	$\exp\left(\mu t + \dfrac{\sigma^2 t^2}{2}\right)$
Gamma (p, λ)	$\dfrac{\lambda e^{-\lambda x} (\lambda x)^{p-1}}{\Gamma(p)}\, 1_{[0,\infty[}(x)$	$\dfrac{p}{\lambda}$	$\dfrac{p}{\lambda^2}$	$\left(\dfrac{\lambda}{\lambda - t}\right)^p$

A.2 Fonction de répartition de la loi normale



$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy\,, \quad x \in \mathbb{R}\,.$$

Table A.2 Tabelle de la fonction de répartition normale.

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,5000	0,50	0,6915	1,00	0,8413	1,50	0,9332	2,00	0,9772	2,50	0,9938	3,00	0,9987
0,01	0,5040	0,51	0,6950	1,01	0,8438	1,51	0,9345	2,01	0,9778	2,51	0,9940	3,01	0,9987
0,02	0,5080	0,52	0,6985	1,02	0,8461	1,52	0,9357	2,02	0,9783	2,52	0,9941	3,02	0,9987
0,03	0,5120	0,53	0,7019	1,03	0,8485	1,53	0,9370	2,03	0,9788	2,53	0,9943	3,03	0,9988
0,04	0,5160	0,54	0,7054	1,04	0,8508	1,54	0,9382	2,04	0,9793	2,54	0,9945	3,04	0,9988
0,05	0,5199	0,55	0,7088	1,05	0,8531	1,55	0,9394	2,05	0,9798	2,55	0,9946	3,05	0,9989
0,06	0,5239	0,56	0,7123	1,06	0,8554	1,56	0,9406	2,06	0,9803	2,56	0,9948	3,06	0,9989
0,07	0,5279	0,57	0,7157	1,07	0,8577	1,57	0,9418	2,07	0,9808	2,57	0,9949	3,07	0,9989
0,08	0,5319	0,58	0,7190	1,08	0,8599	1,58	0,9429	2,08	0,9812	2,58	0,9951	3,08	0,9990
0,09	0,5359	0,59	0,7224	1,09	0,8621	1,59	0,9441	2,09	0,9817	2,59	0,9952	3,09	0,9990
0,10	0,5398	0,60	0,7257	1,10	0,8643	1,60	0,9452	2,10	0,9821	2,60	0,9953	3,10	0,9990
0,11	0,5438	0,61	0,7291	1,11	0,8665	1,61	0,9463	2,11	0,9826	2,61	0,9955	3,11	0,9991
0,12	0,5478	0,62	0,7324	1,12	0,8686	1,62	0,9474	2,12	0,9830	2,62	0,9956	3,12	0,9991
0,13	0,5517	0,63	0,7357	1,13	0,8708	1,63	0,9484	2,13	0,9834	2,63	0,9957	3,13	0,9991
0,14	0,5557	0,64	0,7389	1,14	0,8729	1,64	0,9495	2,14	0,9838	2,64	0,9959	3,14	0,9992
0,15	0,5596	0,65	0,7422	1,15	0,8749	1,65	0,9505	2,15	0,9842	2,65	0,9960	3,15	0,9992
0,16	0,5636	0,66	0,7454	1,16	0,8770	1,66	0,9515	2,16	0,9846	2,66	0,9961	3,16	0,9992
0,17	0,5675	0,67	0,7486	1,17	0,8790	1,67	0,9525	2,17	0,9850	2,67	0,9962	3,17	0,9992
0,18	0,5714	0,68	0,7517	1,18	0,8810	1,68	0,9535	2,18	0,9854	2,68	0,9963	3,18	0,9993
0,19	0,5753	0,69	0,7549	1,19	0,8830	1,69	0,9545	2,19	0,9857	2,69	0,9964	3,19	0,9993
0,20	0,5793	0,70	0,7580	1,20	0,8849	1,70	0,9554	2,20	0,9861	2,70	0,9965	3,20	0,9993
0,21	0,5832	0,71	0,7611	1,21	0,8869	1,71	0,9564	2,21	0,9864	2,71	0,9966	3,21	0,9993
0,22	0,5871	0,72	0,7642	1,22	0,8888	1,72	0,9573	2,22	0,9868	2,72	0,9967	3,22	0,9994
0,23	0,5910	0,73	0,7673	1,23	0,8907	1,73	0,9582	2,23	0,9871	2,73	0,9968	3,23	0,9994
0,24	0,5948	0,74	0,7704	1,24	0,8925	1,74	0,9591	2,24	0,9875	2,74	0,9969	3,24	0,9994
0,25	0,5987	0,75	0,7734	1,25	0,8944	1,75	0,9599	2,25	0,9878	2,75	0,9970	3,25	0,9994
0,26	0,6026	0,76	0,7764	1,26	0,8962	1,76	0,9608	2,26	0,9881	2,76	0,9971	3,26	0,9994
0,27	0,6064	0,77	0,7794	1,27	0,8980	1,77	0,9616	2,27	0,9884	2,77	0,9972	3,27	0,9995
0,28	0,6103	0,78	0,7823	1,28	0,8997	1,78	0,9625	2,28	0,9887	2,78	0,9973	3,28	0,9995
0,29	0,6141	0,79	0,7852	1,29	0,9015	1,79	0,9633	2,29	0,9890	2,79	0,9974	3,29	0,9995
0,30	0,6179	0,80	0,7881	1,30	0,9032	1,80	0,9641	2,30	0,9893	2,80	0,9974	3,30	0,9995
0,31	0,6217	0,81	0,7910	1,31	0,9049	1,81	0,9649	2,31	0,9896	2,81	0,9975	3,31	0,9995
0,32	0,6255	0,82	0,7939	1,32	0,9066	1,82	0,9656	2,32	0,9898	2,82	0,9976	3,32	0,9995
0,33	0,6293	0,83	0,7967	1,33	0,9082	1,83	0,9664	2,33	0,9901	2,83	0,9977	3,33	0,9996
0,34	0,6331	0,84	0,7995	1,34	0,9099	1,84	0,9671	2,34	0,9904	2,84	0,9977	3,34	0,9996
0,35	0,6368	0,85	0,8023	1,35	0,9115	1,85	0,9678	2,35	0,9906	2,85	0,9978	3,35	0,9996
0,36	0,6406	0,86	0,8051	1,36	0,9131	1,86	0,9686	2,36	0,9909	2,86	0,9979	3,36	0,9996
0,37	0,6443	0,87	0,8078	1,37	0,9147	1,87	0,9693	2,37	0,9911	2,87	0,9979	3,37	0,9996
0,38	0,6480	0,88	0,8106	1,38	0,9162	1,88	0,9699	2,38	0,9913	2,88	0,9980	3,38	0,9996
0,39	0,6517	0,89	0,8133	1,39	0,9177	1,89	0,9706	2,39	0,9916	2,89	0,9981	3,39	0,9997
0,40	0,6554	0,90	0,8159	1,40	0,9192	1,90	0,9713	2,40	0,9918	2,90	0,9981	3,40	0,9997
0,41	0,6591	0,91	0,8186	1,41	0,9207	1,91	0,9719	2,41	0,9920	2,91	0,9982	3,41	0,9997
0,42	0,6628	0,92	0,8212	1,42	0,9222	1,92	0,9726	2,42	0,9922	2,92	0,9982	3,42	0,9997
0,43	0,6664	0,93	0,8238	1,43	0,9236	1,93	0,9732	2,43	0,9925	2,93	0,9983	3,43	0,9997
0,44	0,6700	0,94	0,8264	1,44	0,9251	1,94	0,9738	2,44	0,9927	2,94	0,9984	3,44	0,9997
0,45	0,6736	0,95	0,8289	1,45	0,9265	1,95	0,9744	2,45	0,9929	2,95	0,9984	3,45	0,9997
0,46	0,6772	0,96	0,8315	1,46	0,9279	1,96	0,9750	2,46	0,9931	2,96	0,9985	3,46	0,9997
0,47	0,6808	0,97	0,8340	1,47	0,9292	1,97	0,9756	2,47	0,9932	2,97	0,9985	3,47	0,9997
0,48	0,6844	0,98	0,8365	1,48	0,9306	1,98	0,9761	2,48	0,9934	2,98	0,9986	3,48	0,9997
0,49	0,6879	0,99	0,8389	1,49	0,9319	1,99	0,9767	2,49	0,9936	2,99	0,9986	3,49	0,9998

Bibliographie

[1] P. Billingsley, *Probability and Measure*. Wiley, 1986.

[2] N. Bouleau, *Probabilités de l'ingénieur*. Hermann, 1986.

[3] R. C. Dalang, *Une démonstration élémentaire du théorème central limite.* Elem. Math. 61 (2006), n^o 2, pp. 65-73.

[4] R. Durrett, *The Essentials of Probability.* Duxbury Press, 1994.

[5] R. Durrett, *Probability : Theory and Examples.* Second Edition. Duxbury Press, 1995.

[6] G. Grimmett et D. Stirzaker, *Probability and Random Processes.* Third Edition. Oxford University Press, 2001.

[7] S. Morgenthaler, *Introduction à la statistique.* 3^e Edition. Presses polytechniques et universitaires Romandes, 2007.

[8] S. M. Ross, *Initiation aux probabilités.* Presses polytechniques et universitaires Romandes, 2007.

[9] M. Sanz i Solé, *Probabilitats.* Edicions de la Universitat de Barcelona, 1999.

[10] D. Stirzaker, *Elementary Probability.* Cambridge University Press, 1994.

Index