

# **Cours de Probabilités - MDI 104–114**

P. Bianchi, T. Bonald, L. Decreusefond, G. Fort, J. Najim

2017–2018



# Table des matières

<b>Introduction</b>	<b>7</b>
<b>Guide de lecture</b>	<b>11</b>
<b>1. Evénements</b>	<b>13</b>
1.1. Définitions	13
1.2. Probabilités sur un espace discret	14
1.3. Conditionnement et indépendance	18
1.4. Exercices	23
<b>2. Variables aléatoires discrètes</b>	<b>27</b>
2.1. Loi d'une variable discrète	27
2.2. Indépendance des v.a. discrètes	29
2.3. Espérance, moments	31
2.4. Fonction génératrice d'une v.a. à valeurs entières	38
2.5. Exercices	40
<b>3. Chaînes de Markov</b>	<b>47</b>
3.1. Préliminaires	47
3.2. Définition	48
3.3. Matrice de transition	49
3.4. Graphe de transition	51
3.5. Irréductibilité	51
3.6. Périodicité	54
3.7. Théorème ergodique	59
3.8. Exercices	64
<b>4. Mesures et intégration</b>	<b>69</b>
4.1. Introduction	69
4.2. Tribus	70
4.3. Fonctions mesurables	72
4.4. Mesures	76
4.5. Intégration	85
4.6. Exercices	102
<b>5. Variables et vecteurs aléatoires réels</b>	<b>111</b>
5.1. Généralités	111
5.2. Variables aléatoires réelles	112
5.3. Vecteurs aléatoires	120

5.4. Changement de variables . . . . .	126
5.5. Exercices . . . . .	132
<b>6. Loi conditionnelle</b>	<b>149</b>
6.1. Cas général . . . . .	150
6.2. Calculs . . . . .	152
6.3. Espérance conditionnelle . . . . .	155
6.4. Exercices . . . . .	157
<b>7. Fonction caractéristique</b>	<b>161</b>
7.1. Définition et propriétés . . . . .	161
7.2. Fonctions caractéristiques de v.a. usuelles . . . . .	164
7.3. Caractérisation de la loi . . . . .	166
7.4. Caractérisation de l'indépendance . . . . .	166
7.5. Calcul de moments . . . . .	167
7.6. Exercices . . . . .	171
<b>8. Vecteurs gaussiens</b>	<b>173</b>
8.1. Préliminaires . . . . .	173
8.2. Définition, propriétés . . . . .	173
8.3. Caractérisation de l'indépendance . . . . .	175
8.4. Stabilité par transformation affine . . . . .	176
8.5. Somme de vecteurs gaussiens indépendants . . . . .	177
8.6. La loi d'un vecteur gaussien admet-elle une densité ? . . . . .	178
8.7. Exercices . . . . .	180
<b>9. Convergences</b>	<b>183</b>
9.1. Loi des grands nombres . . . . .	183
9.2. Limite centrée . . . . .	183
9.3. Exercices . . . . .	185
<b>10. Simulation</b>	<b>187</b>
10.1. Générateurs de nombres aléatoires . . . . .	187
10.2. Méthode d'inversion de la fonction de répartition . . . . .	189
10.3. Méthode du rejet . . . . .	191
10.4. Méthodes de Monte Carlo . . . . .	193
10.5. Histogrammes . . . . .	195
10.6. Exercices . . . . .	196
<b>Annexes</b>	<b>199</b>
<b>A. Ensembles</b>	<b>201</b>
A.1. Opérations sur les ensembles . . . . .	201
A.2. Espaces d'états dénombrables . . . . .	203
<b>B. Notions utiles d'analyse</b>	<b>205</b>
B.1. Limite supérieure et limite inférieure . . . . .	205
B.2. Séries . . . . .	206

B.3. Convexité . . . . .	207
--------------------------	-----



# Notations

$\perp\!\!\!\perp$	Indépendance d'événement, indépendance de variables aléatoires.
$(\cdot)^T$	Transposée d'une matrice ou d'un vecteur.
$\circ$	Composition.
$ \Omega $	Cardinal d'un ensemble $\Omega$ .
$a_n \uparrow a$	La suite réelle $(a_n)_n$ est croissante et converge vers $a$ .
$a_n \downarrow a$	La suite réelle $(a_n)_n$ est décroissante et converge vers $a$ .
$A_n \uparrow A$	La suite d'ensembles $(A_n)_n$ est croissante et converge vers $A$ <i>i.e.</i> , $\cup_n A_n = A$ .
$A_n \downarrow A$	La suite d'ensembles $(A_n)_n$ est décroissante et converge vers $A$ <i>i.e.</i> , $\cap_n A_n = A$ .
$\det$	Déterminant.
$I_n$	Matrice identité de taille $n$ .
$\mathcal{P}(\Omega)$ ou $2^\Omega$	Tribu des parties sur $\Omega$ .
$\text{Tr}$	Trace.
$x \vee y$	Maximum de $x$ et de $y$ .
$x \wedge y$	Minimum de $x$ et de $y$ .
$\text{argmin}_{x \in C} f(x)$	La valeur de la variable $x$ (supposée atteinte et unique) en laquelle $f$ est minimale sur $C$ .





# Introduction

Les premières formalisations des probabilité datent du XVIII<sup>e</sup> siècle avec les travaux de Jacob Bernoulli (1713) et de Abraham de Moivre (1718). La probabilité d'un événement y était définie comme le rapport du nombre de cas favorables sur le nombre total de cas. Au début du XIX<sup>e</sup> siècle, les « probabilités géométriques » firent leur apparition. Dans ce cadre, la probabilité d'un événement s'exprime comme un rapport de volumes ou d'aires. Ces approches permettaient de faire bon nombre de calculs mais butaient sur certains paradoxes.

Les probabilités sont au départ, une tentative de représentation mathématique de l'incertain. Elles doivent être tout à la fois suffisamment formalisées pour permettre des calculs justes et rigoureux et garder une connexion forte et immédiate avec les phénomènes « physiques » analysés. Cette tension a longtemps posé des problèmes. Notamment, à la fin du XIX<sup>e</sup>, se posait le problème des événements « presque certains » ou « presque impossibles » : y-a-t'il un seuil en dessous duquel un événement de probabilité inférieure à ce seuil ne peut se réaliser ?

Au début du XX<sup>e</sup>, David Hilbert assigna aux mathématiciens, 23 problèmes, ou plutôt 23 défis, pour les années à venir. Parmi ceux-ci figurait l'axiomatisation de la « physique » par laquelle il fallait entendre l'axiomatisation des probabilités.

Le formalisme correct ne se fit jour qu'en 1930 dans les travaux d'Andreï Kolmogorov, qui réussit la synthèse des réflexions de Émile Borel, Jacques Hadamard, Maurice Fréchet et Paul Lévy entre autres.

Le concept de mesure permet d'avoir une vision unifiée des probabilités discrètes et des probabilités dites « continues ». Le vocabulaire de l'intégration permet de simplifier la présentation des différentes notions probabilistes. Par ailleurs, ainsi que l'illustre le deuxième paradoxe de Bertrand, la modélisation de certains phénomènes même simples impose de comprendre finement les liens entre théorie et interprétation physique. Enfin, la simulation, outil indispensable tellement est grande la complexité des systèmes, requiert de « construire » des variables et des processus aléatoires. Tout cela ne peut se faire sans une solide compréhension de la théorie sous-jacente.



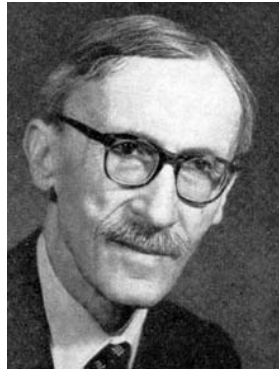
É. Borel (1871-1956),



M. Fréchet (1878-1973)



J. Hadamard (1865-1963),



P. Lévy (1886-1971)



A. Kolmogorov (1903-1987) (DR)

Ce document garde les traces de ses précédentes versions avec un chapitre sur les probabilités discrètes. Comme ces notions sont supposées connues, elles ont été mises à titre de rappel et pour fixer les notations.

# Guide de lecture

Ce support de cours est destinée à plusieurs populations d'étudiants qui se distinguent par leur prérequis en mathématiques générales et en probabilités. La lecture n'est donc pas la même pour tous.

Pour ceux qui connaissent déjà les probabilités discrètes, le cheminement normal du cours (et donc de votre lecture) est le suivant :

- Chaînes de Markov, chapitre 3
- Mesure et intégration, chapitre 4
- Variables et vecteurs aléatoires, chapitre 5
- Loi conditionnelle, chapitre 6
- Fonctions caractéristiques, 7
- Vecteurs gaussiens, 8

Les deux chapitres 9 et 10 peuvent être abordés avant celui sur les fonctions caractéristiques ou en fin de programme après les vecteurs gaussiens.

**Avertissement :** Le lecteur pressé peut être tenté de sauter le chapitre « mesures et intégration ». ECela peut se faire en première lecture mais les renvois à ce chapitre tant du point de vue des notations que des concepts sont fréquents. Ne pas le lire risque d'obscurcir votre compréhension tant du cours de probabilités que du cours d'analyse MDI 103 dont il constitue le fondement.

Pour ceux qui ne maîtrisent pas les manipulations de base sur les ensembles, il est recommandé de se référer à l'annexe A. Les résultats prérequis d'analyse générale notamment sur les séries sont accessibles dans l'annexe B. Les exercices des chapitres sur les probabilités discrètes et la théorie de la mesure sont tous corrigés sur la version en ligne de cet opus, accessible depuis le site pédagogique.



# 1. Événements

## 1.1. Définitions

Une *expérience aléatoire* est une expérience pouvant conduire à plusieurs résultats possibles. Formellement, une expérience aléatoire se décrit par la donnée de l'ensemble  $\Omega$  des résultats possibles. L'ensemble  $\Omega$  est appelé l'*univers* ou l'*espace des états*.

Traditionnellement, un résultat possible de l'expérience est noté  $\omega$ . C'est un élément de l'univers  $\Omega$ . Un tel élément  $\omega \in \Omega$  est parfois appelé une *épreuve* ou une *issue*.

EXEMPLE 1.— a) Jet d'un dé :  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .

b) Deux lancers consécutifs d'une pièce. L'univers est  $\Omega = \{PP, PF, FP, FF\}$  où  $P$  et  $F$  signifient respectivement pile et face.

c) Durée de fonctionnement sans panne d'une machine :  $\Omega = [0, +\infty[$ .

d) Valeur d'un signal continu sur un intervalle de temps  $[t_0, t_1]$  :  $\Omega = \mathcal{C}_b([t_0, t_1])$  est l'ensemble des fonctions continues de  $[t_0, t_1]$  dans  $\mathbf{R}$ .

Un *événement aléatoire* est un événement dont la réalisation dépend du résultat de l'expérience. Formellement, un événement aléatoire se décrit comme un sous-ensemble de  $\Omega$ .

EXEMPLE 2.— Considérons à nouveau les exemples précédents.

a)  $\Omega = \{1, 2, \dots, 6\}$ . L'événement  $A = \ll \text{Le résultat est pair} \gg$  s'identifie au sous-ensemble  $A = \{2, 4, 6\}$ .

b)  $\Omega = \{PP, PF, FP, FF\}$ . L'événement  $A = \ll \text{on obtient deux faces identiques} \gg$  s'identifie au sous-ensemble  $A = \{PP, FF\}$ .

c)  $\Omega = [0, \infty[$ . L'événement  $A = \ll \text{La machine fonctionne au moins } x \text{ unités de temps} \gg$  s'identifie à  $A = [x, +\infty[$ .

d)  $\Omega = \mathcal{C}_b([t_0, t_1])$ . L'événement  $A = \ll \text{L'amplitude du signal n'excède pas } \alpha \gg$  s'écrit  $A = \{\omega \in \Omega : \sup_{t \in [t_0, t_1]} |\omega(t)| \leq \alpha\}$ .

Pour une issue donnée  $\omega \in \Omega$ , on dit qu'un événement  $A$  est *réalisé* si  $\omega \in A$ .

L'espace d'état  $\Omega$  est aussi appelé l'*événement certain* : il est réalisé quelle que soit l'issue.

L'ensemble vide  $\emptyset$  est aussi appelé l'*événement impossible* : il n'est jamais réalisé.

La notation suivante sera d'un usage constant.

**Définition 1.1.** Soit  $\Omega$  un espace d'état et  $A \subset \Omega$  un ensemble. La fonction indicatrice de  $A$  est définie par :

$$\begin{aligned} 1_A : \Omega &\rightarrow \{0, 1\} \\ \omega &\mapsto 1 \text{ si } \omega \in A, 0 \text{ sinon.} \end{aligned}$$

L'exercice 1.1 fournit quelques propriétés importantes de l'indicatrice.

## 1.2. Probabilités sur un espace discret

Comme nous l'avons vu au début de ce chapitre, beaucoup d'expériences aléatoires peuvent être décrites par un univers  $\Omega$  fini ou dénombrable. Citons comme exemples immédiats le tirage à pile ou face ( $\Omega = \{P, F\}$ ), le lancer de dé ( $\Omega = \{1, \dots, 6\}$ ), le nombre de requêtes reçues par un serveur en une unité de temps ( $\Omega = \mathbf{N}$ ), etc.

### 1.2.1. Définition

**Définition 1.2.** Une mesure  $\mu$ , sur un ensemble  $E$  au plus dénombrable, est une application de  $\mathcal{P}(E)$ , l'ensemble des parties de  $E$ , dans  $\mathbb{R}$  qui satisfait les deux propriétés suivantes :

- $\mu(\emptyset) = 0$ ,
- pour toute famille  $(A_j, j \in \mathbf{N}^*)$  de parties deux à deux disjointes de  $E$ ,

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{+\infty} \mu(A_j). \quad (1.1)$$

Les parties de  $E$  s'appellent plus souvent des « événements ».

**Définition 1.3.** Une mesure  $\mu$  est dite mesure de probabilité (ou probabilité) lorsque  $\mu(E) = 1$ . Dans ce cas, on la note usuellement  $\mathbf{P}$  et non  $\mu$ .

### 1.2.2. Propriétés générales

**Proposition 1.1.** Soient  $A, B, (A_n)_{n \in \mathbf{N}^*}$  des ensembles.

- a)  $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$ .
- b)  $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$ .
- c) Si  $A \subset B$ , alors  $\mathbf{P}(A) \leq \mathbf{P}(B)$ .

d) Si  $(A_n)_{n \in \mathbf{N}^*}$  est une partition de  $\Omega$ , alors

$$\mathbf{P}(B) = \sum_{n=1}^{\infty} \mathbf{P}(A_n \cap B) .$$

e) Si  $A_n \uparrow A$ , alors  $\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$ .

Si  $A_n \downarrow A$ , alors  $\mathbf{P}(A) = \lim_{n \rightarrow \infty} \mathbf{P}(A_n)$ .

f) Si  $\mathbf{P}(A_n) = 1$  pour tout  $n \in \mathbf{N}^*$ , alors  $\mathbf{P}(\bigcap_{n=1}^{\infty} A_n) = 1$ .

g) Pour une famille quelconque  $(A_n)_{n \in \mathbf{N}^*}$  dans  $\mathcal{F}$ , on a la *borne de l'union* :

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbf{P}(A_n) .$$

*Démonstration.* a) On applique l'axiome de  $\sigma$ -additivité (1.1) en posant  $A_1 = A$ ,  $A_2 = A^c$  et  $A_n = \emptyset$  pour tout  $n \geq 3$ . Il en résulte que  $1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup A^c \cup \emptyset \cup \emptyset \cup \dots) = \mathbf{P}(A) + \mathbf{P}(A^c) + 0 + 0 + \dots$  et finalement  $1 = \mathbf{P}(A) + \mathbf{P}(A^c)$ .

b) On écrit que  $A \cup B$  s'écrit comme l'union disjointe  $(A \setminus B) \cup (B \setminus A) \cup (A \cap B)$ . La règle de  $\sigma$ -additivité conduit à :

$$\mathbf{P}(A \cup B) = \mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus A) + \mathbf{P}(A \cap B) . \quad (1.2)$$

Par ailleurs,  $A$  s'écrit comme l'union disjointe  $(A \setminus B) \cup (A \cap B)$  et donc  $\mathbf{P}(A) = \mathbf{P}(A \setminus B) + \mathbf{P}(A \cap B)$ . De même,  $\mathbf{P}(B) = \mathbf{P}(B \setminus A) + \mathbf{P}(A \cap B)$ . On a donc :  $\mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus A) = \mathbf{P}(A) + \mathbf{P}(B) - 2\mathbf{P}(A \cap B)$ . En faisant la substitution dans (1.2), nous obtenons le résultat.

c) Si  $A \subset B$ , on a en particulier :  $B = A \cup (B \setminus A)$  et comme l'union est disjointe,  $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \geq \mathbf{P}(A)$ .

d) Comme les  $(A_n)$  sont deux à deux disjoints, il en est de même pour les événements  $(A_n \cap B)$ . Par  $\sigma$ -additivité, on a  $\sum_n \mathbf{P}(A_n \cap B) = \mathbf{P}(\bigcup_n (A_n \cap B)) = \mathbf{P}((\bigcup_n A_n) \cap B) = \mathbf{P}(B)$ , où la dernière égalité provient du fait que  $\bigcup_n A_n = \Omega$ .

e) Soit  $A_n \uparrow A$ . On introduit la suite  $(B_n)$  définie par récurrence de la façon suivante :  $B_1 = A_1$  et  $B_{n+1} = A_{n+1} \setminus B_n$ . On vérifie sans peine que les  $(B_n)$  sont deux à deux disjoints, ce qui implique :

$$\mathbf{P}\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} \mathbf{P}(B_k) . \quad (1.3)$$

On vérifie également que pour tout  $n$ ,  $A_n = \bigcup_{k=1}^n B_k$ , et donc, par passage à la limite,  $A = \bigcup_{k=1}^{\infty} B_k$ . Ainsi, le membre de gauche de (1.3) n'est autre que  $\mathbf{P}(A)$ . Le membre de droite se réécrit comme la limite quand  $n \rightarrow \infty$  de la suite  $\sum_{k=1}^n \mathbf{P}(B_k)$ . Mais comme les  $(B_k)$  sont deux à deux disjoints,  $\sum_{k=1}^n \mathbf{P}(B_k) = \mathbf{P}(\bigcup_{k=1}^n B_k) = \mathbf{P}(A_n)$ . On a donc bien montré que  $\mathbf{P}(A) = \lim_n \mathbf{P}(A_n)$ .

Soit maintenant  $A_n \downarrow A$ . Dans ce cas,  $A_n^c \uparrow A^c$ . En appliquant le résultat précédent,  $\mathbf{P}(A^c) = \lim_n \mathbf{P}(A_n^c)$ . Par la propriété a), cette égalité se réécrit  $1 - \mathbf{P}(A) = \lim_n (1 - \mathbf{P}(A_n))$ , d'où on déduit  $\mathbf{P}(A) = \lim_n \mathbf{P}(A_n)$ .

f) La suite  $(\bigcap_{k=1}^n A_k)$  est décroissante et converge vers  $\bigcap_{k=1}^{\infty} A_k$ . Puisque  $\mathbf{P}(A_n) = 1$  pour tout  $n$ , il s'en suit que  $1 = \lim_n \mathbf{P}(A_n) = \mathbf{P}(\bigcap_{k=1}^{\infty} A_k)$ .

g) On montre d'abord la borne pour un nombre fini d'éléments :

$$\forall n, \mathbf{P}\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n \mathbf{P}(A_k). \quad (1.4)$$

L'inégalité est vraie au rang  $n = 1$ . Supposons qu'elle soit vraie au rang  $n$ ,  $\mathbf{P}(\bigcup_{k=1}^{n+1} A_k) = \mathbf{P}(A_{n+1} \cup (\bigcup_{k=1}^n A_k)) \leq \mathbf{P}(A_{n+1}) + \mathbf{P}(\bigcup_{k=1}^n A_k)$  d'après la propriété b). En injectant l'hypothèse de récurrence dans le membre de droite, la proposition est démontrée au rang  $n + 1$ .

L'inégalité (1.4) implique que  $\mathbf{P}(\bigcup_{k=1}^n A_k) \leq \sum_{k=1}^{\infty} \mathbf{P}(A_k)$  pour tout  $n$ . Or la suite  $(\bigcup_{k=1}^n A_k)$  est croissante, de limite  $\bigcup_{k=1}^{\infty} A_k$ . Par passage à la limite dans la dernière inégalité, on obtient le résultat voulu en invoquant la propriété e).

□

### 1.2.3. Représentation des probabilités sur un espace discret

La propriété suivante établit qu'une probabilité  $\mathbf{P}$  sur un espace discret est *entièrement caractérisée par la valeur qu'elle prend sur les singletons*.

**Proposition 1.2.** Soit  $\Omega$  un espace discret et  $\mathbf{P}$  une mesure de probabilité définie sur la tribu des parties  $\mathcal{P}(\Omega)$ . Alors pour tout événement  $A$ ,

$$\mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\{\omega\}).$$

*Démonstration.* Comme  $\Omega$  est au plus dénombrable, on peut indexer ses éléments sous la forme  $\Omega = \{\omega_1, \omega_2, \dots\}$ . Tout sous-ensemble  $A$  de  $\Omega$  est donc de la forme  $A = \{\omega_{i_1}, \omega_{i_2}, \dots\}$  où  $i_1, i_2, \dots$  sont des entiers. Par conséquent,  $A$  est l'union dénombrable des singletons  $\{\omega_{i_1}\}, \{\omega_{i_2}\}, \dots$ . Par  $\sigma$ -additivité de  $\mathbf{P}$ , on a donc  $\mathbf{P}(A) = \mathbf{P}(\{\omega_{i_1}\}) + \mathbf{P}(\{\omega_{i_2}\}) + \dots$ , ce qui prouve le résultat. □

Ainsi, il suffit de connaître la probabilité des événements élémentaires pour connaître la probabilité de n'importe quel événement. Cette affirmation est caractéristique des probabilités sur un espace discret, elle est clairement fausse dans le cas général des probabilités sur un espace non dénombrable.

La propriété suivante va un peu plus loin : elle établit que, pour qu'une famille de nombres positifs définissent une probabilité, il faut et il suffit que leur somme soit égale à un.



**Proposition 1.3.** Soit  $\Omega$  un ensemble discret. Soit  $(p_\omega)_{\omega \in \Omega}$  une suite de nombres positifs satisfaisant  $\sum_{\omega \in \Omega} p_\omega = 1$ . Alors il existe une (unique) mesure de probabilité  $\mathbf{P}$  sur  $\mathcal{P}(\Omega)$  telle que pour tout  $\omega \in \Omega$ ,  $\mathbf{P}(\{\omega\}) = p_\omega$ .

*Démonstration.* L'unicité est une conséquence de la propriété précédente. Afin de montrer l'existence, il suffit de poser  $\mathbf{P}(A) = \sum_{\omega \in A} p_\omega$ . On montre sans peine que cette application satisfait les axiomes d'une mesure de probabilité.  $\square$

Ainsi, concrètement, *une probabilité sur un espace discret se ramène à une famille de nombres positifs sommant à un* : se donner l'un revient à se donner l'autre.

### 1.2.4. Exemples de probabilités sur un espace discret

#### Cas où $\Omega$ est fini

**Loi uniforme** Soit  $\Omega$  un ensemble fini quelconque. La *probabilité uniforme sur  $\Omega$*  est définie par :

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|}$$

où  $|A|$  représente ici le cardinal de l'ensemble  $A$ . Autrement dit,  $\mathbf{P}(A)$  est le ratio entre le nombre d'issues pour lesquelles  $A$  est réalisé, et le nombre total d'issues. D'après la propriété précédente, on aurait pu définir la probabilité uniforme de façon équivalente comme l'unique probabilité pour laquelle toutes les issues sont *équiprobables*, c'est-à-dire pour tout  $\omega$ ,

$$\mathbf{P}(\{\omega\}) = \frac{1}{|\Omega|} .$$

**Loi de Bernoulli** Soit  $p \in [0, 1]$ . La *probabilité de Bernoulli de paramètre  $p$* , notée  $B(p)$ , est la probabilité définie sur  $\Omega = \{0, 1\}$  par :

$$\mathbf{P}(\{1\}) = p , \quad \mathbf{P}(\{0\}) = 1 - p .$$

La probabilité de Bernoulli permet de décrire la probabilité de succès ou d'échec d'une expérience. Par exemple, elle permet de décrire la probabilité qu'une pièce tombe sur pile : si la pièce est parfaitement équilibrée, on choisira  $p = 1/2$  et la probabilité de Bernoulli se ramène à la loi uniforme sur  $\{0, 1\}$  ; dans le cas d'une pièce non équilibrée ou d'un jeu truqué, le paramètre  $p$  est possiblement différent de  $1/2$ .

**Loi binomiale** Soit  $n \in \mathbf{N}^*$  et  $p \in [0, 1]$ . La *probabilité binomiale de paramètres  $n, p$* , notée  $B(n, p)$ , est la probabilité définie sur  $\Omega = \{0, 1, \dots, n\}$  par :

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

pour tout  $k = 0, \dots, n$  où l'on rappelle que  $\binom{n}{k} = n!/(k!(n-k)!)$ . La probabilité binomiale est utilisée pour décrire le nombre de succès obtenus lorsqu'on réitère  $n$  fois une expérience ayant même probabilité de succès  $p$  (voir l'exercice 2.1).

### Cas où $\Omega$ est dénombrable

**Loi géométrique** Soit  $p \in ]0, 1]$ . La *probabilité géométrique de paramètre  $p$  sur  $\mathbf{N}^*$* , notée  $\mathcal{G}(p)$ , est la probabilité définie sur  $\Omega = \mathbf{N}^*$  par :

$$\mathbf{P}(\{k\}) = p(1-p)^{k-1} \quad (1.5)$$

pour tout  $k \in \mathbf{N}^*$ . Imaginons que l'on réitère autant de fois que nécessaire une certaine expérience ayant une probabilité de succès  $p$ . Alors la probabilité géométrique est utilisée pour décrire le nombre d'expériences qui ont été nécessaires pour obtenir un succès (voir l'exercice 2.1).

REMARQUE.— On peut aussi définir la probabilité géométrique sur  $\mathbf{N}$  (et non  $\mathbf{N}^*$ ) par  $\mathbf{P}(\{k\}) = p(1-p)^k$  pour tout  $k = 0, 1, 2, \dots$ . Dans ce dernier cas, on cherche à décrire non pas l'instant du premier succès, mais le nombre d'échecs qui ont précédé le premier succès.

**Loi de Poisson** Soit  $\alpha > 0$ . La *probabilité de Poisson de paramètre  $\alpha$* , notée  $\mathcal{P}(\alpha)$ , est la probabilité définie sur  $\Omega = \mathbf{N}$  par :

$$\mathbf{P}(\{k\}) = \frac{\alpha^k}{k!} e^{-\alpha}.$$

La probabilité de Poisson est souvent utilisée pour modéliser des quantités telles que le nombre de requêtes reçues par un serveur par unité de temps, ou le nombre de clients qui se présentent à un guichet pendant une unité de temps.

## 1.3. Conditionnement et indépendance

### 1.3.1. Probabilité conditionnelle : définition

De façon informelle, la probabilité d'un événement vise à quantifier l'occurrence de cet événement. La probabilité conditionnelle d'un événement  $A$  sachant un événement  $B$  vise à quantifier l'occurrence de  $A$  lorsque l'on sait que  $B$  s'est produit. D'un point de vue plus formel, on a la définition suivante.

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

**Définition 1.4.** Pour tous événements  $A, B \in \mathcal{F}$  tels que  $\mathbf{P}(B) \neq 0$ , on définit la probabilité conditionnelle de  $A$  sachant  $B$ , et on note  $\mathbf{P}(A|B)$ , la quantité :

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

Si on associe probabilité et « poids », la probabilité d'un ensemble étant son poids relatif par rapport à celui de l'ensemble total, la probabilité conditionnelle de  $A$  sachant  $B$  est le poids de la trace de  $A$  sur  $B$  *relativement* au poids total de  $B$ .

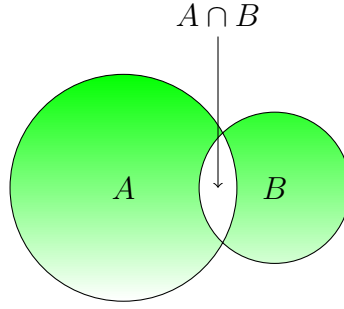


Figure 1.1. – Interprétation graphique du conditionnement.

Considérons le cas où  $\mathbf{P}$  est la probabilité uniforme sur un ensemble  $\Omega$  fini, c'est-à-dire  $\mathbf{P}(A) = |A|/|\Omega|$ . On a alors  $\mathbf{P}(A|B) = |A \cap B|/|B|$ . Cette expression justifie la remarque suivante :  $\mathbf{P}(A|B)$  peut être interprétée comme la probabilité de l'événement  $A \cap B$  dans ce nouvel univers qu'est  $B$ .

**APPLICATION 1.**– Considérons le lancer d'un dé :  $\mathbf{P}$  est la probabilité uniforme sur  $\Omega = \{1, 2, \dots, 6\}$ . Calculer la probabilité d'obtenir « 6 » sachant que le résultat est pair.

**Proposition 1.4.** Soit  $B \in \mathcal{F}$  tel que  $\mathbf{P}(B) \neq 0$ . L'application définie sur  $\mathcal{F}$  par  $A \mapsto \mathbf{P}(A|B)$  est une mesure de probabilité. On la nomme la probabilité conditionnelle à  $B$ .

*Démonstration.* i)  $\mathbf{P}(\Omega|B) = \mathbf{P}(\Omega \cap B)/\mathbf{P}(B) = 1$  et  $\mathbf{P}(\emptyset|B) = \mathbf{P}(\emptyset \cap B)/\mathbf{P}(B) = 0$ .  
ii) Soit  $(A_n)$  une famille d'événement deux à deux disjoints.  $\mathbf{P}(\bigcup_n A_n|B) = \mathbf{P}(\bigcup_n A_n \cap B)/\mathbf{P}(B) = \mathbf{P}(\bigcup_n (A_n \cap B))/\mathbf{P}(B) = \sum_n \mathbf{P}(A_n \cap B)/\mathbf{P}(B) = \sum_n \mathbf{P}(A_n|B)$ .  $\square$

### 1.3.2. Propriétés

La première propriété est connue sous le nom de *formule des probabilités totales*.

**Proposition 1.5.** a) Soient  $A, B \in \mathcal{F}$  tels que  $0 < \mathbf{P}(B) < 1$ . Alors,

$$\mathbf{P}(A) = \mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(A|B^c)\mathbf{P}(B^c).$$

b) Soit  $(B_n)_{n \in \mathbf{N}^*}$  une partition de  $\Omega$  telle que pour tout  $n$ ,  $\mathbf{P}(B_n) \neq 0$ . Alors,

$$\mathbf{P}(A) = \sum_{n=1}^{\infty} \mathbf{P}(A | B_n) \mathbf{P}(B_n) .$$

*Démonstration.*  $A$  s'écrit comme l'union disjointe  $A = (A \cap B) \cup (A \cap B^c)$  donc  $\mathbf{P}(A) = \mathbf{P}(A \cap B) + \mathbf{P}(A \cap B^c)$ . Le résultat provient du fait que  $\mathbf{P}(A \cap B) = \mathbf{P}(A | B) \mathbf{P}(B)$  et  $\mathbf{P}(A \cap B^c) = \mathbf{P}(A | B^c) \mathbf{P}(B^c)$ . La preuve de b) est fondée sur le même principe.  $\square$

EXEMPLE 3.— On dispose de trois pièces de monnaie : l'une est bien équilibrée, l'une comporte deux côtés pile, l'autre deux côtés face. On choisit une pièce au hasard. Evaluons la probabilité de tomber sur pile.

Désignons par  $E$ ,  $2P$  et  $2F$  les événements « la pièce bien équilibrée est choisie », « la pièce comportant deux côtés pile est choisie », etc. D'après la propriété ci-dessus,

$$\begin{aligned} \mathbf{P}(\text{pile}) &= \mathbf{P}(\text{pile} | E) \mathbf{P}(E) + \mathbf{P}(\text{pile} | 2P) \mathbf{P}(2P) + \mathbf{P}(\text{pile} | 2F) \mathbf{P}(2F) \\ &= \frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3} = \frac{1}{2} . \end{aligned}$$

La seconde propriété est connue sous le nom de *formule de Bayes*. La preuve est immédiate.

**Proposition 1.6.** Soient  $A, B \in \mathcal{F}$  deux événements tels que  $\mathbf{P}(A) \neq 0$  et  $\mathbf{P}(B) \neq 0$ . Alors,

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

La formule de Bayes permet typiquement d'évaluer des probabilités du type :

$$\mathbf{P}(\text{une action « a' » a été effectuée} \mid \text{le résultat « r » a été observé})$$

lorsqu'on connaît le modèle  $\mathbf{P}(\text{le résultat « r » est observé} \mid \text{l'action « a » est effectuée})$ .

EXEMPLE 4.— Reprenons l'exemple précédent des trois pièces. Sachant qu'on obtient le résultat pile, quelle est la probabilité que la pièce à deux côtés pile ait été choisie ? La réponse est donnée par la formule de Bayes :

$$\mathbf{P}(2P \mid \text{pile}) = \frac{\mathbf{P}(\text{pile} | 2P) \mathbf{P}(2P)}{\mathbf{P}(\text{pile})} = \frac{1 \times \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3} .$$

### 1.3.3. Événements indépendants

Dans l'exemple précédent, l'événement  $B = \text{« pile est le résultat »}$  apporte une information sur la probabilité que l'événement  $A = \text{« la pièce à deux côtés pile a été choisie »}$ . Avant l'expérience qui a vu  $B$  se réaliser, la probabilité de  $A$  valait  $\frac{1}{2}$ . Après l'expérience, elle vaut  $\frac{2}{3}$ . Le fait que  $B$  soit réalisé ne dit pas si  $A$  est ou non réalisé, mais par contre, il change

notre croyance en  $A$ . À l'inverse, il existe des événements  $A, B$  tels que la réalisation de  $B$  n'apporte aucune information sur  $A$ . De tels événements sont dits *indépendants*. Voici une définition plus formelle.

**Définition 1.5.** Deux événements  $A, B \in \mathcal{F}$  sont dits *indépendants*, noté  $A \perp\!\!\!\perp B$ , si

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

Si  $\mathbf{P}(B) \neq 0$ , la définition revient bien à  $\mathbf{P}(A | B) = \mathbf{P}(A)$  : la réalisation de  $B$  ne modifie pas la croyance en  $A$ .

REMARQUE.— a) Les propriétés suivantes sont équivalentes :  $A \perp\!\!\!\perp B$ ,  $B \perp\!\!\!\perp A$ ,  $A \perp\!\!\!\perp B^c$ ,  $A^c \perp\!\!\!\perp B$ ,  $A^c \perp\!\!\!\perp B^c$ .

b) Si  $\mathbf{P}(B) = 0$  ou  $\mathbf{P}(B) = 1$ , alors  $A$  et  $B$  sont indépendants quel que soit  $A$ .

**Définition 1.6.** Soit  $I$  un ensemble quelconque. Une famille  $(A_i)_{i \in I}$  d'événements est dite indépendante si pour tout sous-ensemble fini  $J \subset I$ , on a :

$$\mathbf{P} \left( \bigcap_{j \in J} A_j \right) = \prod_{j \in J} \mathbf{P}(A_j) .$$

Illustrons la formule ci-dessus lorsque la famille contient trois événements  $A, B, C$ . Les événements  $A, B, C$  sont indépendants si

$$\begin{aligned} \mathbf{P}(A \cap B) &= \mathbf{P}(A)\mathbf{P}(B), \quad \mathbf{P}(A \cap C) = \mathbf{P}(A)\mathbf{P}(C), \quad \mathbf{P}(B \cap C) = \mathbf{P}(B)\mathbf{P}(C), \\ \text{et } \mathbf{P}(A \cap B \cap C) &= \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C) . \end{aligned}$$

Il est important de souligner que la première ligne d'équations ci-dessus n'implique pas la deuxième : ce n'est pas parce que des événements sont deux à deux indépendants qu'ils forment une famille indépendante. L'exercice 1.10 fournit un contre-exemple.

**Définition 1.7.** Soit  $C \in \mathcal{F}$  tel que  $\mathbf{P}(C) \neq 0$ . On dit que  $A$  et  $B$  sont *indépendants conditionnellement à  $C$* , noté  $A \perp\!\!\!\perp B | C$  si  $\mathbf{P}(A \cap B | C) = \mathbf{P}(A | C)\mathbf{P}(B | C)$ .

La notion de famille indépendante conditionnellement à  $C$  se définit selon le même principe.

REMARQUE.— Attention : des propositions  $A \perp\!\!\!\perp B$  et  $A \perp\!\!\!\perp B | C$ , aucune n'implique l'autre. Là encore, l'exercice 1.10 fournit un contre-exemple.

$\Omega$	Probabilité	Expression de $\mathbf{P}$	Notation
$\Omega$ fini	Probabilité uniforme	$\mathbf{P}(x) = \frac{1}{ \Omega }$	
$\{0, 1\}$	Bernoulli de paramètre $p \in [0, 1]$	$\mathbf{P}(\{1\}) = p$ $\mathbf{P}(\{0\}) = 1 - p$	$\mathcal{B}(p)$
$\{1, \dots, n\}$	Binomiale de paramètres $n, p \in [0, 1]$	$\mathbf{P}(\{k\}) = \binom{n}{k} p^k (1 - p)^{n-k}$	$\mathcal{B}(n, p)$
$\mathbf{N}^*$	Géométrique de paramètre $p \in ]0, 1]$	$\mathbf{P}(\{k\}) = p(1 - p)^{k-1}$	$\mathcal{G}(p)$
$\mathbf{N}$	Géométrique de paramètre $p \in ]0, 1]$	$\mathbf{P}(\{k\}) = p(1 - p)^k$	$\mathcal{G}(p)$
$\mathbf{N}$	Poisson de paramètre $\alpha > 0$	$\mathbf{P}(\{k\}) = \frac{\alpha^k}{k!} e^{-\alpha}$	$\mathcal{P}(\alpha)$

Table 1.1. – Quelques exemples de probabilités sur un espace discret.

## 1.4. Exercices

### Pour apprendre

#### EXERCICE 1.1. –

On rappelle que

$$\mathbf{1}_A(x) = 1 \text{ si } x \in A, \text{ 0 sinon.}$$

1.1.a) Montrer que  $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$ ,  $\mathbf{1}_{A \cap B} = \mathbf{1}_A \mathbf{1}_B$ ,  $\mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \mathbf{1}_B$ .

1.1.b) Montrer que pour toute famille dénombrable  $(A_i)_{i \in I}$  d'ensembles deux à deux disjoints,

$$\mathbf{1}_{\cup_{i \in I} A_i} = \sum_{i \in I} \mathbf{1}_{A_i}.$$

#### EXERCICE 1.2. –

On lance simultanément trois dés à 6 faces non pipés.

1.2.a) Quel est l'espace des événements ?

1.2.b) Quelle est la probabilité d'avoir au moins 1 as ?

1.2.c) Montrer que les événements « la somme des faces est paire » et « la somme des faces est impaire » ont même probabilité.

1.2.d) Quelle est la probabilité que la somme des faces soit paire ?

#### EXERCICE 1.3 (Loto). –

Un joueur coche 6 numéros sur une grille de 49 numéros. On tire 6 boules parmi 49 boules numérotées. Quelle est la probabilité pour que le joueur ait exactement  $n$  bons numéros ( $n = 1, \dots, 6$ ) ?

### Pour s'entraîner

#### EXERCICE 1.4. –

Dans un lot de 20 articles, 12 sont parfaits, 6 comportent un défaut mineur et 2 un défaut majeur. Deux articles sont choisis au hasard, calculer les probabilités suivantes :

1.4.a) Les deux sont parfaits,

1.4.b) Les deux ont un défaut majeur,

1.4.c) Au moins l'un d'entre eux est parfait,

1.4.d) Exactement un est parfait

1.4.e) Au plus l'un d'entre eux est parfait.

Un lot de 20 articles est accepté lorsque 3 éléments choisis au hasard n'ont pas de défaut majeur.

1.4.f) Quelle est la probabilité que le lot décrit ci-dessus soit accepté ?

**EXERCICE 1.5.**—

Une boîte contient 4 piles usagées et 6 piles neuves. On tire deux piles au hasard. L'une d'entre elles seulement est testée.

1.5.a) Quelle est la probabilité que l'autre soit bonne si la pile testée est bonne ?

1.5.b) Même question si la pile testée est usagée.

On teste l'ensemble de la boîte par la méthode suivante : les piles sont tirées les unes après les autres au hasard sans remise. À chaque tirage, on teste la pile courante, le protocole s'arrête lorsque l'on a sorti les 4 piles usagées.

1.5.c) Quelle est la probabilité que le test s'arrête au cinquième tirage (au dixième tirage) ?

**EXERCICE 1.6** (Formule de Bayes).—

On suppose que l'on dispose d'un test déterminant d'une maladie donnée. Malheureusement, comme tout test, celui-ci est faillible : 1% des individus que l'on sait sains sont positifs au test et 2% des individus que l'on sait malades sont négatifs au test. On suppose que la maladie atteint 1% de la population testée. Quelle est la probabilité qu'un individu réagissant positivement au test soit effectivement malade ?

## Pour aller plus loin

**EXERCICE 1.7.**—

Lors d'un bal,  $n$  couples dansent. Les cavaliers ont choisi leur cavalière aléatoirement. Quelle est la probabilité qu'aucun des couples d'origine ne soit réuni ?

**EXERCICE 1.8** (Arnaque ou pas ?).—

Dans le jeu « Vegas », il est vendu 500 000 tickets à 3 € chaque. Ces tickets sont distribués aux buralistes sous forme de bandes de 50 tickets attachés les uns aux autres. La répartition des gains est la suivante :

Nb de lots	Gains
1	40 000 €
1	20 000 €
2	10 000 €
5	1 000 €
18	500 €
800	200 €
850	100 €
2 020	50 €
4 000	20 €
9 000	10 €
28 000	6 €
25 000	4 €
47 500	3 €



- 1.8.a) Quel est le montant moyen des gains ?  
 1.8.b) Quelle est la probabilité d'avoir un lot supérieur à 20 € ?  
 1.8.c) Sur 50 tickets, quelle est la probabilité (exacte et approchée) d'avoir 0 ou 1 lot supérieur à 20 € ?  
 1.8.d) M. R. a acheté 100 bandes de 50 tickets et il a constaté qu'aucune d'entre elles ne comportait plus d'un lot supérieur à 20 €. Quelle est la probabilité (approchée, en supposant que 5 000 est négligeable devant 500 000) d'un tel événement ?

**EXERCICE 1.9.**—

Soient  $\mathbf{P}$  et  $\mathbf{Q}$  deux mesures de probabilité sur  $\mathbf{N}$ . On note  $p_i = \mathbf{P}(\{i\})$  et  $q_i = \mathbf{Q}(\{i\})$ . On définit la distance en *variation totale* entre  $\mathbf{P}$  et  $\mathbf{Q}$  par

$$d_{TV}(\mathbf{P}, \mathbf{Q}) = \sup_{A \in \mathbf{N}} |\mathbf{P}(A) - \mathbf{Q}(A)|.$$

- 1.9.a) Montrer que

$$\sum_{i=0}^{+\infty} (p_i - q_i)^+ = \frac{1}{2} \sum_{i=0}^{+\infty} |p_i - q_i|.$$

On pourra utiliser le fait  $\sum_i p_i = \sum_i q_i = 1$ .

- 1.9.b) Montrer que pour toute partie  $A$  de  $\mathbf{N}$ ,

$$|\mathbf{P}(A) - \mathbf{Q}(A)| \leq \sum_{i=0}^{+\infty} (p_i - q_i)^+.$$

- 1.9.c) En choisissant convenablement l'ensemble  $A$ , montrer que

$$d_{TV}(\mathbf{P}, \mathbf{Q}) = \frac{1}{2} \sum_{i=0}^{+\infty} |p_i - q_i|.$$

On suppose maintenant que  $\mathbf{P}$  est donnée par  $\mathbf{P}(\{0\}) = p = 1 - \mathbf{P}(\{1\})$  et que  $\mathbf{Q}$  est une mesure de Poisson de paramètre  $\lambda = -\ln(p)$ , c'est-à-dire que

$$q_i = e^{-\lambda} \frac{\lambda^i}{i!}.$$

- 1.9.d) Calculer  $d_{TV}(\mathbf{P}, \mathbf{Q})$ .

**EXERCICE 1.10.**—

On lance deux dés. Soient les événements :  $A$  = « le premier dé affiche un résultat pair »,  $B$  = « le deuxième dé affiche un résultat pair »,  $C$  = « la somme des deux dés est paire ». Montrer que  $A$ ,  $B$ ,  $C$  sont deux à deux indépendants, mais ne forment pas une famille indépendante (on montrera que  $\mathbf{P}(A \cap B \cap C) \neq \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$ ).



## 2. Variables aléatoires discrètes

### 2.1. Loi d'une variable discrète

#### 2.1.1. Définitions

On se donne un univers  $\Omega$  au plus dénombrable, équipé d'une probabilité  $\mathbf{P}$ . De manière informelle, une variable aléatoire discrète est une grandeur à valeur dans un ensemble discret  $E$  qui dépend du résultat de l'expérience. C'est donc une *fonction* de l'issue  $\omega$  (en ce sens, la terminologie de *variable* est assez malencontreuse). On la notera souvent :

$$\begin{aligned} X : \Omega &\rightarrow E \\ \omega &\mapsto X(\omega) , \end{aligned}$$

où  $\Omega$  est l'univers, supposé muni d'une probabilité  $\mathbf{P}$  et où  $E$  est un ensemble **au plus dénombrable**.

EXEMPLE 5.— Considérons un lancer de  $n$  dés. Une issue  $\omega$  est un  $n$ -uplet sur  $\Omega = \{1, \dots, 6\}^n$ . On peut par exemple définir la variable aléatoire  $X(\omega)$  qui est égale au nombre de « 6 » obtenus : c'est bien une fonction de  $\omega$ .

Nous nous intéressons à la probabilité la forme « la variable  $X$  vaut  $x$  » ou, plus généralement,

$$\text{« la variable } X \text{ appartient à l'ensemble } H \text{ »} = \{\omega \in \Omega : X(\omega) \in H\} = X^{-1}(H) .$$

Nous utiliserons souvent les notations  $[X \in H]$  ou  $\{X \in H\}$  pour désigner l'événement  $\{\omega \in \Omega : X(\omega) \in H\}$ .

**Définition 2.1.** On appelle *loi de la v.a.  $X$*  la fonction définie pour tout  $H \subset E$  par :

$$\mathbf{P}_X(H) = \mathbf{P}(X^{-1}(H)) .$$

Avec une écriture plus compacte :  $\mathbf{P}_X := \mathbb{P} \circ X^{-1}$  où  $\circ$  représente la composition. Nous pouvons écrire de manière équivalente mais sans doute plus « parlante » :

$$\mathbf{P}_X(H) := \mathbb{P}[X \in H] .$$

Autrement dit,  $\mathbf{P}_X(H)$  est la probabilité pour que  $X$  appartienne à  $H$ .

**Proposition 2.1.**  $\mathbf{P}_X$  est une probabilité sur  $E$ .

*Démonstration.* On vérifie les axiomes i) et ii) que doit satisfaire une mesure de probabilité.

i)  $\mathbf{P}_X(E) = \mathbb{P}(X^{-1}(E)) = \mathbb{P}(\Omega) = 1$  et  $\mathbf{P}_X(\emptyset) = \mathbb{P}(X^{-1}(\emptyset)) = \mathbb{P}(\emptyset) = 0$ .

ii) Soit  $(H_n)_{n \in \mathbb{N}^*}$  une famille d'éléments de  $E$  deux à deux disjoints. On a  $X^{-1}(\bigcup_n H_n) = \bigcup_n X^{-1}(H_n)$  et on montre aisément que les événements  $(X^{-1}(H_n))$  sont deux à deux disjoints. Ainsi en appliquant  $\mathbb{P}$  aux deux membres de l'égalité précédente, on obtient  $\mathbf{P}_X(\bigcup_n H_n) = \mathbb{P}(\bigcup_n X^{-1}(H_n)) = \sum_n \mathbb{P}(X^{-1}(H_n))$ .  $\square$

On sait grâce au paragraphe 1.2.3 que  $\mathbf{P}_X$  est entièrement caractérisée par la valeur qu'elle prend sur les singletons. Afin d'alléger les notations, nous écrirons  $\mathbf{P}_X(x)$  au lieu de  $\mathbf{P}_X(\{x\})$ , soit :

$$\mathbf{P}_X(x) = \mathbf{P}[X = x] .$$

**Proposition 2.2.** Pour toute partie  $H$  de l'ensemble d'arrivée  $E$ , on a :

$$\mathbf{P}[X \in H] = \sum_{x \in H} \mathbf{P}_X(x) .$$

EXEMPLE 6.— Soit  $X$  le nombre de « 6 » obtenus lorsqu'on lance  $n$  dés. La probabilité d'obtenir au plus deux « 6 » s'écrit :  $\mathbf{P}[X \in \{0, 1, 2\}] = \mathbf{P}_X(0) + \mathbf{P}_X(1) + \mathbf{P}_X(2)$ .

### 2.1.2. Loi jointe, lois marginales

Soient  $X, Y$  deux v.a.d. de  $\Omega$  dans  $E$  de lois respectives  $\mathbf{P}_X$  et  $\mathbf{P}_Y$ . L'application :

$$\begin{aligned} (X, Y) : \Omega &\rightarrow E \times E \\ \omega &\mapsto (X(\omega), Y(\omega)) \end{aligned}$$

définit une v.a.d. sur  $E \times E$ .

**Définition 2.2.** La loi du couple  $(X, Y)$  est appelée la *loi jointe* de  $X$  et  $Y$ , notée  $\mathbf{P}_{X,Y}$ . Les lois  $\mathbf{P}_X$  et  $\mathbf{P}_Y$  sont appelées les *lois marginales* de  $X$  et  $Y$  respectivement.

D'après ce qui précède, la loi jointe est définie pour tout  $(x, y) \in E \times E$  par  $\mathbf{P}_{X,Y}(x, y) = \mathbf{P}[(X, Y) = (x, y)]$  soit :

$$\mathbf{P}_{X,Y}(x, y) = \mathbf{P}[X = x, Y = y] .$$

**Proposition 2.3.** La loi marginale  $\mathbf{P}_X$  est liée à la loi jointe par :

$$\forall x \in E, \mathbf{P}_X(x) = \sum_{y \in E} \mathbf{P}_{X,Y}(x, y) .$$

*Démonstration.* La famille d'événements de la forme  $[Y = y]$  où  $y$  décrit  $E$ , forme une partition de l'univers  $\Omega$ . D'après la formule des probabilités totales,  $\mathbf{P}[X = x] = \sum_y \mathbf{P}[X = x, Y = y]$ .  $\square$

### Généralisation au cas d'une famille finie de v.a.d.

Soient  $X_1, \dots, X_n$  des v.a. de  $\Omega \rightarrow E$ . Le  $n$ -uplet  $(X_1, \dots, X_n)$  définit une v.a.d. sur  $E^n$ . Sa loi est appelée la *loi jointe* de  $X_1, \dots, X_n$ , notée  $\mathbf{P}_{X_1, \dots, X_n}$ . Pour tout  $k$ , la loi  $\mathbf{P}_{X_k}$  est appelée la loi marginale de  $X_k$ .

**Proposition 2.4.** Pour tout  $k = 1, \dots, n$  et tout  $x_k \in E$ ,

$$\mathbf{P}_{X_k}(x_k) = \sum_{x_1 \dots x_{k-1}, x_{k+1} \dots x_n} \mathbf{P}_{X_1, \dots, X_n}(x_1, \dots, x_n),$$

où la somme s'étend sur l'ensemble des  $(n-1)$ -uplets  $(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$  sur  $E$ .

Ainsi, à partir de la loi jointe, on peut déduire les lois marginales en éliminant les variables non-souhaitées par sommation sur toutes les valeurs possibles prises par celles-ci.

**Définition 2.3.** Une famille  $(X_i)_{i \in I}$  de variables aléatoires sur le même espace  $E$  est dite identiquement distribuée si toutes les variables ont la même loi :  $\forall i \in I, \mathbf{P}_{X_i} = \mathbf{P}_{X_1}$ .

REMARQUE.— Il est évident que deux v.a.  $X$  et  $Y$  différentes peuvent avoir la même loi ( $\mathbf{P}_X = \mathbf{P}_Y$ ). Par exemple, si  $X \in \{0, 1\}$  suit une loi de Bernoulli de paramètre  $1/2$ , alors la v.a.  $Y = 1 - X$  est différente de  $X$  et suit néanmoins la même loi :

$$\mathbf{P}[Y = 0] = \mathbf{P}[1 - X = 0] = \mathbf{P}[X = 1] = \frac{1}{2} = \mathbf{P}[X = 0].$$

## 2.2. Indépendance des v.a. discrètes

Soient  $X$  et  $Y$  deux v.a.d. à valeurs dans  $E$ .

**Définition 2.4.**  $X$  et  $Y$  sont dites *indépendantes* si pour tout  $G, H \subset E$ , les événements  $[X \in G]$  et  $[Y \in H]$  sont indépendants, autrement dit si :

$$\mathbf{P}[X \in G, Y \in H] = \mathbf{P}[X \in G] \mathbf{P}[Y \in H],$$

où  $[X \in G, Y \in H]$  désigne l'ensemble  $[X \in G] \cap [Y \in H]$ .

**Proposition 2.5.** Deux v.a.d  $X$  et  $Y$  sont indépendantes si et seulement si pour tout  $(x, y) \in E^2$ ,

$$\mathbf{P}_{X,Y}(x, y) = \mathbf{P}_X(x) \mathbf{P}_Y(y) \quad (2.1)$$

REMARQUE.— Par définition,  $X$  et  $Y$  sont indépendantes lorsque les événements  $[X \in H]$  et  $[Y \in G]$  sont indépendants quel que soit le choix de  $H$  et  $G$ . Le résultat ci-dessus montre qu'il suffit de vérifier cette propriété sur les singletons  $H = \{x\}$  et  $G = \{y\}$ .

*Démonstration.* Le sens  $\Rightarrow$  est immédiat. On montre la réciproque. On a pour tout  $H, G \subset E$ ,  $\mathbf{P}[X \in H, Y \in G] = \mathbf{P}_{X,Y}(H \times G)$ . Comme  $\mathbf{P}_{X,Y}$  est une mesure de probabilité sur un espace discret,  $\mathbf{P}_{X,Y}(H \times G) = \sum_{(x,y) \in H \times G} \mathbf{P}_{X,Y}(x, y)$ . En appliquant l'hypothèse,  $\mathbf{P}_{X,Y}(H \times G) = \sum_{(x,y) \in H \times G} \mathbf{P}_X(x) \mathbf{P}_Y(y) = \sum_{x \in H} \mathbf{P}_X(x) \sum_{y \in G} \mathbf{P}_Y(y) = \mathbf{P}_X(H) \mathbf{P}_Y(G)$ , ce qui prouve le résultat.  $\square$

Soit  $E'$  un autre espace discret et soient  $f, g : E \rightarrow E'$ . On désigne par  $f(X)$  la v.a.d.  $\omega \mapsto f(X(\omega))$ , c'est à dire  $f(X) = f \circ X$ .

**Proposition 2.6.** Si  $X$  et  $Y$  sont indépendantes, alors  $f(X)$  et  $g(Y)$  sont des v.a. indépendantes.

*Démonstration.* Soient  $H, G$  deux parties de  $E'$ . Les ensembles  $[f(X) \in H]$  et  $[g(Y) \in G]$  s'écrivent respectivement  $[X \in f^{-1}(H)]$  et  $[Y \in g^{-1}(G)]$  et sont donc indépendants.  $\square$

### Généralisation au cas d'une famille finie de v.a.d.

Soient  $X_1, \dots, X_n$  des v.a.d. sur  $E$ . Elles sont dites indépendantes si pour toute suite d'ensembles  $(H_1, \dots, H_n)$ , les événements  $([X_k \in H_k])_{k=1, \dots, n}$  sont indépendants. Autrement dit,

$$\mathbb{P}[X_1 \in H_1, \dots, X_n \in H_n] = \mathbb{P}[X_1 \in H_1] \times \dots \times \mathbb{P}[X_n \in H_n] ,$$

où l'on utilise la notation  $[X_1 \in H_1, \dots, X_n \in H_n] = \bigcap_k [X_k \in H_k]$ .

**Théorème 2.7.**  $X_1, \dots, X_n$  sont indépendantes si et seulement si  $\forall (x_1, \dots, x_n) \in E^n$ ,

$$\mathbf{P}_{X_1 \dots X_n}(x_1, \dots, x_n) = \prod_{k=1}^n \mathbf{P}_{X_k}(x_k) .$$

La preuve suit le même principe que dans le cas  $n = 2$  traité plus haut.

**Définition 2.5.** Une famille de variables aléatoires est dite indépendante si toute sous-famille *finie* est indépendante.

*n.b.* : on utilise souvent l'abréviation *i.i.d.* pour désigner une famille indépendante et *identiquement distribuée* de variables aléatoires.

**Proposition 2.8.** Soit  $(X_i)_{i \in I}$  une famille indépendante de v.a., chacune étant à valeur dans un espace  $E_i$ . On se donne pour tout  $i$  une application mesurable  $f_i$  sur  $E_i$ . Alors la famille de v.a.  $(f_i(X_i))_{i \in I}$  est indépendante.

## 2.3. Espérance, moments

### 2.3.1. Introduction

Un joueur de « pile ou face » gagne 10 euros lorsque la pièce tombe sur *pile* et perd 5 euros lorsqu'elle tombe sur *face*. Soit  $X$  le gain réalisé après l'expérience.  $X$  peut prendre deux valeurs :  $a = 10$  ou  $b = -5$ . On définit l'*espérance* du gain par

$$\begin{aligned}\mathbb{E}(X) &= a \mathbf{P}[X = a] + b \mathbf{P}[X = b] \\ &= 10 \cdot \frac{1}{2} + (-5) \cdot \frac{1}{2} = 2,5 \text{ euros.}\end{aligned}$$

L'espérance est donc une *moyenne pondérée* des gains. D'un point de vue physique, c'est le *centre de gravité* des points  $a$  et  $b$  auquel on a affecté les masses  $\mathbf{P}[X = a]$  et  $\mathbf{P}[X = b]$  respectivement.

Imaginons que le joueur précédent effectue  $n$  lancers de pièce : on note  $X_1, \dots, X_n$  les gains respectifs réalisés à chaque expérience. La *moyenne empirique* des gains est définie par

$$S_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

Nous verrons à la fin de ce chapitre un résultat important appelé la *loi des grands nombres* que nous énonçons pour l'instant de manière informelle : la moyenne empirique  $S_n$  converge vers l'espérance lorsque  $n \rightarrow \infty$ . Naturellement, il conviendrait de préciser de quelle « convergence » il est question (n'oublions pas que l'on parle ici de variables aléatoires et non d'une simple suite de nombres). Mais cette remarque donne une première illustration de l'importance de l'espérance en probabilité.

### 2.3.2. Définition

On suppose dorénavant que  $E$  est une partie au plus dénombrable de  $\mathbb{R}$ .

**Définition 2.6.** On définit l'espérance  $\mathbb{E}(X)$  d'une v.a.d.  $X$  par

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x \in E} x \mathbf{P}[X = x] \\ &= \sum_{x \in E} x \mathbf{P}_X(x).\end{aligned}\tag{2.2}$$

Pour que cette somme ait un sens, il suffit que l'une de ces deux conditions soit

vérifiée :

1. ses termes sont tous positifs :  $\mathbf{P}_X(x) = 0$  pour tout  $x < 0$  (auquel cas  $\mathbb{E}(X)$  peut éventuellement être égal à  $+\infty$ );
2. ses termes sont absolument sommables, c'est à dire :

$$\sum_{x \in E} |x| \mathbf{P}[X = x] < \infty . \quad (2.3)$$

Lorsque la première condition est vraie *i.e.*,  $\mathbf{P}_X(x) = 0$  pour tout  $x < 0$ , nous dirons que la v.a.d.  $X$  est *positive presque partout* et nous noterons  $X \geq 0$  p.p.

Soulignons que l'espérance  $\mathbb{E}(X)$  est une constante, elle ne dépend pas de l'issue  $\omega$ . Elle ne dépend de  $X$  qu'au travers de sa loi  $\mathbf{P}_X$ . En particulier, deux v.a. identiquement distribuées ont même espérance.

Une variable d'espérance nulle est dite *centrée*.

### 2.3.3. Propriétés

Soient  $E, F$  deux espaces discrets avec  $F \subset \mathbb{R}$ . Soit  $f : E \rightarrow F$  une fonction. La composée  $f(X)$  définit une nouvelle variable aléatoire  $\omega \mapsto f(X(\omega))$ . Nous nous intéressons à son espérance.

**Proposition 2.9.** Si  $f$  est positive,

$$\mathbb{E}(f(X)) = \sum_{x \in E} f(x) \mathbf{P}[X = x] . \quad (2.4)$$

La formule reste vraie pour  $f$  quelconque pourvu que  $\sum_{x \in E} |f(x)| \mathbf{P}[X = x] < \infty$ .

*Démonstration.* Donnons d'abord la preuve pour  $f$  positive :

$$\begin{aligned} \mathbb{E}(f(X)) &= \sum_{y \in F} y \mathbf{P}[f(X) = y] = \sum_{y \in F} y \mathbf{P}[X \in f^{-1}(\{y\})] \\ &= \sum_{y \in F} \sum_{x \in f^{-1}(\{y\})} y \mathbf{P}[X = x] \\ &= \sum_{y \in F} \sum_{x \in f^{-1}(\{y\})} f(x) \mathbf{P}[X = x] \\ &= \sum_{x \in E} f(x) \mathbf{P}[X = x] , \end{aligned} \quad (2.5)$$

où on a utilisé le fait que les ensembles de la forme  $f^{-1}(\{y\})$  sont une partition de  $E$ . Dans le cas où  $f$  n'est pas positive, on doit d'abord vérifier que  $\mathbb{E}(f(X))$  est bien définie. En appliquant le résultat déjà démontré à la fonction « valeur absolue », nous avons :

$$\sum_{y \in F} |y| \mathbf{P}[f(X) = y] = \mathbb{E}(|f(X)|)$$



et en appliquant le résultat à la fonction  $|f|$ ,  $\mathbb{E}(|f(X)|) = \sum_{x \in E} |f(x)| \mathbf{P}[X = x]$  qui est fini par hypothèse. L'équivalent pour la v.a.  $f(X)$  de la condition (2.3) est satisfaite :  $\mathbb{E}(f(X))$  est bien définie. La preuve de (2.4) est obtenue par le même calcul qu'en (2.5).  $\square$

REMARQUE.— Si l'on devait évaluer  $\mathbb{E}(f(X))$  en utilisant la définition (2.2) de l'espérance, on devrait au préalable calculer la loi  $\mathbf{P}_{f(X)}$  de la v.a.d.  $f(X)$ . L'équation (2.4) montre que l'espérance  $\mathbb{E}(f(X))$  s'exprime directement en fonction de la loi de la variable  $X$ .

Lorsqu'on choisit pour  $f$  l'indicatrice d'un ensemble  $H$ , on a le corollaire suivant :

$$\mathbb{E}(\mathbf{1}_H(X)) = \mathbf{P}_X(H) . \quad (2.6)$$

La propriété (2.4) permet d'écrire la condition de sommabilité (2.3) de manière plus compacte : on écrira simplement  $\mathbb{E}|X| < \infty$ .

Soient deux  $X$  et  $Y$  deux v.a.d. sur  $E \subset \mathbb{R}$ . Pour tous coefficients réels  $\alpha, \beta$ , la somme  $\alpha X + \beta Y$  est bien une v.a.d. en tant que fonction du couple  $(X, Y)$ . Dans la suite, on utilisera la notation « p.p. » pour « presque partout », c'est-à-dire avec probabilité 1. Par exemple, on écrira que «  $X = a$  p.p. » pour signifier que  $\mathbf{P}(X = a) = 1$ .

**Proposition 2.10.** Soient  $X$  et  $Y$  deux variables aléatoires dans un ensemble  $E \subset \mathbb{R}$  discret. Supposons que  $\mathbb{E}|X| < \infty$  et  $\mathbb{E}|Y| < \infty$ . Soient  $(\alpha, \beta) \in \mathbb{R}^2$  et  $a \in E$ . Alors :

- a)  $\mathbb{E}(\alpha X + \beta Y)$  est bien définie et  $\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$ .
- b) Si  $X \geq 0$  p.p., alors  $\mathbb{E}(X) \geq 0$ .
- c) Si  $X \geq 0$  p.p. et si  $\mathbb{E}(X) = 0$ , alors  $X = 0$  p.p.
- d)  $|\mathbb{E}(X)| \leq \mathbb{E}|X|$ .
- e) Si  $X \leq Y$  p.p., alors  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .
- f) Si  $X = a$  p.p., alors  $\mathbb{E}(X) = a$ .

*Démonstration.* Montrons que  $\mathbb{E}(\alpha X + \beta Y)$  est bien définie. D'après la propriété précédente,

$$\begin{aligned} \mathbb{E}|\alpha X + \beta Y| &= \sum_{(x,y) \in E^2} |\alpha x + \beta y| \mathbf{P}_{X,Y}(x, y) \\ &\leq |\alpha| \sum_{(x,y)} |x| \mathbf{P}_{X,Y}(x, y) + |\beta| \sum_{(x,y)} |y| \mathbf{P}_{X,Y}(x, y) \\ &= |\alpha| \sum_x |x| \sum_y \mathbf{P}_{X,Y}(x, y) + |\beta| \sum_y |y| \sum_x \mathbf{P}_{X,Y}(x, y) \\ &= |\alpha| \sum_x |x| \mathbf{P}_X(x) + |\beta| \sum_y |y| \mathbf{P}_Y(y) . \end{aligned}$$

Ainsi,  $\mathbb{E}|\alpha X + \beta Y| \leq |\alpha| \mathbb{E}|X| + |\beta| \mathbb{E}|Y| < \infty$  par hypothèse. On évalue l'espérance :

$$\begin{aligned} \mathbb{E}(\alpha X + \beta Y) &= \sum_{(x,y)} (\alpha x + \beta y) \mathbf{P}_{X,Y}(x, y) \\ &= \alpha \sum_{(x,y)} x \mathbf{P}_{X,Y}(x, y) + \beta \sum_{(x,y)} y \mathbf{P}_{X,Y}(x, y) , \end{aligned}$$

où la dernière égalité se justifie par le fait que les deux dernières sommes convergent absolument (nous l'avons prouvé plus haut). Par le même calcul que ci-dessus, ces deux sommes sont égales à  $\mathbb{E}(X)$  et  $\mathbb{E}(Y)$  respectivement, ce qui démontre a). Les preuves des autres propositions sont laissées au lecteur.  $\square$

### 2.3.4. Inégalités

**Proposition 2.11.** (Inégalité de Markov) Pour tout  $\epsilon > 0$ ,  $p \geq 1$ ,

$$\mathbf{P}[|X| > \epsilon] \leq \frac{\mathbb{E}(|X|^p)}{\epsilon^p}.$$

*Démonstration.* On donne d'abord la preuve pour  $\epsilon = p = 1$  et  $X \geq 0$ . D'après (2.6),  $\mathbf{P}[X > 1] = \mathbb{E}(\mathbf{1}_{]1, +\infty[}(X)) \leq \mathbb{E}(X)$  car  $\mathbf{1}_{]1, +\infty[}(X) \leq X$ . Dans le cas général, on utilise le fait que  $\mathbf{P}[|X| > \epsilon] = \mathbf{P}[|X|^p/\epsilon^p > 1]$  et on applique le résultat précédent.  $\square$

Lorsque  $p = 2$ , l'inégalité de Markov est aussi connue sous le nom d'inégalité de Bienaymé-Tchebychev.

**Proposition 2.12.** (Inégalité de Cauchy-Schwarz)

$$\mathbf{E}(|XY|) \leq \sqrt{\mathbf{E}(X^2) \mathbf{E}(Y^2)}.$$

*Démonstration.* Si  $\mathbf{E}(X^2) = 0$ , la v.a.  $X^2$  est nulle p.p. donc  $XY = 0$  p.p. ce qui implique que le membre de gauche est nul. L'inégalité est triviale dans ce cas. Le seul cas non-trivial est celui pour lequel  $\mathbf{E}(X^2) \neq 0$  et  $\mathbf{E}(Y^2) \neq 0$ .

On utilise l'inégalité  $U^2 + V^2 \geq 2UV$  en posant  $U = |X|/\sqrt{\mathbf{E}(X^2)}$  et  $V = |Y|/\sqrt{\mathbf{E}(Y^2)}$ . Comme  $\mathbf{E}(U^2) = \mathbf{E}(V^2) = 1$ , on obtient en prenant l'espérance de chaque membre de l'inégalité :  $1 + 1 \geq 2\mathbf{E}(|XY|/\sqrt{\mathbf{E}(X^2) \mathbf{E}(Y^2)})$  ce qui démontre le résultat.  $\square$

### 2.3.5. Moments, variance, covariance

**Définition 2.7.** Soit  $p \geq 0$ . Soit une v.a.d. réelle  $X$  telle que  $\mathbb{E}(|X|^p) < \infty$ . La quantité  $\mathbb{E}(X^p)$  est appelée le *moment d'ordre  $p$*  de  $X$ .

On dit d'une telle variable qu'elle est d'*ordre  $p$* , ou qu'elle *possède un moment d'ordre  $p$* .

REMARQUE.— Le moment d'ordre 1 coïncide avec l'espérance. Une variable bornée possède tous ses moments.

**Proposition 2.13.** Une variable d'ordre  $p$  possède tous ses moments d'ordre inférieur.

*Démonstration.* Soit  $0 \leq q \leq p$ . De l'inégalité  $|x|^q \leq 1 + |x|^p$ , on déduit  $\mathbb{E}|X|^q \leq 1 + \mathbb{E}|X|^p < \infty$ .  $\square$

**Définition 2.8.** La *variance* d'une v.a.d.  $X$  d'ordre 2 est définie par

$$\text{Var}(X) := \mathbf{E} \left( (X - \mathbf{E}(X))^2 \right).$$

Son *écart-type* est la racine carrée de la variance, noté  $\sigma_X := \sqrt{\text{Var}(X)}$ .

EXEMPLE 7.— Un joueur lance une pièce, gagne un euro si le résultat est pile, perd un euro sinon. L'espérance du gain  $X$  est  $\mathbf{E}(X) = 0$ . La variance est  $\text{Var}(X) = 1 \times \frac{1}{2} + 1 \times \frac{1}{2} = 1$  et l'écart-type est 1. Si le joueur gagne ou perd 10 euros à chaque partie, l'espérance de gain est toujours nulle. En revanche, la variance vaut 100 et l'écart type vaut 10. La variance donne donc une information sur l'amplitude des fluctuations de la  $X$  autour de son espérance.

EXEMPLE 8.— La variance d'une loi de Bernoulli  $B(p)$  vaut  $p(1 - p)$ .

**Définition 2.9.** Soient  $X$  et  $Y$  deux v.a.d. d'ordre 2. Leur *covariance* est définie par :

$$\text{Cov}(X, Y) := \mathbb{E} [(X - \mathbf{E}(X))(Y - \mathbf{E}(Y))].$$

L'inégalité de Cauchy-Schwarz garantit que la quantité ci-dessus est bien définie. En statistique et en traitement du signal, on utilise souvent une version renormalisée de la covariance, le *coefficient de corrélation* qui est défini par :

$$\rho_{X,Y} := \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}.$$

Lorsque  $\text{Cov}(X, Y) = 0$ , on dit que  $X$  et  $Y$  sont *décorrélées*.

**Proposition 2.14.** Soient  $X$  et  $Y$  deux v.a.d. d'ordre 2 et  $(\alpha, \beta) \in \mathbb{R}^2$ . On a :

- a)  $\text{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}X)^2$  ;
- b)  $\text{Cov}(X, X) = \text{Var}(X)$  ;
- c)  $\text{Cov}(Y, X) = \text{Cov}(X, Y)$  ;
- d)  $\text{Var}(\alpha X + \beta) = \alpha^2 \text{Var}(X)$  ;
- e)  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$ .

La preuve est laissée à titre d'exercice.

### 2.3.6. Cas des variables indépendantes

**Proposition 2.15.** Soient  $X$  et  $Y$  des v.a. indépendantes telles que  $\mathbf{E}|X|, \mathbf{E}|Y| < \infty$ . Alors  $\mathbf{E}|XY| < \infty$ , et on a l'égalité :

$$\mathbf{E}(XY) = \mathbf{E}(X) \mathbf{E}(Y) .$$

*Démonstration.*  $\mathbf{E}|XY| = \sum_{(x,y)} |xy| \mathbf{P}_{X,Y}(x,y)$  et comme  $X$  et  $Y$  sont indépendantes,  $\mathbf{P}_{X,Y}(x,y) = \mathbf{P}_X(x) \mathbf{P}_Y(y)$ . Ainsi,  $\mathbf{E}|XY| = \sum_x |x| \mathbf{P}_X(x) \sum_y |y| \mathbf{P}_Y(y) = \mathbf{E}|X| \mathbf{E}|Y| < \infty$ . Le même calcul, sans les valeurs absolues, montre que  $\mathbf{E}(XY) = \mathbf{E}(X) \mathbf{E}(Y)$ .  $\square$

Cette propriété admet une généralisation immédiate. Si  $X$  et  $Y$  sont indépendantes, on sait que pour des fonctions  $f$  et  $g$  arbitraires, les v.a.d.  $f(X)$  et  $g(Y)$  restent indépendantes. Par conséquent,

$$\mathbf{E}(f(X)g(Y)) = \mathbf{E}(f(X)) \mathbf{E}(g(Y)) , \quad (2.7)$$

dès lors que les deux sommes du membre de droite sont absolument convergentes. On a même une réciproque à ce résultat.

**Théorème 2.16.** Deux variables aléatoires  $X : \Omega \rightarrow (E, \mathbf{E})$  et  $Y : \Omega \rightarrow (F, \mathcal{F})$  sont indépendantes si et seulement pour toutes les fonctions mesurables bornées  $f : E \rightarrow \mathbf{R}$  et  $g : F \rightarrow \mathbf{R}$ ,

$$\mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)] \mathbf{E}[g(Y)] . \quad (2.8)$$

*Démonstration.* On vient de voir que l'indépendance implique (2.8).

Réciproquement, si l'équation (2.8) est vérifiée, alors quels que soient  $x \in E$  et  $y \in F$ , on obtient (2.1) en spécialisant (2.8) pour  $f = \mathbf{1}_{\{x\}}$  et  $g = \mathbf{1}_{\{y\}}$ .  $\square$

Un cas particulier intéressant est obtenu en posant  $f(x) = x - \mathbf{E}(X)$  et  $g(y) = y - \mathbf{E}(Y)$ . Dans ce cas, le membre de gauche de (2.7) n'est autre que la covariance  $\text{Cov}(X, Y)$  et les deux facteurs du membre de droite sont nuls. On en déduit la propriété suivante :

**Proposition 2.17.** Si  $X$  et  $Y$  sont indépendantes et d'ordre 2, alors

$$\text{Cov}(X, Y) = 0 .$$

Cette propriété implique en particulier que pour des v.a. indépendantes :

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) . \quad (2.9)$$

Notons que deux variables décorréées ne sont pas nécessairement indépendantes. L'exercice 2.4 permet de s'en convaincre.

**Généralisation au cas d'une famille finie de v.a.d.**

**Proposition 2.18.** Pour tout  $k = 1, \dots, n$ , soit  $X_k$  une v.a. sur un espace discret  $E_k$  et  $f_k : E_k \rightarrow E'_k$  une fonction sur  $E'_k \subset \mathbb{R}$  telle que  $\mathbb{E}|f_k(X_k)| < \infty$ . On suppose  $X_1, \dots, X_n$  indépendantes. Alors,

$$\mathbf{E} \left( \prod_{k=1}^n f_k(X_k) \right) = \prod_{k=1}^n \mathbf{E}(f_k(X_k))$$

**Proposition 2.19.** Si  $X_1, \dots, X_n$  sont des v.a.d. indépendantes d'ordre 2, alors

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) .$$

*Démonstration.* La propriété est vraie au rang  $n = 1$ . Supposons la vraie au rang  $n - 1$ . Posons  $Z_n = X_1 + \dots + X_{n-1}$ . Les v.a.  $X_n$  et  $Z_n$  sont indépendantes. Par l'égalité (2.9),  $\text{Var}(X_n + Z_n) = \text{Var}(X_n) + \text{Var}(Z_n)$  or  $\text{Var}(Z_n) = \text{Var}(X_1) + \dots + \text{Var}(X_{n-1})$  par l'hypothèse de récurrence. La propriété est donc démontrée.  $\square$

**2.3.7. Application : Loi faible des grands nombres \***

Soit  $(X_k)_{k \in \mathbb{N}^*}$  une famille indépendante et identiquement distribuée de v.a. sur un ensemble  $E \subset \mathbb{R}$  au plus dénombrable. On s'intéresse au comportement de la moyenne empirique des  $n$  premières variables :

$$S_n = \frac{1}{n} \sum_{k=1}^n X_k .$$

**Théorème 2.20.** Soit  $(X_k)_{k \in \mathbb{N}^*}$  une famille indépendante, identiquement distribuée de v.a.d.. On suppose que  $\mathbf{E}(X_1^2) < \infty$ . Alors,

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbf{P}[|S_n - \mathbf{E}(X_1)| > \epsilon] = 0 .$$

On dit de la variable aléatoire  $S_n$  qu'elle *converge en probabilité* vers  $\mathbf{E}(X_1)$ .

*Démonstration.* En utilisant le fait que  $\mathbf{E}(X_1) = \mathbf{E}(X_k)$ , on a :

$$\begin{aligned} \mathbf{P}[|S_n - \mathbf{E}(X_1)| > \epsilon] &= \mathbf{P} \left[ \left| \sum_k (X_k - \mathbf{E}(X_k)) \right| > n\epsilon \right] \\ &\leq \frac{\mathbf{E}((\sum_k (X_k - \mathbf{E}(X_k)))^2)}{n^2 \epsilon^2} , \end{aligned}$$

en utilisant l'inégalité de Bienaymé-Tchebichev. La somme dans l'espérance est une v.a. centrée, donc son moment d'ordre 2 et sa variance coïncident. Par indépendance de  $X_k$ , sa variance satisfait :  $\text{Var}(\sum_k X_k) = \sum_k \text{Var}(X_k) = n \text{Var}(X_1)$ , où la seconde égalité

provient du fait que toutes les variances sont égales, les  $X_k$  étant identiquement distribués. Finalement,

$$\mathbf{P}[|S_n - \mathbf{E}(X_1)| > \epsilon] \leq \frac{n \operatorname{Var}(X_1)}{n^2 \epsilon^2},$$

et le membre de gauche converge bien vers zéro quand  $n$  tend vers l'infini.  $\square$

Le théorème précédent se nomme « loi faible des grands nombres ». Donnons-en une illustration. Un joueur lance une pièce, gagne un euro si le résultat est pile, perd un euro sinon. Il réitère l'expérience  $n$  fois.  $X_k$  représente son gain à l'instant  $k$  et  $S_n$  la moyenne des gains. L'espérance du gain  $X_k$  est  $\mathbf{E}(X_k) = 0$ . La loi faible des grands nombres implique que  $\mathbf{P}[|S_n| > \epsilon]$  tend vers zéro. Quel que soit  $\epsilon$  aussi petit qu'on veut, le gain moyen est plus petit que  $\epsilon$  avec forte probabilité lorsque  $n$  est grand.

REMARQUE.— Plus loin dans ce cours, nous étendrons la loi faible des grands nombres à des v.a. quelconques, pas nécessairement discrètes. Nous montrerons également un résultat plus puissant appelé « loi forte des grands nombres ». La loi forte établit que *quelle que soit l'issue*  $\omega$ , hormis peut-être pour  $\omega$  dans un ensemble de probabilité nulle, nous avons  $\lim_n S_n(\omega) = \mathbf{E}(X_1)$ .

## 2.4. Fonction génératrice d'une v.a. à valeurs entières

Dans ce paragraphe, on se limite au cas où la v.a.  $X$  est à valeurs dans  $\mathbb{N}$  (ou bien dans un sous-ensemble  $E \subset \mathbb{N}$  : dans ce dernier cas, on étend  $X$  à une fonction dans  $\mathbb{N}$  en imposant que  $\mathbf{P}[X = k] = 0$  pour  $k \notin E$ ).

**Définition 2.10.** La *fonction génératrice* de  $X$ , notée  $\Phi_X$ , est définie pour tout  $s$  dans l'intervalle  $[-1, +1]$  par :

$$\begin{aligned} \Phi_X(s) &= \mathbf{E}(s^X) \\ &= \sum_{k=0}^{\infty} \mathbf{P}[X = k] s^k. \end{aligned}$$

La fonction génératrice est donc la série entière de terme général  $\mathbf{P}[X = k]$ . Le rayon de convergence de cette série est supérieur ou égal à un.

**Proposition 2.21.** Pour toute v.a.d. à valeurs entières, sa fonction génératrice  $\Phi_X$  satisfait les propriétés suivantes.

- a)  $\Phi_X$  est continue sur  $[-1, +1]$  et de classe  $C^\infty$  sur  $] -1, +1[$ .
- b) Pour tout  $n$ ,

$$\mathbf{P}[X = n] = \frac{\Phi_X^{(n)}(0)}{n!},$$

où  $\Phi_X^{(n)}$  est la dérivée  $n$ ème de  $\Phi_X$ .

*Démonstration.* Il suffit d'appliquer des résultats connus sur les séries entières. On sait (voir l'annexe et [Rud95, Théorème 8.1]) qu'une série entière de terme général  $a_k$  et de rayon de convergence  $R$  est de classe  $C^\infty$  sur  $] -r, r[$ , et sa dérivée  $n$ ème vaut  $\sum_{k \geq n} n(n-1) \cdots (n-k+1) a_k s^{n-k}$ . L'application de ce résultat démontre b). Il ne reste qu'à montrer la continuité de  $\Phi_X$  en  $\pm 1$ , ce qui peut être fait par un argument de convergence dominée.  $\square$

Du deuxième résultat, on en déduit le corollaire suivant.

**Corollaire 2.22.** Si  $X$  et  $Y$  sont deux v.a.d. de même fonction génératrice alors  $\mathbf{P}_X = \mathbf{P}_Y$ , i.e.  $X$  et  $Y$  ont la même loi.

Si la fonction génératrice caractérise la loi, elle caractérise *a fortiori* les moments. La propriété suivante permet de déduire les moments de la fonction caractéristique.

*Notation :* Pour toute fonction  $f$  ayant une limite à gauche (resp. à droite) en  $b$ , on note  $f(b-)$  cette limite (resp.  $f(b+)$ ).

**Théorème 2.23.** Une v.a.  $X : \Omega \rightarrow \mathbb{N}$  admet un moment d'ordre  $p$  si et seulement si  $\Phi_X^{(p)}$  admet une limite à gauche en 1. Alors,

$$\Phi_X^{(p)}(1^-) = \mathbf{E}(X(X-1) \cdots (X-p+1)) .$$

*Démonstration.* On traite le cas  $p = 1$ , le cas général suit le même principe. Rappelons que  $\Phi'_X(s) = \sum_{k \geq 1} k p_X(k) s^{k-1}$ . Supposons que  $\mathbf{E}(X) < \infty$ . Comme  $\mathbf{E}(X) = \sum_{k \geq 1} k p_X(k)$ , les termes  $k p_X(k) s^{k-1}$  de la série  $\Phi'_X(s)$  sont dominés par une suite sommable  $k p_X(k)$ . Par convergence dominée,  $\lim_{s \uparrow 1} \Phi'_X(s) = \sum_{k \geq 1} \lim_{s \uparrow 1} k p_X(k) s^{k-1} = \mathbf{E}(X)$ .

Réciproquement, supposons que  $\Phi'_X(1^-)$  existe. Comme  $\Phi'_X$  est croissante sur  $[0, 1[$ , on a pour tout  $s < 1$ ,  $\sum_{k \geq 1} k p_X(k) s^{k-1} \leq \Phi'_X(1^-)$  et comme tous les termes sont positifs,  $\sum_{k=1}^n k p_X(k) s^{k-1} \leq \Phi'_X(1^-)$  quel que soit  $n$ . En faisant  $s \uparrow 1$  dans la dernière inégalité, on en déduit que la suite  $(\sum_{k=1}^n k p_X(k))_n$  est bornée. C'est une suite croissante, elle est donc convergente. On a bien  $\sum_{k=1}^n k p_X(k) < \infty$ , autrement dit  $\mathbf{E}(X) < \infty$ .  $\square$

L'exercice 2.2 fournit des exemples d'applications.

## 2.5. Exercices

### Pour apprendre

#### EXERCICE 2.1. –

Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite i.i.d. de loi de Bernoulli de paramètre  $p$  sur  $\{0, 1\}$ .

- 2.1.a) Pour tout  $n$ , caractériser la loi de  $S_n = \sum_{k=1}^n X_k$ .
- 2.1.b) En déduire l'espérance et la variance d'une variable binomiale de paramètres  $(n, p)$ .
- 2.1.c) On pose  $Y = \min\{n : X_n = 1\}$  lorsque cet ensemble est non-vide,  $Y = +\infty$  sinon. Caractériser la loi de  $Y$ .

#### EXERCICE 2.2. –

On rappelle que  $\Phi_X$  est la fonction génératrice de la variable aléatoire  $X$  (cf. définition 2.10).

- 2.2.a) Calculer  $\Phi_X$ ,  $\mathbf{E}(X)$  et  $\text{Var}(X)$  pour une v.a. de Bernoulli  $B(p)$ , une v.a. de loi géométrique  $\mathcal{G}(p)$ , une v.a. de Poisson  $\mathcal{P}(\lambda)$ .
- 2.2.b) Soient  $X_1, \dots, X_n$  des v.a. indépendantes.  $X_k$  suit une loi de Poisson de paramètre  $\lambda_k$ . Caractériser la loi de  $\sum_{k=1}^n X_k$ .

#### EXERCICE 2.3. –

Soient  $(X_1, X_2, X_3)$  des variables aléatoires indépendantes de même loi à valeurs dans  $\mathbf{N}$ .

On note  $p_i = \mathbf{P}(X_l = i)$ ,  $l = 1, 2, 3$ . On introduit  $Z$  de loi uniforme sur  $\{1, 2\}$ .

- 2.3.a) Quelle est la loi de  $Y = (X_Z, X_{3-Z})$ ?
- 2.3.b) Soit  $W$  le vecteur aléatoire défini par :

$$W = (X_1, X_3) \text{ si } Z = 2 \text{ et } W = (X_3, X_2) \text{ si } Z = 1.$$

Quelle est la loi de  $W$ ?

#### EXERCICE 2.4. –

Soit  $X$  de loi uniforme sur  $\{0, 1\}$  et  $Z$  de loi uniforme sur  $\{-1, +1\}$  indépendante de  $X$ .

Soit  $Y = ZX$ . Montrer que  $X$  et  $Y$  sont décorrélées mais ne sont pas indépendantes.

#### EXERCICE 2.5. –

On veut calculer les moments d'une v.a. de loi hypergéométrique. On se donne donc une urne contenant  $r$  boules rouges et  $b$  boules blanches de sorte que  $N = r + b$ . Muni d'une épuisette à boules, on tire  $m$  boules parmi les  $N$  présentes. On range ces boules dans des cases numérotées de 1 à  $m$ . On note  $X$  le nombre de boules rouges ressorties et

$$X_i = \begin{cases} 1 & \text{si la case } i \text{ contient une boule rouge,} \\ 0 & \text{sinon.} \end{cases}$$

On a donc  $X = \sum_{i=1}^m X_i$ .

- 2.5.a) Soit  $\sigma$  une permutation de  $\{1, \dots, m\}$ . Pourquoi les vecteurs aléatoires  $(X_1, \dots, X_m)$  et  $(X_{\sigma(1)}, \dots, X_{\sigma(m)})$  ont-ils la même loi?
- 2.5.b) Calculer  $\mathbf{P}(X_i = 1)$  et  $\mathbf{P}(X_i X_j = 1)$  pour  $i \neq j$ .
- 2.5.c) En déduire  $\mathbf{E}[X]$  et  $\text{Var}(X)$ .



## Pour s'entraîner

### EXERCICE 2.6. –

Soient  $1 \leq n \leq N$  deux entiers. Soit  $M$  une v.a. de loi binomiale  $(N, \theta)$  et  $X$  une v.a. dont la loi est donnée par

$$\mathbf{P}(X = k | M = m) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}} \text{ pour tout } k \in \{0, \dots, n\}.$$

- 2.6.a) Calculer  $\mathbf{P}(X = k)$ .
- 2.6.b) Calculer la loi de  $M$  sachant  $X = k$ , dite loi *a posteriori* de  $M$ .
- 2.6.c) Pour  $k = 0$ , identifier cette loi.

### EXERCICE 2.7. –

En codage correcteur d'erreurs, les erreurs interviennent au hasard sur l'un quelconque des bits. Si on transmet des mots de  $n$  bits, on pose  $\Omega = \{0, 1\}^n$ , que l'on munit de la loi uniforme. On introduit  $X_i(\omega) = \omega_i$  pour  $i = 1, \dots, n$ . La distance de Hamming entre mots de code  $x = (x_1, \dots, x_n)$  et  $y = (y_1, \dots, y_n)$ , est définie par :

$$d(x, y) = \sum_{i=1}^n \mathbf{1}_{\{x_i \neq y_i\}}.$$

On appelle longueur d'un mot  $x$ , sa distance au mot nul  $0 = (0, \dots, 0)$ .

- 2.7.a) Montrer que sous la loi uniforme sur  $\Omega$ , les variables  $(X_i, i \in \{1, \dots, n\})$  sont indépendantes et identiquement distribuées de loi de Bernoulli de paramètre  $1/2$ .
- 2.7.b) Quelle est la longueur moyenne d'un mot ?
- 2.7.c) Quelle est la variance de la longueur d'un mot ?
- 2.7.d) On choisit deux mots au hasard indépendamment l'un de l'autre, soit  $X$  et  $Y$  les variables aléatoires correspondantes. Calculer

$$\mathbf{E} [d(X, Y)^2].$$

## Pour aller plus loin

### EXERCICE 2.8. –

Un étang contient un nombre de poissons  $N$  inconnu. Pour estimer  $N$ , on prélève un échantillon de  $r$  poissons que l'on marque et que l'on remet dans l'étang. Une semaine plus tard, un autre échantillon de  $s < r$  individus est prélevé. On appelle  $X$  le nombre de poissons marqués lors du premier prélèvement qui sont aussi dans le deuxième échantillon.

- 2.8.a) Calculer la loi de  $X$  (dite loi hypergéométrique).

On note pour la suite de cet exercice

$$p_k = \frac{\binom{r}{k} \binom{N-r}{s-k}}{\binom{N}{s}},$$

pour  $k \leq \min(r, s)$  et  $k \geq \max(s + r - N, 0)$ .

2.8.b) Montrer que  $p_k^2 \geq p_{k-1}p_{k+1}$ .

2.8.c) En déduire qu'il existe une unique valeur de  $k$  telle que  $p_k = \max_j p_j$ .

2.8.d) Soit  $k_0$  tel cette valeur. Par définition,  $p_{k_0+1} < p_{k_0}$  et  $p_{k_0-1} < p_{k_0}$ . En déduire que

$$k_0 = \left\lfloor \frac{(r+1)(s+1)}{N+2} \right\rfloor.$$

On pourra poser pour simplifier les calculs,  $r' = r + 1$ ,  $s' = s + 1$ ,  $N' = N + 2$ .

2.8.e) En déduire une estimation de  $N$ .

**EXERCICE 2.9** (\*\*\*, Erdős et Renyi (1960)).—

On fabrique un graphe sur  $n$  sommets en choisissant ses arêtes n au hasard. Plus précisément, on considère le graphe  $G_{n,p}$  obtenu en choisissant chacune des  $\binom{n}{2}$  arêtes potentielles indépendamment avec probabilité  $p$ . Le but de ce problème est d'étudier la probabilité que  $G_{n,p}$  soit connexe. On s'intéressera au cas où  $p$  est de la forme

$$p = p(n) = \frac{\ln n}{n} + \frac{c}{n}$$

où  $c$  est une constante fixée.

2.9.a) Soit  $(X_i, 1 \leq i \leq n)$  un  $n$ -uplet de variables aléatoires à valeurs dans  $\{0, 1\}$  et soit  $X = \sum_{i=1}^n X_i$ . Montrer que pour tout  $r$  tel que  $r \geq 1$  et  $2r + 1 \leq n$  on a :

$$\sum_{k=0}^{2r+1} (-1)^k F^{(k)} \leq \mathbf{P}(X = 0) \leq \sum_{k=0}^{2r} (-1)^k F^{(k)}$$

où l'on a posé  $F^{(0)} = 1$  et pour  $k \geq 1$

$$F^{(k)} = \sum_{j_1 < j_2 < \dots < j_k} \mathbf{E}[X_{j_1} X_{j_2} \dots X_{j_k}].$$

*Suggestion.* On pourra montrer que

$$\mathbf{P}(X = 0) = \mathbf{E} \left[ \prod_{i=1}^n (1 - X_i) \right]$$

et appliquer une formule de Taylor à la fonction  $\prod_{i=1}^n (1 - x_i)$ .

On dira qu'un sommet est isolé s'il n'est l'extrémité d'aucune arête. Dans un premier temps, on étudie le nombre  $X$  de sommets isolés. On peut écrire  $X = \sum_{i=1}^n X_i$  où  $X_i$  est la variable aléatoire qui vaut 1 si le sommet  $i$  est isolé, 0 sinon.

2.9.b) Que valent  $\mathbf{E}[X_i]$  et  $\mathbf{E}[X]$  ?

2.9.c) On suppose dorénavant  $c$  fixé. Montrer que la quantité  $F^{(k)}$ , pour la variable  $X$ , converge, lorsque  $n$  tend vers l'infini, vers  $e^{-ck}/k!$ .

2.9.d) Montrer que  $\lim_{n \rightarrow \infty} \mathbf{P}(X = 0) = e^{-e^{-c}}$ .

2.9.e) Calculer l'espérance du nombre de composantes connexes à 2 sommets, et constater que celle-ci tend vers zéro quand  $n$  tend vers l'infini.

2.9.f) Plus généralement, soit  $C_t$  le nombre de composantes connexes à  $t$  sommets. Montrer que pour  $2 \leq t \leq n/2$ ,

$$\mathbf{E}[C_t] \leq \frac{1}{t!} \sum_{t-1 \leq k \leq \binom{t}{2}} \binom{\binom{t}{2}}{k} \left( \frac{p}{1-p} \right)^k.$$

En déduire que la probabilité que  $G_{n,p}$  soit connexe tend, quand  $n \rightarrow \infty$ , vers  $e^{-e^{-c}}$ . On admettra que  $\sum_{2 \leq t \leq n/2} \mathbf{E}[C_t] \rightarrow 0$  quand  $n \rightarrow \infty$ .

2.9.g) Que peut-on dire de la probabilité que  $G_{n,p}$  soit connexe ?

**EXERCICE 2.10.**—

On rappelle qu'une suite de variables aléatoires  $(X_n, n \in \mathbf{N})$  converge en probabilité vers la variable aléatoire  $X$  si et seulement si pour tout  $\epsilon > 0$ ,

$$\lim_{n \rightarrow +\infty} \mathbf{P}(|X_n - X| \geq \epsilon) = 0.$$

Soit  $(X_n, n \in \mathbf{N})$  une suite de v.a. de moyenne  $\mu_n$  et de variance  $\sigma_n^2$ . Soit  $(b_n, n \in \mathbf{N})$  une suite de réels positifs tels que  $\sigma_n^2/b_n^2$  tende vers 0. Montrer que

$$\frac{X_n - \mu_n}{b_n} \text{ tend vers 0 en probabilité.}$$

**EXERCICE 2.11** (Borne de Chernoff).—

Soit  $X$  une v.a. de loi de Poisson de paramètre  $\lambda$ .

2.11.a) Montrer que  $\mathbf{P}(X \geq \eta) = \mathbf{P}(\exp(\theta X) \geq \exp(\theta \eta))$  pour tout  $\theta > 0$ .

2.11.b) Montrer que, pour tout  $\theta \geq 0$ ,

$$\mathbf{P}(X \geq \eta) \leq e^{-\eta\theta} \mathbf{E}[\exp(\theta X)]. \quad (2.10)$$

2.11.c) Calculer  $\mathbf{E}[\exp(\theta X)]$ .

2.11.d) Trouver  $\theta$  qui minimise le terme de droite de (2.10).

2.11.e) Trouver  $K$  tel que  $\mathbf{P}(X \geq K\lambda) \leq 0,001$ .

**EXERCICE 2.12.**—

On veut collectionner  $N$  images dont une et une seule apparaît dans chaque tablette de chocolat achetée. Les images sont mises au hasard dans les tablettes. On appelle  $T_i$  le nombre de tablettes nécessaires avant d'avoir  $i$  images distinctes. On pose  $T_0 = 0$ .

2.12.a) Montrer que  $T_{i+1} - T_i$  suit une loi géométrique de paramètre  $1 - i/N$ .

2.12.b) Montrer que les variables aléatoires  $T_0, T_1 - T_0, \dots, T_N - T_{N-1}$  sont indépendantes dans leur ensemble.

2.12.c) Calculer l'espérance et la variance de  $T_N$ . Trouver un équivalent de l'espérance et montrer que la variance est un  $O(N^2)$  quand  $N$  tend vers  $+\infty$ .

2.12.d) En utilisant l'exercice 2.10, montrer que  $T_N/(N \log N)$  tend vers 1 en probabilité.

**EXERCICE 2.13.**—

Les règles du jeu du **not-seven** sont les suivantes : on part d'un score  $X_0 = 0$ . À chaque coup, on lance deux dés non pipés, si la somme des faces égale 7, le score retourne à 0 et la partie est terminée. Sinon, le score augmente de la somme des faces et on a le droit de rejouer ou pas. Si l'on ne rejoue pas, le score est acquis et la partie est terminée. Si l'on rejoue, on relance les deux dés avec la même règle.

2.13.a) Calculer la loi de la somme  $S$  des deux faces. Calculer son espérance.

On considère une suite  $(S_n, n \in \mathbf{N})$  de variables aléatoires indépendantes de même loi que  $S$ .

2.13.b) Soit  $\tau = \inf\{n \geq 1, S_n = 7\}$ , trouver la loi de  $\tau$ . Quelle est la moyenne de  $\tau$  ?

2.13.c) Quelle est la stratégie d'un Initié (celui qui sait le résultat du prochain lancer de dés) ?

2.13.d) Calculer son gain moyen.

2.13.e) On appelle  $X_n$  le score au  $n$ -ième coup en l'absence de stratégie d'arrêt. Montrer que

$$\mathbf{E}[X_{n+1} | X_n = i] = \frac{5}{6}i + \frac{35}{6},$$

où l'espérance conditionnelle par rapport à un événement  $B$  est définie comme l'espérance associée à la loi de probabilité  $A \mapsto \mathbf{P}(A | B)$ .

2.13.f) En déduire que la stratégie optimale consiste à jouer tant que l'on n'a pas atteint 35 et à s'arrêter immédiatement après avoir franchi ce seuil.

2.13.g) Calculer numériquement le gain moyen avec cette stratégie.

**EXERCICE 2.14.**—

Peut-on piper deux dés de sorte que la loi de leur somme soit la loi uniforme sur  $\{2, \dots, 12\}$  ?

**EXERCICE 2.15 (\*\*, Processus de branchement).**—

Soit  $X_0$  une v.a. à valeurs entières. Soit  $(X_{n,j}, n \geq 1, 1 \leq j \leq n)$  une famille dénombrable de variables aléatoires indépendantes, de loi  $\mathbf{P}_{X_0}$ . On note  $\Phi$  la fonction génératrice de  $\mathbf{P}_{X_0}$ . On considère un individu « racine » qui a un nombre  $X_0$  de descendants. Chacun de ses descendants a un nombre aléatoire de descendant, ce nombre est indépendant de celui des autres descendants et de loi  $\mathbf{P}_{X_0}$ . On pose  $Z_n$  le nombre total d'individus au rang  $n$ .

1. Calculer la fonction génératrice de  $Z_n$  en fonction de celle de  $Z_{n-1}$ .
2. Soit  $u_n = \mathbf{P}(Z_n = 0)$ . Montrer que  $u_n = \Phi(u_{n-1})$ .
3. Trouver des conditions nécessaires et suffisantes sur  $\mathbf{P}_{X_0}$  qui garantissent que  $\Phi$  est strictement convexe.
4. Montrer que  $u$  converge vers une limite non nulle si et seulement si  $\mathbf{E}[X_0] < 1$ .

*Ce processus représente tout aussi bien l'évolution de la contamination par un virus ( $X_0$  est le nombre d'individus contaminés par le malade initial), que la transmission d'un nom de famille ( $X_0$  étant alors le nombre d'enfants portant le nom de leur père) et bien d'autres situations.*

EXERCICE 2.16 (\*\*).—

Dans le tri rapide (quicksort), on note  $M_n$  le nombre de comparaisons nécessaires pour ordonner un tableau de  $n$  nombres. Montrer que  $\mathbf{E}[M_n]$  vérifie la relation

$$\mathbf{E}[M_n] = n - 1 + \frac{2}{n} \sum_{k=1}^{n-1} \mathbf{E}[M_k].$$

En déduire que

$$\mathbf{E}[M_n] = 2(n+1) \sum_{i=1}^{n-1} \frac{i}{(i+1)(i+2)}$$

et trouver un équivalent asymptotique de  $M_n$  quand  $n$  tend vers  $+\infty$ .



## 3. Chaînes de Markov

*De par leur universalité et leur relative simplicité d'utilisation, les chaînes de Markov sont l'un des outils les plus fréquents de modélisation de phénomènes aléatoires en temps discret. Ce chapitre n'utilise que les notions que vous avez déjà vues mais permet d'aller plus loin dans les applications des probabilités. Vous rencontrerez des chaînes de Markov dans les cours de communications numériques, de signal, de réseaux, etc.*

La loi des grands nombres repose sur l'hypothèse de variables aléatoires indépendantes et identiquement distribuées. Si ces variables aléatoires représentent l'évolution de l'état d'un système au cours du temps, cette hypothèse n'est généralement pas satisfaite car les états du système sont corrélés dans le temps. Cependant, si le système a une mémoire limitée, à savoir si le changement d'état du système entre les instants  $n$  et  $n + 1$  ne dépend pas de l'état du système au temps  $n - 1$ , alors on peut obtenir une loi des grands nombres, liée à l'état d'équilibre du système : c'est le théorème ergodique. Cette propriété de « mémoire limitée » définit les chaînes de Markov, outil central des probabilités, utilisé dans de très nombreux domaines scientifiques et techniques.

On se restreint ici aux chaînes de Markov en temps discret sur un espace d'états fini ; les résultats se généralisent au temps continu et à un espace d'états dénombrable notamment.

### 3.1. Préliminaires

Dans ce chapitre, plus encore que dans les autres, la notion de probabilité conditionnelle, qui sera vue en détail dans le chapitre 6, joue un rôle primordial. Comme toutes les variables aléatoires sont définies sur un espace fini, les formules sont simples et les constructions élémentaires.

Rappelons que pour un événement  $B$  de probabilité non nulle, la mesure de probabilité

$$A \longmapsto \mathbf{P}(A | B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$$

est appelée la probabilité conditionnelle selon  $B$ . On notera au passage le lemme élémentaire mais indispensable :

$$\mathbf{P}(A \cap B | C) = \mathbf{P}(A | B \cap C) \mathbf{P}(B | C). \quad (3.1)$$

## 3.2. Définition

**Définition 3.1.** Soit  $(X_n, n \in \mathbf{N})$  une suite de variables aléatoires à valeurs dans  $\{1, \dots, N\}$ . La suite  $(X_n)_{n \in \mathbf{N}}$  est une chaîne de Markov si

$$\begin{aligned} \forall n \in \mathbf{N}, \quad \forall x_0, x_1, \dots, x_{n+1} \in \{1, \dots, N\}, \\ \mathbf{P}(X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mathbf{P}(X_{n+1} = x_{n+1} | X_n = x_n). \end{aligned}$$

On dit d'une chaîne de Markov qu'elle « oublie son passé » : conditionnellement à la variable aléatoire  $X_n$  (l'état présent), la variable aléatoire  $X_{n+1}$  (l'état futur) ne dépend pas des variables aléatoires  $X_0, \dots, X_{n-1}$  (les états passés). Plus généralement, toute l'évolution future d'une chaîne de Markov ne dépend que de son état présent :

**Théorème 3.1.** Pour tout  $n \in \mathbf{N}$ , pour tout  $k \geq 1$ , pour tout  $x_0, x_1, \dots, x_{n+k} \in \{1, \dots, N\}$ , on a :

$$\begin{aligned} \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ = \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} | X_n = x_n). \end{aligned}$$

*Démonstration.* Par récurrence sur  $k$ . La propriété est vraie pour  $k = 1$  par définition. On suppose qu'elle est vraie pour un certain  $k \geq 1$ . Alors, pour tout  $n \in \mathbf{N}$ , pour tout  $x_0, x_1, \dots, x_{n+k} \in \{1, \dots, N\}$ , on a

$$\begin{aligned} & \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k+1} = x_{n+k+1} | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ &= \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ & \quad \times \mathbf{P}(X_{n+k+1} = x_{n+k+1} | X_0 = x_0, X_1 = x_1, \dots, X_{n+k} = x_{n+k}), \\ &= \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} | X_n = x_n) \times \mathbf{P}(X_{n+k+1} = x_{n+k+1} | X_{n+k} = x_{n+k}), \\ &= \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} | X_n = x_n) \\ & \quad \times \mathbf{P}(X_{n+k+1} = x_{n+k+1} | X_n = x_n, X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k}), \\ &= \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k+1} = x_{n+k+1} | X_n = x_n), \end{aligned}$$

où l'on a utilisé (3.1), l'hypothèse de récurrence et la définition des chaînes de Markov pour obtenir la seconde égalité, et à nouveau la définition des chaînes de Markov pour obtenir la troisième égalité.  $\square$

Pour prédire l'évolution d'une chaîne de Markov, on peut simplement considérer les maillons de cette chaîne, correspondant aux transitions successives d'un état à l'autre :



**Théorème 3.2.** On a :

$$\begin{aligned} \forall n \in \mathbf{N}, \forall k \geq 1, \quad \forall x_n, x_{n+1}, \dots, x_{n+k} \in \{1, \dots, N\}, \\ \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \mid X_n = x_n) = \mathbf{P}(X_{n+1} = x_{n+1} \mid X_n = x_n) \times \\ \mathbf{P}(X_{n+2} = x_{n+2} \mid X_{n+1} = x_{n+1}) \times \dots \times \mathbf{P}(X_{n+k} = x_{n+k} \mid X_{n+k-1} = x_{n+k-1}). \end{aligned}$$

*Démonstration.* Par la formule des probabilités composées,

$$\begin{aligned} \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_{n+k} = x_{n+k} \mid X_n = x_n) \\ = \mathbf{P}(X_{n+1} = x_{n+1} \mid X_n = x_n) \times \mathbf{P}(X_{n+2} = x_{n+2} \mid X_{n+1} = x_{n+1}, X_n = x_n) \\ \times \dots \times \mathbf{P}(X_{n+k} = x_{n+k} \mid X_n = x_n, \dots, X_{n+k-1} = x_{n+k-1}). \end{aligned}$$

La preuve découle alors de la définition des chaînes de Markov.  $\square$

### 3.3. Matrice de transition

D'après le théorème 3.2, l'évolution d'une chaîne de Markov est entièrement déterminée par ses probabilités de transition,  $\mathbf{P}(X_{n+1} = j \mid X_n = i)$  pour tous  $i, j \in \{1, \dots, N\}$ . Par la suite, on ne considèrera que des chaînes de Markov *homogènes* (sous-entendues en temps), pour lesquelles ces probabilités de transition ne dépendent pas de  $n$  :

$$\forall n \in \mathbf{N}, \quad \mathbf{P}(X_{n+1} = j \mid X_n = i) = \mathbf{P}(X_1 = j \mid X_0 = i).$$

**Définition 3.2.** L'évolution de la chaîne de Markov ne dépend alors que de sa *matrice de transition*  $P$ , définie par :

$$\forall i, j = 1, \dots, N, \quad P_{ij} = \mathbf{P}(X_1 = j \mid X_0 = i).$$

C'est une matrice *stochastique*, au sens où

$$\forall i = 1, \dots, N, \quad \sum_j P_{ij} = 1,$$

soit  $P\mathbf{1} = \mathbf{1}$  en notation vectorielle, avec  $\mathbf{1} = (1, \dots, 1)^T$ .

EXEMPLE 9.— La matrice suivante est stochastique et définit donc une chaîne de Markov :

$$P = \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/3 & 0 & 2/3 \\ 0 & 1/5 & 4/5 \end{pmatrix}.$$

En particulier, la chaîne passe de l'état 1 à l'état 2 avec probabilité  $1/2$ .

Par la loi des probabilités totales, on a pour tout  $n \in \mathbf{N}$ ,

$$\forall i = 1, \dots, N, \quad \mathbf{P}(X_{n+1} = i) = \sum_{j=1}^N \mathbf{P}(X_n = j) \mathbf{P}(X_{n+1} = i | X_n = j).$$

En notant la loi de  $X_n$  comme un vecteur ligne  $\pi(n)$  de dimension  $N$ , avec  $\pi(n)\mathbf{1} = 1$ , cela s'écrit :

$$\forall i = 1, \dots, N, \quad \pi(n+1)_i = \sum_{j=1}^N \pi(n)_j P_{ji},$$

soit  $\pi(n+1) = \pi(n)P$  en notation vectorielle, dont on déduit

$$\forall n \in \mathbf{N}, \quad \pi(n) = \pi(0)P^n. \quad (3.2)$$

En prenant  $\pi(0)$  le vecteur nul sauf la  $i$ -ième composante égale à 1, on obtient

$$\pi(n) = (\mathbf{P}(X_n = j | X_0 = i), 1 \leq j \leq N) = ((P^n)_{ij}, 1 \leq j \leq N)$$

donc

$$\forall n \in \mathbf{N}, \quad (P^n)_{ij} = \mathbf{P}(X_n = j | X_0 = i). \quad (3.3)$$

EXEMPLE 10.— Soit  $P = \begin{pmatrix} 0 & 1 \\ 1/3 & 2/3 \end{pmatrix}$ . On a en particulier :

$$P^2 = \begin{pmatrix} 1/3 & 2/3 \\ 2/9 & 7/9 \end{pmatrix}.$$

Ainsi pour  $\pi(0) = (1, 0)$  (la chaîne est initialement dans l'état 1), alors  $\pi(1) = (0, 1)$  (la chaîne est dans l'état 2 au temps 1) et  $\pi(2) = (1/3, 2/3)$  (la chaîne est dans l'état 1 avec probabilité  $1/3$  et dans l'état 2 avec probabilité  $2/3$  au temps 2).

REMARQUE.— Soit  $P$  la matrice de transition d'une chaîne de Markov  $X_0, X_1, X_2, \dots$ . Comme  $P\mathbf{1} = \mathbf{1}$ , on a  $P^n\mathbf{1} = \mathbf{1}$  pour tout entier  $n \geq 1$  de sorte que  $P^n$  est la matrice de transition d'une chaîne de Markov. Il s'agit de la chaîne de Markov initiale échantillonnée aux temps multiples de  $n$ , soit  $X_0, X_n, X_{2n}, \dots$ .

Supposons qu'il existe une loi de probabilité  $\pi$  telle que :

$$\forall i = 1, \dots, N, \quad \pi_i = \sum_j \pi_j P_{ji}, \quad (3.4)$$

soit  $\pi = \pi P$  sous forme vectorielle. Alors  $\pi$  est une loi stationnaire de la chaîne de Markov au sens où, d'après (3.3),  $\pi(0) = \pi$  implique  $\pi(n) = \pi$  pour tout  $n$ . Nous reviendrons sur les conditions d'existence et d'unicité d'une telle loi stationnaire, et sur la convergence de  $\pi(n)$  vers  $\pi$  lorsque  $\pi(0) \neq \pi$ . Les équations (3.4) s'appellent les *équations d'équilibre* de la chaîne de Markov.

### 3.4. Graphe de transition

On représente généralement une chaîne de Markov par son *graphe de transition*, graphe orienté dont les sommets sont les états de la chaîne de Markov et les arcs sont les transitions possibles entre états : il existe un arc de  $i$  vers  $j$  si et seulement si  $P_{ij} > 0$  ; on associe alors à cet arc le poids  $P_{ij}$  (voir figure 3.1). La somme des poids des arcs sortants de tout sommet doit donc être égale à 1.

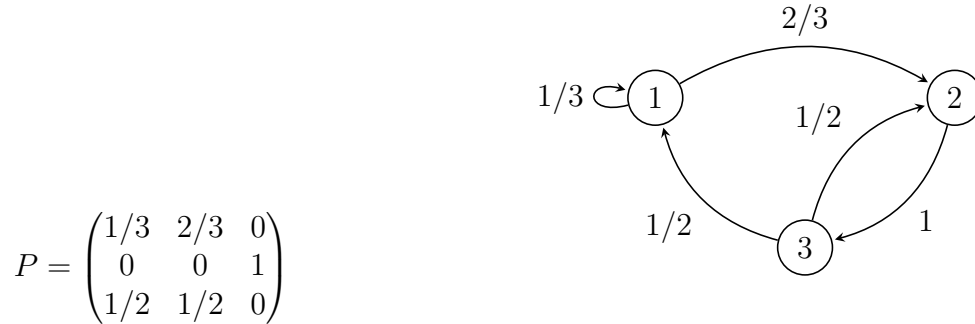


Figure 3.1. – Matrice et graphe de transition d'une chaîne de Markov à trois états.

D'après le théorème 3.2, la probabilité d'une suite d'états de la chaîne de Markov s'obtient en faisant le *produit* des poids le long du chemin correspondant dans le graphe de transition. Ainsi pour l'exemple de la figure 3.1, la probabilité que, partant de l'état initial  $X_0 = 1$ , la chaîne de Markov prenne les états successifs 2, 3, 1, 1, 2 est égale à :

$$\frac{2}{3} \times 1 \times \frac{1}{2} \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{81}.$$

### 3.5. Irréductibilité

**Définition 3.3.** Une chaîne de Markov est dite *irréductible* si son graphe de transition est fortement connexe, c'est-à-dire si pour tous sommets distincts  $i, j$ , il existe un chemin de  $i$  vers  $j$  dans le graphe de transition. La propriété d'irréductibilité ne dépend donc que de la structure du graphe de transition et non des poids de ce graphe.

Une chaîne de Markov irréductible « visite » tous ses états infiniment souvent tandis qu'une chaîne de Markov non irréductible finit par ne prendre qu'un sous-ensemble strict des états possibles. Pour les chaînes de Markov de la figure 3.2 par exemple, la chaîne de gauche prendra les états 1, 2, 3 infiniment souvent tandis que la chaîne de droite finira par ne prendre que les états 2, 3 (on dit que l'ensemble  $\{2, 3\}$  est *absorbant*).

Le résultat suivant permet de caractériser une chaîne de Markov irréductible à partir de sa matrice de transition.

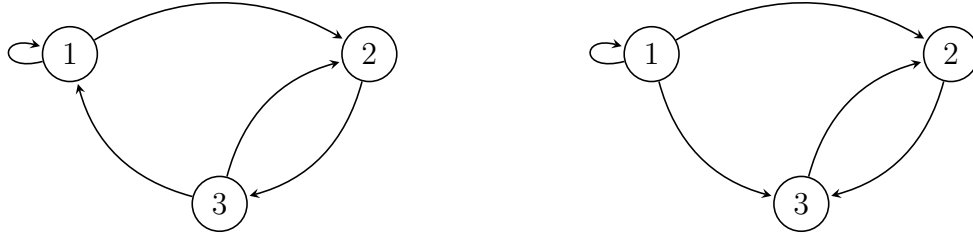


Figure 3.2. – Graphes de transition non valués d’une chaîne de Markov irréductible (à gauche) et d’une chaîne de Markov non irréductible (à droite).

**Proposition 3.3.** Une chaîne de Markov de matrice de transition  $P$  est irréductible si et seulement s’il existe un entier  $n \geq 1$  tel que la matrice  $(I + P)^n$  est strictement positive.

*Démonstration.* On commence par la condition suffisante. Soit  $n \geq 1$  tel que la matrice  $(I + P)^n$  est strictement positive. Comme

$$(I + P)^n = \sum_{k=0}^n \binom{n}{k} P^k,$$

il existe pour toute paire d’états distincts  $i \neq j$  un entier  $k$  tel que  $(P^k)_{ij} > 0$ , et donc un chemin de longueur  $k$  de  $i$  à  $j$  dans le graphe de transition.

Pour la condition nécessaire, il suffit de prendre  $n = N$  car il existe pour toute paire d’états distincts  $i \neq j$  un chemin de longueur inférieure à  $N$  de  $i$  à  $j$  dans le graphe de transition.  $\square$

**Théorème 3.4** (Perron-Frobenius (partiel)). Une chaîne de Markov irréductible admet une unique loi stationnaire.

*Démonstration.* La matrice  $P$  étant stochastique admet  $\mathbf{1} = (1, \dots, 1)^T$  comme vecteur propre droit associé à la valeur propre 1, soit  $P\mathbf{1} = \mathbf{1}$ . La matrice transposée  $P^T$  ayant les mêmes valeurs propres que  $P$ , il existe un vecteur propre gauche  $v$  de  $P$  associé à la valeur propre 1, soit  $vP = v$ .

On montre tout d’abord que le vecteur ligne  $|v|$  satisfait également l’égalité  $|v|P = |v|$ . On a en effet pour tout  $j$ ,

$$|v_j| = \left| \sum_i v_i P_{ij} \right| \leq \sum_i |v_i| P_{ij}.$$

En sommant ces inégalités, on trouve

$$\sum_j |v_j| \leq \sum_j \sum_i |v_i| P_{ij} = \sum_i |v_i| \sum_j P_{ij} = \sum_i |v_i|,$$

de sorte que toutes les inégalités précédentes sont en fait des égalités.

On vérifie ensuite que le vecteur  $|v|$  est strictement positif. Par irréductibilité, il existe en effet d'après la Proposition 3.3 un entier  $n \geq 1$  tel que la matrice  $(I + P)^n$  est strictement positive. Le résultat découle alors du fait que

$$|v|(I + P)^n = |v| \sum_{k=0}^n \binom{n}{k} P^k = 2^n |v|.$$

On sait donc que tout vecteur propre  $v$  de  $P$  pour la valeur propre 1 n'a aucune coordonnée nulle. Il reste à montrer que ce vecteur est de signe constant, soit  $v = |v|$  ou  $v = -|v|$ . Ceci vient du fait que si le vecteur  $v + |v|$  est non nul, c'est un vecteur propre gauche de  $P$  pour la valeur propre 1 ; comme un tel vecteur propre n'a aucune coordonnée nulle, cela implique que toutes les coordonnées du vecteur  $v$  sont positives, c'est-à-dire que  $v = |v|$ .

Enfin, on définit le vecteur ligne  $\pi$  par renormalisation de  $|v|$  :

$$\pi_j = \frac{|v_j|}{\sum_{i=1}^N |v_i|}.$$

C'est un vecteur strictement positif qui satisfait  $\pi P = \pi$  et  $\sum_j \pi_j = 1$ . Par ailleurs, s'il existe un autre vecteur  $\pi'$  qui satisfait ces égalités, alors le vecteur  $\pi - \pi'$  est nul ou vecteur propre gauche de  $P$  pour la valeur propre 1. Mais si le vecteur  $\pi - \pi'$  était vecteur propre gauche de  $P$  pour la valeur propre 1, il serait de signe constant, ce qui est impossible car les vecteurs  $\pi$  et  $\pi'$  sont tous deux positifs et de somme unitaire. On en déduit que  $\pi' = \pi$ , et donc que la loi stationnaire de la chaîne de Markov est unique.  $\square$

La preuve du théorème permet en outre de conclure que la loi stationnaire d'une chaîne de Markov irréductible a pour support l'ensemble des états :

**Corollaire 3.5.** La loi stationnaire  $\pi$  d'une chaîne de Markov irréductible vérifie  $\pi_i > 0$  pour tout  $i = 1, \dots, N$ .

EXEMPLE 11.— Considérons une chaîne de Markov de matrice de transition :

$$P = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

Cette chaîne de Markov est irréductible et admet donc une unique loi stationnaire. Les équations d'équilibre (3.4), soit  $\pi = \pi P$ , s'écrivent :

$$\begin{aligned} \pi_1 &= \frac{1}{3}\pi_1 + \frac{1}{2}\pi_3 \\ \pi_2 &= \frac{2}{3}\pi_1 + \frac{1}{2}\pi_3 \\ \pi_3 &= \pi_2 \end{aligned}$$

Ces équations définissent un vecteur propre gauche de  $P$  pour la valeur propre 1 et ne sont donc pas indépendantes (la solution est unique à une constante multiplicative près). La dernière équation définissant la loi stationnaire  $\pi$  de manière unique est la condition de normalisation :

$$\pi_1 + \pi_2 + \pi_3 = 1.$$

On trouve  $\pi = (3/11, 4/11, 4/11)$ .

L'irréductibilité n'est pas une condition *nécessaire* d'existence et d'unicité d'une loi stationnaire. La chaîne de Markov associée au graphe de transition de droite de la figure 3.2 a pour unique loi stationnaire  $\pi = (0, 1/2, 1/2)$  par exemple. Par contre, l'irréductibilité est une condition nécessaire pour que la loi stationnaire ait pour support l'ensemble des états.

Une chaîne de Markov admet en fait toujours une loi stationnaire (dans le cas d'un nombre fini d'états considéré ici). Il suffit de décomposer le graphe de transition en ensembles absorbants fortement connexes et de considérer la restriction de la chaîne de Markov à chacun de ces ensembles : chacune de ces chaînes de Markov « réduites » définit une chaîne de Markov irréductible, qui admet donc une loi stationnaire. Toute combinaison convexe de ces lois donne une loi stationnaire pour la chaîne de Markov initiale ; il n'y a ainsi pas unicité en général.

EXEMPLE 12. – Le graphe de transition de la figure 3.3 a deux ensembles absorbants fortement connexes :  $\{4, 5\}$  et  $\{6, 7, 8\}$ . Les lois stationnaires associées sont  $\pi^{(1)} = (0, 0, 0, 1/2, 1/2, 0, 0, 0)$  et  $\pi^{(2)} = (0, 0, 0, 0, 0, 1/3, 1/3, 1/3)$  ; on a ainsi  $\pi^{(1)}P = \pi^{(1)}$  et  $\pi^{(2)}P = \pi^{(2)}$ . Toute combinaison convexe de ces deux lois est une loi stationnaire de la chaîne de Markov.

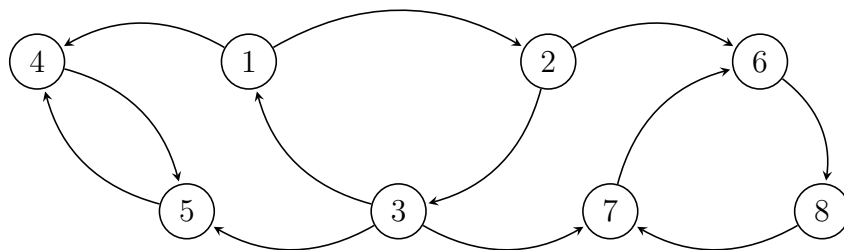


Figure 3.3. – Graphe de transition non valué d'une chaîne de Markov non irréductible, avec deux ensembles absorbants fortement connexes :  $\{4, 5\}$  et  $\{6, 7, 8\}$ .

## 3.6. Périodicité

Certaines chaînes de Markov ont un comportement périodique au sens où les mêmes sous-ensembles d'états sont visités de manière cyclique.

**Définition 3.4.** On appelle *période* d'un état le plus grand commun diviseur des longueurs des cycles du graphe transition passant par cet état. On dit d'un état qu'il est *apériodique* lorsqu'il est de période 1.

**Proposition 3.6.** Tous les états d'une chaîne de Markov irréductible ont la même période.

*Démonstration.* Soit  $i, j$  deux états distincts de périodes respectives  $d_i, d_j$ . La chaîne de Markov étant irréductible, il existe  $n, m$  tels que  $(P^n)_{ij} > 0$  et  $(P^m)_{ji} > 0$ . Comme

$$(P^{n+m})_{ii} \geq (P^n)_{ij}(P^m)_{ji} > 0,$$

il existe un cycle de longueur  $n + m$  passant par  $i$ , si bien que  $d_i$  divise  $n + m$ . Par ailleurs, pour tout cycle de longueur  $k$  passant par  $j$ ,

$$(P^{n+k+m})_{ii} \geq (P^n)_{ij}(P^k)_{jj}(P^m)_{ji} > 0,$$

de sorte qu'il existe un cycle de longueur  $n + k + m$  passant par  $i$ . Ainsi  $d_i$  divise  $n + k + m$ , et comme  $d_i$  divise  $n + m$ ,  $d_i$  divise  $k$ . Finalement, comme  $d_i$  divise toutes les longueurs de cycles passant par  $j$ ,  $d_i$  divise  $d_j$ . Par symétrie,  $d_j$  divise  $d_i$  et  $d_i = d_j$ .  $\square$

On parle donc sans ambiguïté de *période* d'une chaîne de Markov irréductible, qui est la période de tous ses états.

**Théorème 3.7.** Une chaîne de Markov  $(X_n)_{n \in \mathbf{N}}$  irréductible de période  $d$  admet une partition  $C_0, \dots, C_{d-1}$  de ses états telle que

$$\forall k = 0, \dots, d-1, \forall n \in \mathbf{N}, \quad \mathbf{P}(X_n \in C_{k+n \bmod d} \mid X_0 \in C_k) = 1. \quad (3.5)$$

Les chaînes de Markov  $(X_{nd})_{n \in \mathbf{N}}$  restreintes aux ensembles  $C_0, \dots, C_{d-1}$  sont irréductibles, de lois stationnaires respectives  $\pi^{(0)}, \dots, \pi^{(d-1)}$  telles que  $\pi^{(1)} = \pi^{(0)}P, \dots, \pi^{(d)} = \pi^{(d-1)}P$  et  $\pi^{(0)} = \pi^{(d)}P$ . La loi stationnaire de la chaîne de Markov  $(X_n)_{n \in \mathbf{N}}$  est donnée par :

$$\pi = \frac{1}{d}(\pi^{(0)} + \dots + \pi^{(d-1)}). \quad (3.6)$$

*Démonstration.* Pour un état fixé, par exemple l'état 1, on note  $C_k$  l'ensemble des états atteignables à partir de l'état 1 en un temps égal à  $k$  modulo  $d$ , pour tout  $k = 0, \dots, d-1$ . Les ensembles  $C_0, \dots, C_{d-1}$  forment une partition des états. En effet, supposons qu'il existe un état  $i \in C_{k_1} \cap C_{k_2}$  pour  $k_1 < k_2$ ; par irréductibilité, il existe un chemin de longueur  $\ell \geq 0$  de  $i$  à 1 dans le graphe de transition. Il existe donc deux cycles passant par 1 dans le graphe de transition, de longueurs respectives égales à  $k_1 + \ell$  et  $k_2 + \ell$  modulo  $d$ . Comme tous les états sont de période  $d$ ,  $d$  divise  $k_1 + \ell$  et  $k_2 + \ell$ ;  $d$  divise donc  $k_2 - k_1$ , un nombre compris strictement entre 0 et  $d$ , une contradiction. L'état 1 étant arbitrairement choisi, les ensembles  $C_0, \dots, C_{d-1}$  sont visités de manière cyclique, ce qui donne (3.5).

Comme la chaîne de Markov  $(X_n)_{n \in \mathbb{N}}$  est irréductible, les chaînes de Markov  $(X_{nd})_{n \in \mathbb{N}}$  restreintes aux ensembles  $C_0, \dots, C_{d-1}$  sont irréductibles. D'après le Théorème 3.4, chacun a une unique loi stationnaire, respectivement  $\pi^{(0)}, \dots, \pi^{(d-1)}$ . Comme  $\pi^{(0)}$  a pour support  $C_0$  et vérifie  $\pi^{(0)} = \pi^{(0)}P^d$ ,  $\pi^{(0)}P$  a pour support  $C_1$  et vérifie  $(\pi^{(0)}P) = (\pi^{(0)}P)P^d$ , de sorte que  $\pi^{(1)} = \pi^{(0)}P$ , par unicité de la mesure stationnaire. On montre de la même manière que  $\pi^{(2)} = \pi^{(1)}P, \dots, \pi^{(d)} = \pi^{(d-1)}P$  et  $\pi^{(0)} = \pi^{(d)}P$ . Pour finir, on vérifie que la mesure de probabilité  $\pi$  satisfait bien  $\pi = \pi P$ .  $\square$

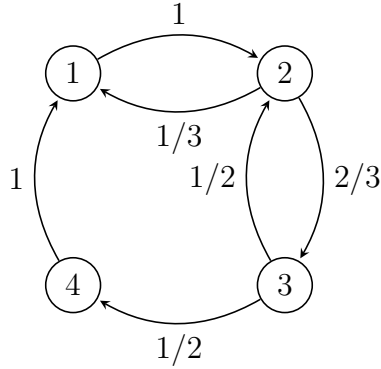


Figure 3.4. – Graphe de transition d'une chaîne de Markov irréductible de période 2.

EXEMPLE 13. – La chaîne de Markov dont le graphe de transition est représenté en figure 3.4 est irréductible de période 2. Elle visite les états  $\{1, 3\}$  et  $\{2, 4\}$  de manière cyclique. Sa matrice de transition est :

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

On vérifie que la matrice

$$P^2 = \begin{pmatrix} 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 2/3 & 0 & 1/3 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

définit bien deux chaînes de Markov irréductibles, restreintes à  $\{1, 3\}$  et  $\{2, 4\}$ , respectivement. Les lois stationnaires associées sont  $\pi^{(0)} = (1/2, 0, 1/2, 0)$  et  $\pi^{(1)} = (0, 3/4, 0, 1/4)$ . On vérifie que  $\pi^{(1)} = \pi^{(0)}P$  et  $\pi^{(0)} = \pi^{(1)}P$ . De plus, la chaîne de Markov initiale a bien pour unique loi stationnaire  $\pi = (\pi^{(0)} + \pi^{(1)})/2$ , soit  $\pi = (1/4, 3/8, 1/4, 1/8)$ .

Une chaîne de Markov irréductible est dite apériodique si sa période est 1. Le résultat suivant en donne une caractérisation à partir de la matrice de transition.

**Théorème 3.8.** Une chaîne de Markov irréductible est apériodique si et seulement si la matrice  $P^n$  est strictement positive pour tout  $n$  suffisamment grand.



*Démonstration.* La condition suffisante est évidente, le pgcd des nombres supérieurs à un entier quelconque étant égal à 1. Montrons la condition nécessaire. Il suffit de montrer le résultat pour les éléments diagonaux de  $P^n$ . En effet, pour deux états distincts  $i, j$ , il existe un chemin de longueur  $m$  de  $i$  à  $j$  par irréductibilité, si bien que

$$(P^{n+m})_{ij} \geq (P^n)_{ii}(P^m)_{ij} > 0$$

pour  $n$  suffisamment grand.

Soit  $C$  l'ensemble des longueurs des cycles passant par un état  $i$  fixé. Notons que  $C$  est stable par addition : pour tous  $n, m \in C$ ,

$$(P^{n+m})_{ii} \geq (P^n)_{ii}(P^m)_{ii} > 0,$$

de sorte que  $n + m \in C$ . Par ailleurs, la chaîne étant apériodique, il existe  $k$  éléments de  $C$  de pgcd égal à 1. Par l'identité de Bezout, il existe une combinaison linéaire à coefficients entiers de ces  $k$  éléments égale à 1. En regroupant les coefficients positifs et les coefficients négatifs, on obtient  $a - b = 1$  où  $a$  et  $b$  sont des entiers naturels, éléments de  $C$ . Si  $b = 0$ , alors  $a = 1$  est un élément de  $C$  et  $C = \mathbf{N} \setminus \{0\}$ . Si  $b > 0$ , soit  $n = bq + r$  la division euclidienne d'un entier  $n$  quelconque par  $b$ . En prenant  $n \geq b^2$ , on obtient  $q \geq b > r$  et  $n = bq + r(a - b) = (q - r)b + ra$ , élément de  $C$  puisque  $a, b \in C$ . On en déduit que  $(P^n)_{ii} > 0$  pour tout  $n \geq b^2$ .  $\square$

Le résultat clé suivant montre que, pour une chaîne de Markov irréductible et apériodique, la loi stationnaire est également la loi limite de la chaîne de Markov :

**Théorème 3.9.** Soit  $\pi$  la loi stationnaire d'une chaîne de Markov irréductible et apériodique. Alors pour toute loi initiale  $\pi(0)$ ,

$$\lim_{n \rightarrow \infty} \pi(n) = \pi.$$

*Démonstration.* Soit  $P^\infty$  la matrice définie par :

$$P^\infty = \begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_n \\ \pi_1 & \pi_2 & \dots & \pi_n \\ \vdots & & & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix}.$$

Cette matrice est strictement positive d'après le Corollaire 3.5. Nous allons montrer que

$$\lim_{n \rightarrow \infty} P^n = P^\infty. \quad (3.7)$$

Cela suffit pour conclure puisqu'alors, d'après (3.3),

$$\forall j = 1, \dots, N, \quad \lim_{n \rightarrow \infty} \pi(n)_j = \sum_i \pi(0)_i P_{ij}^\infty = \sum_i \pi(0)_i \pi_j = \pi_j.$$

On remarque tout d'abord que, comme  $P^\infty = (1, \dots, 1)^T \pi$ ,

$$P^\infty P = P P^\infty = P^\infty \quad (3.8)$$

et

$$(P^\infty)^2 = P^\infty. \quad (3.9)$$

Par ailleurs, on sait par le théorème 3.8 qu'il existe un entier  $k \geq 1$  tel que la matrice  $P^k$  est strictement positive. Soit :

$$\eta = \min_{i,j=1,\dots,N} \frac{(P^k)_{ij}}{P_{ij}^\infty}.$$

Cette constante est strictement positive. De plus, les matrices  $P^k$  et  $P^\infty$  étant stochastiques, on a  $\eta \leq 1$  avec égalité si et seulement si  $P^k = P^\infty$ . Dans ce dernier cas, la propriété (3.7) découle de (3.8). On suppose donc par la suite que  $\eta < 1$ .

Soit  $Q$  la matrice :

$$Q = \frac{1}{1-\eta} (P^k - \eta P^\infty).$$

C'est une matrice stochastique. D'après (3.8)-(3.9),

$$P^\infty Q = Q P^\infty = P^\infty,$$

de sorte que pour tout entier  $n \geq 1$ ,  $(Q - P^\infty)^n = Q^n - P^\infty$  (par récurrence). De la même manière, on a  $P^{kn} - P^\infty = (P^k - P^\infty)^n$  pour tout entier  $n \geq 1$ . Or  $P^k - P^\infty = (1-\eta)(Q - P^\infty)$  par définition de la matrice  $Q$ . On en déduit que

$$P^{kn} - P^\infty = (1-\eta)^n (Q^n - P^\infty).$$

Les matrices  $Q^n$  et  $P^\infty$  étant stochastiques, on a

$$\|P^{kn} - P^\infty\|_\infty \leq (1-\eta)^n,$$

où  $\|\cdot\|_\infty$  désigne le maximum des entrées d'une matrice en valeur absolue. On en déduit que

$$\lim_{n \rightarrow \infty} P^{kn} = P^\infty.$$

Pour conclure, on remarque que pour tout  $\ell = 0, 1, \dots, k-1$ ,

$$P^{kn+\ell} - P^\infty = P^\ell (P^{kn} - P^\infty),$$

de sorte que

$$\lim_{n \rightarrow \infty} P^{kn+\ell} = P^\infty.$$

□

REMARQUE.— D'après la preuve du théorème 3.9, la convergence de la loi d'une chaîne de Markov irréductible et apériodique vers sa loi stationnaire se fait à une vitesse exponentielle.

La loi d'une chaîne de Markov irréductible de période  $d \geq 2$  ne converge pas en général vers la loi stationnaire. Cela dépend de la loi initiale. Si  $\pi(0)$  a pour support l'un des  $d$  ensembles cycliques par exemple, la loi  $\pi(n)$  aura un comportement périodique (voir théorème 3.7). Une condition nécessaire et suffisante pour que  $\pi(n)$  converge vers  $\pi$  est que la mesure de chaque ensemble cyclique soit la même, égale à  $1/d$ , sous la loi initiale  $\pi(0)$ .

EXEMPLE 14. – Pour la chaîne de Markov de l'Exemple 13, la loi stationnaire est la loi limite si  $\pi(0) = (1/2, 1/2, 0, 0)$  par exemple.

## 3.7. Théorème ergodique

On connaît la loi stationnaire mais pas le comportement de la chaîne de Markov pour une trajectoire donnée, soit  $X_0, X_1, \dots$  pour un  $\omega \in \Omega$  fixé. On aimerait connaître par exemple la fraction du temps que la chaîne de Markov passe dans l'état  $i$ . Comme la probabilité que  $X_n = i$  tend vers  $\pi_i$  dès que la chaîne est apériodique, on s'attend à trouver pour toute loi initiale  $\pi(0)$  :

$$\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k=i\}} \xrightarrow{\text{p.s.}} \pi_i \quad \text{lorsque } n \rightarrow \infty. \quad (3.10)$$

Plus généralement, pour toute fonction  $\varphi$  sur  $\{1, \dots, N\}$ , on doit avoir

$$\frac{1}{n} \sum_{k=1}^n \varphi(X_k) \xrightarrow{\text{p.s.}} \sum_i \pi_i \varphi(i) \quad \text{lorsque } n \rightarrow \infty, \quad (3.11)$$

la forme précédente s'obtenant en prenant pour  $\varphi$  la fonction indicatrice  $\varphi(x) = \mathbf{1}_{\{x=i\}}$ . C'est le théorème ergodique. Ce résultat généralise la loi forte des grands nombres, qui n'est valable que pour des variables aléatoires i.i.d. Lorsque  $X_1, X_2, \dots$  sont des variables aléatoires indépendantes de même loi  $\pi$  sur  $\{1, \dots, N\}$  (ce qui définit en particulier une chaîne de Markov), les variables aléatoires  $\varphi(X_1), \varphi(X_2), \dots$  sont en effet également i.i.d. de sorte que

$$\frac{1}{n} \sum_{k=1}^n \varphi(X_k) \xrightarrow{\text{p.s.}} \mathbf{E}(\varphi(X_1)) = \sum_i \pi_i \varphi(i) \quad \text{lorsque } n \rightarrow \infty.$$

la dernière égalité venant du théorème de transfert. Ce résultat s'étend donc à des variables aléatoires non indépendantes, ne suivant pas la même loi, mais vérifiant la propriété de Markov. La loi  $\pi$  intervenant dans la limite est alors la loi stationnaire de la chaîne de Markov, supposée irréductible.

Pour prouver le théorème ergodique, nous commençons par donner une forme de la loi stationnaire faisant intervenir le *temps de retour* dans chaque état. D'après (3.10), la loi stationnaire donne les fréquences asymptotiques des passages dans chaque état. Cela suggère que la probabilité d'être dans un état sous la loi stationnaire est l'inverse du temps moyen de retour à cet état. Pour tout  $i = 1, \dots, N$ , on note donc  $T_i$  le temps d'atteinte de l'état  $i$  :

$$T_i = \min\{n \geq 1 : X_n = i\}.$$

C'est une variable aléatoire sur  $\mathbf{N} \setminus \{0\} \cup \{+\infty\}$ . Le temps de retour est une variable aléatoire de même loi que  $T_i$  conditionnellement à  $X_0 = i$  (on supposera par la suite que  $\pi(0)$  a pour support  $\{1, \dots, N\}$  pour simplifier l'exposé mais le résultat est valable quelle que soit la loi initiale  $\pi(0)$ ). On note  $\mathbf{P}_i$  la mesure de probabilité conditionnelle à  $X_0 = i$  et  $\mathbf{E}_i$  l'espérance associée.

**Théorème 3.10.** La loi stationnaire d'une chaîne de Markov irréductible est donnée par :

$$\forall i = 1, \dots, N, \quad \pi(i) = \frac{1}{\mathbf{E}_i(T_i)}.$$

*Démonstration.* On montre le résultat pour un état fixé, par exemple l'état 1. On note pour tout état  $j = 1, \dots, N$  et tout  $n \geq 1$  :

$$q_j(n) = \mathbf{P}_1(T_1 \geq n, X_n = j).$$

C'est la probabilité que la chaîne de Markov soit dans l'état  $j$  au temps  $n$  sans avoir été dans l'état 1 depuis le temps 0. On a :

$$q_j(1) = P_{1j}, \quad q_j(n) = \sum_{i \neq 1} q_i(n-1)P_{ij} \quad \text{pour } n \geq 2.$$

En notant

$$v_j = \sum_{n=1}^{\infty} q_j(n),$$

on obtient par sommation des équations précédentes :

$$v_j = P_{1j} + \sum_{i \neq 1} v_i P_{ij}. \quad (3.12)$$

Par ailleurs, en remarquant que  $\mathbf{P}_1(T_1 \geq n, X_n = 1) = \mathbf{P}_1(T_1 = n)$ , on obtient

$$v_1 = \sum_{n=1}^{\infty} \mathbf{P}_1(T_1 = n) = \mathbf{P}_1(T_1 < \infty),$$

de sorte que  $v_1 \leq 1$ . On déduit alors de (3.12) que :

$$v_j \geq \sum_{i=1}^N v_i P_{ij}.$$

soit  $v \geq vP$  en notation vectorielle. En particulier,  $v \geq vP^n$  pour tout  $n \geq 1$ . Comme la chaîne est irréductible, il existe un entier  $n \geq 1$  tel que  $(I+P)^n$  est une matrice strictement positive d'après la Proposition 3.3. Comme  $v2^n \geq v(I+P)^n$ , cela implique que si le vecteur  $v$  a une composante infinie, toutes ses composantes sont infinies. Comme  $v_1 \leq 1$ , toutes les composantes du vecteur  $v$  sont finies. En sommant les égalités (3.12), on obtient :

$$\sum_{j=1}^N v_j = 1 + \sum_{j \neq 1} v_j,$$

si bien que  $v_1 = 1$ . Ainsi  $v = vP$  et on obtient la loi stationnaire de la chaîne de Markov par normalisation de  $v$  :

$$\forall i = 1, \dots, N, \quad \pi_i = \frac{v_i}{\sum_j v_j}.$$

Pour finir, on remarque que :

$$\sum_j v_j = \sum_j \sum_{n=1}^{\infty} q_j(n) = \sum_{n=1}^{\infty} \sum_j q_j(n) = \sum_{n=1}^{\infty} \mathbf{P}_1(T_1 \geq n) = \mathbf{E}_1(T_1),$$

de sorte que

$$\pi_1 = \frac{1}{\mathbf{E}_1(T_1)}.$$

□

REMARQUE.— Le résultat précédent donne une preuve probabiliste de l'existence de la loi stationnaire, à comparer à la preuve algébrique du Théorème 3.4.

Le Théorème 3.10 et le Corollaire 3.5 donnent le résultat suivant :

**Corollaire 3.11.** Pour toute chaîne de Markov irréductible,  $\mathbf{E}_i(T_i) < \infty$  pour tout état  $i = 1, \dots, N$ .

On est maintenant en mesure de prouver le théorème ergodique :

**Théorème 3.12.** Soit  $\pi$  la loi stationnaire d'une chaîne de Markov irréductible. Pour toute fonction  $\varphi$  sur  $\{1, \dots, N\}$ , on a :

$$\frac{1}{n} \sum_{k=1}^n \varphi(X_k) \xrightarrow{\text{p.s.}} \sum_i \pi_i \varphi(i) \quad \text{lorsque } n \rightarrow \infty.$$

*Démonstration.* On commence par prouver le résultat lorsque  $\varphi$  est une fonction indicatrice, soit  $\varphi(x) = \mathbf{1}_{\{x=i\}}$ . On note pour tout  $n \geq 1$  :

$$S_n = \sum_{k=1}^n \mathbf{1}_{\{X_k=i\}}.$$

Soit  $U_1, U_2, \dots$  la suite des temps successifs d'atteinte de l'état  $i$ . Conditionnellement au fait que  $X_0 = i$ , c'est une suite de variables aléatoires i.i.d.<sup>1</sup> de même loi que  $T_i$ . Comme  $\mathbf{E}_i(T_i) < \infty$  d'après le Corollaire 3.11, on a par la loi forte des grands nombres :

$$\frac{1}{n} \sum_{k=1}^n U_k \rightarrow \mathbf{E}_i(T_i) \quad \mathbf{P}_i - \text{p.s.} \quad \text{lorsque } n \rightarrow \infty. \quad (3.13)$$

1. Ce résultat fait appel à la propriété dite forte de Markov : après avoir atteint l'état  $i$ , la chaîne de Markov se « régénère » et évolue comme si elle était partie initialement de l'état  $i$ .

Par ailleurs, comme  $S_n$  est égal au nombre de visites de l'état  $i$  entre les temps 1 et  $n$ , on a :

$$\sum_{k=1}^{S_n} U_k \leq n < \sum_{k=1}^{S_n+1} U_k. \quad (3.14)$$

La suite de variables aléatoires  $S_1, S_2, \dots$  tend  $\mathbf{P}_i$ -p.s. vers  $+\infty$ . En effet, il existerait dans le cas contraire un entier  $m \geq 1$  tel que  $S_n < m$  pour tout  $n$  avec probabilité strictement positive ; cela impliquerait que  $\mathbf{P}_i(\sum_{k=1}^m U_k = +\infty) > 0$  d'après l'inégalité précédente, et comme les variables aléatoires  $U_1, U_2, \dots$  sont indépendantes et de même loi que  $T_i$  sous  $\mathbf{P}_i$ , que  $\mathbf{P}_i(T_i = +\infty) > 0$ , une contradiction. D'après (3.13), on a donc

$$\frac{1}{S_n} \sum_{k=1}^{S_n} U_k \rightarrow \mathbf{E}_i(T_i) \quad \mathbf{P}_i - \text{p.s.} \quad \text{lorsque } n \rightarrow \infty.$$

Par les inégalités (3.14), on conclut que :

$$\frac{S_n}{n} \rightarrow \frac{1}{\mathbf{E}_i(T_i)} \quad \mathbf{P}_i - \text{p.s.} \quad \text{lorsque } n \rightarrow \infty. \quad (3.15)$$

Le résultat découle alors du Théorème 3.10.

Il reste à montrer le résultat sous une loi  $\mathbf{P}_j$ , avec  $j \neq i$ . Alors  $U_1$  est le temps d'atteinte de l'état  $i$  depuis l'état  $j$  tandis que  $U_2, U_3, \dots$  sont les temps successifs de retour à l'état  $i$ . On a à nouveau par la loi forte des grands nombres :

$$\frac{1}{n} \sum_{k=2}^n U_k \rightarrow \mathbf{E}_i(T_i) \quad \mathbf{P}_j - \text{p.s.} \quad \text{lorsque } n \rightarrow \infty.$$

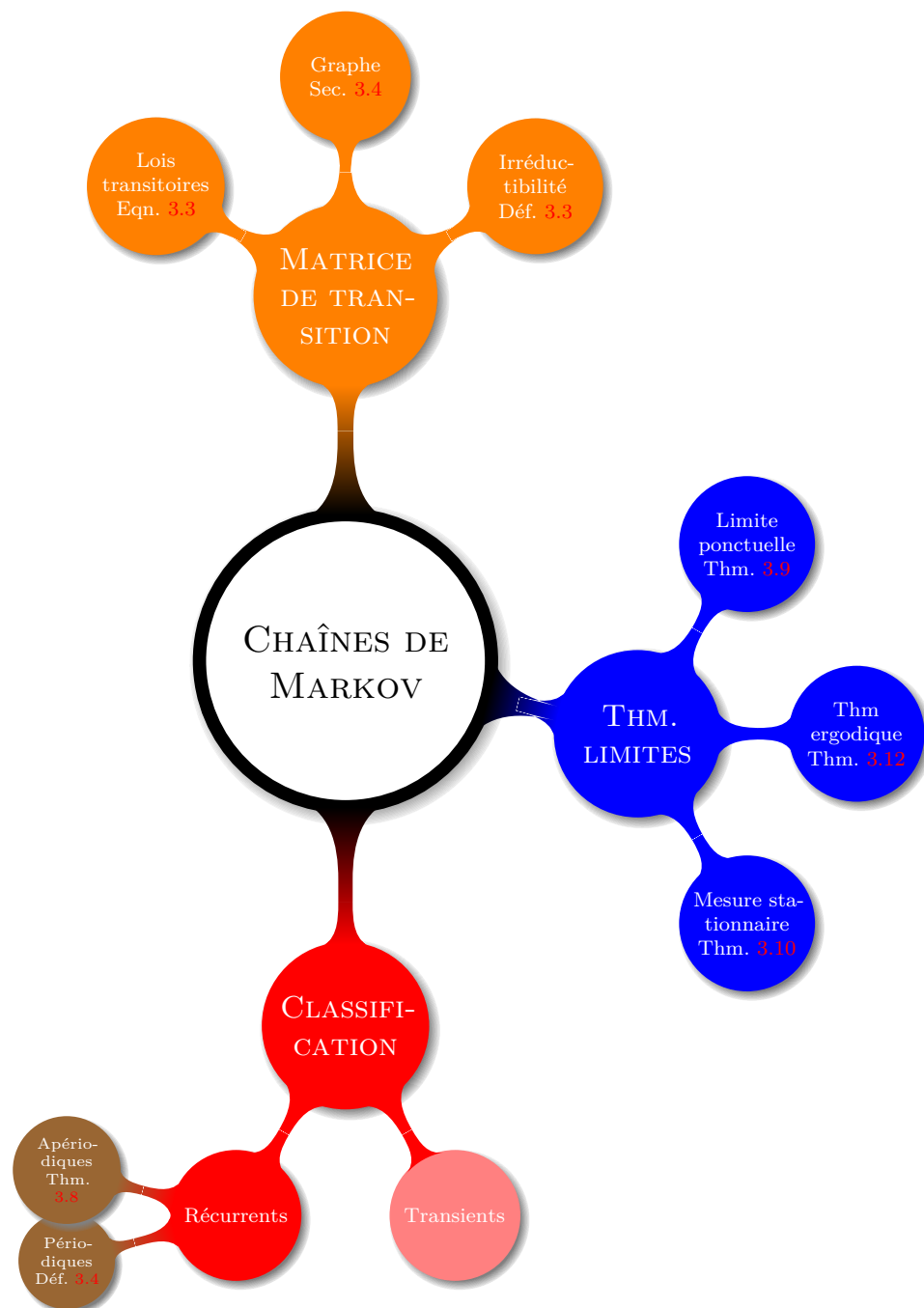
Il reste à montrer que la variable aléatoire  $U_1$  est  $\mathbf{P}_j$ -p.s. finie pour en déduire (3.13) et donc (3.15) pour la loi  $\mathbf{P}_j$ . Supposons donc que  $\mathbf{P}_j(T_i = +\infty) > 0$ . Comme la chaîne est irréductible, il existe un entier  $n \geq 1$  tel que la chaîne passe de l'état  $i$  à l'état  $j$  en  $n$  étapes avec probabilité  $p > 0$ . On a alors  $\mathbf{P}_i(T_i = +\infty) \geq p \times \mathbf{P}_j(T_i = +\infty) > 0$ , une contradiction.

Finalement, le résultat s'étend à toute fonction  $\varphi$  en écrivant

$$\frac{1}{n} \sum_{k=1}^n \varphi(X_k) = \sum_i \varphi(i) \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k=i\}} \xrightarrow{\text{p.s.}} \sum_i \pi_i \varphi(i) \quad \text{lorsque } n \rightarrow \infty.$$

□

REMARQUE.— La convergence du théorème ergodique est presque sûre et donc indépendante de la loi initiale de la chaîne de Markov.



## 3.8. Exercices

### Pour apprendre

EXERCICE 3.1.—

Soit  $(X_n)_{n \in \mathbb{N}}$  une chaîne de Markov de matrice de transition  $P$ . Montrer que  $(X_{2n})_{n \in \mathbb{N}}$  est une chaîne de Markov de matrice de transition  $P^2$ .

EXERCICE 3.2.—

Soit  $(X_n)_{n \in \mathbb{N}}$  une chaîne de Markov de matrice de transition :

$$P = \begin{pmatrix} 1/4 & 1/2 & 1/4 \\ 1/3 & 0 & 2/3 \\ 0 & 1/5 & 4/5 \end{pmatrix},$$

avec  $X_0 = 1$  p.s. Calculer les probabilités des événements  $\{X_1 = 2\}$ ,  $\{X_2 = 1\}$ ,  $\{X_1 = 2, X_2 = 1\}$ .

EXERCICE 3.3.—

Soit  $(X_n)_{n \in \mathbb{N}}$  une chaîne de Markov de matrice de transition :

$$P = \begin{pmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{pmatrix},$$

et de loi initiale  $\pi(0) = (1/2, 1/2)$ . Calculer les lois  $\pi(1)$  et  $\pi(2)$ . Quelle est la loi stationnaire ?

EXERCICE 3.4.—

Soit  $X$  une chaîne de Markov d'espace d'états  $E = \{a, b, c\}$ , de matrice de transition

$$P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{2}{5} & 0 & \frac{3}{5} \end{pmatrix},$$

de loi initiale

$$\pi(0) = \left( \frac{2}{5}, \frac{1}{5}, \frac{2}{5} \right).$$

1. Calculer

$$\mathbf{P}_a(X_1 = b, X_2 = b, X_3 = b, X_4 = a, X_5 = c).$$

2. Calculer

$$\mathbf{P}_a(X_1 = b, X_3 = a, X_4 = c, X_6 = b).$$

3. Calculer

$$\mathbf{P}(X_1 = b, X_2 = b, X_3 = a).$$

4. Calculer

$$\mathbf{P}(X_2 = b, X_5 = b, X_6 = b).$$



## EXERCICE 3.5. –

Soit  $X$  une chaîne de Markov d'espace d'états  $E = \{r, w, b, y\}$  et de matrice de transition

$$P = \begin{pmatrix} 0 & 0 & 10 & \\ 0 & 0,4 & 0,6 & 0 \\ 0,8 & 0 & 0,2 & 0 \\ 0,2 & 0,3 & 0 & 0,5 \end{pmatrix}.$$

1. Calculer

$$\mathbf{P}(X_5 = b, X_6 = r, X_7 = b, X_8 = b \mid X_4 = w).$$

2. Calculer

$$\mathbf{E}[f(X_5)f(X_6) \mid X_4 = y]$$

où

$$f(r) = 2, f(w) = 4, f(b) = 7, f(y) = 3.$$

**Pour s'entraîner**

## EXERCICE 3.6 (Canal binaire symétrique). –

On considère un canal de communication qui transmet des bits avec erreur selon le modèle suivant : un bit à une probabilité  $1 - p$  d'être transmis correctement et  $p$  d'être inversé. On suppose que  $n$  canaux de ce type sont en série. On note  $X_n$  le bit reçu en sortie du  $n$ -canal. On note

$$\pi(n) = (\mathbf{P}(X_n = 0), \mathbf{P}(X_n = 1)).$$

1. Exprimer la relation matricielle entre  $\pi(n)$  et  $\pi(n-1)$  pour tout  $n \geq 1$ . On traitera à part les cas  $p = 0$  et  $p = 1$ .
2. Calculer la probabilité  $p_n$  que l'information soit correctement transmise au rang  $n$ .
3. Que se passe-t-il quand  $n$  tend vers l'infini ?

## EXERCICE 3.7 (Urnes d'Erhenfest). –

On considère deux urnes contenant au total  $m$  boules. A chaque étape, on choisit une boule au hasard et on change cette boule d'urne. On note  $X_n$  le nombre de boules dans l'urne 1 au bout de  $n$  mouvements.

1. Montrer que  $X$  est une chaîne de Markov dont on donnera la matrice de transition.
2. Est-ce que cette chaîne est irréductible, apériodique ?
3. Quelle est la matrice de transition de la chaîne  $Z_n = X_{2n}$  ?

On change maintenant de protocole : on choisit au hasard une boule et une urne au hasard, on met la boule choisie dans l'urne choisie. On note toujours  $X_n$  le nombre de boules dans l'urne 1 au bout de  $n$  mouvements.

4. Calculer la matrice de transition de cette nouvelle chaîne.
5. Est-elle irréductible ? apériodique ?
6. Montrer que sa probabilité invariante est une loi binomiale dont on précisera les paramètres.

**EXERCICE 3.8** (Urnes de Benoulli-Laplace).–

Un modèle simple de diffusion de deux gaz est le suivant. On considère deux urnes 1 et 2 qui contiennent respectivement  $r$  boules blanches et  $r$  boules noires. À chaque étape, on choisit une boule dans chacune des urnes et on permute la position de ces deux boules : celle qui était en 1 passe en 2 et réciproquement. On note  $X_n$  le nombre de boules blanches dans 1 après le  $n$ -mélange.

- 3.8.a) Pour  $n \geq 1$ , calculer la loi de  $X_n$  sachant  $X_{n-1}$ .
- 3.8.b) Calculer  $\mathbf{E}[X_n | X_{n-1}]$ .
- 3.8.c) En déduire une expression et la limite de  $\mathbf{E}[X_n]$  quand  $n$  tend vers l'infini ?
- 3.8.d) Montrer que la loi stationnaire de cette chaîne de Markov est la loi hypergéométrique de paramètres  $(r, r)$ .

**Pour aller plus loin****EXERCICE 3.9** (Labyrinthe).–

Un rat se déplace dans le labyrinthe à sept cases représenté dans la figure 3.5. Il passe d'une case à l'autre uniformément suivant les possibilités qui lui sont offertes, c'est-à-dire que lorsqu'il y a 2 (respectivement 3) sorties dans la case où il se trouve, il va dans chacune des cases possibles avec une probabilité d'un demi (respectivement d'un tiers). Son évolution est sans mémoire : chaque changement ne dépend que de la situation courante, pas du passé. On appelle  $X_n$  la position du rat après son  $n$ -ième mouvement,  $X_0$  est sa position initiale.

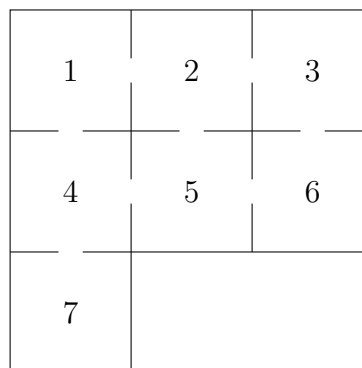


Figure 3.5. – Le labyrinthe.

- 3.9.a) Construire la matrice de transition de cette chaîne de Markov.
- 3.9.b) Ecrire le code Python qui permet de calculer la probabilité stationnaire.
- 3.9.c) Calculer (en Python)  $P^2$ ,  $P^3$ ,  $P^4$ . Que remarquez-vous ? Interpréter.
- 3.9.d) La chaîne de Markov de matrice de transition  $P^2$  est-elle irréductible ?

**EXERCICE 3.10.**–

Un trait (génétique) est déterminé par une paire de gènes, appartenant tous deux à  $G$

ou  $g$ . Un individu peut avoir la paire GG (dominant),  $gG$  ou de manière équivalente Gg (hybride), ou gg (récessif).

Lors de la reproduction sexuée, un descendant hérite au hasard d'un gène de chacun des parents.

1. Pour les 9 situations possibles de parents, calculer la probabilité de chaque configuration pour un descendant. On présentera les résultats sous la forme suivante :

	GG			Gg			gg		
GG	P(GG)	P(Gg)	P(gg)	P(GG)	P(Gg)	P(gg)	P(GG)	P(Gg)	P(gg)
Gg	$\vdots$			$\vdots$			$\vdots$		
gg	$\vdots$			$\vdots$			$\vdots$		

*On expliquera le raisonnement sur un cas et on se contentera de donner les résultats pour les autres cas.*

Un éleveur prend un individu quelconque et le fait se reproduire avec un individu hybride. A la génération suivante, il répète le même processus. On note  $X_n$  le type génétique de l'individu de génération  $n$ .

2. Quelle est la matrice de transition de la chaîne de Markov  $X = (X_n, n \geq 0)$  ?
3. Est-elle irréductible, récurrente ?
4. Calculer sa loi stationnaire.

Pour simplifier les notations, on pose GG = 1, Gg = 2, gg = 3. On pose  $\tau = \inf\{n > 0, X_n = 1\}$  et

$$u_i = \mathbf{E}[\tau \mid X_0 = i].$$

On admet que (grâce au théorème de Fubini) pour toute variable aléatoire  $Z$  discrète positive :

$$\mathbf{E}[Z] = \sum_{k=1}^{\infty} \mathbf{P}(Z \geq k)$$

donc en particulier,

$$u_i = \sum_{n=1}^{\infty} \mathbf{P}(\tau \geq n \mid X_0 = i)$$

5. Ecrire l'événement  $(\tau \geq n)$  en fonction de  $X_1, \dots, X_n$ .
6. Montrer que pour tout  $i = 1, 2, 3$ ,

$$\begin{aligned} \mathbf{P}(\tau \geq n \mid X_0 = i) &= \mathbf{P}(\tau + 1 \geq n \mid X_0 = 2) \mathbf{P}(X_1 = 2 \mid X_0 = i) \\ &\quad + \mathbf{P}(\tau + 1 \geq n \mid X_0 = 3) \mathbf{P}(X_1 = 3 \mid X_0 = i). \end{aligned}$$

7. En déduire le système linéaire dont est solution  $(u_1, u_2, u_3)$ .

8. Le résoudre (pas besoin de détailler les calculs) et donner le nombre moyen de générations nécessaire avant d'obtenir un individu de type dominant si on ne sait pas a priori quel était le type de l'individu initial.

## 4. Mesures et intégration

*Ce chapitre a un statut particulier. D'un côté, il introduit les concepts et les notations au cœur de ce cours et de MDI 103. De l'autre, son contenu en tant que tel n'est pas évalué dans le contrôle final. En bref, il n'y aura pas de questions sur la théorie de la mesure mais toutes les notations et définitions sont d'un usage constant.*

### 4.1. Introduction

Ce chapitre a pour but l'introduction des outils nécessaires à la construction de probabilités sur des ensembles plus généraux que les seuls ensembles discrets. Une probabilité sur l'univers  $\Omega$  est une application  $\mathbf{P}$  qui à un événement  $A$  associe une valeur  $\mathbf{P}(A)$  comprise entre 0 et 1. Formellement,

$$\begin{aligned}\mathbf{P} : \mathcal{F} &\rightarrow [0, 1] \\ A &\mapsto \mathbf{P}(A)\end{aligned}$$

où  $\mathcal{F}$  est le domaine de définition de  $\mathbf{P}$ . Lorsque l'univers  $\Omega$  est au plus dénombrable, nous pouvons choisir  $\mathcal{F}$  comme l'ensemble des parties. En revanche, dans des espaces plus complexes tels que  $\Omega = \mathbf{R}$ , sauf à renoncer à l'axiome du choix [Oxt80], il est impossible de définir  $\mathbf{P}$  sur l'ensemble des parties. Il nous faut donc restreindre le domaine de définition de  $\mathbf{P}$ .

Les conditions qui pèsent sur  $\mathcal{F}$  sont liées aux considérations suivantes :

- a) On veut pouvoir définir les probabilités de l'événement impossible (zéro) et de l'événement certain (un). Donc  $\mathcal{F}$  doit contenir  $\emptyset$  et  $\Omega$ .
- b) Si on sait évaluer la probabilité qu'un événement  $A$  se réalise, on doit logiquement pouvoir parler de la probabilité qu'il ne se réalise pas. Autrement dit, si  $\mathcal{F}$  est stable par passage au complémentaire.
- c) Si  $A$  et  $B$  sont des événements dont on sait évaluer les probabilités, on doit pouvoir donner un sens à  $\mathbf{P}(A \cup B)$ , donc  $\mathcal{F}$  est stable pour l'union. Pour des raisons techniques qui apparaîtront clairement plus loin, nous supposons en outre que  $\mathcal{F}$  est stable par union dénombrable.

Les axiomes ci-dessus sont ceux qui définissent une *tribu*. Le paragraphe 4.2 sera consacré à des rappels sur les tribus. Le paragraphe 4.4 est consacré à la définition des mesures, qui

sont des applications sur  $\mathcal{F}$  à valeurs positives. Les mesures de probabilité en sont un cas particulier. Le paragraphe 4.3 introduit la notion d'applications mesurables qui donnera le cadre formel nécessaire à la définition des variables aléatoires à valeurs dans des espaces plus généraux que les espaces discrets.

## 4.2. Tribus

**Définition 4.1** (Tribu). Une famille  $\mathcal{F}$  de sous-ensembles de  $\Omega$  est appelée une *tribu* sur  $\Omega$  si elle vérifie les propriétés suivantes :

- 4.1.i)  $\Omega \in \mathcal{F}$  ;
- 4.1.ii)  $\forall A \in \mathcal{F}, A^c \in \mathcal{F}$  ;
- 4.1.iii)  $\forall A_1, A_2, \dots \in \mathcal{F}, \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ .

Autrement dit, une tribu est stable par passage au complémentaire et stable par union dénombrable.

Citons quelques exemples de tribus :

- la tribu grossière :  $\mathcal{F} = \{\emptyset, \Omega\}$  ;
- la tribu des parties :  $\mathcal{F} =$  l'ensemble des sous-ensembles de  $\Omega$ , noté  $\mathcal{P}(\Omega)$  ou  $2^\Omega$  ;

Un *espace mesurable* est un couple  $(\Omega, \mathcal{F})$  où  $\Omega$  est un ensemble et  $\mathcal{F}$  est une tribu sur  $\Omega$ . On parle parfois d'*espace probabilisable*.

**Proposition 4.1.** Toute tribu satisfait les propriétés suivantes.

- a)  $\emptyset \in \mathcal{F}$  ;
- b)  $\forall A_1, A_2, \dots \in \mathcal{F}, \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$  ;
- c)  $\forall A, B \in \mathcal{F}, A \cup B \in \mathcal{F}$  et  $A \cap B \in \mathcal{F}$ .

*Démonstration.* — a)  $\Omega \in \mathcal{F}$  donc  $\Omega^c = \emptyset \in \mathcal{F}$  par l'axiome ii).

- b) Les complémentaires  $A_1^c, A_2^c, \dots$  sont dans  $\mathcal{F}$  par l'axiome ii). Donc  $\bigcup_{i=1}^{\infty} A_i^c \in \mathcal{F}$  par l'axiome 4.1.iii). En invoquant à nouveau l'axiome 4.1.ii), le complémentaire de cet ensemble est également dans  $\mathcal{F}$  par l'axiome 4.1.ii). Or d'après les lois de De Morgan, le complémentaire coïncide avec  $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$
- c) Il suffit de poser  $A_1 = A, A_2 = B$  et  $A_i = \emptyset$  pour tout  $i \geq 3$ . Les deux résultats découlent de l'axiome 4.1.iii) et de b) respectivement.

□

En appliquant les axiomes définissant une tribu, il n'est pas difficile de montrer le résultat suivant.

**Lemme 4.2.** L'intersection de deux tribus est une tribu.

On peut alors envisager la notion fondamentale de *tribu engendrée*.

**Théorème 4.3** (Tribu engendrée). Soit  $\mathcal{C}$  une collection d'ensembles sur  $\Omega$ . L'intersection de toutes les tribus sur  $\Omega$  contenant  $\mathcal{C}$  est une tribu sur  $\Omega$ . On la note  $\sigma(\mathcal{C})$  et on l'appelle la tribu engendrée par  $\mathcal{C}$  sur  $\Omega$ .

*Démonstration.* Soit  $S$  l'ensemble des tribus contenant  $\mathcal{C}$ .  $S$  est non-vide puisqu'il contient la tribu des parties. On vérifie que  $\sigma(\mathcal{C}) = \bigcap_{\tau \in S} \tau$  vérifie les trois axiomes d'une tribu.

- *a)* Pour tout  $\tau \in S$ ,  $\emptyset \in \tau$  puisque  $\tau$  est une tribu. Donc  $\emptyset \in \bigcap_{\tau \in S} \tau$ .
- *b)* Soit  $A \in \sigma(\mathcal{C})$ . Pour tout  $\tau \in S$ , on a  $A \in \tau$  par définition de  $\sigma(\mathcal{C})$ . Donc  $A^c \in \tau$  car une tribu est stable par passage au complémentaire. Ainsi  $A^c \in \bigcap_{\tau \in S} \tau$ .
- *c)* Soient  $A_1, A_2, \dots \in \sigma(\mathcal{C})$ . Pour tout  $\tau \in S$ , on a  $A_1, A_2, \dots \in \tau$ . Donc, l'union des  $A_i$  est dans  $\tau$  quelque soit  $\tau \in S$ . Finalement,  $\bigcup_{i=1}^{\infty} A_i \in \bigcap_{\tau \in S} \tau$ .

□

EXEMPLE 15.— Soit  $A \subset \Omega$ , calculons  $\sigma(A)$ . Nécessairement,  $\emptyset$ ,  $A$  et  $\Omega$  appartiennent à  $\sigma(A)$ . D'après l'axiome *b)*,  $A^c \in \sigma(A)$ . En envisageant alors tous les cas possibles, on se convainc aisément que  $\sigma(A) = \{\emptyset, \Omega, A, A^c\}$ .

EXEMPLE 16.— Rappelons nous qu'une tribu est la famille des ensembles que l'on peut « mesurer ». Si  $\Omega = \mathbf{N}$ , la moindre des choses est de pouvoir mesurer les singletons. Dans ce cas, comme toute partie de  $\Omega$  s'écrit la réunion des singletons qui la composent et que cette réunion contient un nombre au plus dénombrable d'éléments, il ressort de l'axiome *4.1.iii)* que toutes les parties de  $\Omega$  appartiennent à la tribu engendrée par les singletons.

EXEMPLE 17.— Le troisième exemple est le plus fondamental pour toutes les applications qui vont suivre. Il va nous permettre de préciser enfin ce que sont les parties de  $\mathbf{R}$  (ou de  $\mathbf{R}^d$ ) susceptible d'être mesurée sans renoncer à l'axiome du choix.

**Définition 4.2** (Tribu borélienne). La *tribu de Borel sur  $\mathbf{R}^d$* , notée  $\mathcal{B}(\mathbf{R}^d)$ , est la tribu engendrée par les « pavés » :

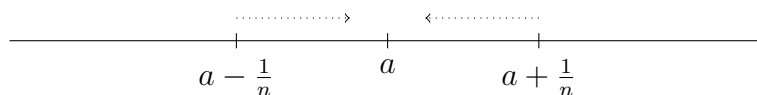
$$\mathcal{B}(\mathbf{R}^d) := \sigma \left( \left\{ \prod_{i=1}^d ]a_i, b_i[ : \forall i = 1, \dots, d, (a_i, b_i) \in \mathbf{R}^2, a_i < b_i \right\} \right).$$

En particulier sur  $\mathbf{R}$ ,  $\mathcal{B}(\mathbf{R})$  est la plus petite tribu qui contient les intervalles de la forme  $]a, b[$ .

Un élément de la tribu de Borel est appelé un *borélien*.

**Théorème 4.4** (Éléments caractéristiques de  $\mathcal{B}(\mathbf{R})$ ). Les ensembles suivants sont dans  $\mathcal{B}(\mathbf{R})$  : le singleton  $\{a\}$ , les intervalles de la forme  $[a, b]$ ,  $] - \infty, b]$ ,  $[a, +\infty[$ , l'ensemble des rationnels, l'ensemble des irrationnels.

*Démonstration.* Le singleton  $\{a\}$  est un borélien car il s'écrit comme une intersection dénombrable de boréliens  $\{a\} = \bigcap_{n \geq 1} ]a - \frac{1}{n}, a + \frac{1}{n}[$ .



L'intervalle  $[a, b]$  est un borélien car il s'écrit comme l'union de trois boréliens  $[a, b] = \{a\} \cup ]a, b[ \cup \{b\}$ . L'intervalle  $] - \infty, b]$  est un borélien car il s'écrit comme une union dénombrable de boréliens  $] - \infty, b] = \bigcup_{n \geq 1} ]b - n, b]$ . La preuve est similaire pour  $[a, +\infty[$ . L'ensemble des rationnels  $\mathbf{Q}$  s'écrit comme l'union de ses singletons  $\mathbf{Q} = \bigcup_{x \in \mathbf{Q}} \{x\}$ . Puisque  $\mathbf{Q}$  est dénombrable et que les singletons sont des boréliens,  $\mathbf{Q}$  est un borélien en vertu de 4.1.iii). L'ensemble des irrationnels est le complémentaire de  $\mathbf{Q}$ , c'est donc aussi un borélien puisque  $\mathcal{B}(\mathbf{R})$  est stable par passage au complémentaire.  $\square$

**Proposition 4.5.**  $\mathcal{B}(\mathbf{R})$  est la tribu engendrée par les intervalles de la forme  $] - \infty, b]$ . Plus généralement,  $\mathcal{B}(\mathbf{R}^d)$  est engendrée par les ensembles de la forme  $\prod_{i=1}^d ] - \infty, b_i]$ .

*Démonstration.* D'après la propriété précédente, la classe  $\mathcal{C}$  des ensembles  $] - \infty, b]$  sont dans  $\mathcal{B}(\mathbf{R})$ , donc  $\sigma(\mathcal{C}) \subset \mathcal{B}(\mathbf{R})$  car  $\sigma(\mathcal{C})$  est la plus petite tribu contenant  $\mathcal{C}$ . Inversement, tout intervalle  $]a, b[$  s'écrit  $] - \infty, a]^c \cap (\bigcup_n ] - \infty, b - \frac{1}{n}]$ , donc  $]a, b[$  est inclus dans  $\sigma(\mathcal{C})$ . Donc  $\mathcal{B}(\mathbf{R}) \subset \sigma(\mathcal{C})$ .  $\square$

REMARQUE.— Dans ce cours, nous utiliserons assez fréquemment la tribu de Borel  $\mathcal{B}(\overline{\mathbf{R}})$  sur  $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty, +\infty\}$ . On admettra que les éléments de  $\mathcal{B}(\overline{\mathbf{R}})$  sont exactement les ensembles de la forme  $H$ ,  $H \cup \{+\infty\}$ ,  $H \cup \{-\infty\}$ ,  $H \cup \{-\infty, +\infty\}$  où  $H \in \mathcal{B}(\mathbf{R})$ . On en déduit que  $\mathcal{B}(\overline{\mathbf{R}})$  est la tribu engendrée par les intervalles de la forme  $[-\infty, b]$  pour  $b$  décrivant  $\overline{\mathbf{R}}$ .

REMARQUE.— A l'inverse, pour n'importe quel ensemble  $E \subset \mathbf{R}^d$ , la tribu borélienne sur  $E$  est la *trace* de la tribu borélienne de  $\mathbf{R}^d$  sur  $E$  : en termes moins sophistiqués, cela signifie que

$$\mathcal{B}(E) = \mathcal{B}(\mathbf{R}^d) \cap E.$$

### 4.3. Fonctions mesurables

On rappelle la notion d'*image réciproque* d'un ensemble par une application. On se donne deux ensembles  $\Omega$  et  $E$  et une application  $X$  de  $\Omega$  dans  $E$ . Pour une partie de  $E$ , notée



$H$ , l'image réciproque de  $H$  par  $X$ , notée  $X^{-1}(H)$  est l'ensemble des éléments de  $\Omega$  dont l'image par  $X$  est dans  $H$  :

$$X^{-1}(H) = \{\omega \in \Omega, X(\omega) \in H\}.$$

Cette notion est définie que  $X$  soit inversible ou pas.

EXEMPLE 18.— Considérons l'application qui à un entier  $n$  associe la partie entière de sa moitié :  $f(n) = [n/2]$ . Dans ces conditions, pour tout entier  $k$ ,  $f^{-1}(\{k\}) = \{2k, 2k+1\}$ .

La notion d'image réciproque (au contraire de la notion d'image directe) se marie bien avec les intersections et les unions d'ensemble.

**Proposition 4.6.** Pour toute famille dénombrable d'ensemble  $(A_n, n \geq 1)$

$$\begin{aligned} X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \bigcup_{n=1}^{\infty} X^{-1}(A_n), \\ X^{-1}\left(\bigcap_{n=1}^{\infty} A_n\right) &= \bigcap_{n=1}^{\infty} X^{-1}(A_n). \end{aligned}$$

Nantis de ces propriétés de base, nous sommes prêts pour donner la définition des fonctions mesurables.

**Définition 4.3.** Une application  $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$  est dite *mesurable* si :

$$\forall H \in \mathcal{E}, X^{-1}(H) \in \mathcal{F}.$$

En langage probabiliste, une application mesurable s'appelle une *variable aléatoire*.

EXEMPLE 19.— Si  $\Omega$  est munie de la tribu grossière, c'est-à-dire que  $\mathcal{F} = \{\emptyset, \Omega\}$  alors une fonction mesurable est nécessairement constante (dès que les singletons sont des mesurables de  $E$ ).

En effet, soit  $\omega \in \Omega$  et considérons  $A = X^{-1}(\{X(\omega)\})$ , l'ensemble des éléments de  $\Omega$  qui ont la même image par  $X$  que le  $\omega$  considéré. Puisque  $A$  est mesurable,  $A$  vaut  $\emptyset$  ou  $\Omega$ . Comme  $\omega \in A$ ,  $A$  ne peut être vide donc  $A = \Omega$ . Autrement dit,  $X$  est constante.

EXEMPLE 20.— De la même manière, on montre que si  $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$  alors les fonctions mesurables ne prennent que deux valeurs : elles sont de la forme

$$X(\omega) = \alpha \mathbf{1}_A(\omega) + \beta \mathbf{1}_{A^c}(\omega),$$

avec  $\alpha$  et  $\beta$ , deux éléments de  $E$ .

REMARQUE.— Si l'espace  $\Omega$  est dénombrable (ou fini), on a vu que la seule tribu sensée était la tribu des parties de  $\Omega$ . En d'autres termes, toute fonction définie sur un tel ensemble est nécessairement mesurable puisque l'image inverse d'un ensemble quelconque de  $E$  est par définition une partie de  $\Omega$ .

La propriété de composition des fonctions mesurables bien que totalement basique est d'un usage constant.

**Théorème 4.7.** Soit  $(E', \mathcal{E}')$  un espace mesurable. Soit  $X : \Omega \rightarrow E$  et  $f : E \rightarrow E'$  deux applications mesurables. La composée  $f \circ X : \Omega \rightarrow E'$  est une application mesurable.

*Démonstration.* Soit  $H' \in \mathcal{E}'$ . L'image réciproque de  $H'$  par  $f \circ X$  est égale à  $X^{-1}(f^{-1}(H'))$ . Comme  $f$  est mesurable,  $f^{-1}(H') \in \mathcal{E}$ . Comme  $X$  est mesurable, l'image réciproque d'un élément de  $\mathcal{E}$  est dans  $\mathcal{F}$ , d'où  $X^{-1}(f^{-1}(H')) \in \mathcal{F}$ .  $\square$

On admettra les deux résultats suivants. Le premier stipule que pour vérifier la mesurabilité d'une application  $X$ , il suffit de vérifier la propriété «  $X^{-1}(H) \in \mathcal{F}$  » non pas pour *tout*  $H \in \mathcal{E}$ , mais seulement pour  $H$  dans une classe plus réduite, qui engendre la tribu d'arrivée.

**Théorème 4.8.** Supposons que  $\mathcal{E} = \sigma(\mathcal{C})$  pour une certaine classe  $\mathcal{C}$ . Soit une fonction  $X : \Omega \rightarrow E$  telle que pour tout  $C \in \mathcal{C}$ ,  $X^{-1}(C) \in \mathcal{F}$ . Alors  $X$  est mesurable.

Le corollaire d'usage pratique est le suivant.

**Corollaire 4.9.** Soit  $d \in \mathbf{N}^*$  et soient  $X_1, \dots, X_d$  une collection de fonctions de  $\Omega$  dans  $\mathbf{R}$ . On définit  $X : \Omega \rightarrow \mathbf{R}^d$  par  $X(\omega) = (X_1(\omega), \dots, X_d(\omega))$ . Les propositions suivantes sont équivalentes :

- a)  $X_1, \dots, X_d$  sont des fonctions mesurables ;
- b)  $X$  est une fonction mesurable.

*Démonstration.* a) $\Rightarrow$ b). Puisque  $\mathcal{B}(\mathbf{R}^d)$  est la tribu engendrée par les pavés, il suffit de vérifier que  $X^{-1}(H) \in \mathcal{F}$  pour  $H = \prod_{k=1}^d H_k$ , où  $H_k$  est un intervalle réel. Or  $X^{-1}(\prod_k H_k) = \bigcap_k X_k^{-1}(H_k)$  est bien un élément de  $\mathcal{F}$  comme intersection d'éléments de  $\mathcal{F}$ .

b) $\Rightarrow$ a) Donnons la preuve pour  $X_1$ . Pour tout intervalle  $]a, b[$ , l'ensemble  $X_1^{-1}(]a, b[)$  est l'image réciproque par  $X$  de  $]a, b[ \times \mathbf{R} \times \dots \times \mathbf{R}$  qui appartient à  $\mathcal{B}(\mathbf{R}^d)$ . Donc  $X_1^{-1}(]a, b[) \in \mathcal{F}$ , ce qui montre que  $X_1$  est mesurable.  $\square$

Le deuxième qui nous simplifiera considérablement la vie fait le lien entre fonctions continues et fonctions mesurables.

**Théorème 4.10.** Une application  $f : \mathbf{R}^d \rightarrow \mathbf{R}^n$  continue est mesurable. On dit alors qu'elle est *borélienne*.

REMARQUE.— On prendra soin de noter que la réciproque est fausse : il existe des fonctions mesurables qui ne sont pas continues. Par exemple,  $\mathbf{1}_{\mathbf{Q}}$  la fonction indicatrice de l'ensemble des rationnels puisque  $\mathbf{Q}$  est mesurable (union dénombrable de singletons) mais manifestement pas continue.

A partir de là, on peut démontrer les propriétés élémentaires des fonctions mesurables.

**Corollaire 4.11.** Si  $X, Y$  sont des fonctions mesurable sur  $\mathbf{R}$ , alors  $X + Y, XY, X \vee Y := \max(X, Y), X \wedge Y := \min(X, Y)$  sont des fonctions mesurables.

*Démonstration.* La fonction

$$\begin{aligned} \mathbf{R} \times \mathbf{R} &\longrightarrow \mathbf{R} \\ (x, y) &\longmapsto x + y \end{aligned}$$

est continue donc borélienne de  $\mathbf{R}^2$  dans  $\mathbf{R}$  en vertu du théorème 4.10. Le résultat se déduit alors du théorème 4.7 par composition de fonctions mesurables. Les autres propriétés se montrent de façon similaire en utilisant les formules

$$x \vee y = \frac{x + y + |x - y|}{2} \text{ et } x \wedge y = \frac{x + y - |x - y|}{2}.$$

□

Jusqu'à présent, toutes les propriétés des fonctions mesurables existent à l'identique pour les fonctions continues. Ce qui fait la richesse du concept de mesurabilité tient aux théorèmes limites. En effet, rien ne garantit qu'une limite simple de fonctions continues soit une fonction continue alors que la limite d'une suite de fonctions mesurables est mesurable !

Nous allons maintenant montrer que si  $(X_n)_n$  est une suite de fonctions mesurables, alors le sup et l'inf de la suite sont aussi des variables aléatoires. Bien sûr, le sup d'une suite réelle n'est pas forcément fini. Pour donner un sens à  $\sup_n X_n$  et  $\inf_n X_n$  en tant que variables aléatoires, nous devons dorénavant nous placer sur  $\overline{\mathbf{R}}$ .

Soit  $(X_n, n \in \mathbf{N})$  une suite de variables aléatoires sur  $\overline{\mathbf{R}}$ . On désigne par respectivement par  $\sup_n X_n, \inf_n X_n, \lim_n X_n$  les fonctions définies sur  $\overline{\mathbf{R}}$  par  $\omega \mapsto \sup_n X_n(\omega), \omega \mapsto \inf_n X_n(\omega)$  et, lorsqu'une telle fonction existe,  $\omega \mapsto \lim_n X_n(\omega)$ .

**Théorème 4.12.** Soit  $(X_n, n \in \mathbf{N})$  une suite de fonctions mesurables sur  $\overline{\mathbf{R}}$ .

- 4.12.a)  $\sup_n X_n, \inf_n X_n$  sont des fonctions mesurables sur  $\overline{\mathbf{R}}$  ;
- 4.12.b) Les fonctions  $\liminf_n X_n$  et  $\limsup_n X_n$  sont mesurables ;
- 4.12.c) Si  $\lim_n X_n$  existe, c'est une fonction mesurable sur  $\overline{\mathbf{R}}$ .

*Démonstration.*

- 4.12.a)** Posons  $X := \sup_n X_n$ .  $\mathcal{B}(\overline{\mathbf{R}})$  est la tribu engendrée par les intervalles  $[-\infty, b]$  où  $b \in \overline{\mathbf{R}}$ . On laisse au lecteur le soin de vérifier que  $X^{-1}([-\infty, b]) = \bigcap_n X_n^{-1}([-\infty, b])$ . Ainsi,  $X^{-1}([-\infty, b])$  est dans  $\mathcal{F}$  comme intersection dénombrable d'éléments de  $\mathcal{F}$ . D'après le critère de mesurabilité du paragraphe précédent,  $X$  est mesurable. Pour montrer que  $\inf_n X_n$  est mesurable, il suffit d'écrire  $\inf_n X_n = -\sup_n (-X_n)$ .
- 4.12.b)** D'après le point précédent, les fonctions  $Y_k = \inf_{n \geq k} X_n$  et  $Z_k = \sup_{n \geq k} X_n$  sont mesurables. Toujours d'après ce point, les variables  $\sup_k Y_k$  et  $\inf_k Z_k$  sont aussi mesurables. Or par définition, elles sont respectivement égales à  $\liminf_n X_n$  et  $\limsup_n X_n$ .
- 4.12.c)** Si  $X_n$  converge, alors  $\lim_n X_n = \liminf_n X_n = \limsup_n X_n$  donc d'après le point précédent, c'est une variable aléatoire.

□

## 4.4. Mesures

### 4.4.1. Définition et exemples

**Définition 4.4.** Une mesure sur  $(E, \mathcal{E})$  est une fonction d'ensemble  $\mu$  telle que

4.4.i)  $\mu : \mathcal{E} \rightarrow [0, +\infty]$  ;

4.4.ii)  $\mu(\emptyset) = 0$  ;

4.4.iii) Pour toute famille  $(A_n, n \in \mathbf{N})$  d'événements deux à deux disjoints,

$$\mu\left(\bigcup_{n \in \mathbf{N}} A_n\right) = \sum_{n \in \mathbf{N}} \mu(A_n) .$$

On prêtera attention à l'intervalle fermé à droite dans l'axiome 4.4.i) : la mesure d'un ensemble  $A$  est une quantité positive, possiblement infinie.

Lorsque  $\mu(E) < \infty$ , on dit que la mesure est *finie*. Si en outre  $\mu(E) = 1$ , la mesure  $\mu$  est une *mesure de probabilité*.

Dans ce dernier cas, on utilise alors les notations probabilistes.

- $(E, \mathcal{E})$  devient  $(\Omega, \mathcal{A})$ .
- Les fonctions mesurables notées  $f$  ou  $g$  deviennent des variables aléatoires notées par des capitales  $X, Y$ , etc.
- La variable  $x \in E$  devient  $\omega \in \Omega$ .
- Les mesures  $\mu, \nu$  deviennent des probabilités notées  $\mathbf{P}, \mathbf{Q}$ .

Pour autant les concepts restent rigoureusement les mêmes !

Un triplet  $(E, \mathcal{E}, \mu)$  où  $\mu$  est une mesure sur  $(E, \mathcal{E})$ , est appelé un *espace mesuré*. C'est un espace de probabilité lorsque  $\mu$  est une mesure de probabilité.

L'axiome 4.4.i) traduit le fait que l'événement impossible a une probabilité nulle d'être réalisé, alors que l'événement certain est réalisé avec une probabilité égale à un. L'axiome 4.4.ii) est connu sous le nom de propriété de  $\sigma$ -additivité.

EXEMPLE 21 (Mesures de Dirac).— Soit  $(E, \mathcal{E})$  un espace mesurable quelconque et soit  $a \in E$ . La mesure de Dirac au point  $a$  est l'application  $\delta_a$  définie sur  $\mathcal{E}$  par  $\delta_a(A) = \mathbf{1}_A(a)$  c'est à dire :

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

EXEMPLE 22 (Mesure de comptage et mesures discrète).— Soit  $I$  un ensemble au plus dénombrable et  $(a_i)_{i \in I}$  une collection de points de  $E$ . La fonction définie pour tout  $A \in \mathcal{E}$  par :

$$\mu(A) := \sum_{i \in I} \delta_{a_i}(A)$$

définit une mesure sur  $\mathcal{E}$ , appelée la mesure de comptage de  $(a_i)_{i \in I}$ . La quantité  $\mu(A)$  est le nombre de points  $a_i$  contenus dans l'ensemble  $A$  :  $\mu(A) = \text{cardinal}\{i : a_i \in A\}$ .

Plus généralement, si  $(\alpha_i)_{i \in I}$  est une suite de coefficients réels positifs, alors :

$$\mu(A) := \sum_{i \in I} \alpha_i \delta_{a_i}(A)$$

définit une mesure sur  $\mathcal{E}$ . La quantité  $\mu(A)$  est égale à la somme des  $\alpha_i$  pour tous les  $i$  tels que  $a_i \in A$ . Une telle mesure est appelée une mesure *discrète*.

Une conséquence immédiate de l'axiome 4.1.iii) est la très utile formule des probabilités totales.

**Théorème 4.13** (Formule des probabilités totales). Pour toute partition  $(B_n, n \geq 1)$  constituée d'ensembles mesurables de  $E$ , pour tout mesurable  $A$ ,

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A \cap B_n).$$

*Démonstration.* Les ensembles  $(A \cap B_n, n \geq 1)$  forment une partition de  $A$ , le résultat découle de l'axiome 4.1.iii).  $\square$

Les propriétés suivantes sont d'un usage constant.

**Théorème 4.14.** Soient  $A, B, (A_n, n \in \mathbf{N}^*)$  des éléments de  $\mathcal{E}$ .

4.14.a) Si  $\mu(A \cap B) < \infty$ , alors  $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$ .

4.14.b) Si  $A \subset B$ , alors  $\mu(A) \leq \mu(B)$ .

4.14.c) Si  $A_n \uparrow A$ , alors  $\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n)$ .

4.14.d) Si  $A_n \downarrow A$  et si  $\mu(A_1) < \infty$ , alors  $\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n)$ .

4.14.e) Pour une famille quelconque  $(A_n, n \in \mathbf{N}^*)$  dans  $\mathcal{E}$ , on a la *borne de l'union* :

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

*Démonstration.*

4.14.a) On remarque que

$$A = A \setminus (A \cap B) \cup (A \cap B).$$

Comme ces deux ensembles sont disjoints, d'après l'axiome 4.1.iii),

$$\mu(A) = \mu(A \setminus (A \cap B)) + \mu(A \cap B).$$

De même,

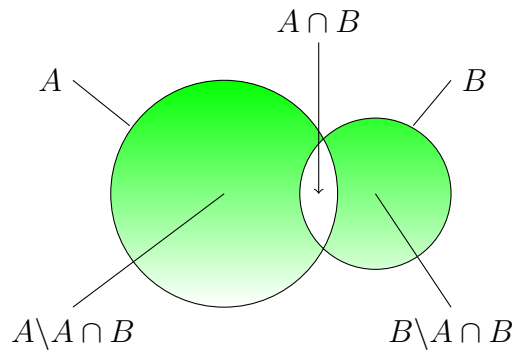
$$\mu(B) = \mu(B \setminus (A \cap B)) + \mu(A \cap B).$$

Enfin,

$$A \cup B = (A \setminus (A \cap B)) \cup (B \setminus (A \cap B)) \cup (A \cap B).$$

Par conséquent,

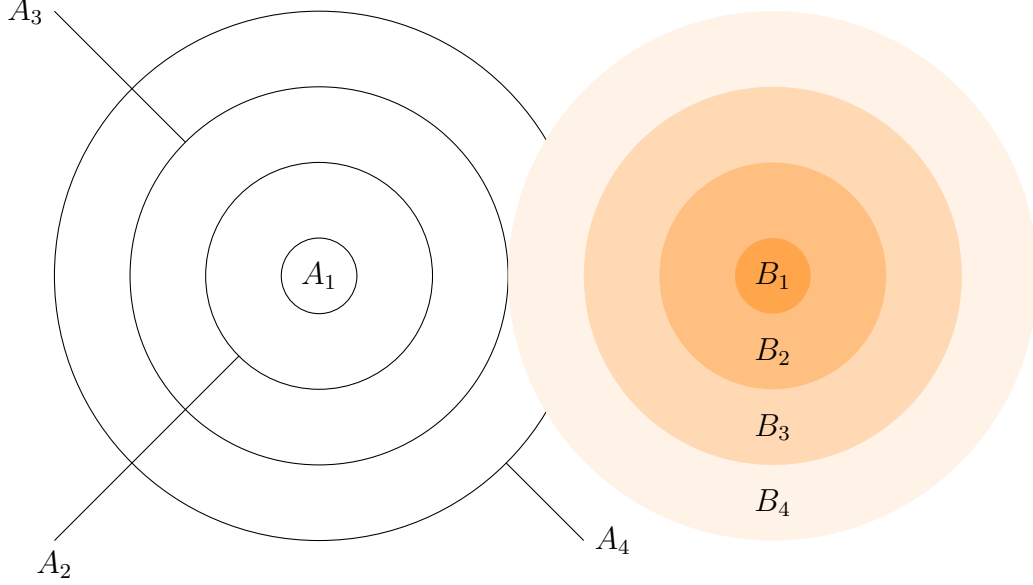
$$\begin{aligned} \mu(A \cup B) &= \mu(A \setminus (A \cap B)) + \mu(B \setminus (A \cap B)) + \mu(A \cap B) \\ &= \mu(A) - \mu(A \cap B) + \mu(B) - \mu(A \cap B) + \mu(A \cap B) \\ &= \mu(A) + \mu(B) - \mu(A \cap B). \end{aligned}$$



4.14.b) Si  $A \subset B$ ,  $B = A \cup (B \setminus A)$ , donc  $\mu(B) = \mu(A) + \mu(B \setminus A)$ . Comme  $\mu$  est à valeurs positives, on en déduit que  $\mu(B \setminus A) \geq 0$  donc  $\mu(B) \geq \mu(A)$ .

4.14.c) Comme la famille des  $A_n$  est croissante, on peut considérer la famille  $B_n$  définie par

$$B_1 = A_1, \quad B_n = A_n \setminus A_{n-1}.$$



On de façon évidente que

$$\bigcup_{k=1}^n B_k = A_n \text{ et } \bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k.$$

Comme les  $B_n$  sont disjoints,

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \mu\left(\bigcup_{k=1}^{\infty} B_k\right) = \lim_{n \rightarrow \infty} \sum_{k=1}^n \mu(B_k) = \lim_{n \rightarrow \infty} \mu\left(\bigcup_{k=1}^n B_k\right) = \lim_{n \rightarrow \infty} \mu(A_n).$$

4.14.d) Comme tout se passe dans l'ensemble  $A_1$ , on ne change rien en supposant que  $E = A_1$  et  $\mu(E) < \infty$ . La famille des  $A_n^c$  est croissante donc on peut appliquer le résultat précédent 4.14.c),

$$\mu\left(\bigcup_{n=1}^{\infty} A_n^c\right) = \lim_{n \rightarrow \infty} \mu(A_n^c).$$

Or,

$$\bigcup_{n=1}^{\infty} A_n^c = \left(\bigcap_{n=1}^{\infty} A_n\right)^c$$

et d'après 4.14.a), on obtient

$$\mu(E) - \mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(E) - \mu(A_n).$$

En simplifiant par  $\mu(E)$ , on obtient le résultat voulu.

4.14.e) Soit la propriété au rang  $k$

$$\mu\left(\bigcup_{n=1}^k A_n\right) \leq \sum_{n=1}^k \mu(A_n).$$

D'après 4.14.a), elle est vraie pour  $k = 2$ . Par récurrence, en utilisant 4.14.a), elle est vraie pour tout  $k$ . Maintenant, la famille  $B_k = \bigcup_{n=1}^k A_n$  est croissante donc d'après 4.14.c)

$$\begin{aligned} \mu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu\left(\bigcup_{k=1}^{\infty} B_k\right) = \lim_{k \rightarrow \infty} \mu(B_k) = \lim_{k \rightarrow \infty} \mu\left(\bigcup_{n=1}^k A_n\right) \\ &\leq \lim_{k \rightarrow \infty} \sum_{n=1}^K \mu(A_n) = \sum_{n=1}^{\infty} \mu(A_n). \end{aligned}$$

Le résultat s'en suit. □

Le lemme qui suit est fondamental dans nombre d'applications.

**Lemme 4.15** (Lemme de Borel-Cantelli). Soit un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$  et  $(A_n, n \geq 1)$  une famille d'éléments de  $\mathcal{E}$ .

$$\sum_{n \geq 1} \mathbf{P}(A_n) < +\infty \implies \mathbf{P}(\limsup_n A_n) = 0, \quad (4.1)$$

où la limite supérieure d'ensembles est définie par

$$\limsup_n A_n = \bigcap_{k \geq 1} \bigcup_{n \geq k} A_n.$$

*Démonstration.* Puisque la limite supérieure d'ensembles est une famille monotone d'ensembles,

$$\mathbf{P}(\limsup_n A_n) = \mathbf{P}\left(\bigcap_{k \geq 1} \bigcup_{n \geq k} A_n\right) = \lim_k \uparrow \mathbf{P}\left(\bigcup_{n \geq k} A_n\right) \leq \lim_k \sum_{n \geq k} \mathbf{P}(A_n).$$

La limite à droite est nulle comme limite du reste d'une série convergente  $\sum_n \mathbf{P}(A_n) < +\infty$ . □

#### 4.4.2. Caractérisation

Dès que  $\Omega$  n'est pas dénombrable, il est impossible de décrire une mesure en donnant sa valeur pour tous les ensembles mesurables. Arrive à notre secours le théorème de classe monotone (théorème 4.19 ci-dessous) qui nous dit, en substance, qu'une mesure est totalement déterminée par sa valeur sur un ensemble d'ensembles suffisamment riche.



**Définition 4.5.** Un  $\pi$ -système (ou algèbre)  $\mathcal{P}$  est une classe de sous-ensembles de  $\Omega$  stable pour l'intersection finie :  $\forall A, B \in \mathcal{P}, A \cap B \in \mathcal{P}$ .

EXEMPLE 23.— Des  $\pi$ -systèmes intéressants sur  $\mathbf{R}^d$  sont la classe des pavés de  $] -\infty, x_1] \times \cdots \times ] -\infty, x_d]$  ou la classe des pavés bornés

$$\mathcal{I}_d = \{[x_1, y_1] \times \cdots \times [x_d, y_d], -\infty < x_i < y_i < \infty, i = 1, \dots, d\}.$$

**Définition 4.6.** Un  $\lambda$ -système (ou classe monotone)  $\mathcal{L}$  est une classe de sous-ensembles de  $\Omega$  vérifiant :

- 4.6.i)  $\Omega \in \mathcal{L}$  ;
- 4.6.ii) Pour tout  $A \in \mathcal{L}, A^c \in \mathcal{L}$  ;
- 4.6.iii) Pour toute suite  $(A_n)_{n \in \mathbf{N}^*}$  d'éléments de  $\mathcal{L}$  deux à deux disjoints,  $\bigcup_n A_n \in \mathcal{L}$ .

REMARQUE.— La définition d'un  $\lambda$ -système est assez semblable à celles d'une tribu, à une différence majeure près : on n'impose pas que toute union dénombrable soit dans  $\mathcal{L}$ , mais seulement les unions dénombrables d'ensembles deux à deux disjoints. Un  $\lambda$ -système est également appelé une *classe monotone* pour la raison suivante : on peut montrer que la limite d'une suite croissante d'éléments de  $\mathcal{L}$  est dans  $\mathcal{L}$ . La lettre grecque  $\lambda$  de  $\lambda$ -système fait référence au L de limite.

On a le résultat pratique suivant (cf. [Kal98, Théorème 1.1]) :

**Théorème 4.16.** Soit  $\mathcal{P}$  un  $\pi$ -système et  $\mathcal{L}$  un  $\lambda$ -système. Si  $\mathcal{P} \subset \mathcal{L}$  alors  $\sigma(\mathcal{P}) \subset \mathcal{L}$ .

Voyons maintenant comment ce résultat abstrait permet de caractériser simplement les mesures. Afin d'éviter des pathologies désagréables, on se restreindra aux mesures *localement finies* :

**Définition 4.7.** Une mesure  $\mu$  sur  $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$  est *localement finie* si pour tout  $x \in \mathbf{R}^d$ , il existe un pavé borné  $I \subset \mathcal{I}_d$  contenant  $x$  tel que  $\mu(I) < \infty$ .

**Lemme 4.17.** Soit  $J \in \mathcal{I}_d$ ,  $\mu$  et  $\nu$  deux mesures sur  $\mathbf{R}^d$  telles que  $\mu(J) = \nu(J) < \infty$ . On note  $\mathcal{B}(J) = \{B \cap J, B \in \mathcal{B}(\mathbf{R}^d)\}$ , les traces sur  $J$  des boréliens de  $\mathbf{R}^d$ . L'ensemble

$$\mathcal{L} = \{A \in \mathcal{B}(J), \mu(A) = \nu(A)\} \tag{4.2}$$

est un  $\lambda$ -système de  $\Omega = J$ .

*Démonstration.* Il s'agit de vérifier les trois points qui caractérisent un  $\lambda$ -système.

4.6.i) Par hypothèse  $\mu(\Omega) = \nu(\Omega)$ .

4.6.ii) Comme l'ensemble de référence est  $\Omega = J$ , pour  $A \subset J$ , son complémentaire, toujours noté  $A^c$ , s'entend comme  $J \setminus A$ . Si  $A \in \mathcal{L}$ , d'après 4.14.a)

$$\mu(A^c) = \mu(J) - \mu(A) = \nu(J) - \nu(A) = \nu(A^c)$$

donc  $A^c \in \mathcal{L}$ .

4.6.iii) Si les  $(A_n, n \in \mathbf{N})$  sont deux à deux disjoints et appartiennent tous à  $\mathcal{L}$  alors d'après 4.4.iii) :

$$\mu\left(\bigcup_n A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \nu(A_n) = \nu\left(\bigcup_n A_n\right),$$

donc la réunion  $\bigcup_n A_n$  appartient à  $\mathcal{L}$ .

□

**Corollaire 4.18.** Deux mesures localement finies sur  $\mathbf{R}^d$  qui coïncident sur  $\mathcal{I}_d$  sont égales :

$$\mu(A) = \nu(A), \forall A \in \mathcal{I}_d \implies \mu(A) = \nu(A), \forall A \in \mathcal{B}(\mathbf{R}^d).$$

*Démonstration.* Soit  $J \in \mathcal{I}_d$ , d'après le lemme 4.17, l'ensemble

$$\mathcal{L} = \{A \in \mathcal{B}(J), \mu(A) = \nu(A)\}$$

est un  $\lambda$ -système. Par ailleurs,  $\mathcal{I}_d$  est un  $\pi$ -système. Par hypothèse,  $\mathcal{I}_d \subset \mathcal{L}$ . En vertu du théorème 4.16,  $\mathcal{B}(J) \subset \mathcal{L}$  donc les deux mesures coïncident sur  $\mathcal{B}(J)$ . Soit

$$J_M = [-M, M] \times \dots \times [-M, M].$$

Soit  $B \in \mathcal{B}(\mathbf{R}^d)$ . On vient de voir que

$$\mu(B \cap J_M) = \nu(B \cap J_M).$$

Les ensembles  $(B \cap J_M, M \geq 0)$  croissent avec  $M$  donc d'après la propriété de monotonie des mesures 4.14.c),

$$\mu(B) = \lim_{M \rightarrow \infty} \mu(B \cap J_M) = \lim_{M \rightarrow \infty} \nu(B \cap J_M) = \nu(B).$$

D'où le résultat.

□

En appliquant une preuve similaire avec le  $\pi$ -système  $\{]-\infty, x], x \in \mathbf{R}\}$ , on obtient le résultat suivant.

**Corollaire 4.19.** Pour identifier une probabilité sur  $\mathbf{R}$ , il faut et il suffit que l'on connaisse  $\mathbf{P}(]-\infty, x])$  pour tout réel  $x$ .

REMARQUE.— Ce résultat s'étend sans changement aux dimensions supérieures : pour identifier une probabilité sur  $\mathbf{R}^d$ , il faut et il suffit que l'on connaisse

$$\mathbf{P}(]-\infty, x_1] \times \dots \times ]-\infty, x_d])$$

pour tout  $d$ -uplet  $(x_1, \dots, x_d)$ .

### 4.4.3. Construction

Un autre théorème fondamental de la théorie de la mesure est le suivant (cf. [Kal98, Corollaire A1.2]) :

**Théorème 4.20** (Kolmogorov). Il existe une unique mesure, notée  $\lambda$ , sur  $\mathbf{R}^d$  muni de la tribu des boréliens, qui coïncide avec la mesure de longueur/surface/volume sur les pavés, c'est-à-dire telle que

$$\lambda([a_1, b_1[ \times \dots \times [a_d, b_d[) = (b_1 - a_1) \dots (b_d - a_d).$$

Cette mesure s'appelle la *mesure de Lebesgue*.

**Corollaire 4.21.** La mesure de Lebesgue est invariante par translation : c'est-à-dire que pour tout  $x \in \mathbf{R}^d$ , pour tout borélien  $A$ ,

$$\lambda(A + x) = \lambda(A),$$

où

$$A + x = \{y + x, y \in A\}.$$

*Démonstration.* Pour  $x \in \mathbf{R}^d$  fixé, pour un pavé  $A = ]a_1, b_1[ \times \dots \times [a_d, b_d[$ , son translaté est donné par

$$A + x = ]a_1 + x_1, b_1 + x_1[ \times \dots \times [a_d + x_d, b_d + x_d[$$

donc

$$\lambda(A + x) = \lambda(A).$$

Considérons la mesure  $\lambda_x$  définie par  $\lambda_x(A) = \lambda(A + x)$ . La mesure  $\lambda_x$  et la mesure de Lebesgue coïncident sur les pavés donc d'après l'unicité résultant du théorème 4.20, elles sont égales d'où le résultat.  $\square$

Cet exemple de transformation de mesure est un cas particulier du concept très important de *mesure image*. Étant donné un ensemble mesuré  $(E, \mathcal{E}, \mu)$  et d'une variable aléatoire  $X$  à valeurs dans  $(F, \mathcal{F})$ , comment construire une probabilité sur  $\mathcal{F}$ ? La réponse somme toute assez naturelle, dans le sens où il n'y en a pas tellement d'autre possible, est donnée par la définition suivante.

**Définition 4.8** (Mesure image). Soit un ensemble mesuré  $(E, \mathcal{E}, \mu)$  et une fonction mesurable  $f$  à valeurs dans  $(F, \mathcal{F})$ . La mesure image de  $\mu$  par  $f$  notée  $\mu_f$  (ou  $f^*\mu$  ou  $f^{-1}\mu$ ) est définie par :

$$\forall A \in \mathcal{F}, \mu_f(A) = \mu(f^{-1}(A)),$$

où, on le rappelle,  $f^{-1}(A)$  est l'image réciproque de  $A$  par  $f$ , c'est-à-dire  $f^{-1}(A) = \{x \in E, f(x) \in A\}$ . En probabilités, comme on omet systématiquement les  $x$ , on

privilégiera donc la notation  $(f^{-1}(A)) = (f \in A)$ , d'où au final,

$$\forall A \in \mathcal{F}, \mu_f(A) = \mu(f \in A).$$

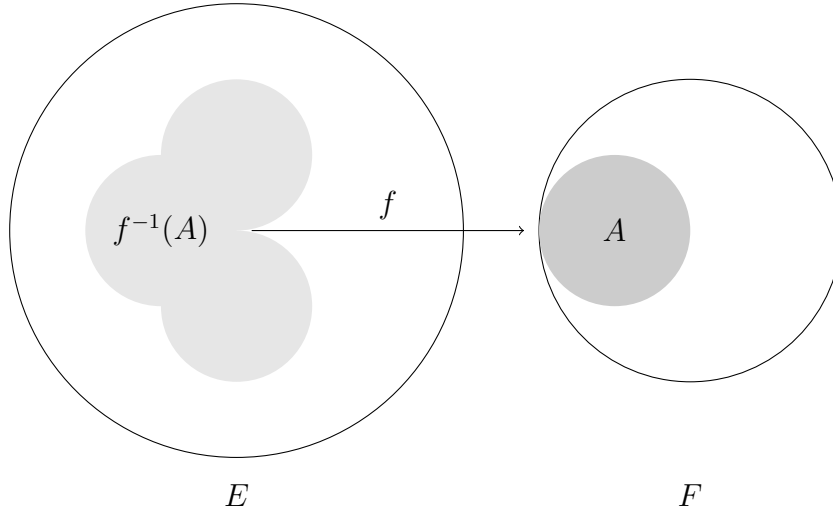


Figure 4.1. – Principe de construction d'une mesure image.

EXEMPLE 24. – Si l'on revient à l'exemple de mesure de Lebesgue « translatée », il faut considérer l'application

$$\begin{aligned} f_x : \mathbf{R}^d &\longrightarrow \mathbf{R}^d \\ y &\longmapsto y - x. \end{aligned}$$

En effet,

$$f_x^{-1}(A) = (y \in \mathbf{R}^d, f(y) \in A) = (y \in \mathbf{R}^d, y - x \in A) = (y \in \mathbf{R}^d, y \in A + x) = A + x.$$

La mesure  $\lambda_x$  introduite ci-dessus est donc la mesure image de la mesure de Lebesgue  $\lambda$  par l'application  $f_x$  qui translate tous les éléments de  $\mathbf{R}^d$  d'une quantité  $x$  fixe. Le corollaire 4.21 implique donc que cette mesure est la même que la mesure de Lebesgue. Ceci traduit le fait que du point de vue de la mesure de Lebesgue, le point 0 ne joue pas de rôle particulier puisque la mesure est uniforme dans tout l'espace.

Si on dispose de deux espaces mesurés  $(E_1, \mathcal{E}_1, \mu)$  et  $(E_2, \mathcal{E}_2, \nu)$ , on veut construire une mesure sur le produit cartésien  $E_1 \times E_2$ . La première difficulté à surmonter est la définition de la tribu sur  $\mathcal{E}_1 \times \mathcal{E}_2$ . Les éléments de  $\mathcal{E}_1 \times \mathcal{E}_2$  sont les produits cartésiens d'un élément de  $\mathcal{E}_1$  et d'un élément de  $\mathcal{E}_2$ . Comme la réunion de deux rectangles n'est pas un rectangle,  $\mathcal{E}_1 \times \mathcal{E}_2$  n'est pas une tribu. Qu'à cela ne tienne, on note  $\mathcal{E}_1 \otimes \mathcal{E}_2$  la plus petite tribu qui contient  $\mathcal{E}_1 \times \mathcal{E}_2$  et le tour est joué ! Le même procédé de preuve que celui qui est utilisé pour le théorème 4.20 permet de montrer le théorème suivant.

**Théorème 4.22.** Il existe une unique mesure, notée  $\mu \otimes \nu$ , dite mesure produit de

$\mu$  et  $\nu$ , sur  $(E_1 \times E_2, \mathcal{E}_1 \times \mathcal{E}_2)$  qui soit telle que

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B), \text{ pour tout } A \in \mathcal{E}_1, B \in \mathcal{E}_2. \quad (4.3)$$

## 4.5. Intégration

### 4.5.1. Construction de l'intégrale

Soit  $(E, \mathcal{E}, \mu)$  un espace mesuré.

**Définition 4.9.** Une fonction  $f$  mesurable de  $(E, \mathcal{E})$  dans  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  est dite étagée si elle prend un nombre fini de valeurs.

Une telle fonction  $f$  est alors de la forme

$$f(x) = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}(x) \quad (4.4)$$

où  $(\alpha_i, i = 1, \dots, n)$  sont les valeurs distinctes prises par la fonction  $f$  et où pour tout  $i$ ,  $A_i = f^{-1}(\{\alpha_i\})$ .

REMARQUE.— Les fonctions en *escalier* sont des cas particuliers des fonctions étagées où les  $A_i$  sont des intervalles.

**Définition 4.10.** La  $\mu$ -intégrale d'une fonction étagée positive  $f$  est la valeur (éventuellement égale à  $+\infty$ ) :

$$\int f \, d\mu = \sum_{i=1}^n \alpha_i \mu(A_i)$$

avec la convention  $0 \cdot \infty = 0$ , où les valeurs  $(\alpha_i, i = 1, \dots, n)$  et les ensembles  $(A_i, i = 1, \dots, n)$  sont donnés par (4.4).

La convention  $0 \cdot \infty = 0$  est indispensable pour l'éventualité où, pour un certain  $i$ ,  $\alpha_i = 0$  et  $\mu(A_i) = \infty$ . Grâce à cette convention, l'intégrale de la fonction nulle est égale à zéro.

Par définition de l'intégrale, on a la conséquence immédiate que pour tout  $A \in \mathcal{E}$ ,

$$\int \mathbf{1}_A d\mu = \mu(A).$$

En ce sens, la notion d'intégrale généralise la notion de mesure.

**Lemme 4.23.** Soient  $a, b \geq 0$  et  $f, g$  deux fonctions étagées positives. Alors

$$\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu.$$

*Démonstration.* A compléter. □

On définit maintenant l'intégrale de toute fonction  $f$  mesurable *positive*.

**Définition 4.11.** Soit  $f : E \rightarrow \overline{\mathbb{R}}$  une fonction mesurable positive ( $f \geq 0$ ). On définit :

$$\int f d\mu = \sup \left\{ \int g d\mu : 0 \leq g \leq f, g \text{ étagée} \right\}.$$

On vérifie aisément que si  $f$  est étagée, cette définition coïncide avec la précédente.

Il reste à désormais à définir l'intégrale dans le cas général de fonctions non-nécessairement positives. Pour toute fonction mesurable  $f$  à valeurs réelles, on introduit les notations

$$\begin{aligned} f^+ &= \max(f, 0) \\ f^- &= \max(-f, 0). \end{aligned}$$

**Définition 4.12.** Soit  $f : E \rightarrow \overline{\mathbb{R}}$  une fonction mesurable, telle que  $\int f^+ d\mu < \infty$  ou  $\int f^- d\mu < \infty$ . On définit

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

La fonction  $f$  est dite  $\mu$ -intégrable si  $\int f^+ d\mu < \infty$  et  $\int f^- d\mu < \infty$ .

L'intégrale de  $f$  est bien définie si  $\int f^+ d\mu < \infty$  ou  $\int f^- d\mu < \infty$  : cette précaution est nécessaire pour éviter la soustraction  $\infty - \infty$  qui n'a pas de sens. La fonction  $f$  est intégrable si les deux conditions sont satisfaites simultanément. En ce sens, une fonction peut ne pas être intégrable, bien que son intégrale soit définie. L'ensemble des fonctions  $\mu$ -intégrables est noté  $\mathcal{L}^1(\mu)$ .

Pour désigner l'intégrale de  $f$ , on utilise indifféremment les notations  $\int f d\mu$ ,  $\int f(x) d\mu(x)$ ,  $\int f(x) \mu(dx)$  ou, de manière à première vue impropre,  $\mu(f)$ .

### 4.5.2. Premières propriétés

**Lemme 4.24.** Soient  $f, g$  deux fonctions mesurables de  $E \rightarrow \overline{\mathbb{R}}$ , dont les intégrales sont définies. Alors

1.  $f \leq g$  implique  $\int f d\mu \leq \int g d\mu$ .
2.  $|\int f d\mu| \leq \int |f| d\mu$ .

*Démonstration.* 1. Supposons  $0 \leq f \leq g$ . Toute fonction  $\varphi$  étagée plus petite que  $f$  est aussi plus petite que  $g$ . On a donc l'inclusion

$$\left\{ \int \varphi d\mu : 0 \leq \varphi \leq f, \varphi \text{ étagée} \right\} \subset \left\{ \int \varphi d\mu : 0 \leq \varphi \leq g, \varphi \text{ étagée} \right\}.$$

Il en découle que le supremum de l'ensemble de gauche est plus petit que le supremum de l'ensemble de droite, ce qui se lit  $\int f d\mu \leq \int g d\mu$ .

Traisons le cas où  $f \leq g$  sont quelconques. On a  $f^+ \leq g^+$  et  $f^- \geq g^-$ , donc, d'après ci-dessus,  $\int f^+ d\mu \leq \int g^+ d\mu$  et  $\int f^- d\mu \geq \int g^- d\mu$ . En soustrayant,  $\int f d\mu \leq \int g d\mu$ .

2. On a  $f \leq |f|$  donc  $\int f d\mu \leq \int |f| d\mu$ . De même  $-f \leq |f|$ , donc  $\int (-f) d\mu \leq \int |f| d\mu$ . Or  $\int (-f) d\mu = \int (-f)^+ d\mu - \int (-f)^- d\mu = \int f^- d\mu - \int f^+ d\mu = -\int f d\mu$ . On a donc  $-\int f d\mu \leq \int |f| d\mu$ , ce qui conclut la preuve.  $\square$

Dans la suite, la notation  $0 \leq f_n \uparrow f$  est utilisée pour désigner une suite de fonctions  $f_n$  positives, convergeant ponctuellement vers  $f$ .

**Lemme 4.25.** Soient  $f, f_n$  ( $n \in \mathbb{N}$ ) des fonctions mesurables de  $E \rightarrow \overline{\mathbb{R}}$  telles que  $0 \leq f_n \uparrow f$ . Alors

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

*Démonstration.* Pour tout  $n$ ,  $f_n \leq f_{n+1} \leq f$ , donc  $\int f_n d\mu$  est une suite croissante majorée par  $\int f d\mu$ . Sa limite existe, et  $\lim \int f_n d\mu \leq \int f d\mu$ . Il reste à démontrer l'autre inégalité.

Soit une fonction étagée  $g$  telle que  $0 \leq g \leq f$ . Soit  $\varepsilon > 0$  et  $B_n = \{f_n \geq (1 - \varepsilon)g\}$ . Décrivons  $g$  sous la forme  $g = \sum_{i=1}^p \alpha_i \mathbf{1}_{A_i}$  (choisir par exemple, les  $\alpha_i$  comme les valeurs distinctes de  $g$  et  $A_i = g^{-1}(\{\alpha_i\})$ ). On a :

$$\begin{aligned} \int f_n d\mu &\geq \int f_n \mathbf{1}_{B_n} d\mu \geq \int (1 - \varepsilon)g \mathbf{1}_{B_n} d\mu \\ &\geq \int \left( \sum_{i=1}^p (1 - \varepsilon)\alpha_i \mathbf{1}_{A_i \cap B_n} \right) d\mu \\ &\geq \sum_{i=1}^p (1 - \varepsilon)\alpha_i \mu(A_i \cap B_n), \end{aligned}$$

où, dans la dernière inégalité, on a utilisé la linéarité de l'intégrale pour les fonctions étagées (Lemme 4.23). Comme  $B_n \uparrow E$ ,  $\mu(A_i \cap B_n)$  tend vers  $\mu(A_i)$  pour tout  $i$ , lorsque  $n \rightarrow \infty$ . Par conséquent,

$$\lim_{n \rightarrow \infty} \int f_n d\mu \geq (1 - \varepsilon) \sum_{i=1}^p \alpha_i \mu(A_i) = (1 - \varepsilon) \int g d\mu.$$

Cela étant vrai pour tout  $\varepsilon > 0$ , nous avons  $\lim_{n \rightarrow \infty} \int f_n d\mu \geq \int g d\mu$ . En prenant le supremum sur l'ensemble des fonctions  $g$  étagées positives minorant  $f$ , on conclut que  $\lim_{n \rightarrow \infty} \int f_n d\mu \geq \int f d\mu$ .  $\square$

Le lemme qui suit établit que toute fonction mesurable (positive) peut être approchée par une limite *simple* de fonctions mesurables. Il jouera un rôle essentiel dans la suite : pour démontrer telle ou telle propriété de l'intégrale, il suffira dans un premier temps de la démontrer pour des fonctions étagées positives, et dans un deuxième temps de choisir une suite de fonctions étagées approchant les fonctions de départ, en passant à la limite grâce au Lemme 4.25.

**Lemme 4.26.** Toute fonction mesurable  $f$  à valeurs dans  $\overline{\mathbf{R}}^+$  est limite croissante simple de fonctions étagées positives.

*Démonstration.* Soit

$$I = \{x : f(x) = +\infty\} \text{ et } E_{n,k} = f^{-1}\left(\left[\frac{k}{n}, \frac{k+1}{n}\right]\right).$$

Posons

$$t_n(x) = \sum_{k=0}^{2^n} \frac{k}{n} \mathbf{1}_{E_{n,k}}(x) + n \mathbf{1}_I(x).$$

Il est clair que  $t_n(x) \leq f(x)$  et que  $t_n$  converge simplement vers  $f$ . En posant,

$$f_1 = t_1, f_2 = f_1 \vee t_2, f_n = f_{n-1} \vee t_n, \dots$$

on construit une suite de fonctions qui prennent chacune un nombre fini de valeurs donc sont étagées, croissante et qui converge simplement vers  $f$ .  $\square$

### 4.5.3. Linéarité

**Théorème 4.27.** Soient  $f, g$  des fonctions mesurables de  $E \rightarrow \overline{\mathbf{R}}$ , et  $(a, b) \in \mathbb{R}^2$ . On suppose que l'une au moins des deux conditions suivantes est satisfaite :

1.  $f, g, a, b \geq 0$   
ou
2.  $f$  et  $g$  sont intégrables.



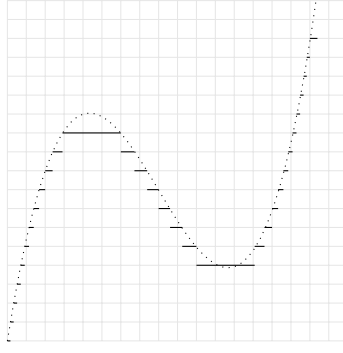


Figure 4.2. – Une fonction (en pointillés) et son approximation par une fonction étagée.

Alors  $\int (af + bg)d\mu = a \int f d\mu + b \int g d\mu$ .

*Démonstration.* 1. On se place sous la première condition. Considérons deux suites de fonctions étagées  $f_n$  et  $g_n$  telles que  $0 \leq f_n \uparrow f$  et  $0 \leq g_n \uparrow g$ . D'après le lemme 4.25,  $a \int f_n d\mu + b \int g_n d\mu \rightarrow a \int f d\mu + b \int g d\mu$ . Par ailleurs,  $a \int f_n d\mu + b \int g_n d\mu = \int (af_n + bg_n) d\mu$ . Il est facile de vérifier que  $0 \leq af_n + bg_n \uparrow af + bg$  et donc  $\int (af_n + bg_n) d\mu \rightarrow \int (af + bg) d\mu$ . Le résultat découle de l'identification des limites.

2. On se place sous la seconde condition. Il est facile de montrer que  $\int (af) d\mu = a \int f d\mu$ , il suffit donc dans la suite de considérer le cas où  $a = b = 1$ . En écrivant  $f + g$  de deux façons différentes,

$$\begin{aligned} f + g &= (f + g)^+ - (f + g)^- \\ &= f^+ - f^- + g^+ - g^-. \end{aligned}$$

Ainsi,

$$(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+.$$

On laisse au lecteur le soin de vérifier l'intégrabilité de  $f + g$ . Finalement, par passage à l'intégrale de l'équation précédente et utilisation de la linéarité dans le cas positif,

$$\int (f + g)^+ d\mu - \int (f + g)^- d\mu = \int f^+ d\mu - \int f^- d\mu - \int g^- d\mu + \int g^+ d\mu,$$

ce qui correspond à l'égalité souhaitée.  $\square$

Deux corollaires découlent de la propriété précédente et du lemme 4.25.

**Corollaire 4.28.**  $f$  est  $\mu$ -intégrable si et seulement si  $\int |f| d\mu < \infty$ .

*Démonstration.* Par définition  $f \in \mathcal{L}^1(\mu)$  précisément si  $\int f^+ d\mu + \int f^- d\mu < \infty$ , soit par linéarité de l'intégrale  $\int (f^+ + f^-) d\mu < \infty$ . La conclusion provient du fait que  $f^+ + f^- = |f|$ .  $\square$

**Corollaire 4.29.** Soit  $(f_n : n \in \mathbb{N})$  une suite de fonctions mesurables à valeurs positives. Alors

$$\sum_{n \in \mathbb{N}} \int f_n d\mu = \int \left( \sum_{n \in \mathbb{N}} f_n \right) d\mu.$$

*Démonstration.* Posons  $g_n = \sum_{k \leq n} f_k$ . On a  $0 \leq g_n \uparrow \sum_k f_k$  donc par le lemme 4.25,  $\int g_n d\mu \rightarrow \int \sum_k f_k d\mu$ . Par ailleurs, la linéarité de l'intégrale implique que  $\int g_n d\mu = \sum_{k \leq n} \int f_k d\mu$  qui converge vers  $\sum_k \int f_k d\mu$ . Il suffit d'identifier les limites.  $\square$

#### 4.5.4. Ensembles $\mu$ -négligeables

Un ensemble  $\mu$ -négligeable (ou simplement “négligeable” s'il n'y a pas d'ambiguïté) est un ensemble  $N \in \mathcal{E}$  tel que  $\mu(N) = 0$ .

**Proposition 4.30.** Soit  $f : \mathbf{E} \rightarrow \bar{\mathbb{R}}$  une fonction mesurable et  $N$  un ensemble  $\mu$ -négligeable. Alors

$$\int f \mathbf{1}_N d\mu = 0.$$

*Démonstration.* Traitons d'abord le cas où  $f \geq 0$ . Soit  $g$  une fonction étagée positive telle que  $g \leq f \mathbf{1}_N$ , disons de la forme  $g = \sum_{i=1}^p \alpha_i \mathbf{1}_{A_i}$  avec  $A_i = g^{-1}(\{\alpha_i\})$ . On a  $\int g d\mu = \sum_i \alpha_i \mu(A_i)$ . Il est facile de voir que chaque terme de la somme est nulle. En effet, pour  $i$  fixé tel que  $\alpha_i \neq 0$ , on  $g(x) = \alpha_i > 0$  et par construction de  $g \leq f \mathbf{1}_N$ , on a  $\mathbf{1}_N(x) \neq 0$ . Donc  $x \in N$ , ce qui montre que  $A_i \subset N$ , et donc  $\mu(A_i) = 0$ . Toute fonction étagée  $g$  minorant  $f \mathbf{1}_N$  est d'intégrale nulle. Donc  $f \mathbf{1}_N$  est d'intégrale nulle. On déduit la propriété dans le cas général en utilisant la décomposition  $\int f \mathbf{1}_N d\mu = \int f^+ \mathbf{1}_N d\mu - \int f^- \mathbf{1}_N d\mu = 0$ .  $\square$

Les ensembles mesurables de mesure nulle jouent un rôle particulier parce qu'ils ne sont pas « visibles » par la mesure même s'ils sont non vides.

**Définition 4.13** (Presque-partout, presque-sûre). On dit d'une propriété qu'elle est vraie  $\mu$ -presque-partout ou  $\mu$ -presque-sûrement, en abrégé  $\mu$ -p.p. ou  $\mu$ -p.s., lorsque son complémentaire est de mesure nulle.

Donnons un exemple. L'écriture “ $0 \leq f_n \uparrow f$   $\mu$ -pp” signifie qu'il existe un ensemble  $N \in \mathcal{E}$  tel que  $\mu(N) = 0$  et tel que pour tout  $x \in E \setminus N$ ,  $f_n(x)$  est une suite positive croissante convergeant vers  $f(x)$ .

Soit  $(E, \mathcal{E}, \mu)$  un espace probabilisé. Soit  $f$  une fonction mesurable positive de  $\mu$ -intégrale nulle. Pour tout  $n > 0$  et  $A_n = \{x : f(x) > 1/n\}$ ,

$$0 = \int f d\mu \geq \int_{A_n} f d\mu \geq \frac{1}{n} \mu(A_n),$$

donc  $\mu(A_n) = 0$  pour tout  $n \geq 1$ . Comme  $\bigcap_{n \geq 1} A_n = A_0$ , par monotonie, on obtient  $\mu(A_0) = 0$ . On a donc démontré le théorème suivant :

**Proposition 4.31.** Soit  $f$  une fonction mesurable positive, de  $\mu$ -intégrale nulle alors  $f$  est nulle  $\mu$ -presque partout.

Mais une fonction nulle « presque partout » n'est pas une fonction nulle. Par exemple, la fonction indicatrice de  $\mathbf{Q}$ , l'ensemble des rationnels, est presque-partout nulle pour la mesure de Lebesgue mais elle est non nulle sur un ensemble dense !

REMARQUE.— Ainsi, l'application  $f \mapsto \int |f| d\mu$  ne définit *pas* une norme sur l'espace vectoriel  $\mathcal{L}^1(\mu)$ , car la nullité de l'intégrale n'implique pas la nullité de la fonction. Afin de construire une norme, il faut que deux fonctions qui sont égales  $\mu$ -pp soient identifiées comme un seul et même élément. Formellement, cela revient à définir la relation d'équivalence suivante.

$f$  est en relation d'égalité p.p. avec  $g$  lorsque  $\mu(f \neq g) = 0$ . On note  $f \mathcal{R} g$ .

On définit alors  $L^1(\mu)$  comme l'ensemble des classes d'équivalence de la relation  $\mathcal{R}$ , et l'application qui à toute classe d'équivalence  $c$  associe la valeur  $\int |f| d\mu$  pour une fonction  $f$  quelconque de la classe  $c$  (le théorème 4.32 énoncé plus bas établit qu'elles ont toutes la même intégrale), définit bien une norme sur  $L^1(\mu)$ . L'espace  $L^1(\mu)$  est un espace vectoriel normé *complet*, c'est à dire que toute suite de Cauchy est convergente dans cet espace. La complétude est une propriété essentielle en analyse fonctionnelle.

Dans la pratique, travailler avec des classes d'équivalence est fastidieux, on peut continuer de penser les fonctions mesurables comme des fonctions ordinaires en prenant garde qu'elles ne sont définies qu'à un ensemble de mesure nulle près.

#### 4.5.5. Résultats fondamentaux

**Théorème 4.32.** Soient  $f, g : E \rightarrow \overline{\mathbb{R}}$  des fonctions dont les intégrales sont définies.

1.  $f \leq g$   $\mu$ -pp implique que  $\int f d\mu \leq \int g d\mu$ .
2.  $f = g$   $\mu$ -pp implique que  $\int f d\mu = \int g d\mu$ .

*Démonstration.* 1. Soit  $N = \{f > g\}$ . On a  $\mu(N) = 0$ . Comme  $f \mathbf{1}_N \leq g \mathbf{1}_N$ , on a  $\int f \mathbf{1}_N d\mu \leq \int g \mathbf{1}_N d\mu$ . Par ailleurs  $\int f \mathbf{1}_{\overline{N}} d\mu = \int g \mathbf{1}_{\overline{N}} d\mu = 0$ . Donc, par linéarité,

$$\int (f \mathbf{1}_N + f \mathbf{1}_{\overline{N}}) d\mu \leq \int (g \mathbf{1}_N + g \mathbf{1}_{\overline{N}}) d\mu,$$

soit  $\int f d\mu \leq \int g d\mu$ .

2. Immédiat à partir du résultat 1. □

**Théorème 4.33** (Théorème de convergence monotone). Soit  $(f_n, n \geq 1)$  une suite de fonctions (mesurables) positives qui converge en croissant,  $\mu$ -p.p. vers  $f$  alors

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int \left( \lim_{n \rightarrow \infty} f_n \right) d\mu = \int f d\mu.$$

La preuve découle directement du lemme 4.25, en traitant les ensembles négligeables de la même façon que dans la preuve du théorème 4.32.

**Théorème 4.34** (Lemme de Fatou). Soit  $(f_n, n \geq 1)$  une suite de fonctions (mesurables) positives,

$$\int \liminf_n f_n d\mu \leq \liminf_n \int f_n d\mu.$$

*Démonstration.* Par définition,

$$\liminf_n f_n(x) = \sup_k \inf_{n \geq k} f_n(x) = \lim_k \inf_{n \geq k} f_n(x).$$

La suite  $(g_k = \inf_{n \geq k} f_n, k \in \mathbf{N})$  est croissante positive donc le théorème de convergence monotone assure que

$$\int \lim_k g_k d\mu = \lim_k \int g_k d\mu.$$

Par conséquent,

$$\int \liminf_n f_n d\mu = \lim_k \int \inf_{n \geq k} f_n d\mu \leq \lim_k \inf_{n \geq k} \int f_n d\mu = \liminf_n \int f_n d\mu.$$

La preuve est terminée.  $\square$

Si la mesure est finie, on peut évidemment remplacer l'hypothèse de positivité par l'hypothèse que les  $f_n$  sont inférieurement bornées. Dans le cas où toutes les fonctions ne sont pas positives, il faut une contrainte de domination.

**Théorème 4.35** (Convergence dominée). Soit  $(f_n, n \geq 1)$  une suite de fonctions (mesurables) qui converge  $\mu$  p.p. vers  $f$ . Si de plus, il existe  $g$  telle que

$$|f_n(x)| \leq g(x), \text{ pour presque tout } x \text{ et } \int g d\mu < \infty$$

alors

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int \left( \lim_{n \rightarrow \infty} f_n \right) d\mu = \int f d\mu.$$

*Démonstration.* Comme  $|f_n| \leq g$  et que l'intégrale de  $g$  est finie, celle de  $|f_n|$  l'est aussi pour tout  $n$  entier. D'autre part,  $g - f_n$  et  $g + f_n$  sont mesurables positives et  $g \pm f_n$  tend vers  $g \pm f$  donc

$$\liminf_n g \pm f_n = g \pm f.$$

Appliquons le lemme de Fatou aux deux suite  $((g \pm f_n), n \geq 1)$ , on obtient

$$\begin{aligned}\int (g - f) d\mu &\leq \liminf_n \int (g - f_n) d\mu \\ \int (g + f) d\mu &\leq \liminf_n \int (g + f_n) d\mu.\end{aligned}$$

De plus, le lemme de Fatou implique aussi que

$$\int |f| d\mu = \int \liminf_n |f_n| d\mu \leq \liminf_n \int |f_n| d\mu \leq \int g d\mu < +\infty,$$

donc  $f$  est intégrable. On peut donc écrire

$$\begin{aligned}\int g d\mu - \int f d\mu &\leq \int g d\mu + \liminf_n \int (-f_n) d\mu \\ &= \int g d\mu - \limsup_n \int f_n d\mu \\ \int g d\mu + \int f d\mu &\leq \int g d\mu + \liminf_n \int f_n d\mu.\end{aligned}$$

On tire de ces inégalités que

$$\int f d\mu \leq \liminf_n \int f_n d\mu \leq \limsup_n \int f_n d\mu \leq \int f d\mu,$$

d'où l'on conclut que la suite  $\int f_n d\mu$  converge et que la limite est  $\int f d\mu$ .  $\square$

### 4.5.6. Exemples

EXEMPLE 25 (Mesure de comptage).— Si  $\mu = \sum_{n \in \mathbf{N}} \delta_n$ , alors

$$\mu(A) = \text{card}(A) = \sum_{n \in A} 1 = \sum_{n \in \mathbf{N}} \mathbf{1}_A(n) = \int \mathbf{1}_A d\mu.$$

Par conséquent, pour  $f : \mathbf{N} \rightarrow \mathbf{R}$ , intégrable,

$$\int f d\mu = \sum_{n \in \mathbf{N}} f(n).$$

Il faut donc garder à l'esprit que l'intégrale de Lebesgue peut se réaliser en ce qui est communément appelé une série.

EXEMPLE 26 (Mesure de Lebesgue).— Par construction, la mesure de Lebesgue est telle que

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

Réécrit dans le langage de la TdM, cela revient à dire

$$\int \mathbf{1}_{[a, b]} d\lambda = b - a = \int_a^b dx, \quad (4.5)$$

où  $\int f(x) dx$  représente l'intégrale de Riemann.

REMARQUE.— Pour les fonctions en escalier, qui sont des cas particuliers de fonctions étagées avec des  $A_i$  qui sont nécessairement des intervalles, les intégrales de Riemann et Lebesgue coïncident. Plus généralement, on admettra que toutes les fonctions Riemann intégrables (fonctions continues sur un compact ou d'intégrale absolument convergente) sont Lebesgue intégrables (voir [Kut98] pour la démonstration). C'est pour cette raison que l'on notera

$$\int f d\lambda = \int f(x) dx,$$

même si maintenant le terme de droite désigne l'intégrale de Lebesgue de  $f$ .

REMARQUE.— L'intégration « à la Lebesgue » permet d'intégrer plus de fonctions qu'avec l'intégrale de Riemann : pour être Riemann intégrable, une fonction (positive) se doit d'être continue sauf en un nombre dénombrable de points. Pour être Lebesgue intégrable, une fonction positive peut se contenter d'être mesurable. La notion de mesurabilité est nettement moins contraignante que la continuité.

Il est par exemple clair que  $\mathbf{Q}$  est mesurable (en tant que réunion dénombrable de singletons), de mesure de Lebesgue nulle (puisque les singletons sont de mesure nulle et en vertu de l'axiome 4.1.iii) de la définition 4.4). Par conséquent,  $\mathbf{1}_{\mathbf{Q}}$  est mesurable mais discontinue en tout point irrationnel : soit  $r \in \mathbf{Q}$ , il existe une suite  $(q_n, n \geq 1)$  de rationnels qui converge vers  $r$  et

$$\lim_{n \rightarrow +\infty} \mathbf{1}_{\mathbf{Q}}(q_n) = 1 \neq 0 = \mathbf{1}_{\mathbf{Q}}(r).$$

Cette fonction est donc Lebesgue intégrable mais pas Riemann intégrable.

Plus généralement, l'intérêt de l'intégrale de Lebesgue se trouve aussi dans la complétude de l'espace des fonctions intégrables, dont il a été question plus haut.

### 4.5.7. Théorème de Fubini-Tonelli

Le théorème de Fubini est celui qui permet de calculer des intégrales multiples en choisissant l'ordre des intégrations. La mesure produit a été définie dans la définition 4.22.

**Théorème 4.36** (Fubini-Tonelli). Soit  $(E \times F, \mathcal{E} \otimes \mathcal{F}, \mu \otimes \nu)$  un espace mesuré produit et  $f$  une fonction mesurable de  $E \times F$  dans  $\mathbf{R}^+$ . On a

- $x \mapsto \int f(x, y) d\nu(y)$  est mesurable,
- $y \mapsto \int f(x, y) d\mu(x)$  est mesurable,
- on peut choisir l'ordre d'intégration :

$$\int f(x, y) d(\mu \otimes \nu)(x, y) = \int \left( \int f(x, y) d\nu(y) \right) d\mu(x) = \int \left( \int f(x, y) d\mu(x) \right) d\nu(y). \quad (4.6)$$

*Démonstration.* Pour  $f = \mathbf{1}_{A \times B}$  avec  $A \in \mathcal{E}$  et  $B \in \mathcal{F}$ ,

$$\int f(x, y) d\nu(y) = \nu(B) \mathbf{1}_A(x),$$

qui est une fonction mesurable puisque  $A \in \mathcal{E}$ . Par un argument de classe monotone, on en déduit que ce résultat reste vrai pour  $f = \mathbf{1}_C$  avec  $C \in \mathcal{E} \otimes \mathcal{F}$  dont on rappelle qu'il s'agit de la tribu engendrée par les ensembles de la forme  $A \times B$  avec  $A \in \mathcal{E}$  et  $B \in \mathcal{F}$ . Par linéarité, la mesurabilité est toujours vérifiée pour les fonctions étagées (cf. définition 4.9) de  $E \times F$  dans  $\mathbf{R}^+$ . Pour une fonction mesurable positive, par construction

$$\int f(x, y) d\nu(y) = \lim_{n \rightarrow \infty} \int f_n(x, y) d\nu(y)$$

où  $(f_n, n \geq 1)$  est une suite de fonctions étagées qui converge vers  $f$ . La fonction

$$x \mapsto \int f(x, y) d\nu(y)$$

apparaît donc comme la limite d'une suite de fonctions mesurables, elle est donc mesurable (cf. 4.12.c). Le même raisonnement s'applique évidemment pour

$$y \mapsto \int f(x, y) d\mu(x).$$

La deuxième partie de la preuve suit rigoureusement le même schéma. Pour  $f = \mathbf{1}_{A \times B}$  avec  $A \in \mathcal{E}$  et  $B \in \mathcal{F}$ ,

$$\int f(x, y) d(\mu \otimes \nu)(x, y) = \mu \otimes \nu(A \times B) = \mu(A) \nu(B),$$

et

$$\int \left( \int f(x, y) d\nu(y) \right) d\mu(x) = \int \nu(B) \mathbf{1}_A(x) d\mu(x) = \nu(B) \int \mathbf{1}_A(x) d\mu(x) = \nu(B) \mu(A),$$

$$\int \left( \int f(x, y) d\mu(x) \right) d\nu(y) = \int \mu(A) \mathbf{1}_B(y) d\nu(y) = \mu(A) \int \mathbf{1}_B(y) d\nu(y) = \mu(A) \nu(B).$$

La formule (4.6) est donc vraie pour ces fonctions  $f$  particulières. Par un argument de classe monotone, il reste vrai pour les indicatrices de  $C \in \mathcal{E} \otimes \mathcal{F}$ . Par linéarité, il est vrai pour les fonctions étagées. Par passage à la limite, il est vrai pour les fonctions mesurables positives.  $\square$

Dans le cas où n'est pas de signe constant, on vérifie d'abord que l'intégrale double de  $|f|$  est finie (en utilisant le théorème précédent pour la calculer) puis on applique le théorème de Fubini proprement dit.

**Théorème 4.37** (Fubini). Soit  $(E \times F, \mathcal{E} \otimes \mathcal{F}, \mu \otimes \nu)$  un espace mesuré produit et

$f$  une fonction mesurable de  $E \times F$  dans  $\mathbf{R}$ . Si

$$\iint |f(x, y)| d\mu(x) d\nu(y) < \infty,$$

alors les conclusions du théorème 4.36 restent inchangées.

*Démonstration.* D'après le théorème de Fubini-Tonnelli,

$$\int \left( \int |f(x, y)| d\nu(y) \right) d\mu(x) < \infty$$

donc la fonction

$$x \mapsto \int |f(x, y)| d\nu(y)$$

est finie  $\mu$ -presque-sûrement : si l'on note

$$N_E = \{x \in E, \int |f(x, y)| d\nu(y) = \infty\}$$

alors  $\mu(N_E) = 0$  et donc, par construction de la mesure produit,  $(\mu \otimes \nu)(N_E \times F) = 0$ . De même,

$$N_F = \{y \in F, \int |f(x, y)| d\mu(x) = \infty\}$$

est de  $\nu$  mesure nulle donc  $(\mu \otimes \nu)(E \times N_F) = 0$ . On peut redéfinir  $f$  par 0 sur ces deux ensembles. En décomposant cette fonction (redéfinie mais toujours notée  $f$ ) en  $f = f^+ - f^-$ , le résultat découle par linéarité de (4.6).  $\square$

REMARQUE.— Il convient de noter que le théorème de Fubini permet de permuter des intégrales mais aussi des séries (puisque une série est au sens de Lebesgue, une intégrale par rapport à la mesure de comptage) ou des séries et des intégrales (ce que vous avez appelé jusqu'à présent le théorème d'intégration terme à terme).

REMARQUE.— Les plus attentifs remarqueront que le théorème suivant se distingue de celui que vous connaissez pour l'intégrale de Riemann en ce qu'il n'est pas nécessaire de vérifier que la fonction dominante et la fonction limite sont continues !

**Théorème 4.38** (Continuité sous le signe somme). Soit  $I$  un ouvert de  $\mathbf{R}^n$  et  $\{f(x, t), t \in I\}$  une famille de fonctions mesurables telle que pour tout  $t \in I$ ,  $f(\cdot, t)$  soit  $\mu$ -intégrable. Si  $\mu(x)$ -p.p.,  $t \mapsto f(x, t)$  est continue sur  $I$ , s'il existe  $G$  une fonction mesurable telle que

- $\mu(x)$  p.p.,  $|f(x, t)| \leq G(x)$  pour tout  $t \in I$ ,
- $\int G d\mu < \infty$ ,

alors l'application  $t \mapsto \int f(x, t) d\mu(x)$  est continue.



*Démonstration.* Soit  $(t_n, n \geq 1)$  une suite d'éléments de  $I$  qui tend vers  $t$ . Alors, on a évidemment

$$\left| \int f(x, t) d\mu(x) - \int f(x, t_n) d\mu(x) \right| \leq \int_E |f(x, t) - f(x, t_n)| d\mu(x).$$

Les assertions suivantes sont immédiates :

- $g_n : x \mapsto |f(x, t) - f(x, t_n)|$  est mesurable ;
- $g_n(x) = |f(x, t) - f(x, t_n)| \xrightarrow{n \rightarrow \infty} 0, \mu\text{-p.p.}$  ;
- $|g_n(x)| = |f(x, t) - f(x, t_n)| \leq 2G(x), \mu\text{-p.p.}$  ;
- $\int G d\mu < \infty$  ;

donc on peut appliquer le théorème de convergence dominée (théorème 4.35) à  $g_n$ , ce qui induit que

$$\int_E |f(x, t) - f(x, t_n)| d\mu(x) \xrightarrow{n \rightarrow \infty} 0.$$

D'où le résultat. □

Une preuve similaire permet d'obtenir le résultat de dérivation sous le signe « somme ».

**Théorème 4.39.** Soit  $I$  un ouvert de  $\mathbf{R}^n$  et  $\{f(x, t), t \in I\}$  une famille de fonctions mesurables telles que pour tout  $t \in I$ ,  $f(\cdot, t)$  soit  $\mu$ -intégrable. Si  $t \mapsto f(x, t)$  est dérivable sur  $I$ ,  $d\mu$  p.p., s'il existe  $G(x)$  une fonction mesurable telle que

- $\mu$  p.p.,  $|\frac{d}{dt} f(x, t)| \leq G(x)$  pour tout  $t \in I$ ,
- $\int G d\mu < \infty$ ,

alors l'application

$$\begin{aligned} I &\longrightarrow \mathbf{R} \\ t &\longmapsto \int f(x, t) d\mu(x) \end{aligned}$$

est dérivable sur  $I$  et

$$\frac{d}{dt} \int f(x, t) d\mu = \int \frac{d}{dt} f(x, t) d\mu(x).$$

### 4.5.8. Fonctions test

Pour caractériser une mesure de probabilité sur  $\mathbf{R}^d$ , il est souvent utile de considérer l'intégrale d'une fonction « test », mesurable positive ou mesurable bornée.

**Théorème 4.40.** Deux mesures de probabilité  $\mathbf{P}$  et  $\mathbf{Q}$  sur  $\mathbf{R}^d$  sont égales si et seulement si pour toute fonction  $f$  mesurable positive (respectivement, mesurable

bornée),

$$\int f d\mathbf{P} = \int f d\mathbf{Q}. \quad (4.7)$$

*Démonstration.* Il suffit de remarquer que pour tout borélien  $B$  de  $\mathbf{R}^d$ , la fonction  $f = \mathbf{1}_B$  est mesurable positive (respectivement, mesurable bornée). L'égalité (4.7) donne  $\mathbf{P}(B) = \mathbf{Q}(B)$ , si bien que  $\mathbf{P} = \mathbf{Q}$ .  $\square$

### 4.5.9. Mesures à densité

**Définition 4.14.** Soit  $\mu$  et  $\nu$  deux mesures sur  $(E, \mathcal{E})$ . On dit que  $\nu$  est absolument continue par rapport à  $\mu$  si et seulement si

$$A \in \mathcal{E}, \mu(A) = 0 \implies \nu(A) = 0.$$

On note  $\nu \ll \mu$ .

**Théorème 4.41** (Radon-Nikodym). Soit  $\mu$  et  $\nu$  deux mesures, finies sur tous compacts, sur  $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$  telles que  $\nu \ll \mu$ . Il existe une unique fonction mesurable  $f \geq 0$  (définie  $\mu$  presque-sûrement), telle que

$$\int_{\mathbf{R}^d} f d\mu < \infty \text{ et } \int_{\mathbf{R}^d} g d\nu = \int_{\mathbf{R}^d} g f d\mu,$$

pour toute fonction  $g$  mesurable bornée. La fonction  $f$  s'appelle la *densité* de  $\nu$  par rapport à  $\mu$ .

Lorsque l'on dit qu'une mesure  $\nu$  est absolue continue sans autre précision, on sous-entend qu'elle l'est par rapport à la mesure de Lebesgue, c'est-à-dire qu'il existe  $f$  telle que

$$d\nu(x) = f(x)dx.$$

**Définition 4.15.** Le support d'une mesure  $\mu$  est le complémentaire du plus grand ouvert de  $\mu$ -mesure nulle.

EXEMPLE 27.— Par exemple, le support de la mesure  $\delta_x$  est le singleton  $\{x\}$ . Pour une mesure à densité (sous-entendu par rapport à la mesure de Lebesgue)  $f$  supposée continue par morceaux, le support de la mesure  $f d\lambda$  est l'adhérence de l'ensemble des valeurs où  $f$  ne s'annule pas.

EXEMPLE 28.— Soit  $\mu = \delta_0$ , les mesures absolument continues par rapport à  $\mu$  sont de la forme  $\nu = \alpha \delta_0$  avec  $\alpha \geq 0$ . Soit  $\nu \ll \mu$ , pour tout ouvert  $A$  ne contenant pas 0,  $\mu(A) = 0$  donc  $\nu(A) = 0$ . Le support de  $\nu$  est donc inclus dans  $\{0\}$ . Ce qui signifie que  $\nu = \alpha \delta_0$ . La

fonction  $g$  est alors définie par  $g(0) = \alpha$  et  $g$  quelconque sur  $\mathbf{R} \setminus \{0\}$ . En effet,

$$\int f d\nu = f(0)\alpha = \int f g d\mu.$$

REMARQUE.— Pour  $\mu$  et  $\nu$ , deux mesures de probabilité telles que  $\nu \ll \mu$ , on remarquera (en prenant  $g = \mathbf{1}$ ) que nécessairement

$$\int_{\mathbf{R}^d} f d\mu = 1. \quad (4.8)$$

De nombreuses probabilités sont définies comme des mesures à densité par rapport à la mesure de Lebesgue. La table 4.2 fournit différents exemples de densités de probabilité  $f$ , sur lesquels nous aurons l'occasion de revenir pendant ce cours.

Table 4.1. – Quelques exemples de densités de probabilité – (Rappel : $\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$ ).			
Domaine	Densité	Expression de $f(x)$	Notation
$\mathbf{R}$	Densité uniforme sur $[a, b]$	$\frac{\mathbf{1}_{[a,b]}(x)}{b-a}$	$\mathcal{U}([a, b])$
$\mathbf{R}^d$	Densité uniforme sur une partie $A \subset \mathbf{R}^d$	$\frac{\mathbf{1}_A(x)}{\int \mathbf{1}_A}$	
$\mathbf{R}$	Densité exponentielle de paramètre $\alpha > 0$	$\alpha e^{-\alpha x} \mathbf{1}_{\mathbf{R}_+}(x)$	$\mathcal{E}(\alpha)$
$\mathbf{R}$	Densité gaussienne de paramètres $m, \sigma^2$ ( $m \in \mathbf{R}, \sigma^2 > 0$ )	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/(2\sigma^2)}$	$\mathcal{N}(m, \sigma^2)$
$\mathbf{R}^d$	Gaussienne multivariée de paramètres $m, \Sigma$ ( $m \in \mathbf{R}^d, \Sigma \in \mathbf{R}^{d \times d}$ définie positive)	$\frac{e^{-\frac{1}{2}(x-m)^T \Sigma^{-1}(x-m)}}{\sqrt{(2\pi)^d \det \Sigma}}$	$\mathcal{N}_d(m, \Sigma)$
$\mathbf{R}$	Densité de Cauchy de paramètres $m, \alpha$ ( $m \in \mathbf{R}, \alpha > 0$ )	$\frac{1}{\pi} \cdot \frac{\alpha}{(x-m)^2 + \alpha^2}$	
$\mathbf{R}$	Densité Gamma de paramètres $a, b$ ( $a > 0, b > 0$ )	$\frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$	$\Gamma(a, b)$

Table 4.2. – Quelques exemples de densités de probabilité – (Rappel : $\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$ ).			
---	--	--	--

A titre d'exemple, la figure 4.3 représente la densité gaussienne de paramètres 0 et 1, notée  $\mathcal{N}(0, 1)$  et appelée gaussienne *centrée réduite* :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} . \quad (4.9)$$

La densité gaussienne porte aussi le nom de densité *normale*, ce qui justifie la notation  $\mathcal{N}$ .

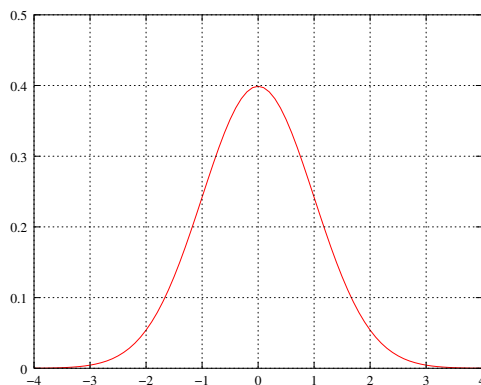
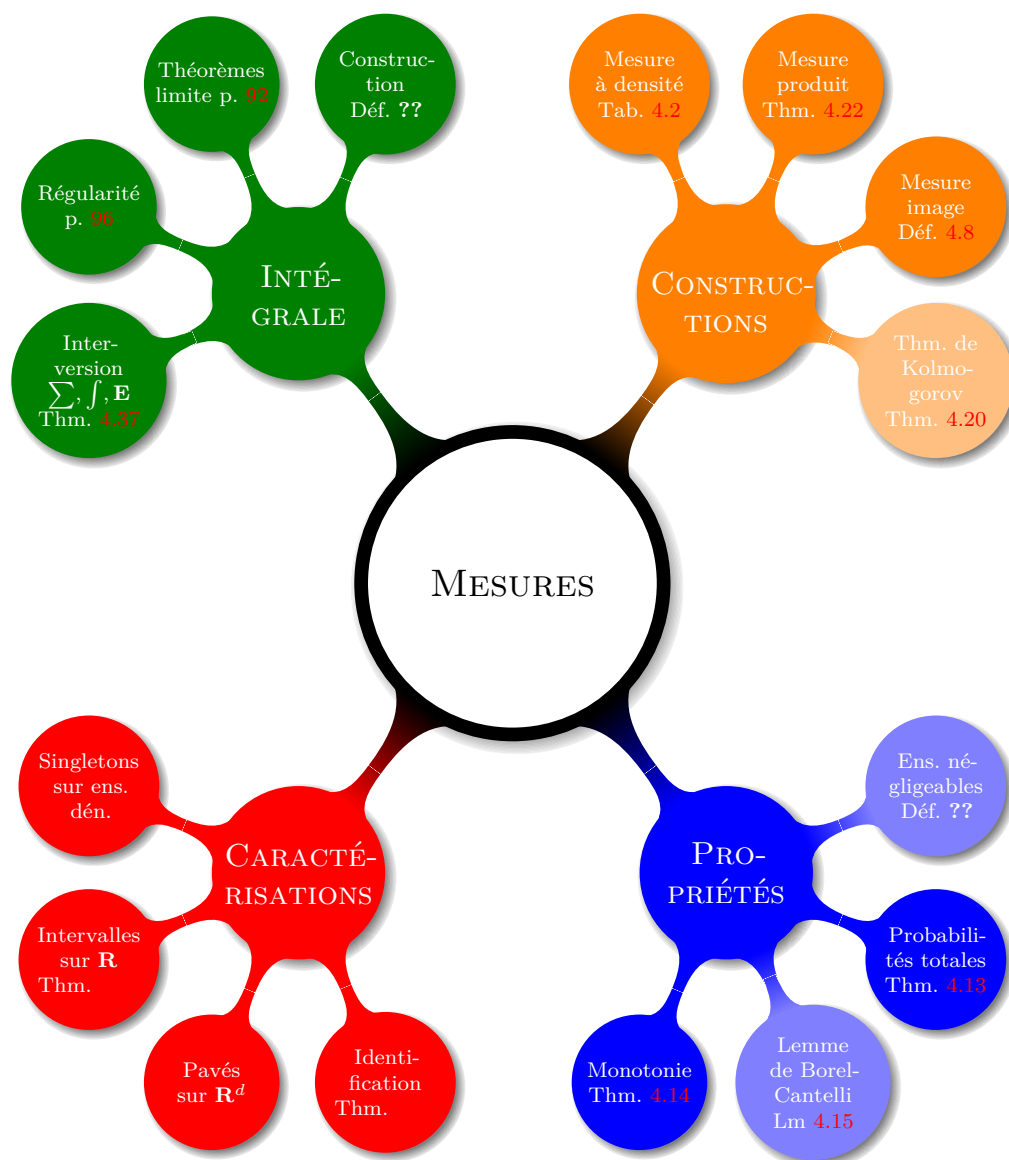


Figure 4.3. – Densité gaussienne centrée réduite  $\mathcal{N}(0, 1)$ .



## 4.6. Exercices

### Pour apprendre

EXERCICE 4.1 (Vrai ou faux?).— 1. Soit la fonction  $f$  définie par

$$\begin{aligned} f : \mathbf{R} &\longrightarrow \mathbf{N} \\ x &\longmapsto [\sqrt{|x|}] \end{aligned}$$

où  $[x]$  est la partie entière de  $x$ . On a

$$f^{-1}([2; 3.5]) = [4; 16[ \cup ] - 16; -4].$$

2. Une fonction constante est toujours mesurable.
3. Une fonction étagée prend un nombre fini de valeurs.
4. Une fonction mesurable qui prend un nombre fini de valeurs est étagée.
5. La mesure d'un singleton est strictement positive.
6. On a toujours

$$\lim_{x \uparrow x_0} \mu([-\infty; x]) = \mu([-\infty; x_0]).$$

7. Pour  $\mu$  mesure de probabilité, on a toujours

$$\lim_{x \downarrow x_0} \mu([-\infty; x]) = \mu([-\infty; x_0]).$$

EXERCICE 4.2.—

Soit  $\mu$  la mesure de comptage sur  $\mathbf{N}$  et pour tout entier  $n$ ,  $A_n = \{k, k \geq n\}$ .

1. Que vaut  $\cap_n A_n$ ?
2. Montre que l'on n'a pas  $\mu(\cap_n A_n) = \lim_{n \rightarrow \infty} \mu(A_n)$ .
3. Pourquoi n'y a-t-il pas de contradiction avec 4.14.d)?

EXERCICE 4.3 (Crible de Poincaré).—

Soit  $\mu$  une mesure finie. Montrer la formule suivante.

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} \mu(A_{i_1} \cap A_{i_2}) + \dots + (-1)^{n+1} \mu\left(\bigcap_{i=1}^n A_i\right).$$

EXERCICE 4.4.—

Soit  $E$  un ensemble muni d'une tribu  $\mathcal{E}$ . Soit  $X : \Omega \rightarrow E$  une fonction. Montrer que la famille  $\{X^{-1}(H) : H \in \mathcal{E}\}$  forme une tribu.

On l'appelle la *tribu engendrée par  $X$*  et on la note  $\sigma(X)$ . De manière informelle, on peut interpréter  $\sigma(X)$  comme l'ensemble des événements dont un observateur qui disposerait seulement de la valeur de  $X$  pourrait décider s'ils sont ou non réalisés.

EXERCICE 4.5 (\*\*\*).—

Soit  $f$  une fonction de  $\mathbf{R}$  dans  $\mathbf{R}$  bornée, croissante, continue à droite. On peut sans restreindre la généralité supposer que  $f$  prend ses valeurs dans  $[0, 1]$ .

1. Pour  $n \geq 1$ , montrer que l'ensemble  $\{x : f(x) \geq f(x_-) + 1/n\}$  est de cardinal fini.
2. En déduire que l'ensemble des points de discontinuité de  $f$  est au plus dénombrable.

## Pour s'entraîner

EXERCICE 4.6.—

Montrer qu'une fonction mesurable de  $(E, \{\emptyset, E\})$  dans  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  est constante. Caractériser les fonctions mesurables de  $(E, \{\emptyset, A, A^c, E\})$  dans  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$  où  $A$  est un sous-ensemble propre de  $E$ .

EXERCICE 4.7.—

On veut montrer que toute mesure de Radon sur  $\mathbf{R}$  (c'est-à-dire  $\mu(K) < +\infty$  quel que soit le compact  $K$ ) invariante par translation est proportionnelle à la mesure de Lebesgue.

1. Montrer que pour tout entier  $\mu([n; n+1]) = \mu([0; 1])$ .
2. En déduire que  $\mu([0; n]) = n \mu([0; 1])$ .
3. Montrer que pour tout entier  $p, q$  positifs,

$$q \mu([0; p/q]) = p \mu([0; 1]).$$

4. En déduire que

$$\mu([0; r]) = r \mu([0; 1])$$

pour  $r$  rationnel.

5. En déduire que En déduire que

$$\mu([0; x]) = x \mu([0; 1])$$

pour tout  $x$  réel.

6. Conclure.

EXERCICE 4.8.—

Soit  $(E, \mathcal{E}, \mu)$  un espace mesurable et  $T$  une application de  $E$  dans lui-même. On dit que  $\mu$  est invariante par  $T$  si

$$\int_E f \circ T d\mu = \int f d\mu$$

pour toute fonction  $f$  mesurable bornée.

1. Montrer que la mesure de Lebesgue sur  $\mathbf{R}$  est invariante par translation.
2. Soit  $E = \mathbf{R}^n$  et

$$d\mu(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(x_1^2 + \dots + x_n^2)\right) dx_1 \dots dx_n.$$

Montrer que  $\mu$  est invariante par rotation.

3. Soit  $E = [0, 1]$  et  $T(x) = 2x - [2x]$  ( $T(x)$  est la partie fractionnaire de  $x$ ). Montrer que la mesure de Lebesgue restreinte à  $E$  est invariante.

EXERCICE 4.9 (\*\*).—

On veut montrer que les hypothèses du théorème de Fubini sont optimales. On considère la fonction définie sur  $[0, 1] \times [0, 1]$  par  $(x^2 - y^2)/(x^2 + y^2)^2$ .

1. Montrer que si le théorème de Fubini est applicable alors nécessairement cette intégrale est nulle.
2. En faisant les calculs explicites des intégrales partielles, montrer que l'on n'obtient pas le même résultat.
3. Montrer enfin, en utilisant Fubini-Tonelli, que la fonction étudiée n'est pas de module intégrable.

## Pour aller plus loin

EXERCICE 4.10. –

Soit  $\mu_1$  et  $\mu_2$  deux mesures absolument continues par rapport à une même mesure  $\mu$ . A quelle condition (suffisante) sur leurs dérivées de Radon-Nikodym,  $\mu_2$  est-elle absolument continue par rapport à  $\mu_1$ ? Que vaut alors  $d\mu_2/d\mu_1$ ?

EXERCICE 4.11. –

Quelles sont les mesures sur  $\mathbf{R}$  absolument continues par rapport à  $\delta_1$ ?

EXERCICE 4.12 (\*\*\*, Ensemble triadique de Cantor). –

L'objectif est de construire un ensemble non dénombrable de mesure de Lebesgue nulle. Soit  $S$  les éléments de  $\{0, 1, 2\}^{\mathbf{N}}$  qui ne se terminent pas par une infinité de 2.

1. Montrer que tout nombre  $x$  de  $[0, 1[$  s'écrit de manière unique sous la forme

$$x = \sum_{n=1}^{+\infty} x_n 3^{-n} \text{ où } (x_n, n \geq 1) \in S.$$

On appelle la suite  $(x_n, n \geq 1)$  le développement triadique de  $x$ .

2. On appelle  $C$ , l'ensemble de Cantor, constitué des réels de  $[0, 1[$  qui n'ont pas de 1 dans leur développement triadique. Montrer que  $C = \bigcap_{n=1}^{\infty} E_n$  où les  $E_n$  sont des ensembles que l'on construira (voir la figure 4.4).
3. Montrer que la mesure de Lebesgue de  $C$  est nulle.
4. Montrer que  $C$  est non dénombrable.
5. Montrer que  $C^c$  est partout dans  $[0, 1[$  : quel que soit  $\epsilon > 0$ , pour tout  $x \in [0, 1[$ , il existe  $y \in C^c$  tel que  $|x - y| < \epsilon$ .
6. En déduire que l'intérieur de  $C$  est vide.

EXERCICE 4.13 (\*\*\*, Fonction de Cantor). –

À partir de l'ensemble de Cantor, on va maintenant construire une fonction continue, croissante, nulle en 0, qui vaut 1 en 1 et dont la dérivée est presque-partout nulle...

- La fonction  $f_0$  est définie par  $f_0(x) = x$ .
- La fonction  $f_1$  est continue, affine par morceaux, est telle que  $f_1(0) = 0$ ,  $f_1(1) = 1$



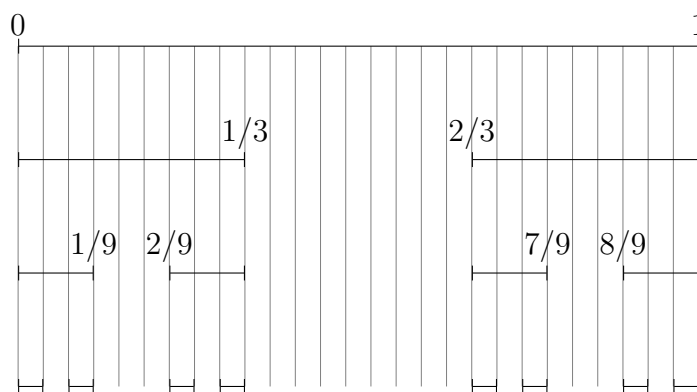


Figure 4.4. – Les premières étapes de la construction de l'ensemble de Cantor.

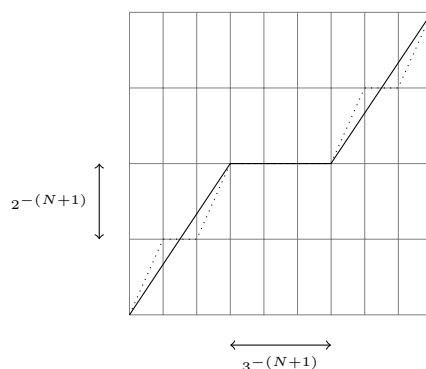


Figure 4.5. – Vue partielle de deux étapes successives dans la construction de la fonction de Cantor.

et vaut  $1/2$  sur  $E_1^c$  donc

$$f_1(x) = \begin{cases} \frac{3}{2}x & \text{pour } x \leq \frac{1}{3} \\ \frac{1}{2} & \text{pour } \frac{1}{3} \leq x \leq \frac{2}{3} \\ \frac{1}{2} + \frac{3}{2}\left(x - \frac{2}{3}\right) & \text{pour } x \geq \frac{2}{3}. \end{cases}$$

— Au rang  $n$ ,  $f_n$  est continue, affine par morceaux, égale à  $j2^{-n}$  sur le  $j$ -ième intervalle de  $E_n^c$  et telle que  $f_n(0) = 0$  et  $f_n(1) = 1$ .

1. Montrer que  $\|f_n - f_{n+1}\|_\infty \leq 2^{-(n+1)}$ .
2. En déduire que la suite  $(f_n, n \geq 1)$  est de Cauchy dans l'ensemble des fonctions continues muni de la norme uniforme. Soit  $f$  sa limite.
3. Montrer que  $f$  est croissante, vaut 0 en 0 et 1 en 1, est dérivable et de dérivée nulle sur  $C^c$ .

EXERCICE 4.14 (\*\*\*, Construction d'un ensemble non-mesurable).—

Soit  $E = [0, 1]$  muni de la mesure de Lebesgue notée  $\mu$ . Pour  $A \subset E \setminus \{1\}$  et  $x \in \mathbf{R}$ , on pose

$$\tau_x(A) = \{t + x - [t + x], t \in A\}$$

où  $[a]$  est la partie entière de  $a$ .

1. Montrer que si  $A$  est mesurable alors  $\tau_x(A)$  l'est aussi et  $\mu(\tau_x(A)) = \mu(A)$ .
2. Soit  $\mathcal{R}$  la relation d'équivalence définie par  $x\mathcal{R}y$  ssi  $x - y \in \mathbf{Q}$ . On construit  $F$  en choisissant un et un seul représentant de chaque classe d'équivalence. Montrer que les  $(\tau_r(F), r \in [0, 1] \cap \mathbf{Q})$  forment une partition de  $[0, 1]$  et en déduire que  $F$  n'est pas mesurable.

EXERCICE 4.15.—

La suite récurrente  $x = (x_n, n \geq 0)$  définie par

$$x_0 \in ]0, 1[, x_{n+1} = 4x_n(1 - x_n),$$

appartient à la classe des *systèmes dynamiques* dits chaotiques. En effet, si l'on part d'une condition initiale quelconque, l'orbite (la suite des valeurs prises par la suite pour une condition initiale donnée) présente de façon évidente un caractère hautement erratique.

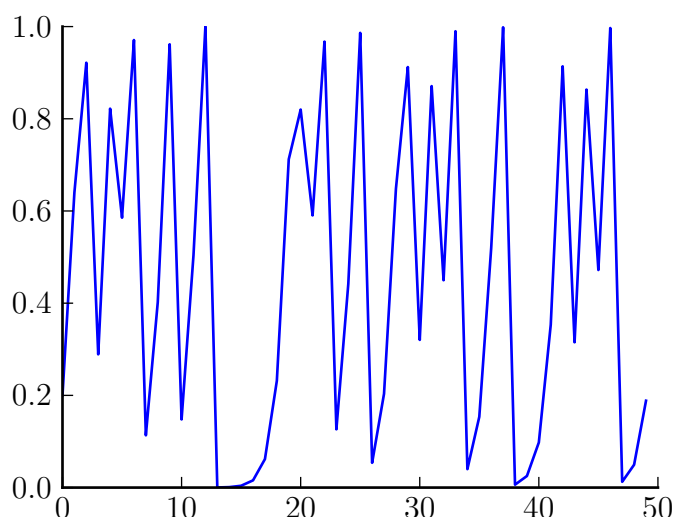


Figure 4.6. — L'orbite de la suite  $x$  pour  $x_0 = 0,2$ .

Qui plus est, pour des conditions initiales très proches, les orbites sont proches au début mais finissent par se séparer franchement au bout de quelques itérations.

Ce comportement est en partie dû au fait qu'à l'intersection de la première bissectrice et de la courbe représentative de  $f(x) = 4x(1 - x)$ , la dérivée est en module supérieure à 1 donc le potentiel point fixe  $x = 3/4$  est de type répulsif :  $f'(3/4) = -2$ .

On ne peut donc rien dire de déterministe sur le comportement asymptotique de cette suite. En revanche, on peut essayer de voir ce que sont les « zones » de  $[0, 1]$  qu'elle visite

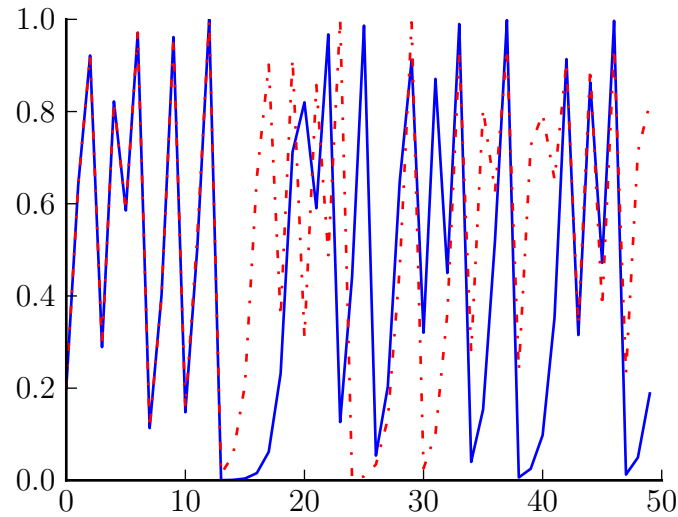


Figure 4.7. – L'orbite de la suite  $x$  pour  $x_0 = 0,2$ , comparée à l'orbite pour  $x_0 = 0,2001$ .

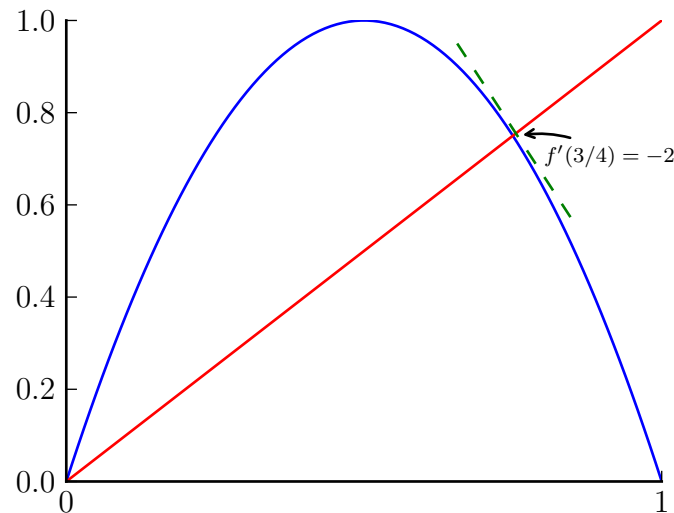


Figure 4.8. – Le point  $(3/4, 3/4)$  est répulsif.

le plus fréquemment, c'est-à-dire de regarder

$$\hat{\mu}([a, b]) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{[a, b]}(x_n)$$

pour tous les intervalles  $[a, b]$ . Cette quantité correspond à la fréquence empirique des passages dans l'intervalle  $[a, b]$ . Si l'on fait l'analogie avec les temps de séjour dans un état d'une chaîne de Markov,  $\hat{\mu}$  doit se comporter comme une mesure sur  $[0, 1]$ . L'objectif de cet exercice est de montrer que c'est bien le cas et d'identifier au passage la dite mesure. Comme les techniques mises en place relèvent de ce que l'on appelle la *théorie ergodique*, les premières questions sont génériques et s'appliquent à de nombreux autres systèmes dynamiques.

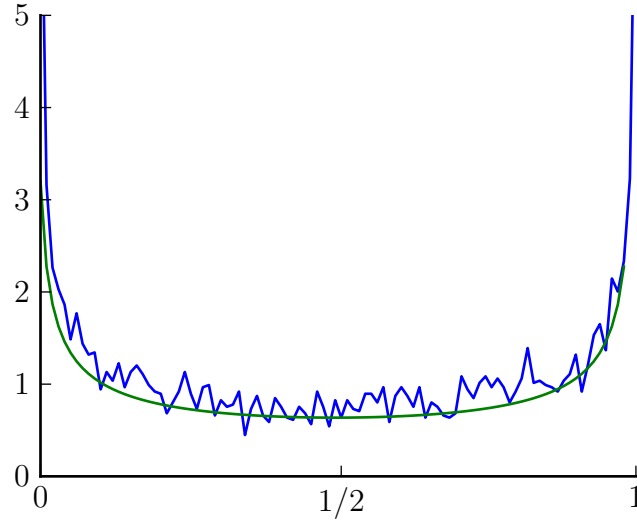


Figure 4.9. —  $\hat{\mu}$  (courbe verte, obtenue pour  $N = 50\,000$ ) et  $\mu$  (courbe rouge, calculée dans la question 4.15.i).

Soit  $(E, \mathcal{E}, \mathbf{P})$  un espace probabilisé et  $T$  une application mesurable de  $E$  dans lui-même. On suppose que  $\mathbf{P}$  est invariante par  $T$ , c'est-à-dire que

$$\mathbf{P}(T^{-1}(A)) = \mathbf{P}(A) \text{ pour tout } A \in \mathcal{E}.$$

- 4.15.a) Montrer que l'ensemble des mesurables invariants par  $T$ , c'est-à-dire qui vérifie  $T^{-1}(A) = A$ , est une tribu (notée  $\mathcal{I}$  par la suite).
- 4.15.b) Soit  $f$  une fonction mesurable de  $E$  dans  $\mathbf{R}$ . Montrer que si  $f$  est invariante par  $T$  (c'est-à-dire  $f \circ T = f$ ) alors  $f$  est mesurable de  $(E, \mathcal{I})$  dans  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ .

Le système dynamique  $(E, T, \mathbf{P})$  est dit ergodique lorsque

$$\mathcal{I} \subset \sigma\{A \in \mathcal{E}, \mathbf{P}(A) = 0 \text{ ou } \mathbf{P}(A) = 1\}.$$

- 4.15.c) Montrer que  $(E, T, \mathbf{P})$  est ergodique si et seulement si les fonctions invariantes par  $T$  sont constantes presque partout.

On dit que  $T$  est mélangeante si et seulement si pour tout couple  $f, g$  d'éléments de  $L^2(\mathbf{P})$ ,

$$\lim_{n \rightarrow +\infty} \int_E f \circ T^n g \, d\mathbf{P} = \int_E f \, d\mathbf{P} \int_E g \, d\mathbf{P}. \quad (4.10)$$

- 4.15.d) Montrer que si  $T$  est mélangeante alors  $(E, T)$  est ergodique.
- 4.15.e) Montrer pour  $f \in L^2(\mathbf{P})$ , pour tout entier  $n \geq 1$ ,

$$\int_E f \circ T^n \, d\mathbf{P} = \int_E f \, d\mathbf{P}. \quad (4.11)$$

- 4.15.f) Montrer que si la condition de mélange (4.10) est vérifiée pour  $f, g$  appartenant à un sous-ensemble dense de  $L^2(\mathbf{P})$  alors  $T$  est mélangeante.

On veut maintenant étudier le système dynamique donnée par l'équation d'évolution :

$$x_{n+1}^a = T(x_n^a) \text{ où } T(x) = 4x(1-x), \quad x_0^a = a \in [0, 1].$$

On veut montrer en particulier que pour presque tout  $a \in [0, 1]$ ,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^n f(x_j^a) = \int_0^1 f(u) (\pi \sqrt{u} \sqrt{1-u})^{-1} du.$$

On admet *le théorème de Birkhoff* qui stipule que si  $(E, T, \mathbf{P})$  est un système ergodique alors

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^n f \circ T^j(x) = \int_E f d\mathbf{P} \quad (4.12)$$

pour presque tout  $x$ . Il nous faut donc trouver une mesure invariante  $\mu$  par  $T$  et montrer que le système dynamique  $([0, 1], T, \mu)$  est ergodique. Pour ce faire on considère un autre système dynamique :

$$E_1 = [0, 1[, \quad T_1 x = 2x \text{ si } 0 \leq x \leq 1/2, \quad T_1(x) = 2 - 2x \text{ pour } 1/2 \leq x < 1.$$

(où  $[x]$  est la partie entière de  $x$ ) muni de la mesure de Lebesgue sur  $[0, 1[$ , notée  $\lambda$ .

4.15.g) Montrer que  $\lambda$  est invariante par  $T_1$ .

4.15.h) En admettant (ou se souvenant, cf. séries de Fourier) que la famille de fonctions  $e_k(x) = \exp(2i\pi kx)$  pour  $k \in \mathbf{Z}$  est une famille dense de  $L^2(d\lambda)$ , montrer que  $T_1$  est mélangeante.

Soit  $\Theta$  l'application de  $E_1$  dans  $[0, 1]$  définie par :

$$\Theta(x) = \sin^2(\pi x/2).$$

4.15.i) Identifier  $\mu$  la mesure image de  $\lambda$  par  $\Theta$ .

4.15.j) Montrer que  $([0, 1[, T, \mu)$  est ergodique et conclure.



# 5. Variables et vecteurs aléatoires réels

## 5.1. Généralités

### 5.1.1. Loi d'une variable aléatoire

On se place sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbf{P})$ . Soit  $(E, \mathcal{E})$  un espace mesurable.

**Définition 5.1.** Une *variable aléatoire*  $X$  sur  $E$  est une application mesurable de  $\Omega$  dans  $E$ .

Grâce à la notion de mesurabilité, nous assurons que les événements du type «  $X$  appartient à  $H$  » sont bien des événements de la tribu  $\mathcal{F}$ , c'est-à-dire des sous-ensembles de  $\Omega$  dont nous pouvons évaluer la probabilité. La théorie de la mesure fournit le cadre mathématique nécessaire à la construction d'une théorie complète des probabilités (voir également la table 5.1).

Lorsque l'ensemble d'arrivée est une partie de  $\mathbf{R}$  ou de  $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty, +\infty\}$ , on parle de *variable aléatoire réelle* (en abrégé, v.a.r.). Lorsque l'ensemble d'arrivée est une partie de  $\mathbf{R}^d$  avec  $d \geq 2$ , on parle de *vecteur aléatoire* ou de variable aléatoire *multivariée*. Une quantité scalaire ou vectorielle  $a$  constante par rapport à  $\omega$  est parfois qualifiée de *déterministe*.

**Définition 5.2.** On appelle *loi de la v.a.  $X$*  la fonction  $\mathbf{P}_X$  définie par :

$$\begin{aligned} \mathbf{P}_X : \mathcal{E} &\rightarrow [0, 1] \\ H &\mapsto \mathbf{P}(X^{-1}(H)) . \end{aligned}$$

En écriture plus compacte,  $\mathbf{P}_X := \mathbf{P} \circ X^{-1}$ . Une autre manière d'écrire cette définition est :

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

Autrement dit,  $\mathbf{P}_X(H)$  est la probabilité pour que  $X$  appartienne à  $H$ .

**Proposition 5.1.**  $\mathbf{P}_X$  est une mesure de probabilité sur  $(E, \mathcal{E})$ .

*Démonstration.* On vérifie les axiomes i) et ii) que doit satisfaire une mesure de probabilité.  
i)  $\mathbf{P}_X(E) = \mathbf{P}(X^{-1}(E)) = \mathbf{P}(\Omega) = 1$  et  $\mathbf{P}_X(\emptyset) = \mathbf{P}(X^{-1}(\emptyset)) = \mathbf{P}(\emptyset) = 0$ .

ii) Soit  $(H_n)_{n \in \mathbb{N}^*}$  une famille d'éléments de  $\mathcal{E}$  deux à deux disjoints. On a  $X^{-1}(\bigcup_n H_n) = \bigcup_n X^{-1}(H_n)$  et on montre aisément que les événements  $(X^{-1}(H_n))$  sont deux à deux disjoints. Ainsi en appliquant  $\mathbf{P}$  aux deux membres de l'égalité précédente, on obtient  $\mathbf{P}_X(\bigcup_n H_n) = \mathbf{P}(\bigcup_n X^{-1}(H_n)) = \sum_n \mathbf{P}(X^{-1}(H_n))$ .  $\square$

### 5.1.2. Discussion

En pratique, on s'intéresse le plus souvent à des probabilités de la forme  $\mathbf{P}(X \in H) = P_X(H)$  où  $X$  est une v.a. et  $H$  un ensemble. Ainsi, on manipule la loi  $\mathbf{P}_X = \mathbf{P} \circ X^{-1}$  bien plus souvent que la probabilité  $\mathbf{P}$  elle-même. Or  $\mathbf{P}_X$  est une probabilité sur l'espace où  $X$  prend ses valeurs. De ce fait, dans les problèmes que nous rencontrerons, l'espace de probabilité  $(\Omega, \mathcal{F}, \mathbf{P})$  n'est souvent pas spécifié. Il s'agit d'un espace abstrait, suffisamment riche pour modéliser le problème d'intérêt, mais sans nécessairement de signification « physique » en rapport avec l'expérience décrite.

Par exemple, nous rencontrerons fréquemment des énoncés qui débutent par une phrase du type : « Soit  $X$  une variable aléatoire de Bernoulli de paramètre  $p$  sur  $\{0, 1\}$  ». Un tel énoncé suppose implicitement la donnée d'un espace de probabilité  $(\Omega, \mathcal{F}, \mathbf{P})$  tel que  $X$  est une v.a. sur cet espace et tel que  $P_X$  est une mesure de probabilité de Bernoulli, c'est-à-dire  $P_X(\{1\}) = p$ ,  $P_X(\{0\}) = 1 - p$ . Toutefois, cet énoncé ne précise ni la nature de  $(\Omega, \mathcal{F}, \mathbf{P})$ , ni l'expression de  $X(\omega)$  en fonction de  $\omega$  : cela est sans importance du moment que  $\mathbf{P}_X$  est la loi voulue. Spécifier  $\Omega$  n'est d'aucune utilité. Si nous tenions malgré tout à le faire, nous aurions maintes possibilités. Dans l'exemple précédent, nous pourrions naturellement poser  $\Omega = \{0, 1\}$ ,  $\mathbf{P}$  la probabilité de Bernoulli et  $X(\omega) = \omega$ . Mais nous pourrions tout aussi bien choisir pour  $\Omega$  l'intervalle  $[0, 1]$ , pour  $\mathbf{P}$  la mesure de Lebesgue sur  $[0, 1]$  et poser  $X(\omega) = \mathbf{1}_{[0, p]}(\omega)$ . Dans les deux cas, on pourra vérifier que  $X$  est bien une v.a. de Bernoulli de paramètre  $p$  : peu importe donc la solution choisie.

## 5.2. Variables aléatoires réelles

Dans ce paragraphe, on traite le cas de variables aléatoires  $X$  à valeurs dans  $E = \mathbf{R}$  muni de la tribu de Borel.

### 5.2.1. Fonction de répartition

Comme  $\mathbf{P}_X$  est une mesure sur  $\mathbf{R}$ , on sait qu'elle est totalement caractérisée (voir le chapitre 4) par les valeurs de  $\mathbf{P}_X(-\infty, b]$  pour  $b$  parcourant  $\mathbf{R}$ .



**Définition 5.3.** Soit  $X$  une v.a.r., la fonction

$$\begin{aligned} F_X : \mathbf{R} &\longrightarrow [0, 1] \\ x &\longmapsto \mathbf{P}_X([-\infty, x]) = \mathbf{P}(X \leq x) \end{aligned}$$

s'appelle la fonction de répartition de  $X$ .

En vertu des propriétés de monotonie des mesures,  $F_X$  possède les propriétés suivantes, que l'on pourra vérifier à titre d'exercice :

- $\lim_{x \rightarrow -\infty} F_X(x) = 0$ ,
- $\lim_{x \rightarrow +\infty} F_X(x) = 1$ ,
- $F_X$  est croissante, continue à droite, i.e.,  $\lim_{y \downarrow x} F_X(y) = F_X(x)$ .

On a, d'après les propriétés de monotonie des mesures,

$$F_X(x_-) = \lim_{n \rightarrow +\infty} \mathbf{P}\left(\bigcup_{n=1}^{\infty} ]-\infty, x - \frac{1}{n}]\right) = \mathbf{P}_X([-\infty, x[).$$

Par conséquent,  $F_X(x_-) = \mathbf{P}_X([-\infty, x[)$  et donc

$$F_X(x) - F_X(x_-) = \mathbf{P}(X = x).$$

En d'autres termes, si  $F_X$  est continue en  $x$ ,  $\mathbf{P}(X = x) = 0$ . Comme  $F_X$  est bornée, le nombre de ces points de discontinuité est au plus dénombrable (voir exercice 4.5). Soit  $\{x_n, n \in \mathbf{N}^*\}$  ces points. On peut alors parler de  $F_X^c$ , la régularisée de  $F_X$  :

$$\begin{aligned} F_X^c(x) &= F_X(x) - \sum_{n=1}^{\infty} \left( F_X(x_n) - F_X(x_{n-}) \right) \mathbf{1}_{[x_n, +\infty[}(x). \\ &= F_X(x) - \sum_{n=1}^{\infty} \Delta F_X(x) \mathbf{1}_{[x_n, +\infty[}(x). \end{aligned}$$

La fonction  $F_X^c$  est continue et croissante par définition. Elle est d'après un théorème de Lebesgue, dérivable sauf sur un ensemble de mesure de Lebesgue nulle. Dans la suite, nous ne nous préoccupons pas de savoir ce qui se passe si elle n'est pas dérivable en tout point.

**Théorème 5.2.** Soit  $X$  une v.a.r. de fonction de répartition  $F_X$ . Si  $F_X^c$  est dérivable sur  $\mathbf{R}$ , alors

$$d\mathbf{P}_X(x) = \sum_{n=1}^{\infty} \Delta F_X(x_n) d\delta_{x_n} + (F_X^c)'(x) dx. \quad (5.1)$$

Si  $F_X$  est continue alors

$$d\mathbf{P}_X(x) = (F_X^c)'(x) dx$$

et  $(F_X^c)'$  s'appelle la densité de la loi de  $X$ .

*Démonstration.* Remarquons que

$$\mathbf{1}_{[x_n, +\infty[}(x) = \delta_{x_n}([-\infty, x]),$$

puisque le terme de gauche ne vaut 1 que si  $x \geq x_n$ , soit de manière équivalente  $x_n \in ]-\infty, x]$ . Partant de l'écriture,

$$F_X(x) = F_X^c(x) + \sum_{n=1}^{\infty} \Delta F_X(x) \mathbf{1}_{[x_n, +\infty[}(x),$$

si  $F_X^c$  est dérivable en tout point alors on a

$$\mathbf{P}(X \in ]-\infty, x]) = \int_{-\infty}^x (F_X^c)'(s) ds + \sum_{n=1}^{\infty} \Delta F_X(x_n) \delta_{x_n}([-\infty, x]).$$

Les deux mesures de part et d'autre de l'égalité (5.1) coïncident donc sur les ensembles de la forme  $] - \infty, x]$  pour tout  $x$  réel. C'est suffisant (cf. thm. 4.19) pour assurer que ces deux mesures sont égales.  $\square$

### 5.2.2. Variables aléatoires réelles à densité

Lorsque  $F_X$  est dérivable donc a fortiori continue,  $F_X = F_X^c$  et

$$d\mathbf{P}_X(x) = F_X'(x)dx.$$

On dit que  $\mathbf{P}_X$  admet  $F_X'$  pour densité, on dit de manière raccourcie que  $X$  est de densité  $F_X'$ . Réciproquement, si l'on se donne une v.a.r. de densité  $f$ , i.e. si  $f$  est supposée satisfaire la relation

$$\mathbf{P}(X \in H) = \int_H f.$$

En posant  $H = ]-\infty, x]$ , on en déduit que la fonction de répartition de  $X$  est l'intégrale de la densité :

$$F_X(x) = \int_{-\infty}^x f. \quad (5.2)$$

Notons bien que la condition «  $F_X$  est de classe  $C^1$  » est suffisante, mais pas nécessaire pour que la loi de  $X$  soit à densité. D'ailleurs, les fonctions  $F_X$  rencontrées en pratique ne sont pas toujours de classe  $C^1$  et peuvent néanmoins admettre une densité. Le paragraphe suivant fournit une condition nécessaire et suffisante.

#### Conditions d'existence d'une densité\*

Une fonction  $F$  est dite *absolument continue* si pour tout  $\epsilon > 0$ , il existe  $\delta > 0$  tel que pour toute suite finie d'intervalles disjoints  $[a_1, b_1], \dots, [a_p, b_p]$ ,

$$\sum_{k=1}^p (b_k - a_k) < \delta \Rightarrow \sum_{k=1}^p |F(b_k) - F(a_k)| < \epsilon.$$

Le résultat suivant est admis (voir [Bil95, Théorèmes 31.7 et 31.8]).

**Théorème 5.3.** Les trois propositions suivantes sont équivalentes :

- i)  $X$  admet une densité ;
- ii)  $F_X$  est absolument continue ;
- iii)  $P_X(A) = 0$  pour tout ensemble négligeable  $A \in \mathcal{F}$ .

### 5.2.3. Retour sur les variables discrètes

Soit  $X : \Omega \rightarrow \mathbf{R}$  une variable aléatoire prenant ses valeurs dans un ensemble au plus dénombrable, disons par exemple :

$$X(\Omega) =: \{x_1, x_2, x_3, \dots\} \subset \mathbf{R} .$$

Bien que l'espace d'arrivée  $\mathbf{R}$  ne soit pas discret,  $X(\Omega)$  l'est, et on ne perdrait guère à restreindre le domaine d'arrivée de  $X$  à  $\{x_1, x_2, \dots\}$  plutôt que  $\mathbf{R}$ . C'est pourquoi nous utiliserons le terme de *variable aléatoire discrète* pour désigner  $X$ .

Le chapitre 2 suffit donc à étudier ce type de v.a. sans qu'il soit besoin d'avoir recours aux notions nouvelles de théorie de la mesure que nous venons d'introduire. Nul besoin par exemple de donner la fonction de répartition pour caractériser la loi de  $X$  : nous savons déjà que la donnée des seuls coefficients  $\mathbf{P}(X = x_n)$  suffit.

Toutefois, il est intéressant, à titre d'exercice, de voir comment le formalisme général présenté dans ce chapitre permet de couvrir le traitement des variables discrètes, et de comprendre ce que deviennent les notions de loi et de fonction de répartition dans ce cas particulier.

**Proposition 5.4.** La loi  $\mathbf{P}_X$  est donnée par :

$$\mathbf{P}_X = \sum_{n=1}^{\infty} \mathbf{P}(X = x_n) \delta_{x_n} .$$

*Démonstration.*  $\mathbf{P}_X(H) = \mathbf{P}_X(H \cap X(\Omega)) = \mathbf{P}_X(\bigcup_n H \cap \{x_n\}) = \sum_n \mathbf{P}_X(H \cap \{x_n\})$ . Or  $\mathbf{P}_X(H \cap \{x_n\})$  est égal à  $\mathbf{P}_X(x_n)$  si  $x_n \in H$ , à 0 sinon. Donc  $\mathbf{P}_X(H \cap \{x_n\}) = \mathbf{P}_X(x_n) \delta_{x_n}(H)$ .  $\square$

La fonction de répartition est donnée par :

$$F_X(x) = \sum_{n=1}^{\infty} \mathbf{P}(X = x_n) \mathbf{1}_{[x_n, +\infty[}(x) .$$

Il s'agit donc d'une fonction en escalier, dont l'amplitude des sauts est donnée par les coefficients  $\mathbf{P}(X = x_n)$ . On note que  $X$  n'admet pas de densité puisque  $F_X$  est discontinue.

Dans le cas particulier où  $X$  est une v.a. constante, disons  $X(\omega) = a$  pour tout  $\omega$ , la loi de  $X$  coïncide avec un Dirac au point  $a$ . Une telle loi est dite *dégénérée*. La fonction de répartition associée est un échelon :  $F_X = \mathbf{1}_{[a, +\infty[}$ .

### 5.2.4. Espérance et moments

Au chapitre 2, nous avons défini l'espérance  $\mathbf{E}[X]$  d'une v.a.r. discrète  $X \in \{x_1, x_2, \dots\}$  comme le barycentre des  $x_n$  pondérés par la « masse »  $\mathbf{P}(X = x_n)$  :

$$\sum_n x_n \mathbf{P}(X = x_n) .$$

Cette définition est très spécifique au cas discret, et il nous faut maintenant la généraliser. Par exemple, si  $X$  est une v.a.r. de densité  $f_X$ , la notion précédente de barycentre devient :

$$\int_{\mathbf{R}} x f_X(x) dx ,$$

et on pourrait ainsi définir l'espérance d'une v.a. à densité, en remplaçant la somme par une intégrale, et la loi discrète par la densité. Cette seconde définition resterait elle aussi très spécifique au cas des variables à densité. Le chapitre 4 permet de fournir une définition générale de l'espérance qui admet les deux exemples ci-dessus comme cas particuliers.

Soit  $(\Omega, \mathcal{F}, \mathbf{P})$  un espace de probabilité et  $X : \Omega \rightarrow \overline{\mathbf{R}}$  une variable aléatoire.

**Définition 5.4.** L'espérance de la v.a.  $X$  est définie par :

$$\mathbf{E}[X] := \int X d\mathbf{P} .$$

Il s'agit donc de l'intégrale de  $X$ , vue comme fonction sur  $\Omega$ , par rapport à la mesure de probabilité  $\mathbf{P}$ .

REMARQUE.— Il y a un cas particulier d'une importance fondamentale à cette définition. Si  $X$  est une variable aléatoire à valeurs dans un espace quelconque et que l'on considère la variable aléatoire réelle

$$Y = \mathbf{1}_A(X) = \begin{cases} 1 & \text{si } X \in A \\ 0 & \text{sinon.} \end{cases}$$

D'après la construction de l'intégrale par rapport à une mesure quelconque (cf. définition ??),

$$\mathbf{P}(X \in A) = \mathbf{E}[\mathbf{1}_A(X)] . \quad (5.3)$$

Terminologie « analyse »		Terminologie « probabilités »	
Fonction mesurable	$f$	Variable aléatoire	$X$
Mesure	$\mu$	Mesure	$\mathbf{P}$
Intégrale	$\int f(x)d\mu(x)$	Espérance	$\mathbf{E}[X] = \int X(\omega)d\mathbf{P}(\omega)$

Table 5.1. – Correspondance des notations et de la terminologie « théorie de la mesure et de l'intégration » vs « théorie des probabilités ».

L'espérance est toujours bien définie pour  $X \geq 0$ . Dans le cas général, elle est bien définie lorsque  $\mathbf{E}[X^+]$  ou  $\mathbf{E}[X^-]$  sont finies. La v.a.  $X$  est intégrable lorsque  $\mathbf{E}[|X|] < \infty$ .

Un cas particulier important est obtenu en posant  $X(\omega) = \mathbf{1}_A(\omega)$  où  $A \in \mathcal{F}$ . Comme  $\mathbf{1}_A$  est une fonction simple, son intégrale de Lebesgue par rapport à  $\mathbf{P}$  est immédiatement donnée par :

$$\mathbf{P}(A) = \mathbf{E}[\mathbf{1}_A] . \quad (5.4)$$

Au sens de l'égalité précédente, l'espérance peut être interprétée comme une extension de la notion de mesure de probabilité.

### Théorème de transfert

Dans la pratique, nous avons généralement accès à la loi  $\mathbf{P}_X$  et non à la loi  $\mathbf{P}$  ni à l'expression de  $X(\omega)$  en fonction de  $\omega$  : le théorème de transfert permet d'exprimer  $\mathbf{E}[X]$  en fonction de  $\mathbf{P}_X$ . Il permet en outre d'exprimer l'espérance d'une variable aléatoire  $g(X)$  non pas en fonction de la loi  $\mathbf{P}_{g(X)}$  qui n'est généralement pas disponible directement, mais en fonction de la loi  $\mathbf{P}_X$ . On a donc le diagramme suivant.

$$\begin{array}{ccc} (\Omega, \mathbf{P}) & \xrightarrow{X} & (E, \mathbf{P}_X) \\ & \searrow g \circ X & \downarrow g \\ & & \mathbf{R} \end{array}$$

Soit  $(E, \mathcal{E})$  un espace mesurable.

**Théorème 5.5.** Soit  $X : \Omega \rightarrow E$  et  $g : E \rightarrow \mathbf{R}$  deux fonctions mesurables telles que  $\mathbf{E}[g(X)]$  est définie. Alors,

$$\mathbf{E}[g(X)] = \int g(x)d\mathbf{P}_X(x) .$$

*Démonstration.* On donne d'abord la preuve dans le cas où  $g$  est une fonction simple positive, de la forme  $g = \sum_k \alpha_k \mathbf{1}_{A_k}$ . Alors  $\int g d\mathbf{P}_X = \sum_k \alpha_k \mathbf{P}_X(A_k)$ . Or par définition,

$\mathbf{E}[g(X)] = \int (g \circ X) d\mathbf{P}$ . Comme  $g \circ X(\omega) = \sum_k \alpha_k \mathbf{1}_{X^{-1}(A_k)}(\omega)$ , on obtient :  $\mathbf{E}[g(X)] = \sum_k \alpha_k \mathbf{P}(X^{-1}(A_k))$  et l'égalité est donc démontrée pour les fonctions simples.

Donnons maintenant la preuve pour  $g$  fonction positive quelconque. D'après le lemme 4.26, il existe une suite de fonctions simples  $g_n \uparrow g$ . Par le théorème de convergence monotone,  $\int g_n dP_x \rightarrow \int g dP_x$ . Or d'après la preuve précédente,  $\int g_n dP_x = \mathbf{E}[g_n(X)] = \int (g_n \circ X) d\mathbf{P}$ . On montre facilement que  $g_n \circ X \uparrow g \circ X$ , donc, toujours d'après le théorème de convergence monotone,  $\int (g_n \circ X) d\mathbf{P} \uparrow \int (g \circ X) d\mathbf{P} = \mathbf{E}[g(X)]$ . On conclut que  $\mathbf{E}[g(X)] = \int g dP_x$ .

Le cas  $g$  quelconque se traite facilement en écrivant  $\mathbf{E}[g(X)] = \mathbf{E}[g(X)^+] - \mathbf{E}[g(X)^-]$  et en appliquant le résultat précédent aux fonctions  $g(X)^+$  et  $g(X)^-$  respectivement.  $\square$

EXEMPLE 29. – En particulier, notons bien la conséquence suivante :  $\mathbf{E}[X] = \int x dP_X(x)$ . Pour  $X$  une variable aléatoire de loi exponentielle de paramètre  $\lambda$ , i.e. de densité

$$f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0, +\infty[}(x),$$

la moyenne (ou espérance) est donnée par

$$\mathbf{E}[X] = \int_{\mathbf{R}} x f_X(x) dx = \int_0^\infty x \cdot \lambda e^{-\lambda x} dx.$$

On procède alors à une intégration par parties (en dérivant  $x \mapsto x$  et en intégrant  $f_X$ ) et il vient

$$\int_0^\infty x \cdot \lambda e^{-\lambda x} dx = [-x e^{-\lambda x}]_0^\infty - \int_0^\infty (-e^{-\lambda x}) dx.$$

Le terme tout intégré est nul par « croissance comparée » de la fonction linéaire et de la fonction exponentielle. Il reste

$$\int_0^\infty x \cdot \lambda e^{-\lambda x} dx = \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda}.$$

### 5.2.5. Cas des variables à densité

Soit  $X$  une v.a.r. de densité  $f_X$ . On rappelle que la loi d'une telle v.a. est donnée par  $\mathbf{P}_X(A) = \int_A f_X(x) dx$ .

**Théorème 5.6.** Dès que  $\mathbf{E}[g(X)]$  est bien définie, on a :

$$\mathbf{E}[g(X)] = \int g(x) f_X(x) dx. \quad (5.5)$$

*Démonstration.* Commençons par le cas où  $g$  est une fonction simple positive, disons  $g = \sum_k \alpha_k \mathbf{1}_{A_k}$ . D'après le théorème de transfert,  $\mathbf{E}[g(X)] = \sum_k \alpha_k \mathbf{P}_X(A_k)$  et comme  $\mathbf{P}_X(A_k) = \int \mathbf{1}_{A_k} f_X$ , nous avons bien  $\mathbf{E}[g(X)] = \int (\sum_k \alpha_k \mathbf{1}_{A_k}) f_X = \int g \cdot f_X$ .

Traisons maintenant le cas où  $g$  est une fonction positive quelconque. Soit  $0 \leq g_n \uparrow g$  une suite de fonctions simples. Le théorème de convergence monotone implique que

$\int g_n d\mathbf{P}_X \rightarrow \int g d\mathbf{P}_X = \mathbf{E}[g(X)]$ . Mais d'après la preuve précédente,  $\int g_n d\mathbf{P}_X = \int g_n f_X$  tend vers  $\int g \cdot f_X$  toujours d'après le théorème de convergence monotone. Cela conclut la preuve pour les fonctions  $g$  positives. Le cas  $g$  quelconque se traite en décomposant  $g = g^+ - g^-$  et en appliquant le résultat précédent à  $g^+$  et  $g^-$ .  $\square$

REMARQUE.— D'après (5.4) appliquée à  $\mathbf{P}_X$  au lieu de  $\mathbf{P}$ , on a  $\mathbf{P}_X(H) = \int_H d\mathbf{P}_X(x)$ . Dans le cas où  $\mathbf{P}_X$  est de densité  $f_X$ , on a en outre  $\mathbf{P}_X(H) = \int_H f_X(x)dx$ . Ainsi, pour écrire que  $\mathbf{P}_X$  est de densité  $f_X$ , on utilise souvent la notation symbolique «  $d\mathbf{P}_X(x) = f_X(x) dx$  ».

### 5.2.6. Inégalités

**Proposition 5.7** (Inégalité de Markov). Pour tout  $\epsilon > 0$ ,  $p \geq 1$ ,

$$\mathbf{P}(|X| > \epsilon) \leq \frac{\mathbf{E}[|X|^p]}{\epsilon^p}.$$

*Démonstration.* Voir paragraphe 2.3.4.  $\square$

**Proposition 5.8** (Inégalité de Hölder). Soient  $p, q \geq 0$  tels que  $\frac{1}{p} + \frac{1}{q} = 1$ . Alors,

$$\mathbf{E}[|XY|] \leq (\mathbf{E}[|X|^p])^{\frac{1}{p}} (\mathbf{E}[|Y|^q])^{\frac{1}{q}}.$$

Lorsque  $p = q = 2$ , l'inégalité de Hölder se ramène à l'inégalité de Cauchy-Schwarz :

$$\mathbf{E}[|XY|] \leq \sqrt{\mathbf{E}[X^2] \mathbf{E}[Y^2]}.$$

*Démonstration.* Il suffit de donner la preuve pour des v.a. positives. On utilise l'inégalité  $ab \leq a^p/p + b^q/q$  valable  $\forall a, b \geq 0$  (pour démontrer cette inégalité, poser  $(s, t) = (p \ln a, q \ln b)$ ,  $ab = \exp(\frac{s}{p} + \frac{t}{q}) \leq \frac{1}{p}e^s + \frac{1}{q}e^t$  par convexité de  $\exp$ , ce qui est bien l'inégalité voulue). En posant  $a = X/\mathbf{E}[X^p]$  et  $b = Y/\mathbf{E}[Y^q]$  et en passant à l'espérance, on tombe bien sur l'inégalité de Hölder après un calcul simple.  $\square$

**Proposition 5.9** (Inégalité de Jensen). Soit  $\varphi : \mathbf{R} \rightarrow \mathbf{R}$  une fonction convexe. Soit  $X$  une v.a.r. telle que  $\mathbf{E}[|X|] < \infty$  et  $\mathbf{E}[|\varphi(X)|] < \infty$ . Alors :

$$\varphi(\mathbf{E}[X]) \leq \mathbf{E}[\varphi(X)].$$

*Démonstration.* Rappelons que toute fonction convexe définie sur  $\mathbf{R}$  est continue et que de plus,  $\forall x \in \mathbf{R}$ ,  $\exists \alpha$ ,  $\forall t$ ,  $\varphi(t) \geq \varphi(x) + \alpha(t - x)$ . Cela signifie que le graphe est au dessus d'une droite qui touche le graphe au point  $x$ . Soit  $\alpha$  une constante telle que pour tout  $t$ ,  $\varphi(t) \geq \varphi(\mathbf{E}[X]) + \alpha(t - \mathbf{E}[X])$ . On intègre les deux membres de cette inégalité par rapport à la loi  $\mathbf{P}_X$ . Par monotonie de l'intégrale de fonctions  $\mathbf{P}_X$ -intégrables,  $\mathbf{E}[\varphi(X)] \geq \varphi(\mathbf{E}[X])$ .  $\square$

### 5.2.7. Moments, variance

**Définition 5.5.** Soit  $p \geq 0$ . Soit une v.a.r.  $X$  telle que  $\mathbf{E}[|X|^p] < \infty$ . La quantité  $\mathbf{E}[X^p]$  est appelée le *moment d'ordre  $p$*  de  $X$ .

On dit d'une telle variable qu'elle est d'*ordre  $p$* , ou qu'elle *possède un moment d'ordre  $p$* . L'ensemble de telles variables est noté  $\mathcal{L}^p(\mathbf{P})$ . Les propriétés des moments sont identiques à celles vue dans le cas discret. Nous les résumons ici sans preuves.

**Proposition 5.10.** Une variable d'ordre  $p$  possède tous ses moments d'ordre inférieur.

Notons que certaines v.a. possèdent tous leurs moments, c'est par exemple le cas des variables gaussiennes ou des variables à valeur dans un ensemble borné. À l'inverse, certaines v.a. n'admettent aucun moment (voir l'exercice 5.7).

**Définition 5.6.** La *variance* d'une v.a.r.  $X$  d'ordre 2 est définie par

$$\text{Var}(X) := \mathbf{E}[(X - \mathbf{E}[X])^2].$$

Son *écart-type* est la racine carrée de la variance, noté  $\sigma_X := \sqrt{\text{Var}(X)}$ .

**Définition 5.7.** Soient  $X$  et  $Y$  deux v.a.r. d'ordre 2. Leur *covariance* est définie par :

$$\text{Cov}(X, Y) := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])].$$

On utilise parfois le *coefficient de corrélation* défini par  $\rho_{X,Y} = \text{Cov}(X, Y)/(\sigma_X \sigma_Y)$ . Lorsque  $\text{Cov}(X, Y) = 0$ , on dit que  $X$  et  $Y$  sont *décorrélées*.

**Proposition 5.11.** Soient  $X$  et  $Y$  deux v.a.r. d'ordre 2 et  $(\alpha, \beta) \in \mathbf{R}^2$ . On a :

- a)  $\text{Var}(X) = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$  ;
- b)  $\text{Cov}(X, X) = \text{Var}(X)$  ;
- c)  $\text{Cov}(Y, X) = \text{Cov}(X, Y)$  ;
- d)  $\text{Var}(\alpha X + \beta) = \alpha^2 \text{Var}(X)$  ;
- e)  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$ . En particulier, si  $X$  et  $Y$  sont indépendantes alors  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ .

## 5.3. Vecteurs aléatoires

On se place sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbf{P})$ . On se donne une fonction mesurable  $X : \Omega \rightarrow \mathbf{R}^d$ , où  $d$  est un entier et  $\mathbf{R}^d$  est muni de sa tribu de Borel. Pour tout  $\omega \in \Omega$ , on notera  $X_1(\omega), \dots, X_d(\omega)$  les coordonnées du vecteur  $X(\omega)$  dans la base canonique de  $\mathbf{R}^d$ . Nous savons d'après le corollaire 4.9 que  $X$  est mesurable si et seulement si  $X_1, \dots, X_d$  le



sont. Se donner un vecteur aléatoire est donc équivalent à se donner une collection de  $d$  variables aléatoires réelles.

### 5.3.1. Fonction de répartition

On rappelle que la loi du vecteur aléatoire  $X$  est la mesure de probabilité définie pour tout  $H \in \mathcal{B}(\mathbf{R}^d)$  par  $\mathbf{P}_X(H) = \mathbf{P}(X \in H)$ , aussi appelée *loi jointe* des variables aléatoires  $X_1, \dots, X_d$ .

**Définition 5.8.** La *fonction de répartition* de  $X$  est l'application  $F_X : \mathbf{R}^d \rightarrow \mathbf{R}_+$  définie pour tout  $(x_1, \dots, x_d) \in \mathbf{R}^d$  par :

$$\begin{aligned} F_X(x_1, \dots, x_d) &= \mathbf{P}(X_1 \leq x_1, \dots, X_d \leq x_d) \\ &= \mathbf{P}_X \left( \prod_{k=1}^d ] - \infty, x_k ] \right). \end{aligned}$$

D'après la remarque qui suit le théorème 4.19, on sait que la fonction de répartition ainsi définie caractérise la loi du vecteur aléatoire  $X$ .

**Théorème 5.12.** Deux mesures de probabilité sur  $\mathcal{B}(\mathbf{R}^d)$  ayant même fonction de répartition sont égales.

### Vecteurs aléatoires à densité

Par définition,  $X$  admet une densité  $f_X : \mathbf{R}^d \rightarrow \mathbf{R}^+$  si  $\mathbf{P}_X(H) = \int_H f_X$  pour tout  $H \in \mathcal{B}(\mathbf{R}^d)$ . Dans ce cas,  $F_X(x_1, \dots, x_d)$  est l'intégrale de  $f_X$  sur le pavé  $\prod_k ] - \infty, x_k ]$ . D'après le théorème de Fubini 4.37, on peut écrire de manière équivalente :

$$F_X(x_1, \dots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_X(u_1, \dots, u_d) du_1 \cdots du_d.$$

Si  $F_X$  est de classe  $C^d$ , alors :

$$f_X(x_1, \dots, x_d) = \frac{\partial^d F_X(x_1, \dots, x_d)}{\partial x_1 \cdots \partial x_d}.$$

### 5.3.2. Variables aléatoires indépendantes

Soit  $X = (X_1, \dots, X_d)$  un vecteur aléatoire sur  $(\Omega, \mathcal{F})$ . On rappelle que les v.a.  $X_1, \dots, X_d$  sont dites indépendantes si pour tout  $H_1, \dots, H_d \in \mathcal{B}(\mathbf{R})$ ,

$$\mathbf{P}(X_1 \in H_1, \dots, X_d \in H_d) = \mathbf{P}(X_1 \in H_1) \times \cdots \times \mathbf{P}(X_d \in H_d).$$

Le membre de droite est égal au produit  $\prod_i \mathbf{P}_{X_i}(H_i)$  où  $\mathbf{P}_{X_i}$  est la loi marginale de  $X_i$ . Le membre de gauche est égal à la loi jointe  $\mathbf{P}_X$  évaluée en  $H_1 \times \cdots \times H_d$ . D'après le paragraphe précédent, l'unique loi satisfaisant la propriété ci-dessus est la loi produit. Ainsi, les v.a.  $X_1, \dots, X_d$  sont indépendantes si et seulement si *la loi jointe est égale au produit des lois marginales* :

$$\mathbf{P}_X = \bigotimes_{i=1}^d \mathbf{P}_{X_i} . \quad (5.6)$$

On note  $F_X$  la fonction de répartition de  $X = (X_1, \dots, X_d)$  et  $F_{X_i}$  celle de la v.a.r.  $X_i$ .

**Théorème 5.13.** Les propositions suivantes sont équivalentes.

- i)  $X_1, \dots, X_d$  sont indépendantes ;
- ii) Pour tout  $x_1, \dots, x_d \in \mathbf{R}$ ,

$$F_X(x_1, \dots, x_d) = F_{X_1}(x_1) \times \cdots \times F_{X_d}(x_d) ; \quad (5.7)$$

- iii) Pour toutes fonctions mesurables  $h_1, \dots, h_d : \mathbf{R} \rightarrow \mathbf{R}$  telles que les v.a.  $h_i(X_i)$  sont toutes positives *ou* toutes intégrables,

$$\mathbf{E}[h_1(X_1) \times \cdots \times h_d(X_d)] = \mathbf{E}[h_1(X_1)] \times \cdots \times \mathbf{E}[h_d(X_d)] ; \quad (5.8)$$

- iv) Pour toutes fonctions  $h_1, \dots, h_d : \mathbf{R} \rightarrow \mathbf{R}^+$  continues à support compact, (5.8) est satisfaite.

Si en outre chaque v.a.r.  $X_i$  admet une densité  $f_{X_i}$ , alors les propositions précédentes sont équivalentes à :

- v) Le vecteur  $X$  admet une densité  $f_X$  donnée pour tout  $x_1, \dots, x_d \in \mathbf{R}$  par :

$$f_X(x_1, \dots, x_d) = f_{X_1}(x_1) \times \cdots \times f_{X_d}(x_d) ; \quad (5.9)$$

*Démonstration.* i)  $\Rightarrow$  iii). On donne la preuve dans le cas où  $d = 2$  (le cas général se traite de manière similaire). Posons  $Y_1 = h_1(X_1)$  et  $Y_2 = h_2(X_2)$ . D'après le paragraphe 5.3.2, nous savons que  $Y_1$  et  $Y_2$  sont indépendantes. Supposons  $Y_1, Y_2$  positives. Puisque  $\mathbf{P}_{Y_1, Y_2} = \mathbf{P}_{Y_1} \otimes \mathbf{P}_{Y_2}$ , le théorème de Fubini implique que  $\mathbf{E}[Y_1 Y_2] = \int y_1 y_2 d\mathbf{P}_{Y_1, Y_2}(y_1, y_2)$ . Donc,  $\mathbf{E}[Y_1 Y_2] = \int (\int y_1 y_2 d\mathbf{P}_{Y_1}(y_1)) d\mathbf{P}_{Y_2}(y_2) = \int y_2 \mathbf{E}[Y_1] d\mathbf{P}_{Y_2}(y_2) = \mathbf{E}[Y_1] \mathbf{E}[Y_2]$ . La propriété est prouvée pour des v.a.  $Y_i$  positives. Dans le cas de v.a. signées, on a d'après ce qui précède :  $\mathbf{E}[|Y_1 Y_2|] = \mathbf{E}[|Y_1|] \mathbf{E}[|Y_2|] < \infty$ . La fonction  $(y_1, y_2) \mapsto y_1 y_2$  est intégrable et le théorème de Fubini s'applique là encore.

iii)  $\Rightarrow$  iv). Immédiat.

iv)  $\Rightarrow$  ii). Fixons  $x_1, \dots, x_d$ . Soit  $h_{i,n}$  la fonction continue égale à un sur l'intervalle  $] -\infty, x_i]$ , à zéro sur  $[x_i + \frac{1}{n}, +\infty[$ , et linéaire sur  $[x_i, x_i + \frac{1}{n}]$ . Pour tout  $i$ ,  $h_{i,n} \uparrow \mathbf{1}_{]-\infty, x_i]}$  et donc  $\mathbf{E}[h_{i,n}(X_i)] \uparrow F_{X_i}(x_i)$  par convergence monotone. De même,  $\prod_i h_{i,n} \uparrow \prod_i \mathbf{1}_{]-\infty, x_i]}$  et

donc  $\mathbf{E} [\prod_i h_{i,n}(X_i)] \uparrow \prod_i F_{X_i}(x_i)$ . Or, par hypothèse,  $\mathbf{E} [\prod_i h_{i,n}(X_i)] = \prod_i \mathbf{E} [h_{i,n}(X_i)]$ . Le résultat est obtenu par passage à la limite.

ii)  $\Rightarrow$  i). La fonction de répartition associée à la loi produit  $\otimes_i \mathbf{P}_{X_i}$  est égale au produit des  $F_{X_i}$ . Donc  $\otimes_i \mathbf{P}_{X_i}$  et  $\mathbf{P}_X$  ont la même fonction de répartition. Puisque la fonction de répartition caractérise la loi, ces lois sont égales, ce qui prouve (5.6).

Dans le cas où chaque  $X_i$  admet une densité  $f_{X_i}$  on montre que ii)  $\Leftrightarrow$  iv). ii) équivaut à :  $\forall x = (x_1, \dots, x_d)$ ,  $F_X(x) = \prod_i \left( \int_{-\infty}^{x_i} f_{X_i} \right)$ . Par le théorème de Fubini, cela équivaut à :  $F_X(x) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} (\otimes_i f_{X_i})$ , ce qui revient à dire que  $X$  est de densité  $\otimes_i f_{X_i}$ .  $\square$

Du théorème ci-dessus, on retiendra en particulier la propriété importante suivante : *l'espérance d'un produit de variables aléatoires indépendantes est égale au produit des espérances*. Plus précisément, si  $X_1, \dots, X_d$  des v.a.r. indépendantes telles que  $\mathbf{E} [|X_i|] < \infty$  pour tout  $i$ , alors le produit  $X_1 \times \dots \times X_d$  est une v.a. intégrable et on a :

$$\mathbf{E} [X_1 \times \dots \times X_d] = \mathbf{E} [X_1] \times \dots \times \mathbf{E} [X_d] .$$

En corollaire, l'égalité ci-dessus implique que si  $X_1$  et  $X_2$  sont indépendantes alors

$$\text{Cov}(X_1, X_2) = 0 .$$

Autrement dit, *des variables indépendantes sont décorrélées*. Attention ! La réciproque est fausse.

### Généralisation à une collection de vecteurs aléatoires

Le Théorème 5.13 admet une généralisation immédiate au cas où  $X_1, \dots, X_d$  sont elles-mêmes des vecteurs aléatoires de dimensions  $n_1, \dots, n_d$  respectivement. Il suffit d'adapter l'énoncé au fait que pour tout  $i$ ,  $F_{X_i}$  et  $f_{X_i}$  sont cette fois des fonctions de  $\mathbf{R}^{n_i} \rightarrow \mathbf{R}$ .

**Théorème 5.14.** Les propositions suivantes sont équivalentes.

- i)  $X_1, \dots, X_d$  sont indépendantes ;
- ii) Pour tout  $(x_1, \dots, x_d) \in \mathbf{R}^{n_1} \times \dots \times \mathbf{R}^{n_d}$ , (5.7) est vérifiée ;
- iii) Pour toutes fonctions mesurables  $h_i : \mathbf{R}^{n_i} \rightarrow \mathbf{R}$  ( $i = 1, \dots, d$ ) telles que les v.a.  $h_i(X_i)$  sont toutes positives ou toutes intégrables, (5.8) est vérifiée ;
- iv) Pour toutes fonctions  $h_i : \mathbf{R}^{n_i} \rightarrow \mathbf{R}$  ( $i = 1, \dots, d$ ) continues à support compact, (5.8) est vérifiée.

Si en outre chaque v.a.r.  $X_i$  admet une densité  $f_{X_i}$ , alors les trois propositions précédentes sont équivalentes à :

- v) Le vecteur  $X$  admet une densité  $f_X$  donnée par (5.9) pour tout  $(x_1, \dots, x_d) \in \mathbf{R}^{n_1} \times \dots \times \mathbf{R}^{n_d}$ .

### Généralisation au cas d'une famille de v.a.

Une famille de variables aléatoires est un ensemble de variables aléatoires de cardinal non nécessairement fini. Dans ce cas, on restreint l'indépendance aux seules sous-familles finies.

**Définition 5.9.** Une famille de variables aléatoires est dite indépendante si toute sous-famille *finie* est indépendante.

*n.b.* : on utilise souvent l'abréviation *i.i.d.* pour désigner une famille indépendante et identiquement distribuée de variables aléatoires.

On retiendra de tout ceci le lemme fondamental suivant.

**Lemme 5.15.** Soit  $(X_i)_{i \in I}$  une famille de v.a. indépendantes et soit  $J$  et  $K$  deux sous-ensembles disjoints de  $I$ . Les familles  $(X_i, i \in J)$  et  $(X_i, i \in K)$  sont indépendantes.

*Démonstration.* Pour 3 v.a. réelles indépendantes,  $X$ ,  $Y$  et  $Z$ . On doit prouver que la v.a.r.  $X$  est indépendante du vecteur aléatoire  $(Y, Z)$ . D'après théorème 5.14, il suffit de prouver que

$$\mathbf{P}((X \leq x) \cap (Y \leq y, Z \leq z)) = \mathbf{P}(X \leq x)\mathbf{P}(Y \leq y, Z \leq z). \quad (5.10)$$

Or l'indépendance des trois variables induit que

$$\mathbf{P}((X \leq x) \cap (Y \leq y, Z \leq z)) = \mathbf{P}(X \leq x)\mathbf{P}(Y \leq y)\mathbf{P}(Z \leq z).$$

De plus l'indépendance de  $Y$  et  $Z$  implique que

$$\mathbf{P}(Y \leq y)\mathbf{P}(Z \leq z) = \mathbf{P}(Y \leq y, Z \leq z).$$

En mettant bout à bout les deux dernières identités, on obtient (5.10). □

EXEMPLE 30.— Soit  $(X_n, n \geq 1)$  des variables aléatoires indépendantes. En considérant  $J = \{1\}$  et  $K = \{2, \dots, 10\}$ , le lemme précédent indique que  $X_1$  est indépendante de  $\sum_{n=2}^{10} X_n$ .

### 5.3.3. Moments, matrice de covariance

Soient  $X_1, \dots, X_d$  des v.a. réelles. On s'intéresse au vecteur-colonne  $X = (X_1, \dots, X_d)^T$ .

**Définition 5.10.** L'espérance de  $X$  est définie comme le vecteur des espérances :

$$\mathbf{E}(X) := \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_d) \end{pmatrix}.$$

Elle est bien définie si et seulement si toutes les composantes  $X_k$  admettent une espérance.

Un vecteur aléatoire  $X$  est dit d'ordre  $p$  si toutes ses composantes  $X_1, \dots, X_d$  sont d'ordre  $p$ . Cela revient à dire que  $\mathbf{E}(\|X\|^p) < \infty$  où  $\|\cdot\|$  est une norme sur  $\mathbf{R}^d$ .

**Définition 5.11.** On appelle *matrice de covariance* d'un vecteur  $X$  d'ordre 2 la matrice notée  $\text{Cov}(X)$  dont le coefficient  $(i, j)$  vaut  $\text{Cov}(X_i, X_j)$  :

$$\text{Cov}(X) := (\text{Cov}(X_i, X_j))_{i,j=1\dots d}.$$

En particulier, le  $i$ ème coefficient diagonal de  $\text{Cov}(X)$  vaut  $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$ . On notera donc les deux propriétés utiles suivantes :

- La diagonale de  $\text{Cov}(X)$  est égale au vecteur des variances ;
- Dans le cas où les v.a.  $X_1, \dots, X_d$  sont décorrélées, la matrice de covariance est diagonale.

La preuve de la proposition suivante est laissée à titre d'exercice.

**Proposition 5.16.** Soit  $X$  un vecteur aléatoire d'ordre 2 de taille  $d$ ,  $A$  une matrice constante de taille  $n \times d$  et  $b$  un vecteur constant de taille  $n$ . Alors

- a)  $\mathbf{E}(AX + b) = A\mathbf{E}(X) + b$  ;
- b)  $\text{Cov}(AX + b) = A \text{Cov}(X) A^T$  ;
- c)  $\text{Cov}(X) = \mathbf{E}((X - \mathbf{E}(X))(X - \mathbf{E}(X))^T)$ .

**Proposition 5.17.**  $\text{Cov}(X)$  est une matrice symétrique semi-définie positive.

*Démonstration.* On voit immédiatement que  $\text{Cov}(X)$  est symétrique car  $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$ . Pour tout  $x$  vecteur-colonne de  $\mathbf{R}^d$ , on calcule

$$x^T \text{Cov}(X) x = x^T \mathbf{E}((X - \mathbf{E}(X))(X - \mathbf{E}(X))^T) x = \mathbf{E}(x^T (X - \mathbf{E}(X))(X - \mathbf{E}(X))^T x) = E \left[ \left( x^T X \right)^2 \right] \geq 0.$$

□

## 5.4. Changement de variables

### 5.4.1. Introduction

Soit  $X$  est un vecteur aléatoire sur  $\mathbf{R}^d$  admettant une densité connue  $f_X$ . On pose :

$$Y := \phi(X)$$

pour une certaine fonction borélienne  $\phi$ . L'objectif de cette section est de déterminer la loi de  $Y$  et, si elle existe, sa densité. Les exercices 5.4 et 5.5 montrent que, *dans les cas simples* ( $d = 1$ ), la réponse est immédiatement donnée par calcul et différentiation de la fonction de répartition de  $Y$ . Dans les cas plus complexes, l'expression de la densité de  $Y$  est obtenue grâce à la *formule du changement de variable*.

### 5.4.2. Formule du changement de variable

**Définition 5.12.** Soit une fonction  $\phi$  définie sur une partie ouverte  $O$  de  $\mathbf{R}^n$  à valeurs dans  $\mathbf{R}^n$ , c'est-à-dire que l'on a

$$\begin{aligned} \phi : O \subset \mathbf{R}^n &\longrightarrow \mathbf{R}^n \\ x &\longmapsto \phi(x) = (\phi_1(x), \dots, \phi_n(x)). \end{aligned}$$

Si toutes les dérivées partielles existent sur  $O$ , la jacobienne de  $\phi$  au point  $x$ , est la matrice  $J_\phi(x)$  où

$$\begin{aligned} J_\phi(x) &= \left( \frac{\partial \phi_i}{\partial x_j}(x), 1 \leq i \leq n, 1 \leq j \leq n \right) \\ &= \begin{pmatrix} \frac{\partial \phi_1}{\partial x_1}(x) & \vdots & \vdots & \vdots \\ \vdots & \frac{\partial \phi_i}{\partial x_j}(x) & \vdots & \vdots \\ \vdots & \vdots & \frac{\partial \phi_n}{\partial x_n}(x) & \vdots \end{pmatrix}. \end{aligned}$$

Le jacobien de  $\phi$  est le déterminant de  $J_\phi$ .

**Définition 5.13.** Soit  $O$  un ouvert de  $\mathbf{R}^n$ ,  $\phi : O \subset \mathbf{R}^n \rightarrow \mathbf{R}^n$ ,  $\phi$  est un  $\mathcal{C}^1$ -difféomorphisme de  $O$  sur  $\Delta \subset \mathbf{R}^n$ , lorsque

- les dérivées partielles de  $\phi$  existent et sont continues sur  $O$ ,
- $\phi$  est une bijection de  $O$  sur  $\Delta$ ,

— le jacobien de  $\phi$  ne s'annule pas sur  $O$ .

On rappelle la propriété suivante des difféomorphismes :

$$\det J_{\phi^{-1}} = \frac{1}{(\det J_{\phi}) \circ \phi^{-1}}.$$

**Proposition 5.18.** Soient  $U$  et  $V$  deux ouverts de  $\mathbf{R}^d$  et  $\phi : U \rightarrow V$  une application bijective, continûment différentiable sur  $U$ . Alors  $\phi$  est un difféomorphisme si et seulement si  $J_{\phi}(x) \neq 0, \forall x \in U$ .

**Théorème 5.19.** (*Formule du changement de variable*). Soient  $U$  et  $V$  deux ouverts de  $\mathbf{R}^d$  et  $\phi : U \rightarrow V$  un difféomorphisme. Alors si  $f$  est une fonction définie sur  $V$  à valeurs positives,

$$\int_U f \circ \phi = \int_V \frac{f}{|(\det J_{\phi}) \circ \phi^{-1}|} . \quad (5.11)$$

REMARQUE.— Dans l'égalité (5.11), on suppose implicitement que  $f : U \rightarrow \mathbf{R}$  est une fonction borélienne. Dans le cas où  $f$  n'est pas nécessairement positive, alors l'égalité (5.11) est satisfaite au moins par la valeur absolue  $|f|$  et, dans le cas où les deux membres de l'égalité sont finis, les barres de valeur absolue peuvent être enlevées.

### 5.4.3. Application au calcul de densité

Revenons au problème initialement posé. On souhaite déterminer la loi de

$$Y = \phi(X)$$

où  $X$  est un vecteur aléatoire. On fait les hypothèses suivantes :

- $X$  admet une densité  $f_X$ .
- $X(\Omega) \subset U \subset \mathbf{R}^d$  où  $U$  est un ouvert ;
- $\phi : U \rightarrow V$  est un difféomorphisme.

Soit  $h$  une fonction arbitraire, positive, définie sur  $\mathbf{R}^d$ . On évalue l'espérance :

$$\begin{aligned} \mathbf{E}[h(Y)] &= \mathbf{E}[(h \circ \phi)(X)] \\ &= \int_U (h \circ \phi) f_X \\ &= \int_U (h \times (f_X \circ \phi^{-1})) \circ \phi \\ &= \int_V h \frac{f_X \circ \phi^{-1}}{|(\det J_{\phi}) \circ \phi^{-1}|} \end{aligned}$$

où la dernière égalité provient de la formule du changement de variable. Ainsi, on peut écrire  $\mathbf{E}[h(Y)] = \int h f_Y$  où :

$$f_Y = \frac{f_X \circ \phi^{-1}}{|(\det J_{\phi}) \circ \phi^{-1}|} \mathbf{1}_V . \quad (5.12)$$

Le calcul précédent étant valable pour toute fonction positive  $h$ , il l'est en particulier lorsque  $h$  est de la forme  $h = \mathbf{1}_H$  pour un certain ensemble  $H \in \mathcal{B}(\mathbf{R})$ . L'égalité  $\mathbf{E}[h(Y)] = \int h f_Y$  se lit alors :

$$\mathbf{P}(Y \in H) = \int_H f_Y .$$

On en conclut que  $Y$  est un vecteur aléatoire de densité  $f_Y$  donnée par (5.12).

EXEMPLE 31.— Soit  $(X_1, X_2)$  deux variables aléatoires réelles indépendantes, de même loi

$$d\mathbf{P}(x) = \mathbf{1}_{[1, \infty[}(x) \frac{1}{x^2} dx.$$

On pose  $U = X_1 X_2$  et  $V = X_1 / X_2$ .

1. Calculer la loi du vecteur  $(U, V)$ .
2. Calculer la loi de  $U$  et celle de  $V$ .
3.  $U$  et  $V$  sont-elles indépendantes ?

On part de l'hypothèse que la loi du couple  $(U, V)$  a une densité par rapport à la mesure de Lebesgue, ce qui revient à trouver  $h : \mathbf{R}^2 \rightarrow \mathbf{R}^+$  telle que pour toute fonction  $f$  continue bornée de  $\mathbf{R}^2$  dans  $\mathbf{R}$ , on ait

$$\mathbf{E}[f(U, V)] = \int_{\mathbf{R}^2} f(x, y) h(x, y) dx dy.$$

Posons

$$\begin{aligned} \phi : \mathbf{R}^2 &\rightarrow \mathbf{R}^2 \\ (x, y) &\mapsto (xy, x/y). \end{aligned}$$

On a

$$\mathbf{E}[f(U, V)] = \mathbf{E}[(f \circ \phi)(X, Y)] = \int (f \circ \phi)(x, y) d\mathbf{P}_{X, Y}(x, y),$$

où la deuxième égalité découle du théorème de transfert. Maintenant, les v.a.  $X$  et  $Y$  sont indépendantes, ce qui équivaut à dire que

$$d\mathbf{P}_{X, Y}(x, y) = d\mathbf{P}_X(x) \otimes d\mathbf{P}_Y(y).$$

Par hypothèse,

$$d\mathbf{P}_X(x) = \mathbf{1}_{[1, \infty[}(x) \frac{1}{x^2} dx \text{ et } d\mathbf{P}_Y(y) = \mathbf{1}_{[1, \infty[}(y) \frac{1}{y^2} dy,$$

donc

$$d\mathbf{P}_{X, Y}(x, y) = \mathbf{1}_{[1, \infty[}(x) \frac{1}{x^2} \mathbf{1}_{[1, \infty[}(y) dx dy.$$

On a donc obtenu

$$\mathbf{E}[f(U, V)] = \int_{[1, +\infty[^2} (f \circ \phi)(x, y) \frac{1}{x^2 y^2} dx dy.$$

Rappelons nous que nous voulons aboutir à une identité de la forme

$$\mathbf{E}[f(U, V)] = \int_{\mathbf{R}^2} f(x, y) h(x, y) dx dy.$$



On est naturellement enclin à utiliser le théorème de changement de variables, pour cela, il nous faut calculer  $\Delta$ , l'ensemble image de  $[1, +\infty[^2$  par  $\phi$  et le jacobien de  $\phi$ . Posons  $u = xy$  et  $v = x/y$ ,

$$\det J_\phi(x, y) = \det \begin{pmatrix} y & x \\ \frac{1}{y} & -\frac{x}{y^2} \end{pmatrix} = -2 \frac{x}{y} = -2v.$$

Si  $x$  et  $y$  sont tous deux plus grands que 1 alors  $u$  l'est, et  $v$  est strictement positif. Par ailleurs,

$$\begin{cases} u = xy \\ v = x/y \end{cases} \iff \begin{cases} x^2 = uv \\ y^2 = u/v \end{cases}.$$

On déduit de ces dernières équations que  $u \geq v$  et  $uv \geq 1$ . On vérifie alors facilement que  $\phi$  est une bijection de  $[1, +\infty[^2$  sur

$$\Delta = \{(u, v), u \geq v \geq 0 \text{ et } uv \geq 1\}.$$

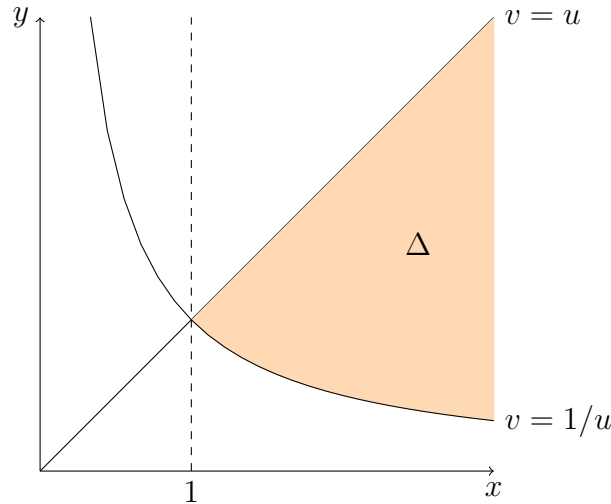


Figure 5.1. – Le domaine  $\Delta$  dans le plan  $(u, v)$ .

On tire du théorème de changement de variables que

$$\begin{aligned} \int_{[1, +\infty[^2} (f \circ \phi)(x, y) \frac{1}{x^2 y^2} dx dy &= \int_{\Delta} f(u, v) \frac{1}{uv \cdot u/v} \left| \frac{1}{-2v} \right| du dv \\ &= \int_{\Delta} f(u, v) \frac{1}{2u^2 v} du dv, \end{aligned}$$

d'où par identification,

$$d\mathbf{P}_{(U, V)}(u, v) = \frac{1}{2u^2 v} \mathbf{1}_{\Delta}(u, v) du dv.$$

Pour calculer la loi de  $U$ , on veut exprimer  $\mathbf{E}[f(U)]$  pour toute fonction continue bornée de  $\mathbf{R}$  dans  $\mathbf{R}$ . On remarque alors que l'application  $\tilde{f}(x, y) = f(x)$  est continue bornée de

$\mathbf{R}^2$  dans  $\mathbf{R}$  donc

$$\begin{aligned}\mathbf{E}[f(U)] &= \mathbf{E}[\tilde{f}(U, V)] = \int_{\Delta} f(u) \frac{1}{2u^2v} du dv \\ &= \int_1^{+\infty} f(u) \frac{1}{2u^2} \left( \int_{1/u}^u \frac{1}{v} dv \right) du,\end{aligned}$$

d'après le théorème de Fubini. Par conséquent,

$$\begin{aligned}d\mathbf{P}_U(u) &= \frac{1}{2u^2} \left( \int_{1/u}^u \frac{1}{v} dv \right) \mathbf{1}_{[1, +\infty[}(u) du \\ &= \frac{\ln u}{u^2} \mathbf{1}_{[1, +\infty[}(u) du.\end{aligned}$$

De même,

$$\begin{aligned}\mathbf{E}[f(V)] &= \int_{\Delta} f(v) \frac{1}{2u^2v} du dv \\ &= \int_0^{+\infty} f(v) \frac{1}{2v} \left( \int_{\Delta_v} \frac{1}{u^2} du \right) dv,\end{aligned}$$

où

$$\begin{aligned}\Delta_v &= \{u (u, v) \in \Delta\} \\ &= \begin{cases} [1/v, +\infty[ & \text{si } 0 \leq v \leq 1 \\ [v, +\infty[ & \text{si } v \geq 1. \end{cases}\end{aligned}$$

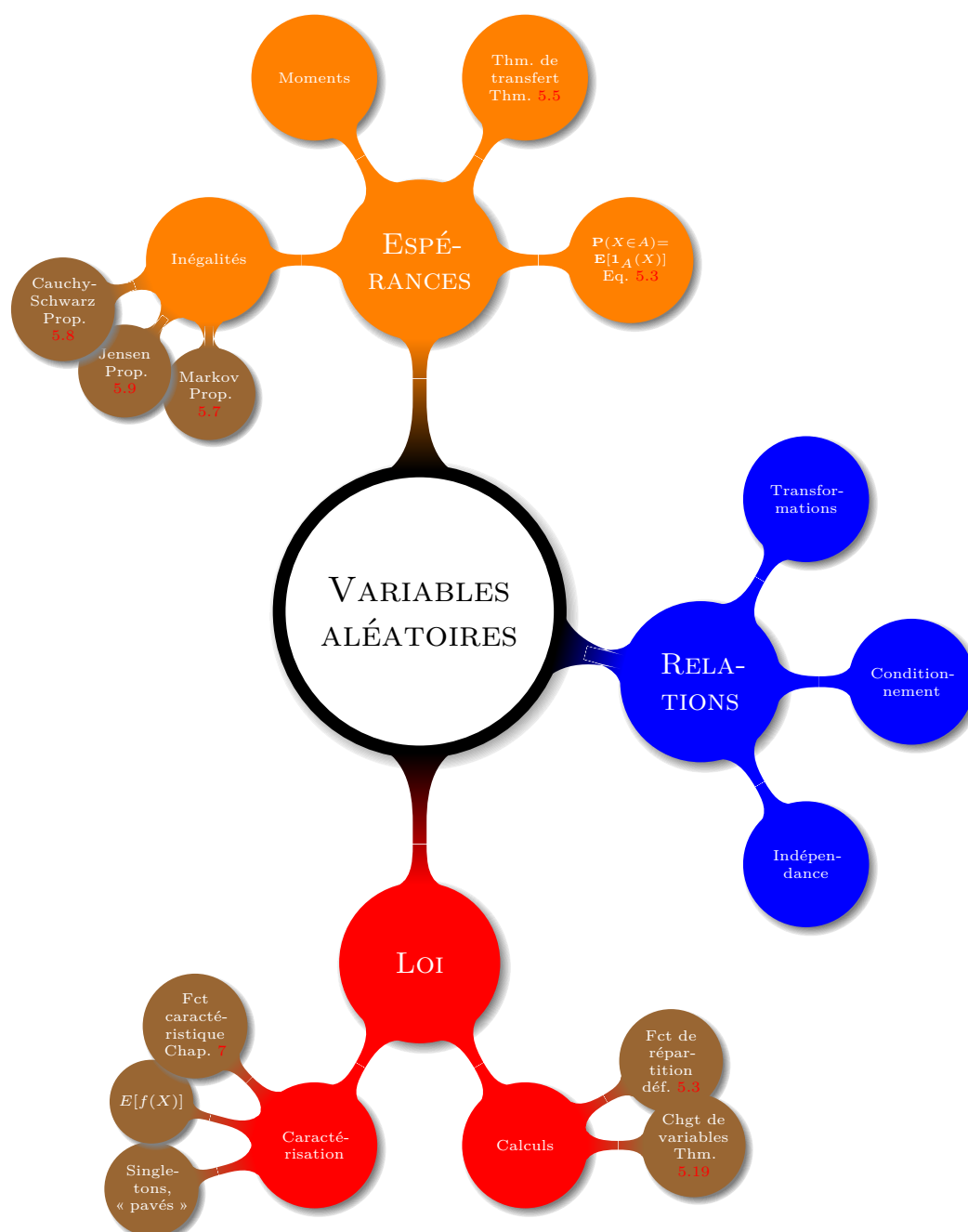
Par conséquent,

$$\begin{aligned}d\mathbf{P}_V(v) &= \frac{1}{2v} \left( \int_{1/v}^{+\infty} \frac{1}{u^2} du \mathbf{1}_{[0, 1]}(v) + \int_v^{+\infty} \frac{1}{u^2} du \mathbf{1}_{]v, +\infty[}(v) \right) \\ &= \frac{1}{2v} \left( v \mathbf{1}_{[0, 1]}(v) + \frac{1}{v} \mathbf{1}_{]1, +\infty[}(v) \right).\end{aligned}$$

Comme

$$d\mathbf{P}_{(U, V)} \neq d\mathbf{P}_U \otimes d\mathbf{P}_V,$$

les v.a.  $U$  et  $V$  ne sont pas indépendantes.



## 5.5. Exercices

### Pour apprendre

EXERCICE 5.1. –

Soit  $X$  la v.a. dont la loi est donnée par

$$\mathbf{P}(X = n) = \frac{6}{\pi^2} \frac{1}{n^2}, \text{ pour } n \geq 1.$$

Montrer que  $\mathbf{P}(X < +\infty) = 1$  mais que  $X$  n'a pas d'espérance.

EXERCICE 5.2. –

Soit  $X$  une v.a.r. sur  $(\Omega, \mathcal{F}, \mathbf{P})$  de fonction de répartition  $F_X$ .

1. Démontrer les égalités :  $\mathbf{P}[a < X \leq b] = F_X(b) - F_X(a)$ ,  $\mathbf{P}[a < X < b] = F_X(b^-) - F_X(a)$ ,  $\mathbf{P}[a \leq X \leq b] = F_X(b) - F_X(a^-)$ .
2. Calculer  $F_X$  dans les cas suivants :  $X$  suit la loi exponentielle de paramètre  $\alpha$ ,  $X$  suit la loi uniforme sur l'intervalle  $[a, b]$ .

EXERCICE 5.3. –

Nous dirons qu'une v.a.r.  $X$  est *symétrique* lorsque  $X$  et  $-X$  ont la même loi. Si  $X$  est une v.a.r. de densité  $f$ , montrer que  $X$  est symétrique si et seulement si  $f(x) = f(-x)$  pour tout  $x$  hors d'un ensemble négligeable.

EXERCICE 5.4. –

Soit  $X$  une v.a.r. à densité et  $(a, b) \in \mathbb{R}^2$ . Exprimer la densité de la v.a.r.  $Y := aX + b$  en fonction de la densité de  $X$ .

EXERCICE 5.5. –

Soit  $X \sim \mathcal{N}(0, 1)$  et  $Y = X^2$ .

1. Calculer la fonction de répartition  $F_Y$  de  $Y$  en fonction de celle de  $X$ .
2. En déduire que  $Y$  admet une densité, que l'on exprimera.

EXERCICE 5.6. –

Soit  $\mu_n$  la suite de mesure sur  $[0, 1]$  donnée par

$$d\mu_n(x) = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{j/n}.$$

Pour  $f$  continue sur  $[0, 1]$ , quelle est la limite de  $\int f(t) d\mu_n(t)$  quand  $n$  tend vers  $+\infty$  ?

EXERCICE 5.7. – 1. Soit  $X$  une v.a.r. de densité  $f$  telle que  $\lim_{|x| \rightarrow \infty} |x|^p f(x) = 0$  pour tout  $p > 0$ . Montrer que  $X$  possède tous ses moments. En déduire qu'une v.a.r. gaussienne possède tous ses moments.

2. Soit  $X$  une v.a.r. suivant une loi de Cauchy (voir la table 4.2). Montrer que  $\mathbf{E}(|X|)$  diverge. Plus généralement, montrer que  $X$  ne possède aucun moment.

EXERCICE 5.8. –

Soient  $X_1, \dots, X_d$  des v.a. i.i.d. de loi  $\mathcal{N}(0, 1)$ .

1. Quelle est la loi du vecteur aléatoire  $X = (X_1, \dots, X_d)$  ?
2. Même question pour des variables  $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$  indépendantes,  $\sigma_i^2 > 0$ .

## Pour s'entraîner

EXERCICE 5.9 (\*). –

Un nombre est choisi au hasard dans l'intervalle  $[0, 10]$  suivant une loi  $\mathbf{P}$  donnée par

$$d\mathbf{P}(t) = K t \mathbf{1}_{[0,10]}(t) dt,$$

où  $K$  est une constante à calculer. On note par  $X$  sa partie entière et par  $Y$  sa partie fractionnaire.

1. Calculer la loi du vecteur  $(X, Y)$ . Est-ce que les composantes sont indépendantes ?
2. Calculer la matrice de covariance de  $(X, Y)$ .

EXERCICE 5.10 (\*\*). –

Pour  $a > 0$ , on définit

$$\Gamma(a) = \int_0^\infty e^{-t} t^{a-1} dt.$$

Une v.a.r.  $X$  est dite de loi gamma de paramètres  $a$  et  $\lambda > 0$  si sa loi est donnée par

$$dP_X(t) = \mathbf{1}_{[0,\infty[}(t) \frac{\lambda^a}{\Gamma(a)} e^{-\lambda t} t^{a-1} dt,$$

notée par  $X \sim G(a, \lambda)$ .

1. Calculer l'espérance et la variance de  $X$ .
2. Soit  $Y$  une autre v.a.r. indépendante de  $X$ , de loi  $G(b, \lambda)$ . Montrer que  $X + Y$  et  $\frac{X}{X + Y}$  sont indépendantes, calculer leur loi.
3. En déduire que

$$\beta(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

EXERCICE 5.11 (\*). –

Soit  $X$  une v.a. réelle de fonction de répartition  $F_X$  et  $F_X^{-1}$  l'inverse à droite de  $F_X$  défini par :

$$F_X^{-1}(y) = \inf\{u; F_X(u) \geq y\}.$$

Soit  $U$  une v.a. de loi uniforme sur  $[0, 1]$ , montrer que  $F_X^{-1}(U)$  a la loi de  $X$ . Cette relation permet de générer des v.a. de loi arbitraire à partir de variables de loi uniforme sur  $[0, 1]$ . Ceci est très fréquemment utilisé en simulation et connu sous le nom de méthode d'inversion.

Trouver comment générer des variables de loi exponentielle et de Cauchy avec cette méthode.

## EXERCICE 5.12 (\*).—

La difficulté qui apparaît lors de la mise en oeuvre de la méthode précédente est l'inversion de la fonction de répartition. On a fréquemment la densité de façon explicite mais pas la fonction de répartition. Dans ce cas, on applique la méthode de rejet. Soit  $f_X$  la densité de  $X$  et  $g$  une densité qui majore à une constante près  $f_X$  et pour laquelle on sait facilement générer des v.a. dont la loi a pour densité  $g$ . On procède de la manière suivante : soit  $a$  tel que  $f_X(u) \leq ag(u)$  pour tout  $u$ . On tire une v.a. de loi de densité  $g$ , soit  $Y$  le résultat de ce tirage. On tire, indépendamment, une v.a. de loi uniforme sur  $[0, 1]$  et on note  $U$  le résultat de ce tirage. Si  $U \leq f(Y)/ag(Y)$  alors le résultat est  $Y$  sinon on recommence au début.

1. Quel est l'espace de probabilité sous-jacent sur lequel sont définies les v.a.  $Z$  et  $Y$ .
2. Montrer que  $\mathbf{P}(Y \leq t) = F_X(t)$ .
3. Soit  $X$  et  $Y$  deux v.a. indépendantes de loi exponentielle de paramètre  $\mu$ . Calculer la densité de la loi de  $Z = X - Y$ .
4. En déduire une façon d'engendrer des v.a. de loi de densité :

$$\frac{\mu}{2\gamma(1 + 1/\alpha)} \exp(-\mu|x|^\alpha)$$

où  $\alpha \geq 1$  et  $\mu > 0$ .

## EXERCICE 5.13.—

Soit  $U$  et  $V$  deux v.a. indépendantes de loi uniforme sur  $[0, 1]$ . Posons :

$$X = \sqrt{-2 \ln(U)} \cos(2\pi V) \text{ et } Y = \sqrt{-2 \ln(U)} \sin(2\pi V).$$

Montrer que  $X$  et  $Y$  sont des v.a. gaussiennes centrées, réduites, indépendantes.

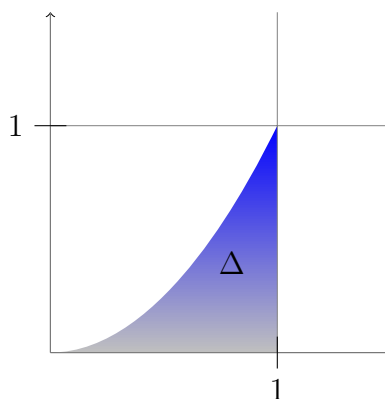
EXERCICE 5.14 (Simulation dans un cercle - \*).— 1. Comment utiliser la méthode de rejet pour simuler le tirage d'un point « au hasard » dans un cercle de rayon  $R$ ?

2. Calculer la loi jointe de module et de l'argument d'un point « au hasard » dans le même cercle. En déduire une autre façon de simuler le tirage d'un point choisi uniformément dans le cercle.
3. Quelle est la meilleure méthode?

## EXERCICE 5.15 (\*\*).—

Soit un couple de v.a.  $(X, Y)$  de densité conjointe  $f(x, y) = cxy^2$  si  $(x, y) \in \Delta$  et 0 sinon, où  $c$  est une constante positive et  $\Delta$  est donné par

$$\Delta = \{(x, y) \in [0, 1]^2, 0 < y < x^2 < 1\}.$$



1. Déterminer la constante  $c$ .
2. Déterminer les lois marginales des variables  $X$  et  $Y$ .
3. Soit  $U$  une v.a. de loi exponentielle de paramètre 8, et  $V$  une v.a. de loi exponentielle de paramètre 3, indépendante de  $U$ . On pose

$$W = \exp(-U), \quad Z = \exp(-V)W^2.$$

Vérifier que  $f$  est la densité de la loi du couple  $(W, Z)$ .

EXERCICE 5.16 (\*\*).—

Mots-clés : statistiques d'ordre, permutations aléatoires

Soit  $(X_1, \dots, X_n)$  des v.a. i.i.d. de fonction de répartition  $F$ . On définit par récurrence sur  $p$ , la suite de v.a.  $X_{(p)}$  par

$$\begin{aligned} X_{(1)} &= \min_{1 \leq j \leq n} X_j \\ \tau_1 &= \inf\{j, X_j = X_{(1)}\} \\ X_{(2)} &= \min_{j \neq \tau_1} X_j \\ \tau_2 &= \inf\{j \neq \tau_1, X_j = X_{(2)}\} \\ &\vdots \\ X_{(n)} &= \max_j X_j \\ \tau_n &= \max\{j, X_j = X_{(n)}\}. \end{aligned}$$

1. Montrer que presque sûrement,  $X_i \neq X_j$  pour  $i \neq j$ .
2. Calculer la fonction de répartition de  $X_{(1)}$  et de  $X_{(n)}$ .
3. Soit  $\tau$  la permutation définie par  $\tau(i) = \tau_i$ . Calculer la loi de  $\tau$ .
4. Calculer la loi de  $X_{(k)}$ .
5. Soit  $\alpha \in ]0, 1[$  et  $F_\alpha^n(x) = \mathbf{P}(X_{([\alpha n])} \leq x)$ . On définit  $x_\alpha$  par

$$x_\alpha = \inf\{x, F(x) \geq \alpha\}.$$

Montrer que

$$F_\alpha^n \xrightarrow{n \rightarrow +\infty} \begin{cases} 1 & \text{si } x \geq x_\alpha \\ 0 & \text{sinon.} \end{cases}$$

EXERCICE 5.17 (Recouvrement d'un cercle - \*\*\*).—

Soit  $U = (U_1, \dots, U_n)$  des v.a. i.i.d. de loi uniforme sur  $[0, 1]$ . Soit  $W = (W_1, \dots, W_n)$  la statistique d'ordre (cf. exercice 5.16) associée à  $U$ , i.e.,

$$W_i = U_{(i)}, \text{ pour tout } i = 1, \dots, n.$$

On pose

$$V_1 = 1 + W_1 - W_n, \quad V_2 = W_2 - W_1, \quad \dots, \quad W_n = W_n - W_{n-1}.$$

On considère aussi  $X_1, \dots, X_n$  des v.a. indépendantes de loi exponentielle de paramètre

1. On pose  $S_n = n^{-1} \sum_{j=1}^n X_j$ .

1. Montrer que la loi de  $W$  est donnée par

$$dP_W(w_1, \dots, w_n) = n! \mathbf{1}_A(w_1, \dots, w_n) dw_1 \dots dw_n,$$

où

$$A = \{(x_1, \dots, x_n), 0 \leq x_1 \leq x_2 \leq \dots \leq x_n \leq 1\}.$$

2. Calculer la loi de  $V = (nV_1, \dots, nV_{n-1})$ .

3. Calculer la loi de  $(X_1, \dots, X_{n-1}, S_n)$ .

4. Montrer que la loi de

$$\left( \frac{X_1}{S_n}, \dots, \frac{X_{n-1}}{S_n} \right)$$

est la même que celle de  $V$ .

5. Soit  $N_\alpha$  le nombre minimum d'arcs de longueur  $\alpha$  nécessaires pour recouvrir la circonférence du cercle unité. Montrer que

$$(N_\alpha \leq n) = (\max_{k \leq n} V_k \leq \alpha).$$

EXERCICE 5.18 (\*\*\*).—

On considère  $E = \{x = (x_1, x_2) \in \mathbf{R}^2, x_1^2 + x_2^2 \leq 1\}$  et on considère  $\Omega$  l'ensemble des familles finies de points de  $E$ , c'est-à-dire qu'un  $\omega \in \Omega$  est une famille finie de points de  $E$ . On munit  $E$  de la tribu borélienne et d'une probabilité  $\mathbf{P}$ . Pour toute partie  $A$  de  $E$  on définit la variable aléatoire  $N(A)(\omega)$  qui représente le nombre de points de  $\omega$  qui sont dans  $A$ . Les seules hypothèses que l'on fait sur  $\mathbf{P}$  sont :

— Pour toute partie borélienne  $A$  de  $E$ ,

$$\mathbf{P}(N(A) = k) = e^{-m(A)} \frac{m(A)^k}{k!}, \text{ pour tout } k \in \mathbf{N},$$

où  $m$  est la mesure de Lebesgue sur  $\mathbf{R}^2$ .

— Si  $(A_i, i \in \mathbf{N})$  sont des boréliens disjoints deux à deux, les v.a.  $(N(A_i), i \in \mathbf{N})$  sont indépendantes dans leur ensemble.

On appelle le triplet  $(E, \mathbf{P}, N)$  un processus de Poisson ponctuel d'intensité  $m$ .

1. Calculer la moyenne et la variance de  $N(A)$  pour  $A$  borélien de  $E$ . Calculer la probabilité que  $A$  ne contienne pas de points de  $\omega$ .

2. Soient  $A \subset B$  deux boréliens, calculer la loi de la variable aléatoire  $(N(A), N(B))$ .



3. Pour  $C = \{x, a^2 < x_1^2 + x_2^2 \leq b^2\}$ , calculer la loi de  $N(C)$ .
4. On pose  $U(\omega) = \inf\{\alpha, N(B(0, \alpha))(\omega) > 0\}$  où  $B(0, \alpha)$  est la boule fermée de centre  $O$  et de rayon  $r$ . Calculer  $\mathbf{P}(U > x)$  pour tout  $x$ .
5. On fixe  $r > 0$ , on considère  $A_\alpha^r$  le secteur angulaire composé des points distants de  $O$  de moins de  $r$  et d'argument compris entre 0 et  $\alpha$ . On pose  $V^r = \inf\{\beta, N(A_\beta^r) > 0\}$  avec la convention  $V^r = 0$  si  $B(0, r)$  ne contient pas de point de  $\omega$ . Calculer  $\mathbf{P}(V > x)$  pour tout  $x \in [0, 2\pi[$ .
6. Calculer la loi de l'argument du point le plus proche de  $O$ .
7. On suppose  $n$  fixé, pour  $k \in \{0, \dots, n-1\}$ , on appelle  $B_{k,n}^r$  le secteur angulaire des éléments de  $E$  de module inférieur à  $r$  et d'argument supérieur à  $2k\pi/n$  et strictement inférieur à  $2(k+1)\pi/n$ . Calculer la loi de  $(N(B_{1,n}^1), \dots, N(B_{n-1,n}^1))$  conditionnellement à  $N(E) = k$ .
8. On admet que les secteurs angulaires définis précédemment engendrent la tribu borélienne de  $E$  quand  $r$  parcourt  $[0, 1]$  et  $n$  décrit  $\mathbf{N}$ . Montrer que si on met  $k$  points répartis uniformément dans  $E$  la loi de

$$(N(B_{1,n}^1), \dots, N(B_{n-1,n}^1))$$

est celle que l'on vient de trouver. En déduire (en utilisant l'exercice 5.28) une façon de simuler un processus Poisson ponctuel d'intensité  $m$ .

9. Dans l'avant-dernière question, que se passe-t-il si on change  $m$  en une constante fois  $m$  ?
10. Calculer  $\mathbf{E}[e^{-sN(A)}]$  pour tout borélien. Pour  $f$  fonction mesurable positive de  $E$  dans  $\mathbf{R}^+$ , on pose

$$N(f)(\omega) = \sum_{\xi \in \omega} f(\xi).$$

Calculer  $\mathbf{E}[e^{-sN(f)}]$ .

11. Chaque point de  $\omega$  est effacé avec probabilité  $p$  et conservé avec probabilité  $1 - p$  et ce indépendamment des autres. On appelle  $N_p(A)$  le nombre de points qui restent dans  $A$  après l'opération d'effacement. Montrer que  $(E, \mathbf{P}, N_p)$  est un processus de Poisson ponctuel d'intensité  $(1 - p)m$ . Calculer  $\mathbf{E}[e^{-sN(A)}]$  pour tout borélien.

EXERCICE 5.19 (Processus de Poisson).—

L'un des modèles stochastiques les plus utilisés est le processus de Poisson. Nous allons ici le décrire et exhiber quelques unes de ses propriétés. Soit  $(S_n, n \geq 1)$  une suite de v.a.r. indépendantes, identiquement distribuées, de loi exponentielle de paramètre  $\lambda$ . On note

$$T_1 = S_1 \text{ et } T_{n+1} = T_n + S_{n+1}.$$

Les instants  $(T_n, n \geq 1)$  sont usuellement vus comme des instants d'arrivée. Les durées  $S_n$  s'appellent logiquement inter-arrivées. On pose

$$N_t = \sum_{n=1}^{+\infty} \mathbf{1}_{[0,t]}(T_n).$$

1. Calculer la loi de  $(T_1, \dots, T_n)$ .

2. Calculer la loi de  $T_n$ .
3. Montrer que  $(N_t = k) = (T_n \leq t < T_{n+1})$ .
4. Calculer la loi de  $N_t$ .
5. Soit  $W_t = t - T_{N_t}$  et  $Z_t = T_{N_t+1} - t$ . Calculer la loi de  $(W_t, Z_t)$ .
6. Montrer que  $W_t$  et  $Z_t$  sont indépendantes et que  $Z_t$  suit une loi exponentielle de paramètre  $\lambda$ .
7. En quoi, ce résultat est-il surprenant ?

EXERCICE 5.20. —

Soit  $W$  une v.a. de loi de Poisson de paramètre  $\lambda > 0$  :

$$\mathbf{P}(W = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

1. Montrer que pour toute fonction positive  $f$  :

$$\lambda \mathbf{E}[f(W + 1)] = \mathbf{E}[W f(W)]. \quad (5.13)$$

2. Réciproquement, soit  $W$  une v.a. discrète, à valeurs dans  $\mathbf{N}$ , telle que pour toute fonction positive, l'identité 5.13 soit satisfaite. En appliquant 5.13 à des fonctions  $f$  judicieusement choisies, montrer que

$$\mathbf{P}(W = j) = \frac{\lambda}{j} \mathbf{P}(W = j - 1),$$

pour tout  $j \geq 1$ .

3. En déduire la loi de  $W$ .

EXERCICE 5.21. —

On tire un nombre  $X$  uniformément sur  $[0, 1]$ . On tire ensuite des nombres  $Y_1, Y_2, \dots$  indépendamment les uns des autres et indépendamment de  $X$ , uniformément sur  $[0, 1]$ . Le jeu s'arrête dès que  $Y_i > X$ . Vous gagnez alors  $(i - 1)\text{€}$ . On appelle  $G$  le gain. Pour  $k$  entier, on définit

$$\varphi_k(x, y_1, \dots, y_{k+1}) = \begin{cases} \mathbf{1}_{\{y_1 > x\}} & \text{si } k = 0 \\ \mathbf{1}_{\{y_1 \leq x, \dots, y_k \leq x, y_{k+1} > x\}} & \text{si } k > 0. \end{cases}$$

1. Pour  $k$  entier, montrer que

$$\int_{[0,1]^{k+2}} \varphi_k(x, y_1, \dots, y_{k+1}) dy_1 dy_2 \dots dy_{k+1} dx = \frac{1}{k+1} - \frac{1}{k+2}.$$

*On traitera séparément les cas  $k = 0$  et  $k > 0$ .*

2. Calculer la loi de  $G$ .
3. Calculer l'espérance de  $G$ .

## EXERCICE 5.22. –

Pour tout  $a$  réel strictement positif,  $G_a$  désigne une variable aléatoire de loi gamma de paramètres  $(a, 1)$  : la densité  $g_a$  de sa loi est donnée par

$$g_a(x) = \frac{1}{\Gamma(a)} x^{a-1} e^{-x} \mathbf{1}_{\mathbf{R}^+}(x),$$

où

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

En particulier,  $G_1$  suit une loi exponentielle de paramètre 1. On admet que

$$\mathbf{E} \left[ e^{itG_a} \right] = (1 - it)^{-a}, \text{ pour tout } t \in \mathbf{R}.$$

De plus, pour  $a, b$  réels strictement positifs,  $B_{a,b}$  désigne une variable aléatoire de loi bêta de paramètres  $(a, b)$  : la densité  $h_{a,b}$  de sa loi est donnée par

$$h_{a,b}(y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1} \mathbf{1}_{[0,1]}(y).$$

1. Calculer la loi du couple  $(G_{a+b}B_{a,b}, G_{a+b})$  lorsque les v.a.  $G_{a+b}$  et  $B_{a,b}$  sont indépendantes.
2. En déduire que pour deux variables  $G_{a+b}, B_{a,b}$  indépendantes, la loi de  $B_{a,b}G_{a+b}$  est identique à celle de  $G_a$ .
3. Soit  $n \geq 0$ . Montrer par récurrence, que lorsque les variables aléatoires  $B_{a,1}, \dots, B_{a+n,1}, G_{a+n+1}$  sont indépendantes, la loi de

$$P_n = G_{a+n+1} \prod_{j=0}^n B_{a+j,1}$$

est la même que celle de  $G_a$ .

*On utilisera la question précédente et les hypothèses d'indépendance. On évitera les longs calculs.*

4. Soit  $X$  une v.a. de loi exponentielle de paramètre 1 indépendante de  $G_a$ , montrer que  $G_a + X$  a la même loi que  $G_{a+1}$ .
5. En déduire que pour tout entier  $n$ ,  $G_{a+n}$  a même loi que

$$H_n = G_a + X_1 + X_2 + \dots + X_n,$$

où les  $X_i$  sont des v.a. dont on précisera les propriétés.

On pose  $W_n = G_a + X_1 + X_2 + \dots + X_n$  où les  $X_i$  sont indépendantes, identiquement distribuées de loi exponentielle de paramètre 1. On suppose de plus que les v.a.  $G_a$  et  $\{X_k, k \geq 1\}$  sont définies sur le même espace de probabilité.

6. Quelle est la limite presque-sûre de  $(n^{-1}W_n, n \geq 1)$  ?
7. Montrer que la suite  $(n^{-1}G_{a+n}, n \geq 1)$  converge en loi, vers une loi que l'on précisera.

## EXERCICE 5.23. –

On rappelle que

$$\int_0^1 u^{-1/2}(1-u)^{-1/2} du = \pi.$$

Soit  $X = (X_1, X_2)$  un vecteur gaussien de  $\mathbf{R}^2$ , centré, de matrice de covariance (ou dispersion)  $\Gamma = I$ . On pose

$$U = \frac{X_1^2}{X_1^2 + X_2^2} \text{ et } V = X_1^2 + X_2^2.$$

1. Calculer la densité de la loi de  $(U, V)$ .
2. Donner les densités marginales de  $U$  et  $V$ . On précisera les constantes de normalisation.
3. Soit  $Z = \frac{X_2^2}{X_1^2}$ . Exprimer  $Z$  en fonction de  $U$  puis calculer la densité de la loi de  $Z$ .

On note  $R_\theta$  la rotation d'angle  $\theta$  dans  $\mathbf{R}^2$ . Si  $x \in \mathbf{R}^2$ ,

$$R_\theta x = \begin{pmatrix} x_1 \cos \theta - x_2 \sin \theta \\ x_1 \sin \theta + x_2 \cos \theta \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

où  $x_1$  et  $x_2$  sont les composantes de  $x$  dans la base canonique de  $\mathbf{R}^2$ .

Soit  $X = (X_1, X_2)$  une v.a. à valeurs dans  $\mathbf{R}^2$  telle que pour tout  $\theta \in [-\pi, \pi]$ ,  $R_\theta X$  a même loi que  $X$ . C'est-à-dire que

$$\mathbf{E}[g(R_\theta X)] = \mathbf{E}[g(X)], \quad (5.14)$$

pour toute fonction  $g$  mesurable bornée de  $\mathbf{R}^2$  dans  $\mathbf{R}$ . On suppose que la loi de  $X$  a une densité par rapport à la mesure de Lebesgue, notée  $v$ .

4. Montrer que pour toute fonction  $g$  mesurable bornée de  $\mathbf{R}^2$  dans  $\mathbf{R}$ , pour tout  $\theta \in [-\pi, \pi]$ ,

$$\int_{\mathbf{R}^2} g(x)v(x)dx = \int_{\mathbf{R}^2} g(y)v(R_\theta y)dy.$$

On admet qu'alors il existe  $w : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ , mesurable, telle que

$$v(x) = w(\|x\|) \text{ pour tout } x \in \mathbf{R}^2.$$

5. Montrer que dans ce cas,

$$\int_0^{+\infty} w(r) r dr = \frac{1}{2\pi}.$$

On suppose maintenant que  $X = (X_1, X_2)$  est un vecteur gaussien centré de matrice de covariance (ou dispersion)  $\Gamma$ .

6. Soit  $\theta \in [-\pi, \pi]$ , quelle est la loi de  $R_\theta X$  ?
7. Montrer que  $R_\theta X$  a même loi que  $X$  pour tout  $\theta$  si et seulement si  $\Gamma R_\theta = R_\theta \Gamma$ .

8. Supposons que  $\Gamma R_\theta = R_\theta \Gamma$  pour tout  $\theta \in [-\pi, \pi]$ . En écrivant les équations satisfaites par les coefficients de  $\Gamma$ , montrer que  $\Gamma$  est la matrice d'une homothétie positive (c'est-à-dire qu'il existe  $\sigma^2$  tel que  $\Gamma = \sigma^2 I$ ).

EXERCICE 5.24.—

Soit  $N$  un processus de Poisson (cf. exercice 5.19) d'intensité  $\lambda$ , on note  $T_n$  le  $n$ -ième instant de saut. Par convention,  $T_0 = 0$ . Soit  $(Z_n, n \geq 1)$ , une suite de variables aléatoires de même loi telles que pour tout  $n$ ,  $T_n$  et  $Z_n$  sont indépendantes. Soit  $g$  la densité de la loi commune aux  $Z_n$ .

1. Montrer que pour toute fonction  $f$ ,

$$E[f(T_n, Z_n)] = \int_0^{+\infty} \int f(t, z) g(z) \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} dz dt.$$

2. En déduire que

$$E\left[\sum_{n \geq 1} f(T_n, Z_n)\right] = \lambda \int_0^{+\infty} \int f(t, z) g(z) dz dt.$$

3. On suppose que les communications téléphoniques d'un abonné durent un temps aléatoire de loi exponentielle de moyenne 3 minutes. Ces durées sont indépendantes entre elles. Au siècle dernier, le coût d'une communication était fonction de sa durée  $t$  selon la formule suivante :

$$c(t) = \alpha \text{ si } t \leq t_0, \text{ et } c(t) = \alpha + \beta(t - t_0) \text{ si } t \geq t_0.$$

Déduire de ce qui précède que le coût moyen d'une heure totale de communication est donné par :

$$\lambda \int_0^1 c(t) \lambda e^{-\lambda t} dt$$

avec  $\lambda = 20$ . (Indication : Considérer  $Z_n = T_{n+1} - T_n$  et expliquer pourquoi on peut appliquer le résultat précédent.)

EXERCICE 5.25.—

Soit  $N$  un processus de Poisson sur  $\mathbf{R}^+$ . Soit  $f: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ . Considérons

$$\int f(s) dN_s = \sum_{n \geq 1} f(T_n).$$

1. Montrer que  $N_t - N_s$  a même loi que  $N_{t-s}$  pour tout couple  $(t, s)$  avec  $t \geq s$ .  
2. Montrer que

$$\mathbf{E} \left[ \exp \left( - \int \mathbf{1}_{]a, b]}(s) dN_s \right) \right] = \exp \left( - \int 1 - e^{-\mathbf{1}_{]a, b]}(s) \lambda ds \right).$$

3. En déduire  $\mathbf{E}[\exp(-\int f(s) dN_s)]$  pour toute fonction  $f$  positive.  
4. Pour  $B \subset \mathbf{R}^+$ , calculer de deux manières différentes

$$\frac{d}{dt} \mathbf{E} \left[ \exp \left( - \int (f + t \mathbf{1}_B)(s) dN_s \right) \right]_{t=0}.$$

EXERCICE 5.26 (Somme aléatoire).—

Soient  $X = (X_n, n \geq 1)$  une suite de v.a. indépendantes de loi exponentielle de paramètre  $\lambda$ . Soit  $N$  une v.a. indépendante de la suite  $X$  de loi géométrique de paramètre  $\rho$ . Calculer la loi de  $Z$  où

$$Z = \sum_{j=1}^N X_j.$$

EXERCICE 5.27.—

Soient  $X, Y$  deux v.a. indépendantes suivant la loi  $\mathcal{N}(0, 1)$ . Caractériser la loi du vecteur  $(X + Y, X - Y)$ .

EXERCICE 5.28.—

Comment simuler le tirage de points uniformément répartis dans un triangle scalène en utilisant le moins possible le générateur de nombres pseudo-aléatoires. Même question avec un disque.

EXERCICE 5.29.—

Soit  $D$  une variable aléatoire de loi uniforme sur  $[0, 3]$ , c'est-à-dire

$$dP_D(x) = \frac{1}{3} \mathbf{1}_{[0,3]}(x) dx.$$

Soient  $s$  et  $t$  deux réels positifs tels que  $0 \leq t + s \leq 3$ .

1. Pour  $x \in [0, 3]$ , simplifier l'expression  $(t - (x - s)^+)^+$  où  $x^+ = \max(x, 0)$ .
2. Calculer la loi de  $R = (t - (D - s)^+)^+$ .

EXERCICE 5.30.—

Soit  $(X_1, X_2)$ , une variable aléatoire à valeurs dans  $\mathbf{R}^2$  et  $N$  une deuxième variable aléatoire indépendante de  $(X_1, X_2)$  et de loi  $\alpha\delta_1 + (1 - \alpha)\delta_2$ , où  $\alpha \in ]0, 1[$ .

1. Calculer  $E[X_N], \sigma_{X_N}^2$  en termes de celles de  $X_1$  et de  $X_2$ .
2. On suppose que  $X_1$  et  $X_2$  sont indépendantes et de même loi, calculer la loi de  $X_N$ .

EXERCICE 5.31.—

Trois personnes  $A, B$  et  $C$  arrivent à la poste en même temps pour téléphoner. Il y a deux cabines téléphoniques qu'occupent  $A$  et  $B$  tout de suite.  $C$  remplace le premier sorti. On désigne par  $X_1, X_2, X_3$  les temps d'occupation de la cabine par  $A, B$  et  $C$  respectivement. On suppose que  $(X_1, X_2, X_3)$  sont indépendantes, de même loi exponentielle de paramètre  $\alpha$ .

1. Calculer la probabilité que  $C$  sorte le dernier.
2. Donner la loi du temps  $T$  passé par  $C$  à la poste.
3. Donner la loi de probabilité de l'instant du dernier départ ; l'instant 0 étant l'instant d'arrivée des trois personnes à la poste.

EXERCICE 5.32 (Castor et Pollux).—

Castor et Pollux se sont donnés rendez-vous en convenant de ne pas attendre l'autre plus de dix minutes. Ils arrivent tous les deux indépendamment à un instant « au hasard » entre midi et 13 heures. On note  $X$ , respectivement  $Y$ , l'heure d'arrivée de Castor, respectivement celle de Pollux. On note  $W$  le temps d'attente de Castor.

1. Quelle est la probabilité qu'ils se rencontrent ?
2. Exprimer en fonction de  $X$  et  $Y$ , la valeur du temps d'attente de Castor. On pourra utilement faire un dessin en identifiant dans le pavé  $[0, 1] \times [0, 1]$ , différentes zones où l'expression de  $W$  est simple – voir Figure 5.2.

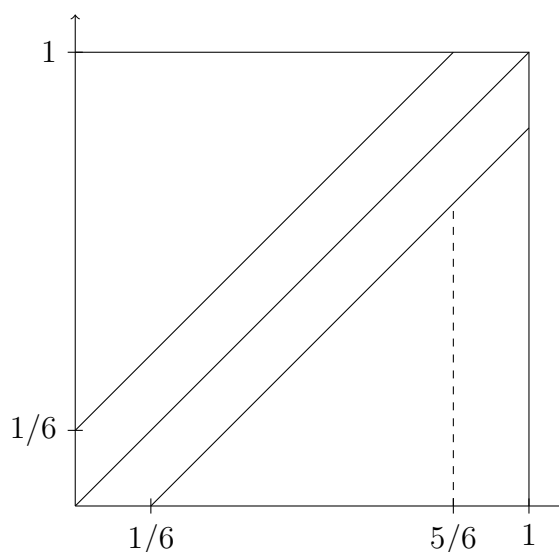


Figure 5.2. – Castor et Pollux

3. Quelle est la loi du temps d'attente de Castor ?
4. Quel est le temps d'attente moyen de Castor ?
5. Quelle est la loi du temps d'attente de Castor sachant qu'il y a rencontre ?

EXERCICE 5.33. –

Soit

$$d\mathbf{P}(x, y) = c \exp\left(-\frac{x^2 + y^2}{2}\right) \mathbf{1}_{\{x > y\}} dx dy$$

une mesure sur le plan  $\mathbf{R}^2$ .

1. Trouver la constante  $c$  pour que  $\mathbf{P}$  soit une probabilité.
2. Soit  $(\Omega, \mathcal{F}, \mathbf{P})$  un espace de probabilité et  $(X, Y) : \Omega \rightarrow \mathbf{R}^2$  une variable aléatoire de loi  $\mathbf{P}$ . Trouver la loi de  $X$  et celle de  $Y$ .
3. Sont-elles indépendantes ?
4. On définit les nouvelles variables aléatoires  $U = X^2 + Y^2$  et  $V = Y$ . Calculer la loi du vecteur  $(U, V)$ .
5. Les variables  $U$  et  $V$  sont-elles indépendantes ?

EXERCICE 5.34. –

Soient  $X$  et  $Y$  deux v.a. réelles indépendantes sur  $(\Omega, \mathcal{F}, \mathbf{P})$ , de même loi uniforme sur  $[0, a]$  ( $a > 0$  réel, fixé). On note par  $R = \sqrt{X^2 + Y^2}$ ,  $Z = Y/X$  et par  $\mathbf{P}_a$  une nouvelle probabilité définie par

$$\mathbf{P}_a(A) = \mathbf{P}(A \mid R < a),$$

pour tout  $A \in \mathcal{F}$ .

1. Pour tout borélien  $B$  de  $[0, a]^2$ , exprimer  $\mathbf{P}((X, Y) \in B)$  à l'aide de la surface  $S(B)$  de  $B$ .
2. Montrer que  $R$  et  $Z$  sont indépendantes pour la probabilité  $\mathbf{P}_a$  mais pas pour  $\mathbf{P}$ .
3. Trouver deux fonctions simples  $f$  et  $g$  telles que pour  $\mathbf{P}_a$ ,  $f(R)$  et  $g(Z)$  soient uniformes ; sont-elles indépendantes ?

EXERCICE 5.35. —

Soient  $X$  et  $Y$  deux v.a. indépendantes de loi uniforme sur  $[0, 1]$ .

1. Quelle est la loi du couple  $(X, Y)$  ?
2. Quelle est la loi du couple  $(\min(X, Y), \max(X, Y))$  ?

EXERCICE 5.36. —

Soient  $Z = (X, Y)$  la loi de densité  $\pi^{-1} \mathbf{1}_D(x, y)$  où  $D$  est le disque unité de  $\mathbf{R}^2$ .

1. Calculer les lois marginales de  $X$  et  $Y$ .
2. Ces deux variables sont-elles indépendantes ?
3. Calculer la loi du couple  $(\min(X, Y), \max(X, Y))$ .

## Pour aller plus loin

EXERCICE 5.37. —

Soit  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré. Soit  $I$  un intervalle de  $\mathbb{R}$  et une famille de fonctions mesurables  $\{f(\cdot, t)\}_{t \in I}$ ,  $f(\cdot, t) : \Omega \rightarrow \mathbb{R}$ . On suppose que

- $\forall \omega \in \Omega, t \mapsto f(\omega, t)$  est continue sur  $I$ .
- Il existe une application mesurable  $g : \Omega \rightarrow \mathbb{R}^+$  telle que  $\int g \, d\mu < +\infty$  et

$$\forall t \in I, \quad \forall \omega \in \Omega, \quad |f(\omega, t)| \leq g(\omega).$$

Montrer que  $t \mapsto \int f(\omega, t) \, d\mu(\omega)$  est continue sur  $I$ .

EXERCICE 5.38. —

Soit  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré. Soit  $I$  un intervalle de  $\mathbb{R}$  et une famille de fonctions mesurables  $\{f(\cdot, t)\}_{t \in I} : f(\cdot, t) : \Omega \rightarrow \mathbb{R}$ . On suppose que

- $x \mapsto f(x, 0)$  est intégrable.
- $\forall \omega \in \Omega, t \mapsto f(\omega, t)$  est dérivable sur  $I$ .
- Il existe une application mesurable  $g : \Omega \rightarrow \mathbb{R}^+$  telle que  $\int g \, d\mu < +\infty$  et

$$\forall t \in I, \quad \forall \omega \in \Omega, \quad \left| \frac{\partial f(\omega, t)}{\partial t} \right| \leq g(\omega).$$

En utilisant l'inégalité des accroissements finis, montrer que  $t \mapsto \int f(\omega, t) \, d\mu(\omega)$  est bien définie sur  $I$ . Montrer que cette fonction est dérivable et que l'on a

$$\frac{\partial}{\partial t} \int f(\omega, t) \, d\mu(\omega) = \int \frac{\partial f(\omega, t)}{\partial t} \, d\mu(\omega).$$



EXERCICE 5.39. —

Soit  $(E, \mathcal{E}, \mu)$  un espace mesurable et  $T$  une application de  $E$  dans lui-même. On dit que  $\mu$  est invariante par  $T$  si

$$\int_E f \circ T d\mu = \int f d\mu$$

pour toute fonction  $f$  mesurable bornée.

1. Montrer que la mesure de Lebesgue sur  $\mathbf{R}$  est invariante par translation.
2. Soit  $E = \mathbf{R}^n$  et

$$d\mu(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(x_1^2 + \dots + x_n^2)\right) dx_1 \dots dx_n.$$

Montrer que  $\mu$  est invariante par rotation.

3. Soit  $E = [0, 1]$  et  $T(x) = 2x - [2x]$  ( $T(x)$  est la partie fractionnaire de  $x$ ). Montrer que la mesure de Lebesgue restreinte à  $E$  est invariante.

EXERCICE 5.40. —

Montrer que toute mesure de Radon sur  $\mathbf{R}$  (c'est-à-dire  $\mu(K) < +\infty$  quel que soit le compact  $K$ ) invariante par translation est proportionnelle à la mesure de Lebesgue.

EXERCICE 5.41. —

Soit  $(E, \mathcal{E}, \mu)$  un ensemble mesuré,  $(F, \mathcal{F})$  un ensemble et une tribu et  $T$  une application mesurable de  $E$  dans  $F$ . On définit la mesure  $T^*\mu$  (appelée mesure image de  $\mu$  par  $T$ ) par

$$\forall B \in \mathcal{F}, \quad (T^*\mu)(B) = \mu(T^{-1}(B)).$$

ou de manière équivalente par

$$\int_F f d(T^*\mu) = \int_E f \circ T d\mu.$$

pour toute fonction  $f$  mesurable bornée de  $F$  dans  $\mathbf{R}$ . Soit  $E = \mathbf{R}/\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$ , muni de  $\mu$  la mesure uniforme.

1. Montrer que  $\mu$  est invariante par translation.
2. Considérons l'application  $T$  de  $E$  dans  $O_2(\mathbf{R})$  (le groupe des transformations orthogonales de  $\mathbf{R}^2$ ) donnée par :

$$T(\theta, \epsilon) = \begin{pmatrix} \cos 2\pi\theta & \sin 2\pi\theta \\ (-1)^\epsilon \sin 2\pi\theta & (-1)^{1-\epsilon} \cos 2\pi\theta \end{pmatrix}$$

Quelle est la mesure de l'ensemble des symétries (respectivement des rotations d'angle inférieur à  $\theta_0$  donné) sous  $T^*\mu$ ?

3. Montrer que  $T^*\mu$  est invariante par translation.
4. On considère  $S$  l'application de  $O_2(\mathbf{R})$  dans  $\mathbb{C}$  qui à une transformation orthogonale associe la valeur propre de plus grandes parties réelle et imaginaire. Décrire  $S^*(T^*\mu)$ .

EXERCICE 5.42. —

Soient  $a, m \in \mathbb{R}^d$  et  $\Sigma$  une matrice  $d \times d$  définie positive. Soit  $A$  une matrice  $d \times d$ . Si  $X \sim \mathcal{N}_d(m, \Sigma)$ , quelle est la loi de  $a + AX$ ? En déduire que

si  $X \sim \mathcal{N}_d(m, \Sigma)$  alors  $\sqrt{\Sigma}^{-1}(X - m) \sim \mathcal{N}_d(0, I)$ .  
 si  $X \sim \mathcal{N}_d(0, I)$ , alors  $m + \sqrt{\Sigma}X \sim \mathcal{N}_d(m, \Sigma)$ .

EXERCICE 5.43. —

Soient  $X, Y$  deux v.a.r. de loi jointe  $f_{X,Y}$  sur  $\mathbb{R}^2$ . Exprimer la densité de probabilité de  $X + Y$  en fonction de  $f_{X,Y}$ . Dans le cas où  $X$  et  $Y$  sont indépendantes, montrer cette densité est égale au produit de convolution des densités marginales.

EXERCICE 5.44. —

En quoi la fonction définie sur  $[0, 1] \times [0, 1]$  par  $(x^2 - y^2)/(x^2 + y^2)^2$  montre-t-elle que les hypothèses du théorème de Fubini sont optimales ?

EXERCICE 5.45. —

Soit  $X_1, \dots, X_n$  des variables aléatoires indépendantes et identiquement distribuées, de loi uniforme sur  $[0, \theta]$ , avec  $\theta > 0$ . On pose :

$$\hat{\theta}_n = \max(X_1, \dots, X_n).$$

1. Calculer la fonction de répartition de  $\hat{\theta}_n$ .
2. En déduire que la loi de  $\hat{\theta}_n$  admet une densité de la forme :

$$f_{\hat{\theta}_n}(x) = C_n x^{n-1} \mathbf{1}_{[0, \theta]}(x),$$

où  $C_n$  est une constante que l'on précisera.

3. Calculer l'espérance de  $\hat{\theta}_n$  et montrer que celle-ci tend vers  $\theta$  lorsque  $n \rightarrow +\infty$ .
4. Déterminer la loi de la variable aléatoire  $n(\theta - \hat{\theta}_n)$ .
5. Montrer celle-ci tend vers une loi exponentielle lorsque  $n \rightarrow +\infty$ .
6. On utilise  $\hat{\theta}_n$  pour estimer le paramètre  $\theta$ , supposé inconnu. Déduire de la question précédente une valeur approchée de la probabilité que l'erreur relative de l'estimation, soit  $\left| \frac{\hat{\theta}_n - \theta}{\theta} \right|$ , soit supérieure à 1% pour  $n = 100$ .

EXERCICE 5.46. —

Soit  $(E, \mathcal{E}, \mathbf{P})$  un espace probabilisé et  $T$  une application mesurable de  $E$  dans lui-même. On suppose que  $\mathbf{P}$  est invariante par  $T$ , c'est-à-dire que

$$\mathbf{P}(T^{-1}(A)) = \mathbf{P}(A) \text{ pour tout } A \in \mathcal{E}.$$

1. Montrer que l'ensemble des mesurables invariants par  $T$ , c'est-à-dire qui vérifie  $T^{-1}(A) = A$ , est une tribu (notée  $\mathcal{I}$  par la suite).
2. Soit  $f$  une fonction mesurable de  $E$  dans  $\mathbf{R}$ . Montrer que si  $f$  est invariante par  $T$  (c'est-à-dire  $f \circ T = f$ ) alors  $f$  est mesurable de  $(E, \mathcal{I})$  dans  $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ .
3. Le système dynamique  $(E, T, \mathbf{P})$  est dit ergodique lorsque

$$\mathcal{I} \subset \sigma\{A \in \mathcal{E}, \mathbf{P}(A) = 0 \text{ ou } \mathbf{P}(A) = 1\}.$$

Montrer que  $(E, T, \mathbf{P})$  est ergodique si et seulement si les fonctions invariantes par  $T$  sont constantes presque partout.

4. On dit que  $T$  est mélangeante si et seulement si pour tout couple  $f, g$  d'éléments de  $L^2(d\mathbf{P})$ ,

$$\lim_{n \rightarrow +\infty} \int_E f \circ T^n g d\mathbf{P} = \int_E f d\mathbf{P} \int_E g d\mathbf{P}. \quad (5.15)$$

Montrer que si  $T$  est mélangeante alors  $(E, T)$  est ergodique.

5. Montrer que si la condition de mélange (5.15) est vérifiée pour  $f, g$  appartenant à un sous-ensemble dense de  $L^2(d\mathbf{P})$  alors  $T$  est mélangeante.

On veut maintenant étudier le système dynamique donnée par l'équation d'évolution :

$$x_{n+1}^a = T(x_n^a) \text{ où } T(x) = 4x(1-x), \quad x_0^a = a \in [0, 1].$$

On veut montrer en particulier que pour presque tout  $a \in [0, 1]$ ,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^n f(x_j^a) = \int_0^1 f(u) (\pi \sqrt{u} \sqrt{1-u})^{-1} du.$$

On admet que si  $(E, T, \mathbf{P})$  est un système ergodique alors

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^n f \circ T^j(x) = \int_E f d\mathbf{P}$$

pour presque tout  $x$ . Il nous faut donc trouver une mesure invariante  $\mu$  par  $T$  et montrer que le système dynamique  $([0, 1], T, \mu)$  est ergodique. Pour ce faire on considère un autre système dynamique :

$$E_1 = [0, 1[, \quad T_1 x = 2x \text{ si } 0 \leq x \leq 1/2, \quad T_1(x) = 2 - 2x \text{ pour } 1/2 \leq x < 1.$$

(où  $[x]$  est la partie entière de  $x$ ) muni de la mesure de Lebesgue sur  $[0, 1[$ , notée  $\lambda$ .

1. Montrer que  $\lambda$  est invariante par  $T_1$ .
2. En admettant (ou se souvenant, cf. séries de Fourier) que la famille de fonctions  $e_k(x) = \exp(2i\pi kx)$  pour  $k \in \mathbf{Z}$  est une famille dense de  $L^2(d\lambda)$ , montrer que  $T_1$  est mélangeante.
3. Soit  $\Theta$  l'application de  $E_1$  dans  $[0, 1]$  définie par :

$$\Theta(x) = \sin^2(\pi x/2).$$

4. Identifier  $\mu$  la mesure image de  $\lambda$  par  $\Theta$ .
5. Montrer que  $([0, 1[, T, \mu)$  est ergodique et conclure.



## 6. Loi conditionnelle

Lorsque l'on modélise un phénomène, la situation la plus simple se présente lorsque les variables aléatoires en jeu sont toutes indépendantes les unes des autres. Malheureusement, c'est rarement le cas. Il nous faut donc un moyen de représenter des variables aléatoires dépendantes. Cela s'appelle la loi conditionnelle.

A titre d'exemple considérons la loi uniforme sur l'ensemble à 12 éléments suivant :

$$T = \{(i, j) \in \mathbf{N} \times \mathbf{N}, i + 2j \leq 5\}.$$

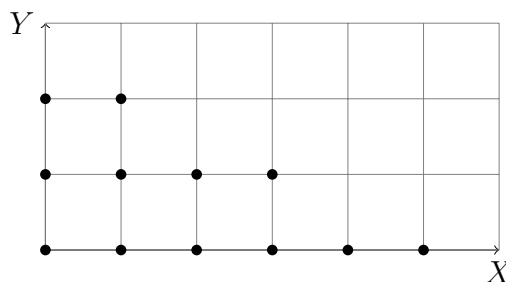


Figure 6.1. – L'ensemble  $T$

On note  $X$  l'abscisse et  $Y$  l'ordonnée. Si l'on sait que  $X = 0$ , alors  $Y$  peut prendre les valeurs 0, 1, 2 tandis que si  $X = 2$ ,  $Y$  ne peut prendre que les valeurs 0 et 1. Connaître  $X$  donne donc une information sur  $Y$  qui se quantifie ainsi :

$$\mathbf{P}(Y = 1 | X = 0) = \frac{\mathbf{P}(Y = 1, X = 0)}{\mathbf{P}(X = 0)} = \frac{1/12}{3/12} = \frac{1}{3}.$$

De même,

$$\mathbf{P}(Y = 0 | X = 0) = \mathbf{P}(Y = 2 | X = 0) = \frac{1}{3}.$$

On résume ceci en disant que la loi de  $Y$  conditionnellement à  $X = 0$  est la loi uniforme sur  $\{0, 1, 2\}$ , ce qui formellement s'écrit :

$$\mathbf{P}_{Y|X=0} = \frac{1}{3}\delta_0 + \frac{1}{3}\delta_1 + \frac{1}{3}\delta_2.$$

Le même raisonnement permet de montrer que la loi conditionnelle de  $Y$  sachant  $X = 5$  est celle d'une variable aléatoire constante égale à 0, c'est-à-dire

$$\mathbf{P}_{Y|X=5} = \delta_0.$$

## 6.1. Cas général

Evidemment les affaires se corsent si la loi de  $X$  est à densité puisqu'alors l'événement  $(X = x)$  est de mesure nulle. Pourtant, on peut très bien imaginer un couple de points répartis au hasard dans le triangle  $T'$  :

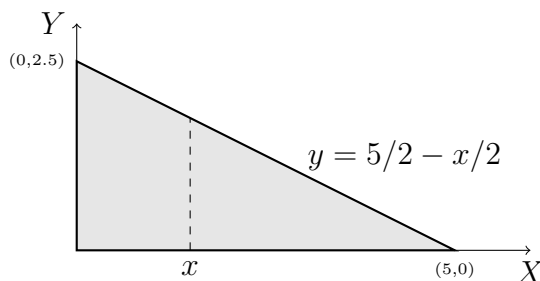


Figure 6.2. – L'ensemble  $T'$

Dans ces conditions, on a bien envie de dire que conditionnellement à  $X = x$ , la loi de  $Y$  est la loi uniforme sur  $[0, 5/2 - x/2]$ .

Pour ce faire, il nous faut introduire la notion de noyau de probabilités.

**Définition 6.1** (Noyau de probabilités). Un noyau de probabilités  $\mathcal{Q}$  est une application

$$\begin{aligned} \mathbf{R}^n \times \mathcal{B}(\mathbf{R}^p) &\longrightarrow [0, 1] \\ (x, A) &\longmapsto \mathcal{Q}_x(A) \end{aligned}$$

telle que :

— Pour tout  $x \in \mathbf{R}^n$ , l'application

$$\begin{aligned} \mathcal{B}(\mathbf{R}^p) &\longrightarrow [0, 1] \\ A &\longmapsto \mathcal{Q}_x(A) \end{aligned}$$

est une probabilité sur l'ensemble des boréliens de  $\mathbf{R}^p$ .

— Pour tout  $A \in \mathcal{B}(\mathbf{R}^p)$ , l'application

$$\begin{aligned} (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n)) &\longrightarrow [0, 1] \\ x &\longmapsto \mathcal{Q}_x(A) \end{aligned}$$

est mesurable.

A partir de là, nous pouvons définir la notion de loi conditionnelle. Rappelons que le support d'une probabilité, est le complémentaire du plus grand ouvert de probabilité nulle.

**Définition 6.2** (Loi conditionnelle). Soit  $X$  une variable aléatoire dans  $\mathbf{R}^n$  et  $Y$  une variable aléatoire dans  $\mathbf{R}^p$ . La loi conditionnelle de  $Y$  sachant  $X$  est l'unique famille de noyaux  $(\mathbf{P}_{Y|X=x}, x \in \text{support } \mathbf{P}_X)$  définie par

$$\mathbf{E} [\psi(X)\theta(Y)] = \mathbf{E} \left[ \psi(X) \int_{\mathbf{R}^p} \theta(y) \, d\mathbf{P}_{Y|X=x}(y) \right] \quad (6.1)$$

pour toutes fonctions  $\theta$  et  $\psi$  mesurables bornées. En forme intégrale, cela donne

$$\mathbf{E} [\psi(X)\theta(Y)] = \int_{\mathbf{R}^n} \psi(x) \int_{\mathbf{R}^p} \theta(y) \, d\mathbf{P}_{Y|X=x}(y) \, d\mathbf{P}_X(x).$$

Nous ne nous attarderons pas sur l'existence et l'unicité de cette famille qui sont garantis par des théorèmes généraux dans des espaces bien plus ésotériques que  $\mathbf{R}^n$  et  $\mathbf{R}^p$  [Kal98]. En tout état de cause, la relation ci-dessus permet de déterminer  $d\mathbf{P}_{Y|X=x}$  mais elle est généralisable à des fonctions qui ne sont pas à variables séparées.

**Théorème 6.1.** Soit  $X$  une variable aléatoire dans  $\mathbf{R}^n$  et  $Y$  une variable aléatoire dans  $\mathbf{R}^p$ . Soit  $(\mathbf{P}_{Y|X=x}, x \in \text{support } \mathbf{P}_X)$  définie comme ci-dessus. Pour toute fonction  $\psi : \mathbf{R}^n \times \mathbf{R}^p \rightarrow \mathbf{R}$ , mesurable bornée, on a

$$\mathbf{E} [\psi(X, Y)] = \int_{\mathbf{R}^n \times \mathbf{R}^p} \psi(x, y) d\mathbf{P}_{Y|X=x}(y) \, d\mathbf{P}_X(x).$$

Il y a deux cas particuliers dans lesquels le calcul de la loi conditionnelle est particulièrement simple.

**Théorème 6.2.** Soit  $X$  et  $Y$  deux variables aléatoires indépendantes. Pour tout  $x \in \text{supp } \mathbf{P}_X$ ,  $\mathbf{P}_{Y|X=x} = \mathbf{P}_Y$ .

*Démonstration.* La méthode générale est d'identifier  $\mathbf{P}_{Y|X=x}$  à travers la relation (6.1). Pour toute paire de fonctions bornées  $\psi$  et  $\theta$ ,

$$\begin{aligned} \mathbf{E} [\psi(X)\theta(Y)] &= \mathbf{E} [\psi(X)] \mathbf{E} [\theta(Y)] \\ &= \int \psi(x) \, d\mathbf{P}_X(x) \int \theta(y) \, d\mathbf{P}_Y(y) \\ &= \int \psi(x) \left( \int \theta(y) \, d\mathbf{P}_Y(y) \right) \, d\mathbf{P}_X(x), \end{aligned}$$

où l'on a simplement rentré l'intégrale en  $y$  dans l'intégrale en  $x$  puisque c'est une constante. Si  $\mathbf{P}_{Y|X=x} = \mathbf{P}_Y$ , on a bien

$$\mathbf{E} [\psi(X)\theta(Y)] = \int \psi(x) \left( \int \theta(y) \, d\mathbf{P}_{Y|X=x}(y) \right) \, d\mathbf{P}_X(x).$$

D'où le résultat. □

**Théorème 6.3.** Soit  $X$  et  $Y$  deux variables aléatoires telles que  $Y = T(X)$ . Pour tout  $x \in \text{supp } \mathbf{P}_X$ ,  $\mathbf{P}_{Y|X=x} = \delta_{T(x)}$ .

*Démonstration.* Pour toute paire de fonctions bornées  $\psi$  et  $\theta$ ,

$$\begin{aligned} \mathbf{E}[\psi(X)\theta(Y)] &= \int_{\mathbf{R}^n} \psi(x)\theta(T(x)) \, d\mathbf{P}_X(x) \\ &= \int_{\mathbf{R}^n} \psi(x) \left( \int \psi(y) \, d\delta_{T(x)}(y) \right) \, d\mathbf{P}_X(x). \end{aligned}$$

Encore une fois, par identification, on obtient  $\mathbf{P}_{Y|X=x} = \delta_{T(x)}$ .  $\square$

## 6.2. Calculs

**Théorème 6.4.** Soit  $X$  et  $Y$  deux variables aléatoires discrètes. Pour tout  $x \in \text{support } \mathbf{P}_X$ ,

$$\mathbf{P}_{Y|X=x} = \sum_{y \in \text{supp } \mathbf{P}_Y} \mathbf{P}(Y = y | X = x) \delta_y.$$

Rappelons que l'intégrale d'une fonction  $\theta$  de  $Y$  par rapport à  $\mathbf{P}_{Y|X=x}$  est définie par

$$\int \theta(y) \, d\mathbf{P}_{Y|X=x}(y) = \sum_{y \in \text{supp } \mathbf{P}_Y} \theta(y) \mathbf{P}(Y = y | X = x). \quad (6.2)$$

*Démonstration.* On procède par identification. L'idée est de faire apparaître la loi de  $X$  au forceps puis de compenser cette modification par une multiplication par une bonne fonction (encore une occurrence du célèbre théorème «  $1 = a/a$  pour  $a \neq 0$  », d'où le besoin de se restreindre à des valeurs de  $x$  dans le support de  $\mathbf{P}_X$ ). Pour toute paire de fonctions  $\psi$  et  $\theta$  bornées, on a

$$\begin{aligned} \mathbf{E}[\psi(X)\theta(Y)] &= \sum_{x \in \text{supp } \mathbf{P}_X} \sum_{y \in \text{supp } \mathbf{P}_Y} \psi(x)\theta(y)\mathbf{P}(X = x, Y = y) \\ &= \sum_{x \in \text{supp } \mathbf{P}_X} \psi(x) \left( \sum_{y \in \text{supp } \mathbf{P}_Y} \theta(y) \frac{\mathbf{P}(X = x, Y = y)}{\mathbf{P}(X = x)} \right) \mathbf{P}(X = x) \\ &= \sum_{x \in \text{supp } \mathbf{P}_X} \psi(x) \left( \sum_{y \in \text{supp } \mathbf{P}_Y} \theta(y) \mathbf{P}(Y = y | X = x) \right) \mathbf{P}(X = x). \end{aligned}$$

On reconnaît bien dans le terme entre parenthèse, l'intégrale de  $\theta$  selon la mesure  $\mathbf{P}_{Y|X=x}$ , voir (6.2). D'où le résultat.  $\square$



**Théorème 6.5.** Soit  $X : \Omega \rightarrow \mathbf{R}^n$  et  $Y : \Omega \rightarrow \mathbf{R}^p$ , de densité jointe  $f_{X,Y}$ . La loi conditionnelle de  $Y$  sachant  $X$  a pour densité

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad x \in \text{supp } \mathbf{P}_X.$$

On notera le parallèle entre cette démonstration et la précédente.

*Démonstration.* Pour toute paire de fonctions  $\psi$  et  $\theta$  bornées, on a

$$\begin{aligned} \mathbf{E}[\psi(X)\theta(Y)] &= \int_{\mathbf{R}^n \times \mathbf{R}^p} \psi(x)\theta(y)f_{X,Y}(x,y) \, dx \, dy \\ &= \int_{\mathbf{R}^n} \psi(x) \left( \int_{\mathbf{R}^p} \theta(y) \frac{f_{X,Y}(x,y)}{f_X(x)} \, dy \right) f_X(x) \, dx \\ &= \int_{\mathbf{R}^n} \psi(x) \left( \int_{\mathbf{R}^p} \theta(y) \frac{f_{X,Y}(x,y)}{f_X(x)} \, dy \right) d\mathbf{P}_X(x). \end{aligned}$$

Le résultat s'en suit par identification.  $\square$

EXEMPLE 32.— Dans le cas de notre triangle  $T'$ , la densité jointe est

$$f_{X,Y}(x,y) = \frac{1}{6} \mathbf{1}_{T'}(x,y).$$

A  $x$  fixé, le couple  $(x,y)$  appartient à  $T'$  si et seulement si  $y$  appartient à  $[0, 5/2 - x/2]$ . Par conséquent,

$$f_X(x) = \int f_{X,Y}(x,y) \, dy = \frac{1}{6} \int_0^{5/2-x/2} dy = \frac{5/2 - x/2}{6}.$$

En vertu du théorème qui précède, la loi de  $Y$  conditionnellement à  $X = x$  a pour densité

$$f_{Y|X=x}(y) = \frac{\mathbf{1}_{T'}(x,y)/6}{(5/2 - x/2)/6} = \frac{1}{5/2 - x/2} \mathbf{1}_{[0, 5/2-x/2]}(y).$$

Comme prévu, c'est bien la loi uniforme sur l'intervalle  $[0, 5/2 - x/2]$ .

Les autres cas ( $X$  discrète,  $Y$  à densité ou l'inverse) se traitent de manière analogue. Plus généralement, si la loi du couple  $(X,Y)$  admet une densité par rapport à une mesure produit  $\mu \otimes \nu$ , on peut reproduire les mêmes calculs.

**Théorème 6.6.** Soit  $E$  et  $F$  deux espaces munis d'une tribu et d'une mesure finie sur tous les compacts,  $\mu$  pour  $E$  et  $\nu$  pour  $F$ . Soit  $X : \Omega \rightarrow E$  et  $Y : \Omega \rightarrow F$ , de densité jointe  $f_{X,Y}$  par rapport à la mesure  $\mu \otimes \nu$ . La loi conditionnelle de  $Y$  sachant  $X$  est donnée par

$$d\mathbf{P}_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} \, d\nu(y). \quad (6.3)$$

Cela signifie que la loi conditionnelle de  $Y$  sachant  $X = x$  est absolument continue par rapport à  $\nu$  donc la densité est donnée par

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

*Démonstration.* Remarquons d'abord que même dans ce formalisme, la loi de  $X$  a une densité par rapport à  $\mu$ , donnée par

$$f_X(x) = \int_F f_{X,Y}(x, y) d\nu(y) \text{ et } d\mathbf{P}_X(x) = f_X(x) d\mu(x).$$

Pour toute paire de fonctions  $\psi$  et  $\theta$  bornées, on a

$$\begin{aligned} \mathbf{E} [\psi(X)\theta(Y)] &= \int_{E \times F} \psi(x)\theta(y) f_{X,Y}(x, y) d\mu(x) d\nu(y) \\ &= \int_E \psi(x) \left( \int_F \theta(y) \frac{f_{X,Y}(x, y)}{f_X(x)} d\nu(y) \right) f_X(x) d\mu(x) \\ &= \int_E \psi(x) \left( \int_F \theta(y) \frac{f_{X,Y}(x, y)}{f_X(x)} d\nu(y) \right) d\mathbf{P}_X(x). \end{aligned}$$

Le résultat s'en suit par identification.  $\square$

EXEMPLE 33.— On considère  $X$  une variable aléatoire de loi bêta de paramètres  $(a, b)$ , c'est-à-dire de densité

$$f_X(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \mathbf{1}_{[0,1]}(x).$$

On se donne aussi  $Y$  telle que la loi conditionnelle de  $Y$  sachant  $X$  est donnée par

$$\mathbf{P}_{Y|X=x} = x\delta_1 + (1-x)\delta_0, \text{ pour tout } x \in [0, 1], \quad (6.4)$$

en d'autres termes, la loi de  $Y$  sachant  $X = x$  est une loi de Bernoulli de paramètre  $x$ . D'après (6.4), pour tout fonction  $\psi$  bornée,

$$\begin{aligned} \int \psi(y) d\mathbf{P}_{Y|X=x}(y) &= x\psi(1) + (1-x)\psi(0) \\ &= \int_{\{0,1\}} \psi(y) x^y (1-x)^{1-y} d\mu(y), \end{aligned}$$

où  $\mu = \delta_0 + \delta_1$ . Ce qui signifie que  $\mathbf{P}_{Y|X=x}$  est une mesure à densité par rapport à  $\mu$ , de densité  $f_{Y|X=x} = x^y(1-x)^{1-y}$ . La loi de  $(X, Y)$  est donc de densité

$$f_{X,Y}(x, y) = f_X(x) f_{Y|X=x} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a+y-1} (1-x)^{b-y} \mathbf{1}_{[0,1]}(x)$$

par rapport à la mesure produit  $\lambda \otimes \mu$ ,  $\lambda$  désignant ici la mesure de Lebesgue. En réutilisant (6.4), on peut calculer la loi conditionnelle de  $X$  sachant  $Y$ . On constate qu'il s'agit d'une loi absolument continue par rapport à la loi de  $X$ , donc absolument continue par rapport à

la mesure de Lebesgue. D'autre part, on a à une constante multiplicative près  $C$  (dépendant de  $y$  mais non de  $x$ ),

$$d\mathbf{P}_{X|Y=y} = \frac{f_{X,Y}(x,y)}{f_Y(y)} dx = Cx^{a-1+y}(1-x)^{b-y} dx.$$

La loi de  $X$  sachant  $Y = y$  est donc une loi bêta de paramètres  $(a+y, b+1-y)$ . Comme l'on reconnaît une loi connue, on en déduit immédiatement la constante de normalisation sans avoir eu à faire les calculs ! De manière plus générale, on remarque que la densité de la loi conditionnelle  $f_{X|Y=y}$  est donnée directement par la densité de la loi jointe  $f_{X,Y}$  à une constante multiplicative près.

## 6.3. Espérance conditionnelle

Munie de la loi conditionnelle, on peut calculer les espérances dites « conditionnelles » de  $Y$  sachant  $X$ .

**Définition 6.3** (Espérance conditionnelle). Soit  $X$  et  $Y$  deux variables aléatoires et  $\mathbf{P}_{Y|X=x}$  la loi conditionnelle de  $Y$  sachant  $X$ . Pour tout  $x \in \text{supp } X$ , on appelle « espérance conditionnelle de  $Y$  sachant  $X = x$  », que l'on note  $\mathbf{E}[Y | X = x]$ , la quantité

$$\mathbf{E}[Y | X = x] = \int y d\mathbf{P}_{Y|X=x}(y).$$

Par définition de la loi conditionnelle, on a (en prenant  $\theta = \psi = 1$  dans (6.1)),

$$\int_{\mathbf{R}^n} \mathbf{E}[Y | X = x] d\mathbf{P}_X(x) = \mathbf{E}[Y].$$

Cette relation s'utilise surtout pour calculer  $\mathbf{E}[Y]$  à partir de  $\mathbf{E}[Y | X = x]$ , voir exercices. De la même manière, pour toute fonction mesurable bornée  $\theta$  (en prenant  $\psi = 1$  dans (6.1)),

$$\int_{\mathbf{R}^n} \mathbf{E}[\theta(Y) | X = x] d\mathbf{P}_X(x) = \mathbf{E}[\theta(Y)].$$

Dans le cas discret, la formule de l'espérance conditionnelle est analogue à celle de l'espérance, en remplaçant la probabilité par la probabilité conditionnelle.

**Corollaire 6.7.** Soit  $X$  et  $Y$  deux variables aléatoires discrètes. Pour tout  $x \in \text{supp } X$ , l'espérance conditionnelle de  $Y$  sachant  $X = x$  est donnée par

$$\mathbf{E}[Y | X = x] = \sum_{y \in \text{supp } \mathbf{P}_Y} y \mathbf{P}(Y = y | X = x).$$

*Démonstration.* On a vu au théorème 6.4 que

$$\mathbf{P}_{Y|X=x} = \sum_{y \in \text{supp } \mathbf{P}_Y} \mathbf{P}(Y = y | X = x) \delta_y.$$

En vertu de la définition des mesures de Dirac et de la définition 6.3,

$$\mathbf{E}[Y | X = x] = \int y \, d\mathbf{P}_{Y|X=x}(y) = \sum_{y \in \text{supp } \mathbf{P}_Y} y \, \mathbf{P}(Y = y | X = x).$$

□

## 6.4. Exercices

### Pour apprendre

EXERCICE 6.1.—

Soit  $X$  et  $Y$  deux variables aléatoires indépendantes de loi respective  $B(N, p)$  et  $B(M, p)$ . Calculer la loi conditionnelle de  $X$  sachant  $X + Y$ .

### Pour s'entraîner

EXERCICE 6.2.—

Soit

$$d\mathbf{P}(x, y) = c \exp\left(-\frac{x^2 + y^2}{2}\right) \mathbf{1}_{\{x > y\}} dx dy$$

une mesure sur le plan  $\mathbf{R}^2$ .

1. Trouver la constante  $c$  pour que  $\mathbf{P}$  soit une probabilité.
2. Soit  $(\Omega, \mathcal{F}, \mathbf{P})$  un espace de probabilité et  $(X, Y) : \Omega \rightarrow \mathbf{R}^2$  une variable aléatoire de loi  $\mathbf{P}$ . Trouver la loi de  $X$  et celle de  $Y$ .
3. Sont-elles indépendantes ?
4. On définit les nouvelles variables aléatoires  $U = X^2 + Y^2$  et  $V = Y$ . Calculer la loi du vecteur  $(U, V)$ .
5. Les variables  $U$  et  $V$  sont-elles indépendantes ?

EXERCICE 6.3 (Points dans un hexagone - \*\*).—

En radio-mobiles, on est souvent amené à simuler des usagers répartis de façon uniforme dans une cellule hexagonale (voir la figure 6.3 pour les éléments caractéristiques d'une telle cellule). Comment faire en utilisant un minimum d'appels au générateur de nombres aléatoires ?

*On rappelle pour simplifier les calculs que pour un hexagone de longueur de côté 1, l'aire est  $A = 3\sqrt{3}/2$ .*

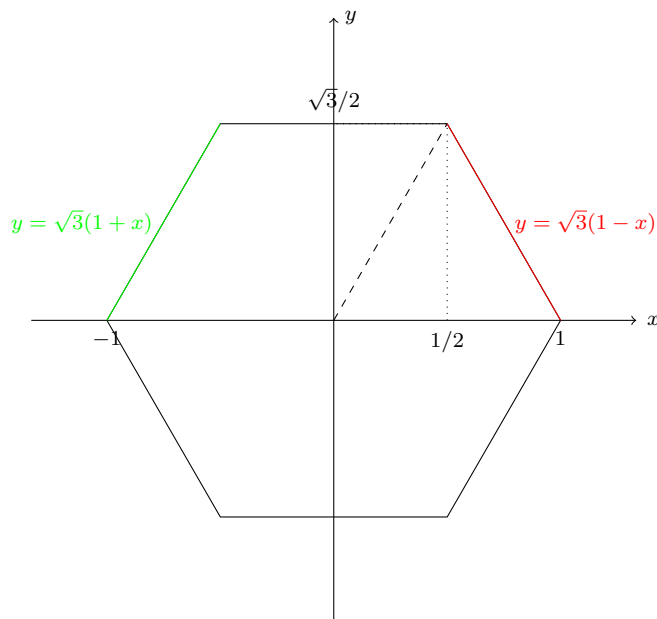
### Pour aller plus loin

EXERCICE 6.4.—

Soit  $X_1, \dots, X_n, Y_1, \dots, Y_n, U$  des variables aléatoires réelles indépendantes et de même loi. La loi de  $U$  est la loi uniforme sur  $\{1, \dots, n\}$ . On pose  $W_n = X_1 + \dots + X_n$ . On note  $X = (X_1, \dots, X_n)$ . On construit  $X'$  de la façon suivante :

$$X' = (X_1, \dots, X_{i-1}, Y_i, X_{i+1}, \dots) \text{ si } U = i.$$

Figure 6.3. – Hexagone régulier.



On pose

$$W'_n = W_n - X_U + Y_U.$$

1. Montrer que  $X$  et  $X'$  ont même loi.
2. En déduire que  $W_n$  et  $W'_n$  ont même loi.

On suppose dorénavant que pour tout  $i \in \{1, \dots, n\}$ ,

$$\mathbf{P}(X_i = 1) = \mathbf{P}(Y_i = 1) = p = 1 - \mathbf{P}(X_i = 0) = 1 - \mathbf{P}(Y_i = 0).$$

3. Identifier la loi de  $W_n$ .
4. Montrer que l'on peut construire  $\tilde{W}_{n-1}$  indépendante de  $X_1$  et de même de loi que  $W_{n-1}$ , telle que

$$W_n = X_i + \tilde{W}_{n-1}.$$

5. Calculer  $\mathbf{P}(X_1 = 1 \mid W_n = m)$  pour  $m \in \{0, \dots, n\}$ .
6. Calculer

$$\mathbf{P}(W_n - W'_n = \epsilon \mid W_n = m) \text{ pour } \epsilon = -1, 0, 1.$$

7. En déduire que

$$\mathbf{E}[W_n - W'_n \mid W_n = m] = \frac{m}{n} - p.$$

**EXERCICE 6.5.**–

Soit  $\begin{pmatrix} X \\ Y \end{pmatrix}$  un vecteur gaussien de matrice de covariance

$$\Gamma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

1. Quelle est la loi de  $X$  ?

2. Quelle est la covariance de  $X$  et  $Y$  ? A quelle condition sur  $\rho$  sont-elles indépendantes ?
3. Calculer les valeurs propres de  $\Gamma$ . A quelle condition sur  $\rho$ ,  $\Gamma$  est-elle définie positive ?

On suppose dorénavant cette condition vérifiée.

4. Rappeler comment est construit  $\Gamma^{1/2}$ . On note

$$\Gamma^{1/2} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{12} & \gamma_{22} \end{pmatrix},$$

où l'on ne cherchera pas à identifier les coefficients  $\gamma_{ij}$ .

En déduire que  $\Gamma^{1/2}$  est inversible, que

$$(\Gamma^{1/2})^{-1} = \Gamma^{-1}, \det \Gamma^{1/2} = (\det \Gamma)^{1/2} \text{ et que } \|(\Gamma^{1/2})^{-1}z\|^2 = \Gamma^{-1}z.z$$

pour tout  $z \in \mathbf{R}^2$ .

5. Calculer la densité de la loi du couple  $(X, Y)$  en fonction de  $\det \Gamma$  et de  $\Gamma^{-1}$ .

*Indication : on introduira un couple  $\begin{pmatrix} U \\ V \end{pmatrix}$ , vecteur gaussien de matrice de covariance  $\mathcal{N}(0, I_2)$  et on utilisera la représentation de  $(X, Y)$  en fonction de  $(U, V)$ .*

6. Montrer que la loi conditionnelle de  $Y$  sachant  $X = x$  est une loi gaussienne de paramètres que l'on précisera.





## 7. Fonction caractéristique

On note  $\langle a, b \rangle$  le produit scalaire de deux vecteurs  $a, b$  de  $\mathbb{R}^d$ , et  $\|a\|$  la norme euclidienne de  $a$ .  $A^T$  désigne la transposée de la matrice  $A$ . Par convention, les vecteurs sont des vecteurs-colonne.

Pour un nombre complexe  $x$ ,  $|x|$  désigne son module et  $\bar{x}$  son conjugué. On note  $i = \sqrt{-1}$ .

**Fonctions à valeur dans  $\mathbb{C}$**  Dans ce chapitre, nous sommes amenés à utiliser l'intégrale, par rapport à des mesures de probabilité, de fonctions à valeur dans  $\mathbb{C}$ . Toute fonction  $f$  à valeur dans  $\mathbb{C}$  peut s'écrire sous la forme :

$$f = f_R + if_I \quad (7.1)$$

où  $f_R, f_I$  sont les fonctions à valeur dans  $\mathbb{R}$  désignant respectivement la partie réelle et la partie imaginaire. On dira qu'une fonction  $f$  à valeur dans  $\mathbb{C}$  est *mesurable* si les fonctions  $f_R$  et  $f_I$  sont mesurables. Elle est par définition *intégrable par rapport à une mesure  $\mu$*  si  $f_R$  et  $f_I$  sont intégrables. On définit alors l'intégrale de  $f$  par :

$$\int f \, d\mu := \int f_R \, d\mu + i \int f_I \, d\mu .$$

A l'aide de la décomposition (7.1) et des inégalités :

$$|f_R| \leq |f|, \quad |f_I| \leq |f|, \quad |f| \leq |f_R| + |f_I| ,$$

il est aisé de vérifier que des propriétés établies pour des fonctions à valeur dans  $\mathbb{R}$  restent valables pour des fonctions à valeur dans  $\mathbb{C}$ . Par exemple :

- une fonction mesurable  $f$  à valeur dans  $\mathbb{C}$  est intégrable par rapport à  $\mu$  si et seulement

$$\int |f| \, d\mu < \infty ;$$

- le théorème de convergence dominée (voir théorème 4.35) reste vrai si les fonctions  $f, f_n$  sont supposées à valeur dans  $\mathbb{C}$ .

### 7.1. Définition et propriétés

Dans cette section,  $X = (X_1, \dots, X_d)$  désigne un vecteur aléatoire à valeur dans  $\mathbb{R}^d$ , défini sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Définition 7.1.** On appelle *fonction caractéristique* de la v.a.  $X$  la fonction  $\phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$  définie par

$$\begin{aligned}\phi_X(t) &:= \mathbf{E}(\exp(i\langle t, X \rangle)) = \int_{\mathbb{R}^d} \exp(i\langle t, x \rangle) dP_X(x) \\ &= \mathbf{E}\left(\exp\left(i \sum_{k=1}^d t_k X_k\right)\right) = \mathbf{E}\left(\prod_{k=1}^d \exp(it_k X_k)\right) .\end{aligned}$$

Comme  $|\exp(i\langle t, X \rangle)| = 1$ , la fonction  $\phi_X$  est bien définie sur  $\mathbb{R}^d$ . Elle est définie pour les v.a. discrètes comme les v.a. à valeur réelle ou vectorielle. Dans le cas particulier où  $X$  est à valeur dans un espace au plus dénombrable  $E$ , on a

$$\phi_X(t) = \sum_{e \in E} \exp(i\langle t, e \rangle) \mathbf{P}[X = e] ;$$

dans le cas où  $X$  est une v.a. à densité  $f_X$  (par rapport à la mesure de Lebesgue  $\lambda_d$  sur  $\mathbb{R}^d$ ), on a

$$\phi_X(t) = \int \exp(i\langle t, x \rangle) f_X(x) dx .$$

La définition de la fonction caractéristique ne dépend que de la loi de la v.a.  $X$ . En particulier, si deux v.a. ont même loi alors elles ont même fonction caractéristique.

*Démonstration.* En conséquence du théorème de transfert (théorème 5.5), on a  $\mathbf{E}(g(X)) = \mathbf{E}(g(Y))$  pour toute fonction mesurable bornée à valeur dans  $\mathbb{R}$  (et donc par une extension triviale, à valeur dans  $\mathbb{C}$ ). En particulier, en appliquant cette propriété à la fonction  $g(x) = \exp(i\langle t, x \rangle)$  ( $t$  étant fixé), on en déduit que  $\phi_X(t) = \phi_Y(t)$ .  $\square$

Vue sous l'angle de transformée d'une loi, la fonction caractéristique d'une v.a. n'est autre que la *transformée de Fourier* de sa loi, qui est définie de la façon suivante :

**Définition 7.2.** Soit  $\mu$  une mesure finie sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . La *transformée de Fourier* de  $\mu$  est la fonction  $\hat{\mu} : \mathbb{R}^d \rightarrow \mathbb{C}$  définie par

$$\hat{\mu}(t) := \int \exp(i\langle t, x \rangle) d\mu(x) .$$

Autrement dit, comme alternative à la définition 7.1, on peut simplement écrire que la fonction caractéristique d'une v.a.  $X$  est  $\phi_X = \hat{P}_X$ .

**REMARQUE.**— Dans le cas où  $\mu$  est une mesure sur  $\mathbb{R}$  admettant une densité  $f$  i.e.,  $\mu(dx) = f(x)dx$ , la définition 7.2 implique que  $\hat{\mu}(t) = \int e^{itx} f(x)dx$ . Ainsi,  $\hat{\mu}(-t)$  est égale à la transformée de Fourier  $\hat{f}$  de la fonction intégrable  $f$ , vue en cours MDI-103. Cette remarque justifie la terminologie « transformée de Fourier ». En ce sens, la définition 7.2 est une extension de la transformée de Fourier des fonctions vue en cours d'analyse.

**Proposition 7.1.** La fonction caractéristique d'une v.a.  $X$  vérifie les propriétés suivantes :

- a) Pour tout  $t \in \mathbb{R}^d$ ,  $|\phi_X(t)| \leq 1$ . En outre,  $\phi_X(0) = 1$ .
- b) La fonction  $\phi_X$  est continue sur  $\mathbb{R}^d$ .
- c) Si  $b \in \mathbb{R}^p$  est un vecteur déterministe et  $A$  est une matrice déterministe de taille  $p \times d$  alors  $AX + b \in \mathbb{R}^p$  et pour tout  $t \in \mathbb{R}^p$ ,

$$\phi_{AX+b}(t) = \exp(i\langle t, b \rangle) \phi_{AX}(t) = \exp(i\langle t, b \rangle) \phi_X(A^T t) .$$

En particulier, si  $X$  est une v.a. à valeur dans  $\mathbb{R}$  et que  $a, b$  sont des réels alors

$$\phi_{aX+b}(t) = \exp(itb) \phi_X(at) .$$

- d) Si  $X, Y$  sont deux v.a. indépendantes définies sur le même espace de probabilité  $(\Omega, \mathcal{F}, \mathbf{P})$  et à valeurs dans  $\mathbb{R}^d$ , alors  $\phi_{X+Y} = \phi_X \phi_Y$ . Plus généralement, si  $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}^d$  sont des v.a. indépendantes, alors :

$$\phi_{X_1+\dots+X_n} = \prod_{k=1}^n \phi_{X_k}$$

*Démonstration.* **a))** On a  $|\phi_X(t)| \leq \mathbf{E}(|\exp(i\langle t, X \rangle)|) = 1$ . De plus,  $\phi_X(0) = \mathbf{E}(\exp(0)) = 1$ .

**b))** La continuité en tout point  $t \in \mathbb{R}^d$  est une conséquence du théorème de convergence dominée (voir théorème 4.35). Soit  $t \in \mathbb{R}^d$ ; nous écrivons pour toute suite  $(h_\ell)_\ell$  à valeur dans  $\mathbb{R}^d$  telle que  $\lim_{\ell \rightarrow +\infty} h_\ell = 0$ ,

$$\phi_X(t + h_\ell) - \phi_X(t) = \int (\exp(i\langle t + h_\ell, x \rangle) - \exp(i\langle t, x \rangle)) dP_X(x) .$$

Par continuité du produit scalaire et de la fonction exponentielle,

$$\lim_{\ell \rightarrow +\infty} (\exp(i\langle t + h_\ell, x \rangle) - \exp(i\langle t, x \rangle)) = 0 .$$

De plus, pour tout  $h \in \mathbb{R}^d$ ,  $|\exp(i\langle t + h, x \rangle) - \exp(i\langle t, x \rangle)| \leq 2$  et  $\int 2 d\mathbf{P}_X = 2 < \infty$ . Par suite, le théorème de convergence dominée entraîne  $\lim_{\ell \rightarrow +\infty} \phi_X(t + h_\ell) = \phi_X(t)$ , ce qui établit la continuité en  $t$  pour tout  $t \in \mathbb{R}^d$ .

**c))** On écrit en utilisant les propriétés du produit scalaire et de la fonction exponentielle, et en utilisant le fait que  $A, b$  sont déterministes

$$\begin{aligned} \phi_{AX+b}(t) &= \mathbf{E}(\exp(i\langle AX + b, t \rangle)) = \exp(i\langle b, t \rangle) \mathbf{E}(\exp(i\langle AX, t \rangle)) \\ &= \exp(i\langle b, t \rangle) \mathbf{E}(\exp(i\langle X, A^T t \rangle)) = \exp(i\langle b, t \rangle) \phi_X(A^T t) . \end{aligned}$$

**d))** On a pour tout  $t \in \mathbb{R}^d$ ,

$$\begin{aligned} \phi_{X+Y}(t) &= \mathbf{E}(\exp(i\langle t, X \rangle) \exp(i\langle t, Y \rangle)) = \mathbf{E}(\exp(i\langle t, X \rangle)) \mathbf{E}(\exp(i\langle t, Y \rangle)) \\ &= \phi_X(t) \phi_Y(t) \end{aligned}$$

en utilisant la caractérisation de l'indépendance donnée par le théorème 5.13 (établie pour des fonctions à valeur réelle et donc valable aussi pour des fonctions à valeur dans  $\mathbb{C}$ ). La généralisation à une somme de  $n$  v.a. indépendantes est immédiate.

□

## 7.2. Fonctions caractéristiques de v.a. usuelles

La table 7.1 fournit l'expression des fonctions caractéristiques de plusieurs loi usuelles. Les sept premières expressions sont la conséquence de calculs triviaux que nous omettons ici. Nous fournissons ci-dessous la preuve des trois dernières.

Nous commençons par établir l'expression de la fonction caractéristique d'une v.a. normale  $\mathcal{N}(0, 1)$ ; nous en déduisons celle d'une loi  $\mathcal{N}_d(0, I)$  puis d'une loi  $\mathcal{N}_d(\mu, \Gamma)$ . Soit  $Y \sim \mathcal{N}(0, 1)$ . Les conditions de dérivation sous l'intégrale sont vérifiées (voir exercice 5.38) et on a

$$\begin{aligned}\phi'_Y(t) &= \frac{1}{\sqrt{2\pi}} \int ix \exp(itx) \exp(-1/2 x^2) dx \\ &= -\frac{i}{\sqrt{2\pi}} \int (it - x) \exp(itx) \exp(-1/2 x^2) dx - \frac{t}{\sqrt{2\pi}} \int \exp(itx) \exp(-1/2 x^2) dx \\ &= -\frac{i}{\sqrt{2\pi}} \left[ \exp(itx) \exp(-1/2 x^2) \right]_{-\infty}^{+\infty} - \frac{t}{\sqrt{2\pi}} \int \exp(itx) \exp(-1/2 x^2) dx = -t \phi_Y(t).\end{aligned}$$

La résolution de l'équation différentielle  $\phi'_Y(t) = -t \phi_Y(t)$  sachant que l'on doit avoir  $\phi_Y(0) = 1$ , donne le résultat.

Soit  $Z \sim \mathcal{N}_d(0, I)$ . Alors par le théorème de Fubini (voir Chapitre 5.3),

$$\begin{aligned}\phi_Z(t) &= \frac{1}{\sqrt{2\pi}^d} \int_{\mathbb{R}^d} \exp(i\langle t, z \rangle) \exp(-1/2 \sum_{k=1}^d z_k^2) dz_1 \cdots dz_d \\ &= \prod_{k=1}^d \left( \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(it_k z_k) \exp(-1/2 z_k^2) dz_k \right) = \prod_{k=1}^d \phi_Y(t_k) \\ &= \prod_{k=1}^d \exp(-1/2 t_k^2) = \exp(-1/2 t^T t) .\end{aligned}$$

Enfin, nous savons que si  $Z \sim \mathcal{N}_d(0, I)$  alors  $X = \mu + \sqrt{\Gamma}Z$  suit une loi  $\mathcal{N}_d(\mu, \Gamma)$  (voir exercice 5.42). L'expression de la fonction caractéristique d'une loi  $\mathcal{N}_d(\mu, \Gamma)$  est maintenant la conséquence de la Proposition 7.1-c).

Soit  $X \sim \Gamma(a, b)$ . Par application du théorème de dérivation sous l'intégrale (voir exercice 5.38) suivie d'une intégration par parties, nous avons  $\phi'_X(t) = \frac{ia}{b-it} \phi_X(t)$  dont nous déduisons l'expression de  $\phi_X$  en utilisant la condition  $\phi_X(0) = 1$ . A noter que si  $a$  n'est pas un entier, on prend la détermination continue valant 1 en 0.

Loi	Expression de $\Phi_X(t)$
Mesure de Dirac $\delta_a$	$\exp(ita)$
Bernoulli $\mathcal{B}(p)$	$1 - p + p \exp(it)$
Binomiale $\mathcal{B}(n, p)$	$(1 - p + p \exp(it))^n$
Géométrique $\mathcal{G}(p)$ sur $\mathbb{N}$	$\frac{p}{1 - (1 - p) \exp(it)}$
Poisson $\mathcal{P}(\lambda)$	$e^{\lambda(e^{it} - 1)}$
Uniforme $\mathcal{U}([a, b])$	$\frac{\exp(itb) - \exp(ita)}{it(b - a)}$
Exponentielle $\mathcal{E}(\lambda)$	$\frac{\lambda}{\lambda - it}$
Gaussienne réelle $\mathcal{N}(\mu, \sigma^2)$ avec $\sigma^2 \geq 0$	$\exp(it\mu - 1/2 \sigma^2 t^2)$
Gaussienne multivariée $\mathcal{N}_n(\mu, \Gamma)$	$\exp(i\langle t, \mu \rangle - 1/2 t^T \Gamma t)$
Gamma $\Gamma(a, b)$	$\left( \frac{b}{b - it} \right)^a$

Table 7.1. — Quelques fonctions caractéristiques utiles. Se rapporter aux tables 4.1 et 1.1 pour la définition des lois. Pour la fonction caractéristique d'une loi  $\mathcal{N}_n(\mu, \Gamma)$  lorsque  $\Gamma$  est définie positive, voir section 7.2; lorsque  $\Gamma$  est positive, voir chapitre 8. Par convention,  $\mathcal{N}(a, 0)$  est la mesure de Dirac en  $a$

### 7.3. Caractérisation de la loi

Nous avons établi (théorème 4.40) que deux lois  $\mu, \nu$  sont égales si et seulement si  $\mathbf{E}(f(X)) = \mathbf{E}(f(Y))$  pour toute fonction continue bornée (ou pour toute fonction continue à support compact); ici  $X \sim \mu$  et  $Y \sim \nu$ . Le théorème suivant donne une autre caractérisation :  $\mu, \nu$  sont égales si et seulement si  $\mathbf{E}(f(X)) = \mathbf{E}(f(Y))$  pour toute fonction  $f$  de la forme  $x \mapsto \exp(i\langle t, x \rangle)$ ,  $t \in \mathbb{R}^n$ . La preuve de ce résultat repose sur le lemme suivant.

**Théorème 7.2.** La fonction caractéristique d'une v.a. détermine sa loi *i.e.*, pour deux vecteurs aléatoires  $X, Y$  à valeur dans  $\mathbb{R}^d$  on a équivalence :

- a) pour tout  $t \in \mathbb{R}^d$ ,  $\phi_X(t) = \phi_Y(t)$ .
- b) les vecteurs aléatoires  $X$  et  $Y$  ont même loi.

*Démonstration.* Le sens réciproque est trivial étant donnée la définition de la fonction caractéristique : on ne se préoccupe que du sens direct. Plaçons nous pour simplifier dans le cas  $d = 1$ . A titre de remarque, signalons que la preuve est déjà connue des élèves dans le cas particulier où  $X$  et  $Y$  sont deux lois à densité  $f_X$  et  $f_Y$  : dans le cas où les fonctions caractéristiques sont elles-mêmes intégrables, les densités se déduisent des fonctions caractéristiques par transformée de Fourier de  $\phi_X$  et  $\phi_Y$ . Ainsi,  $\phi_X = \phi_Y$  implique que  $f_X = f_Y$  presque partout, et donc que  $P_X = P_Y$ .

Dans le cas général, la preuve repose sur la formule d'inversion (voir l'exercice 7.4) :

$$P_X([a, b]) = \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t) dt, \quad (7.2)$$

qui vraie pour tous  $a, b$  en lesquels  $P_X$  n'a pas de masse, c'est-à-dire pour tous  $a, b$  hors de l'ensemble  $D_X := \{x : P_X(\{x\}) > 0\}$ . D'après l'exercice 4.5, cet ensemble  $D_X$  est au plus dénombrable. Donc pour tout  $a$  et  $b$  hors de  $D_X \cup D_Y$ , on a  $P_X([a, b]) = P_Y([a, b])$ . Puisque  $F_X(b) = F_X(a) + P_X([a, b])$ , il suffit de faire tendre  $a$  vers  $-\infty$  pour obtenir :

$$F_X(b) = F_Y(b) \quad (7.3)$$

pour tout point  $b$  hors d'un ensemble au plus dénombrable. En utilisant la continuité à droite des fonctions de répartition, on conclut que (7.3) est vraie en tout point. Puisque le fonction de répartition détermine entièrement la loi, le résultat est démontré. □

### 7.4. Caractérisation de l'indépendance

Soient  $(X_k)_{k \leq p}$  des vecteurs aléatoires à valeur dans  $\mathbb{R}^{n_k}$ , définies sur le même espace de probabilité  $(\Omega, \mathcal{F}, \mathbf{P})$ .

**Théorème 7.3.** Les v.a.  $X_1, \dots, X_p$  sont indépendantes si et seulement si pour tout  $t = (t_1, \dots, t_p) \in \mathbb{R}^{n_1 + \dots + n_p}$ ,  $\phi_X(t) = \prod_{k=1}^p \phi_{X_k}(t_k)$ .

*Démonstration.* Supposons que les v.a. sont indépendantes. On a

$$\phi_X(t) = \mathbf{E} \left( \prod_{k=1}^p \exp(i \langle t_k, X_k \rangle) \right) = \prod_{k=1}^p \mathbf{E} (\exp(i \langle t_k, X_k \rangle)) = \prod_{k=1}^p \phi_{X_k}(t_k) ,$$

en utilisant la caractérisation de l'indépendance donnée par le théorème 5.13 dans la seconde égalité.

Considérons la réciproque. Soient  $Y_1, \dots, Y_p$  des v.a. indépendantes et telles que pour tout  $k$ ,  $X_k$  et  $Y_k$  ont même loi. Comme les v.a. sont indépendantes, pour tout  $t = (t_1, \dots, t_p) \in \mathbb{R}^{n_1 + \dots + n_p}$ ,  $\phi_{(Y_1, \dots, Y_p)}(t) = \prod_{k=1}^p \phi_{Y_k}(t_k)$ ; de plus, comme  $X_k$  et  $Y_k$  ont même loi, d'après le Théorème 7.2  $\phi_{X_k}(t_k) = \phi_{Y_k}(t_k)$  pour tout  $t_k \in \mathbb{R}^{n_k}$ . Donc pour tout  $t = (t_1, \dots, t_p) \in \mathbb{R}^{n_1 + \dots + n_p}$ ,

$$\phi_{(Y_1, \dots, Y_p)}(t) = \prod_{k=1}^p \phi_{X_k}(t_k) = \phi_X(t) .$$

Le Théorème 7.2 entraîne que  $(Y_1, \dots, Y_p)$  et  $(X_1, \dots, X_p)$  ont même loi donc en particulier, les v.a.  $(X_k)_{1 \leq k \leq p}$  sont indépendantes.  $\square$

## 7.5. Calcul de moments

### 7.5.1. Moments et fonction caractéristique

Puisque la fonction caractéristique « caractérise la loi », elle détermine également les moments de cette loi. Il se trouve que la fonction caractéristique est un outil souvent commode pour évaluer les moments d'une loi.

**Théorème 7.4.** Soit  $X$  une variable aléatoire *réelle* possédant un moment d'ordre  $p$  ( $p > 0$ ). Alors  $\phi_X$  est de classe  $C^p$  et l'on a

$$\frac{\partial^p \phi_X(t)}{\partial t^p} = i^p \mathbf{E} (X^p \exp(itX)) .$$

En particulier, les moments sont liés aux dérivées en zero :

$$\mathbf{E}(X^p) = (-1)^p i^p \frac{\partial^p \phi_X(0)}{\partial t^p} .$$

*Démonstration.* Nous établissons le résultat suivant :

$$\phi_X(t) = 1 + it\mathbf{E}(X) - \frac{t^2}{2}\mathbf{E}(X^2) + \cdots + i^p \frac{t^p}{p!}\mathbf{E}(X^p) + \xi(t) \quad \text{où } \lim_{t \rightarrow 0} \xi(t)/t^p = 0.$$

Cette égalité se justifie par le développement :

$$\begin{aligned} \exp(itx) &= \sum_{k=0}^p i^k \frac{t^k}{k!} x^k \\ &= \sum_{k \geq p} i^k \frac{t^k}{(k-p)!} x^k \frac{1}{(k-p+1) \cdots (k-1)k} \\ &= \sum_{k \geq p} i^k \frac{t^k}{(k-p)!} x^k \int_0^1 \int_0^{u_p} \int_0^{u_{p-1}} \cdots \int_0^{u_2} u_1^{k-p} du_1 du_2 \cdots du_p \\ &= i^p t^p x^p \int_0^1 \int_0^{u_p} \int_0^{u_{p-1}} \cdots \int_0^{u_2} \sum_{k \geq p} i^{k-p} \frac{t^{k-p}}{(k-p)!} x^{k-p} u_1^{k-p} du_1 du_2 \cdots du_p \\ &= i^p t^p x^p \int_0^1 \int_0^{u_p} \int_0^{u_{p-1}} \cdots \int_0^{u_2} \left( \exp(iu_1 tx) - 1 \right) du_1 du_2 \cdots du_p, \end{aligned}$$

où l'on a utilisé le théorème Fubini (voir Chapitre 5.3). On montre que l'espérance de ce dernier terme est  $o(t^p)$  par application du théorème de convergence dominée.  $\square$

## Généralisation aux vecteurs aléatoires

De même, on peut montrer le résultat suivant :

**Théorème 7.5.** Soit  $X$  un vecteur aléatoire admettant un moment d'ordre  $p$  ( $p > 0$ ).  $t \mapsto \phi_X(t)$  est de classe  $C^p$  et on a

$$\frac{\partial^p \phi_X(t)}{\partial t_1 \cdots \partial t_p} = i^p \mathbf{E} \left( X_{t_1} X_{t_2} \cdots X_{t_p} \exp(i \langle t, X \rangle) \right).$$

On en déduit aussi une méthode pour le calcul de moments à partir de l'expression de la fonction caractéristique.

## 7.5.2. Applications

### Application 1

Nous montrons que si  $X$  est une v.a. réelle gaussienne d'espérance  $\mu$  et de variance  $\sigma^2$  ( $\sigma^2 > 0$ ) i.e.,  $X \sim \mathcal{N}(\mu, \sigma^2)$  alors tous les moments impairs de  $X - \mu$  sont nuls et les



moments pairs s'expriment à l'aide de  $\sigma^2$  : pour tout  $q \in \mathbb{N}$ ,

$$\mathbf{E} \left( (X - \mu)^{2q+1} \right) = 0, \quad (7.4)$$

$$\mathbf{E} \left( (X - \mu)^{2q} \right) = \sigma^{2q} (2q-1)(2q-3) \cdots 3 = \sigma^{2q} \frac{(2q)!}{2^q q!}. \quad (7.5)$$

*Démonstration.* On veut calculer  $\mathbf{E}((X - \mu)^q)$  pour tout  $q \in \mathbb{N}$ . Nous avons établi (exercice 5.42) que si  $X \sim \mathcal{N}(\mu, \sigma^2)$  alors  $\sigma^{-1}(X - \mu) \sim \mathcal{N}(0, 1)$ . Par suite, nous allons établir

$$\mathbf{E}(Y^{2q+1}) = 0, \quad \mathbf{E}(Y^{2q}) = (2q-1)(2q-3) \cdots 3 = \frac{(2q)!}{2^q q!}, \quad (7.6)$$

où  $Y \sim \mathcal{N}(0, 1)$ . On écrit

$$\mathbf{E}(\exp(itY)) = \mathbf{E} \left( \sum_{n \geq 0} \frac{(itY)^n}{n!} \right) = \mathbf{E} \left( \lim_{N \rightarrow +\infty} \sum_{k=0}^N \frac{(it)^k}{k!} Y^k \right).$$

Pour permuter limite et espérance, on applique le théorème de convergence dominée (théorème 4.35) en remarquant que  $|\sum_{k=0}^N \frac{(it)^k}{k!} Y^k| \leq \exp(|t||Y|)$  et que l'espérance de ce majorant est finie. Par suite,

$$\mathbf{E}(\exp(itY)) = \lim_{N \rightarrow +\infty} \sum_{k=0}^N \frac{(it)^k}{k!} \mathbf{E}(Y^k) = \sum_{k \geq 0} \frac{(it)^k}{k!} \mathbf{E}(Y^k).$$

D'autre part, d'après le Tableau 7.1

$$\mathbf{E}(\exp(itY)) = \exp(-1/2 t^2) = \sum_{n \geq 0} \frac{(-1)^n t^{2n}}{2^n n!}.$$

On en déduit (7.6) par identification.  $\square$

Par convention, une loi gaussienne d'espérance  $\mu$  et de variance nulle est une masse de Dirac en  $\mu$  ( $\mu \in \mathbb{R}$ ). Si  $X \sim \delta_\mu$  alors tous ses moments centrés sont nuls et les égalités ci-dessus restent vraies.

Ainsi, on a établi que si  $X \sim \mathcal{N}(\mu, \sigma^2)$ , avec  $\sigma^2 \geq 0$ , on a (7.4) et (7.5).

## Application 2

Soit  $m \in \mathbb{R}^d$  un vecteur déterministe et  $\Gamma$  une matrice  $d \times d$  positive déterministe. Montrons

a) que la fonction

$$t \mapsto \exp \left( i \langle t, m \rangle - 1/2 t^T \Gamma t \right) \quad (7.7)$$

est la fonction caractéristique d'un vecteur aléatoire,

b) et que ce vecteur aléatoire a pour espérance  $m$  et pour matrice de covariance  $\Gamma$ .

*Démonstration.* (a) Nous avons établi que lorsque  $\Gamma$  est définie positive, la fonction donnée par (7.7) est la transformée de Fourier d'une loi  $\mathcal{N}_d(m, \Gamma)$  (voir Tableau 7.1). L'exercice suivant justifie le fait que lorsque  $\Gamma$  est juste supposée positive, la fonction définie par (7.7) peut encore être lue comme la transformée de Fourier d'une loi.

(b) Nous montrons l'expression de l'espérance et de la matrice de covariance par application des résultats du Théorème 7.5.1.

Soit  $k \in \{1, \dots, n\}$ . En dérivant  $t \mapsto \phi_X(t)$  par rapport à la  $k$ -ième composante  $t_k$ , on sait d'une part que

$$\partial_{t_k} \phi_X|_{t=0} = i\mathbf{E}(X_k)$$

et d'autre part,

$$\partial_{t_k} \phi_X|_{t=0} = \partial_{t_k} \left( \exp(it^T m - 1/2 t^T \Gamma t) \right) |_{t=0} = i m_k .$$

Ainsi,  $\mathbf{E}(X_k) = m_k$  pour tout  $k \in \{1, \dots, n\}$ .

On considère la dérivée partielle d'ordre 2 :  $\partial_{t_j t_k} \phi_X(t)$ . D'une part, on sait que

$$\partial_{t_j t_k} \phi_X|_{t=0} = -\mathbf{E}(X_j X_k) .$$

D'autre part,

$$\partial_{t_j t_k} \phi_X|_{t=0} = \partial_{t_j t_k} \left( \exp(it^T m - 1/2 t^T \Gamma t) \right) |_{t=0} = -\Gamma_{j,k} - m_j m_k .$$

Par suite,

$$\Gamma_{j,k} = \mathbf{E}(X_j X_k) - m_j m_k = \mathbf{E}(X_j X_k) - \mathbf{E}(X_j) \mathbf{E}(X_k) = \text{Cov}(X_j, X_k) .$$

Ainsi,  $\Gamma$  est la matrice de covariance du vecteur aléatoire  $X$ . □

## 7.6. Exercices

### Pour apprendre

EXERCICE 7.1.—

Dans ce qui suit,  $X$  et  $Y$  sont deux variables indépendantes, on demande de calculer la loi de leur somme.

1.  $X \sim B(n, p)$  et  $Y \sim B(m, p)$ .
2.  $X \sim \text{Geom}(p)$  et  $Y \sim \text{Geom}(p')$ .
3.  $X \sim \text{Po}(\lambda)$  et  $Y \sim \text{Po}(\mu)$ .
4.  $X \sim \mathcal{N}(m, \sigma^2)$  et  $Y \sim \mathcal{N}(r, \nu^2)$ .
5.  $X \sim \mathcal{E}(\lambda)$  et  $Y \sim \mathcal{E}(\mu)$ .

### Pour s'entraîner

EXERCICE 7.2.—

Soit  $(X_n, n \geq 1)$  une suite de v.a. indépendantes, de loi exponentielle de paramètre  $\lambda$ . Soit  $T_n = X_1 + \dots + X_n$ .

1. Calculer la loi de  $(T_1, T_2, \dots, T_n)$ .
2. En déduire la loi de  $T_n$ .
3. Calculer directement la fonction caractéristique de  $T_n$ .

EXERCICE 7.3.—

Soit  $m \in \mathbb{R}^d$  un vecteur déterministe et  $\Gamma$  une matrice  $d \times d$  positive déterministe. On écrit  $\Gamma = U\Delta U^T$  où  $U$  est une matrice orthogonale et  $\Delta$  est une matrice diagonale. On suppose que  $\Delta_{r+1,r+1} = \dots = \Delta_{d,d} = 0$ .

Montrer que la transformée de Fourier de la loi de  $m + \sum_{k=1}^r \sqrt{\Delta_{k,k}} Y_k U_{\cdot,k}$  où les v.a.  $(Y_j)_j$  sont i.i.d. de loi  $\mathcal{N}(0, 1)$  est donnée par (7.7).  $U_{\cdot,k}$  désigne la colonne  $k$  de la matrice  $U$ .

### Pour aller plus loin

EXERCICE 7.4.—

Soit  $X$  une v.a.r. de loi  $\mathbf{P}_X$  et de fonction caractéristique  $\phi_X$ . On veut démontrer la formule d'inversion (7.2). On note  $I_T$  la quantité à l'intérieur de la limite. Soient  $a < b$  deux réels.

1. En invoquant le théorème de Fubini, justifier l'égalité

$$I_T = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left[ \int_{-T}^T \frac{e^{it(x-a)} - e^{it(x-b)}}{it} \right] d\mathbf{P}_X(x) .$$

2. On pose  $S(T) = \int_0^T (\sin x)/x \, dx$ . On note  $\text{sgn}(x)$  le signe de  $x$  (1 si  $x > 0$ , -1 si  $x < 0$  et 0 si  $x = 0$ ). Montrer que pour tout  $T > 0$ ,

$$\int_0^T \frac{\sin t\theta}{t} dt = \text{sgn}(\theta) S(T|\theta|) .$$

3. En déduire que  $I_T = \int_{-\infty}^{+\infty} \psi(T, x) d\mathbf{P}_X(x)$  où

$$\psi(T, x) = \frac{\text{sgn}(x-a)}{\pi} S(T|x-a|) - \frac{\text{sgn}(x-b)}{\pi} S(T|x-b|) .$$

4. On admettra (ou on se souviendra) que  $S(T)$  tend vers  $\pi/2$  quand  $T \rightarrow \infty$ . Montrer que l'intégrande  $\psi$  est bornée et que :

$$\lim_{T \rightarrow +\infty} \psi(T, x) = \begin{cases} 0 & \text{si } x < a \text{ ou } x > b \\ \frac{1}{2} & \text{si } x = a \text{ ou } x = b \\ 1 & \text{si } a < x < b . \end{cases}$$

5. En utilisant le théorème de convergence dominée, en déduire la formule d'inversion (7.2) pour tout points  $a, b$  tels que  $\mathbf{P}_X(\{a\}) = \mathbf{P}_X(\{b\}) = 0$ .

## 8. Vecteurs gaussiens

On note  $\langle a, b \rangle$  le produit scalaire de deux vecteurs  $a, b$  de  $\mathbf{R}^d$ ; et on note  $A^T$  la transposée de la matrice  $A$ . **Par convention, les vecteurs sont des vecteurs-colonne.**

### 8.1. Préliminaires

La loi gaussienne (ou normale) de paramètres  $m, \sigma^2$  ( $\sigma \geq 0, m \in \mathbf{R}$ ) - notée  $\mathcal{N}(m, \sigma^2)$  - est définie comme suit :

**si  $\sigma > 0$  :** la loi de densité par rapport à la mesure de Lebesgue donnée par

$$\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{(x-m)^2}{\sigma^2}\right). \quad (8.1)$$

**si  $\sigma = 0$  :** la mesure de Dirac en  $m$ .

La fonction caractéristique d'une loi gaussienne  $\mathcal{N}(m, \sigma^2)$  est donnée par (voir Tableau 7.1)

$$t \mapsto \exp(itm - \frac{1}{2} t^2 \sigma^2). \quad (8.2)$$

### 8.2. Définition, propriétés

#### 8.2.1. Définition

Dans la suite de ce chapitre,  $m \in \mathbf{R}^d$  est un vecteur déterministe et  $\Gamma$  est une matrice de covariance  $d \times d$  (en particulier,  $\Gamma$  est une matrice symétrique, positive, d'après le paragraphe 5.3.3). Nous écrirons  $m = (m_1, \dots, m_d)$  et noterons  $\Gamma_{i,j}$  l'élément  $(i, j)$  de la matrice  $\Gamma$ .

Soit  $X$  un vecteur aléatoire à valeur dans  $\mathbf{R}^d$  défini sur  $(\Omega, \mathcal{F}, \mathbf{P})$  et possédant des moments d'ordre 2.

**Définition 8.1.**  $X$  est un *vecteur gaussien* (ou *variable gaussienne multivariée* ou *variable normale multivariée*) si et seulement si pour tout  $a \in \mathbf{R}^d$ , la loi de

$\langle a, X \rangle$  est une loi gaussienne (éventuellement de variance nulle).

### 8.2.2. Fonction caractéristique

Nous avons vu au chapitre 7 que la fonction caractéristique déterminait la loi de  $X$ . Le théorème suivant peut être lu comme une alternative à la définition de vecteur gaussien.

**Théorème 8.1.** Les deux conditions sont équivalentes

- a)  $X$  est un vecteur gaussien d'espérance  $m$  et de matrice de covariance  $\Gamma$ .
- b) la fonction caractéristique du vecteur aléatoire  $X$  est  $t \mapsto \exp(i\langle t, m \rangle - 1/2 t^T \Gamma t)$ .

Dans ce cas, on écrira  $X \sim \mathcal{N}_d(m, \Gamma)$ .

*Démonstration.* Supposons a)). Alors pour tout  $t \in \mathbf{R}^d$ ,  $\langle t, X \rangle$  est une v.a.r. gaussienne d'espérance  $\langle t, m \rangle$  et de variance  $t^T \Gamma t$ . On en déduit l'expression de la fonction caractéristique en appliquant le formulaire, tableau 7.1.

Réciproquement, supposons b)). Identifions la loi de  $\langle t, X \rangle$ , pour tout  $t \in \mathbf{R}^d$ , en calculant la fonction caractéristique de cette v.a. à valeur dans  $\mathbf{R}$ . Soit  $y \in \mathbf{R}$  :

$$\begin{aligned} \phi_{\langle t, X \rangle}(y) &= \mathbf{E}(\exp(iy \langle t, X \rangle)) \\ &= \mathbf{E}(\exp(i\langle yt, X \rangle)) = \phi_X(yt) = \exp(i\langle yt, m \rangle - 1/2 (yt)^T \Gamma (yt)) \\ &= \exp(iy\langle t, m \rangle - 1/2 y^2 (t^T \Gamma t)) . \end{aligned}$$

On reconnaît la fonction caractéristique d'une loi gaussienne réelle d'espérance  $\langle t, m \rangle$  et de variance (éventuellement nulle)  $t^T \Gamma t$  (voir le formulaire Tableau 7.1). Donc  $X$  est un vecteur gaussien ; l'expression de son espérance et de sa matrice de covariance sont une conséquence de la section 7.5.2.  $\square$

Ce théorème, combiné au théorème 7.2, montre que la loi d'un vecteur gaussien est entièrement caractérisée par son espérance  $m$  et sa matrice de covariance  $\Gamma$ .

L'expression de la fonction caractéristique d'un vecteur gaussien est à rapprocher de l'expression obtenue dans la section 7.2 pour les variables gaussiennes multivariées  $\mathcal{N}_d(m, \Gamma)$  dans le cas où  $\Gamma$  est définie positive.

Puisque la fonction caractéristique caractérise la loi, il est légitime de se demander si, réciproquement, la loi d'un vecteur gaussien possède une densité par rapport à la mesure de Lebesgue sur  $\mathbf{R}^d$ . Ce n'est pas toujours le cas et tout dépend si  $\Gamma$  est inversible ou pas. Nous reviendrons sur ce point en section 8.6.

### 8.2.3. Exemples et contre-exemple

Puisque les v.a. constantes sont des lois gaussiennes (de variance nulle), tout vecteur constant est un exemple de vecteur gaussien.

Un exemple de vecteur gaussien moins trivial est obtenu en considérant des v.a.  $X_1, \dots, X_d$  indépendantes de même loi  $\mathcal{N}(0, 1)$  et en posant  $X = (X_1, \dots, X_d)$ .

*Démonstration.*  $X$  est une loi gaussienne multivariée  $\mathcal{N}_d(0, I)$  (voir exercice 5.8) donc, d'après le tableau 7.1, sa fonction caractéristique est donnée par  $\exp(-1/2 \|t\|^2)$ . D'après le théorème 8.1,  $X$  est un vecteur gaussien.  $\square$

Plus généralement, on peut obtenir un vecteur gaussien par concaténation de v.a. gaussiennes indépendantes (on peut établir rapidement ce résultat en appliquant le résultat énoncé en section 8.5).

Si  $X$  est un vecteur gaussien, alors tout sous-vecteur est encore un vecteur gaussien. En particulier, toute composante d'un vecteur gaussien est un vecteur gaussien réel *i.e.*, c'est une loi gaussienne (donc soit une v.a. constante, soit une v.a. de densité de la forme (8.1)). La proposition suivante précise le lien entre les paramètres du vecteur gaussien et les paramètres de la loi gaussienne de chaque composante.

**Proposition 8.2.** Soit  $X \sim \mathcal{N}_d(m, \Gamma)$ . Pour tout  $k \in \{1, \dots, n\}$ ,  $X_k \sim \mathcal{N}(m_k, \Gamma_{k,k})$ .

*Démonstration.* On a pour tout  $t \in \mathbf{R}$  :

$$\phi_{X_k}(t) = \phi_X((0, \dots, 0, t, 0, \dots, 0)) = \exp(itm_k - 1/2 t^2 \Gamma_{k,k}) .$$

D'après le théorème 7.2, la fonction caractéristique caractérise la loi et à droite, on reconnaît la fonction caractéristique d'une loi  $\mathcal{N}(m_k, \Gamma_{k,k})$  (voir Eq. (8.2)).  $\square$

Réciproquement, est-il vrai qu'un vecteur aléatoire tel que toutes ses composantes sont des v.a. gaussiennes est un vecteur gaussien ? la réponse est non comme le montre l'exercice 8.3.

## 8.3. Caractérisation de l'indépendance

Nous savons que si les composantes  $X_1, \dots, X_d$  d'un vecteur aléatoire  $X$  sont indépendantes, alors ces v.a. sont décorrélées et la matrice de covariance de  $X$  est une matrice diagonale.

Le théorème suivant établit un résultat plus fort en considérant la réciproque : si les composantes  $X_1, \dots, X_d$  du vecteur gaussien  $X$  sont décorrélées (*i.e.*, la matrice de covariance de  $X$  est diagonale) alors ces composantes sont indépendantes.

Nous insistons sur le fait que la décorrélation deux à deux d'une famille de v.a. n'entraîne pas nécessairement l'indépendance mutuelle de ces v.a. (voir par exemple, l'exercice 8.3) et que le résultat est ici établi sous des hypothèses précises sur la loi jointe de cette famille de v.a.

**Théorème 8.3.** Soient  $(X_k)_{k \leq d}$  des v.a. réelles définies sur  $(\Omega, \mathcal{F}, \mathbf{P})$ . Les deux conditions sont équivalentes :

- a) Le vecteur aléatoire  $(X_1, \dots, X_d)$  est un vecteur gaussien et  $\text{Cov}(X_i, X_j) = 0$  pour tout  $i \neq j$ .
- b) Les v.a.  $X_1, \dots, X_d$  sont des v.a. gaussiennes **indépendantes**.

*Démonstration.* Soit  $\Gamma$  la matrice du vecteur aléatoire  $X = (X_1, \dots, X_d)$ . Supposons que  $\Gamma$  est de la forme  $\text{diag}(\sigma_1^2, \dots, \sigma_d^2)$ ,  $\sigma_k \geq 0$ . Alors, d'après le théorème 8.1, la fonction caractéristique de  $X$  est

$$\phi_X(t) = \exp(i \sum_{k=1}^d t_k m_k) \exp(-1/2 \sum_{k=1}^d t_k^2 \sigma_k^2) = \prod_{k=1}^d \exp(it_k m_k - 1/2 t_k^2 \sigma_k^2).$$

Or on sait que chaque composante  $X_k$  suit une loi  $\mathcal{N}(m_k, \sigma_k^2)$  (cf. Proposition 8.2). Donc  $\phi_X(t) = \prod_{k=1}^d \phi_{X_k}(t_k)$ . Ainsi, par le théorème 7.3, les v.a. sont indépendantes.

Réciproquement, si les v.a.  $(X_k)_{k \leq d}$  sont indépendantes, alors leur covariance est nulle. Le fait que pour tout  $t \in \mathbf{R}^d$ ,  $\langle t, X \rangle$  est une v.a. gaussienne est établi en section 8.2.3.  $\square$

**Corollaire 8.4.** Soit  $X = (X_1, \dots, X_d) \sim \mathcal{N}_d(m, \Gamma)$ . Les v.a.  $X_1, \dots, X_d$  sont indépendantes si et seulement si la matrice de covariance  $\Gamma$  est diagonale.

L'exercice 8.3 illustre l'importance de la condition sur la loi jointe des v.a. pour que la décorrélation entraîne l'indépendance : dans cet exemple, les composantes  $X_1$  et  $X_2$  sont deux v.a. gaussiennes mais le vecteur  $(X_1, X_2)$  n'est pas un vecteur gaussien ; ces v.a. sont décorrélées mais elles ne sont pas indépendantes.

## 8.4. Stabilité par transformation affine

**Proposition 8.5.** Soient  $b \in \mathbf{R}^p$  un vecteur déterministe et  $A$  une matrice  $p \times d$  déterministe. Soit  $X \sim \mathcal{N}_d(m, \Gamma)$ . Alors  $AX + b$  est un vecteur gaussien (à valeur dans  $\mathbf{R}^p$ ), d'espérance  $Am + b$  et de matrice de covariance  $A\Gamma A^T$ .



*Démonstration.* Calculons la fonction caractéristique de  $AX + b$ . Soit  $t \in \mathbf{R}^p$ . On a

$$\phi_{AX+b}(t) = \exp(i\langle t, b \rangle) \phi_X(A^T t) .$$

Par le Théorème 8.1, il vient

$$\begin{aligned} \phi_X(A^T t) &= \exp(i\langle A^T t, m \rangle) \exp(-1/2 (A^T t)^T \Gamma (A^T t)) \\ &= \exp(i\langle t, Am \rangle) \exp(-1/2 t^T (A \Gamma A^T) t) . \end{aligned}$$

Ainsi,

$$\phi_{AX+b}(t) = \exp(i\langle t, b + Am \rangle) \exp(-1/2 t^T (A \Gamma A^T) t) ,$$

et en utilisant encore le Théorème 8.1, on en déduit que  $AX + b$  est un vecteur gaussien d'espérance  $Am + b$  et de matrice de covariance  $A \Gamma A^T$ .  $\square$

## Construction d'un vecteur gaussien

Le théorème 8.3 prouve l'existence d'un vecteur gaussien centré réduit (*i.e.*, d'espérance nulle et de matrice de covariance égale à l'identité). Nous montrons comment n'importe quel vecteur gaussien  $\mathcal{N}_d(m, \Gamma)$  s'obtient par transformation affine d'un vecteur gaussien centré réduit.

- Etape 1 : construction d'un vecteur de loi  $\mathcal{N}_d(0, I)$ . Le théorème 8.3 montre que le vecteur  $Y := (Y_1, \dots, Y_d)$  où  $(Y_k)_{k \leq d}$  sont des v.a. gaussiennes centrées réduites indépendantes a pour loi  $\mathcal{N}_d(0, I)$ .
- Etape 2 : transformation affine de  $Y$ . Soit une matrice  $\sqrt{\Gamma}$  telle que  $\sqrt{\Gamma} \sqrt{\Gamma}^T = \Gamma$  (comme  $\Gamma$  est une matrice positive,  $\sqrt{\Gamma}$  existe toujours<sup>1</sup> La proposition 8.5 entraîne que  $m + \sqrt{\Gamma} Y$  a pour loi  $\mathcal{N}_d(m, \Gamma)$ ).

## 8.5. Somme de vecteurs gaussiens indépendants

La proposition suivante montre que la somme de **vecteurs gaussiens indépendants** est encore un vecteur gaussien, dont l'espérance (resp. la matrice de covariance) est la somme des espérances (resp. des matrices de covariance).

Il est important de noter que ce résultat n'est vrai que si les variables sont indépendantes ; considérons en effet le contre-exemple suivant. Soit  $X \sim \mathcal{N}(0, 1)$  et  $Y = -X$  ; notons que  $Y \sim \mathcal{N}(0, 1)$  de sorte que  $X$  et  $Y$  sont deux vecteurs gaussiens. Alors  $X + Y = 0$  (avec probabilité 1) et donc  $X + Y$  est un vecteur gaussien de variance nulle (la variance

1. puisque  $\Gamma$  est une matrice de covariance, il existe une matrice orthogonale  $Q$  et une matrice diagonale  $\Delta$  dont les éléments diagonaux  $\Delta_{j,j}$  sont positifs ou nuls telles que  $\Gamma = Q \Delta Q^T$ . On peut prendre  $\sqrt{\Gamma} = Q \sqrt{\Delta} Q^T$  où  $\sqrt{\Delta}$  est la matrice diagonale dont les éléments diagonaux sont  $\sqrt{\Delta_{j,j}}$ .

n'est donc pas égale à la somme des variances). Mais  $X$  et  $Y$  ne sont pas indépendantes puisque

$$\mathbf{E}(XY) = -\mathbf{E}(X^2) = -1 \neq 0 = \mathbf{E}(X)\mathbf{E}(Y) .$$

**Proposition 8.6.** Soient  $X^{(1)}, \dots, X^{(p)}$  des vecteurs aléatoires indépendants tels que  $X^{(\ell)} \sim \mathcal{N}_d(m^{(\ell)}, \Gamma^{(\ell)})$ . Alors

$$X^{(1)} + \dots + X^{(p)} \sim \mathcal{N}_d\left(\sum_{\ell=1}^p m^{(\ell)}; \sum_{\ell=1}^p \Gamma^{(\ell)}\right) .$$

*Démonstration.* Pour tout  $t \in \mathbf{R}^d$ ,

$$\begin{aligned} \phi_{X^{(1)} + \dots + X^{(p)}}(t) &= \mathbf{E}\left(\exp(i\langle t, X^{(1)} \rangle + \dots + i\langle t, X^{(p)} \rangle)\right) \\ &= \prod_{\ell=1}^p \mathbf{E}\left(\exp(i\langle t, X^{(\ell)} \rangle)\right) = \prod_{\ell=1}^p \phi_{X^{(\ell)}}(t) \end{aligned}$$

puisque les v.a.  $(X^{(\ell)})_{\ell \leq p}$  sont indépendantes. En utilisant le formulaire Tableau 7.1

$$\phi_{X^{(\ell)}}(t) = \exp\left(i\langle t, m^{(\ell)} \rangle - \frac{1}{2} t^T \Gamma^{(\ell)} t\right) ,$$

donc

$$\phi_{X^{(1)} + \dots + X^{(p)}}(t) = \exp\left(i\langle t, \sum_{\ell=1}^p m^{(\ell)} \rangle - \frac{1}{2} t^T \left\{ \sum_{\ell=1}^p \Gamma^{(\ell)} \right\} t\right) ,$$

et l'on conclut par application du théorème 7.2 et du formulaire Tableau 7.1.  $\square$

## 8.6. La loi d'un vecteur gaussien admet-elle une densité ?

Soit  $X \sim \mathcal{N}_d(m, \Gamma)$ . On distingue deux cas :

**Lorsque  $\Gamma$  est inversible.** Nous avons établi au chapitre 7 (voir Tableau 7.1 et théorème 7.2) que la loi d'un vecteur gaussien admet une densité :

$$\frac{1}{\sqrt{2\pi}^d} \frac{1}{\sqrt{\det(\Gamma)}} \exp\left(-\frac{1}{2} (x - m)^T \Gamma^{-1} (x - m)\right) .$$

**Lorsque  $\Gamma$  est non-inversible.** La loi du vecteur gaussien n'admet pas de densité. On peut montrer que

$$\mathbf{P}[X - m \in \text{Im}(\Gamma)] = 1 ; \quad (8.3)$$

autrement dit, la v.a.  $X$  prend ses valeurs dans l'espace  $m + \text{Im}(\Gamma)$  avec probabilité 1.

*Démonstration.* Notons  $r$  le rang de  $\Gamma$  et  $u_{r+1}, \dots, u_d$  une base orthonormal de l'espace orthogonal de  $\text{Im}(\Gamma)$ . Alors pour tout  $r + 1 \leq k \leq d$ ,

$$\text{Var}(\langle u_k, (X - m) \rangle) = u_k^T \mathbf{E} \left( (X - m)(X - m)^T \right) u_k = u_k^T \Gamma u_k = 0 .$$

Donc la v.a.  $\langle u_k, X - m \rangle$  est constante avec probabilité 1, et comme son espérance est nulle, elle vaut zero avec probabilité 1. Ainsi,  $X - m$  est, avec probabilité 1, orthogonal au vecteur  $u_k$  pour tout  $r + 1 \leq k \leq d$ . Ce qui conclut la démonstration.  $\square$

Par suite, la loi ne peut pas posséder de densité par rapport à la mesure de Lebesgue sur  $\mathbf{R}^d$ . On dit dans ce cas que le vecteur gaussien est *dégénéré*.

*Démonstration.* En effet, supposons qu'elle en possède une, notée  $f$ . On a alors :

$$1 = \mathbf{P} [X - m \in \text{Im}(\Gamma)] = \int_{m + \text{Im}(\Gamma)} f(x) d\lambda_d(x) = 0$$

où la dernière égalité vient du fait que  $\dim(\text{Im}(\Gamma)) < d$  et donc  $\lambda_d(m + \text{Im}(\Gamma)) = 0$ . Cela conduit à une contradiction.  $\square$

## 8.7. Exercices

### Pour apprendre

EXERCICE 8.1. –

Soit deux v.a. indépendantes  $X \sim \mathcal{N}(0, 1)$  et  $Y$  de loi  $dP_Y = \frac{1}{2}(\delta_{-1} + \delta_1)$ .

1. Montrer que  $Z = YX$  est une v.a. gaussienne.
2. Montrer que  $X$  et  $Z$  sont non corrélées.
3. Si  $(X, Z)$  était un vecteur gaussien, quelle serait sa loi ?
4. Calculer la loi de  $(X, Z)$ .
5. Est-ce que  $(X, Z)$  est un vecteur gaussien ?
6. Est-ce que  $X$  et  $Z$  sont indépendantes ?

EXERCICE 8.2. –

Soit  $X_1 \sim \mathcal{N}(0, 1)$  et  $a$  le réel positif tel que  $\int_0^a \exp(-0.5x^2)dx = \sqrt{2\pi}/4$ . Soit  $X_2$  la v.a. définie par

$$X_2 = \begin{cases} X_1 & \text{si } |X_1| > a \\ -X_1 & \text{si } |X_1| \leq a \end{cases}.$$

Montrer que  $X_2 \sim \mathcal{N}(0, 1)$

Vérifier que le vecteur  $(X_1, X_2)$  n'est pas un vecteur gaussien.

Montrer que  $\text{Cov}(X_1, X_2) = 0$  mais que les v.a.  $X_1, X_2$  ne sont pas indépendantes.

### Pour s'entraîner

EXERCICE 8.3. –

Soit  $X$  et  $Y$  deux gaussiennes centrées réduites indépendantes. Montrer que les v.a.  $X + Y$  et  $\sin(X - Y)$  sont indépendantes.

EXERCICE 8.4 (Sphere hardening). –

Soit  $X_N$  un vecteur gaussien de  $\mathbf{R}^N$ , centré, réduit. Soit  $\|X_N\|$ , la norme euclidienne de  $X_N$  et  $X'_N = \|X_N\|/\sqrt{N}$ .

1. Calculer  $\mathbf{E}[(X'_N)^2]$ .
2. Calculer  $\text{Var}[(X'_N)^2]$ .
3. Montrer que, pour tout  $\eta > 0$ ,

$$\mathbf{P}(|X'_N - 1| \geq \eta) \xrightarrow{N \rightarrow +\infty} 0.$$

On pourra utiliser l'inégalité de Bienaymé-Tchebycev.

## Pour aller plus loin

EXERCICE 8.5.—

On rappelle que pour  $a > 0$ ,  $b > 0$ ,

$$B(a, b) = \int_0^1 u^{a-1}(1-u)^{b-1} du = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

On suppose que  $X_1, \dots, X_n$  sont des v.a.r., gaussiennes, indépendantes, de même loi  $\mathcal{N}(m, \sigma^2)$ . On pose

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \Sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$$

et

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

1. Soit  $I_n(z)$  la suite de fonctions définies par

$$I_0(z) = \frac{1}{\sqrt{z}}, \quad I_n(z) = \int_0^z \frac{1}{\sqrt{z-w}} I_{n-1}(w) dw \text{ pour } n \geq 1.$$

Montrer que

$$I_n(z) = \frac{\Gamma\left(\frac{1}{2}\right)^{n+1}}{\Gamma\left(\frac{n+1}{2}\right)} z^{n/2-1}.$$

2. Soit  $Y_1, \dots, Y_n$  des v.a.r., indépendantes, de même loi gaussienne  $\mathcal{N}(0, 1)$ . Calculer la loi de

$$Z = \sum_{i=1}^n Y_i^2.$$

3. Calculer la loi de  $\bar{X}$ .
4. Calculer la loi de  $(n/\sigma^2)\Sigma^2$ .
5. Montrer que  $\bar{X}$  est indépendante du vecteur  $Z = (X_1 - \bar{X}, \dots, X_n - \bar{X})$  et que  $\bar{X}$  est indépendante de  $S^2$ .
6. Maintenant on veut calculer la loi de  $(n/\sigma^2)S^2$ . Pour cela, supposer d'abord que  $m = 0$  et trouver une matrice orthogonale  $A$  telle que  $Y = AX$  et que

$$nS^2 = \sum_{i=1}^n Y_i^2 - Y_1^2.$$

Ensuite traiter le cas où  $m \neq 0$ .

EXERCICE 8.6.—

Soit  $\Gamma$  une matrice carrée réelles à  $n$  lignes. Montrer qu'il existe un vecteur gaussien de matrice de covariance  $\Gamma$  si et seulement si  $\Gamma s \cdot s \geq 0$  pour tout  $s \in \mathbf{R}^n$ . On pourra s'aider de l'exercice 5.13.

## EXERCICE 8.7. –

Dans la représentation canonique des vecteurs gaussiens, montrer que l'on peut remplacer la matrice  $A$  par une matrice de la forme  $AO$  où  $O$  est orthogonale sans changer la loi de  $AY$ .

## EXERCICE 8.8 (Polynômes d'Hermite). –

Soit  $X$  une v.a.r. gaussienne centrée, réduite et  $\varphi(t, x) = \exp(tx)$ .

1. Trouver  $g(t)$  telle que  $g(t)\mathbf{E}[\varphi(t, X)] = 1$ .

2. On pose

$$\psi(t, x) = g(t)\varphi(t, x).$$

Montrer que

$$\mathbf{E}[\psi(t, X)\psi(s, X)] = \exp(\sigma^2 ts).$$

3. Montrer que

$$\psi(t, x) = \sum_{n=0}^{\infty} \left( \sum_{k=0}^{[n/2]} \frac{x^{n-2k}}{(n-2k)!} \frac{(-\sigma^2)^k}{2^k k!} \right) t^n.$$

4. On pose

$$P_n(x) = \sum_{k=0}^{[n/2]} \frac{x^{n-2k}}{(n-2k)!} \frac{(-\sigma^2)^k}{2^k k!}.$$

Montrer que

$$\mathbf{E}[P_n(X)P_m(X)] = \delta_{n,m}.$$

## 9. Convergences

On fixe dans ce qui suit un espace probabilisé  $(\Omega, \mathcal{A}, \mathbf{P})$ .

### 9.1. Loi des grands nombres

**Définition 9.1.** On dit qu'une suite  $(X_n, n \geq 1)$  de v.a. converge  $\mathbf{P}$ -presque-sûrement (ou  $\mathbf{P}$ -presque-partout) vers une v.a.  $X$  lorsqu'il existe un ensemble  $A$  tel que  $\mathbf{P}(A^c) = 0$  et pour tout  $\omega \in A$ ,

$$X_n(\omega) \xrightarrow{n \rightarrow +\infty} X(\omega).$$

En d'autres termes, il s'agit de la convergence simple à un ensemble de mesure nulle près.

**Théorème 9.1** (Loi forte des grands nombres). Soit  $(X_n, n \geq 1)$  une suite de v.a. indépendantes, identiquement distribuées telles que  $\mathbf{E}[|X_1|] < \infty$  alors

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{n \rightarrow +\infty} \mathbf{E}[X_1], \quad \mathbf{P} - \text{p.p.}$$

### 9.2. Limite centrée

**Définition 9.2.** Pour un ensemble ouvert  $A \in \mathbf{R}^k$ , on note  $\partial A$  sa frontière définie par

$$\partial A = \bar{A} - A.$$

Pour un intervalle  $]a, b[$ , on a alors  $\partial A = \{a, b\}$ . Pour un pavé ouvert de  $\mathbf{R}^k$ , la frontière au sens topologique correspond à la notion intuitive de bord.

REMARQUE.— Si  $Y$  a même loi que  $X$  et si  $(X_n, n \geq 1)$  converge en loi vers  $X$  alors  $(X_n, n \geq 1)$  converge aussi vers  $Y$ . La convergence en loi, malgré sa présentation, n'est

pas une convergence de variables aléatoires mais une convergence des mesures associées aux v.a..

**Théorème 9.2.** La convergence presque sûre implique la convergence en loi mais la réciproque est fausse.

*Démonstration.* Si  $(X_n, n \geq 1)$  converge p.s. vers  $X$  alors pour toute fonction continue bornée,

- $f(X_n) \xrightarrow{n \rightarrow +\infty} f(X)$ , presque-sûrement,
- pour tout  $n \geq 1$ ,  $|f(X_n)| \leq \|f\|_\infty$
- et  $\mathbf{E}[\|f\|_\infty] < \infty$ ,

donc toutes les hypothèses du théorème de convergence dominée sont satisfaites, d'où

$$\mathbf{E}[f(X_n)] \xrightarrow{n \rightarrow +\infty} \mathbf{E}[f(X)].$$

D'après la première caractérisation de la convergence en loi, cela signifie que  $(X_n, n \geq 1)$  converge en loi vers  $X$ .

Construisons un contre-exemple à la réciproque. Soit  $X$  une v.a. gaussienne de moyenne nulle. Comme la densité gaussienne est paire,  $-X$  suit la même loi que  $X$ . Considérons pour tout  $n \geq 1$ , la suite  $X_n = X$ . Il est clair que  $X_n$  converge en loi vers  $X$  donc vers  $-X$ . En revanche,  $X_n$  ne converge vers  $-X$  que sur l'ensemble  $(X = -X)$ , c'est-à-dire l'ensemble  $(X = 0)$ , qui est de probabilité nulle puisque la loi gaussienne est absolument continue.  $\square$

**Définition 9.3.** On dit qu'une suite  $(X_n, n \geq 1)$  de v.a., à valeurs dans  $\mathbf{R}^k$ , converge en loi vers  $X$  lorsque l'une des propriétés équivalentes suivantes est vérifiée :

- Pour toute fonction continue bornée  $f$  de  $\mathbf{R}^k$  dans  $\mathbf{R}$ ,

$$\mathbf{E}[f(X_n)] \xrightarrow{n \rightarrow +\infty} \mathbf{E}[f(X)],$$

- pour tout ensemble ouvert  $A \in \mathbf{R}^k$  tel que  $\mathbf{P}(X \in \partial A) = 0$ ,

$$\mathbf{P}(X_n \in A) \xrightarrow{n \rightarrow +\infty} \mathbf{P}(X \in A),$$

- pour tout  $t \in \mathbf{R}^k$ ,

$$\mathbf{E}[e^{it \cdot X_n}] \xrightarrow{n \rightarrow +\infty} \mathbf{E}[e^{it \cdot X}].$$

**Théorème 9.3.** Soit  $(X_n, n \geq 1)$  une suite de v.a. indépendantes, identiquement



distribuées telles que  $\mathbf{E}[|X_1|^2] < \infty$  alors

$$\frac{\sqrt{n}}{\sigma} \left( \frac{1}{n} \sum_{j=1}^n X_j - \mathbf{E}[X_1] \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, 1), \text{ en loi}$$

où

$$\sigma^2 = \text{Var}(X_1).$$

## 9.3. Exercices

EXERCICE 9.1. — 1. Pour  $z$  réel positif, on pose

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

Soient  $0 < z_m < z_M$ , montrer que pour  $k$  entier strictement positif,  $z \in ]z_m, z_M[$ , il existe une constante  $c_k$  (que l'on ne cherchera pas à expliciter) telle que

$$\begin{aligned} |\ln^k(x) x^{z-1} e^{-x}| &\leq c_k e^{-x} \text{ pour } x \geq 1 \\ &\leq c_k \ln(x)^k x^{z_m-1} \text{ pour } x \leq 1. \end{aligned}$$

2. On admet que  $\ln^k(x) x^{z_m-1}$  est intégrable sur  $[0, 1]$ . Montrer que  $\Gamma$  est  $k$  fois dérivable sur  $\mathbf{R}^+$ .
3. Pour  $a, b$  des réels strictement positifs et  $k$  réel positif, montrer que

$$\frac{b^{-a}}{\Gamma(a)} \int_0^{+\infty} x^k \ln(x) x^{a-1} e^{-bx} dx = b^k \left( \frac{\Gamma'(a)}{\Gamma(a)} - \ln(b) \right).$$

4. Soit  $X$  la variable aléatoire dont la densité est donnée par

$$f_{\beta, \lambda, \mu}(x) = K x^\beta e^{-\lambda x^\mu} \mathbf{1}_{\mathbf{R}^+}(x).$$

On ne demande pas de calculer  $K$ . Calculer la loi de  $Y = X^\mu$ .

5. Soit  $(X_1, \dots, X_n)$   $n$  v.a.r. indépendantes et de même loi que  $X$ . Quelle est la limite presque sûre, notée  $\bar{S}$ , du couple

$$S_n = \left( \frac{1}{n} \sum_{j=1}^n \ln(X_j), \frac{1}{n} \sum_{j=1}^n X_j \right).$$

6. Quelle est la limite de

$$\frac{1}{\sqrt{n}} (S_n - \bar{S}).$$



# 10. Simulation

## 10.1. Générateurs de nombres aléatoires

Les ordinateurs usuels n'étant pas équipés d'une source physique d'aléa (mesurant l'état d'un photon par exemple), une technique courante pour simuler des variables aléatoires consiste à générer une suite de nombres  $u_1, u_2, \dots$  à valeurs dans l'ensemble  $\{0, 1, \dots, m-1\}$ , pour un entier  $m$  grand, à partir d'une valeur initiale  $u_0$  et d'une opération arithmétique simple du type  $u_{n+1} = au_n + b \pmod{m}$ . Pour des entiers  $a$  et  $b$  bien choisis, la suite  $u_1, u_2, \dots$  ressemble à une suite de variables aléatoires indépendantes de loi uniforme sur  $\{0, 1, \dots, m-1\}$  ; elle est cependant parfaitement déterministe et, de ce fait, plutôt qualifiée de suite de nombres *pseudo-aléatoires*. On notera en particulier que cette suite est périodique, de période au plus  $m$  ; il faut ainsi veiller à ne pas faire appel à plus de  $m$  fois consécutives au même générateur aléatoire, sous peine de biaiser les résultats de la simulation.

REMARQUE.— Savoir si la suite générée « ressemble » à une suite parfaitement aléatoire est un problème difficile. Idéalement, il faudrait qu'aucun test statistique ne permette de distinguer lune de l'autre avec une probabilité supérieure à  $1/2$ . La formalisation de cette propriété pour des suites grandes mais finies fait appel à la théorie de la complexité.

La valeur initiale  $u_0$  de la suite est déterminante (aux deux sens du terme). Considérons un programme informatique faisant appel 10 fois de suite au générateur aléatoire. Lorsque la valeur de  $u_0$  est fixe, deux appels successifs du même programme donneront deux suites de nombres  $u_1, u_2, \dots, u_{10}$  strictement identiques. Pour que deux appels du même programme donnent deux suites de nombres  $u_1, u_2, \dots, u_{10}$  distinctes, il faut changer la valeur initiale  $u_0$ . Une technique classique consiste à initialiser la valeur de  $u_0$  à partir de l'horloge de l'ordinateur au moment de l'appel du programme. Comme deux appels successifs du même programme ne commencent jamais exactement en même temps, les deux suites de nombres générés  $u_1, u_2, \dots, u_{10}$  seront distinctes. C'est un aspect important lorsque l'on souhaite faire plusieurs expériences aléatoires indépendantes. La valeur initiale  $u_0$  s'appelle la *graine* du générateur de nombres aléatoires.

Après normalisation de la suite  $u_1, u_2, \dots$  par  $m$ , et pourvu que  $m$  soit assez grand, on obtient une suite de nombres sur  $[0, 1]$  ayant les caractéristiques d'une suite de variables aléatoires indépendantes de loi uniforme sur  $[0, 1]$ . On peut à partir de cette suite générer des variables aléatoires réelles de loi quelconque. Nous verrons deux méthodes génériques : l'inversion de la fonction de répartition et la méthode du rejet. Des méthodes

ad-hoc existent également pour certaines lois, comme les lois gaussiennes (voir Exercice 10.4).

**Exemple du langage Java.** Un objet Java de classe `Random` fournit un générateur de nombres aléatoires avec les valeurs  $m = 2^{48}$ ,  $a = 25\,214\,903\,917$  et  $b = 11$ . La méthode `nextInt(int n)` donne un nombre entier de loi uniforme sur  $\{0, 1, \dots, m - 1\}$ ; la méthode `nextDouble` donne un nombre réel de loi uniforme sur  $[0, 1]$ . Le programme ci-dessous illustre le caractère déterminant de la graine du générateur, spécifiée en argument lors de la construction de l'objet (si rien n'est spécifié, le générateur s'initialise sur l'horloge locale de la machine) :

```
import java.util.Random;

// Suite aléatoire de 10 chiffres puis un nombre entre 0 et 1

public class Test {
    public static void main(String[] arg) {
        System.out.print("Expérience 1 : ");
        Random GenerateurAleatoire1 = new Random();
        for (int n=0;n<10;n++)
            System.out.print(GenerateurAleatoire1.nextInt(10)+", ");
        System.out.println(GenerateurAleatoire1.nextDouble());
        System.out.print("Expérience 2 : ");
        Random GenerateurAleatoire2 = new Random(0);
        for (int n=0;n<10;n++)
            System.out.print(GenerateurAleatoire2.nextInt(10)+", ");
        System.out.println(GenerateurAleatoire2.nextDouble());
    }
}
```

Deux appels successifs à ce programme donnent les résultats suivants :

```
> java Test
Expérience 1 : 2,7,0,3,7,9,0,5,4,0,0.2427885171081926
Expérience 2 : 0,8,9,7,5,3,1,1,9,4,0.3332183994766498
> java Test
Expérience 1 : 3,9,2,9,9,3,4,9,3,6,0.8265674975187485
Expérience 2 : 0,8,9,7,5,3,1,1,9,4,0.3332183994766498
```

## 10.2. Méthode d'inversion de la fonction de répartition

Cette méthode permet de générer une variable aléatoire réelle  $X$  de fonction de répartition  $F$  quelconque à partir d'une variable aléatoire  $U$  de loi uniforme sur  $[0, 1]$ . On suppose tout d'abord que la fonction  $F$  est continue et strictement croissante sur  $[a, +\infty[$ , avec  $a \in \mathbf{R}$  et  $F(a) = 0$ ; elle définit alors une bijection de  $]a, +\infty[$  vers  $]0, 1[$  et on définit la variable aléatoire  $X = F^{-1}(U)$ . Vérifions que  $X$  a bien pour fonction de répartition  $F$  :

$$\begin{aligned} \forall x > a, \quad \mathbf{P}(X \leq x) &= \mathbf{P}(F^{-1}(U) \leq x), \\ &= \mathbf{P}(U \leq F(x)), \\ &= F(x). \end{aligned}$$

EXEMPLE 34.— Pour une loi exponentielle de paramètre  $\lambda$ , on a  $F(x) = 1 - e^{-\lambda x}$  sur  $[0, +\infty[$  de sorte que :

$$X = -\frac{1}{\lambda} \ln(1 - U).$$

Graphiquement, cela revient à générer une variable aléatoire uniforme sur  $[0, 1]$  sur l'axe des ordonnées et à prendre son antécédent par la fonction  $F$  sur l'axe des abscisses, comme illustré par la Figure 10.1.

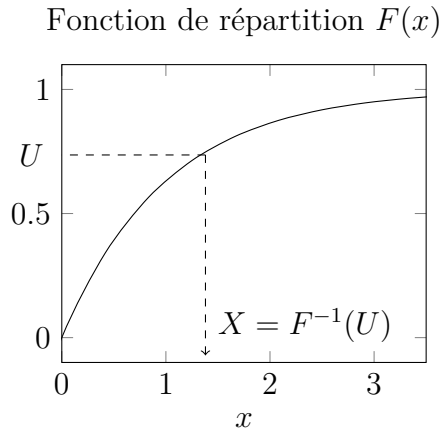


Figure 10.1. — Génération d'une variable aléatoire de loi exponentielle de paramètre 1.

Le principe est le même pour une fonction de répartition  $F$  quelconque (croissante sur  $\mathbf{R}$ , continue à droite), en prenant pour fonction inverse la fonction  $F^{-1}$  définie par :

$$\forall y \in ]0, 1[, \quad F^{-1}(y) \equiv \inf\{x \in \mathbf{R} : y \leq F(x)\},$$

puisque l'on a toujours  $F^{-1}(y) \leq x$  si et seulement si  $y \leq F(x)$ .

EXEMPLE 35.— Pour une loi de Bernoulli de paramètre  $p$ , on a :

$$\forall x \in [0, 1], \quad F(x) = \begin{cases} 1 - p & \text{si } x < 1, \\ 1 & \text{si } x = 1, \end{cases}$$

de sorte que :

$$\forall y \in ]0, 1[, \quad F^{-1}(y) = \begin{cases} 0 & \text{si } y \leq 1 - p, \\ 1 & \text{sinon,} \end{cases}$$

et

$$X = 1_{\{U > 1-p\}}.$$

Ainsi  $X$  vaut 0 si  $U < 1 - p$  et 1 sinon.

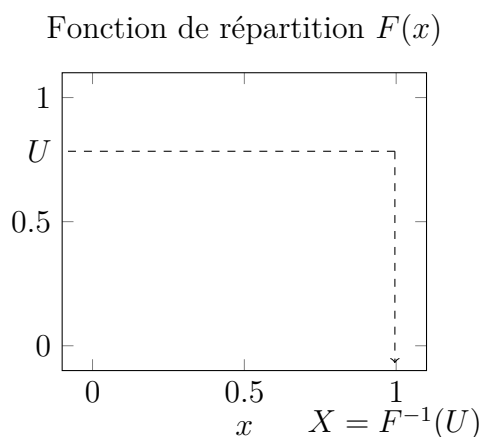


Figure 10.2. – Loi de Bernoulli de paramètre  $p = 0.8$ .

Les variables aléatoires  $U$  et  $V \equiv 1 - U$  suivant toutes deux des lois uniformes sur  $[0, 1]$ , on peut en principe partir indifféremment de l'une ou l'autre variable pour générer  $X$ . Dans les deux exemples précédents, on obtient respectivement pour la loi exponentielle de paramètre  $\lambda$  et pour la loi de Bernoulli de paramètre  $p$  :

$$X = -\frac{1}{\lambda} \ln(V) \quad \text{et} \quad X = 1_{\{V < p\}}.$$

En pratique, l'ordinateur ne peut générer une loi parfaitement uniforme du fait de sa précision finie. Il se peut en particulier que, la probabilité que  $U = 0$  soit très faible mais non nulle alors que la valeur  $U = 1$  n'est pas possible (c'est le cas en Java avec la méthode `nextDouble`). Le choix de  $U$  ou  $V \equiv 1 - U$  peut alors dépendre de ces considérations.

**Exemple du langage Java.** La classe suivante fournit des méthodes pour générer des variables de lois uniforme, Bernoulli ou exponentielle :

```
import java.util.Random;

public class Alea {
    static Random alea = new Random();
```

```

static double uniforme(){
    return alea.nextDouble();
}

static boolean bernoulli(double p){
    return (uniforme()<p);
}

static double exponentielle(double lambda){
    return (-Math.log(1-uniforme())/lambda);
}
}

```

## 10.3. Méthode du rejet

La méthode précédente permet en principe de simuler n'importe quelle loi de probabilité. Il faut cependant pouvoir inverser la fonction de répartition, ce qui n'est pas toujours facile. Pour une loi Gamma par exemple, on ne dispose même pas d'une forme analytique de cette fonction. On peut avoir recours à des méthodes numériques, mais celles-ci peuvent être coûteuses en temps de calcul et introduire des erreurs difficiles à contrôler.

Une autre option est d'appliquer la méthode du rejet, qui impose certaines contraintes sur la loi de probabilité que l'on souhaite simuler. Supposons pour simplifier que celle-ci admet une fonction de densité  $f$  sur  $\mathbf{R}$ . Il faut alors trouver une autre fonction de densité  $g$  sur  $\mathbf{R}$ , associée à une loi de probabilité que l'on sait simuler (par la méthode d'inversion de la fonction de répartition par exemple) et telle que  $f(x) \leq \theta g(x)$  pour une certaine constante  $\theta > 0$ . Remarquons que, puisque  $f$  et  $g$  sont des fonctions de densité,  $\theta \geq 1$  avec égalité si et seulement si  $f = g$ . Soit  $U_1, U_2, \dots$  une suite de variables aléatoires indépendantes de loi uniforme sur  $[0, 1]$  et  $Y_1, Y_2, \dots$  une suite de variables aléatoires indépendantes de densité  $g$ , indépendante de la précédente. On pose :

$$X = Y_K \quad \text{avec } K = \min\{k \geq 1 : f(Y_k) > \theta U_k g(Y_k)\}.$$

Malgré les apparences, la variable aléatoire  $X$  n'a pas la même loi que les variables aléatoires  $Y_1, Y_2, \dots$  car l'indice  $K$  est lui-même aléatoire. Nous allons voir que  $X$  a bien pour densité  $f$ . Notons que l'indice  $K$  correspond au coût de simulation, puisqu'il est fait appel  $2K$  fois au générateur aléatoire. Le résultat suivant montre en particulier que  $\mathbf{E}(K) = \theta$  si bien qu'il faut choisir la constante  $\theta$  aussi petite que possible pour minimiser le coût de simulation.

**Proposition 10.1.** La variable aléatoire  $X$  a pour densité  $f$ , la variable aléatoire  $K$  suit une loi géométrique de paramètre  $1/\theta$ , et ces deux variables sont indépendantes.

*Démonstration.* Soit  $A_k = \{f(Y_k) > \theta U_k g(Y_k)\}$ . On a pour tout  $x \in \mathbf{R}$  et tout  $k \in \mathbf{N}^*$  :

$$\begin{aligned} \mathbf{P}(\{X \leq x\} \cap \{K = k\}) &= \mathbf{P}(\{Y_k \leq x\} \cap A_k \cap A_1^c \cap \dots \cap A_{k-1}^c), \\ &= \mathbf{P}(\{Y_k \leq x\} \cap A_k)(1-p)^{k-1}, \end{aligned}$$

où l'on a utilisé l'indépendance des couples de variables aléatoires  $(U_1, Y_1), (U_2, Y_2), \dots$  et noté  $p$  la probabilité commune des événements  $A_1, A_2, \dots$ . On écrit ensuite :

$$\begin{aligned} \mathbf{P}(\{Y_k \leq x\} \cap A_k) &= \mathbf{E} \left[ 1_{\{Y_k \leq x\}} 1_{A_k} \right], \\ &= \int_{-\infty}^x \left( \int_0^1 1_{\{f(y) > \theta u g(y)\}} du \right) g(y) dy, \\ &= \int_{-\infty}^x \frac{f(y)}{\theta g(y)} g(y) dy, \\ &= \frac{1}{\theta} \int_{-\infty}^x f(y) dy, \end{aligned}$$

de sorte que :

$$\mathbf{P}(\{X \leq x\} \cap \{K = k\}) = \int_{-\infty}^x f(y) dy \times \frac{1}{\theta} (1-p)^{k-1}.$$

En faisant d'une part tendre  $x$  vers  $+\infty$ , on obtient :

$$\mathbf{P}(K = k) = \frac{1}{\theta} (1-p)^{k-1},$$

si bien que  $K$  suit une loi géométrique de paramètre  $p = 1/\theta$ ; en sommant d'autre part sur  $k$ , on obtient :

$$\mathbf{P}(X \leq x) = \int_{-\infty}^x f(y) dy,$$

si bien que la variable aléatoire  $X$  a pour densité  $f$ . Comme

$$\mathbf{P}(\{X \leq x\} \cap \{K = k\}) = \mathbf{P}(X \leq x) \mathbf{P}(K = k),$$

ces variables aléatoires sont indépendantes. □

**EXEMPLE 36.**— La loi Gamma de paramètres  $(2, \lambda)$  a pour densité la fonction  $f(x) = x e^{-\lambda x}$  sur  $\mathbf{R}^+$ . On borne cette fonction, à une constante multiplicative  $\theta$  près, par la densité d'une loi exponentielle de paramètre  $\lambda/2$ , que l'on sait facilement simuler :

$$\forall x \geq 0, \quad f(x) \leq \frac{\theta}{2} e^{-\frac{\lambda}{2} x}.$$

La plus petite valeur de  $\theta$  est  $4/e$ . Les résultats obtenus sont illustrés par la Figure 10.3.

**Exemple du langage Java.** On obtient pour la loi Gamma de paramètres  $(2, \lambda)$  :



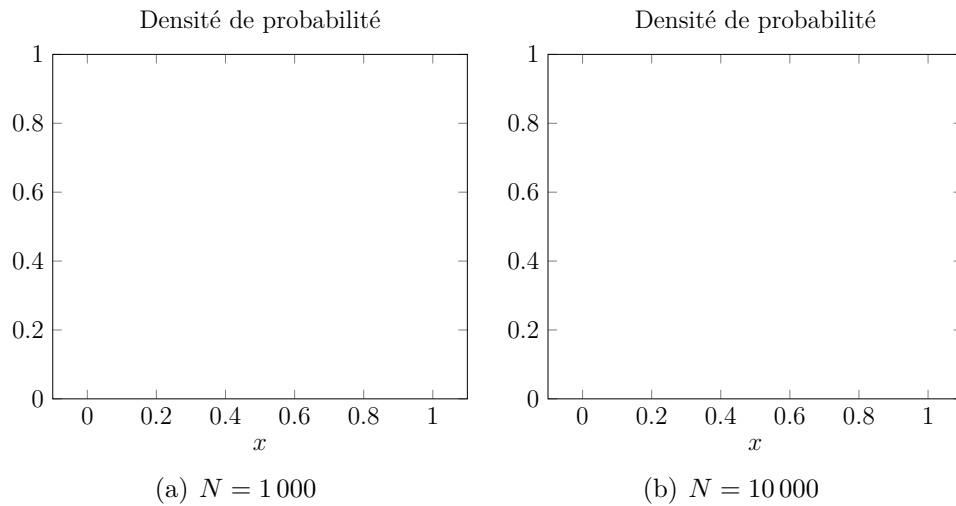


Figure 10.3. – Loi empirique issue de  $N$  échantillons de la loi Gamma de densité  $f(x) = xe^{-x}$  sur  $\mathbb{R}_+$  obtenus par la méthode du rejet.

```
import java.util.Random;

public class Alea {
    static Random alea = new Random();

    static double uniforme(){
        return alea.nextDouble();
    }

    static double exponentielle(double lambda){
        return (-Math.log(1-uniforme())/lambda);
    }

    static double gamma2(double lambda){
        double y=exponentielle(lambda/2);
        while (y*Math.exp(-lambda*y/2+1)<4*uniforme())
            y=exponentielle(lambda/2);
        return y;
    }
}
```

## 10.4. Méthodes de Monte Carlo

Les méthodes de Monte Carlo sont des techniques d'analyse numérique reposant sur la loi forte des grands nombres. Supposons par exemple que l'on cherche à calculer l'intégrale d'une fonction  $g$  continue sur  $[0, 1]$ . En notant  $U_1, U_2, \dots$  une suite de variables aléatoires

indépendantes de loi uniforme sur  $[0, 1]$ , on sait par la loi forte des grands nombres que :

$$\frac{1}{N} \sum_{j=1}^N g(U_j) \xrightarrow{\text{p.s.}} \mathbf{E}[g(U_1)] = \int_0^1 g(u) du.$$

On peut donc utiliser l'estimateur suivant, pour  $N$  grand :

$$\int_0^1 g(u) du \approx \frac{1}{N} \sum_{j=1}^N g(U_j).$$

**Estimation du nombre  $\pi$ .** Un exemple classique illustrant la méthode de Monte Carlo consiste à estimer le nombre  $\pi$  à partir de l'expression suivante :

$$\pi \approx \frac{4}{N} \sum_{j=1}^N 1_{\{U_j^2 + V_j^2 < 1\}},$$

où  $U_1, U_2, \dots$  et  $V_1, V_2, \dots$  désignent deux suites indépendantes de variables aléatoires indépendantes de loi uniforme sur  $[0, 1]$ . On a en effet par la loi forte des grands nombres :

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^N 1_{\{U_j^2 + V_j^2 < 1\}} &\xrightarrow{\text{p.s.}} \mathbf{E}[1_{\{U_1^2 + V_1^2 < 1\}}], \\ &= \int_0^1 \int_0^1 1_{\{u^2 + v^2 < 1\}} du dv, \\ &= \frac{\pi}{4}. \end{aligned}$$

Ainsi en jettant  $N$  points au hasard dans un carré de côté 1, la fraction des points tombant dans le quart de disque de rayon 1 inclus dans ce carré tend vers  $\pi/4$ , comme illustré par la Figure 10.4. La Figure 10.5 montre l'évolution de la qualité de l'estimation en fonction de  $N$ .

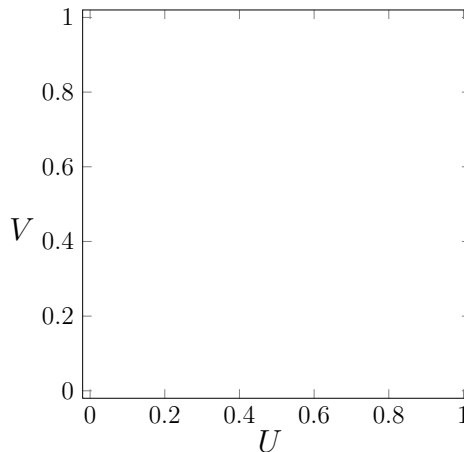


Figure 10.4. – Estimation du nombre  $\pi$  par une méthode de Monte Carlo sur la base du tirage de  $N = 100$  points aléatoires.

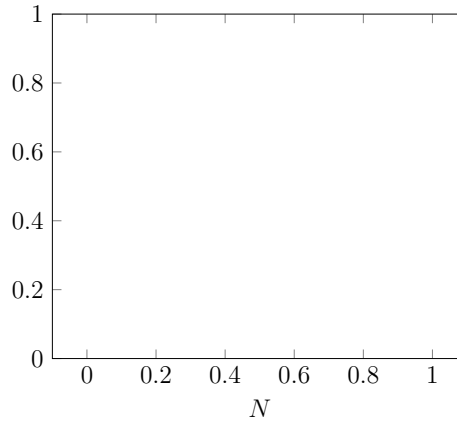


Figure 10.5. – Evolution de l'estimation du nombre  $\pi$  en fonction du nombre de points  $N$ .

## 10.5. Histogrammes

Pour vérifier si la simulation produit bien la loi de probabilité recherchée, on peut tracer l'histogramme obtenu à partir d'un nombre  $N$  d'échantillons  $x_1, \dots, x_N$ , censés correspondre à  $N$  réalisations d'une variable aléatoire réelle  $X$ . On prend pour simplifier des intervalles de même largeur  $\delta > 0$  et on note  $g_k$  la fraction des échantillons se trouvant dans l'intervalle  $[k\delta, (k+1)\delta[$  :

$$\forall k \in \mathbf{Z}, \quad g_k = \frac{1}{N} \sum_{j=1}^N 1_{\{k\delta \leq x_j < (k+1)\delta\}}.$$

Lorsque le nombre d'échantillons  $N$  est grand,  $g_k$  doit être proche de la probabilité que  $X$  soit dans l'intervalle  $[k\delta, (k+1)\delta[$ . Par la loi forte des grands nombres, on sait en effet que pour toute suite de variables aléatoires indépendantes  $X_1, X_2, \dots$  de même loi que  $X$  :

$$\frac{1}{N} \sum_{j=1}^N 1_{\{k\delta \leq X_j < (k+1)\delta\}} \xrightarrow{\text{p.s.}} \mathbf{E} \left[ 1_{\{k\delta \leq X < (k+1)\delta\}} \right] = \mathbf{P}(k\delta \leq X < (k+1)\delta).$$

Lorsque la variable aléatoire  $X$  a pour densité  $f$ , et pour une largeur d'intervalle  $\delta$  suffisamment faible, on peut comparer  $f$  à la fonction de densité  $g$  constante par morceaux définie par :

$$\forall x \in \mathbf{R}, \quad g(x) = \frac{1}{\delta} \sum_{k \in \mathbf{Z}} g_k 1_{[k\delta, (k+1)\delta[}(x).$$

C'est cette fonction qui est tracée en Figure 10.3, pour des intervalles de largeur  $\delta = 0,1$ .

**Illustration du théorème central limite.** Soit  $Y_1, Y_2, \dots$  une suite de variables indépendantes suivant une loi de Bernoulli de paramètre  $p$ . On sait par la loi forte des grands nombres que leur moyenne empirique tend vers  $p$  et par le théorème central limite que la

loi de l'écart de cette moyenne empirique par rapport à  $p$ , après renormalisation adéquate, tend vers une loi gaussienne. Plus spécifiquement,

$$\frac{\sqrt{n}}{\sigma} \left( \frac{1}{n} \sum_{j=1}^n Y_j - p \right) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, 1), \text{ en loi,}$$

avec  $\sigma^2 = \text{Var}(Y_1) = p(1-p)$ . Ce résultat est illustré par la Figure 10.6, où l'histogramme de l'écart normalisé de la moyenne empirique de  $n$  variables  $Y_1, \dots, Y_n$  par rapport à  $p$ , soit

$$X = \frac{\sqrt{n}}{\sigma} \left( \frac{1}{n} \sum_{j=1}^n Y_j - p \right),$$

est représenté avec une précision  $\delta = 0,1$  à partir de  $N = 10\,000$  échantillons de cette variable aléatoire.

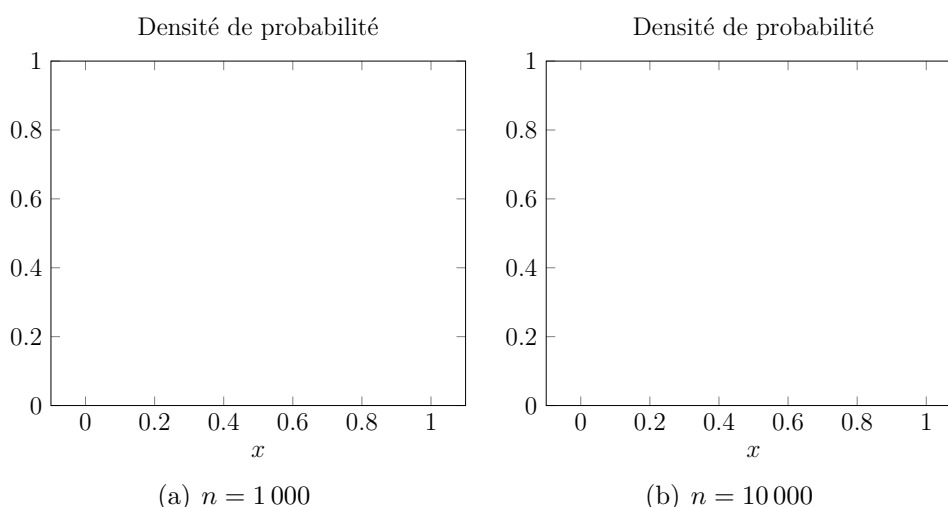


Figure 10.6. – Histogramme formé à partir de  $N = 10\,000$  échantillons des écarts normalisés de la moyenne empirique de  $n$  variables indépendantes de Bernoulli de paramètre  $1/2$  par rapport à  $1/2$ , et comparaison avec la loi normale centrée réduite.

## 10.6. Exercices

EXERCICE 10.1. –

Donner une fonction générant une variable aléatoire de loi de Cauchy à partir d'une variable aléatoire de loi uniforme sur  $[0, 1]$ .

EXERCICE 10.2. –

Écrire en pseudo-code une fonction générant une variable aléatoire discrète à valeurs dans  $\{1, \dots, N\}$  dont la loi est donnée par un vecteur de probabilité  $(p_1, \dots, p_N)$ .

EXERCICE 10.3. –

Écrire en pseudo-code une fonction générant une variable aléatoire à valeurs dans  $[0, 1]$ , de fonction de répartition  $F(x) = (x + 1)/2$ .

EXERCICE 10.4. –

Soit  $Z$  et  $\Theta$  deux variables aléatoires indépendantes de lois respectives exponentielle de paramètre 1 et uniforme sur  $[0, 2\pi]$ . Montrer que les variables aléatoires  $X = \sqrt{Z/2} \cos(\Theta)$  et  $Y = \sqrt{Z/2} \sin(\Theta)$  sont indépendantes et de même loi normale centrée réduite. En déduire une méthode de simulation d'une loi gaussienne quelconque.

EXERCICE 10.5. –

Appliquer la méthode du rejet à la loi géométrique. Interpréter le résultat.

EXERCICE 10.6. –

Appliquer la méthode du rejet à la loi bêta.



# Annexes





# A. Ensembles

## A.1. Opérations sur les ensembles

Un événement est décrit comme un sous-ensemble de  $\Omega$ . Les opérations sur les événements se ramènent donc aux opérations habituelles sur les ensembles.

### A.1.1. Rappels

Soient  $A, B, C$  des ensembles. On rappelle les définitions suivantes.

**Définition A.1.** Les définitions et notations suivantes sont usuelles.

- Le *complémentaire* de  $A$  est défini par  $A^c = \{\omega \in \Omega : \omega \notin A\}$ . L'événement  $A^c$  est réalisé si et seulement si  $A$  ne l'est pas. On a  $\emptyset = \Omega^c$ .
- L' *union* de  $A$  et  $B$  est définie par  $A \cup B = \{\omega \in \Omega : \omega \in A \text{ ou } \omega \in B\}$ . L'événement  $A \cup B$  est réalisé si et seulement si  $A$  OU  $B$  le sont.
- L' *intersection* de  $A$  et  $B$  est définie par  $A \cap B = \{\omega \in \Omega : \omega \in A \text{ et } \omega \in B\}$ . L'événement  $A \cap B$  est réalisé si et seulement si  $A$  ET  $B$  le sont.
- L'ensemble  $A \setminus B = A \cap B^c$  est appelé «  $A$  privé de  $B$  ». Il s'agit de l'ensemble des éléments de  $A$  qui n'appartiennent pas à  $B$ .
- L'ensemble  $A \times B = \{(a, b) : a \in A, b \in B\}$  est appelé le produit cartésien de  $A$  et de  $B$ .
- L'ensemble des parties d'un ensemble  $\Omega$  est noté  $\mathcal{P}(\Omega)$  ou  $2^\Omega$ .

**Proposition A.1.** On rappelle les propriétés suivantes.

**Commutativité :**  $A \cup B = B \cup A$  et  $A \cap B = B \cap A$ .

**Associativité :**  $A \cup (B \cup C) = (A \cup B) \cup C$  et  $A \cap (B \cap C) = (A \cap B) \cap C$ .

**Distributivité :**

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) \quad \text{et} \quad (A \cap B) \cup C = (A \cup B) \cap (A \cup C) .$$

**Lois de Morgan :**  $(A \cup B)^c = A^c \cap B^c$  et  $(A \cap B)^c = A^c \cup B^c$ .

Deux ensembles  $A$  et  $B$  sont dits *incompatibles* ou *disjoints* si  $A \cap B = \emptyset$ .

### A.1.2. Familles d'ensembles

Soit  $I$  un ensemble non vide. Soit  $(A_i)_{i \in I}$  une famille d'ensembles indexés par  $I$ . On appelle respectivement *union* et *intersection* de la famille  $(A_i)_{i \in I}$  les ensembles :

$$\begin{aligned} \bigcup_{i \in I} A_i &:= \{\omega : \exists i \in I, \omega \in A_i\} \\ \bigcap_{i \in I} A_i &:= \{\omega : \forall i \in I, \omega \in A_i\} . \end{aligned}$$

Les éléments de la famille  $(A_i)_{i \in I}$  sont dits *deux à deux disjoints* si pour tout  $i \neq j$ ,  $A_i$  et  $A_j$  sont disjoints.

**Proposition A.2.** (Distributivité)

$$\left( \bigcup_{i \in I} A_i \right) \cap B = \bigcup_{i \in I} (A_i \cap B) \text{ et } \left( \bigcap_{i \in I} A_i \right) \cup B = \bigcap_{i \in I} (A_i \cup B) .$$

(Lois de De Morgan)

$$\left( \bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} A_i^c \text{ et } \left( \bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} A_i^c .$$

Le produit cartésien  $\prod_{i \in I} A_i$  est défini comme l'ensemble des familles  $(a_i)_{i \in I}$  où  $a_i \in A_i$ .

Un ensemble  $I$  est dit *dénombrable* s'il est en bijection avec  $\mathbb{N}$  (citons par exemple  $\mathbb{N}$ ,  $\mathbb{N}^*$ ,  $\mathbb{Z}$  ou  $\mathbb{Q}$ ). Il est dit *au plus dénombrable* (ou parfois *discret*) s'il est fini ou dénombrable. Une famille  $(A_i)_{i \in I}$  est dite *dénombrable* si  $I$  est dénombrable.

**Définition A.2.** Une *partition* d'un ensemble  $\Omega$  est une famille  $(A_i)_{i \in I}$  d'ensembles deux à deux disjoints telle que  $\cup_{i \in I} A_i = \Omega$ .

### A.1.3. Suites et limites

Une suite d'ensembles  $(A_n)_{n \in \mathbb{N}}$  est dite *croissante* si pour tout  $n \in \mathbb{N}$ ,  $A_n \subset A_{n+1}$ .

L'union  $\bigcup_{n \in \mathbb{N}} A_n$  est aussi appelée la *limite* de la suite croissante  $A_n$  et on la note  $\lim_{n \rightarrow \infty} A_n$ . On utilise la notation  $A_n \uparrow A$  pour signifier que  $(A_n)_{n \in \mathbb{N}}$  est une suite croissante et que  $A = \lim_{n \rightarrow \infty} A_n$ .

Une suite d'ensembles  $(A_n)_{n \in \mathbb{N}}$  est dite *décroissante* si pour tout  $n \in \mathbb{N}$ ,  $A_{n+1} \subset A_n$ .

L'intersection  $\bigcap_{n \in \mathbb{N}} A_n$  est aussi appelée la *limite* de la suite décroissante  $A_n$  et on la note  $\lim_{n \rightarrow \infty} A_n$ . On utilise la notation  $A_n \downarrow A$  pour signifier que  $(A_n)_{n \in \mathbb{N}}$  est une suite décroissante et que  $A = \lim_{n \rightarrow \infty} A_n$ .

Une suite est dite *monotone* si elle est croissante ou décroissante.

REMARQUE.— L'annexe B.1 montre que la notion de limite d'une suite d'ensembles peut être étendue à une classe plus large que les seules suites monotones. Pour toute suite  $(A_n)_n$  on définit  $\limsup_n A_n$  et  $\liminf_n A_n$  comme la limite des suites respectivement décroissantes et croissantes  $\bigcup_{k \geq n} A_k$  et  $\bigcap_{k \geq n} A_k$ . Lorsque  $\limsup_n A_n = \liminf_n A_n$ , on note cette quantité  $\lim_n A_n$ . On vérifie que dans le cas de suites monotones, cette définition est bien cohérente avec celle donnée plus haut.

## A.2. Espaces d'états dénombrables

**Définition A.3.** Un ensemble  $E$  est dit dénombrable s'il est en bijection avec  $\mathbf{N}$ , l'ensemble des entiers naturels.  
Il est dit *au plus* dénombrable s'il est inclus dans un ensemble dénombrable.

Quelques exemples :

- Les ensembles de cardinal fini sont évidemment au plus dénombrables. Ceci recouvre non seulement les ensembles de la forme  $\{1, \dots, n\}$  mais aussi des produits cartésiens d'ensembles de cette forme ou des ensembles comme celui des permutations sur un ensemble à  $n$  éléments.
- $2\mathbf{N}$ , l'ensemble des entiers pairs, est dénombrable.
- $\mathcal{P}(\mathbf{N})$ , l'ensemble des parties de  $\mathbf{N}$ , ne l'est pas. Il est en bijection avec l'ensemble des réels  $\mathbf{R}$ . En effet, pour une partie  $A$  de  $\mathbf{N}$ , on peut définir « la suite indicatrice »  $\chi_A(n)$  par

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

Puis à cette suite, on peut associer le réel  $x_A$  défini par :

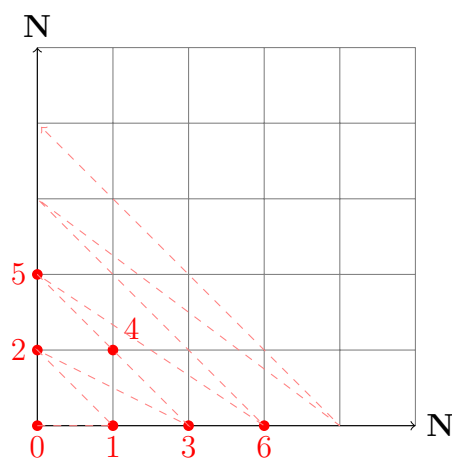
$$x_A = \sum_{n=1}^{\infty} 2^{-n} \chi_A(n).$$

Il est évident que  $\mathcal{P}(\mathbf{N})$  est alors en bijection avec  $\{0, 1\}^{\mathbf{N}}$ . Etant donné qu'à tout réel entre 0 et 1, on peut associer son développement diadique, il existe une injection  $[0, 1]$  dans  $\{0, 1\}^{\mathbf{N}}$  donc  $\mathcal{P}(\mathbf{N})$  n'est pas dénombrable.

La dénombrabilité est stable par réunion et produit cartésien.

**Théorème A.3.** Si  $E$  et  $F$  sont deux ensembles dénombrables alors  $E \times F$  et  $E \cup F$  sont des ensembles dénombrables.

Tout est basé sur le résultat essentiel qui stipule de  $\mathbf{N} \times \mathbf{N}$  est en bijection avec  $\mathbf{N}$ . La bijection est construite comme indiquée sur la figure A.1. Par exemple, l'élément  $(2, 0)$  est envoyé sur 3 et l'élément  $(0, 2)$  est envoyé sur 5.

Figure A.1. – Bijection de  $\mathbf{N} \times \mathbf{N}$  avec  $\mathbf{N}$ .

**Corollaire A.4.** L'ensemble des entiers relatifs,  $\mathbf{Z}$ , est dénombrable.

Le théorème suivant est loin d'être trivial (cf. [Kut98]).

**Théorème A.5.** S'il existe une injection de  $E$  dans  $F$  et une injection de  $F$  dans  $E$  alors il existe une bijection de  $E$  dans  $F$ .

**Corollaire A.6.** L'ensemble des rationnels,  $\mathbf{Q}$  est dénombrable.

*Démonstration.* Il existe évidemment une injection de  $\mathbf{N}$  dans  $\mathbf{Q}$  et par construction de  $\mathbf{Q}$ , il existe une injection de  $\mathbf{Q}$  dans  $\mathbf{Z} \times \mathbf{Z}$ , qui d'après le théorème précédent est en bijection avec  $\mathbf{N}$ . Par conséquent,  $\mathbf{Q}$  est en bijection avec  $\mathbf{N}$ .  $\square$

## B. Notions utiles d'analyse

### B.1. Limite supérieure et limite inférieure

#### B.1.1. Limite inférieure et limite supérieure d'une suite

La limite inférieure et limite supérieure sont des quantités qu'on définit naturellement pour des suites réelles. La limite inférieure d'une suite, communément appelée *liminf*, est sa plus petite valeur d'adhérence ; la limite supérieure (*limsup*) est, elle, sa plus grande valeur d'adhérence. Ces quantités sont toujours définies (elles peuvent néanmoins prendre les valeurs  $\pm\infty$ ) et c'est là leur intérêt principal. En effet, contrairement à la limite d'une suite, que l'on ne peut pas manipuler *a priori* (il faut d'abord montrer la convergence de la suite), on peut toujours manipuler la *liminf* et la *limsup*. Ces notions s'étendent naturellement à des suites de fonctions réelles, et à des familles d'ensembles.

Notons  $\overline{\mathbf{R}}$  la droite réelle complétée :  $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty, \infty\}$ .

**Définition B.1.** La limite inférieure et la limite supérieure d'une suite numérique  $(u_n, n \in \mathbf{N})$  sont les éléments de  $\overline{\mathbf{R}}$ , notés  $\liminf_{n \rightarrow \infty} u_n$  et  $\limsup_{n \rightarrow \infty} u_n$  et définis par :

$$\liminf_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} \inf_{p \geq n} u_p, \quad \limsup_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} \sup_{p \geq n} u_p.$$

On remarque immédiatement que la limite inférieure (resp. supérieure) d'une suite  $(u_n)$  existe toujours dans  $\overline{\mathbf{R}}$  : c'est simplement la limite de la suite croissante  $\alpha_n = \inf_{p \geq n} u_p$  (resp. de la suite décroissante  $\beta_n = \sup_{p \geq n} u_p$ ).

On rappelle que  $\ell \in \overline{\mathbf{R}}$  est une valeur d'adhérence de la suite  $(u_n)$  s'il existe une sous-suite extraite  $(u_{\phi(n)})$  de  $(u_n)$  telle que  $\lim_{n \rightarrow \infty} u_{\phi(n)} = \ell$

**Lemme B.1.** La limite inférieure de la suite  $(u_n)$  est sa plus petite valeur d'adhérence ; sa limite supérieure est sa plus grande valeur d'adhérence.

**Corollaire B.2.** Si  $\limsup_{n \rightarrow \infty} u_n = \liminf_{n \rightarrow \infty} u_n = \ell \in \overline{\mathbf{R}}$ , alors la suite  $(u_n)$  converge vers  $\ell$ .

### B.1.2. Limite supérieure et inférieure d'une suite de fonctions

Etant donnée une suite de fonctions  $(f_n)$  à valeurs  $\mathbf{R}$  ou  $\overline{\mathbf{R}}$ , on peut définir ses limites inférieure et supérieure, en posant, à  $x$  fixé :

$$\underline{f}(x) = \liminf_n f_n(x), \quad \overline{f}(x) = \limsup_{n \rightarrow \infty} f_n(x).$$

Les fonctions  $\underline{f}$  et  $\overline{f}$  définies ainsi pour chaque  $x$  sont naturellement à valeurs  $\overline{\mathbf{R}}$ .

### B.1.3. Limite supérieure et inférieure d'une famille d'ensembles

Soit  $(A_n)_{n \in \mathbf{N}}$  une famille d'ensembles, on définit la limsup et la liminf de la famille  $(A_n)_{n \in \mathbf{N}}$  par :

$$\begin{aligned} \limsup_n A_n &= \bigcap_{n \in \mathbf{N}} \bigcup_{k \geq n} A_k \\ \liminf_n A_n &= \bigcup_{n \in \mathbf{N}} \bigcap_{k \geq n} A_k \end{aligned}$$

**Remarque 4 :** On interprète facilement la limsup de  $A_n$  comme étant l'ensemble des points qui appartiennent à une infinité de  $A_n$ . De même, la liminf des  $A_n$  s'interprète comme l'ensemble des points qui appartiennent à tous les  $A_n$  sauf un nombre fini d'entre eux. On en déduit que  $\liminf_n A_n \subset \limsup_n A_n$ .

## B.2. Séries

### B.2.1. Généralités sur les séries

1. Soit  $(u_n, n \geq 0)$  une suite numérique et  $S_n = \sum_{i=0}^n u_i$  la somme partielle à l'ordre  $n$ . La série  $\sum_{n \geq 0} u_n$  est dite convergente si la limite  $S$  de  $S_n$  existe ; cette limite est notée  $\sum_{n \geq 0} u_n$  :

$$S = \lim_{n \rightarrow \infty} S_n = \sum_{n \geq 0} u_n.$$

Le nombre  $u_n$  est appelé *terme général* de la série et la limite  $S$  d'une série convergente est appelée sa *somme*.

2. Le terme général  $u_n$  d'une série convergente tend vers zéro car  $u_n = S_n - S_{n-1}$ . La réciproque est fautive : la série de terme général  $\frac{1}{n}$  (défini pour  $n \geq 1$ ) *diverge*, i.e. la limite de  $S_n$  est égale à  $\infty$  car  $\sum_{i=1}^n \frac{1}{i} \geq \int_1^n \frac{dt}{t} = \ln(n)$ .
3. La série  $\sum_n u_n$  est dite *absolument convergente* si la série  $\sum_n |u_n|$  converge.
4. Soit  $\sum_n u_n$  une série de terme général positif :  $u_n \geq 0$ . Alors  $S_n$  est croissante et sa limite existe toujours, bien que pouvant être infinie. On la note encore  $S = \sum_n u_n$  mais on ne parlera de série convergente que dans le cas où  $S < \infty$ .

### B.2.2. Séries entières - rappels et calculs de sommes

On rappelle qu'étant donnée une suite réelle  $(u_n)$ , il existe un nombre  $R \in [0, \infty]$  tel que la série  $\sum_{n \geq 0} u_n x^n$  converge absolument si  $|x| < R$  et diverge si  $|x| > R$ . Le nombre  $R$  est appelé rayon de convergence de la série entière  $\sum_{n \geq 0} u_n x^n$ ; il est donné par le critère de Cauchy :

$$\frac{1}{R} = \limsup_{n \rightarrow \infty} |u_n|^{1/n}.$$

Deux exemples bien connus sont la fonction exponentielle :

$$\exp(x) = \sum_{n \geq 0} \frac{x^n}{n!}, \quad R = \infty,$$

et la série géométrique :

$$\frac{1}{1-x} = \sum_{n \geq 0} x^n.$$

La fonction  $f(x) = \sum_{n \geq 0} u_n x^n$  définie pour tout  $x$  tel que  $|x| < R$  est infiniment dérivable dans l'intervalle  $] -R, R[$  et sa dérivée est donnée par la dérivation terme à terme de la série :

$$f'(x) = \sum_{n \geq 1} n u_n x^{n-1}$$

Cette propriété permet le calcul des sommes suivantes :

$$\frac{1}{(1-x)^2} = \sum_{n \geq 1} n x^{n-1} \quad \text{et} \quad \frac{1}{(1-x)^3} = \sum_{n \geq 2} n(n-1) x^{n-2}.$$

Ces sommes permettent le calcul de quantités utiles en probabilité :

$$\begin{aligned} \sum_{n \geq 0} n x^n &= x \sum_{n \geq 1} n x^{n-1} = \frac{x}{(1-x)^2}, \\ \sum_{n \geq 0} n^2 x^n &= x^2 \sum_{n \geq 2} n(n-1) x^{n-2} + x \sum_{n \geq 1} n x^{n-1}, \\ &= \frac{2x^2}{(1-x)^3} + \frac{x}{(1-x)^2} = \frac{x(x+1)}{(1-x)^3}. \end{aligned}$$

## B.3. Convexité

Ce qui suit n'est qu'un tout petit aperçu de la très riche théorie des fonctions convexes. On pourra se référer aux ouvrages [Roc] ou [ET99] pour plus de détails.

Étant donné un espace vectoriel normé  $X$ , on dit qu'un ensemble  $C \subset X$  est convexe si

$$\forall x, y \in C, \forall \alpha, \beta \geq 0 : \alpha + \beta = 1, \quad \alpha x + \beta y \in C.$$

**Définition B.2.** Une fonction d'un ensemble convexe  $C \subset X$  à valeurs  $\overline{\mathbf{R}}$  est dite convexe si et seulement si la propriété suivante est vérifiée :

$$\forall x, y \in C ; \forall \alpha, \beta \geq 0, \alpha + \beta = 1, \quad f(\alpha x + \beta y) \leq \alpha f(x) + \beta f(y) .$$

La fonction est dite strictement convexe dès lors que l'inégalité précédente est stricte pour  $0 < \alpha < 1$  et  $x \neq y$ .

Une fonction est dite concave si son opposé est convexe ou de manière équivalente si

$$\forall x, y \in C ; \forall \alpha, \beta \geq 0, \alpha + \beta = 1, \quad f(\alpha x + \beta y) \geq \alpha f(x) + \beta f(y) .$$

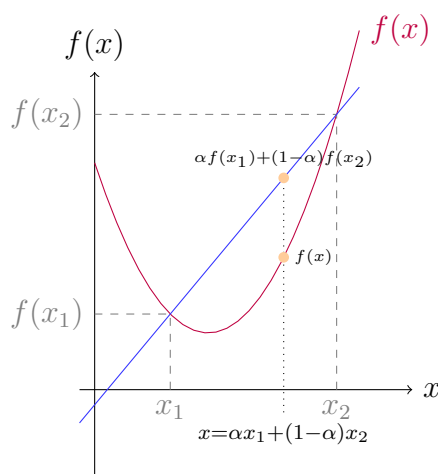


Figure B.1. – Illustration de la propriété de convexité.

EXEMPLE 37. – Les fonctions affines sont convexes (mais non strictement convexes), les fonctions de la forme  $f(x) = x^2 + ax + b$  aussi. Plus généralement, une fonction deux fois différentiable est convexe si et seulement si sa dérivée seconde est positive. On en déduit par exemple que

$$f(x) = |x|^p, \text{ pour } p > 1, \quad f(x) = x \ln(x) - x + 1, \quad x > 0$$

sont convexes. Du premier exemple, en appliquant la définition de la convexité pour  $\alpha = 1/2$ , on déduit que

$$|x + y|^p \leq 2^{p-1}(|x|^p + |y|^p).$$

Les fonctions convexes à valeurs réelles ont de bonnes propriétés de régularité, en particulier, elles admettent en tout point une dérivée à gauche et une dérivée à droite :

**Lemme B.3.** Soit  $f : \mathbf{R} \rightarrow \mathbf{R}$  une fonction convexe, alors les dérivées

$$f'_g(x) = \lim_{u \uparrow x} \frac{f(x) - f(u)}{x - u} \quad \text{et} \quad f'_d(x) = \lim_{v \downarrow x} \frac{f(v) - f(x)}{v - x}$$

existent et vérifient  $f'_g(x) \leq f'_d(x)$ .



**Preuve.** Supposons que  $u < v$ , alors

$$\frac{f(x) - f(u)}{x - u} \leq \frac{f(v) - f(x)}{v - x} . \quad (\text{B.1})$$

Cette inégalité traduit simplement le fait que pour une fonction convexe, le coefficient directeur de la droite entre  $u$  et  $x$  est plus petit que celui entre  $v$  et  $x$ . On notera que  $u$  et  $v$  peuvent être du même côté par rapport à  $x$  ou de part et d'autre de  $x$ . L'idée pour établir l'inégalité (B.1) est d'exprimer le point intermédiaire comme barycentre des deux autres. Dans le cas où on a, par exemple,  $u < x < v$ , alors

$$x = \left( \frac{v - x}{v - u} \right) u + \left( \frac{x - u}{v - u} \right) v \quad \Rightarrow \quad f(x) \leq \left( \frac{v - x}{v - u} \right) f(u) + \left( \frac{x - u}{v - u} \right) f(v) ,$$

ce qui entraîne immédiatement (B.1) en notant que

$$f(x) = \left( \frac{v - x}{v - u} \right) f(x) + \left( \frac{x - u}{v - u} \right) f(x) ;$$

les cas  $x < u < v$  et  $u < v < x$  se traitent de la même manière.

Considérons maintenant  $u < x < v$ , alors le ratio  $\frac{f(x) - f(u)}{x - u}$  est croissant lorsque  $u \uparrow x$  et majoré par  $\frac{f(v) - f(x)}{v - x}$  pour tout  $v > x$ . Par suite la limite

$$f'_g(x) = \lim_{u \uparrow x} \frac{f(x) - f(u)}{x - u}$$

existe. Le même raisonnement s'adapte pour  $f'_d(x)$ , et l'inégalité entre les deux dérivées s'obtient immédiatement comme passage à la limite dans (B.1).

Une propriété des fonctions convexes définies sur  $\mathbf{R}$  est particulièrement utile. Pour l'exprimer, introduisons la famille de fonctions affines  $(\Delta_{a,b}; a, b \in \mathbf{R})$  définies par  $\Delta_{a,b}(x) = ax + b$  et considérons les fonctions affines (les droites) qui minorent  $f : \Delta_{a,b} \leq f$ , i.e.  $\Delta_{a,b}(x) \leq f(x)$  pour tout  $x$  réel.

**Lemme B.4.** Soit  $f : \mathbf{R} \rightarrow \mathbf{R}$  une fonction convexe, alors

$$f(x) = \sup \{ \Delta_{a,b}(x), \Delta_{a,b} \leq f \} .$$

Autrement dit,  $f$  est en chaque point le suprémum de l'ensemble des droites, évaluées en ce point, qui minorent  $f$ .

**Preuve.** Considérons maintenant la fonction affine  $\Delta(y) = a(y - x) + f(x)$  avec  $f'_g(x) \leq a \leq f'_d(x)$ . On a  $\Delta(x) = f(x)$ , reste à vérifier que  $\Delta \leq f$  : cela conclura la preuve du lemme. Si  $y > x$ , alors  $\frac{f(y) - f(x)}{y - x} \geq f'_d(x) \geq a$  et  $\Delta(y) \leq f(y)$  ; un raisonnement similaire s'applique dans le cas où  $y < x$ .

Du point de vue de l'optimisation, les fonctions convexes ont une propriété particulièrement intéressante.

**Théorème B.5.** Soit  $f$  une fonction convexe, continue, à valeurs dans  $\mathbf{R}$  et soit  $C$  un ensemble convexe. Si l'une des deux conditions suivantes est vérifiée :

- L'ensemble  $C$  est borné,
- La fonction  $f$  est coercive :

$$\|x\| \rightarrow \infty \implies f(x) \rightarrow +\infty,$$

alors  $f$  admet au moins un minimum global sur  $C$ . Si  $f$  est strictement convexe sur  $C$  alors ce minimum est unique.

# Bibliographie

- [Bil95] P. Billingsley, *Probability and measure*, Wiley Series in Probability and Mathematical Statistics, John Wiley, 1995.
- [ET99] I. Ekeland and R. Témam, *Convex analysis and variational problems*, english ed., Classics in Applied Mathematics, vol. 28, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999, Translated from the French. MR 1727362 (2000j:49001)
- [Kal98] Olav Kallenberg, *Foundations of modern probability*, Springer-Verlag, 1998.
- [Kut98] K. Kuttler, *Modern analysis*, Studies in Advanced Mathematics, CRC Press, 1998 (English).
- [Oxt80] J. C. Oxtoby, *Measure and category*, second ed., Graduate Texts in Mathematics, vol. 2, Springer-Verlag, New York, 1980, A survey of the analogies between topological and measure spaces. MR MR584443 (81j:28003)
- [Roc] R. Tyrrell Rockafellar, *Convex analysis*, Princeton Landmarks in Mathematics.
- [Rud95] W. Rudin, *Principes d'analyse mathématique*, Dunod, 1995, Traduit de l'anglais par G. Auliac.