

# **Cours de Probabilités - MDI 104**

P. Bianchi, T. Bonald, L. Decreusefond

2022–2023



# Table des matières

<b>Introduction</b>	<b>5</b>
<b>Guide de lecture</b>	<b>7</b>
<b>1. Probabilités discrètes</b>	<b>9</b>
1.1. Événements . . . . .	9
1.2. Probabilités sur un espace discret . . . . .	10
1.3. Conditionnement et indépendance . . . . .	14
1.4. Variables aléatoires discrètes . . . . .	18
1.5. Espérance, moments . . . . .	22
1.6. Exercices . . . . .	28
<b>2. Éléments de théorie de la mesure</b>	<b>33</b>
2.1. Tribus, mesures, intégrale : pourquoi ? . . . . .	33
2.2. Tribus . . . . .	34
2.3. Mesures . . . . .	36
2.4. Fonctions mesurables . . . . .	42
2.5. Exercices . . . . .	47
<b>3. Intégration</b>	<b>53</b>
3.1. L'intégrale de Lebesgue . . . . .	53
3.2. Exemples fondamentaux . . . . .	62
3.3. Intégrale par rapport à la mesure de Lebesgue . . . . .	66
3.4. Espaces produit et théorème de Fubini . . . . .	68
3.5. Exercices . . . . .	73
<b>4. Variables aléatoires</b>	<b>83</b>
4.1. Notations, définitions . . . . .	83
4.2. Variables aléatoires réelles . . . . .	85
4.3. Variables aléatoires indépendantes . . . . .	93
4.4. Changement de variables . . . . .	96
4.5. Exercices . . . . .	101
<b>5. Espérance conditionnelle</b>	<b>117</b>
5.1. Loi conditionnelle . . . . .	117
5.2. Espérance conditionnelle . . . . .	122
5.3. Exercices . . . . .	125
<b>6. Fonction caractéristique</b>	<b>131</b>
6.1. Définition et propriétés . . . . .	131
6.2. Caractérisation de la loi . . . . .	133
6.3. Calcul de moments . . . . .	134

6.4. Exercices . . . . .	138
<b>7. Vecteurs gaussiens</b>	<b>141</b>
7.1. Préliminaires . . . . .	141
7.2. Définition, propriétés . . . . .	141
7.3. Caractérisation de l'indépendance . . . . .	143
7.4. Stabilité par transformation affine . . . . .	144
7.5. Densité d'un vecteur gaussien . . . . .	145
7.6. Exercices . . . . .	146
<b>8. Convergences</b>	<b>153</b>
8.1. Loi des grands nombres . . . . .	153
8.2. Théorème central limite . . . . .	154
8.3. Exercices . . . . .	157
<b>A. Annexes</b>	<b>161</b>
A.1. Rappels sur les ensembles . . . . .	161
A.2. Rappels sur les séries . . . . .	163
A.3. Limite supérieure et limite inférieure . . . . .	164
A.4. Fonctions convexes . . . . .	164
A.5. Théorème des classes monotones et conséquences* . . . . .	165
<b>B. Correction des exercices</b>	<b>169</b>

# Introduction

Les premières formalisations des probabilités 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 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.



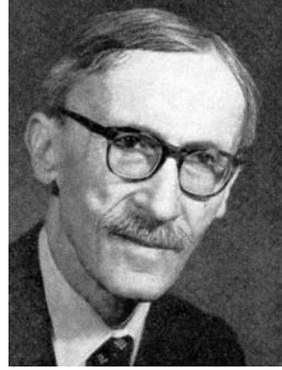
É. Borel (1871-1956),



M. Fréchet (1878-1973)



J. Hadamard (1865-1963),



P. Lévy (1886-1971)



A. Kolmogorov (1903-1987) (DR)

# Guide de lecture

Les paragraphes suivis d'une astérisque \* sont hors des connaissances exigibles à l'examen. Ils peuvent être omis en première lecture.

Le premier chapitre 1 concerne les probabilités sur un espace discret. Il s'agit essentiellement de rappels. Sa lecture est recommandée, mais peut être omise par les élèves qui se sentent à l'aise avec le programme des classes préparatoires.

Les chapitres 2 et 3 introduisent les éléments essentiels de théorie de la mesure et de l'intégration. Ces chapitres relèvent de l'analyse mathématique, ils ne font pas usage de langage probabiliste mais fournissent les outils indispensables à la construction de la théorie des probabilités qui est l'objet des chapitres suivants. Ils constituent également un pré-requis du cours d'analyse MDI 103.

Dans le chapitre 2, il est certes important de comprendre la définition d'une *fonction mesurable* et son utilité en probabilité. Toutefois, un élève qui ressentirait des difficultés ne doit pas se laisser rebuter par les problèmes techniques liés à la mesurabilité. Il saura adapter en conséquence sa lecture du paragraphe 2.4.

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 fournis dans l'annexe.

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. Probabilités discrètes

Ce chapitre est fourni à titre de rappel.

## 1.1. Événements

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*.

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  $\Omega$ .

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 :

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

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

On peut sans trop de difficulté technique développer une théorie des probabilités dans le cas simple où  $\Omega$  est *au plus dénombrable*, c'est à dire fini ou dénombrable. On parle aussi d'espace *discret*. Ce chapitre a pour objet l'étude du cas discret. Le cas général nécessite d'introduire au préalable la théorie de la mesure et de l'intégration. Donnons toutefois quelques exemples d'expériences aléatoires plus générales que celles couvertes par le cas discret :

<i>Expérience aléatoire</i>	<i>Univers</i>	<i>Exemple d'événement</i>
Jet de dé	$\Omega = \{1, 2, 3, 4, 5, 6\}$	$A = \{2, 4, 6\}$ « le résultat est pair »
Deux lancers consécutifs d'une pièce	$\Omega = \{PP, PF, FP, FF\}$ où $P = \text{pile}$ , $F = \text{face}$	$A = \{PP, FF\}$ « on obtient deux faces identiques »

Table 1.1. – Exemples de modélisation dans le cas discret.

<i>Expérience aléatoire</i>	<i>Univers</i>	<i>Exemple d'événement</i>
Durée de fonctionnement sans panne d'une machine	$\Omega = [0, +\infty)$	$A = [x, +\infty)$ « La machine fonctionne pendant au moins $x$ unités de temps »
Valeur d'un signal sur un intervalle de temps	$\Omega = \mathcal{C}_b([t_0, t_1])$ ensemble des fonctions continues sur $[t_0, t_1]$	$A = \{\omega \in \Omega : \sup_{t \in [t_0, t_1]}  \omega(t)  \leq \alpha\}$ « l'amplitude du signal n'excède pas $\alpha$ »

Table 1.2. – Exemples de modélisation non couvertes par le cas discret.

## 1.2. Probabilités sur un espace discret

Jusqu'à la fin de ce chapitre, nous nous restreignons au cas où  $\Omega$  est un ensemble au plus dénombrable.  $\mathcal{P}(\Omega)$  représente l'ensemble des parties de  $\Omega$ .

### 1.2.1. Définition et propriétés

**Définition 1.2.** Une mesure de probabilité  $\mathbb{P}$  sur un ensemble  $\Omega$  au plus dénombrable est une application de  $\mathcal{P}(\Omega)$  dans  $\mathbb{R}$  qui satisfait les deux propriétés suivantes :

- $\mathbb{P}(\emptyset) = 0$ ,  $\mathbb{P}(\Omega) = 1$ ,
- pour toute famille  $(A_j, j \in \mathbb{N}^*)$  de parties deux à deux disjointes de  $\Omega$ ,

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

Si  $(A_n)$  est une suite croissante d'événements (c'est à dire  $A_n \subset A_{n+1}$ ), l'ensemble  $\bigcup_n A_n$  est aussi appelé la limite de  $A_n$  et noté  $\lim_n A_n$ . Si  $(A_n)$  est une suite décroissante d'événements (c'est à dire  $A_{n+1} \subset A_n$ ), l'ensemble  $\bigcap_n A_n$  est aussi appelé la limite de  $A_n$  et noté  $\lim_n A_n$ . La notation  $A_n \uparrow A$  (resp.  $A_n \downarrow A$ ) signifie que  $(A_n)$  est une suite croissante (resp. décroissante) et que  $A = \lim_n A_n$ .

**Proposition 1.1.** Soit  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$  une mesure de probabilité sur  $\Omega$ . Soient  $A, B, (A_n)_{n \in \mathbb{N}^*}$  des parties de  $\Omega$ .

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

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

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

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

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

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

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

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

*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  $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c \cup \emptyset \cup \emptyset \cup \dots) = \mathbb{P}(A) + \mathbb{P}(A^c) + 0 + 0 + \dots$  et finalement  $1 = \mathbb{P}(A) + \mathbb{P}(A^c)$ .

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 :

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

Par ailleurs,  $A$  s'écrit comme l'union disjointe  $(A \setminus B) \cup (A \cap B)$  et donc  $\mathbb{P}(A) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B)$ . De même,  $\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)$ . On a donc :  $\mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B)$ . En faisant la substitution dans (3.12), 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,  $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A)$ .

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  $\sum_n \mathbb{P}(A_n \cap B) = \mathbb{P}(\bigcup_n (A_n \cap B)) = \mathbb{P}((\bigcup_n A_n) \cap B) = \mathbb{P}(B)$ , 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 :

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

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  $\mathbb{P}(A)$ . Le membre de droite se réécrit comme la limite quand  $n \rightarrow \infty$  de la suite  $\sum_{k=1}^n \mathbb{P}(B_k)$ . Mais comme les  $(B_k)$  sont deux à deux disjoints,  $\sum_{k=1}^n \mathbb{P}(B_k) = \mathbb{P}(\bigcup_{k=1}^n B_k) = \mathbb{P}(A_n)$ . On a donc bien montré que  $\mathbb{P}(A) = \lim_n \mathbb{P}(A_n)$ .

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

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

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

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

L'inégalité est vraie au rang  $n = 1$ . Supposons qu'elle soit vraie au rang  $n$ ,  $\mathbb{P}\left(\bigcup_{k=1}^{n+1} A_k\right) = \mathbb{P}(A_{n+1} \cup (\bigcup_{k=1}^n A_k)) \leq \mathbb{P}(A_{n+1}) + \mathbb{P}(\bigcup_{k=1}^n A_k)$  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  $\mathbb{P}(\bigcup_{k=1}^n A_k) \leq \sum_{k=1}^{\infty} \mathbb{P}(A_k)$  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).

□

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

La propriété suivante établit qu'une probabilité  $\mathbb{P}$  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  $\mathbb{P}$  une mesure de probabilité définie sur la tribu des parties  $\mathcal{P}(\Omega)$ . Alors pour tout événement  $A$ ,

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

*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  $\mathbb{P}$ , on a donc  $\mathbb{P}(A) = \mathbb{P}(\{\omega_{i_1}\}) + \mathbb{P}(\{\omega_{i_2}\}) + \dots$ , 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 fautive 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  $\mathbb{P}$  sur  $\mathcal{P}(\Omega)$  telle que pour tout  $\omega \in \Omega$ ,  $\mathbb{P}(\{\omega\}) = p_\omega$ .

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

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.

### 1.2.3. Exemples

#### Cas où $\Omega$ est fini

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

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

où  $|A|$  représente ici le cardinal de l'ensemble  $A$ . Autrement dit,  $\mathbb{P}(A)$  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  $\omega$ ,

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

**Probabilité 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 :

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

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  $1/2$ .

**Probabilité binomiale** Soit  $n \in \mathbb{N}^*$  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 :

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

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  $p$ .

#### Cas où $\Omega$ est dénombrable

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

$$\mathbb{P}(\{k\}) = p(1-p)^{k-1} \tag{1.5}$$

pour tout  $k \in \mathbb{N}^*$ . 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 ??).

REMARQUE.— On peut aussi définir la probabilité géométrique sur  $\mathbb{N}$  (et non  $\mathbb{N}^*$ ) par  $\mathbb{P}(\{k\}) = p(1-p)^k$  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.

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

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

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  $\mathbb{P}$  une probabilité sur  $\Omega$ .

**Définition 1.3.** Pour tous événements  $A, B$  tels que  $\mathbb{P}(B) \neq 0$ , on définit la probabilité conditionnelle de  $A$  sachant  $B$ , et on note  $\mathbb{P}(A | B)$ , la quantité :

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

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  $B$ .

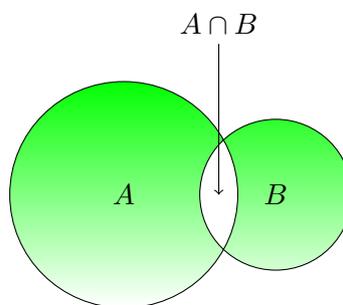


Figure 1.1. – Interprétation graphique du conditionnement.

Considérons le cas où  $\mathbb{P}$  est la probabilité uniforme sur un ensemble  $\Omega$  fini, c'est-à-dire  $\mathbb{P}(A) = |A|/|\Omega|$ . On a alors  $\mathbb{P}(A|B) = |A \cap B|/|B|$ . Cette expression justifie la remarque suivante :  $\mathbb{P}(A|B)$  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 :  $\mathbb{P}$  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  $B$  un événement tel que  $\mathbb{P}(B) \neq 0$ . L'application définie sur  $\mathcal{P}(E)$  par  $A \mapsto \mathbb{P}(A|B)$  est une mesure de probabilité. On la nomme la probabilité conditionnelle à  $B$ .

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

### 1.3.2. Propriétés

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

**Proposition 1.5.** a) Soient  $A, B$  des événements tels que  $0 < \mathbb{P}(B) < 1$ . Alors,

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

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

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

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

EXEMPLE 1.– 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,

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

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

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

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

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

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

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

EXEMPLE 2.– 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 :

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

### 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  $A$ . A 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 1.4.** Deux événements  $A, B$  sont dits *indépendants*, noté  $A \perp\!\!\!\perp B$ , si

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

Si  $\mathbb{P}(B) \neq 0$ , la définition revient bien à  $\mathbb{P}(A|B) = \mathbb{P}(A)$  : 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  $\mathbb{P}(B) = 0$  ou  $\mathbb{P}(B) = 1$ , alors  $A$  et  $B$  sont indépendants quel que soit  $A$ .

**Définition 1.5.** 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 :

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

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

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \quad \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C), \quad \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C),$$

$$\text{et } \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

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.

EXEMPLE 3.— On lance deux dés. On note  $A$  l'événement « le premier lancer est pair »,  $B$  l'événement « le second lancer est pair » et  $C$  l'événement « la somme des deux lancers est paire ». Les événements  $A, B, C$  ne sont pas indépendants car  $\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ . Pourtant,  $A, B$  sont indépendants,  $A, C$  sont indépendants,  $B, C$  sont indépendants.

**Définition 1.6.** Soit  $C$  un événement tel que  $\mathbb{P}(C) \neq 0$ . On dit que  $A$  et  $B$  sont *indépendants conditionnellement à  $C$* , noté  $A \perp\!\!\!\perp B \mid C$  si  $\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid C)$ .

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

REMARQUE.— Attention : des propositions  $A \perp\!\!\!\perp B$  et  $A \perp\!\!\!\perp B \mid C$ , aucune n'implique l'autre. Voir à nouveau l'exemple 3 pour un contre-exemple.

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

Table 1.3. – Quelques exemples de probabilités sur un espace discret.

## 1.4. Variables aléatoires discrètes

### 1.4.1. Rappel : image réciproque

On rappelle la notion d'*image réciproque* d'un ensemble par une application. On se donne deux ensembles arbitraires  $\Omega$  et  $E$  et une application  $X$  de  $\Omega$  dans  $E$  (*nb.* : nous ne supposons pas dans ce paragraphe que  $\Omega$  et  $E$  sont discrets). 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 4.– 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 1.7.** Pour toute famille dénombrable d'ensemble  $(A_n, n \geq 1)$

$$X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} X^{-1}(A_n),$$

$$X^{-1}\left(\bigcap_{n=1}^{\infty} A_n\right) = \bigcap_{n=1}^{\infty} X^{-1}(A_n).$$

### 1.4.2. Définitions

On se donne un univers  $\Omega$  au plus dénombrable, équipé d'une probabilité  $\mathbb{P}$ . Soit  $E$  un autre ensemble au plus dénombrable (par exemple  $E = \mathbb{N}$ , pour fixer les idées).

De manière informelle, une variable aléatoire discrète (v.a.d.) est une grandeur à valeur dans  $E$  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 :

$$X : \Omega \rightarrow E$$

$$\omega \mapsto X(\omega).$$

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  $\omega$ .

En probabilités, on s'intéresse plus particulièrement à évaluer la probabilité d'événements de la forme « la variable  $X$  vaut  $x$  » ou, plus généralement,

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

pour un sous-ensemble  $H \subset E$  arbitraire. Nous utiliserons souvent la notation  $\{X \in H\}$  pour désigner l'événement  $\{\omega \in \Omega : X(\omega) \in H\}$ . La probabilité  $\mathbb{P}(\{X \in H\})$  de l'événement  $\{X \in H\}$  est souvent notée en omettant les accolades, soit  $\mathbb{P}(X \in H)$ . De même, la probabilité  $\mathbb{P}(\{X = x\})$  de l'ensemble  $\{X = x\}$  est souvent notée sans les accolades, par  $\mathbb{P}(X = x)$ .

**Définition 1.7.** Soit  $X$  une v.a.d. On appelle *loi* de  $X$  la fonction définie pour tout  $H \subset E$  par :

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

Autrement dit,  $\mathbf{P}_X(H) = \mathbb{P}(X^{-1}(H))$ . On aurait donc pu utiliser la définition identique mais plus abstraite :

$$\mathbf{P}_X = \mathbb{P} \circ X^{-1},$$

où  $\circ$  représente la composition.

**Proposition 1.8.**  $\mathbf{P}_X$  est une mesure de 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 1.2.2 que  $\mathbf{P}_X$  est entièrement caractérisée par la valeur qu'elle prend sur les singletons. Afin d'alléger les notations, nous écrivons  $\mathbf{P}_X(x)$  au lieu de  $\mathbf{P}_X(\{x\})$ , soit :

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

Ceci se lit, pour tout  $H \subset E$ ,

$$\begin{aligned} \mathbf{P}_X(H) &= \sum_{x \in H} \mathbf{P}_X(x) \quad \text{ou, écrit d'une autre façon :} & (1.6) \\ \mathbb{P}(X \in H) &= \sum_{x \in H} \mathbb{P}(X = x) . \end{aligned}$$

### 1.4.3. Loi jointe, lois marginales, Indépendance

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 1.8.** 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 l'équation (1.6) appliquée au couple  $(X, Y)$ , la loi jointe est entièrement caractérisée par les coefficients

$$\mathbf{P}_{X,Y}(x, y) = \mathbb{P}((X, Y) = (x, y)) = \mathbb{P}(X = x, Y = y)$$

pour tout  $(x, y) \in E \times E$ .

**Proposition 1.9.** 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,  $\mathbf{P}(X = x) = \sum_y \mathbb{P}(X = x, Y = y)$ .  $\square$

**Définition 1.9.**  $X$  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 :

$$\mathbb{P}(X \in G, Y \in H) = \mathbb{P}(X \in G)\mathbb{P}(Y \in H) ,$$

où  $\{X \in G, Y \in H\}$  désigne l'ensemble  $\{X \in G\} \cap \{Y \in H\}$ .

**Proposition 1.10.** Deux v.a.d  $X$  et  $Y$  sont indépendantes si et seulement si pour tout  $(x, y) \in E^2$ ,

$$\mathbf{P}_{X,Y}(x, y) = \mathbf{P}_X(x) \mathbf{P}_Y(y) \tag{1.7}$$

*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 1.11.** Si  $X$  et  $Y$  sont indépendantes, alors  $f(X)$  et  $g(Y)$  sont 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., avec  $X_i : \Omega \rightarrow E_i$  pour tout  $i$ , où  $E_i$  est un espace au plus plus dénombrable. Le  $n$ -uplet  $(X_1, \dots, X_n)$  définit une v.a.d. sur  $E_1 \times \dots \times E_n$ . 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 1.12.** Pour tout  $k = 1, \dots, n$  et tout  $x_k \in E$ ,

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

où la somme s'étend sur l'ensemble des  $(n-1)$ -uplets  $(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$  de  $E_1 \times E_{k-1} \times E_{k+1} \times \cdots \times E_n$ .

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 1.10.** Les v.a.d.  $X_1, \dots, X_n$  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, les v.a.d.  $X_1, \dots, X_n$  sont indépendantes si

$$\mathbb{P}(X_1 \in H_1, \dots, X_n \in H_n) = \mathbb{P}(X_1 \in H_1) \times \cdots \times \mathbb{P}(X_n \in H_n), \quad (1.8)$$

où l'on utilise la notation  $\{X_1 \in H_1, \dots, X_n \in H_n\} = \bigcap_k \{X_k \in H_k\}$ . La définition 1.5 de l'indépendance d'événements requiert que la relation (1.8) soit aussi valide si certains événements  $\{X_k \in H_k\}$  sont supprimés. Il est inutile de l'imposer ici : en prenant  $H_k$  égal à l'espace entier, on a  $\{X_k \in H_k\} = \Omega$ , donc l'événement  $\{X_k \in H_k\}$  est automatiquement supprimé des deux membres de l'équation (1.8).

**Théorème 1.13.** Les v.a.d.  $X_1, \dots, X_n$  sont indépendantes si et seulement si  $\forall (x_1, \dots, x_n) \in E_1 \times \cdots \times E_n$ ,

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

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

**Définition 1.11.** Une famille arbitraire de variables aléatoires est dite indépendante si toute sous-famille *finie* est indépendante.

Une famille de v.a.d. est dite identiquement distribuée si toutes les v.a.d. de la famille ont la même loi. On utilise souvent l'abréviation *i.i.d.* pour désigner une famille indépendante et identiquement distribuée de variables aléatoires.

**Proposition 1.14.** 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.

## 1.5. Espérance, moments

### 1.5.1. Définition

**Définition 1.12.** Soit  $E$  une partie au plus dénombrable de  $\mathbb{R}$ . On définit l'espérance  $\mathbb{E}(X)$  d'une v.a.d.  $X : \Omega \rightarrow E$  par

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x \in E} x \mathbb{P}(X = x) \\ &= \sum_{x \in E} x \mathbf{P}_X(x).\end{aligned}\tag{1.9}$$

**Conditions d'existence.** Lorsque  $E$  est de cardinal infini, l'espérance  $\mathbb{E}(X)$  peut ne pas être définie. Néanmoins la somme a toujours un sens dès que l'une des deux conditions suivantes est 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  $+\infty$ ) ; Lorsque cette condition est vraie *i.e.*,  $\mathbf{P}_X(x) = 0$  pour tout  $x < 0$ , nous dirons que la v.a.d.  $X$  est *positive  $\mathbb{P}$ -presque partout* et nous noterons  $X \geq 0$   $\mathbb{P}$ -p.p.
2. ses termes sont absolument sommables, c'est à dire :

$$\sum_{x \in E} |x| \mathbb{P}(X = x) < \infty.\tag{1.10}$$

En effet dans ce cas, la somme  $\sum_{x \in E} x \mathbb{P}(X = x)$  est bien définie, et ne dépend pas de l'ordre dans lequel on somme les termes.

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

**EXEMPLE 6.**— 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

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

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  $\mathbb{P}(X = a)$  et  $\mathbb{P}(X = b)$  respectivement.

**EXEMPLE 7.**— Soit  $X(\omega) = \mathbf{1}_A(\omega)$  la variable aléatoire sur  $E = \{0, 1\}$  égale à l'indicatrice d'un certain événement  $A \subset \Omega$ . On a  $\mathbb{E}(\mathbf{1}_A) = 0 \times \mathbb{P}(\mathbf{1}_A = 0) + 1 \times \mathbb{P}(\mathbf{1}_A = 1)$ , et comme  $\mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(A)$ , on obtient la formule :

$$\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A).$$

On voit donc que *l'espérance généralise la notion de probabilité*.

## 1.5.2. Propriétés

Soient  $E, F$  deux espaces au plus dénombrables, 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. La proposition suivante porte parfois le nom de « Théorème de transfert ».

**Proposition 1.15.** Si  $f$  est positive,

$$\mathbb{E}(f(X)) = \sum_{x \in E} f(x) \mathbb{P}(X = x) . \quad (1.11)$$

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

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

$$\begin{aligned} \mathbb{E}(f(X)) &= \sum_{y \in F} y \mathbb{P}(f(X) = y) = \sum_{y \in F} y \mathbb{P}(X \in f^{-1}(\{y\})) \\ &= \sum_{y \in F} \sum_{x \in f^{-1}(\{y\})} y \mathbb{P}(X = x) \\ &= \sum_{y \in F} \sum_{x \in f^{-1}(\{y\})} f(x) \mathbb{P}(X = x) \\ &= \sum_{x \in E} f(x) \mathbb{P}(X = x) , \end{aligned} \quad (1.12)$$

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 :

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

et en appliquant le résultat à la fonction  $|f|$ ,  $\mathbb{E}(|f(X)|) = \sum_{x \in E} |f(x)| \mathbb{P}(X = x)$  qui est fini par hypothèse. L'équivalent pour la v.a.  $f(X)$  de la condition (1.10) est satisfaite :  $\mathbb{E}(f(X))$  est bien définie. La preuve de (1.11) est obtenue par le même calcul qu'en (1.12).  $\square$

REMARQUE.— Si l'on devait évaluer  $\mathbb{E}(f(X))$  en utilisant la définition (1.9) de l'espérance, on devrait au préalable calculer la loi  $\mathbf{P}_{f(X)}$  de la v.a.d.  $f(X)$ . L'équation (1.11) montre que l'espérance  $\mathbb{E}(f(X))$  s'exprime directement en fonction de la loi de la variable  $X$ .

REMARQUE.— La propriété (1.11) permet d'écrire la condition de sommabilité (1.10) 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 «  $\mathbb{P}$ -p.p. » pour « presque partout », c'est-à-dire avec probabilité 1. Par exemple, on écrira que «  $X = a$   $\mathbb{P}$ -p.p. » pour signifier que  $\mathbf{P}(X = a) = 1$ .

**Proposition 1.16.** 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  $X \geq 0$   $\mathbb{P}$ -p.p., alors  $\mathbb{E}(X) \geq 0$ .
- c) Si  $X \geq 0$   $\mathbb{P}$ -p.p. et si  $\mathbb{E}(X) = 0$ , alors  $X = 0$   $\mathbb{P}$ -p.p.
- d)  $|\mathbb{E}(X)| \leq \mathbb{E}|X|$ .
- e) Si  $X \leq Y$   $\mathbb{P}$ -p.p., alors  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .
- f) Si  $X = a$   $\mathbb{P}$ -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.  $\square$

### 1.5.3. Inégalités

**Proposition 1.17.** (Inégalité de Markov) Pour tout  $\epsilon > 0$ ,  $p \geq 1$ ,

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

*Démonstration.* On donne d'abord la preuve pour  $\epsilon = p = 1$  et  $X \geq 0$ . On a  $\mathbb{P}(X > 1) = \mathbb{E}(\mathbf{1}_{]1,+\infty[}(X)) \leq \mathbb{E}(X)$  car  $\mathbf{1}_{]1,+\infty[}(X) \leq X$ . Dans le cas général, on utilise le fait que  $\mathbb{P}(|X| > \epsilon) = \mathbb{P}(|X|^p / \epsilon^p > 1)$  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 1.18.** (Inégalité de Cauchy-Schwarz)

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

*Démonstration.* Si  $\mathbb{E}(X^2) = 0$ , la v.a.  $X^2$  est nulle  $\mathbb{P}$ -p.p. donc  $XY = 0$   $\mathbb{P}$ -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  $\mathbb{E}(X^2) \neq 0$  et  $\mathbb{E}(Y^2) \neq 0$ .

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

#### 1.5.4. Moments, variance, covariance

**Définition 1.13.** 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 1.19.** 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 1.14.** La *variance* d'une v.a.d.  $X$  d'ordre 2 est définie par

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

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

EXEMPLE 8.— Un joueur lance une pièce, gagne un euro si le résultat est pile, perd un euro sinon. L'espérance du gain  $X$  est  $\mathbb{E}(X) = 0$ . 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 9.— La variance d'une loi de Bernoulli  $B(p)$  vaut  $p(1-p)$ .

**Définition 1.15.** Soient  $X$  et  $Y$  deux v.a.d. d'ordre 2. Leur *covariance* est définie par :

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

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

- a)  $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$  ;
- 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.

### 1.5.5. Cas des variables indépendantes

**Proposition 1.21.** Soient  $X$  et  $Y$  des v.a. indépendantes telles que  $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$ . Alors  $\mathbb{E}|XY| < \infty$ , et on a l'égalité :

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

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

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,

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X)) \mathbb{E}(g(Y)), \quad (1.13)$$

dès lors que les deux sommes du membre de droite sont absolument convergentes. Nous noterons au passage que ce résultat admet une réciproque : si  $X$  et  $Y$  vérifient l'équation (1.13) pour toutes fonctions  $f, g$  bornées, elles sont indépendantes. Pouvez vous en donner la preuve? (Indication : choisir pour  $f$  et  $g$  des fonctions indicatrices).

Un cas particulier intéressant est obtenu en posant  $f(x) = x - \mathbb{E}(X)$  et  $g(y) = y - \mathbb{E}(Y)$ . Dans ce cas, le membre de gauche de (1.13) 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 1.22.** 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 :

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) . \quad (1.14)$$

Notons que deux variables décorréélées ne sont pas nécessairement indépendantes. L'exercice 1.5 permet de s'en convaincre.

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

**Proposition 1.23.** 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  $E'_k$ , où  $E'_k \subset \mathbb{R}$  telle que  $\mathbb{E}|f_k(X_k)| < \infty$ . On suppose  $X_1, \dots, X_n$  indépendantes. Alors,

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

**Proposition 1.24.** 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 (1.14),  $\text{Var}(X_n + Z_n) = \text{Var}(X_n) + \text{Var}(Z_n)$  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.  $\square$

## 1.6. Exercices

### S'exercer au dénombrement

**EXERCICE 1.1.**–

**(Un peu de dénombrement).**

1. 25 chevaux participent au tiercé. Combien y a-t-il de tiercés possibles? (rq : un tiercé est une suite ordonnée)
2. Combien de mots de 5 lettres (pas nécessairement prononçables) peut-on faire avec les 26 lettres de l'alphabet français?
3. Combien y a-t-il de tirages de 5 cartes dans un jeu de 52 cartes?
4. De combien de façon peut-on répartir  $r$  boules numérotées de 1 à  $r$ , dans  $n$  urnes?
5. On tire quatre cartes au hasard, dans un jeu de 52 cartes. Quelle est la probabilité d'avoir exactement deux rois?

**EXERCICE 1.2** ((Encore un peu de dénombrement).)–

$n$  étudiants sont dans une salle, chacun d'eux ayant un jour anniversaire numéroté de 1 à 365 (on ne tient pas compte des années bissextiles).

1. Le professeur parie que deux étudiants au moins sont nés le même jour. On s'intéresse au nombre d'étudiants  $n$  minimal qu'il doit y avoir dans la salle pour que la probabilité que le professeur gagne soit supérieure à 0.5. Pour répondre à cette question, on pourra
  - a) Proposer un univers  $\Omega$  et une probabilité  $\mathbb{P}$  qui décrive cette expérience. On choisira  $\Omega$  de sorte qu'il soit raisonnable de prendre pour  $\mathbb{P}$  la probabilité uniforme sur  $\Omega$ .
  - b) calculer la probabilité de "2 étudiants au moins sont nés le même jour".
  - c) Tracer cette quantité en fonction de  $n$  et conclure.
2. Quelle est la probabilité des évènements suivants :
  - a) Aucun étudiant n'est né le 15 novembre
  - b) Au moins un étudiant est né le 15 novembre
  - c) Un étudiant exactement est né le 15 novembre
  - d) Deux étudiants au moins sont nés le 15 novembre.

### Pour apprendre

**EXERCICE 1.3** (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?

**EXERCICE 1.4.**–

Une tortue donne naissance à un nombre  $N$  de bébés : on note  $M$  le nombre de mâles et  $F$  le

nombre de femelles. On suppose que  $N$  est une variable aléatoire suivant une loi de Poisson de paramètre  $\lambda > 0$ , c'est à dire :

$$\forall k \in \mathbb{N}, \quad \mathbb{P}[N = k] = \frac{\lambda^k}{k!} e^{-\lambda}$$

et on suppose que chaque bébé a (indépendamment des autres) une chance sur deux d'être une femelle.

1. Calculer la loi du couple  $(N, F)$ .  
*Indication* : On pourra écrire  $F = Y_1 + Y_2 + \dots + Y_N$  où  $(Y_n)_{n \in \mathbb{N}^*}$  est une suite de variables aléatoires i.i.d., indépendantes de  $N$ , et dont on précisera la loi.
2. Calculer la loi de  $F$ , puis celle de  $M$ .
3. Trouver une relation simple entre  $M$ ,  $F$  et  $N$ . Calculer la loi du couple  $(M, F)$ .
4. Les variables  $M$  et  $F$  sont-elles indépendantes ?

#### EXERCICE 1.5. –

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 1.6. –

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

- 1.6.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 ?
- 1.6.b) Calculer  $\mathbb{P}(X_i = 1)$  et  $\mathbb{P}(X_i X_j = 1)$  pour  $i \neq j$ .
- 1.6.c) En déduire  $\mathbb{E}(X)$  et  $\text{Var}(X)$ .

### Pour s'entraîner

#### EXERCICE 1.7. –

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  $\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ ).

#### EXERCICE 1.8. –

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

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

1.8.a) Calculer  $\mathbb{P}(X = k)$ .

1.8.b) Calculer la loi de  $M$  sachant  $X = k$ , dite loi *a posteriori* de  $M$ .

1.8.c) Pour  $k = 0$ , identifier cette loi.

### EXERCICE 1.9. –

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  $0 = (0, \dots, 0)$ .

1.9.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  $1/2$ .

1.9.b) Quelle est la longueur moyenne d'un mot ?

1.9.c) Quelle est la variance de la longueur d'un mot ?

1.9.d) On choisit deux mots au hasard indépendamment l'un de l'autre, soit  $X$  et  $Y$  les variables aléatoires correspondantes. Calculer

$$\mathbb{E}(d(X, Y)^2).$$

### Pour aller plus loin

#### EXERCICE 1.10 (\*\*, 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  $u_n = \mathbb{P}(Z_n = 0)$ . 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  $\mathbb{E}(X_0) < 1$ .

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 1.11** (\*\*\*, Erdős et Renyi (1960)).—

On fabrique un graphe sur  $n$  sommets en choisissant ses arêtes « 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.

- 1.11.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 :

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

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

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

*Suggestion.* On pourra montrer que

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

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.

- 1.11.b) Que valent  $\mathbb{E}(X_i)$  et  $\mathbb{E}(X)$  ?  
 1.11.c) On suppose dorénavant  $c$  fixé. Montrer que la quantité  $F^{(k)}$ , pour la variable  $X$ , converge, lorsque  $n$  tend vers l'infini, vers  $e^{-ck}/k!$ .  
 1.11.d) Montrer que  $\lim_{n \rightarrow \infty} \mathbb{P}(X = 0) = e^{-e^{-c}}$ .  
 1.11.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.  
 1.11.f) Plus généralement, soit  $C_t$  le nombre de composantes connexes à  $t$  sommets. Montrer que pour  $2 \leq t \leq n/2$ ,

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

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

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

**EXERCICE 1.12** (Borne de Chernoff).—

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

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

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

$$\mathbb{P}(X \geq \eta) \leq e^{-\eta\theta} \mathbb{E}(\exp(\theta X)). \quad (1.15)$$

1.12.c) Calculer  $\mathbb{E}(\exp(\theta X))$ .

1.12.d) Trouver  $\theta$  qui minimise le terme de droite de (1.15).

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

**EXERCICE 1.13.**–

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.

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

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

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

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

1.13.d) Calculer son gain moyen.

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

$$\mathbb{E}(X_{n+1} | X_n = i) = \frac{5}{6}i + \frac{35}{6},$$

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

1.13.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.

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

**EXERCICE 1.14.**–

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

## 2. Éléments de théorie de la mesure

### 2.1. Tribus, mesures, intégrale : pourquoi ?

La théorie de la mesure fournit le cadre théorique indispensable à la construction rigoureuse d'une théorie des probabilités qui va au delà des univers discrets. Le terme de *mesure* doit être compris comme une formalisation des notions de « volume », « aire », ou « longueur » : une mesure est une application qui, à un ensemble, associe un nombre. Si cet ensemble s'écrit comme une union dénombrable de sous-ensembles disjoints, sa mesure doit être égale à la somme des sous-ensembles qui le composent. C'est la condition de  $\sigma$ -additivité, qui s'impose naturellement si on veut définir une notion de volume, d'aire ou de longueur d'un ensemble.

Par exemple, la mesure de Lebesgue sur  $\mathbb{R}$ , notée  $\lambda$ , satisfait  $\lambda([a, b]) = b - a$  pour tout segment  $[a, b]$ . C'est la façon la plus naturelle de mesurer un segment. Et par  $\sigma$ -additivité, tout ensemble s'écrivant comme union disjointe de segments, a pour mesure la somme des longueurs des segments qui le composent. Un tel objet  $\lambda$  semble naturel, pourtant sa construction est délicate. A moins de renoncer au célèbre « axiome du choix », *il n'existe pas d'application  $\sigma$ -additive  $\lambda$ , définie sur  $\mathcal{P}(\mathbb{R})$ , qui satisfait  $\lambda([a, b]) = b - a$  pour tout segment  $[a, b]$ . Autrement dit, on ne peut pas définir une application « longueur » qui s'appliquerait à *tout* sous-ensemble de  $\mathbb{R}$ . On est forcé de restreindre le domaine de définition de  $\lambda$  à une certaine classe  $\mathcal{B}(\mathbb{R})$  de parties de  $\mathbb{R}$ , strictement plus petite que  $\mathcal{P}(\mathbb{R})$  :*

$$\lambda : \mathcal{B}(\mathbb{R}) \rightarrow [0, +\infty].$$

L'ensemble  $\mathcal{B}(\mathbb{R})$  est la *tribu de Borel*. Grâce à cette restriction du domaine de définition, on peut enfin définir une application  $\lambda$ ,  $\sigma$ -additive, et qui à chaque segment associe sa longueur.

Plus généralement, une mesure sur un espace arbitraire  $\Omega$  est une application  $\mu : \mathcal{F} \rightarrow [0, +\infty]$ ,  $\sigma$ -additive, définie sur un ensemble  $\mathcal{F}$  de parties de  $\Omega$ . Comme on l'a vu plus haut, on ne peut pas systématiquement considérer que  $\mathcal{F} = \mathcal{P}(\Omega)$  sous peine de ne pas pouvoir définir des mesures pourtant naturelles, comme la mesure de Lebesgue. L'ensemble de définition  $\mathcal{F}$  doit tout de même avoir certaines propriétés (en particulier, toute union dénombrable d'éléments de  $\mathcal{F}$  doit être dans  $\mathcal{F}$ , ne serait-ce que pour pouvoir énoncer la condition de  $\sigma$ -additivité de  $\mu$ ). On dit que  $\mathcal{F}$  est une *tribu*.

La théorie de la mesure s'inscrit dans le cadre de l'analyse : elle n'est pas historiquement motivée par le besoin de formuler une théorie des probabilités. C'est *a posteriori* que des mathématiciens comme A. Kolmogorov ont remarqué l'utilité de la théorie de la mesure pour fonder celle des probabilités. Un lien évident entre probabilité et mesure est la  $\sigma$ -additivité : pour définir une probabilité  $\mathbb{P}$  en tant qu'application d'ensemble, il est naturel d'imposer que la probabilité d'une union disjointe d'événements est la somme des probabilités. Une probabilité est donc une mesure, au sens de l'analyse. Après avoir énoncé les principaux résultats de théorie de la mesure, nous

serons ainsi capables de construire des mesures de probabilités sur  $\mathbb{R}$ ,  $\mathbb{R}^n$  ou même des espaces plus complexes. Nous serons capables de définir des variables aléatoires, en nous appuyant sur la notion de *fonctions mesurables*. Nous serons capables enfin de définir une notion d'espérance, en nous appuyant sur *l'intégrale de Lebesgue par rapport à une mesure*.

## 2.2. Tribus

### 2.2.1. Définition et propriétés générales

Soit  $\Omega$  un ensemble arbitraire.

**Définition 2.1** (Tribu). Une famille  $\mathcal{F}$  de sous-ensembles de  $\Omega$  est appelée une *tribu* sur  $\Omega$  si elle vérifie les propriétés suivantes :

- i)  $\Omega \in \mathcal{F}$  ;
- ii)  $\forall A \in \mathcal{F}, A^c \in \mathcal{F}$  ;
- 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 deux exemples évidents 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 2.1.** 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 iii). En invoquant à nouveau l'axiome ii), le complémentaire de cet ensemble est également dans  $\mathcal{F}$  par l'axiome 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 iii) et de b) respectivement.

□

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

**Lemme 2.2.** L'intersection de deux tribus est une tribu.

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

**Théorème 2.3** (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-vidé 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 10.— 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  $\sigma(\{A\}) = \{\emptyset, \Omega, A, A^c\}$ .

### 2.2.2. La tribu de Borel sur $\mathbb{R}^d$

L'exemple qui suit est fondamental. Il va nous permettre de préciser enfin ce que sont les parties de  $\mathbb{R}$  (ou de  $\mathbb{R}^d$ ) susceptibles d'être mesurée.

**Définition 2.2** (Tribu borélienne). La *tribu de Borel sur  $\mathbb{R}^d$* , notée  $\mathcal{B}(\mathbb{R}^d)$ , est la tribu engendrée par les « pavés » :

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

En particulier sur  $\mathbb{R}$ ,  $\mathcal{B}(\mathbb{R})$  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*.

On rappelle qu'un sous-ensemble  $U \subset \mathbb{R}$  est *ouvert* si pour tout  $x \in U$ , il existe  $(a, b) \in \mathbb{Q}^2$ ,  $a < b$ , tel que  $x \in [a, b]$  et  $[a, b] \subset U$ . La même définition s'applique pour les ouverts de  $\mathbb{R}^d$ , en remplaçant le segment  $[a, b]$  par un pavé à bords rationnels. Un ensemble est dit *fermé* s'il est le complémentaire d'un ouvert.

**Théorème 2.4.**  $\mathcal{B}(\mathbb{R}^d)$  est égale à la tribu engendrée par les ouverts de  $\mathbb{R}^d$ .

*Démonstration.* Donnons la preuve pour  $d = 1$  (le cas  $d > 1$  se traite de façon identique). Soit  $U$  un ouvert de  $\mathbb{R}$ . Pour tout  $x \in U$ , il existe un segment  $[a_x, b_x]$  à extrémités rationnelles tel que  $x \in [a_x, b_x]$  et  $[a_x, b_x] \subset U$ . Donc

$$\bigcup_{x \in U} [a_x, b_x] = U.$$

Or l'ensemble des intervalles à extrémités rationnelles est dénombrable. L'union dans l'égalité ci-dessus est donc une union dénombrable de boréliens. Donc  $U \in \mathcal{B}(\mathbb{R})$ . Par conséquent, la tribu engendrée par les ouverts est incluse dans  $\mathcal{B}(\mathbb{R})$ . L'autre inclusion est évidente.  $\square$

En particulier, tout ouvert de  $\mathbb{R}^d$  est un borélien, et par passage au complémentaire, tout fermé est un borélien. Les ensembles suivants sont dans  $\mathcal{B}(\mathbb{R})$  : le singleton  $\{a\}$ , les intervalles de la forme  $]a, b]$ ,  $]a, b[$ ,  $[a, b[$ ,  $] - \infty, b]$ ,  $[a, +\infty[$ , l'ensemble des rationnels, l'ensemble des irrationnels. En conclusion, la tribu de Borel est très riche, au point que l'on peut se demander s'il existe des parties de  $\mathbb{R}$  qui sont pas boréliennes. De tels ensembles existent en effet, nous en donnerons un exemple au paragraphe 2.3.5.

REMARQUE.— Pour les élèves familiers des notions de topologie, le théorème 2.4 suggère une généralisation à la dimension infinie. Sur n'importe quel espace muni d'une topologie (c'est à dire d'une collection d'ouverts), on peut définir la tribu de Borel comme la tribu engendrée par les ouverts, sans qu'il soit besoin de parler d'intervalles ou de pavés.

**Proposition 2.5.**  $\mathcal{B}(\mathbb{R})$  est la tribu engendrée par les intervalles de la forme  $] - \infty, b]$ . Plus généralement,  $\mathcal{B}(\mathbb{R}^d)$  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  $\mathcal{B}(\mathbb{R})$ , donc  $\sigma(\mathcal{C}) \subset \mathcal{B}(\mathbb{R})$  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  $\mathcal{B}(\mathbb{R}) \subset \sigma(\mathcal{C})$ .  $\square$

Dans ce cours, nous définirons la tribu de Borel  $\mathcal{B}(\overline{\mathbb{R}})$  sur  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$  comme l'ensemble des éléments de la forme  $H$ ,  $H \cup \{+\infty\}$ ,  $H \cup \{-\infty\}$ ,  $H \cup \{-\infty, +\infty\}$  où  $H$  décrit  $\mathcal{B}(\mathbb{R})$ . De manière équivalente,  $\mathcal{B}(\overline{\mathbb{R}})$  est la tribu engendrée par les intervalles de la forme  $[-\infty, b]$  pour  $b$  décrivant  $\overline{\mathbb{R}}$ .

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

$$\mathcal{B}(E) = \{H \cap E : H \in \mathcal{B}(\mathbb{R}^d)\}.$$

## 2.3. Mesures

### 2.3.1. Définition et propriétés

Soit  $(\Omega, \mathcal{F})$  un espace mesurable.

**Définition 2.3.** Une mesure sur  $(\Omega, \mathcal{F})$  est une fonction d'ensemble  $\mu$  telle que

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

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

2.3.iii) Pour toute famille  $(A_n, n \in \mathbb{N})$  d'éléments de  $\mathcal{F}$  deux à deux disjoints,

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

On prêtera attention à l'intervalle fermé à droite dans l'axiome 2.3.i) : la mesure d'un ensemble  $A$  est une quantité positive, possiblement infinie.

Lorsque  $\mu(\Omega) < \infty$ , on dit que la mesure est *finie*. Si en outre  $\mu(\Omega) = 1$ , la mesure  $\mu$  est une *mesure de probabilité*. Un triplet  $(\Omega, \mathcal{F}, \mu)$  où  $\mu$  est une mesure sur  $(\Omega, \mathcal{F})$ , est appelé un *espace mesuré*. C'est un espace de probabilité lorsque  $\mu$  est une mesure de probabilité.

L'axiome 2.3.iii) est connu sous le nom de propriété de  $\sigma$ -additivité. Une conséquence immédiate de cet axiome est le résultat suivant. Lorsque  $\mu$  est une mesure de probabilité, ce résultat est connu sous le nom de *formule des probabilités totales*.

**Théorème 2.6.** Soit  $(B_n, n \geq 1)$  des éléments de  $\mathcal{F}$  formant une partition de  $\Omega$ . Pour tout  $A \in \mathcal{F}$ ,

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

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

Les propriétés suivantes sont d'un usage constant.

**Théorème 2.7.** Soient  $A, B, (A_n, n \in \mathbb{N}^*)$  des éléments de  $\mathcal{F}$ .

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

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

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

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

2.7.e) Pour une famille quelconque  $(A_n, n \in \mathbb{N}^*)$  dans  $\mathcal{F}$ , 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.* La preuve est rigoureusement identique à celle de la Proposition 1.1, il suffit de changer  $\mathbb{P}$  en  $\mu$  (on notera seulement, en ce qui concerne l'item 2.7.d), que l'hypothèse  $\mu(A_1) < \infty$

permet d'éviter l'éventuel problème d'une soustraction  $+\infty - (+\infty)$ ). Nous renvoyons le lecteur à la Proposition 1.1.  $\square$

### 2.3.2. Premier exemple : mesures discrètes

EXEMPLE 11 (Mesure de Dirac).— Soit  $(\Omega, \mathcal{F})$  un espace mesurable quelconque et soit  $a \in \Omega$ . La mesure de Dirac au point  $a$  est l'application  $\delta_a$  définie sur  $\mathcal{F}$  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 12 (Mesure de comptage).— Soit  $I$  un ensemble au plus dénombrable et  $(a_i)_{i \in I}$  une collection de points de  $\Omega$ . La fonction définie pour tout  $A \in \mathcal{F}$  par :

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

définit une mesure sur  $\mathcal{F}$ , 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  $A$  :  $\mu(A) = \text{cardinal}\{i : a_i \in A\}$ .

EXEMPLE 13 (Mesure discrète).— Plus généralement, si  $(\alpha_i)_{i \in I}$  est une suite de coefficients réels positifs, alors :

$$\mu(A) := \sum_{i \in I} \alpha_i \delta_{a_i}(A) \tag{2.1}$$

définit une mesure sur  $\mathcal{F}$ . 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*.

Lorsqu'une mesure  $\mu$  est telle que  $\mu(\{a\}) > 0$  pour un certain singleton  $a \in \Omega$ , on dit que  $\mu$  admet une masse  $\mu(\{a\})$  au point  $a$ . On voit que c'est le cas pour les mesures discrètes : la mesure (2.1) a une masse  $\alpha_i$  au point  $a_i$ .

### 2.3.3. La mesure de Lebesgue sur $\mathbb{R}^d$

**Théorème 2.8.** Il existe une unique mesure  $\lambda_d$  sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  telle que pour réels  $a_1 < b_1, \dots, a_d < b_d$ ,

$$\lambda_d([a_1, b_1] \times \dots \times [a_d, b_d]) = (b_1 - a_1) \times \dots \times (b_d - a_d).$$

L'application  $\lambda_d$  est appelée la *mesure de Lebesgue* sur  $\mathbb{R}^d$ .

Le résultat d'existence est admis. Il est non-trivial, et sa genèse repose sur une longue lignée de travaux (C. Jordan, E. Borel, H. Lebesgue, C. Carathéodory). L'unicité est une conséquence du théorème général 2.9 que nous verrons plus bas : le fait de connaître la valeur d'une mesure sur les pavés détermine la mesure partout.

Lorsque  $d = 1$ , ou lorsqu'il n'y a pas d'ambiguïté sur la dimension de l'espace, on pourra simplement noter  $\lambda_d = \lambda$ . Si  $E$  est un borélien de  $\mathbb{R}^d$ , la mesure de Lebesgue sur  $E$ , notée  $\lambda_E$  est simplement définie par

$$\lambda_E(H) = \lambda_d(H \cap E).$$

Ainsi, on pourra par exemple parler de la mesure de Lebesgue sur l'intervalle  $[0, 1]$ .

Les mesures de Lebesgue sur  $\mathbb{R}$ ,  $\mathbb{R}^2$ ,  $\mathbb{R}^3$  correspondent respectivement aux applications « longueur », « aire », « volume ».

Notons que la mesure de Lebesgue n'admet de masse en aucun point. En effet, si  $a \in \mathbb{R}$  est un point quelconque,  $\lambda(\{a\}) = \lim_n \lambda([a - \frac{1}{n}, a + \frac{1}{n}]) = \lim_n \frac{2}{n} = 0$ .

### 2.3.4. Caractérisation des mesures, fonction de répartition

Une mesure discrète est facile à identifier (il suffit de connaître ses masses en chaque point). Il n'en va pas de même pour des mesures plus générales. Pour montrer que deux mesures  $\mu$  et  $\nu$  sur  $(\Omega, \mathcal{F})$  sont égales, il faut a priori montrer que  $\mu(A) = \nu(A)$  pour tout  $A \in \mathcal{F}$ , ce qui est souvent délicat. Le résultat principal de ce paragraphe montre que l'on peut généralement se simplifier la tâche, en vérifiant cette égalité seulement sur un sous-ensemble de  $\mathcal{F}$ . Autrement dit, une mesure est entièrement déterminée par sa valeur sur un certain ensemble d'ensembles suffisamment riche.

Nous commençons par énoncer le résultat principal de façon abstraite, puis nous fournissons des exemples d'application.

**Définition 2.4.** 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 14.— Donnons deux exemples de  $\pi$ -systèmes intéressants sur  $\mathbb{R}^d$

$$\begin{aligned} \mathcal{P}_1 &= \{[a_1, b_1] \times \cdots \times [a_d, b_d], (a_1, \dots, a_d) \in \mathbb{R}^d, (b_1, \dots, b_d) \in \mathbb{R}^d\} \\ \mathcal{P}_2 &= \{] - \infty, a_1] \times \cdots \times ] - \infty, a_d], (a_1, \dots, a_d) \in \mathbb{R}^d\}. \end{aligned}$$

Vérifier à titre d'exercice que ce sont bien des  $\pi$ -systèmes. Ici, nous utilisons la convention que  $[x, y] = \emptyset$  lorsque  $x > y$ .

**Définition 2.5.** Une mesure  $\mu$  sur un espace mesurable  $(\Omega, \mathcal{F})$  est dite  $\sigma$ -finie s'il existe une collection de  $(A_n)_{n \in \mathbb{N}}$  d'éléments de  $\mathcal{F}$ , telle que  $\Omega = \bigcup_n A_n$  et  $\mu(A_n) < \infty$  pour tout  $n$ .

Evidemment toute mesure finie est  $\sigma$ -finie, à commencer par les mesures de probabilité. La mesure de Lebesgue n'est pas finie, mais est par contre  $\sigma$ -finie (par exemple dans  $\mathbb{R}$ , on a bien  $\mathbb{R} = \bigcup_n [-n, n]$  et  $\lambda([-n, n]) < \infty$ ).

Si en outre  $\mathcal{A} \subset \mathcal{F}$  est une certaine collection d'ensembles, et que les  $A_n$  de la définition 2.5 peuvent être choisis au sein de  $\mathcal{A}$ , on dira que  $\mu$  est  $\sigma$ -finie sur  $\mathcal{A}$ .

**Théorème 2.9.** Soit  $\mu, \nu$  deux mesures sur  $(\Omega, \mathcal{F})$ . Soit  $\mathcal{P}$  un  $\pi$ -système engendrant  $\mathcal{F}$  i.e.,  $\mathcal{F} = \sigma(\mathcal{P})$ . Supposons que l'une des deux conditions suivantes est satisfaite :

1.  $\mu, \nu$  sont deux mesures de probabilité ;
2.  $\mu, \nu$  sont  $\sigma$ -finies sur  $\mathcal{P}$ .

Si  $\mu$  et  $\nu$  coïncident sur  $\mathcal{P}$ , alors  $\mu = \nu$ .

La preuve est fournie dans l'annexe A.5.2, mais on pourra admettre ce théorème en première lecture.

**Conséquence 1 : unicité de la mesure de Lebesgue.** L'ensemble  $\mathcal{P}_1$  est un  $\pi$ -système qui, par définition, engendre la tribu de borel. Deux mesures  $\sigma$ -finies sur  $\mathcal{P}_1$  qui coïncident sur  $\mathcal{P}_1$  sont forcément égales. Ainsi, il ne peut y avoir qu'une unique mesure qui à tout pavé non-vide  $[a_1, b_1] \times \dots \times [a_d, b_d]$  associe  $\prod_i (b_i - a_i)$ .

**Conséquence 2 : la mesure de Lebesgue est invariante par translation.**

**Corollaire 2.10.** Pour tout  $x \in \mathbb{R}^d$ , pour tout borélien  $A$ ,

$$\lambda(A + x) = \lambda(A),$$

où  $A + x := \{y + x, y \in A\}$ .

*Démonstration.* Pour  $x \in \mathbb{R}^d$  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 2.9, elles sont égales d'où le résultat.  $\square$

**Conséquence 3 : fonction de répartition d'une mesure de probabilité.**

**Définition 2.6.** On appelle *fonction de répartition* d'une mesure de probabilité  $\mu$  sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  la fonction  $F_\mu : \mathbb{R}^d \rightarrow [0, 1]$  définie par

$$F_\mu(x_1, \dots, x_d) = \mu(]-\infty, x_1] \times \dots \times ]-\infty, x_d]).$$

**Corollaire 2.11.** Deux mesures de probabilité sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  ayant la même fonction de répartition sont égales.

*Démonstration.* L'ensemble  $\mathcal{P}_2$  est un  $\pi$ -système qui engendre la tribu de borel d'après la proposition 2.5. Dire que deux mesures ont même fonction de répartition revient à dire qu'elles coïncident sur  $\mathcal{P}_2$ . Elles sont donc égales d'après le théorème 2.9.  $\square$

Ainsi, pour caractériser une mesure dans  $\mathbb{R}^d$ , il suffit de connaître sa fonction de répartition. On comprend l'intérêt : une fonction sur  $\mathbb{R}^d$  est un objet bien plus simple qu'une fonction sur  $\mathcal{B}(\mathbb{R}^d)$ .

### 2.3.5. Existence d'ensembles non-boréliens\*

Ce paragraphe est uniquement destiné aux élèves curieux de savoir ce que peut être une partie de  $\mathbb{R}$  non-borélienne.

**Construction.** Notons  $\oplus$  l'addition modulo un sur l'intervalle  $I = [0, 1[$ , soit  $x \oplus y = x + y$  si  $x + y < 1$ , et  $x \oplus y = x + y - 1$  si  $x + y \geq 1$ . Nous définissons une relation d'équivalence  $\mathcal{R}$  sur  $I$  : nous dirons que  $x \mathcal{R} y$  s'il existe un nombre rationnel  $r \in \mathbb{Q} \cap I$  tel que  $x \oplus r = y$ . La classe d'équivalence d'un point  $x \in I$  est donc l'ensemble de  $y \in I$  tels que  $x \oplus r = y$  pour un certain rationnel  $r$  de  $I$ . On notera  $\mathcal{C}$  l'ensemble des classes d'équivalence. Enfin, on se donne une fonction  $s : \mathcal{C} \rightarrow I$  qui à chaque classe d'équivalence  $c \in \mathcal{C}$  associe un élément arbitraire de la classe  $c$ . On définit finalement :

$$H := s(\mathcal{C}) = \{s(c) : c \in \mathcal{C}\}.$$

**Théorème 2.12.**  $H \notin \mathcal{B}(\mathbb{R})$ .

*Démonstration.* On commence par remarquer que

$$I = \bigcup_{r \in \mathbb{Q} \cap I} r \oplus H, \quad (2.2)$$

où  $r \oplus H := \{r \oplus x : x \in H\} = \{r \oplus s(c) : c \in \mathcal{C}\}$ . Il est clair que le membre de droite est inclus dans  $I$  par définition de l'addition modulo un. Inversement, soit  $x$  un élément quelconque de  $I$  et soit  $c \in \mathcal{C}$  sa classe d'équivalence. Par définition de la relation d'équivalence, il existe  $r \in \mathbb{Q} \cap I$  tel que  $x = r \oplus s(c)$ , et l'autre inclusion est démontrée. L'union (2.2) est évidemment dénombrable, et elle est aussi disjointe. En effet, supposons que  $(r \oplus H) \cap (r' \oplus H) \neq \emptyset$  pour deux rationnels  $r, r'$ . Il existe alors  $c, c' \in \mathcal{C}$  tels que  $r \oplus s(c) = r' \oplus s(c')$ . Par les propriétés de l'addition modulo un, cela signifie que  $s(c) \mathcal{R} s(c') : s(c)$  et  $s(c')$  font partie de la même classe d'équivalence, d'où  $c = c'$ . Ceci implique que  $r \oplus s(c) = r' \oplus s(c)$ , et donc que  $r = r'$ . Ainsi, l'union est disjointe.

Supposons par l'absurde que  $H \in \mathcal{B}(\mathbb{R})$ . Les ensembles  $r \oplus H$  sont alors des boréliens. En notant  $\lambda$  la mesure de Lebesgue, nous avons par  $\sigma$ -additivité :

$$1 = \lambda(I) = \sum_{r \in \mathbb{Q} \cap I} \lambda(r \oplus H) = \sum_{r \in \mathbb{Q} \cap I} \lambda(H),$$

où la dernière égalité provient de l'invariance par translation de la mesure de Lebesgue (voir le corollaire 2.10). Si  $\lambda(H) > 0$ , nous avons la contradiction  $1 = +\infty$ . Si  $\lambda(H) = 0$ , nous avons la contradiction  $1 = 0$ .  $\square$

REMARQUE.— La construction de  $H$  repose sur l'existence d'une fonction  $s$  qui à chaque classe d'équivalence associe un de ses représentants. Le choix de cet élément n'est pas spécifié, et rien

ne dit qu'on puisse expliciter une telle fonction (on pourrait naïvement penser à choisir pour  $s(c)$  le plus petit ou le plus grand élément de  $c$ , sauf que l'inf ou le sup ne font pas *a priori* partie de la classe). Nous avons donc *supposé* l'existence d'une telle fonction  $s$ . C'est l'axiome du choix, communément admis, qui rend cette construction légitime. Pour mémoire, un modèle alternatif possible, proposé par R. Solovay, consisterait à renoncer à l'axiome du choix, et à se donner pour *axiome* que toutes les parties de  $\mathbb{R}$  sont boréliennes.

## 2.4. Fonctions mesurables

### 2.4.1. Définition

*nb.* : Il est utile de relire le paragraphe 1.4.1 afin d'être à l'aise avec la notion d'image réciproque d'un ensemble par une application.

Soit  $(\Omega, \mathcal{F})$  et  $(E, \mathcal{E})$  deux espaces mesurables.

**Définition 2.7.** Une application  $X : \Omega \rightarrow E$  est dite  $\mathcal{F}/\mathcal{E}$ -*mesurable*, ou simplement *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 15.— 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  $\Omega$ ). 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 16.— 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  $E$ .

EXEMPLE 17.— Si  $X : \Omega \rightarrow E$  est une application sur  $(E, \mathcal{E})$ , l'ensemble

$$\{X^{-1}(H) : H \in \mathcal{E}\}$$

est la plus petite tribu qui rend  $X$  mesurable (on pourra vérifier à titre d'exercice qu'il s'agit bien d'une tribu, en utilisant la proposition 1.7). On la note  $\sigma(X)$  et on l'appelle la tribu engendrée par  $X$ .

### 2.4.2. Opérations préservant la mesurabilité

Ce paragraphe est destinée à montrer que l'on peut manipuler des fonctions mesurables (sommer, composer, passer à la limite, etc.) tout en préservant la mesurabilité.

**Théorème 2.13.** 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  $X^{-1}(f^{-1}(H')) \in \mathcal{F}$ .  $\square$

Le résultat suivant 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.

**Lemme 2.14.** 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.

*Démonstration.* Posons  $\mathcal{G} = \{H \in \mathcal{E} : X^{-1}(H) \in \mathcal{F}\}$ . On vérifie facilement en utilisant la proposition 1.7 que  $\mathcal{G}$  est une tribu. Par hypothèse,  $\mathcal{C} \subset \mathcal{G}$ , donc  $\sigma(\mathcal{C}) \subset \sigma(\mathcal{G}) = \mathcal{G}$ . On a bien montré  $\mathcal{E} = \mathcal{G}$ , ce qui signifie que  $X$  est mesurable.  $\square$

Le lemme 2.14 admet deux corollaires pratiques.

**Corollaire 2.15.** Soit  $d \in \mathbb{N}^*$  et soient  $X_1, \dots, X_d$  une collection de fonctions de  $\Omega$  dans  $\mathbb{R}$ . On définit  $X : \Omega \rightarrow \mathbb{R}^d$  par  $X(\omega) = (X_1(\omega), \dots, X_d(\omega))$ . Les propositions suivantes sont équivalentes :

- a)  $X_1, \dots, X_d$  sont des fonctions  $\mathcal{F}/\mathcal{B}(\mathbb{R})$ -mesurables ;
- b)  $X$  est une fonction  $\mathcal{F}/\mathcal{B}(\mathbb{R}^d)$ -mesurable.

*Démonstration.* a) $\Rightarrow$ b). Puisque  $\mathcal{B}(\mathbb{R}^d)$  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  $]a, b[ \times \mathbb{R} \times \dots \times \mathbb{R}$  qui appartient à  $\mathcal{B}(\mathbb{R}^d)$ . Donc  $X_1^{-1}(]a, b[) \in \mathcal{F}$ , ce qui montre que  $X_1$  est mesurable.  $\square$

On dit d'une application  $f : \mathbb{R}^d \rightarrow \mathbb{R}^n$ ,  $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^n)$ -mesurable, qu'elle est *borélienne*.

**Corollaire 2.16.** Toute application  $f : \mathbb{R}^d \rightarrow \mathbb{R}^n$  continue est borélienne.

*Démonstration.* On rappelle qu'une fonction continue  $f : \mathbb{R}^d \rightarrow \mathbb{R}^n$  est une fonction qui satisfait la propriété : pour tout ouvert  $U \subset \mathbb{R}^n$ ,  $f^{-1}(U)$  est un ouvert de  $\mathbb{R}^d$ . Par le théorème 2.4,  $\mathcal{B}(\mathbb{R}^n)$  est la tribu engendrée par les ouverts de  $\mathbb{R}^n$ . Or pour tout ouvert  $U \subset \mathbb{R}^n$ ,  $f^{-1}(U)$  est un ouvert de  $\mathbb{R}^d$ , donc  $f^{-1}(U) \in \mathcal{B}(\mathbb{R}^d)$ . On conclut en utilisant le lemme 2.14.  $\square$

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

En combinant les résultats précédents, on peut en déduire quelques propriétés élémentaires.

**Corollaire 2.17.** Si  $X, Y$  sont des fonctions mesurables sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , alors  $X + Y$ ,  $XY$ ,  $X \vee Y := \max(X, Y)$ ,  $X \wedge Y := \min(X, Y)$  sont des fonctions mesurables.

*Démonstration.* La fonction

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

est continue donc borélienne de  $\mathbb{R}^2$  dans  $\mathbb{R}$  en vertu du théorème 2.16. Le résultat se déduit alors du théorème 2.13 par composition de fonctions mesurables. Les autres propriétés se montrent de façon similaire en utilisant les formules

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

$\square$

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  $\overline{\mathbb{R}}$ .

On rappelle que si  $(u_n)$  est une suite réelle, sa limite n'est pas forcément définie, mais en revanche les quantités

$$\begin{aligned} \liminf_{n \rightarrow \infty} u_n &:= \lim_{n \rightarrow \infty} \inf_{k \geq n} u_k \\ \limsup_{n \rightarrow \infty} u_n &:= \lim_{n \rightarrow \infty} \sup_{k \geq n} u_k \end{aligned}$$

le sont toujours. En outre  $\liminf_n u_n \leq \limsup_n u_n$  et la limite de  $u_n$  existe si et seulement s'il y a égalité (auquel cas,  $\lim_n u_n = \liminf_n u_n = \limsup_n u_n$ ). Voir l'annexe pour plus de détails.

Soit  $(X_n, n \in \mathbb{N})$  une suite de variables aléatoires sur  $\overline{\mathbb{R}}$ . On désigne par respectivement par  $\sup_n X_n$ ,  $\inf_n X_n$ ,  $\lim_n X_n$  les fonctions définies sur  $\overline{\mathbb{R}}$  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 2.18.** Soit  $(X_n, n \in \mathbb{N})$  une suite de fonctions mesurables sur  $\overline{\mathbb{R}}$ .

2.18.a)  $\sup_n X_n$ ,  $\inf_n X_n$  sont des fonctions mesurables sur  $\overline{\mathbb{R}}$  ;

2.18.b) Les fonctions  $\liminf_n X_n$  et  $\limsup_n X_n$  sont mesurables ;

2.18.c) Si  $\lim_n X_n$  existe, c'est une fonction mesurable sur  $\overline{\mathbb{R}}$ .

*Démonstration.*

- 2.18.a) Posons  $X := \sup_n X_n$ .  $\mathcal{B}(\overline{\mathbb{R}})$  est la tribu engendrée par les intervalles  $[-\infty, b]$  où  $b \in \overline{\mathbb{R}}$ . 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  $\inf_n X_n = -\sup_n (-X_n)$ .
- 2.18.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  $\limsup_n X_n$ .
- 2.18.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. □

### 2.4.3. Mesure image

Soit un espace mesuré  $(\Omega, \mathcal{F}, \mu)$ . Soit  $(E, \mathcal{E})$  un espace mesurable.

**Théorème 2.19.** Soit  $X : \Omega \rightarrow E$  une application  $\mathcal{F}/\mathcal{E}$ -mesurable. L'application d'ensemble

$$\begin{aligned} \mu X^{-1} : \mathcal{E} &\rightarrow [0, +\infty] \\ H &\mapsto \mu(X^{-1}(H)) \end{aligned}$$

est une mesure sur  $(E, \mathcal{E})$ . On l'appelle la mesure image de  $\mu$  par  $X$ .

*Démonstration.* Laissée à titre d'exercice. Il suffit de vérifier les axiomes définissant une mesure, en s'aidant de la proposition 1.7. □

•

REMARQUE. – Nous achevons ce chapitre en utilisant cette fois un langage probabiliste. Au départ de toute modélisation probabiliste se trouve un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ . Un élément  $A \in \mathcal{F}$  est appelé un événement, sa probabilité est  $\mathbb{P}(A)$ . Une variable aléatoire  $X$  est, comme nous l'avons vu, une fonction mesurable de  $\Omega$  dans un espace d'arrivée  $E$  muni d'une tribu  $\mathcal{E}$ . L'ensemble  $E$  représente les valeurs pouvant être prises par la variable aléatoire. Il s'agit généralement de  $E = \mathbb{R}$  ou  $\mathbb{R}^d$ , muni de sa tribu de Borel. Les événements  $A$  dont on cherche à évaluer la probabilité sont du type  $X^{-1}(H)$  où  $H \in \mathcal{E}$ . En effet,  $\mathbb{P}(X^{-1}(H))$  représente la probabilité de l'événement  $\{X \in H\}$ . La mesure image  $H \mapsto \mathbb{P}(X^{-1}(H))$  est appelée la *loi* de la variable aléatoire  $X$ . On la note  $\mathbf{P}_X = \mathbb{P}X^{-1}$ .

Dans la suite, il nous restera à définir l'espérance d'une variable aléatoire  $X$  à valeurs réelles. Nous la définirons comme une intégrale :

$$\mathbb{E}(X) := \int_{\Omega} X(\omega) d\mathbb{P}(\omega).$$

L'intégrale ci-dessus est appelée *intégrale de Lebesgue* de  $X$  par rapport à la mesure  $\mathbb{P}$ . C'est un objet nouveau, plus sophistiqué que l'intégrale de Riemann vue au lycée. Sa définition est l'objet du chapitre suivant.

## 2.5. Exercices

**EXERCICE 2.1** (Fonction indicatrice).—

Si  $A$  est une partie d'un ensemble  $\Omega$ , on note  $\mathbf{1}_A : \Omega \rightarrow \{0, 1\}$  la fonction indicatrice de l'événement  $A$ , égale à un sur  $A$  et à zéro sur son complémentaire.

1. Que valent les fonctions  $\mathbf{1}_\Omega$  et  $\mathbf{1}_\emptyset$  ?
2. Déterminer  $\mathbf{1}_{\bar{A}}$ ,  $\mathbf{1}_{A \cap B}$ , en fonction de  $\mathbf{1}_A$  et  $\mathbf{1}_B$ .
3. Si  $(A_n)_{n \geq 1}$  forme une suite d'ensembles 2 à 2 disjoints, montrer que  $\mathbf{1}_{\bigcup_{n=1}^{\infty} A_n} = \sum_{n=1}^{\infty} \mathbf{1}_{A_n}$ .

**EXERCICE 2.2** (Image réciproque).— 1. On considère la fonction  $f$  de  $E = \{a, b, c, d\}$  dans  $F = \{1, 2, 3, 4, 5\}$  donnée par :

$$\begin{aligned} f : a &\mapsto 1 \\ b &\mapsto 3 \\ c &\mapsto 3 \\ d &\mapsto 5 \end{aligned}$$

Ecrire  $f^{-1}(A)$  pour les ensembles  $A$  suivants :  $\{1\}, \{2\}, \{3\}, \{1, 2, 3\}, F$ .

2. Soit  $f : E \rightarrow F$  et soient  $A, B$  dans  $F$ . Montrer que  $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$  et  $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$ . Montrer que pour  $A_1, A_2, \dots$  dans  $F$ ,

$$f^{-1}(\cup_i A_i) = \bigcup_i f^{-1}(A_i) \quad \text{et} \quad f^{-1}(\cap_i A_i) = \bigcap_i f^{-1}(A_i).$$

3. Soit  $f : [0, \pi] \rightarrow \mathbb{R}$  donnée par  $f(x) = \sin x$ . Calculer  $f^{-1}([0, \frac{1}{2}])$ .

**EXERCICE 2.3** (Mesure de Dirac).—

Soit  $(\Omega, \mathcal{F})$  un ensemble,  $a \in \Omega$  et  $\delta_a$  l'application de  $\mathcal{F}$  dans  $\mathbb{R}^+$  vérifiant

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

Vérifier que  $\delta_a$  est une probabilité (appelée “mesure de Dirac en  $a$ ”).

**EXERCICE 2.4.**—

Soit  $(\Omega, \mathcal{F})$  un espace mesurable. Soit  $I$  un ensemble fini ou dénombrable. Soit  $(\alpha_i : i \in I)$  une famille de réels positifs et  $(\mu_i : i \in I)$  une famille de mesures sur  $(\Omega, \mathcal{F})$ . L'application d'ensemble

$$\mu = \sum_{i \in I} \alpha_i \mu_i$$

est-elle une mesure ?

**EXERCICE 2.5.**—

Soit  $(\Omega, \mathcal{F}, \mu)$  un espace de probabilité. Soit  $D = \{\omega \in \Omega : \mu(\{\omega\}) \neq 0\}$ .

1. Vérifier que  $D$  s'écrit comme union des ensembles  $D_n = \{\omega : \mu(\{\omega\}) > \frac{1}{n}\}$ .
2. Montrer que  $D_n$  est fini pour tout  $n \geq 1$ .

3. Montrer que l'ensemble des points de discontinuité d'une fonction de répartition est au plus dénombrable.

**EXERCICE 2.6.**–

Tracer et donner l'expression de la fonction de répartition des probabilités  $\mu_1$  et  $\mu_2$  définies sur  $\mathbb{R}$  par  $\mu_1 = \delta_2$  et  $\mu_2 = 0.2 \delta_{-1} + 0.4 \delta_0 + 0.4 \delta_2$ .

**EXERCICE 2.7.**–

Soit  $F$  la fonction de répartition donnée par

$$F(t) = \begin{cases} 0, & t \leq 0, \\ t/4 & 0 \leq t < 1, \\ 1/2 & 1 \leq t < 2, \\ 2/3 + \alpha(1 - \exp(-(t-2))) & t \geq 2. \end{cases}$$

- Déterminer  $\alpha$  et tracer le graphe de  $F$ .
- Montrer que la probabilité  $\mu$  sur  $\mathbb{R}$  dont  $F$  est la fonction de répartition s'écrit comme une somme de mesures de Dirac et d'une mesure ayant une densité par rapport à la mesure de Lebesgue.

**EXERCICE 2.8** (Vrai ou faux?).– 1. Soit la fonction  $f$  définie par

$$\begin{aligned} f : \mathbb{R} &\longrightarrow \mathbb{N} \\ x &\longmapsto [\sqrt{|x|}] \end{aligned}$$

où  $[x]$  est la partie entière de  $x$ . On a

$$f^{-1}([2; 3.5]) = [4; 16[ \cup ] - 16; -4].$$

- Une fonction constante est toujours mesurable.
- Une fonction étagée prend un nombre fini de valeurs.
- Une fonction mesurable qui prend un nombre fini de valeurs est étagée.
- La mesure d'un singleton est strictement positive.
- On a toujours

$$\lim_{x \uparrow x_0} \mu([-\infty; x]) = \mu([-\infty; x_0]).$$

- Pour  $\mu$  mesure de probabilité, on a toujours

$$\lim_{x \downarrow x_0} \mu([-\infty; x]) = \mu([-\infty; x_0]).$$

**EXERCICE 2.9.**–

Soit  $\mu$  la mesure de comptage sur  $\mathbb{N}$  et pour tout entier  $n$ ,  $A_n = \{k, k \geq n\}$ .

- Que vaut  $\cap_n A_n$ ?
- Montre que l'on n'a pas  $\mu(\cap_n A_n) = \lim_{n \rightarrow \infty} \mu(A_n)$ .
- Pourquoi n'y a-t-il pas de contradiction avec 2.7.d)?

EXERCICE 2.10 (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 2.11.—

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 2.12 (\*\*\*)—

Soit  $f$  une fonction de  $\mathbb{R}$  dans  $\mathbb{R}$  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 aller plus loin

EXERCICE 2.13.—

**(Mesure extérieure).** Soit  $E$  un ensemble. On appelle *mesure extérieure* sur  $E$ , toute application  $\mu^* : \mathcal{P}(E) \rightarrow [0, +\infty]$  une définie sur la tribu des parties de  $E$ , et satisfaisant les conditions suivantes :

1.  $\mu^*(\emptyset) = 0$ ;
2.  $\mu^*$  est croissante :  $A \subset B \Rightarrow \mu^*(A) \leq \mu^*(B)$ ;
3.  $\mu^*$  est sous-additive : pour toute suite  $A_n$  de parties de  $E$ ,

$$\mu^*\left(\bigcup_n A_n\right) \leq \sum_n \mu^*(A_n).$$

On dit qu'une partie  $B$  de  $E$  est  $\mu^*$ -mesurable si la propriété suivante est satisfaite :

$$\forall A \in \mathcal{P}(E), \mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c).$$

On note  $\mathcal{M}(\mu^*)$  l'ensemble des ensembles  $\mu^*$ -mesurables. Démontrer les affirmations suivantes :

1.  $\mathcal{M}(\mu^*)$  est une tribu.
2. La tribu  $\mathcal{M}(\mu^*)$  contient toutes les parties  $A$  tels que  $\mu^*(A) = 0$ .
3. La restriction de  $\mu^*$  à  $\mathcal{M}(\mu^*)$  est une mesure.

EXERCICE 2.14.—

**(Existence de la mesure de Lebesgue)** Sur l'ensemble des parties de  $\mathbb{R}$ , on définit l'application :

$$\lambda^*(A) = \inf \left\{ \sum_{i \in \mathbb{N}} b_i - a_i : A \subset \bigcup_{i \in \mathbb{N}} ]a_i, b_i[ \right\},$$

où l'infimum porte sur tous les recouvrements dénombrables de  $A$  par des intervalles ouverts de la forme  $]a_i, b_i[$ , avec  $a_i \leq b_i$ .

1. Montrer que  $\lambda^*$  est une mesure extérieure sur  $\mathbb{R}$ .
2. Montrer que la tribu  $\mathcal{M}(\lambda^*)$  contient  $\mathcal{B}(\mathbb{R})$ . La restriction de  $\lambda^*$  à  $\mathcal{B}(\mathbb{R})$  est appelée la mesure de Lebesgue, on la note  $\lambda$ .
3. Montrer que  $\lambda([a, b]) = b - a$  pour tout segment  $[a, b]$ ,  $a \leq b$ .

**EXERCICE 2.15.**—

**(La mesure de Lebesgue est invariante par translation).** On appelle mesure invariante par translation toute mesure  $\mu$  sur  $\mathcal{B}(\mathbb{R})$  telle que, quel que soit le borélien  $H$  et quel que soit  $x \in \mathbb{R}$ ,  $\mu(H) = \mu(x + H)$ , où  $x + H$  est l'ensemble défini par  $x + H = \{x + y : y \in H\}$ . On souhaite montrer que la mesure de Lebesgue  $\lambda$  est invariante par translation. Si  $\mathcal{P}$  est une collection de sous-ensembles de  $\mathbb{R}$ , on dit d'une mesure  $\mu$  qu'elle est  $\sigma$ -finie sur  $\mathcal{P}$  s'il existe une famille d'ensembles  $A_1, A_2, \dots$  appartenant à  $\mathcal{P}$  telle que  $\Omega = \cup_k A_k$  et  $\mu(A_k) < \infty$  pour tout  $k$ . Si  $\mathcal{P}$  est un  $\pi$ -système (c'est à dire que  $\mathcal{P}$  est stable pour l'intersection), alors deux mesures  $\sigma$ -finies sur  $\mathcal{P}$  qui coïncident sur  $\mathcal{P}$  coïncident sur  $\sigma(\mathcal{P})$ .

1. On pose  $x \in \mathbb{R}$ . Montrer que  $\mu_x : H \mapsto \lambda(x + H)$  définit une mesure sur  $\mathcal{B}(\mathbb{R})$ .
2. Montrer que  $\mu_x$  coïncide avec  $\lambda$  sur les intervalles de la forme  $]a, b]$  où  $a \leq b$ .
3. Montrer que  $\mu_x = \lambda$ .
4. Conclure.

**EXERCICE 2.16.**—

**(Un ensemble non-borélien)** On note  $\oplus$  l'addition modulo 1 sur  $I : \forall (x, y) \in I^2$ ,  $x \oplus y = x + y$  si  $x + y < 1$  et  $x + y - 1$  sinon. Pour tout ensemble  $A$ , on note  $x \oplus A = \{x \oplus y : y \in A\}$ . On rappelle que la mesure de Lebesgue  $\lambda$  est invariante par translation.

1. Montrer que

$$I = \bigcup_{r \in \mathbb{Q} \cap I} (r \oplus A).$$

2. Vérifier qu'il s'agit d'une union dénombrable d'ensembles disjoints.
3. Montrer que pour tout ensemble  $A \subset I$  et tout  $x \in I$ ,  $\lambda(x \oplus A) = \lambda(A)$ .
4. En déduire une contradiction et conclure.
5. On se donne une fonction  $f : \mathcal{C} \rightarrow \mathbb{R}$  telle que pour tout  $c \in \mathcal{C}$ ,  $f(c) \in c$ . Montrer que

$$I = \bigcup_{x \in \mathbb{Q} \cap I} (x + f(\mathcal{C})),$$

où  $f(\mathcal{C}) = \{f(c) : c \in \mathcal{C}\}$ .

6. Montrer qu'il s'agit d'une union disjointe.
7. Supposons que  $f(\mathcal{C})$  est borélien. Evaluer la mesure de Lebesgue de  $I$  de deux manières différentes, par la question 3).
8. En déduire que  $f(\mathcal{C}) \notin \mathcal{B}(\mathbb{R})$ .

**EXERCICE 2.17 (\*\*\*, Ensemble triadique de Cantor).**—

L'objectif est de construire un ensemble non dénombrable de mesure de Lebesgue nulle. Soit  $S$  les éléments de  $\{0, 1, 2\}^{\mathbb{N}}$  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 2.1).
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.

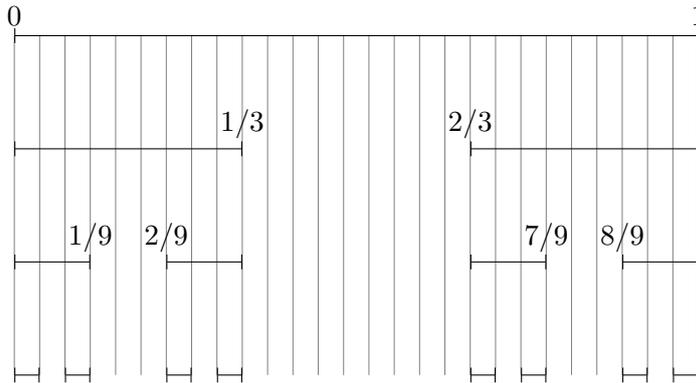


Figure 2.1. – Les premières étapes de la construction de l'ensemble de Cantor.

EXERCICE 2.18 (\*\*\*, 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  $f_1(0) = 0$ ,  $f_1(1) = 1$  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  $C^c$ .

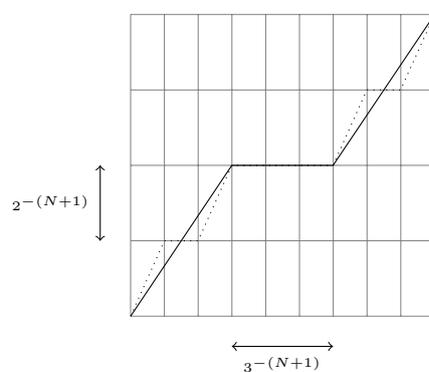


Figure 2.2. – Vue partielle de deux étapes successives dans la construction de la fonction de Cantor.

## 3. Intégration

Soit  $(\Omega, \mathcal{F}, \mu)$  un espace mesuré. On suppose  $\mathbb{R}$  et  $\overline{\mathbb{R}}$  équipés de leurs tribus de Borel. On utilise la convention :

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

### 3.1. L'intégrale de Lebesgue

#### 3.1.1. Définition

**Définition 3.1.** Une fonction  $f$  mesurable de  $(\Omega, \mathcal{F})$  dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  est dite étagée si elle prend un nombre fini de valeurs.

Une telle fonction  $f$  est alors de la forme

$$f(x) = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}(x) \tag{3.1}$$

où  $(\alpha_i, i = 1, \dots, n)$  sont les valeurs distinctes prises par la fonction  $f$  et où pour tout  $i$ ,  $A_i = f^{-1}(\{\alpha_i\})$ .

**Définition 3.2.** La  $\mu$ -intégrale d'une fonction étagée positive  $f$  est la valeur (éventuellement égale à  $+\infty$ ) :

$$\int f \, d\mu = \sum_{i=1}^n \alpha_i \mu(A_i)$$

où les valeurs  $(\alpha_i, i = 1, \dots, n)$  et les ensembles  $(A_i, i = 1, \dots, n)$  sont donnés par (3.1).

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  $A \in \mathcal{F}$ ,

$$\int \mathbf{1}_A \, d\mu = \mu(A) .$$

En ce sens, la notion d'intégrale généralise la notion de mesure.

**Lemme 3.1.** 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.* Pour  $f = \mathbf{1}_A$  et  $g = \mathbf{1}_B$ ,

$$af + bg = a\mathbf{1}_{A \setminus A \cap B} + b\mathbf{1}_{B \setminus A \cap B} + (a + b)\mathbf{1}_{A \cap B}.$$

Par conséquent,

$$\begin{aligned} \int (af + bg) d\mu &= a\mu(A \setminus A \cap B) + b\mu(B \setminus A \cap B) + (a + b)\mu(A \cap B) \\ &= a\mu(A) + b\mu(B) = a \int f d\mu + b \int g d\mu. \end{aligned}$$

Le résultat général s'en déduit par récurrence double sur le nombre de valeurs de  $f$  et  $g$ .  $\square$

On définit maintenant l'intégrale de toute fonction  $f$  mesurable *positive*.

**Définition 3.3.** Soit  $f : \Omega \rightarrow \overline{\mathbb{R}}$  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 3.4.** Soit  $f : \Omega \rightarrow \overline{\mathbb{R}}$  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 dite *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,  $\mu(f)$ .

## 3.1.2. Premières propriétés

**Lemme 3.2.** Soient  $f, g$  deux fonctions mesurables de  $\Omega \rightarrow \overline{\mathbb{R}}$ , 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  $\int f d\mu \leq \int g d\mu$ .

Traitons 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 3.3.** Soient  $f, f_n$  ( $n \in \mathbb{N}$ ) des fonctions mesurables de  $\Omega \rightarrow \overline{\mathbb{R}^+}$  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 :

$$\begin{aligned} \int f_n d\mu &\geq \int f_n \mathbf{1}_{B_n} d\mu \geq \int (1 - \varepsilon)g \mathbf{1}_{B_n} d\mu \\ &= \int \left( \sum_{i=1}^p (1 - \varepsilon)\alpha_i \mathbf{1}_{A_i \cap B_n} \right) d\mu \\ &= \sum_{i=1}^p (1 - \varepsilon)\alpha_i \mu(A_i \cap B_n), \end{aligned}$$

où, dans la dernière égalité, on a utilisé la linéarité de l'intégrale pour les fonctions étagées. Comme  $B_n \uparrow \Omega$ ,  $\mu(A_i \cap B_n)$  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 3.3.

**Lemme 3.4.** Toute fonction mesurable  $f$  à valeurs dans  $\overline{\mathbb{R}^+}$  est limite croissante simple de fonctions étagées positives.

*Démonstration.* Il suffit de considérer

$$f_n(x) = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbf{1}_{[\frac{k}{2^n}, \frac{k+1}{2^n}[}(f(x)) + n \mathbf{1}_{[n, +\infty)}(f(x)).$$

On laisse le soin au lecteur de vérifier que  $f_n(x) \leq f_{n+1}(x)$  et que  $f_n(x) \rightarrow f(x)$ .  $\square$

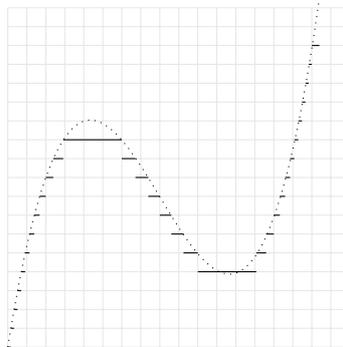


Figure 3.1. – Une fonction (en pointillés) et son approximation par une fonction étagée.

### 3.1.3. Linéarité

**Théorème 3.5.** Soient  $f, g$  des fonctions mesurables de  $\Omega \rightarrow \overline{\mathbb{R}}$ , 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.

$$\text{Alors } \int (af + bg)d\mu = a \int f d\mu + b \int g d\mu.$$

*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 3.3,  $a \int f_n d\mu + b \int g_n d\mu \rightarrow a \int f d\mu + b \int g d\mu$ . 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 3.3.

**Corollaire 3.6.**  $f$  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 3.7.** 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 3.3,  $\int g_n d\mu \rightarrow \int \sum_k f_k d\mu$ . 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.  $\square$

### 3.1.4. Ensembles $\mu$ -négligeables

Un ensemble  $\mu$ -négligeable (ou simplement “négligeable” s’il n’y a pas d’ambiguïté) est un ensemble  $N \in \mathcal{F}$  tel que  $\mu(N) = 0$ .

**Proposition 3.8.** Soit  $f : \Omega \rightarrow \overline{\mathbb{R}}$  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 a  $g(x) = \alpha_i > 0$  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 3.5** (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 convergant vers  $f(x)$ .

**Proposition 3.9.** Soit  $f$  une fonction mesurable positive, de  $\mu$ -intégrale nulle alors  $f$  est nulle  $\mu$ -presque partout.

*Démonstration.* 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  $\mu(A_0) = 0$ .  $\square$

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! Autre exemple : sur  $\mathbb{R}$ , la fonction  $f(x) = x$  est nulle presque partout pour la mesure de Dirac en zéro.

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 3.10 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.

**Théorème 3.10.** 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  $f \mathbf{1}_{\overline{N}} \leq g \mathbf{1}_{\overline{N}}$ , on a  $\int f \mathbf{1}_{\overline{N}} d\mu \leq \int g \mathbf{1}_{\overline{N}} d\mu$ . Par ailleurs  $\int f \mathbf{1}_N d\mu = \int g \mathbf{1}_N d\mu = 0$ . 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. □

**Proposition 3.11.** Soit  $f : \Omega \rightarrow \overline{\mathbb{R}}$  une fonction mesurable positive, telle que  $\int f d\mu < \infty$ . Alors  $f$  est finie  $\mu$ -p.p.

*Démonstration.* Soit  $n$  un entier. Soit  $g$  la fonction égale à  $n$  sur l'ensemble  $\{f = +\infty\}$  et à zéro sinon, autrement dit  $g = n \mathbf{1}_{\{f = +\infty\}}$ . On a  $g \leq f$ , donc  $\int g d\mu \leq \int f d\mu$ , ce qui se lit  $n\mu(\{f = \infty\}) \leq \int f d\mu$ . Ceci est vrai pour tout  $n$ . En faisant tendre  $n$  vers l'infini, on voit qu'il est nécessaire que  $\mu(\{f = \infty\}) = 0$ . □

### 3.1.5. Convergence monotone, convergence dominée

**Théorème 3.12** (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 3.3, en traitant les ensembles négligeables de la même façon que dans la preuve du théorème 3.10.

**Lemme 3.13** (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  $(g_k = \inf_{n \geq k} f_n, k \in \mathbb{N})$  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. □

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.

Le lemme de Fatou est l'ingrédient qui permet de démontrer le théorème de convergence dominée de Lebesgue, fondamental en théorie de l'intégration.

**Théorème 3.14** (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

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int (\lim_{n \rightarrow \infty} f_n) d\mu = \int f d\mu.$$

*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  $\int f d\mu$ .  $\square$

### 3.1.6. Théorème de transfert

Dans ce paragraphe, on se donne un espace mesuré  $(\Omega, \mathcal{F}, \mu)$  et un autre espace mesurable  $(E, \mathcal{E})$ . Soit une application mesurable  $X : \Omega \rightarrow E$ . On rappelle que la mesure image de  $\mu$  par  $X$ , définie pour tout  $A \in \mathcal{E}$  par  $\mu X^{-1}(A) = \mu(X^{-1}(A))$ , définit une mesure sur  $(E, \mathcal{E})$ . Si  $f : E \rightarrow \mathbb{R}$  est une fonction mesurable,  $f \circ X(\omega) = f(X(\omega))$  définit une fonction mesurable de  $\Omega \rightarrow \mathbb{R}$ . Le résultat suivant stipule que l'intégrale par rapport à  $\mu$  de cette variable aléatoire  $f \circ X$  est égale à l'intégrale de  $f$  par rapport à la mesure image.

**Théorème 3.15** (Théorème de transfert). Soit  $f : E \rightarrow \mathbb{R}$  une fonction mesurable.

$$\int_{\Omega} f \circ X d\mu = \int_E f d\mu X^{-1},$$

dès lors que l'une de ces deux intégrales est bien définie.

*Démonstration.* Dans le cas où  $f \geq 0$  est étagée, disons  $f = \sum_{k=1}^K \alpha_k \mathbf{1}_{A_k}$ ,

$$\int f d\mu X^{-1} = \sum_k \alpha_k \mu(X^{-1}(A_k)).$$

Par ailleurs,  $f \circ X$  est également étagée :

$$f \circ X(\omega) = f(X(\omega)) = \sum_{k=1}^K \alpha_k \mathbf{1}_{A_k}(X(\omega)) = \sum_{k=1}^K \alpha_k \mathbf{1}_{X^{-1}(A_k)}(\omega).$$

Par conséquent,  $\int f \circ X d\mu = \sum_k \alpha_k \mu(X^{-1}(A_k))$  et le résultat est démontré.

Dans le cas où  $f \geq 0$  est non-étagée, on choisit  $0 \leq f_n \uparrow f$  avec  $f_n$  étagée. On remarque que  $0 \leq f_n \circ X \uparrow f \circ X$ , et il suffit de passer à la limite l'égalité  $\int f_n \circ X d\mu = \int f_n d\mu X^{-1}$  en utilisant le théorème de convergence monotone pour chaque membre. Dans le cas où  $f$  n'est pas de signe constant, on utilise la décomposition habituelle  $f = f^+ - f^-$ . Cette partie est laissée au lecteur.  $\square$

## 3.2. Exemples fondamentaux

### 3.2.1. Intégration par rapport à une mesure discrète

Soit  $(\Omega, \mathcal{F})$  un espace mesurable. Soit  $I$  un ensemble au plus dénombrable. Commençons par le lemme suivant.

**Lemme 3.16.** Soient  $(\mu_i)_{i \in I}$  une collection de mesures sur  $(\Omega, \mathcal{F})$ . Alors  $\mu := \sum_{i \in I} \mu_i$  est une mesure et pour toute fonction mesurable  $f \geq 0$ ,

$$\int f d\mu = \sum_{i \in I} \int f d\mu_i.$$

*Démonstration.* Considérons le cas où  $f$  est étagée, disons  $f = \sum_{k=1}^K \beta_k \mathbf{1}_{A_k}$ . Alors

$$\int f d\mu = \sum_{k=1}^K \beta_k \mu(A_k) = \sum_{k=1}^K \sum_{i \in I} \beta_k \mu_i(A_k) = \sum_{i \in I} \sum_{k=1}^K \beta_k \mu_i(A_k) = \sum_{i \in I} \int f d\mu_i,$$

où l'interversion des deux sommes provient du fait que les éléments sont positifs. Supposons désormais que  $f$  est positive. Soit  $0 \leq f_n \uparrow f$  une suite de fonctions étagées convergeant vers  $f$ . On a  $\int f_n d\mu \rightarrow \int f d\mu$  par convergence monotone. D'après ce qui précède,  $\int f_n d\mu = \sum_{i \in I} \int f_n d\mu_i$ . Les termes de la somme convergence  $\int f_n d\mu_i \rightarrow \int f d\mu_i$  quand  $n \rightarrow \infty$  par convergence monotone. Si  $I$  est fini, on a immédiatement que  $\int f_n d\mu \rightarrow \sum_{i \in I} \int f d\mu_i$ , le résultat est démontré. Si  $I$  est dénombrable, on peut l'identifier à  $\mathbb{N}$ . On a alors, pour tout  $q \in \mathbb{N}$ ,

$$\sum_{i \in I} \int f_n d\mu_i \geq \sum_{i \leq q} \int f_n d\mu_i$$

et en passant à la limite sur  $n$ ,  $\lim_n \sum_{i \in I} \int f_n d\mu_i \geq \sum_{i \leq q} \int f d\mu_i$ . Comme cela est vrai pour tout  $q$ , on peut passer à la limite en  $q$  et on obtient  $\lim_n \sum_{i \in I} \int f_n d\mu_i \geq \sum_{i \in I} \int f d\mu_i$ . Comme par ailleurs  $\int f_n d\mu_i \geq \int f d\mu_i$ , on a bien l'égalité  $\lim_n \sum_{i \in I} \int f_n d\mu_i = \sum_{i \in I} \int f d\mu_i$ , et le résultat est démontré.  $\square$

**Théorème 3.17.** Soient  $(a_i)_{i \in I}$  une famille d'éléments de  $\Omega$  et  $(\alpha_i)_{i \in I}$  des réels positifs. Soit  $\mu$  la mesure discrète définie par

$$\mu = \sum_{i \in I} \alpha_i \delta_{a_i}. \quad (3.2)$$

Pour toute fonction mesurable  $f : \Omega \rightarrow \overline{\mathbb{R}}_+$  (positive),

$$\int f d\mu = \sum_{i \in I} \alpha_i f(a_i).$$

L'égalité reste vraie pour  $f : \Omega \rightarrow \overline{\mathbb{R}}$  dès lors que  $\sum_{i \in I} \alpha_i |f(a_i)| < \infty$ .

*Démonstration.* On montre d'abord que si  $a \in \Omega$ , alors

$$\int f d\delta_a = f(a), \quad (3.3)$$

pour toute fonction mesurable  $f$ . Ceci provient directement du théorème 3.10. En effet la fonction  $f$  coïncide  $\delta_a$ -p.p. avec la fonction constante égale à  $f(a)$ . Leurs intégrales sont donc égales.

Lorsque  $f$  est positive, le résultat du théorème découle du lemme 3.16 en remarquant que  $\int f d(\alpha_i \delta_{a_i}) = \alpha_i f(a_i)$  (c'est la généralisation immédiate de l'égalité (3.3)). Dans le cas où  $f$  n'est pas à valeurs positives, on procède comme d'habitude en décomposant  $f = f^+ - f^-$ .  $\square$

EXEMPLE 18 (Mesure de comptage).— Si  $\mu = \sum_{k \in \mathbb{N}} \delta_k$ , alors le résultat du théorème 3.17 se lit

$$\int f d\mu = \sum_{k \in \mathbb{N}} f(k).$$

Ainsi, l'intégrale de Lebesgue généralise les concepts de somme ou de série. Il faut garder à l'esprit qu'une série est un cas particulier d'intégrale : c'est une intégrale par rapport à la mesure de comptage sur  $\mathbb{N}$ . A l'avenir, il n'est donc plus utile dorénavant d'établir des théorèmes pour les séries, et des théorèmes pour les intégrales. On pourra relire les résultats fondamentaux précédents à la lumière de cette remarque. Par exemple, le théorème de convergence dominée fournit les conditions d'une permutation limite/somme :

$$\lim_{n \rightarrow \infty} \sum_{k \in \mathbb{N}} f_n(k) = \sum_{k \in \mathbb{N}} \lim_{n \rightarrow \infty} f_n(k)$$

dès lors que  $f_n(k)$  est dominée uniformément en  $n$  par des coefficients  $g(k)$  absolument sommables.

### 3.2.2. Mesures à densité

Dorénavant, nous utiliserons souvent la notation :

$$\int_A f d\mu := \int f \mathbf{1}_A d\mu.$$

**Théorème 3.18.** Soit  $(\Omega, \mathcal{F}, \nu)$  un espace mesuré, avec  $\nu$  une mesure  $\sigma$ -finie. Soit  $f : \Omega \rightarrow \mathbb{R}_+$  une fonction mesurable.

1. L'application  $\mu$  définie pour tout  $A \in \mathcal{F}$  par

$$\mu(A) = \int_A f d\nu \quad (3.4)$$

est une mesure sur  $(\Omega, \mathcal{F})$ .

2. Soit  $g : \Omega \rightarrow \mathbb{R}_+$  une autre fonction telle que  $\mu(A) = \int_A g d\nu$  pour tout  $A \in \mathcal{F}$ . Alors  $g = f$   $\nu$ -p.p.

*Démonstration.* 1) La preuve est laissée au lecteur. 2) Puisque  $\nu$  est  $\sigma$ -finie, il existe une suite  $A_n \uparrow \Omega$  telle que  $\nu(A_n) < \infty$  pour tout  $n$ . Posons  $B_n = \{f + \frac{1}{n} < g < n\}$ . On a :  $A_n \cap B_n \uparrow \{f < g\}$ .

Or, par définition de  $f$  et  $g$ ,  $\int \mathbf{1}_{A_n \cap B_n} f d\nu = \int \mathbf{1}_{A_n \cap B_n} g d\nu$ . Ces deux intégrales sont finies (elles sont majorées par  $n\nu(A_n)$ , qui est fini grâce au choix de  $A_n$ ). On a donc, par linéarité de l'intégrale :  $0 = \int \mathbf{1}_{A_n \cap B_n} (g - f) d\nu \geq \int \mathbf{1}_{A_n \cap B_n} \frac{1}{n} d\nu = \frac{1}{n} \nu(A_n \cap B_n)$ . Donc  $\nu(A_n \cap B_n) = 0$ . En passant à la limite,  $A_n \cap B_n \uparrow \{f < g\}$ , on obtient que  $\nu(f < g) = 0$ . Par symétrie,  $\nu(g < f)$  et *in fine*  $\nu(f - g) = 0$ .  $\square$

**Définition 3.6** (Mesure à densité). Soient  $\nu, \mu$  deux mesures sur  $(\Omega, \mathcal{F})$ . On dit que  $\mu$  admet une densité par rapport à  $\nu$  s'il existe une fonction mesurable  $f : \Omega \rightarrow \mathbb{R}_+$  telle que l'égalité (3.4) est satisfaite pour tout  $A \in \mathcal{F}$ . Dans ce cas,  $f$  est appelée une *densité de  $\mu$  par rapport à  $\nu$* .

D'après le théorème 3.18, la densité est unique à un ensemble  $\nu$ -négligeable près.

Dans les applications, le contexte le plus fréquent est le cas où  $\Omega = \mathbb{R}^d$  et  $\nu$  est la mesure de Lebesgue sur  $\mathbb{R}^d$ . Lorsqu'on dit qu'une mesure  $\mu$  sur  $\mathbb{R}^d$  admet une densité, sans préciser la mesure de référence  $\nu$ , on sous-entend que  $\nu$  est la mesure de Lebesgue. Si en outre  $\mu$  est une mesure de probabilité, la densité  $f$  vérifie  $\int f d\lambda_d = 1$  (appliquer la définition avec  $A = \mathbb{R}^d$ ). On dit alors de  $f$  que c'est une *densité de probabilité*. La table 3.1 fournit différents exemples de densités de probabilité, sur lesquels nous aurons l'occasion de revenir pendant ce cours.

**Théorème 3.19.** Soient  $\mu, \nu$  deux mesures sur  $(\Omega, \mathcal{F})$  telles que  $\mu$  admet une densité  $f$  par rapport à  $\nu$ . Soit  $g : \Omega \rightarrow \mathbb{R}$  une application mesurable, positive ou  $\mu$ -intégrable. Alors,

$$\int g d\mu = \int g f d\nu.$$

*Démonstration.* Dans le cas où  $g \geq 0$  est étagée, disons  $g = \sum_{k=1}^K \alpha_k \mathbf{1}_{A_k}$ ,

$$\int g d\mu = \sum_k \alpha_k \mu(A_k) = \sum_k \alpha_k \int \mathbf{1}_{A_k} f d\nu = \int \sum_k \alpha_k \mathbf{1}_{A_k} f d\nu = \int g f d\nu.$$

Dans le cas où  $g \geq 0$  est non-étagée, on choisit  $0 \leq g_n \uparrow g$  avec  $g_n$  étagée. D'une part,  $\int g_n d\mu \rightarrow \int g d\mu$ . D'autre part, puisque  $0 \leq g_n f \uparrow g f$ , on a aussi  $\int g_n f d\nu \rightarrow \int g f d\nu$ . Le résultat est démontré. Dans le cas où  $g$  n'est pas de signe constant, mais  $\mu$ -intégrable, on utilise la décomposition habituelle, en remarquant que  $(g f)^\pm = g^\pm f$ .  $\square$

On se pose la question suivante : comment peut on reconnaître qu'une mesure  $\mu$  admet une densité par rapport à  $\nu$  ?

**Définition 3.7.** Soit  $\mu$  et  $\nu$  deux mesures sur  $(\Omega, \mathcal{F})$ . On dit que  $\mu$  est *absolument continue* par rapport à  $\nu$ , noté  $\mu \ll \nu$ , si pour tout  $A \in \mathcal{F}$ ,

$$\nu(A) = 0 \implies \mu(A) = 0.$$

D'après la proposition 3.8, il est évident que si  $\mu$  est de la forme (3.4), alors  $\mu \ll \nu$ . Le théorème suivant, admis, fournit une réciproque.

**Théorème 3.20** (Radon-Nikodym). Soient  $\mu$  et  $\nu$  deux mesures  $\sigma$ -finies telles que  $\nu \ll \mu$ . Alors  $\mu$  admet une densité par rapport à  $\nu$ .

On emploie souvent la notation symbolique

$$d\mu = f d\nu$$

pour désigner le fait que  $\mu$  admet une densité  $f$  par rapport à  $\nu$ . De même, lorsque  $\mu \ll \nu$ , on désigne souvent par  $f = \frac{d\mu}{d\nu}$  la densité de  $\mu$  par rapport à  $\nu$ .  $\frac{d\mu}{d\nu}$  est appelée la dérivée de Radon-Nikodym. Elle est définie de manière unique à un ensemble  $\nu$ -négligeable près.

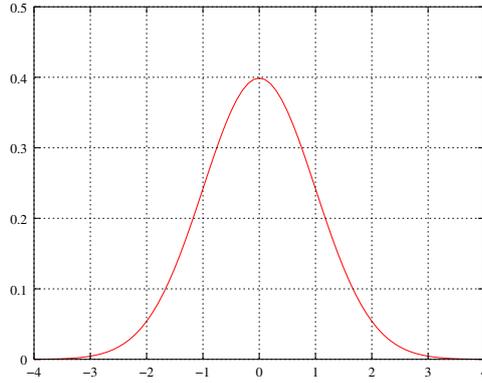
Domaine	Densité	Expression de $f(x)$	Notation
$\mathbb{R}$	Densité uniforme sur $[a, b]$	$\frac{\mathbf{1}_{[a,b]}(x)}{b-a}$	$\mathcal{U}([a, b])$
$\mathbb{R}^d$	Densité uniforme sur une partie $A \subset \mathbb{R}^d$	$\frac{\mathbf{1}_A(x)}{\int \mathbf{1}_A}$	
$\mathbb{R}$	Densité exponentielle de paramètre $\alpha > 0$	$\alpha e^{-\alpha x} \mathbf{1}_{\mathbb{R}_+}(x)$	$\mathcal{E}(\alpha)$
$\mathbb{R}$	Densité gaussienne de paramètres $m, \sigma^2$ ( $m \in \mathbb{R}, \sigma^2 > 0$ )	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/(2\sigma^2)}$	$\mathcal{N}(m, \sigma^2)$
$\mathbb{R}^d$	Gaussienne multivariée de paramètres $m, \Sigma$ ( $m \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d \times d}$ définie positive)	$\frac{e^{-\frac{1}{2}(x-m)^T \Sigma^{-1} (x-m)}}{\sqrt{(2\pi)^d \det \Sigma}}$	$\mathcal{N}_d(m, \Sigma)$
$\mathbb{R}$	Densité de Cauchy de paramètres $m, \alpha$ ( $m \in \mathbb{R}, \alpha > 0$ )	$\frac{1}{\pi} \cdot \frac{\alpha}{(x-m)^2 + \alpha^2}$	
$\mathbb{R}$	Densité Gamma de paramètres $a, b$ ( $a >, b > 0$ )	$\frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$	$\Gamma(a, b)$

Table 3.1. – Quelques exemples de densités de probabilité – (Rappel :  $\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$ ).

A titre d'exemple, la figure 3.2 représente la densité gaussienne de paramètres 0 et 1, notée  $\mathcal{N}(0, 1)$  et appelée gaussienne *centrée réduite* :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (3.5)$$

La densité gaussienne porte aussi le nom de densité *normale*, ce qui justifie la notation  $\mathcal{N}$ .

Figure 3.2. – Densité gaussienne centrée réduite  $\mathcal{N}(0, 1)$ .

EXEMPLE 19. – Soit  $\mu$  la mesure discrète donnée par l'équation (3.2). Soit  $\nu = \sum_{i \in I} \delta_{a_i}$  la mesure de comptage des  $(a_i)$ . Alors, par application du théorème 3.17,  $\mu$  peut également s'écrire sous la forme (3.4). Toute fonction  $f$  telle que  $f(a_i) = \alpha_i$  est une densité de  $\mu$  par rapport à  $\nu$ . Cet exemple est certes artificiel, mais il montre qu'une mesure discrète est une mesure à densité par rapport à la mesure de comptage.

### 3.3. Intégrale par rapport à la mesure de Lebesgue

Dans ce paragraphe, nous étudions le cas où  $(\Omega, \mathcal{F}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  et où on intègre par rapport à la mesure de Lebesgue  $\lambda_d$ . La notation courte  $\int f$  est également utilisée pour désigner  $\int f d\lambda_d$ .

#### 3.3.1. Lien avec l'intégrale de Riemann

La mesure de Lebesgue sur  $\mathbb{R}$  est telle que  $\lambda([a, b]) = b - a$ . Autrement dit,

$$\int \mathbf{1}_{[a, b]} d\lambda = b - a = \int_a^b dx, \quad (3.6)$$

où  $\int f(x) dx$  représente l'intégrale de Riemann. La remarque se généralise aux fonctions en escalier (on rappelle qu'une fonction en escalier est une fonction étagée avec des  $A_i$  qui sont des intervalles) : les intégrales de Riemann et Lebesgue coïncident.

On rappelle qu'une fonction  $f$  sur un intervalle  $[a, b]$  est Riemann-intégrable si elle s'écrit comme limite uniforme d'une suite de fonctions en escalier. Par définition, son intégrale de Riemann est la limite des intégrales de ces fonctions en escalier. On peut démontrer sans grande difficulté qu'une telle fonction est mesurable, et que son intégrale de Riemann coïncide avec son intégrale de Lebesgue (voir [?] pour la démonstration). Autrement dit, *quand l'intégrale de Riemann existe, elle coïncide avec l'intégrale de Lebesgue*. On s'autorisera donc la notation

$$\int f d\lambda = \int f(x) dx,$$

même si maintenant le terme de droite désigne l'intégrale de Lebesgue de  $f$ .

Par conséquent, les manipulations habituelles de l'intégrale de Riemann restent valables pour l'intégrale de Lebesgue : lien entre primitive et intégrale, dérivation d'une intégrale, intégration par parties, etc. On s'en épargnera ici la redémonstration.

REMARQUE.— L'intégration « à la Lebesgue » permet d'intégrer plus de fonctions qu'avec l'intégrale de Riemann : pour être Riemann intégrable, une fonction doit être la limite uniforme de fonctions en escalier. Pour être Lebesgue intégrable, une fonction (positive) peut se contenter d'être mesurable, ce qui est bien moins contraignant. Par exemple, l'indicatrice des nombres rationnels  $\mathbf{1}_{\mathbb{Q}}$  est Lebesgue intégrable sur tout segment, mais n'est pas Riemann-intégrable.

REMARQUE.— Les avantages de l'intégrale de Lebesgue ne se limitent pas au fait qu'on peut intégrer plus de fonctions (là n'était pas la motivation d'H. Lebesgue). Avec l'intégrale de Riemann, le théorème de convergence dominée, qui dit en substance

$$f_n \rightarrow f \Rightarrow \int f_n \rightarrow \int f,$$

nécessite, outre la domination, d'ajouter l'hypothèse que la limite  $f$  s'écrit elle-même comme limite uniforme de fonctions en escalier. Cette hypothèse, délicate à vérifier, restreint la portée du résultat. C'est un caillou dans la chaussure. Elle n'est pas nécessaire avec l'intégrale de Lebesgue : toute limite de fonctions mesurables est forcément mesurable.

Enfin, un élément du succès de l'intégrale de Lebesgue, découvert plus tardivement mais essentiel en analyse, est la *complétude* de l'espace des fonctions intégrables, dont il a été question plus haut. Ce point apparaîtra clairement aux élèves dans le cours d'analyse.

### 3.3.2. Formule du changement de variable

Rappelons d'abord quelques notions. Soient  $U, V$  deux ouverts de  $\mathbb{R}^d$ . Pour toute fonction  $\phi : U \rightarrow V$ , on notera  $\phi_1, \dots, \phi_d$  les composantes de  $\phi$ , c'est à dire :

$$\phi(x) = (\phi_1(x), \dots, \phi_d(x)).$$

Dire que l'application  $\phi$  est de classe  $\mathcal{C}^1$  revient à dire que ses dérivées partielles existent et sont continues. La jacobienne de  $\phi$  au point  $x \in U$  est la matrice  $J_\phi(x)$  définie par

$$J_\phi(x) := \left( \frac{\partial \phi_i}{\partial x_j}(x), 1 \leq i \leq d, 1 \leq j \leq d \right)$$

$$= \begin{pmatrix} \frac{\partial \phi_1}{\partial x_1}(x) & & & & \\ & \vdots & & & \\ \dots & \frac{\partial \phi_i}{\partial x_j}(x) & \dots & & \\ & \vdots & & & \\ & & & & \frac{\partial \phi_d}{\partial x_d}(x) \end{pmatrix}.$$

Le jacobien de  $\phi$  (ou déterminant jacobien) en  $x$  est le déterminant de  $J_\phi(x)$ .

**Définition 3.8.** Soient  $U, V$  deux ouverts de  $\mathbb{R}^d$ . Une application  $\phi : U \rightarrow V$  est un  $\mathcal{C}^1$ -difféomorphisme si

- $\phi$  est bijective ;
- $\phi$  est de classe  $\mathcal{C}^1$  ;
- son inverse  $\phi^{-1}$  est de classe  $\mathcal{C}^1$ .

**Proposition 3.21.** Soient  $U$  et  $V$  deux ouverts de  $\mathbb{R}^d$  et  $\phi : U \rightarrow V$  une application bijective de classe  $\mathcal{C}^1$ . Alors  $\phi$  est un difféomorphisme si et seulement si  $\det J_\phi(x) \neq 0, \forall x \in U$ .

On rappelle la propriété suivante des difféomorphismes :

$$\det J_{\phi^{-1}} = \frac{1}{(\det J_\phi) \circ \phi^{-1}}.$$

**Théorème 3.22.** (*Formule du changement de variable*). Soient  $U$  et  $V$  deux ouverts de  $\mathbb{R}^d$  et  $\phi : U \rightarrow V$  un difféomorphisme. Pour toute fonction  $f$  borélienne positive définie sur  $V$ ,

$$\int_U f \circ \phi = \int_V \frac{f}{|(\det J_\phi) \circ \phi^{-1}|}. \quad (3.7)$$

REMARQUE.— Dans le cas où  $f$  n'est pas nécessairement positive, alors l'égalité (3.7) 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.

### 3.4. Espaces produit et théorème de Fubini

Le théorème de Fubini est celui qui permet de calculer des intégrales multiples en choisissant l'ordre des intégrations.

#### 3.4.1. Tribu produit

On se donne deux espaces mesurables  $(E, \mathcal{E})$  et  $(F, \mathcal{F})$ .

Si on veut construire une mesure sur le produit cartésien  $E \times F$ , la première difficulté à surmonter est la définition de la tribu sur cet espace. Sans trop réfléchir, on pourrait équiper  $E \times F$  des ensembles de la forme  $A \times B$ , où  $A \in \mathcal{E}$  et  $B \in \mathcal{F}$ . Mais ceci ne constitue pas une tribu. Et même si c'était le cas, on peut souhaiter mesurer autre chose que des ensembles s'écrivant comme des produits cartésiens. Qu'à cela ne tienne, on note  $\mathcal{E} \otimes \mathcal{F}$  la plus petite tribu qui contient les produits cartésiens  $A \times B$ .

**Définition 3.9.** La *tribu produit* des espaces  $(E, \mathcal{E})$  et  $(F, \mathcal{F})$  est définie comme la tribu sur  $E \times F$  engendrée par les ensembles de la forme  $A \times B$  où  $A \in \mathcal{E}$  et  $B \in \mathcal{F}$ . On la note  $\mathcal{E} \otimes \mathcal{F}$ .

Naturellement,  $\mathcal{E} \otimes \mathcal{F}$  ne se limite pas aux ensembles de la forme  $A \times B$ . Il ne s'agit donc pas d'un produit cartésien au sens usuel.

**Proposition 3.23.**  $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$ .

*Démonstration.* Tout produit cartésien d'intervalles  $A \times B$  est dans la tribu  $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$  donc  $\mathcal{B}(\mathbb{R}^2) \subset \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ . On montre l'autre inclusion. Si  $A$  est un intervalle, on vérifie que  $\{B : A \times B \in \mathcal{B}(\mathbb{R}^2)\}$  forme une tribu sur  $\mathbb{R}$  : la preuve est laissée à titre d'exercice. Cette tribu contient les intervalles et donc contient  $\mathcal{B}(\mathbb{R})$ . Autrement dit, pour tout  $B \in \mathcal{B}(\mathbb{R})$ , l'ensemble  $\{A : A \times B \in \mathcal{B}(\mathbb{R}^2)\}$  contient tous les intervalles. Comme il s'agit là encore d'une tribu, cet ensemble contient  $\mathcal{B}(\mathbb{R})$ . On a finalement montré que pour tout  $A \in \mathcal{B}(\mathbb{R})$ ,  $B \in \mathcal{B}(\mathbb{R})$ ,  $A \times B \in \mathcal{B}(\mathbb{R}^2)$ . Donc  $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \subset \mathcal{B}(\mathbb{R}^2)$ .  $\square$

**Lemme 3.24.** Soit  $f : E \times F \rightarrow \overline{\mathbb{R}}$  une fonction  $\mathcal{E} \otimes \mathcal{F}/\mathcal{B}(\overline{\mathbb{R}})$ -mesurable. Pour tout  $x \in E$ , la fonction  $y \mapsto f(x, y)$  est mesurable. Pour tout  $y \in F$ , la fonction  $x \mapsto f(x, y)$  est mesurable.

*Démonstration.* Pour  $x$  fixé, on définit la fonction  $T_x : y \mapsto (x, y)$  de  $F$  dans  $E \times F$ . Pour tout  $A \in \mathcal{E}$  et  $B \in \mathcal{F}$ ,  $T_x^{-1}(A \times B)$  vaut  $B$  si  $x \in A$  et vaut  $\emptyset$  sinon, et donc appartient à  $\mathcal{F}$  dans tous les cas. Puisque la classe des ensembles  $A \times B$  engendre  $\mathcal{E} \otimes \mathcal{F}$ ,  $T_x$  est mesurable d'après le lemme 2.14 et donc  $T_x^{-1}(G) \in \mathcal{F}$ . Donc  $T_x$  est  $\mathcal{F}/(\mathcal{E} \otimes \mathcal{F})$ -mesurable. La fonction  $y \mapsto f(x, y)$  s'écrit comme la composée  $f \circ T_x$  de deux fonctions mesurables. Elle est donc mesurable.  $\square$

### 3.4.2. Mesure produit

**Théorème 3.25.** Soient  $\mu$  et  $\nu$  deux mesures  $\sigma$ -finies sur  $(E, \mathcal{E})$  et  $(F, \mathcal{F})$  respectivement. Il existe une unique mesure sur  $\mathcal{E} \otimes \mathcal{F}$ , appelée mesure-produit et notée  $\mu \otimes \nu$  telle que pour tout  $A \in \mathcal{E}$ ,  $B \in \mathcal{F}$  :

$$\mu \otimes \nu(A \times B) = \mu(A) \cdot \nu(B) . \quad (3.8)$$

*Démonstration. Unicité.* Soient  $\pi_1$  et  $\pi_2$  deux mesures sur  $\mathcal{E} \otimes \mathcal{F}$  satisfaisant (3.8). Elles coïncident sur l'ensemble des produits cartésiens du type  $A \times B$ . Or cet ensemble est un  $\pi$ -système qui engendre  $\mathcal{E} \otimes \mathcal{F}$ . Donc  $\pi_1 = \pi_2$  par le théorème 2.9.

*Existence.* On construit cette mesure. Pour tout  $G \in \mathcal{E} \otimes \mathcal{F}$ , posons :

$$\mu \otimes \nu(G) := \int \left( \int \mathbf{1}_G(x, y) d\nu(y) \right) d\mu(x) . \quad (3.9)$$

En toute rigueur, il faudrait vérifier que cette définition est bien légitime, en ce sens que la fonction

$$x \mapsto \int \mathbf{1}_G(x, y) d\nu(y)$$

est bien mesurable. C'est le point technique de la preuve. La preuve est fournie dans l'annexe A.5.3 et peut être omise. La suite, en revanche, est facile. Une fois l'objet (3.9) défini, on vérifie facilement qu'il satisfait la propriété annoncée :

$$\mu \otimes \nu(A \times B) = \int \left( \int \mathbf{1}_A \mathbf{1}_B d\nu \right) d\mu = \int \nu(B) \mathbf{1}_A d\mu = \nu(B) \mu(A).$$

Il reste à vérifier que c'est une mesure. Soit  $(G_n)_n$  une suite d'éléments deux à deux disjoints de  $\mathcal{E} \otimes \mathcal{F}$ . On a

$$\begin{aligned} \mu \otimes \nu \left( \bigcup_n G_n \right) &= \int \left( \int \sum_n \mathbf{1}_{G_n}(x, y) d\nu(y) \right) d\mu(x) \\ &= \int \sum_n \left( \int \mathbf{1}_{G_n}(x, y) d\nu(y) \right) d\mu(x) \\ &= \sum_n \int \left( \int \mathbf{1}_{G_n}(x, y) d\nu(y) \right) d\mu(x) = \sum_n \mu \otimes \nu(G_n), \end{aligned}$$

où les deux permutations somme/intégrale sont justifiées par le corollaire 3.7. Par le théorème de convergence monotone,  $\mu \otimes \nu(\cup_n G_n) = \sum_n \int f_{G_n} d\mu = \sum_n \mu \otimes \nu(G_n)$ . Donc  $\mu \otimes \nu$  est bien une mesure.  $\square$

La mesure de Lebesgue sur  $\mathbb{R}^2$  satisfait  $\lambda_2(A \times B) = \lambda_1(A)\lambda_1(B)$ . Elle est donc égale à la mesure produit  $\lambda_1 \otimes \lambda_1$ .

### 3.4.3. Théorème de Fubini-Tonelli

**Théorème 3.26** (Fubini-Tonelli). Soient  $(E, \mathcal{E}, \mu)$  et  $(F, \mathcal{F}, \nu)$  deux espace mesurés, où  $\mu, \nu$  sont  $\sigma$ -finies. Soit  $f : E \times F \rightarrow \overline{\mathbb{R}}$  une fonction  $\mathcal{E} \otimes \mathcal{F}/\mathcal{B}(\overline{\mathbb{R}})$ -mesurable à valeurs positives. Alors, les applications

$$x \mapsto \int f(x, y) d\nu(y) \quad \text{et} \quad y \mapsto \int f(x, y) d\mu(x)$$

sont mesurables, et on a l'égalité

$$\int f(x, y) d(\mu \otimes \nu)(x, y) = \int \left( \int f(x, y) d\nu(y) \right) d\mu(x) = \int \left( \int f(x, y) d\mu(x) \right) d\nu(y). \quad (3.10)$$

*Démonstration.* • Dans le cas où  $f = \mathbf{1}_G$ , avec  $G \in \mathcal{E} \otimes \mathcal{F}$ , le résultat est déjà démontré. Le premier point, la mesurabilité de  $x \mapsto \int f(x, y) d\nu(y)$ , a été évoqué dans la preuve du théorème 3.25. Le second point découle de l'expression de la mesure produit, donnée par l'équation (3.9). Cette équation se lit précisément  $\int f d\mu \otimes \nu = \int \int f(x, y) d\nu(y) d\mu(x)$  avec  $f = \mathbf{1}_G$ . Par ailleurs, en inversant les rôles joués par  $\mu$  et  $\nu$ , on aurait tout aussi bien pu poser, à la place de l'équation (3.9),

$$\mu \otimes \nu(G) = \int \int \mathbf{1}_G(x, y) d\mu(x) d\nu(y). \quad (3.11)$$

Ceci démontre que  $\int f d\mu \otimes \nu = \int \int f(x, y) d\mu(x) d\nu(y)$ .

• Le résultat étant valable pour les fonctions indicatrices, il est valable également pour les fonctions étagées positives par linéarité.

• Traitons le cas d'une fonction  $f \geq 0$  générale. Soit  $0 \leq f_n \uparrow f$  une suite de fonctions étagées convergeant vers  $f$ . Pour tout  $x$ ,  $\int f_n(x, y) d\nu(y)$  est mesurable (d'après ce qui précède) et converge vers  $\int f(x, y) d\nu(y)$  par le théorème de convergence monotone. Donc  $x \mapsto \int f(x, y) d\nu(y)$  est mesurable comme limite de fonctions mesurables. De plus, grâce au point précédent,

$$\int f_n d(\mu \otimes \nu) = \int \left( \int f_n(x, y) d\nu(y) \right) d\mu(x).$$

Il suffit d'appliquer le théorème de convergence monotone aux deux membres de cette égalité, on obtient  $\int f d(\mu \otimes \nu) = \int (\int f(x, y) d\nu(y)) d\mu(x)$ . En inversant les rôles de  $\mu$  et  $\nu$  dans la preuve, on montre de même que  $\int f d(\mu \otimes \nu) = \int (\int f(x, y) d\mu(x)) d\nu(y)$ .  $\square$

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 obtient la même conclusion, c'est le théorème de Fubini proprement dit. Il y a toutefois une subtilité : elle est liée au fait que pour  $y$  fixé, l'intégrale

$$\int f(x, y) d\nu(y)$$

n'a pas de raison d'être bien définie pour tout  $x$ , puisque  $f$  n'est pas positive. Ce n'est pas grave : le théorème de Fubini établit que cette intégrale est bien définie (et même de valeur finie) pour tout  $x$  hors d'un ensemble  $\mu$ -négligeable. On peut alors introduire une fonction de  $x$  égale à zéro sur cet ensemble négligeable, et égale à  $\int f(x, y) d\nu(y)$  ailleurs. Nous noterons toujours cette fonction

$$x \mapsto \int f(x, y) d\nu(y)$$

et nous dirons qu'elle est *bien définie*  $\mu$ -p.p. pour traduire le fait qu'elle est mesurable, et qu'elle prend la valeur  $\int f(x, y) d\nu(y)$  hors d'un ensemble  $\mu$ -négligeable. Nous dirons qu'elle est *finie*  $\mu$ -p.p. si en outre cette valeur finie.

**Théorème 3.27** (Fubini). Soient  $(E, \mathcal{E}, \mu)$  et  $(F, \mathcal{F}, \nu)$  deux espace mesurés, où  $\mu, \nu$  sont  $\sigma$ -finies. Soit  $f : E \times F \rightarrow \overline{\mathbb{R}}$  une fonction  $\mathcal{E} \otimes \mathcal{F}/\mathcal{B}(\overline{\mathbb{R}})$ -mesurable telle que

$$\iint |f(x, y)| d\mu(x) d\nu(y) < \infty.$$

Alors, les applications

$$x \mapsto \int f(x, y) d\nu(y) \quad \text{et} \quad y \mapsto \int f(x, y) d\mu(x)$$

sont finies  $\mu$  et  $\nu$  presque partout respectivement, et l'égalité (3.10) est satisfaite.

*Démonstration.* D'après le théorème de Fubini-Tonelli,

$$\int \left( \int |f(x, y)| d\nu(y) \right) d\mu(x) < \infty$$

donc la fonction

$$x \mapsto \int |f(x, y)| d\nu(y)$$

est finie  $\mu$ -presque-sûrement d'après la proposition 3.11. : si l'on note

$$N_E = \left\{ x \in E, \int |f(x, y)| d\nu(y) = \infty \right\}$$

alors  $\mu(N_E) = 0$  et donc, par construction de la mesure produit,  $(\mu \otimes \nu)(N_E \times F) = 0$ . De même,

$$N_F = \left\{ y \in F, \int |f(x, y)| d\mu(x) = \infty \right\}$$

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 de la linéarité de (3.10).  $\square$

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!

### 3.4.4. Produit d'ordre supérieur

Tout ce qui précède se généralise immédiatement aux fonctions de plus de deux variables. Soient  $(E_1, \mathcal{E}_1, \mu_1), \dots, (E_d, \mathcal{E}_d, \mu_d)$  des espaces mesurés, où les mesures  $\mu_1, \dots, \mu_d$  sont  $\sigma$ -finies. L'espace produit  $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_d$ , aussi noté  $\mathcal{E}^{\otimes d}$  si tous les  $\mathcal{E}_i$  coïncident avec  $\mathcal{E}$ , est la tribu sur  $E_1 \times \dots \times E_d$  engendrée par les produits cartésiens de la forme  $\prod_{i=1}^d A_i$  où  $A_i \in \mathcal{E}_i$ . On peut démontrer que :  $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d}$ .

De même que dans le cas  $d = 2$ , il existe une unique mesure sur  $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_d$ , notée  $\bigotimes_{i=1}^d \mu_i$ , telle que pour tout  $A_i \in \mathcal{E}_i$ ,

$$\bigotimes_{i=1}^d \mu_i (A_1 \times \dots \times A_d) = \prod_{i=1}^d \mu_i(A_i) .$$

Enfin, le théorème de Fubini se lit ainsi. Pour toute fonction mesurable  $f : \prod_i E_i \rightarrow \mathbb{R}$  positive ou intégrable par rapport à  $\bigotimes_{i=1}^d \mu_i$  et pour toute permutation  $(i_1, \dots, i_d)$  de  $\{1, \dots, d\}$  :

$$\int f d \left( \bigotimes_{i=1}^d \mu_i \right) = \int_{E_{i_d}} \dots \left( \int_{E_{i_1}} f(x_1, \dots, x_d) d\mu_{i_1}(x_{i_1}) \right) \dots d\mu_{i_d}(x_{i_d}) .$$

## 3.5. Exercices

### Preliminaire

EXERCICE 3.1.–

Soit  $f : \Omega \rightarrow \mathbb{R}$  une fonction à valeurs positives. Pour tout  $n$  on définit

$$g_n(x) = \begin{cases} \frac{k}{2^n} & \text{si } \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n} \quad (k = 0, \dots, n2^n - 1) \\ n & \text{si } n \leq f(x) \end{cases}$$

Montrer que  $g_n$  est une suite croissante de fonctions étagées qui converge simplement vers  $f$ .

### Propriétés de l'intégrale

EXERCICE 3.2.–

Soit  $f : \Omega \rightarrow [0, +\infty]$  telle que  $\int f d\mu = 0$ . Montrer que  $f = 0$   $\mu$ -p.p.

*Indication : Remarquer que  $\{f > 0\} = \bigcup_n \{f \geq \frac{1}{n}\}$  et minorer  $\mu(\{f \geq \frac{1}{n}\})$ .*

EXERCICE 3.3.–

Soit  $f : \Omega \rightarrow [0, +\infty]$  telle que  $\int f d\mu < \infty$ . Montrer que  $f$  est finie  $\mu$ -p.p.

EXERCICE 3.4.–

Soient  $\mu$  et  $\nu$  deux mesures sur  $(\Omega, \mathcal{F})$  et  $\alpha, \beta \geq 0$ .

1. Montrer que pour toute fonction  $f$  mesurable positive

$$\int f d(\alpha\mu + \beta\nu) = \alpha \int f d\mu + \beta \int f d\nu.$$

2. Soit  $\mu = \sum_{i=1}^{\infty} \alpha_i \nu_i$  pour  $\alpha_i \geq 0$  et  $\nu_i$  une mesure sur  $(\Omega, \mathcal{F})$ , montrer que pour  $f \geq 0$

$$\int f d\mu = \sum_{i=1}^{\infty} \alpha_i \int f d\nu_i.$$

3. Montrer que le résultat reste vrai pour  $f$  quelconque dès lors que  $\sum_{i=1}^{\infty} \alpha_i \int |f| d\nu_i < \infty$ .

*Indication : Calculer  $\int f^+ d\mu$  et  $\int f^- d\mu$  par le procédé précédent.*

EXERCICE 3.5.–

Soit  $(\Omega, \mathcal{F})$  un espace mesurable et  $a, a_1, a_2, \dots$  des points de  $\Omega$ . Soit  $f : \Omega \rightarrow \mathbb{R}$  une fonction mesurable.

1. Montrer que  $\int f d\delta_a = f(a)$ .
2. Soit  $(\alpha_i)_{i \in \mathbb{N}^*}$  une suite de coefficients positifs. Soit  $\mu = \sum_i \alpha_i \delta_i$ . Montrer que pour  $f \geq 0$

$$\int f d\mu = \sum_{i=1}^{\infty} \alpha_i f(a_i).$$

3. Montrer que le résultat reste vrai pour  $f$  quelconque dès lors que  $\sum_{i=1}^{\infty} \alpha_i |f(a_i)| < \infty$ .

EXERCICE 3.6. –

Soit  $(\Omega, \mathcal{F}, \nu)$  un espace mesuré. On équipe  $\mathbb{R}$  de sa tribu de Borel. Soit  $f : \Omega \rightarrow \mathbb{R}$  une fonction positive et  $\nu$ -intégrable. On considère l'application  $\mu$  définie sur  $\mathcal{F}$  par :

$$\mu(A) = \int_A f d\nu = \int f \mathbf{1}_A d\nu \quad (3.12)$$

1. Montrer que  $\mu$  est une mesure. A quelle condition sur  $f$  est-ce une mesure de probabilité ?  
On suppose cette condition satisfaite dorénavant.

Supposons qu'il existe une autre fonction  $g$  telle que  $\mu(A) = \int_A g d\nu$  pour tout  $A \in \mathcal{F}$ . On veut montrer que  $f = g$   $\nu$ -p.p.

2. Montrer que pour tout  $A$ ,  $\int (f - g) \mathbf{1}_{\{f > g\}} \mathbf{1}_A d\nu = 0$ .
3. En déduire que  $(f - g)^+$  est égale à zéro  $\nu$ -pp.
4. Conclure.

*On dit qu'une mesure  $\mu$  admet une densité par rapport à  $\nu$  s'il existe une fonction  $f$  telle que  $\mu$  et  $\nu$  satisfont (3.12). Cette fonction  $f$  s'appelle la densité, elle est définie de manière unique à un ensemble négligeable près. On la note  $f = \frac{d\mu}{d\nu}$ .*

EXERCICE 3.7. –

Soit  $\mu$  la mesure de probabilité discrète définie par  $\mu(\{i\}) = \frac{\alpha^i}{i!} e^{-\alpha}$  pour tout  $i \in \mathbb{N}$ .

1. Quel nom porte cette mesure ?
2. Ecrire  $\mu$  comme une somme pondérée de mesures de Dirac.
3. Soit  $\nu = \sum_i \delta_i$  la mesure de comptage sur  $\mathbb{N}$ . Montrer que  $\mu$  admet une densité par rapport à  $\nu$  et fournir cette densité.

## Théorèmes de convergence

EXERCICE 3.8. –

Calculer les limites des intégrales suivantes :

1.  $\int_0^{\infty} (1 + \frac{x}{n})^n e^{-bx} dx$ , pour  $b > 1$  ;
2.  $\int_0^{\infty} \frac{dx}{x^n + e^x}$ .

EXERCICE 3.9. –

Soit  $\{f_n\}_{n \in \mathbb{N}_*}$  la famille de fonctions définies sur  $\mathbb{R}^+$  par

$$f_n(x) = \left(1 - \frac{x}{n}\right)^n \text{ si } 0 \leq x < n, \quad f_n(x) = 0 \text{ si } x \geq n.$$

3. Montrer que pour tout  $x \in \mathbb{R}^+$ ,  $\{f_n(x)\}_n$  est croissante et calculer sa limite  $f(x)$ .
4. Calculer  $\int_{\mathbb{R}^+} f_n(x) dx$  ;  $\lim_{n \rightarrow +\infty} \int_{\mathbb{R}^+} f_n(x) dx$  ;  $\int_{\mathbb{R}^+} f(x) dx$ , et commenter les résultats.

EXERCICE 3.10.–

Pour tout  $n \in \mathbb{N}$ , on considère sur  $\mathbb{R}^+$  l'application  $f_n(x) = 1_{[n, +\infty[}(x)$ .

3. Montrer que pour tout  $x \in \mathbb{R}^+$ ,  $\{f_n(x)\}_n$  est décroissante et calculer sa limite  $f(x)$ .

4. A-t-on  $\lim_{n \rightarrow +\infty} \int_{\mathbb{R}^+} f_n(x) dx = \int_{\mathbb{R}^+} f(x) dx$  ?

EXERCICE 3.11.–

Montrer, en utilisant un changement de variables, que  $\lim_{n \rightarrow \infty} \int_1^\infty n e^{-x^n} dx = \int_1^\infty \frac{e^{-x}}{x} dx$ .

EXERCICE 3.12.–

Démontrer le théorème suivant.

**Théorème 3.28** (Continuité sous le signe somme). Soit  $I$  un ouvert de  $\mathbb{R}^n$  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.

EXERCICE 3.13.–

Démontrer le résultat suivant.

**Théorème 3.29** (Dérivation sous le signe somme). Soit  $I$  un ouvert de  $\mathbb{R}^n$  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

$$\begin{aligned} I &\longrightarrow \mathbb{R} \\ t &\longmapsto \int f(x, t) d\mu(x) \end{aligned}$$

est dérivable sur  $I$  et

$$\frac{d}{dt} \int f(x, t) d\mu = \int \frac{d}{dt} f(x, t) d\mu(x).$$

## Intégrale et fonction primitive

EXERCICE 3.14.– 1. Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction intégrable à valeurs positives. La mesure de Lebesgue sur  $\mathbb{R}$  est notée  $\lambda$ . Soit pour tout  $x \in \mathbb{R}$ ,

$$F(x) = \int f \mathbf{1}_{]-\infty, x]} d\lambda = \int_{-\infty}^x f(t) dt$$

Montrer que si  $f$  est continue en  $x$ , alors  $F$  est dérivable en  $x$  et  $F'(x) = f(x)$ .

2. Soit  $F : \mathbb{R} \rightarrow \mathbb{R}$  une fonction positive de classe  $C_1$ . Montrer que pour tout  $x > a$ ,

$$\int_a^x F'(t) dt = F(x) - F(a).$$

3. Soit  $\mu$  une mesure de fonction de répartition  $F$ . Si  $F$  est de classe  $C_1$ , montrer que  $\mu$  admet pour densité  $F'$  par rapport à  $\lambda$ .

*Le résultat se généralise aisément : Toute mesure  $\mu$  sur la droite réelle dont la fonction de répartition  $F$  est  $C_1$  par morceaux, admet pour densité  $F'$ .*

*Hors programme : On peut démontrer un résultat plus fort, connu sous le nom de théorème fondamental de l'analyse : une fonction  $F$  est "absolument continue" si et seulement si elle est dérivable  $\lambda$ -pp, et pour tout  $x$ ,  $\int_a^x F'(t) dt = F(x) - F(a)$ .*

*La notion se généralise aux mesures sur un espace mesurable quelconque  $(\Omega, \mathcal{F})$ . Une mesure  $\mu$  est dite "absolument continue" par rapport à  $\nu$  si pour tout  $A \in \mathcal{F}$ ,  $\nu(A) = 0$  implique  $\mu(A) = 0$ . Le théorème de Radon-Nykodim établit que  $\mu$  est absolument continue par rapport à  $\nu$  si et seulement si  $\mu$  admet une densité par rapport à  $\nu$ .*

EXERCICE 3.15.–

Soit  $f(x) = ce^{-\alpha x} \mathbf{1}_{\mathbb{R}_+}(x)$  où  $\alpha, c > 0$ . Soit  $\mu$  la mesure de densité  $f$  (par rapport à la mesure de Lebesgue). Déterminer  $c$  en fonction de  $\alpha$  pour que  $\mu$  soit une mesure de probabilité. Calculer la fonction de répartition de  $\mu$  et donner son graphe.

## Histoire

EXERCICE 3.16.–

En 1901, Henri Lebesgue a 25 ans. Emile Borel, dans ses travaux, cherche à attribuer une mesure à des sous-ensembles du segment  $[0, 1]$ , plus généraux que des intervalles, et engendrés à partir de ceux-ci par des unions dénombrables ou des passages au complémentaire (ce que N. Bourbaki nommera plus tard une *tribu*), et demande à cette mesure de vérifier la propriété de  $\sigma$ -additivité. Pourtant, en 1901, l'existence d'une telle mesure n'est pas démontrée. Elle est à l'état de projet. E. Borel estime que la preuve serait "*longue et fastidieuse*".

**Lire la note d'H. Lebesgue (1901) disponible sur le site pédagogique du cours, et répondre aux questions suivantes.**

1. Cette note comporte deux contributions révolutionnaires. La première est une nouvelle définition de l'intégrale : c'est l'objet de la note. Quelle est la seconde ?
2. "Considérons un ensemble de points de  $(a, b)$  ; on peut d'une infinité de manières enfermer ces points dans une infinité dénombrable d'intervalles ; la limite inférieure de la somme des longueurs de ces intervalles est la mesure de l'ensemble." Pour une partie  $E \subset (a, b)$ , notons  $\lambda^*(E)$  la "mesure" de  $E$ , au sens où l'entend H. Lebesgue. Traduire en langage mathématique la phrase précédente, c'est à dire, écrire formellement  $\lambda^*(E)$ .
3. "Un ensemble  $E$  est dit mesurable si sa mesure augmentée de celle de l'ensemble des points ne faisant pas partie de  $E$  donne la mesure de  $(a, b)$ ." Traduire en langage mathématique ce qu'est pour H. Lebesgue un ensemble  $E$  mesurable.
4. "Une infinité d'ensembles mesurables  $E_i$  étant donnée, l'ensemble des points qui font partie de l'un au moins d'entre eux est mesurable." Acceptons cette affirmation d'H. Lebesgue (on ne demande pas de la démontrer). Montrer qu'alors, l'ensemble  $\mathcal{L}$  des parties mesurables au sens d'H. Lebesgue forme une tribu sur  $(a, b)$ .
5. Montrer que  $\mathcal{L}$  contient la tribu de Borel sur  $(a, b)$ , notée  $\mathcal{B}((a, b))$ . On pourra utiliser le fait que  $\mathcal{B}((a, b))$  est la tribu engendrée par les intervalles de  $(a, b)$ .
6. Expliquer la note en bas de page 87 : que dit H. Lebesgue sur le lien entre  $\mathcal{L}$  et les ensembles construits par son ami E. Borel ?
7. "si les  $E_i$  n'ont deux à deux aucun point commun, la mesure de l'ensemble [des points qui font partie de l'un au moins d'entre eux] est la somme des mesures  $E_i$ ". Comment a-t-on appelé cette propriété dans le cours ?
8. On appelle  $\lambda$  la restriction de  $\lambda^*$  à  $\mathcal{L}$ . Que peut-on dire du triplet  $((a, b), \mathcal{L}, \lambda)$  ?
9. Afin de pouvoir définir son intégrale, H. Lebesgue introduit la notion de fonction *sommable* (*n.b.* : plus tard, H. Lebesgue utilisera le terme de fonction "mesurable"). Montrer qu'une fonction  $f : (a, b) \rightarrow [m, M]$  est sommable (au sens d'H. Lebesgue) si et seulement si elle est  $\mathcal{L}/\mathcal{B}([m, M])$ -mesurable (au sens du cours). On pourra s'aider du lemme 2.14 du poly.
10. Voyons comment H. Lebesgue construit son intégrale d'une fonction  $y(x)$  "sommable". Ecrire formellement les ensembles  $E_0$  et  $E_i$  comme des images réciproques faisant intervenir  $y$ .
11. H. Lebesgue introduit les deux sommes :

$$I = m_0 \lambda_0 + \sum_i m_i \lambda_i ; \quad J = m_0 \lambda_0 + \sum_i m_{i-1} \lambda_i,$$

où  $\lambda_i = \lambda(E_i)$  est la mesure définie plus haut par Lebesgue. Montrer que ces deux sommes s'écrivent comme la  $\lambda$ -intégrale (dans le sens du cours) de deux fonctions étagées que l'on écrira.

12. Si  $\int y d\lambda$  représente l'intégrale définie en cours, montrer en utilisant votre cours, que

$$\left| \int y d\lambda - I \right| \leq C \max_i (m_i - m_{i-1}),$$

où  $C$  est une constante que l'on écrira. Montrer une inégalité du même type avec  $J$  au lieu de  $I$ .

13. En déduire que l'intégrale définie dans la note correspond bien à l'intégrale  $\int y d\lambda$  vue en cours.
14. En fin de note, H. Lebesgue décrit plusieurs avantages de la construction proposée, par rapport à celle de Riemann. Citez deux de ces avantages qui vous paraissent importants. Quel est celui qui semble avoir motivé ce travail ?

15. 1) Toute limite de fonction sommable est sommable. La classe des fonctions sommable au sens de Lebesgue est stable pour la convergence simple ! Ce n'est pas le cas de la classe des fonctions intégrables au sens de Riemann.
- 2) Si une fonction  $F$  est dérivable sur  $(a, b)$ , de dérivée  $F'$  bornée, alors  $F'$  est sommable au sens de Lebesgue. En outre,  $F(x)$  s'écrit comme l'intégrale de  $F'$  sur l'intervalle  $(a, x)$ .
16. Inversement, quel théorème d'intégration célèbre ne figure pas dans cette note ?

### Théorème de Fubini

EXERCICE 3.17. –

Soit  $f : [0, 1] \times [1, +\infty) \rightarrow \mathbb{R}$  la fonction définie par  $f(x, y) = e^{-xy} - 2e^{-2xy}$ .

1. Montrer que  $\int_0^1 \int_1^\infty f(x, y) dx dy = \int_0^1 (e^{-x} - e^{-2x}) dx$ .
2. Calculer l'autre intégrale, montrer qu'elle est de signe opposé.
3. Cela effreint-il le théorème de Fubini ?

EXERCICE 3.18. –

Soit la fonction

$$f : [0, a] \times \mathbb{R}^+ \longrightarrow \mathbb{R} \\ (x, y) \longmapsto e^{-xy} \sin x.$$

1. Montrer que  $f$  est intégrable.
2. En utilisant le théorème de Fubini, montrer que

$$\int_0^a \frac{\sin x}{x} dx = \frac{\pi}{2} - \cos a \int_0^\infty \frac{e^{-ay}}{1+y^2} dy - \sin a \int_0^\infty \frac{ye^{-ay}}{1+y^2} dy.$$

3. En déduire que

$$\left| \int_0^a \frac{\sin x}{x} dx - \frac{\pi}{2} \right| \leq \frac{2}{a}.$$

EXERCICE 3.19. –

Soit la fonction

$$f : [0, a] \times \mathbb{R}^+ \longrightarrow \mathbb{R} \\ (x, y) \longmapsto e^{-xy} \sin x.$$

1. Montrer que  $f$  est intégrable.
2. En utilisant le théorème de Fubini, montrer que

$$\int_0^a \frac{\sin x}{x} dx = \frac{\pi}{2} - \cos a \int_0^\infty \frac{e^{-ay}}{1+y^2} dy - \sin a \int_0^\infty \frac{ye^{-ay}}{1+y^2} dy.$$

3. En déduire que

$$\left| \int_0^a \frac{\sin x}{x} dx - \frac{\pi}{2} \right| \leq \frac{2}{a}.$$

## EXERCICE 3.20.–

Soit la fonction

$$f : [0, a] \times \mathbb{R}^+ \longrightarrow \mathbb{R}$$

$$(x, y) \longmapsto e^{-xy} \sin x.$$

1. Montrer que  $f$  est intégrable.
2. En utilisant le théorème de Fubini, montrer que

$$\int_0^a \frac{\sin x}{x} dx = \frac{\pi}{2} - \cos a \int_0^\infty \frac{e^{-ay}}{1+y^2} dy - \sin a \int_0^\infty \frac{ye^{-ay}}{1+y^2} dy.$$

3. En déduire que

$$\left| \int_0^a \frac{\sin x}{x} dx - \frac{\pi}{2} \right| \leq \frac{2}{a}.$$

**Pour aller plus loin****EXERCICE 3.21.–**

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.

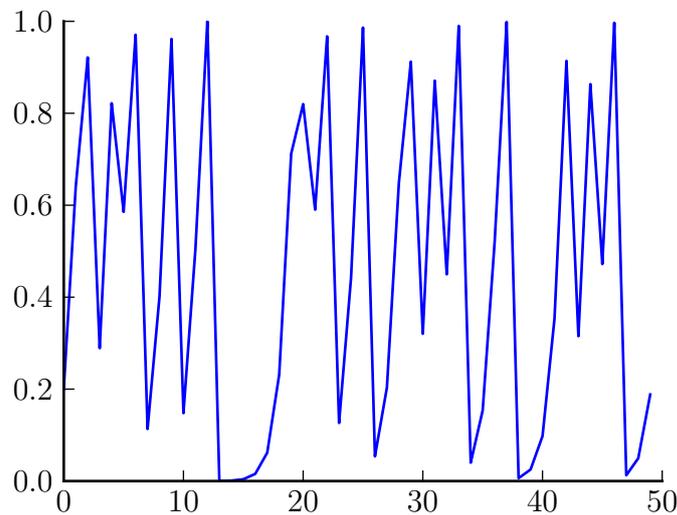


Figure 3.3. – L'orbite de la suite  $x$  pour  $x_0 = 0,2$ .

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.

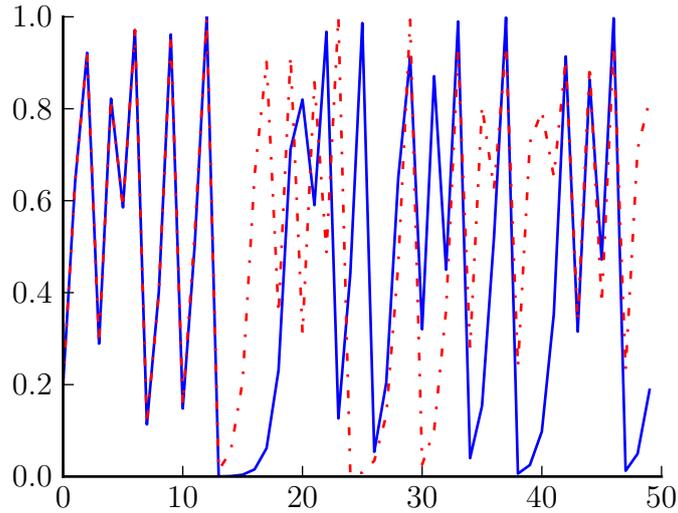


Figure 3.4. – L'orbite de la suite  $x$  pour  $x_0 = 0,2$ , comparée à l'orbite pour  $x_0 = 0,2001$ .

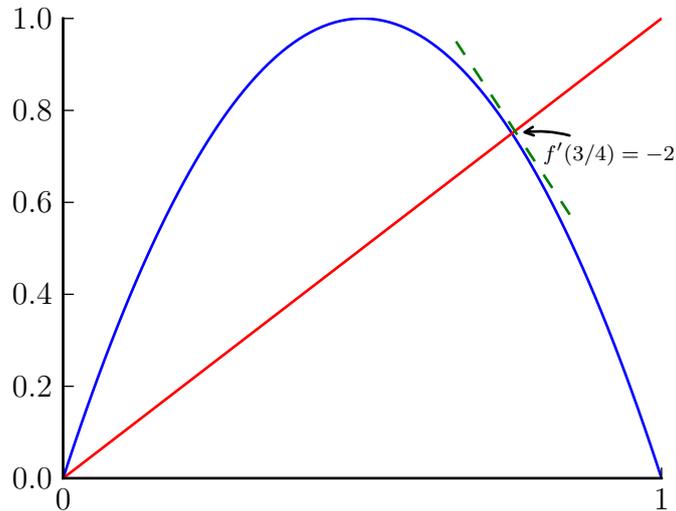


Figure 3.5. – Le point  $(3/4, 3/4)$  est répulsif.

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 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.

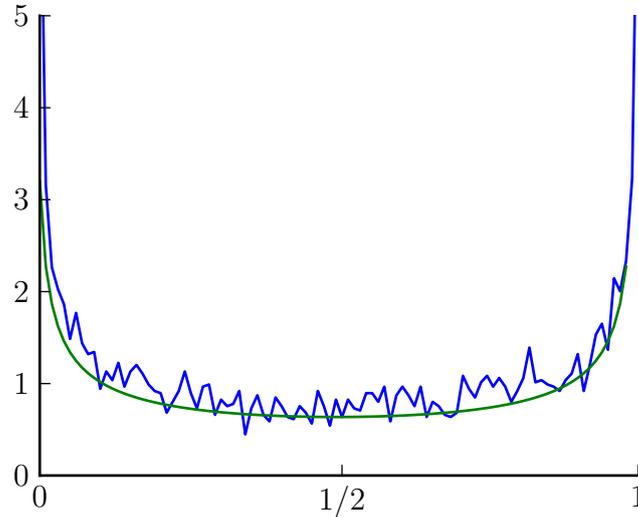


Figure 3.6. —  $\hat{\mu}$  (courbe verte, obtenue pour  $N = 50\,000$ ) et  $\mu$  (courbe rouge, calculée dans la question 3.21.i).

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

$$\mathbf{P}(T^{-1}(A)) = \mathbf{P}(A) \text{ pour tout } A \in \mathcal{E}.$$

- 3.21.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).
- 3.21.b) Soit  $f$  une fonction mesurable de  $E$  dans  $\mathbb{R}$ . 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  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

Le système dynamique  $(E, T, \mathbf{P})$  est dit ergodique lorsque

$$\mathcal{I} \subset \sigma\{A \subset \mathcal{E}, \mathbf{P}(A) = 0 \text{ ou } \mathbf{P}(A) = 1\}.$$

- 3.21.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  $L^2(\mathbf{P})$ ,

$$\lim_{n \rightarrow +\infty} \int_E f \circ T^n g \, d\mathbf{P} = \int_E f \, d\mathbf{P} \int_E g \, d\mathbf{P}. \quad (3.13)$$

- 3.21.d) Montrer que si  $T$  est mélangeante alors  $(E, T)$  est ergodique.
- 3.21.e) Montrer pour  $f \in L^2(\mathbf{P})$ , pour tout entier  $n \geq 1$ ,

$$\int_E f \circ T^n \, d\mathbf{P} = \int_E f \, d\mathbf{P}. \quad (3.14)$$

3.21.f) Montrer que si la condition de mélange (3.13) 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

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^n f \circ T^j(x) = \int_E f d\mathbf{P} \quad (3.15)$$

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  $\lambda$ .

3.21.g) Montrer que  $\lambda$  est invariante par  $T_1$ .

3.21.h) En admettant (ou se souvenant, cf. séries de Fourier) que la famille de fonctions  $e_k(x) = \exp(2i\pi kx)$  pour  $k \in \mathbb{Z}$  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 :

$$\Theta(x) = \sin^2(\pi x/2).$$

3.21.i) Identifier  $\mu$  la mesure image de  $\lambda$  par  $\Theta$ .

3.21.j) Montrer que  $([0, 1[, T, \mu)$  est ergodique et conclure.

## 4. Variables aléatoires

### 4.1. Notations, définitions

#### 4.1.1. Généralités

On se place sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ . Soit  $(E, \mathcal{E})$  un espace mesurable.

On rappelle qu'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.

Lorsque l'ensemble d'arrivée est une partie de  $\mathbb{R}$  ou de  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ , on parle de *variable aléatoire réelle* (en abrégé, v.a.r.). Lorsque l'ensemble d'arrivée est une partie de  $\mathbb{R}^d$  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*.

On appelle *loi de la v.a.  $X$* , et on note  $\mathbf{P}_X$ , la mesure image de  $\mathbb{P}$  par  $X$ , soit

$$\mathbf{P}_X = \mathbb{P}X^{-1}.$$

En clair,  $\mathbf{P}_X(H) = \mathbb{P}(X \in H)$  pour tout  $H \in \mathcal{E}$ , où  $\mathbb{P}(X \in H)$  représente la probabilité de l'ensemble  $\{X \in H\} = X^{-1}(H)$ . Nous avons vu que  $\mathbf{P}_X$  est une mesure de probabilité, mais cette fois sur  $(E, \mathcal{E})$ , et non pas sur  $(\Omega, \mathcal{F})$ .

EXEMPLE 20 (Variables aléatoires discrètes).— Soit  $X : \Omega \rightarrow E$  une variable aléatoire prenant ses valeurs dans un ensemble au plus dénombrable, disons par exemple :  $X(\Omega) =: \{x_1, x_2, x_3, \dots\}$ . Par extension de la terminologie du chapitre 1, nous utiliserons le terme de *variable aléatoire discrète*. La loi de  $X$  est donnée par  $\mathbf{P}_X(H) = \mathbf{P}_X(H \cap \{x_1, x_2, \dots\}) = \sum_n \mathbf{P}_X(H \cap \{x_n\})$ . 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  $\mathbf{P}_X(H \cap \{x_n\}) = \mathbf{P}_X(x_n)\delta_{x_n}(H)$ . Nous obtenons finalement que la loi d'une v.a.d. est donnée par

$$\mathbf{P}_X = \sum_{n=1}^{\infty} \mathbb{P}(X = x_n) \delta_{x_n}.$$

Rappelons que le chapitre 1.4 suffit à étudier ce type de v.a. sans besoin d'avoir recours aux notions nouvelles de théorie de la mesure.

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*.

Nous dirons que  $X$  admet une densité de probabilité  $f_X$  par rapport à une mesure de référence  $\nu$  sur  $(E, \mathcal{E})$  si  $\mathbf{P}_X$  admet une densité  $f_X$  par rapport à  $\nu$ , c'est à dire

$$\mathbb{P}(X \in H) = \int_H f_X(x) d\nu(x) \quad (\forall H \in \mathcal{E}).$$

La densité  $f_X : E \rightarrow \mathbb{R}$  est une fonction mesurable positive, qui satisfait  $\int f_X d\nu = 1$ . Elle est définie de manière unique à un ensemble  $\nu$ -négligeable près. Dans le cas le plus fréquent,  $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . Si on dit que  $X : \Omega \rightarrow \mathbb{R}^d$  admet une densité  $f_X$ , sans préciser la mesure de référence  $\nu$ , il est sous-entendu que cette dernière est la mesure de Lebesgue sur  $\mathbb{R}^d$ . On a alors :

$$\mathbb{P}(X \in H) = \int_H f_X(x) dx \quad (\forall H \in \mathcal{B}(\mathbb{R}^d))$$

en notant  $dx = d\lambda_d(x)$  la mesure de Lebesgue sur  $\mathbb{R}^d$ . Dans ce cas,  $\mathbb{P}(X \in H) = 0$  pour tout ensemble  $H$  négligeable pour la mesure de Lebesgue. En particulier, pour un singleton  $\{x\}$ ,  $\mathbb{P}(X = x) = 0$ . Autrement dit, une variable à densité n'a de « masse » en aucun point.

#### 4.1.2. Variables aléatoires sur $\mathbb{R}^d$

On considère dans ce paragraphe le cas où  $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . On se donne une v.a.  $X : \Omega \rightarrow \mathbb{R}^d$ . 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  $\mathbb{R}^d$ . Nous savons d'après le corollaire 2.15 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.

La loi de  $X$  est aussi appelée la *loi jointe* des v.a.  $X_1, \dots, X_d$ . Les *lois marginales* sont les lois  $\mathbf{P}_{X_1}, \dots, \mathbf{P}_{X_d}$ . On a immédiatement le lien suivant entre la loi marginale de  $X_1$  et la loi jointe  $\mathbf{P}_X$  :

$$\mathbf{P}_{X_1}(H) = \mathbf{P}_X(H \times \mathbb{R} \times \dots \times \mathbb{R}) \quad (\forall H \in \mathcal{B}(\mathbb{R})).$$

Naturellement, une relation similaire est satisfaite pour  $\mathbf{P}_{X_2}, \dots, \mathbf{P}_{X_d}$ .

La *fonction de répartition* d'une v.a.  $X : \Omega \rightarrow \mathbb{R}^d$ , notée  $F_X$ , est par définition la fonction de répartition de sa loi  $\mathbf{P}_X$ . On rappelle que  $F_X(x_1, \dots, x_d) = \mathbf{P}_X(]-\infty, x_1] \times \dots \times ]-\infty, x_d])$  pour tout  $(x_1, \dots, x_d) \in \mathbb{R}^d$ . Ceci revient à l'écriture plus simple :

$$F_X(x_1, \dots, x_d) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d).$$

D'après le corollaire 2.11, deux variables ayant la même fonction de répartition ont la même loi.

Dans le cas particulier où  $X$  admet une densité  $f_X$  par rapport à la mesure de Lebesgue, la fonction de répartition et la densité ont le lien  $F_X(x_1, \dots, x_d) = \int_{]-\infty, x_1] \times \dots \times ]-\infty, x_d]} f_X$ . Par le théorème de Fubini-Tonelli, on peut écrire de manière équivalente :

$$F_X(x_1, \dots, x_d) = \int_{-\infty}^{x_d} \dots \int_{-\infty}^{x_1} f_X(u_1, \dots, u_d) du_1 \dots du_d,$$

où l'ordre dans lequel on intègre les  $u_i$  n'a pas d'importance. Si  $F_X$  est de classe  $C^d$ , on sait alors déduire l'expression de la densité en fonction de  $F_X$  par dérivation :

$$f_X(x_1, \dots, x_d) = \frac{\partial^d F_X(x_1, \dots, x_d)}{\partial x_1 \cdots \partial x_d},$$

où l'ordre de dérivation n'a pas d'importance.

### 4.1.3. Commentaire

En pratique, l'espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$  est rarement 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. Ce que nous manipulons en pratique, ce n'est pas la mesure  $\mathbb{P}$ , c'est la loi de  $X$ . 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  $(\Omega, \mathcal{F}, \mathbb{P})$  tel que  $X$  est une v.a. sur cet espace et tel que  $\mathbf{P}_X$  est la mesure de Bernoulli, c'est-à-dire  $\mathbf{P}_X(\{1\}) = p$ ,  $\mathbf{P}_X(\{0\}) = 1 - p$ . Toutefois, cet énoncé ne précise ni la nature de  $(\Omega, \mathcal{F}, \mathbb{P})$ , 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'a pas d'utilité (en tout cas à l'échelle de ce cours). Si nous tenions malgré tout à le faire, nous aurions maintes possibilités. Dans l'exemple précédent, nous pourrions naturellement poser  $\Omega = \{0, 1\}$ ,  $\mathbb{P}$  la probabilité de Bernoulli et  $X(\omega) = \omega$ . Mais nous pourrions tout aussi bien choisir pour  $\Omega$  l'intervalle  $[0, 1]$ , pour  $\mathbb{P}$  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.

## 4.2. Variables aléatoires réelles

Dans ce paragraphe, on traite le cas de variables aléatoires  $X$  à valeurs dans  $E = \mathbb{R}$  muni de la tribu de Borel.

### 4.2.1. Fonction de répartition

La fonction de répartition d'une v.a.r.  $X$  est donnée pour tout  $x \in \mathbb{R}$  par

$$F_X(x) = \mathbb{P}(X \leq x).$$

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 peut en ajouter une quatrième. Puisque  $F_X$  est croissante bornée, elle admet une limite à gauche en tout point  $x$ . On note  $F_X(x^-) = \lim_{y \uparrow x} F_X(y)$  cette limite. Toujours d'après les propriétés de monotonie des mesures,

$$F_X(x^-) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{n=1}^{\infty} ]-\infty, x - \frac{1}{n}]\right) = \mathbf{P}_X(]-\infty, x[).$$

Par conséquent,  $F_X(x_-) = \mathbb{P}(X < x)$  et donc

$$F_X(x) - F_X(x_-) = \mathbb{P}(X = x).$$

En d'autres termes, si  $F_X$  est continue en  $x$ ,  $\mathbb{P}(X = x) = 0$ . Si  $F_X$  admet une discontinuité en un point  $x$ , alors au contraire la loi de  $X$  possède une masse en ce point, et la valeur de cette masse correspond à l'amplitude du « saut ».

EXEMPLE 21.— La variable aléatoire constante  $X(\omega) = a$  où  $a \in \mathbb{R}$  a pour fonction de répartition un échelon  $F_X(x) = \mathbf{1}_{[a, +\infty)}$ . Plus généralement, une v.a. discrète sur  $\mathbb{R}$  prenant les valeurs  $a_1, a_2, \dots$ , a pour fonction de répartition une fonction en escalier,  $F_X(x) = \sum_n \mathbb{P}(X = a_n) \mathbf{1}_{[a_n, +\infty)}$ .

EXEMPLE 22.— La fonction de répartition d'un v.a.  $X$  de loi uniforme sur  $[0, 1]$  (c'est à dire admettant pour densité  $f_X(x) = \mathbf{1}_{[0, 1]}(x)$ ) est égale à  $F(x) = x$  sur  $[0, 1]$ , zéro sur  $\mathbb{R}_-$ , et 1 sur  $[1, +\infty)$ . Cette fonction est bien continue : une variable à densité n'a de masse en aucun point.

#### 4.2.2. Existence d'une densité

A quelle condition peut-on affirmer qu'une v.a.r.  $X$  dont on connaît la fonction de répartition  $F_X$ , admet une densité par rapport à la mesure de Lebesgue ?

Si  $F_X$  est discontinue, la réponse est non, d'après ce qui précède (une v.a.r. à densité n'a de masse en aucun point, donc sa fonction de répartition est continue).

Si  $F_X$  est de classe  $\mathcal{C}_1$ , la réponse est oui. On sait en effet grâce au théorème fondamental de l'analyse, qu'une fonction de classe  $\mathcal{C}_1$  s'écrit comme l'intégrale de sa dérivée. On peut donc écrire lorsque  $F_X$  est  $\mathcal{C}^1$

$$F_X(x) = \int_{-\infty}^x F'_X(u) du. \quad (4.1)$$

Donc la fonction de répartition de  $X$  coïncide avec la fonction de répartition de la mesure

$$H \mapsto \int_H F'_X,$$

et deux mesures ayant la même fonction de répartition sont égales, on a bien le résultat que  $\mathbf{P}_X(H) = \int_H F'_X$ . Autrement dit,  $X$  admet pour densité  $F'_X$ . Ce raisonnement se généralise facilement au cas où  $F_X$  est continue et  $\mathcal{C}_1$  par morceaux. Nous avons donc la propriété suivante.

**Proposition 4.1.** Soit  $X$  une v.a.r. dont la fonction de répartition  $F_X$  est continue, et de classe  $\mathcal{C}^1$  par morceaux. Alors  $X$  admet pour densité la dérivée presque partout de  $F_X$ , soit

$$f_X(x) = F'_X(x).$$

La condition suffisante ci-dessus est amplement suffisante pour les besoins de ce cours. Elle nous permettra de couvrir tous les cas pratiques rencontrés. Le reste de ce paragraphe ne s'adresse qu'aux élèves curieux d'aller plus loin.

REMARQUE (Hors programme).— Trouver une condition nécessaire et suffisante revient à se demander à quelle condition une fonction  $F_X$  est dérivable presque partout et peut s'écrire comme l'intégrale de sa dérivée (équation (4.1)) ? On remarquera qu'il ne suffit pas qu'une fonction soit dérivable presque-partout pour que (4.1) soit vraie<sup>1</sup>. Pour cela, il faut et il suffit que  $F_X$  soit absolument continue, c'est à dire qu'elle vérifie la propriété suivante : pour tout  $\epsilon > 0$ , il existe  $\delta > 0$  tel que pour toute suite finie d'intervalles disjoints  $[a_1, b_1], \dots, [a_p, b_p]$ ,

$$\sum_{k=1}^p (b_k - a_k) < \delta \Rightarrow \sum_{k=1}^p |F_X(b_k) - F_X(a_k)| < \epsilon .$$

Le résultat suivant, à mettre en regard du théorème 3.20, est prouvé dans [?, Théorèmes 31.7 et 31.8].

**Théorème 4.2.** (Hors programme) 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  $\lambda$ -négligeable  $A \in \mathcal{F}$ .

REMARQUE.— Il existe des v.a.r. qui ne sont ni discrètes, ni à densité par rapport à la mesure de Lebesgue. Considérer par exemple la variable aléatoire  $X = \max(Y, 1)$ , où  $Y$  est une v.a. de loi exponentielle. On pourra calculer la fonction de répartition de  $X$  et vérifier qu'elle n'est pas continue (donc  $X$  n'admet pas de densité) ni en escalier (donc  $X$  n'est pas discrète).

### 4.2.3. Espérance

Au chapitre 1.4, nous avons défini l'espérance  $\mathbb{E}(X)$  d'une v.a.r. discrète  $X \in \{x_1, x_2, \dots\}$  comme le barycentre des  $x_n$  pondérés par la « masse »  $\mathbb{P}(X = x_n)$  :

$$\mathbb{E}(X) = \sum_n x_n \mathbb{P}(X = x_n) . \quad (4.2)$$

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 :

$$\int x f_X(x) dx ,$$

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. Nous fournissons ci-dessous une définition générale et naturelle, qui admet comme cas particuliers les deux exemples ci-dessus.

1. Chercher « escalier du diable » sur internet pour un contre-exemple. On peut même construire une fonction continue, strictement croissante sur  $[0, 1]$ , et dont la dérivée est nulle presque partout.

Terminologie « analyse »		Terminologie « probabilités »	
Fonction mesurable	$f$	Variable aléatoire	$X$
Mesure	$\mu$	Mesure	$\mathbb{P}$
Intégrale	$\int f(x)d\mu(x)$	Espérance	$\mathbb{E}(X) = \int X(\omega)d\mathbb{P}(\omega)$

Table 4.1. – Correspondance des notations et de la terminologie « théorie de la mesure et de l'intégration » vs « théorie des probabilités ».

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilité et  $X : \Omega \rightarrow \overline{\mathbb{R}}$  une variable aléatoire.

**Définition 4.1.** L'espérance de la v.a.  $X$  est définie par :

$$\mathbb{E}(X) := \int X d\mathbb{P} .$$

Il s'agit donc de l'intégrale de  $X$ , vue comme fonction sur  $\Omega$ , par rapport à la mesure de probabilité  $\mathbb{P}$ .

L'espérance est toujours bien définie pour  $X \geq 0$ . Dans le cas général, elle est bien définie lorsque  $\mathbb{E}(X^+)$  ou  $\mathbb{E}(X^-)$  sont finies. La v.a.  $X$  est intégrable lorsque  $\mathbb{E}(|X|) < \infty$ .

REMARQUE.– Si  $A \in \mathcal{F}$ , l'espérance de la v.a.  $\omega \mapsto \mathbf{1}_A(\omega)$  est

$$\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A) .$$

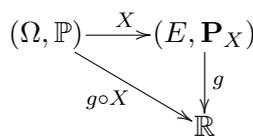
En ce sens, l'espérance est une extension de la notion de probabilité. En particulier, si  $X$  est une v.a. et si  $H$  est un borélien, en posant  $A = \{X \in H\}$ ,

$$\mathbb{E}(\mathbf{1}_{X \in H}) = \mathbb{P}(X \in H) .$$

On retiendra donc que si on sait calculer  $\mathbb{E}(h(X))$  pour n'importe quelle fonction mesurable positive  $h$ , on sait en particulier calculer la loi de  $X$ .

### Théorème de transfert

Dans la pratique, nous avons généralement accès à la loi  $\mathbf{P}_X$  et non à la loi  $\mathbb{P}$  ni à l'expression de  $X(\omega)$  en fonction de  $\omega$  : le théorème de transfert permet d'exprimer  $\mathbb{E}(X)$  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.



Nous rappelons le théorème de transfert vu dans le chapitre précédent, mais réécrit ici avec les notations du probabiliste.

**Théorème 4.3.** Soit  $(E, \mathcal{E})$  un espace mesurable. Soit  $X : \Omega \rightarrow E$  et  $g : E \rightarrow \mathbb{R}$  deux fonctions mesurables telles que  $\mathbb{E}(g(X))$  est définie. Alors,

$$\mathbb{E}(g(X)) = \int g(x) d\mathbf{P}_X(x) .$$

En particulier, notons bien la conséquence suivante :

$$\mathbb{E}(X) = \int x d\mathbf{P}_X(x) .$$

L'espérance est le barycentre de la loi  $\mathbf{P}_X$ .

### Cas des variables discrètes

Il convient de vérifier que la définition de l'espérance (4.1) coïncide bien, dans le cas des v.a.d., avec l'expression (4.2) vue dans le premier chapitre.

**Proposition 4.4.** Soit  $X$  une v.a.d. c'est à dire prenant ses valeurs dans un sous-ensemble de  $\mathbb{R}$  au plus dénombrable  $X(\Omega) = \{x_1, x_2, \dots\}$ . Alors

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} \mathbb{P}(X = x_n) x_n ,$$

dès lors que  $X \geq 0$ , ou que  $\sum_n \mathbb{P}(X = x_n) |x_n| < \infty$ .

*Démonstration.* D'après l'exemple 20, la loi de  $X$  est donnée par  $\mathbf{P}_X = \sum_n \mathbb{P}(X = x_n) \delta_{x_n}$ . En appliquant le théorème de transfert,  $\mathbb{E}(X) = \int x d\mathbf{P}_X(x)$ . On applique enfin le théorème 3.17 qui fournit l'expression de l'intégrale par rapport à une mesure discrète.  $\square$

### Cas des variables à densité

Soit  $X : \Omega \rightarrow \mathbb{R}^d$  une v.a.r. de densité  $f_X$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$ . On rappelle que la loi d'une telle v.a. est donnée par  $\mathbf{P}_X(A) = \int_A f_X(x) dx$ . Soit  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction borélienne.

Le résultat ci-dessous est une réécriture du théorème 3.18 en langage probabiliste.

**Théorème 4.5.** Dès que  $\mathbb{E}(g(X))$  est bien définie, on a :

$$\mathbb{E}(g(X)) = \int g(x) f_X(x) dx . \quad (4.3)$$

En particulier, notons la conséquence suivante pour une v.a.r.  $X$  à densité :

$$\mathbb{E}(X) = \int x f_X(x) dx .$$

L'espérance est le barycentre de la densité.

EXEMPLE 23.– 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

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx.$$

On procède alors à une intégration par parties (en dérivant  $x \mapsto x$  et en intégrant  $f_X$ ) et il vient

$$\int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \left[ -x e^{-\lambda x} \right]_0^{\infty} - \int_0^{\infty} (-e^{-\lambda x}) dx.$$

Le terme tout intégré est nul par « croissance comparée » de la fonction linéaire et de la fonction exponentielle. Il reste

$$\int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda}.$$

#### 4.2.4. Inégalités

**Proposition 4.6** (Inégalité de Markov). Pour tout  $\epsilon > 0$ ,  $p \geq 1$ ,

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

*Démonstration.* Voir paragraphe 1.5.3. □

**Proposition 4.7** (Inégalité de Hölder). Soient  $p, q \geq 0$  tels que  $\frac{1}{p} + \frac{1}{q} = 1$ . Alors,

$$\mathbb{E}(|XY|) \leq (\mathbb{E}(|X|^p))^{\frac{1}{p}} (\mathbb{E}(|Y|^q))^{\frac{1}{q}}.$$

Lorsque  $p = q = 2$ , l'inégalité de Hölder se ramène à l'inégalité de Cauchy-Schwarz :

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

*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  $a = X/\mathbb{E}(X^p)$  et  $b = Y/\mathbb{E}(Y^q)$  et en passant à l'espérance, on tombe bien sur l'inégalité de Hölder après un calcul simple. □

**Proposition 4.8** (Inégalité de Jensen). Soit  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction convexe. Soit  $X$  une v.a.r. telle que  $\mathbb{E}(|X|) < \infty$  et  $\mathbb{E}(|\varphi(X)|) < \infty$ . Alors :

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)).$$

*Démonstration.* Rappelons que toute fonction convexe définie sur  $\mathbb{R}$  est continue et que de plus,  $\forall x \in \mathbb{R}, \exists \alpha, \forall t, \varphi(t) \geq \varphi(x) + \alpha(t - x)$ . 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  $t$ ,  $\varphi(t) \geq \varphi(\mathbb{E}(X)) + \alpha(t - \mathbb{E}(X))$ . 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,  $\mathbb{E}(\varphi(X)) \geq \varphi(\mathbb{E}(X))$ .  $\square$

#### 4.2.5. Moments, variance, covariance

**Définition 4.2.** Soit  $p \geq 0$ . Soit une v.a.r.  $X$  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  $p$* . L'ensemble de telles variables est noté  $\mathcal{L}^p(\mathbb{P})$ . Les propriétés des moments sont identiques à celles vue dans le cas discret. Nous les résumons ici sans preuves.

**Proposition 4.9.** 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. A l'inverse, certaines v.a. n'admettent aucun moment (voir l'exercice 4.16).

**Définition 4.3.** La *variance* d'une v.a.r.  $X$  d'ordre 2 est définie par

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

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

**Définition 4.4.** Soient  $X$  et  $Y$  deux v.a.r. d'ordre 2. Leur *covariance* est définie par :

$$\text{Cov}(X, Y) := \mathbb{E} \left( (X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) \right).$$

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 *décorrélés*.

**Proposition 4.10.** Soient  $X$  et  $Y$  deux v.a.r. d'ordre 2 et  $(\alpha, \beta) \in \mathbb{R}^2$ . On a :

- $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$  ;
- $\text{Cov}(X, X) = \text{Var}(X)$  ;
- $\text{Cov}(Y, X) = \text{Cov}(X, Y)$  ;
- $\text{Var}(\alpha X + \beta) = \alpha^2 \text{Var}(X)$  ;
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$ . En particulier, si  $X$  et  $Y$  sont indépendantes alors  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ .

**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 4.5.** 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  $\mathbb{E}(\|X\|^p) < \infty$  où  $\|\cdot\|$  est une norme sur  $\mathbb{R}^d$ .

**Définition 4.6.** 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 4.11.** 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)  $\mathbb{E}(AX + b) = A\mathbb{E}(X) + b$  ;
- b)  $\text{Cov}(AX + b) = A\text{Cov}(X)A^T$  ;
- c)  $\text{Cov}(X) = \mathbb{E}((X - \mathbf{E}(X))(X - \mathbf{E}(X))^T)$ .

**Proposition 4.12.**  $\text{Cov}(X)$  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  $\mathbb{R}^d$ , on calcule

$$\begin{aligned} x^T \text{Cov}(X)x &= x^T \mathbf{E}((X - \mathbf{E}(X))(X - \mathbf{E}(X))^T)x = \mathbf{E}(x^T (X - \mathbf{E}(X))(X - \mathbf{E}(X))^T x) \\ &= E \left[ (x^T X)^2 \right] \geq 0. \end{aligned}$$

□

## 4.3. Variables aléatoires indépendantes

### 4.3.1. Définition

Si besoin, on relira au préalable la définition 1.5 des événements indépendants (cette définition n'est pas spécifique au cas discret).

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilité. Soient  $(E_1, \mathcal{E}_1), \dots, (E_d, \mathcal{E}_d)$  des espaces mesurables.

**Définition 4.7** (Indépendance d'une famille finie de v.a.). Soit pour tout  $i = 1, \dots, d$ ,  $X_i : \Omega \rightarrow E_i$  une variable aléatoire. Les v.a.  $X_1, \dots, X_d$  sont dites *indépendantes* si les événements  $\{X_1 \in H_1\}, \dots, \{X_d \in H_d\}$  sont indépendants pour tout  $H_1 \in \mathcal{E}_1, \dots, H_d \in \mathcal{E}_d$ .

Ceci revient à écrire :

$$\mathbb{P}(X_1 \in H_1, \dots, X_d \in H_d) = \mathbb{P}(X_1 \in H_1) \times \dots \times \mathbb{P}(X_d \in H_d) \quad (\forall H_1 \in \mathcal{E}_1, \dots, H_d \in \mathcal{E}_d). \quad (4.4)$$

La définition 1.5 de l'indépendance d'événements requiert que la relation (4.4) soit aussi valide si certains événements  $\{X_i \in H_i\}$  sont supprimés. Il est inutile de l'imposer ici : en prenant  $H_i = E_i$ , on a  $\{X_i \in H_i\} = \Omega$ , donc l'événement  $\{X_i \in H_i\}$  est automatiquement supprimé des deux membres de l'équation (4.4).

Le membre de gauche de (4.4) est égal à la loi jointe  $\mathbf{P}_X$  évaluée sur le pavé  $H_1 \times \dots \times H_d$ . 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  $X_i$ . Mais d'après le théorème 3.25, qui coïncide avec  $\prod_i \mathbf{P}_{X_i}(H_i)$  sur les pavés est égale à la mesure produit  $\bigotimes_{i=1}^d \mathbf{P}_{X_i}$ . On a donc le résultat suivant, qui est une simple traduction de la définition précédente :

**Théorème 4.13.** Les propositions suivantes sont équivalentes.

- i)  $X_1, \dots, X_d$  sont indépendantes ;
- ii) La loi jointe est la mesure-produit des lois marginales *i.e.*,

$$\mathbf{P}_X = \bigotimes_{i=1}^d \mathbf{P}_{X_i} \quad \text{où } X = (X_1, \dots, X_d). \quad (4.5)$$

**Définition 4.8** (Indépendance d'une famille quelconque de v.a.). Une famille quelconque 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.

La proposition ci-dessous établit que si on compose par des fonctions des v.a. indépendantes, on conserve des v.a. indépendantes. La preuve, facile, est laissée au lecteur.

**Proposition 4.14.** 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.

EXEMPLE 24.— Soient  $X_1, X_2, X_3$  des v.a.r. indépendantes. Par application immédiate de la définition, les v.a.  $X_1$  et  $(X_2, X_3)$  sont indépendantes, et par composition, les v.a.  $X_1$  et  $X_2 + X_3$  sont indépendantes.

### 4.3.2. Indépendance et espérance

**Théorème 4.15.** Les propositions suivantes sont équivalentes.

- i)  $X_1, \dots, X_d$  sont indépendantes ;
- ii) Pour toutes fonctions mesurables  $h_1 : E_1 \rightarrow \mathbb{R}, \dots, h_d : E_d \rightarrow \mathbb{R}$  telles que les v.a.  $h_i(X_i)$  sont toutes positives ou toutes intégrables,

$$\mathbb{E}(h_1(X_1) \times \dots \times h_d(X_d)) = \mathbb{E}(h_1(X_1)) \times \dots \times \mathbb{E}(h_d(X_d)) . \quad (4.6)$$

*Démonstration.*  $\Rightarrow$  : Par le théorème de transfert, nous avons les deux égalités suivantes :

$$\begin{aligned} \mathbb{E} \left( \prod_i h_i(X_i) \right) &= \int \prod_i h_i(x_i) d\mathbf{P}_X(x_1, \dots, x_d) \\ \prod_i \mathbb{E}(h_i(X_i)) &= \prod_i \int h_i(u) d\mathbf{P}_{X_i}(u) . \end{aligned}$$

Si l'égalité (4.5) est satisfaite, alors les deux membres de droites coïncident, par le théorème de Fubini.

$\Leftarrow$  : Il suffit d'écrire l'égalité pour  $h_i = \mathbf{1}_{H_i}$  ( $i = 1, \dots, d$ ), où  $H_i \in \mathcal{E}_i$ . L'égalité se lit alors  $\mathbb{P}(X_1 \in H_1, \dots, X_d \in H_d) = \prod_i \mathbb{P}(X_i \in H_i)$ , et donc, par définition,  $X_1, \dots, X_d$  sont indépendantes.  $\square$

Du théorème ci-dessus, on retiendra en particulier que *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  $\mathbb{E}(|X_i|) < \infty$  pour tout  $i$ , alors le produit  $X_1 \times \dots \times X_d$  est une v.a. intégrable et on a :

$$\mathbb{E}(X_1 \times \dots \times X_d) = \mathbb{E}(X_1) \times \dots \times \mathbb{E}(X_d) .$$

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 aléatoires réelles indépendantes sont décorréélées*. Attention ! La réciproque est fautive.

### 4.3.3. Indépendance et densités

Comme dans le paragraphe précédent, soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilité, et soient  $X_i : \Omega \rightarrow E_i$  ( $i = 1, \dots, d$ ) une collection de v.a. de  $\Omega$  dans l'espace mesurable  $(E_i, \mathcal{E}_i)$ . On suppose en outre ici que chaque v.a.  $X_i$  admet une densité (par rapport à une mesure de référence arbitraire, qui est dans la plupart des cas pratiques la mesure de Lebesgue).

**Théorème 4.16.** Supposons que pour tout  $i = 1, \dots, d$ ,  $X_i$  admet une densité  $f_{X_i}$  par rapport à une mesure de référence  $\nu_i$  sur  $(E_i, \mathcal{E}_i)$ . Alors les deux propositions suivantes sont équivalentes :

1.  $X_1, \dots, X_d$  sont indépendantes ;
2. Le vecteur  $X = (X_1, \dots, X_d)$  admet une densité par rapport à la mesure  $\bigotimes_{i=1}^d \nu_i$ , donnée pour tout  $(x_1, \dots, x_d) \in E_1 \times \dots \times E_d$  par

$$f_X(x_1, \dots, x_d) = \prod_{i=1}^d f_{X_i}(x_i). \quad (4.7)$$

*Démonstration.* Dire que  $X_1, \dots, X_d$  sont indépendantes, revient à dire que  $\mathbf{P}_X = \bigotimes_{i=1}^d \mathbf{P}_{X_i}$ . Par définition de la mesure produit, ceci est encore équivalent à dire que

$$\mathbf{P}_X(H_1 \times \dots \times H_d) = \prod_{i=1}^d \mathbf{P}_{X_i}(H_i)$$

pour tout  $H_1, \dots, H_d$ , avec  $H_i \in \mathcal{E}_i$ . Par hypothèse,  $\mathbf{P}_{X_i}(H_i) = \int f_{X_i} \mathbf{1}_{H_i} d\nu_i$  pour tout  $i$ , donc, par le théorème de Fubini, l'égalité ci-dessus revient à

$$\mathbf{P}_X(H_1 \times \dots \times H_d) = \int \left( \prod_{i=1}^d f_{X_i}(x_i) \right) \mathbf{1}_{H_1 \times \dots \times H_d}(x) d \bigotimes_{i=1}^d \nu_i(x_1, \dots, x_d).$$

ce qui signifie que la mesure  $\mathbf{P}_X$  coïncide sur les pavés du type  $H_1 \times \dots \times H_d$  avec la mesure de densité  $\prod_{i=1}^d f_{X_i}(x_i)$  par rapport à  $\bigotimes_i \nu_i$ . Comme les pavés du type  $H_1 \times \dots \times H_d$  forment un  $\pi$ -système qui engendre la tribu produit  $\bigotimes_i \mathcal{E}_i$ , cela revient à dire que les deux mesures coïncident sur toute la tribu produit d'après le théorème 2.9. Autrement dit, que  $\mathbf{P}_X$  est bien égale à la mesure de densité (4.8).  $\square$

### 4.3.4. Indépendance et fonctions de répartition

Dans ce paragraphe, on suppose que  $(E_i, \mathcal{E}_i) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , c'est à dire que les  $X_i$  sont des v.a. réelles.

**Théorème 4.17.** Supposons que pour tout  $i = 1, \dots, d$ ,  $X_i : \Omega \rightarrow \mathbb{R}$  est une v.a.r. Alors les deux propositions suivantes sont équivalentes :

1.  $X_1, \dots, X_d$  sont indépendantes :

2. Pour tout  $(x_1, \dots, x_d) \in \mathbb{R}^d$ ,

$$F_X(x_1, \dots, x_d) = \prod_{i=1}^d F_{X_i}(x_i). \quad (4.8)$$

où  $F_{X_i}$  et  $F_X$  représentent les fonctions de répartition de  $X_i$  et de  $X = (X_1, \dots, X_d)$  respectivement.

*Démonstration.* On remarque que  $\mathbf{P}_X$  a pour fonction de répartition  $F_X$  (par définition) et que la loi produit  $\otimes_{i=1}^d \mathbf{P}_{X_i}$  a pour fonction de répartition  $(x_1, \dots, x_d) \mapsto \prod_{i=1}^d F_{X_i}(x_i)$ . Et puisque deux mesures sont égales si et seulement si elles ont la même fonction de répartition, on a directement le résultat :

$$\mathbf{P}_X = \otimes_{i=1}^d \mathbf{P}_{X_i} \Leftrightarrow F_X(x_1, \dots, x_d) = \prod_{i=1}^d F_{X_i}(x_i) \quad (\forall x_1, \dots, x_d).$$

□

## 4.4. Changement de variables

### 4.4.1. Méthode

Soit  $X$  est un vecteur aléatoire sur  $\mathbb{R}^d$  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. L'exercices 4.9 montrent que, dans les cas simples ( $d = 1$ ), la réponse est immédiatement donnée par calcul et différentiation de la fonction de répartition de  $Y$  (inutile d'aller chercher plus loin).

Dans les cas plus complexes, l'expression de la densité de  $Y$  est obtenue grâce à la *formule du changement de variable* (théorème 3.22). La méthode est la suivante.

On fait les hypothèses suivantes :

- $X$  admet une densité  $f_X$ .
- $X(\Omega) \subset U \subset \mathbb{R}^d$  où  $U$  est un ouvert ;
- $\phi : U \rightarrow V$  est un difféomorphisme.

Soit  $h$  une fonction arbitraire, positive, définie sur  $\mathbb{R}^d$ . On évalue l'espérance :

$$\begin{aligned} \mathbb{E}(h(Y)) &= \mathbb{E}((h \circ \phi)(X)) \\ &= \int_U (h \circ \phi) f_X \\ &= \int_U (h \times (f_X \circ \phi^{-1})) \circ \phi \\ &= \int_V h \frac{f_X \circ \phi^{-1}}{|(\det J_\phi) \circ \phi^{-1}|} \end{aligned}$$

où la dernière égalité provient de la formule du changement de variable. Ainsi, on peut écrire  $\mathbb{E}(h(Y)) = \int h f_Y$  où :

$$f_Y = \frac{f_X \circ \phi^{-1}}{|(\det J_\phi) \circ \phi^{-1}|} \mathbf{1}_V. \quad (4.9)$$

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  $H \in \mathcal{B}(\mathbb{R})$ . L'égalité  $\mathbb{E}(h(Y)) = \int h f_Y$  se lit alors :

$$\mathbb{P}(Y \in H) = \int_H f_Y.$$

On en conclut que  $Y$  est un vecteur aléatoire de densité  $f_Y$  donnée par (4.9).

Notons que la méthode de calcul est plus importante à retenir que le résultat final. Il s'agira surtout de s'exercer à la mettre en œuvre au travers des exercices.

**Corollaire 4.18** (Changement de variable par transformation affine). Soit  $X$  un vecteur aléatoire sur  $\mathbb{R}^d$  de densité  $f_X$  par rapport à  $\lambda_d$ . Soient  $A$  une matrice inversible  $d \times d$  et  $b$  un vecteur de  $\mathbb{R}^d$  (*n.b.* :  $A$  et  $b$  sont déterministes). On pose  $Y = AX + b$ . Si  $A$  est inversible, alors  $Y$  admet une densité  $f_Y$  par rapport à  $\lambda_d$ , donnée pour tout  $y \in \mathbb{R}^d$  par

$$f_Y(y) = \frac{1}{|\det(A)|} f_X(A^{-1}(y - b)).$$

*Démonstration.* Si  $A$  est inversible, l'application  $\phi : x \mapsto Ax + b$  est un difféomorphisme de  $\mathbb{R}^d$  dans  $\mathbb{R}^d$ . L'inverse est  $\phi^{-1}(y) = A^{-1}(y - b)$  et la matrice jacobienne est  $A$ . Il suffit alors d'appliquer la formule (4.9).  $\square$

#### 4.4.2. Exemple

Soit  $(X_1, X_2)$  deux variables aléatoires réelles indépendantes, de même loi

$$d\mathbb{P}(x) = \mathbf{1}_{[1, \infty[}(x) \frac{1}{x^2} dx.$$

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  $h : \mathbb{R}^2 \rightarrow \mathbb{R}^+$  telle que pour toute fonction  $f$  continue bornée de  $\mathbb{R}^2$  dans  $\mathbb{R}$ , on ait

$$\mathbb{E}(f(U, V)) = \int_{\mathbb{R}^2} f(x, y) h(x, y) dx dy.$$

Posons

$$\begin{aligned} \phi : \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ (x, y) &\mapsto (xy, x/y). \end{aligned}$$

On a

$$\mathbb{E}(f(U, V)) = \mathbb{E}((f \circ \phi)(X, Y)) = \int (f \circ \phi)(x, y) d\mathbf{P}_{X, Y}(x, y),$$

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

$$\mathbb{E}(f(U, V)) = \int_{[1, +\infty[^2} (f \circ \phi)(x, y) \frac{1}{x^2 y^2} dx dy.$$

Rappelons nous que nous voulons aboutir à une identité de la forme

$$\mathbb{E}(f(U, V)) = \int_{\mathbb{R}^2} f(x, y) h(x, y) dx dy.$$

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

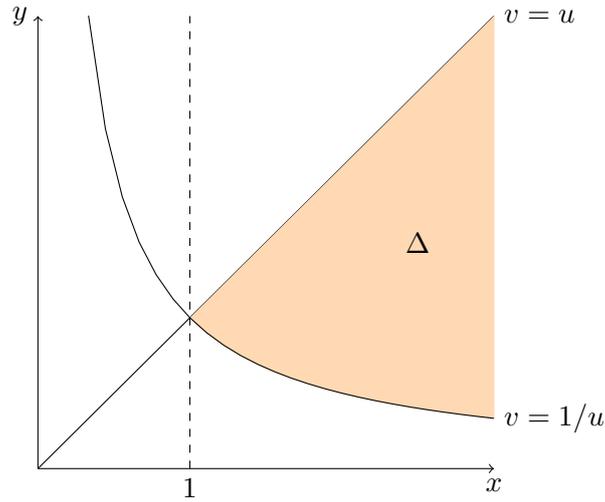
$$\Delta = \{(u, v), u \geq v \geq 0 \text{ et } uv \geq 1\}.$$

On tire du théorème de changement de variables que

$$\begin{aligned} \int_{[1, +\infty[^2} (f \circ \phi)(x, y) \frac{1}{x^2 y^2} dx dy &= \int_{\Delta} f(u, v) \frac{1}{uv \cdot u/v} \left| \frac{1}{-2v} \right| dudv \\ &= \int_{\Delta} f(u, v) \frac{1}{2u^2 v} dudv, \end{aligned}$$

d'où par identification,

$$d\mathbf{P}_{(U, V)}(u, v) = \frac{1}{2u^2 v} \mathbf{1}_{\Delta}(u, v) dudv.$$

Figure 4.1. – Le domaine  $\Delta$  dans le plan  $(u, v)$ .

Pour calculer la loi de  $U$ , on veut exprimer  $\mathbb{E}(f(U))$  pour toute fonction continue bornée de  $\mathbb{R}$  dans  $\mathbb{R}$ . On remarque alors que l'application  $\tilde{f}(x, y) = f(x)$  est continue bornée de  $\mathbb{R}^2$  dans  $\mathbb{R}$  donc

$$\begin{aligned}\mathbb{E}(f(U)) &= \mathbb{E}(\tilde{f}(U, V)) = \int_{\Delta} f(u) \frac{1}{2u^2v} du dv \\ &= \int_1^{+\infty} f(u) \frac{1}{2u^2} \left( \int_{1/u}^u \frac{1}{v} dv \right) du,\end{aligned}$$

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,

$$\begin{aligned}\mathbb{E}(f(V)) &= \int_{\Delta} f(v) \frac{1}{2u^2v} du dv \\ &= \int_0^{+\infty} f(v) \frac{1}{2v} \left( \int_{\Delta_v} \frac{1}{u^2} du \right) dv,\end{aligned}$$

où

$$\begin{aligned}\Delta_v &= \{u \mid (u, v) \in \Delta\} \\ &= \begin{cases} [1/v, +\infty[ & \text{si } 0 \leq v \leq 1 \\ [v, +\infty[ & \text{si } v \geq 1. \end{cases}\end{aligned}$$

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.

## 4.5. Exercices

### 4.5.1. Fonction de répartition, densité

#### EXERCICE 4.1.–

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 4.2.–

Soit  $A$  une v.a. exponentielle de paramètre  $\alpha > 0$ . Déterminer la probabilité pour que le polynôme  $X^2 + 2X(A - 2) + 4A - 2$  ait toutes ses racines réelles.

#### EXERCICE 4.3.–

Soit  $X$  une v.a. exponentielle de paramètre  $\alpha > 0$ . Soit  $a$  un réel strictement positif. Soit  $Y = \min(X, a)$

1. Tracer la fonction de répartition  $F_Y$  de  $Y$ .
2. Ecrire la fonction de répartition sous la forme  $F_Y = F_Y^c + p\mathbf{1}_{[a, +\infty[}$  où  $F_Y^c$  est de classe  $C_1$  par morceaux et où  $p$  est une masse à déterminer.
3. En déduire que la loi de  $Y$  s'écrit comme somme d'une mesure de Dirac et d'une mesure à densité par rapport à la mesure de Lebesgue. Donner la densité en question.

#### EXERCICE 4.4.–

Soit  $X$  une v.a.r. de fonction de répartition  $F_X$ .

1. En fonction de  $F_X$ , exprimer les probabilités  $\mathbb{P}(a < X \leq b)$ ,  $\mathbb{P}(a < X < b)$ ,  $\mathbb{P}(a \leq X \leq b)$ .
2. Exprimer  $F_X$  lorsque  $X$  suit une loi uniforme sur l'intervalle  $[a, b]$ .

#### EXERCICE 4.5.–

Soient  $a$  et  $b$  deux réels et  $X$  une v.a. à valeur dans  $\mathbb{R}$ . Déterminer la fonction de répartition de  $Y = aX + b$  en fonction de celle de  $X$ . Dans le cas où  $X$  admet une densité par rapport à la mesure de Lebesgue, déterminer (si elle existe) la densité de  $Y = aX + b$ .

#### EXERCICE 4.6.–

Soit  $F$  une fonction de répartition sur  $\mathbb{R}$  supposée continue et strictement croissante. Soit  $U$  de loi uniforme  $U$  sur  $[0, 1]$ .

1. Rappeler quelle est la fonction de répartition  $F_U$  de la variable  $U$ .
2. On pose  $X = F^{-1}(U)$ . Calculer la fonction de répartition de  $X$ .
3. Quel intérêt ce résultat présente-t-il ?

Ce résultat peut être généralisé sans difficulté au cas où  $F_X$  est strictement croissante sur  $]a, b[$  où  $a < b$  appartiennent à  $[-\infty, +\infty]$ , avec  $F(a) = 0$  et  $F(b) = 1$ . Il peut même être généralisé à toute fonction de répartition, à condition de remplacer  $F_X^{-1}$  par l'*inverse généralisée*  $F_X^{-1}(t) = \inf\{y \in \mathbb{R} : F_X(y) \leq t\}$ .

- (4) Proposer un algorithme pour obtenir des réalisations d'une v.a. exponentielle de paramètre  $\lambda > 0$ .

**EXERCICE 4.7.**–

Soit  $X$  une v.a. de loi normale  $\mathcal{N}(m, \sigma^2)$  avec  $\sigma^2 > 0$ , c'est à dire que  $X$  admet pour densité :

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

Donner la densité de  $Y = \exp(X)$ .

**EXERCICE 4.8.**–

Soit  $X$  une v.a. de Cauchy sur  $\mathbb{R}$  de densité donnée par  $f(x) = c(1+x^2)^{-1}\mathbf{1}_{\mathbb{R}}(x)$ . Que vaut  $c$ ? Quelle est la loi de  $Y = 1/X$ ?

**EXERCICE 4.9.**–

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 4.10.**–

Soit  $X$  une v.a. de loi exponentielle de paramètre  $\lambda > 0$ .

1. Quelle est la loi de  $X^2$ ?
2. Quelle est la loi de  $Y$  définie par

$$Y = \begin{cases} X & \text{si } X \leq 1, \\ 2X & \text{si } X > 1. \end{cases}$$

**EXERCICE 4.11.**–

Soit  $X$  une v.a. de fonction de répartition  $F_X$ , telle que  $X$  et  $2X$  ont même fonction de répartition. Montrer que pour tout  $n \geq 1$ , et pour tout  $x$ ,  $F_X(2^n x) = F_X(x)$ . En déduire la loi de  $X$ .

**4.5.2. Espérance****EXERCICE 4.12.**–

Pour un patient atteint d'une certaine pathologie, la durée de vie après diagnostic est modélisée par une v.a.  $X$  de loi exponentielle de paramètre  $\lambda$ . Quelle est l'espérance de vie après diagnostic? Donner l'écart-type et la médiane de survie.

**EXERCICE 4.13.**–

Donner, si elles sont définies, les espérances de v.a. dont les densités (par rapport à la mesure de Lebesgue) sont les suivantes :

- Cauchy :  $f(x) \propto 1/(1+x^2)$
- Gaussienne :  $f(x) = e^{-x^2/2}/\sqrt{2\pi}$
- Uniforme :  $f(x) \propto \mathbf{1}_{[a,b]}$

**EXERCICE 4.14.**—

Un charcutier fabrique chaque jour  $x$  kilos de choucroute : il la vend  $p$  euros le kilo, pour un coût de fabrication de  $c < p$  euros le kilo. La demande varie d'un jour à l'autre : on la modélise par une variable positive  $X$  à densité. On note  $F_X$  la *fonction de répartition* de  $X$ . S'il lui reste de la choucroute invendue en fin de journée, il l'offre gratuitement à une association. On note  $P$  le profit réalisé en une journée par la vente de choucroute (*n.b.* : il peut arriver que  $P(\omega) < 0$ ).

1. Exprimer  $P$  en fonction de  $x$ ,  $c$ ,  $p$  et  $X$ .
2. Montrer que

$$\mathbb{E}(P) = \alpha x - \beta \int_0^x F_X(t) dt$$

où  $\alpha$  et  $\beta$  sont deux constantes que l'on exprimera en fonction de  $p$  et  $c$ .

3. Quelle quantité doit produire le charcutier pour maximiser l'espérance de son profit ?
4. Trouver un nombre  $a$  tel que le profit est plus grand que  $a$  avec probabilité 0,95. On exprimera  $a$  en fonction de  $c$ ,  $p$  et  $F_X$ .
5. Application : Exprimer  $a$  lorsque  $X$  suit une loi exponentielle de paramètre  $\lambda > 0$  et lorsque le charcutier produit la quantité optimale  $x$  prescrite à la question 2.

**EXERCICE 4.15 (Caractérisation de la loi par des fonctions-test).**—

Soient  $X, Y$  deux v.a. réelles telles que  $\mathbb{E}(h(X)) = \mathbb{E}(h(Y))$  pour toute  $h$  continue bornée. Montrer que  $X$  et  $Y$  ont la même loi.

- EXERCICE 4.16.**—
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 3.1). Montrer que  $\mathbf{E}(|X|)$  diverge. Plus généralement, montrer que  $X$  ne possède aucun moment.

**4.5.3. Vecteurs aléatoires et indépendance****EXERCICE 4.17.**—

Soient  $X_1, \dots, X_d$  des v.a. i.i.d. de loi  $\mathcal{N}(0, 1)$ .

1. Quelle est la densité 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,  $\sigma_i^2 > 0$ .

**EXERCICE 4.18.**—

On considère le contexte de l'exercice 4.12. Un traitement permet d'allonger significativement la durée de vie et les trois quart des patients répondent positivement au traitement. On modélise la durée de vie comme une variable aléatoire  $Z$  définie par

$$Z = YB + X(1 - B)$$

où  $Y$  suit une loi exponentielle de paramètre  $\lambda/2$  et où  $B$  est une variable de Bernoulli dont on précisera le paramètre. Les variables  $X, Y, B$  sont supposées indépendantes. Calculer l'espérance de vie.

**EXERCICE 4.19.**—

Soit  $(X, Y)$  un couple de v.a.r. dont la loi jointe est la loi uniforme sur le disque unité  $\{(x, y), x^2 + y^2 \leq 1\}$  : la densité est donnée par

$$f(x, y) = \alpha \mathbf{1}_{\mathcal{C}}(x, y) \quad \mathcal{C} := \{(x, y), x^2 + y^2 \leq 1\}.$$

1. Que vaut la constante  $\alpha$  ?
2. Déterminer la loi marginale de  $X$ . Sans calculs, quelle est la loi de  $Y$  ?
3. Les v.a.  $X$  et  $Y$  sont-elles indépendantes ?
4. Calculer la loi du couple  $(\min(X, Y), \max(X, Y))$ .

**EXERCICE 4.20.**—

Un jeu se déroule de la manière suivante. On fait une expérience aléatoire de probabilité de succès  $p$ . Si le résultat est positif, le gain du joueur est choisi aléatoirement selon une loi de Poisson de paramètre  $\lambda > 0$ . Sinon, le gain est une variable exponentielle de paramètre  $\alpha > 0$ .

1. Montrer que l'on peut décrire le gain  $Z$  comme une variable aléatoire répondant à l'expression

$$Z = UX + (1 - U)Y$$

où  $U$  est une variable de Bernoulli. Préciser les lois de  $X$  et  $Y$ .

2. On suppose que les v.a.  $U, X, Y$  sont indépendantes. Calculer  $\mathbb{E}(h(Z))$  pour n'importe quelle fonction  $h$  mesurable bornée.
3. Exprimer la loi de  $Z$  sous la forme de la somme d'une mesure discrète et d'une mesure à densité par rapport à la mesure de Lebesgue.
4. Montrer que  $Z$  admet une densité  $f$  par rapport à une mesure de référence  $\mu$  que l'on exprimera.

**EXERCICE 4.21.**—

Le cycle d'un feu de circulation est le suivant : le feu est vert sur l'intervalle  $[0, v]$  et orange ou rouge sur  $[v, v + r]$  ( $v, r > 0$ ). L'instant d'arrivée  $U$  d'un automobiliste est supposé uniformément réparti sur le cycle  $[0, r + v]$ .

1. Exprimer en fonction de  $U$  le temps d'attente  $T$  de cet automobiliste au feu dans le cas où aucun automobiliste ne se trouve devant le feu à l'instant où il arrive.
2. Déterminer la probabilité  $\mathbb{P}(T = 0)$  pour que l'automobiliste passe immédiatement.
3. Montrer que pour toute fonction  $h : \mathbb{R} \rightarrow \mathbb{R}$  bornée,

$$\mathbb{E}(h(T)) = \frac{v}{v+r} h(0) + \frac{1}{v+r} \int_0^r h(u) du.$$

4. Conclure que  $T$  n'est ni une v.a. à densité sur  $\mathbb{R}$ , ni une v.a. discrète : on écrira sa loi sous la forme  $\sum_{i=1}^K f_i d\mu_i$  pour des densités  $f_i$  et des mesures  $\mu_i$  à préciser.

## EXERCICE 4.22.–

On considère un système constitué de  $n$  composants. On suppose que les durées de vie des composants sont des v.a. exponentielles  $T_1, \dots, T_n$  de paramètres respectifs  $\lambda_1, \dots, \lambda_n$  (strictement positifs) ; et que ces v.a. sont indépendantes.

1. On suppose que le système est en parallèle : il fonctionne lorsqu'au moins un des composants fonctionne. Exprimer la durée de vie  $T$  du système en fonction de v.a.  $T_i, i \leq n$ . Déterminer la fonction de répartition de  $T$ .
2. On suppose que le système est en série : il fonctionne lorsque tous les composants fonctionnent. Déterminer la fonction de répartition de  $T$  et reconnaître sa loi.

## EXERCICE 4.23.–

Une cerise est placée sur la circonférence d'un gâteau rond que l'on partage en deux au hasard en pratiquant deux découpes suivant des rayons. Si on prend la position de la cerise comme origine des angles, les positions  $U$  et  $V$  des deux coups de couteau sont des variables uniformément réparties sur  $[0, 2\pi]$  et indépendantes.

1. Exprimer la taille  $T$  de la part contenant la cerise en fonction de  $U$  et de  $V$ .
2. Calculer son espérance et la probabilité qu'elle soit plus grosse que l'autre.
3. Quelle doit être la position d'un gourmand qui doit choisir entre la part avec la cerise et la part sans la cerise, avant le découpage ?

## EXERCICE 4.24.–

Soient  $U$  et  $V$  deux v.a. de loi uniforme sur l'intervalle  $[0, 1]$ .

1. Ecrire la densité jointe de  $U$  et  $V$ .
2. On pose  $X = \min(U, V)$ . Donner la fonction de répartition et la densité de  $X$ .
3. On pose  $Y = \max(U, V)$ . Ecrire la fonction de répartition du couple  $(X, Y)$ .
4. Donner la densité du couple  $(X, Y)$ .

## EXERCICE 4.25.–

Soient  $X, Y$  deux v.a. indépendantes de loi exponentielle de paramètre 1. Exprimer la fonction de répartition, puis la densité, de la v.a.  $X/Y$ .

## EXERCICE 4.26.–

**(Méthode du rejet).** Soient  $f$  et  $g$  deux densités sur  $\mathbb{R}$  et  $c$  un réel tel que  $f(x) \leq cg(x)$  pour tout  $x \in \mathbb{R}$ . On suppose que l'on sait faire des tirages de v.a. selon la densité  $g$  : on souhaite en déduire une méthode permettant d'effectuer des tirages selon la densité  $f$ . On adopte la convention  $0/0 = 1$ .

Soit  $U$  une v.a. de loi uniforme sur  $[0, 1]$  et  $Y$  une v.a. de densité  $g$ .  $U, Y$  sont indépendantes. Soit aussi  $X$  de densité  $f$ .

1. Montrer que  $c \geq 1$ .
2. Calculer la probabilité de l'événement  $U \leq \frac{f(Y)}{cg(Y)}$ .
3. De même, calculer la probabilité

$$\mathbb{P} \left[ U \leq \frac{f(Y)}{cg(Y)}, Y \leq t \right].$$

Soit  $U_1, U_2, \dots$  une suite de v.a. uniformes sur  $[0, 1]$  et  $Y_1, Y_2, \dots$  une suite de v.a. i.i.d. de densité  $g$ . Les variables  $\{U_k, Y_k\}_{k \geq 1}$  sont indépendantes. On note :

$$B_k = \begin{cases} 1 & \text{si } U_k \leq f(Y_k)/(cg(Y_k)) \\ 0 & \text{sinon} \end{cases}$$

4. Quelle est la loi de  $B_k$  ?
5. On pose  $T = \inf\{k \geq 1 : B_k = 1\}$ . Quelle est la loi de  $T$  ?
6. On pose  $X' := Y_T$ . Montrer que  $X'$  est de densité  $f$ .
7. Ecrire en pseudo-code un algorithme permettant de générer une variable de densité  $f$ , en supposant que l'on sait par ailleurs générer des variables selon la loi  $g$  et selon la loi uniforme.
8. Sachant qu'un ordinateur met un temps  $\tau$  pour générer un couple  $(U, Y)$  de v.a. selon la loi uniforme et la loi  $g$ , calculer l'espérance du temps mis pour générer une variable de densité  $f$  en utilisant l'algorithme précédent (on négligera le temps des opérations usuelles : multiplication, division, test, etc.).

EXERCICE 4.27 (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 4.28.—

Soit  $X$  une v.a. de loi  $\mathcal{N}(0, 1)$ . On pose

$$Y = \begin{cases} X & |X| > a, \\ -X & |X| \leq a, \end{cases} \quad \text{i.e. } Y = X\mathbf{1}_{|X|>a} - X\mathbf{1}_{|X|\leq a}$$

où  $a$  est tel que  $\int_0^a x^2 \exp(-0.5x^2) dx = \sqrt{2\pi}/4$ . Montrer que  $\text{Cov}(X, Y) = 0$  mais que  $X$  et  $Y$  ne sont pas indépendantes.

EXERCICE 4.29.—

**Matrice de covariance.** Soient  $X_1, \dots, X_d$  des v.a. réelles d'ordre deux. L'espérance du vecteur-colonne  $X = (X_1, \dots, X_d)^T$  est définie comme le vecteur des espérances :

$$\mathbb{E}(X) := \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_d) \end{pmatrix}.$$

La matrice de covariance de  $X$ , notée  $\text{Cov}(X)$ , a pour coefficient  $(i, j)$  la covariance  $\text{Cov}(X_i, X_j)$ .

1. Soit  $A$  une matrice et  $b$  un vecteur déterministes. Montrer que  $\mathbb{E}(AX + b) = A\mathbb{E}(X) + b$ .
2. Que vaut le  $i$ ème coefficient de la diagonale de  $\text{Cov}(X)$  ?
3. Quelle est la structure de  $\text{Cov}(X)$  lorsque les variables  $X_i$  sont décorrélées ?

4. Pour quelle raison la matrice suivante n'est-elle PAS une matrice de covariance :

$$\begin{bmatrix} 6 & 2 \\ 3 & 4 \end{bmatrix}$$

5. Montrer que  $\text{Cov}(X) = \mathbb{E}(X_c X_c^T)$  où  $X_c := X - \mathbb{E}(X)$  est le vecteur recentré.
6. Montrer que  $\text{Cov}(AX + b) = A\text{Cov}(X)A^T$ .
7. Montrer que  $\text{Cov}(X)$  est une matrice symétrique semi-définie positive, c'est à dire que pour tout vecteur  $x$ ,  $x^T \text{Cov}(X)x \geq 0$ .
8. Soient  $X$  et  $Y$  deux vecteurs aléatoires d'ordre deux, supposé indépendants. Montrer que  $\text{Cov}(X + Y) = \text{Cov}(X) + \text{Cov}(Y)$ .

**EXERCICE 4.30.**— 1. Soient  $X$  et  $Y$  deux variables aléatoires réelles indépendantes admettant chacune une densité (notées respectivement  $f_X$  et  $f_Y$ ). Prouver que  $X + Y$  est aussi une variable aléatoire à densité et donner sa densité.

2. Ce résultat est-il encore vrai en général si  $X$  et  $Y$  ne sont plus supposées indépendantes ?

3. On suppose que  $X$  et  $Y$  sont indépendantes et de loi normale  $\mathcal{N}(0, \sigma_X^2)$  et  $\mathcal{N}(0, \sigma_Y^2)$ , avec  $\sigma_X$  et  $\sigma_Y$  positifs.  $Z = X + Y$  est-elle une variable aléatoire à densité ? Si oui, donner sa densité<sup>2</sup>.

#### 4.5.4. Changement de variable

**EXERCICE 4.31.**—

Soit  $X, Y$  deux v.a. indépendantes, chacune de même loi uniforme sur  $[0, 1]$ . On pose

$$U := \sqrt{-2 \log X} \cos(2\pi Y), \quad V := \sqrt{-2 \log X} \sin(2\pi Y)$$

1. Quelle est la loi du couple  $(X, Y)$  ?
2. Quelle est la loi du couple  $(U, V)$  ?
3. Les v.a.  $U$  et  $V$  sont-elles indépendantes ?

**EXERCICE 4.32 (Loi Gamma et loi Bêta).**—

Soient  $\lambda, a$  deux réels strictement positifs. On dit qu'une variable aléatoire suit la loi Gamma  $\Gamma(\lambda, a)$  lorsqu'elle admet la densité :

$$\forall x \in \mathbb{R}, \quad \frac{\lambda^a x^{a-1} e^{-\lambda x}}{\Gamma(a)} \mathbf{1}_{\mathbb{R}_+^*}(x).$$

On se donne  $X$  et  $Y$  deux variables aléatoires indépendantes, suivant respectivement la loi  $\Gamma(\lambda, a)$  et  $\Gamma(\lambda, b)$ , avec  $a, b > 0$ .

---

2. On utilisera la relation suivante (très utile dès que l'on manipule des lois normales, donc à retenir!), valable dès que  $a \neq 0$ , avec bien sûr  $\Delta = b^2 - 4ac$  :

$$ax^2 + bx + c = a \left( x + \frac{b}{2a} \right)^2 - \frac{\Delta}{4a}$$

1. Quelle loi classique de votre cours est un cas particulier de la loi Gamma ?
2. Soit  $Z \hookrightarrow \mathcal{N}(0, 1)$ . Quelle est la loi<sup>3</sup> de  $Z^2$  ? (à exprimer en termes de loi Gamma !)
3. Calculer la loi de  $(U, V) = \left(X + Y, \frac{X}{X+Y}\right)$ , en n'oubliant pas de calculer la constante de normalisation de la densité  $f_U$  de  $U$ .
4. En déduire que :

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

$$\text{où } B(a, b) := \int_0^1 t^{a-1}(1-t)^{b-1} dt.$$

**EXERCICE 4.33.**–

Soient  $X_1, X_2$  deux v.a. gaussiennes centrées réduite  $\mathcal{N}(0, 1)$ , indépendantes. On pose

$$U := \frac{X_1^2}{X_1^2 + X_2^2} \quad V := X_1^2 + X_2^2.$$

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  $Z = X_2^2/X_1^2$ . Exprimer  $Z$  en fonction de  $U$  puis calculer la densité de la loi de  $Z$ .

**4.5.5. Pour s'entraîner****EXERCICE 4.34 (\*)**.–

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 4.35 (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 4.36.**–

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  $0 \leq t + s \leq 3$ .

3. À toute fin utile, on rappelle que  $\Gamma(1/2) = \sqrt{\pi}$ .

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 4.37. –

Soit  $(X_1, X_2)$ , une variable aléatoire à valeurs dans  $\mathbb{R}^2$  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 4.38. –

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 4.39. –

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  $n = 100$ .

#### 4.5.6. Pour aller plus loin

EXERCICE 4.40 (\*\*\*) –

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 4.41 (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 4.40) associée à  $U$ , i.e.,

$$W_i = U_{(i)}, \text{ pour tout } i = 1, \dots, n.$$

On pose

$$V_1 = 1 + W_1 - W_n, V_2 = W_2 - W_1, \dots, W_n = W_n - W_{n-1}.$$

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 4.42 (\*\*\*)—

On considère  $E = \{x = (x_1, x_2) \in \mathbb{R}^2, x_1^2 + x_2^2 \leq 1\}$  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  $E$ ,

$$\mathbf{P}(N(A) = k) = e^{-m(A)} \frac{m(A)^k}{k!}, \text{ pour tout } k \in \mathbb{N},$$

où  $m$  est la mesure de Lebesgue sur  $\mathbb{R}^2$ .

— Si  $(A_i, i \in \mathbb{N})$  sont des boréliens disjoints deux à deux, les v.a.  $(N(A_i), i \in \mathbb{N})$  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  $\mathbb{N}$ . 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 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  $\mathbb{E}(e^{-sN(A)})$  pour tout borélien. Pour  $f$  fonction mesurable positive de  $E$  dans  $\mathbb{R}^+$ , on pose

$$N(f)(\omega) = \sum_{\xi \in \omega} f(\xi).$$

Calculer  $\mathbb{E}(e^{-sN(f)})$ .

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  $\mathbb{E}\left(e^{-sN(A)}\right)$  pour tout borélien.

EXERCICE 4.43 (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 4.44.—

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  $f$  :

$$\lambda \mathbb{E}(f(W + 1)) = \mathbb{E}(W f(W)). \quad (4.10)$$

2. Réciproquement, soit  $W$  une v.a. discrète, à valeurs dans  $\mathbf{N}$ , telle que pour toute fonction positive, l'identité 4.10 soit satisfaite. En appliquant 4.10 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 4.45.—

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 4.46.—

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

$$g_a(x) = \frac{1}{\Gamma(a)} x^{a-1} e^{-x} \mathbf{1}_{\mathbb{R}^+}(x),$$

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

$$\mathbb{E} \left( e^{itG_a} \right) = (1 - it)^{-a}, \text{ pour tout } t \in \mathbb{R}.$$

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 4.47. –

Soit  $N$  un processus de Poisson (cf. exercice 4.43) 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 :

$$\lambda \int_0^{t_0} c(t)\lambda e^{-\lambda t} dt$$

avec  $\lambda = 20$ . (Indication : Considérer  $Z_n = T_{n+1} - T_n$  et expliquer pourquoi on peut appliquer le résultat précédent.)

EXERCICE 4.48. –

Soit  $N$  un processus de Poisson sur  $\mathbb{R}^+$ . Soit  $f: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ . 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

$$\mathbb{E}\left(\exp\left(-\int \mathbf{1}_{]a, b]}(s)dN_s\right)\right) = \exp\left(-\int 1 - e^{-\mathbf{1}_{]a, b]}(s)\lambda ds\right).$$

3. En déduire  $\mathbb{E}(\exp(-\int f(s)dN_s))$  pour toute fonction  $f$  positive.

4. Pour  $B \subset \mathbb{R}^+$ , calculer de deux manières différentes

$$\frac{d}{dt} \mathbb{E} \left( \exp \left( - \int (f + t \mathbf{1}_B)(s) dN_s \right) \right)_{t=0}.$$

EXERCICE 4.49.—

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  $\mathbb{R}$ . 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  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .
3. Le système dynamique  $(E, T, \mathbf{P})$  est dit ergodique lorsque

$$\mathcal{I} \subset \sigma\{A \subset \mathcal{E}, \mathbf{P}(A) = 0 \text{ ou } \mathbf{P}(A) = 1\}.$$

Montrer que  $(E, T, \mathbf{P})$  est ergodique si et seulement si les fonctions invariants 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  $L^2(d\mathbf{P})$ ,

$$\lim_{n \rightarrow +\infty} \int_E f \circ T^n g d\mathbf{P} = \int_E f d\mathbf{P} \int_E g d\mathbf{P}. \quad (4.11)$$

Montrer que si  $T$  est mélangeante alors  $(E, T)$  est ergodique.

5. Montrer que si la condition de mélange (4.11) 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  $k \in \mathbb{Z}$  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.

## 5. Espérance conditionnelle

La principale difficulté de ce chapitre est de comprendre non seulement de manière formelle, mais aussi intuitive, le concept d'espérance conditionnelle. Les nombreux exemples sont destinés à aider le lecteur.

### 5.1. Loi conditionnelle

#### 5.1.1. Introduction

Soient  $X, Y$  deux variables aléatoires. On réalise une expérience aléatoire. On suppose qu'un observateur prend connaissance de la valeur  $X(\omega)$ . Il ne connaît pas le résultat  $\omega$  de l'expérience, et ne peut généralement pas déterminer  $Y(\omega)$ , en revanche sa connaissance de  $X(\omega)$  modifie sa croyance en la valeur prise par  $Y$ . *Avant* l'expérience, l'observateur ne peut connaître que la loi  $\mathbf{P}_Y$  de  $Y$ . *Après* l'expérience, si l'observateur s'aperçoit que la réalisation de  $X$  vaut  $x$ , sa connaissance de  $Y$  s'affine en une nouvelle loi  $\mathbf{P}_{Y|X=x}$ , que l'on appelle *loi conditionnelle de  $Y$  sachant  $X = x$* .

La formalisation de ce phénomène est simple dans le cas discret. A titre d'exemple considérons la loi uniforme sur l'ensemble à 12 éléments suivant :

$$T = \{(i, j) \in \mathbb{N} \times \mathbb{N}, i + 2j \leq 5\}.$$

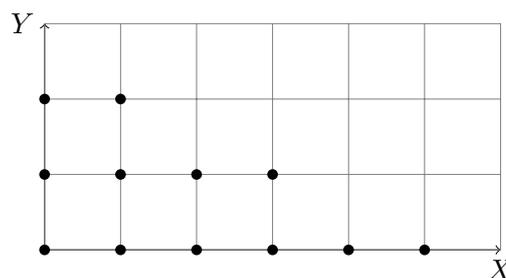


Figure 5.1. – L'ensemble  $T$

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

$$\mathbf{P}_{Y|X=5} = \delta_0.$$

Evidemment les affaires se corsent si la loi de  $X$  est à densité puisqu'alors l'événement  $\{X = x\}$  est de mesure nulle. On ne peut plus calculer de probabilité conditionnelle à l'événement  $\{X = x\}$ , cela n'a plus de sens. Pourtant, on peut très bien imaginer un couple de points répartis au hasard dans le triangle  $T'$  : Dans ces conditions, on a bien envie de dire que conditionnellement à  $X = x$ ,

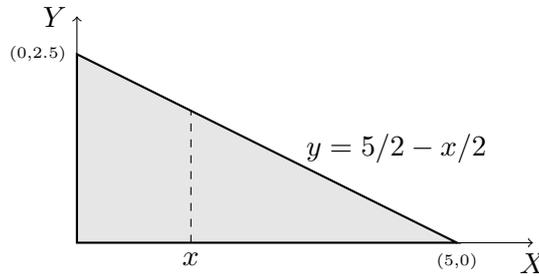


Figure 5.2. – L'ensemble  $T'$

la loi de  $Y$  est la loi uniforme sur  $[0, 5/2 - x/2]$ .

### 5.1.2. Définition

Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilité. Soient  $(E, \mathcal{E})$  et  $(F, \mathcal{G})$  deux espaces mesurables. Soient  $X : \Omega \rightarrow E$  et  $Y : \Omega \rightarrow F$  deux variables aléatoires. On note  $\mathbf{P}_{X,Y}$  la loi jointe et  $\mathbf{P}_X$  la loi marginale de  $X$ .

**Théorème 5.1.** Pour tout  $G \in \mathcal{G}$ , il existe une fonction mesurable  $x \mapsto \mathbf{P}_{Y|X=x}(G)$ , définie de manière unique  $\mathbf{P}_X$ -pp, telle que

$$\forall H \in \mathcal{E}, \mathbf{P}_{X,Y}(H \times G) = \int_H \mathbf{P}_{Y|X=x}(G) d\mathbf{P}_X(x). \quad (5.1)$$

*Démonstration.* Quels que soient les ensembles  $H \in \mathcal{E}$ ,  $G \in \mathcal{G}$ , nous avons l'implication

$$\mathbf{P}_X(H) = 0 \Rightarrow \mathbf{P}_{X,Y}(H \times G) = 0$$

(en effet, si  $\mathbb{P}(X \in H) = 0$  alors  $\mathbb{P}(X \in H, Y \in G) = 0$ ). En d'autres termes, pour  $G$  fixé, la mesure  $H \mapsto \mathbf{P}_{X,Y}(H \times G)$  est *absolument continue* par rapport à la mesure  $\mathbf{P}_X$ . D'après le

théorème 3.20, la première admet une densité par rapport à la seconde. Cela signifie que pour tout  $G$ , il existe une fonction mesurable  $f_G(x)$  telle que pour tout  $H$ ,

$$\mathbf{P}_{X,Y}(H \times G) = \int_H f_G(x) d\mathbf{P}_X(x). \quad (5.2)$$

La densité  $f_G(x)$  est notée  $\mathbf{P}_{Y|X=x}(G)$ . Le premier point est prouvé.  $\square$

REMARQUE.— Pour chaque  $G$ , la fonction  $x \mapsto \mathbf{P}_{Y|X=x}(G)$  n'est définie de façon unique qu'à un ensemble  $\mathbf{P}_X$ -négligeable près, il en existe donc plusieurs *versions*, toutes égales presque partout.

REMARQUE.— Peut-on dire que l'application  $G \mapsto \mathbf{P}_{Y|X=x}(G)$  est une mesure de probabilité? Tâchons par exemple de vérifier la  $\sigma$ -additivité. Si  $(G_n)$  est une collection d'ensembles disjoints, nous avons :

$$\begin{aligned} \mathbf{P}_{X,Y} \left( H \times \bigcup_n G_n \right) &= \sum_n \mathbf{P}_{X,Y}(H \times G_n) \\ &= \sum_n \int_H \mathbf{P}_{Y|X=x}(G_n) d\mathbf{P}_X(x) \\ &= \int_H \sum_n \mathbf{P}_{Y|X=x}(G_n) d\mathbf{P}_X(x) \end{aligned}$$

où la permutation somme-intégrale est justifiée par le corollaire 3.7. Par unicité de la densité presque partout, nous en déduisons que  $\mathbf{P}_{Y|X=x}(\cup_n G_n) = \sum_n \mathbf{P}_{Y|X=x}(G_n)$   $\mathbf{P}_X$ -pp. Dans le cas où  $(F, \mathcal{G}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ , il se trouve que l'on peut retirer ce « presque partout » à condition de choisir le version adéquate de  $x \mapsto \mathbf{P}_{Y|X=x}(G)$  pour chaque  $G$  [?, Théorème 33.3]. Dans ce cas, l'application  $G \mapsto \mathbf{P}_{Y|X=x}(G)$  est une mesure de probabilité pour *tout*  $x$ . On l'appelle la *loi conditionnelle de Y sachant X = x*.

### 5.1.3. Propriétés

Dans le cas où  $X, Y$  sont indépendantes, la connaissance de  $X(\omega)$  n'apporte, pour un certain observateur de l'expérience, aucune information nouvelle sur la réalisation  $Y$ . Il est donc logique que la loi conditionnelle de  $Y$  sachant  $X$  coïncide avec la loi de  $Y$ . Autrement dit, la connaissance de  $X$  ne modifie pas la croyance en l'événement  $\{Y \in G\}$ .

**Proposition 5.2.** Soient  $X, Y$  deux v.a. indépendantes. Alors pour tout  $x$   $\mathbf{P}_X$ -pp,

$$\mathbf{P}_{Y|X=x} = \mathbf{P}_Y.$$

*Démonstration.* Par définition,  $\mathbf{P}_{X,Y}(H \times G) = \mathbf{P}_X(H)\mathbf{P}_Y(G) = \int_H \mathbf{P}_Y(G) d\mathbf{P}_X(x)$ . Donc la fonction constante  $\mathbf{P}_Y(G)$  est bien une densité de  $\mathbf{P}_{X,Y}(\cdot \times G)$  par rapport à  $\mathbf{P}_X$ .  $\square$

A l'inverse, supposons que  $Y = f(X)$  pour une certaine fonction  $f$ , la connaissance de  $X(\omega)$  permet de déterminer  $Y(\omega)$  exactement. Autrement dit, une fois  $X$  connue, il n'y a plus d'aléa sur  $Y$  : tout se passe comme si  $Y$  était déterministe. Il est donc logique que la loi conditionnelle de  $Y$  sachant  $X$  coïncide avec une mesure de Dirac.

**Proposition 5.3.** Supposons que  $Y = f(X)$  pour une certaine fonction mesurable  $f$ . Pour tout  $x$   $\mathbf{P}_X$ -pp,

$$\mathbf{P}_{Y|X=x} = \delta_{f(x)}.$$

*Démonstration.* On a  $\mathbb{P}(X \in H, Y \in G) = \mathbb{P}(X \in H \cap f^{-1}(G)) = \int_H \mathbf{1}_G(f(x)) d\mathbf{P}_X(x)$ . En identifiant les densités presque partout,  $\mathbf{P}_{Y|X=x}(G) = \mathbf{1}_G(f(x)) = \delta_{f(x)}(G)$ .  $\square$

#### 5.1.4. Cas des vecteurs aléatoires à densité

Soit  $\mu$  une mesure de référence sur  $(E, \mathcal{E})$  et  $\nu$  une mesure de référence sur  $(F, \mathcal{G})$  (typiquement  $\mu, \nu$  sont des mesures de Lebesgue ou, dans le cas discret, des mesures de comptage).

**Théorème 5.4.** Soient  $(X, Y)$  un couple de v.a. dont la loi jointe admet une densité  $f_{X,Y}$  par rapport à la mesure produit  $\mu \otimes \nu$ , soit :

$$d\mathbf{P}_{X,Y}(x, y) = f_{X,Y}(x, y) d\mu(x) d\nu(y).$$

On définit la *densité conditionnelle* de  $Y$  sachant  $X = x$  par

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

où  $f_X(x) = \int f_{X,Y}(x, y) d\nu(y)$  est la densité de  $X$ . Alors, pour tout  $x$   $\mathbf{P}_X$ -pp, la fonction  $f_{Y|X=x} : F \rightarrow \mathbb{R}_+$  est bien définie, et  $\mathbf{P}_{Y|X=x}$  est de densité  $f_{Y|X=x}$  par rapport à  $\nu$  :

$$\mathbf{P}_{Y|X=x}(G) = \int_G f_{Y|X=x}(y) d\nu(y) \quad (\forall G \in \mathcal{G}). \quad (5.3)$$

Bien que les conditions d'applications du théorème 5.4 ne soient pas toujours satisfaites<sup>1</sup>, elles le sont dans un grand nombre de cas pratiques. Et on voit bien l'intérêt de la formule (5.3) pour les calculs : elle fournit une expression *explicite* de la loi conditionnelle, calculable directement en fonction de la densité jointe. Les manipulations sont donc plus simples qu'en utilisant la définition (5.1), qui passe par le biais d'une propriété ("c'est l'unique fonction telle que...").

#### Variables à densité par rapport à la mesure de Lebesgue

L'exemple typique est le cas où  $X$  et  $Y$  sont des v.a. sur  $\mathbb{R}^d$  et  $\mathbb{R}^n$  respectivement, telles que  $(X, Y)$  admet une densité  $f_{X,Y}$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}^{d+n}$ . Dans ce cas,  $\mu = \lambda_d$ ,  $\nu = \lambda_n$  et  $\mu \otimes \nu = \lambda_{d+n}$ .

1. Pensons par exemple au cas où  $Y = X$  et où  $X$  admet une densité par rapport à  $\lambda$  : le couple  $(X, X)$  n'admet pas de densité par rapport à  $\lambda \otimes \lambda$ .

EXEMPLE 25.— Avec ces outils, nous sommes maintenant capables de traiter le cas de la figure 5.2 du paragraphe d'introduction. 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  $[0, 5/2 - x/2]$ .

### Variables aléatoires discrètes

Un autre cas d'application du théorème 5.4 est le cas où  $X, Y$  sont des v.a. discrètes, disons  $X : \Omega \rightarrow \mathbb{N}$  et  $Y : \Omega \rightarrow \mathbb{N}$  pour fixer les idées. Dans ce cas, nous savons que  $X$  admet une densité par rapport à la mesure de comptage sur  $\mathbb{N}$ ,  $\sum_n \delta_n$ , donnée par

$$f_X(n) = \mathbb{P}(X = n)$$

pour tout  $n$ . De même, le couple  $(X, Y)$  admet une densité par rapport à la mesure de comptage sur  $\mathbb{N}^2$ ,  $\sum_{n,k} \delta_{(n,k)}$ , donnée par

$$f_{X,Y}(n, k) = \mathbb{P}(X = n, Y = k).$$

La densité conditionnelle de  $Y$  sachant  $X = n$  est donc la fonction

$$f_{Y|X=n}(k) = \frac{\mathbb{P}(X = n, Y = k)}{\mathbb{P}(X = n)} = \mathbb{P}(Y = k | X = n)$$

et elle est bien définie pour tout  $n$  tel que  $\mathbb{P}(X = n) > 0$ .

### Autre exemple

Dans l'exemple suivant, nous traitons le cas où l'une des v.a. est discrète, et l'autre à densité par rapport à la mesure de Lebesgue.

EXEMPLE 26.— 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

$$\mathbf{P}_{Y|X=x} = x\delta_1 + (1-x)\delta_0, \text{ pour tout } x \in [0, 1], \quad (5.4)$$

en d'autres termes, la loi de  $Y$  sachant  $X = x$  est une loi de Bernoulli de paramètre  $x$ . D'après (5.4), 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 (5.4), 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.

## 5.2. Espérance conditionnelle

### 5.2.1. Définition

De la même façon que nous avons défini la probabilité conditionnelle d'un événement  $\{Y \in G\}$  sachant  $X = x$ , nous pouvons définir, en généralisant immédiatement le théorème 5.1, l'espérance conditionnelle de  $Y$  sachant  $X = x$ .

**Théorème 5.5.** Soit  $Y : \Omega \rightarrow \mathbb{R}$  une v.a. réelle, positive ou intégrable. Il existe une fonction mesurable  $x \mapsto \mathbb{E}(Y|X = x)$ , unique  $\mathbf{P}_X$ -pp, telle que

$$\forall H \in \mathcal{E}, \mathbb{E}(Y \mathbf{1}_{\{X \in H\}}) = \int_H \mathbb{E}(Y|X = x) d\mathbf{P}_X(x). \quad (5.5)$$

La quantité  $\mathbb{E}(Y|X = x)$  est appelée *espérance conditionnelle de  $Y$  sachant  $X = x$* .

*Démonstration.* La preuve est identique à celle du théorème 5.1. Soit  $Y \geq 0$ . Pour tout  $H \in \mathcal{E}$ , on a l'implication

$$\mathbf{P}_X(H) = 0 \Rightarrow \mathbb{E}(Y \mathbf{1}_H(X)).$$

On vérifie sans peine que  $H \mapsto \mathbb{E}(Y \mathbf{1}_H(X))$  est une mesure. Cette mesure est donc absolument continue par rapport à  $\mathbf{P}_X$  d'après l'implication qui précède. D'après le théorème 3.20, elle admet

une densité par rapport à  $\mathbf{P}_X$ . En notant  $x \mapsto \mathbb{E}(Y|X = x)$  cette densité, on obtient directement le résultat. Dans le cas où  $Y$  est intégrable mais n'est pas de signe constant, il suffit de poser  $\mathbb{E}(Y|X = x) = \mathbb{E}(Y^+|X = x) - \mathbb{E}(Y^-|X = x)$ . Cette quantité est bien définie et satisfait les propriétés annoncées.  $\square$

REMARQUE.— Si on pose  $H = E$  dans l'équation (5.5), on obtient la formule utile :

$$\mathbb{E}(Y) = \int \mathbb{E}(Y|X = x) d\mathbf{P}_X(x),$$

ce que l'on peut traduire par : « l'espérance  $\mathbb{E}(Y)$  est égale à l'espérance de l'espérance conditionnelle ».

### 5.2.2. Espérance conditionnelle et loi conditionnelle

Le résultat ci-dessous est une généralisation du théorème de transfert. On munit  $E \times F$  de sa tribu produit.

**Théorème 5.6.** Soit  $Y : \Omega \rightarrow F$  une v.a. et soit  $\psi : E \times F \rightarrow \mathbb{R}$  une fonction mesurable telle que  $\mathbb{E}|\psi(X, Y)| < \infty$ . Alors, pour tout  $x \in \mathbf{P}_X$ -pp.,

$$\mathbb{E}(\psi(X, Y)|X = x) = \int \psi(x, y) d\mathbf{P}_{Y|X=x}(y).$$

Dans le cas où les hypothèses du théorème 5.4 sont satisfaites (variables à densité), le résultat précédent se lit :

$$\mathbb{E}(\psi(X, Y)|X = x) = \int \psi(x, y) f_{Y|X=x}(y) d\nu(y),$$

où  $f_{Y|X=x}(y)$  est la densité conditionnelle de  $Y$  sachant  $X = x$ , et  $\nu$  est une mesure de référence. Nous incitons l'élève à rédiger lui-même la preuve dans ce cas particulier : elle est facile, et utile pour comprendre les principaux mécanismes. La démonstration ci-dessous concerne le cas général.

*Démonstration.* Notons  $\phi(x) = \int \psi(x, y) d\mathbf{P}_{Y|X=x}(y)$  le membre de droite. Par identification dans l'équation (5.5), il suffit de montrer que pour tout  $H \in \mathcal{E}$ ,

$$\mathbb{E}(\psi(X, Y)\mathbf{1}_{\{X \in H\}}) = \int_H \phi(x) d\mathbf{P}_X(x). \quad (5.6)$$

Traitons d'abord le cas où  $\psi$  est la fonction indicatrice d'un pavé  $\psi(x, y) = \mathbf{1}_{A \times B}(x, y)$  où  $A \in \mathcal{E}$ ,  $B \in \mathcal{G}$ . On a  $\phi(x) = \mathbf{P}_{Y|X=x}(B)\mathbf{1}_A(x)$  donc

$$\begin{aligned} \int \mathbf{1}_H(x) \phi(x) d\mathbf{P}_X(x) &= \int \mathbf{1}_{H \cap A}(x) \mathbf{P}_{Y|X=x}(B) d\mathbf{P}_X(x) \\ &= \mathbf{P}_{X, Y}((H \cap A) \times B) \end{aligned}$$

par définition de la loi conditionnelle. Or  $\mathbf{P}_{X, Y}((H \cap A) \times B) = \mathbb{E}(\mathbf{1}_H(X)\mathbf{1}_{A \times B}(X, Y))$ . L'égalité 5.6 est donc satisfaite pour tout  $H \in \mathcal{E}$  et tout  $\psi = \mathbf{1}_{A \times B}$ .

Pour  $H \in \mathcal{E}$  fixé, considérons deux mesures  $\mu_H, \nu_H$  définies sur  $(E \times F, \mathcal{E} \otimes \mathcal{G})$  par  $\mu_H(K) = \mathbb{E}(\mathbf{1}_K(X, Y)\mathbf{1}_{\{X \in H\}})$  et  $\nu_H(K) = \int_H \int \mathbf{1}_K(x, y) d\mathbf{P}_{Y|X=x}(y) d\mathbf{P}_X(x)$ . D'après ce qui précède, ces deux mesures coïncident sur les pavés du type  $A \times B$ . Comme les pavés forment un  $\pi$ -système engendrant  $\mathcal{E} \otimes \mathcal{G}$ , les mesures sont égales. Ceci montre que l'égalité 5.6 est satisfaite pour tout  $H \in \mathcal{E}$  et tout  $\psi$  de la forme  $\psi = \mathbf{1}_K$  où  $K \in \mathcal{E} \otimes \mathcal{G}$ .

Puisque l'égalité 5.6 est satisfaite pour les fonctions indicatrices, par linéarité de l'intégrale, elle l'est aussi pour les fonctions  $\psi$  étagées positives. Dans le cas où  $\psi$  est une fonction mesurable positive quelconque, l'affirmation reste vraie par l'argument habituel de convergence monotone (choisir une suite de fonctions étagées  $\psi_n$  telle que  $0 \leq \psi_n \uparrow \psi$ ). Enfin, lorsque  $\psi$  n'est pas de signe constant, on conclut avec la décomposition  $\psi = \psi^+ - \psi^-$ .  $\square$

REMARQUE.— Comme corollaire immédiat, si  $Y$  est une v.a.r.,

$$\mathbb{E}(Y|X = x) = \int y d\mathbf{P}_{Y|X=x}(y). \quad (5.7)$$

L'espérance conditionnelle est l'intégrale par rapport à la loi conditionnelle.

**Corollaire 5.7.** Soit  $g : E \rightarrow \mathbb{R}$  une fonction mesurable et  $Y$  une v.a.r. intégrable. On suppose que  $\mathbb{E}|g(X)Y| < \infty$ . Alors,

$$\mathbb{E}(g(X)Y|X = x) = g(x)\mathbb{E}(Y|X = x).$$

*Démonstration.* Utiliser le théorème 5.6 avec  $\psi(x, y) = g(x)y$ .  $\square$

**Corollaire 5.8.** Si  $Y$  est une v.a.r. positive ou intégrable, indépendante de  $X$ , alors

$$\mathbb{E}(Y|X = x) = \mathbb{E}(Y).$$

*Démonstration.* Utiliser l'équation (5.7) avec le fait que, si  $X$  et  $Y$  sont indépendantes,  $\mathbf{P}_{Y|X=x} = \mathbf{P}_Y$ .  $\square$

## 5.3. Exercices

### 5.3.1. Variables discrètes

EXERCICE 5.1.—

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  $X + Y$ .

EXERCICE 5.2.—

Un jeu télévisé se déroule comme suit. Un candidat est face à trois portes. Une récompense se trouve derrière l'une d'elles. Seul le présentateur connaît l'emplacement de la récompense. Le candidat désigne une porte en première intention (sans l'ouvrir pour le moment). Dans un deuxième temps, le présentateur de l'émission ouvre au hasard l'une des deux portes non désignées par le candidat, en prenant garde toutefois de ne jamais ouvrir la porte derrière laquelle est cachée la récompense.

On suppose que le candidat désigne la porte n°1. On introduit la variable aléatoire  $X \in \{1, 2, 3\}$  la position de la récompense et la v.a.  $Y$  représentant la porte ouverte par le présentateur.

1. Donner les valeurs de  $P(Y = 2|X = 1)$  et  $P(Y = 3|X = 2)$ .
2. Calculer  $P(X = 3|Y = 2)$ .
3. Après avoir ouvert la porte  $Y$ , le présentateur demande au candidat s'il souhaite maintenir ou modifier son choix. Que doit faire le candidat pour optimiser ses chances ? Justifier.

EXERCICE 5.3.—

Soient  $0 \leq r \leq p \leq 1$  tels que  $1 - 2p + r \geq 0$ . Soient  $X_1, X_2$  des v.a. de Bernoulli telles que

$$\mathbb{P}(X_1 = 1, X_2 = 1) = r, \quad \mathbb{P}(X_1 = 0, X_2 = 1) = p - r, \quad \mathbb{P}(X_1 = 1, X_2 = 0) = p - r.$$

1. Que vaut  $\mathbb{P}(X_1 = 0, X_2 = 0)$  ?
2. Donner les lois de  $X_1$  et de  $X_2$ .
3. Calculer  $\mathbb{E}(X_1|X_2 = x)$ .

EXERCICE 5.4.—

Soit  $X$  une v.a. géométrique de paramètre  $p$  ( $0 < p < 1$ ), soit  $\mathbb{P}(X = k) = (1 - p)^{k-1}p$  pour  $k \geq 1$ .

1. Montrer que pour tout  $k, n \geq 0$ ,  $\mathbb{P}(X > k + n|X > k) = \mathbb{P}(X > n)$ . On dit que la loi géométrique est sans mémoire.
2. Soient  $X_1, X_2$ , deux v.a. indépendantes de loi géométrique de paramètre resp.  $p_1$  et  $p_2$ . Quelle est la loi de  $Y = \min(X_1, X_2)$  ?

EXERCICE 5.5.—

Le quart d'une population a été vacciné contre une maladie. Au cours d'une épidémie, on constate qu'il y a parmi les malades un vacciné pour quatre non vaccinés. On sait de plus qu'il y avait

un malade sur douze parmi les vaccinés. Quelle était la probabilité de tomber malade pour un individu non vacciné ?

EXERCICE 5.6. –

Soient  $X, Y, T$  des v.a. discrètes à valeurs réelles. Que peut-on dire des quantités suivantes (sous réserves d'hypothèses de sommabilité adéquates) ?

1.  $\mathbb{E}(f(T)|T = t)$
2.  $\mathbb{E}(XY|T = t)$  si  $X = f(T)$
3.  $\mathbb{E}(f(X)|T = t)$  avec  $X$  et  $T$  indépendantes
4.  $\mathbb{E}(S_{10}|S_8 = s)$  avec  $S_n = \sum_{i=1}^n X_i$  et les  $X_i$  iid
5.  $\mathbb{E}(S_{31}|X_1 = x)$  avec  $S_n = \sum_{i=1}^n X_i$  et les  $X_i$  iid
6.  $\mathbb{E}(\Pi_4|\Pi_2 = p)$  avec  $\Pi_n = \prod_{i=1}^n X_i$  et les  $X_i$  iid.

### 5.3.2. Cas général

EXERCICE 5.7. –

Soit  $(X, Y)$  un couple de loi uniforme sur le disque unité. Donner la loi conditionnelle de  $Y$  sachant  $X$ . Calculer la variance conditionnelle  $\text{Var}(Y|X = x)$ .

EXERCICE 5.8. –

Soit  $(X, Y)$  un couple de densité  $f(x, y) = 2\mathbf{1}_{0 \leq y \leq x \leq 1}$ .

1. Quelle est la loi de  $X$  ?
2. Donner la densité conditionnelle de  $Y$  sachant  $X$ .
3. Calculer  $\mathbb{P}(X^2 + Y^2 \leq 1|X = x)$ .
4. En déduire  $\mathbb{P}(X^2 + Y^2 \leq 1)$ .

EXERCICE 5.9. –

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  $\mathbb{R}^2$ .

1. Trouver la constante  $c$  pour que  $\mathbf{P}$  soit une probabilité.
2. Soit  $(\Omega, \mathcal{F}, \mathbf{P})$  un espace de probabilité et  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$  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 5.10. –

On cherche à estimer un paramètre  $\Theta$  non-observé, que l'on modélise comme une v.a.r. On

réalise un expérience aléatoire qui conduit au résultat  $X$ . Un estimateur de  $\Theta$  est une variable aléatoire fonction du résultat de l'expérience, notée  $\hat{\Theta}(X)$  où  $\hat{\Theta} : \mathbb{R} \rightarrow \mathbb{R}$  est une certaine fonction à déterminer. On définit l'erreur quadratique moyenne par

$$EQM(\hat{\Theta}) = \mathbb{E}((\hat{\Theta}(X) - \Theta)^2).$$

On définit la v.a.  $\mathbb{E}(\Theta|X)$  comme la fonction qui à tout  $\omega$  associe la valeur de  $\mathbb{E}(\Theta|X = x)$  au point  $x = X(\omega)$ .

1. Montrer l'identité  $EQM(\hat{\Theta}) = \mathbb{E}((\hat{\Theta}(X) - \mathbb{E}(\Theta|X))^2) + \mathbb{E}((\mathbb{E}(\Theta|X) - \Theta)^2)$
2. Quel est l'estimateur optimal au sens de l'EQM?

**EXERCICE 5.11.**—

À un arrêt de bus, un jour donné, les bus passent à intervalles réguliers de durée  $T$ . Dans les conditions idéales,  $T = t_0$ . Cependant, le trafic routier fait que  $T$  est aléatoire, et on peut modéliser la loi de  $T$  par une loi Pareto de paramètres  $(t_0, \alpha)$ ,  $T \sim \mathcal{P}(t_0, \alpha)$  avec  $\alpha > 0$  fixé, c'est-à-dire

$$\forall t \in \mathbb{R}, \quad \mathbb{P}(T > t) = \begin{cases} \left(\frac{t_0}{t}\right)^\alpha & \text{si } t \geq t_0 \\ 1 & \text{sinon.} \end{cases}$$

1. Quelle est la fonction de répartition de  $T$ ?
2. Montrer que  $T$  admet une densité  $f_T$ , que l'on explicitera.

Soit  $n$  voyageurs attendant à l'arrêt de bus, avec  $n \geq 1$ . On note  $X_i$  le temps d'attente du voyageur  $i$ , pour  $i = 1, \dots, n$ . On admet que si  $T = t$ , alors  $X_i$  suit une loi uniforme sur  $[0, t]$ . Autrement dit la loi conditionnelle de  $X_i$  sachant  $\{T = t\}$ , notée  $P_{X_i|t}$ , est une loi uniforme sur  $[0, t]$ .

On note  $X = (X_1, \dots, X_n)$ . Les v.a.  $X_1, \dots, X_n$  sont indépendantes conditionnellement à  $\{T = t\}$ , c'est-à-dire que la loi conditionnelle de  $X$  sachant  $\{T = t\}$  est la mesure produit :

$$P_{X|T=t} = P_{X_1|T=t} \otimes \dots \otimes P_{X_n|T=t}.$$

3. Donner la densité de  $P_{X_1|T=t}$ , la loi de  $X_1$  sachant  $\{T = t\}$ .
4. Donner l'espérance du temps d'attente  $\mathbb{E}(X_1)$  du premier passager en fonction de  $\alpha > 0$ .  
*Indication* : Commencer par calculer  $\mathbb{E}(X_1|T = t)$ . Pour certaines valeurs de  $\alpha$ ,  $\mathbb{E}(X_1) = +\infty$ .
5. Pour  $t \geq t_0$ , calculer la densité de la loi conditionnelle  $P_{X|t}$  de  $X$  sachant  $\{T = t\}$ .  
On notera  $f_{X|T=t}(x)$  cette densité conditionnelle dans la suite.
6. Soit  $x = (x_1, \dots, x_n)$  le vecteur des temps d'attente effectifs des  $n$  voyageurs. Montrer que la loi de  $T$  sachant  $\{X = x\}$ , notée  $P_{T|X=x}$ , est une loi de Pareto dont on précisera les paramètres  $(t_n, \alpha_n)$  en fonction des observations  $(x_1, \dots, x_n)$ . On pourra faire intervenir  $m(x) = \max_{i=1, \dots, n} x_i$ .  
*Indication* : Calculer la densité conditionnelle de  $T$  sachant  $\{X = x\}$ .
7. La compagnie de bus s'intéresse à la durée moyenne  $\mathbb{E}T$ . Grâce aux réseaux sociaux, elle a accès aux temps d'attente de  $n$  passagers d'un même bus. Quelle est l'espérance conditionnelle de  $T$ , sachant cette information? Autrement dit, que vaut  $\mathbb{E}(T|X = x)$ ?

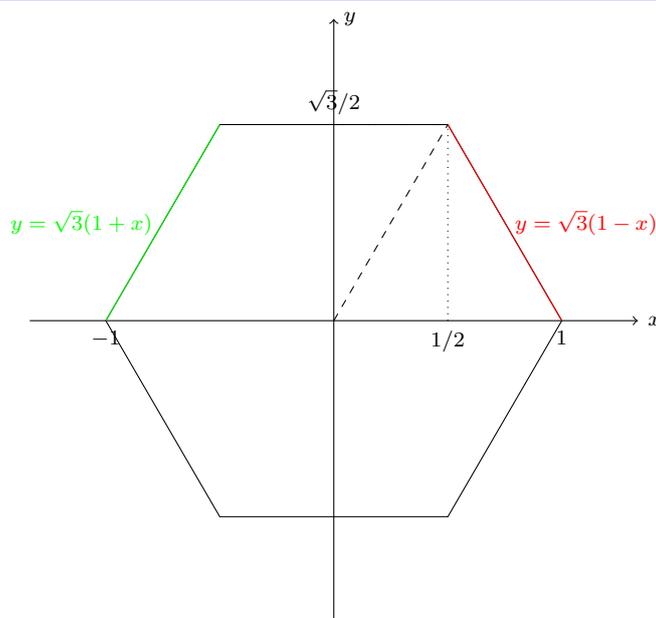
**Pour s'entraîner**

EXERCICE 5.12 (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 5.3 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  $A = 3\sqrt{3}/2$ .

Figure 5.3. – Hexagone régulier.

**Pour aller plus loin**

EXERCICE 5.13.—

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 :

$$X' = (X_1, \dots, X_{i-1}, Y_i, X_{i+1}, \dots) \text{ si } U = i.$$

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

$$\mathbb{E}(W_n - W'_n \mid W_n = m) = \frac{m}{n} - p.$$

**EXERCICE 5.14.**—

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

$$(\Gamma^{1/2})^{-1} = \Gamma^{-1}, \det \Gamma^{1/2} = (\det \Gamma)^{1/2} \text{ et que } \|(\Gamma^{1/2})^{-1}z\|^2 = \Gamma^{-1}z \cdot z$$

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  $\mathcal{N}(0, \mathbf{I}_2)$  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.



## 6. Fonction caractéristique

**Notations.** 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 :

$$f = f_R + if_I \quad (6.1)$$

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 (6.1) 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 3.14) reste vrai si les fonctions  $f, f_n$  sont supposées à valeur dans  $\mathbb{C}$  (en remplaçant la valeur absolue par le module).

### 6.1. Définition et propriétés

**Définition 6.1.** Soit  $\mu$  une mesure finie sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . On appelle *fonction caractéristique* de  $\mu$  la fonction  $\phi_\mu : \mathbb{R}^d \rightarrow \mathbb{C}$  définie par

$$\phi_\mu(t) := \int \exp(i\langle t, x \rangle) \, d\mu(x) .$$

Notons que  $\phi_\mu$  est bien définie sur  $\mathbb{R}^d$  puisque  $|\exp(i\langle t, x \rangle)| = 1$ .

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 6.1 implique que  $\phi_\mu(t) = \int e^{itx} f(x)dx$ . Ainsi,  $\phi_\mu(-t)$  est égale à la transformée de Fourier de la fonction intégrable  $f$ , vue en cours d'analyse. C'est pourquoi, même dans le cas général, la fonction  $\phi_\mu$  est aussi appelée la *transformée de Fourier* de  $\mu$ .

Soit  $X = (X_1, \dots, X_d)$  un vecteur aléatoire à valeur dans  $\mathbb{R}^d$ , défini sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ . La *fonction caractéristique* de  $X$ , notée  $\phi_X$ , est par définition la fonction caractéristique de sa loi. Autrement dit, on note simplement  $\phi_X$  la fonction  $\phi_{\mathbf{P}_X}$ . On a les écritures suivantes :

$$\begin{aligned}\phi_X(t) &= \mathbb{E}(\exp(i\langle t, X \rangle)) = \int_{\mathbb{R}^d} \exp(i\langle t, x \rangle) d\mathbf{P}_X(x) \\ &= \mathbb{E}\left(\exp\left(i \sum_{k=1}^d t_k X_k\right)\right) = \mathbb{E}\left(\prod_{k=1}^d \exp(it_k X_k)\right).\end{aligned}$$

**Proposition 6.1.** La fonction caractéristique d'une v.a.  $X$  vérifie les propriétés suivantes :

- Pour tout  $t \in \mathbb{R}^d$ ,  $|\phi_X(t)| \leq 1$ . En outre,  $\phi_X(0) = 1$ .
- La fonction  $\phi_X$  est continue sur  $\mathbb{R}^d$ .
- 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

$$\phi_{aX+b}(t) = \exp(itb) \phi_X(at).$$

- Si  $X, Y$  sont deux v.a. indépendantes définies sur le même espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$  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  $|\phi_X(t)| \leq \mathbb{E}(|\exp(i\langle t, X \rangle)|) = 1$ . De plus,  $\phi_X(0) = \mathbb{E}(\exp(0)) = 1$ .

b)) La continuité en tout point  $t \in \mathbb{R}^d$  est une conséquence du théorème de convergence dominée (voir théorème 3.14). Soit  $t \in \mathbb{R}^d$ ; nous écrivons pour toute suite  $(h_\ell)_\ell$  à valeur dans  $\mathbb{R}^d$  telle que  $\lim_{\ell \rightarrow +\infty} h_\ell = 0$ ,

$$\phi_X(t + h_\ell) - \phi_X(t) = \int (\exp(i\langle t + h_\ell, x \rangle) - \exp(i\langle t, x \rangle)) d\mathbf{P}_X(x).$$

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

$$\begin{aligned}\phi_{AX+b}(t) &= \mathbb{E}(\exp(i\langle AX + b, t \rangle)) = \exp(i\langle b, t \rangle) \mathbb{E}(\exp(i\langle AX, t \rangle)) \\ &= \exp(i\langle b, t \rangle) \mathbb{E}(\exp(i\langle X, A^T t \rangle)) = \exp(i\langle b, t \rangle) \phi_X(A^T t) .\end{aligned}$$

d)) On a pour tout  $t \in \mathbb{R}^d$ ,

$$\begin{aligned}\phi_{X+Y}(t) &= \mathbb{E}(\exp(i\langle t, X \rangle) \exp(i\langle t, Y \rangle)) = \mathbb{E}(\exp(i\langle t, X \rangle)) \mathbb{E}(\exp(i\langle t, Y \rangle)) \\ &= \phi_X(t) \phi_Y(t)\end{aligned}$$

en utilisant la caractérisation de l'indépendance donnée par le théorème 4.15 (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. □

La table 6.1 fournit l'expression des fonctions caractéristiques de plusieurs loi usuelles (voir l'exercice 6.1).

## 6.2. Caractérisation de la loi

**Théorème 6.2.** 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, on ne se préoccupe que du sens direct. Plaçons nous pour simplifier dans le cas  $d = 1$ . La preuve repose sur la formule d'inversion (voir l'exercice 6.5) :

$$\mathbf{P}_X(]a, b]) = \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t) dt , \quad (6.2)$$

qui vraie pour tous  $a, b$  en lesquels  $\mathbf{P}_X$  n'a pas de masse, c'est-à-dire pour tous  $a, b$  hors de l'ensemble  $D_X := \{x : \mathbf{P}_X(\{x\}) > 0\}$ . D'après l'exercice 2.12, cet ensemble  $D_X$  est au plus dénombrable. Donc pour tout  $a$  et  $b$  hors de  $D_X \cup D_Y$ , on a  $\mathbf{P}_X(]a, b]) = \mathbf{P}_Y(]a, b])$ . Puisque  $F_X(b) = F_X(a) + \mathbf{P}_X(]a, b])$ , il suffit de faire tendre  $a$  vers  $-\infty$  pour obtenir :

$$F_X(b) = F_Y(b) \quad (6.3)$$

pour tout point  $b$  hors d'un ensemble au plus dénombrable. En utilisant la continuité à droite des fonctions de répartition, on conclut que (6.3) est vraie en tout point. Puisque la fonction de répartition détermine entièrement la loi, le résultat est démontré. □

**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  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Théorème 6.3.** 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

$$\phi_X(t) = \mathbb{E} \left( \prod_{k=1}^p \exp(i\langle t_k, X_k \rangle) \right) = \prod_{k=1}^p \mathbb{E} (\exp(i\langle t_k, X_k \rangle)) = \prod_{k=1}^p \phi_{X_k}(t_k) ,$$

en utilisant la caractérisation de l'indépendance donnée par le théorème 4.15 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 6.2  $\phi_{X_k}(t_k) = \phi_{Y_k}(t_k)$  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 6.2 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.  $\square$

### 6.3. Calcul de moments

La fonction caractéristique est un outil souvent commode pour évaluer les moments.

**Théorème 6.4.** 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

$$\frac{\partial^p \phi_X(t)}{\partial t^p} = i^p \mathbb{E} (X^p \exp(itX)) .$$

En particulier, les moments sont liés aux dérivées en zéro :

$$\mathbb{E}(X^p) = (-1)^p i^p \frac{\partial^p \phi_X(0)}{\partial t^p} .$$

*Démonstration.* Nous établissons le résultat suivant :

$$\phi_X(t) = 1 + it\mathbb{E}(X) - \frac{t^2}{2}\mathbb{E}(X^2) + \dots + i^p \frac{t^p}{p!}\mathbb{E}(X^p) + \xi(t) \quad \text{où } \lim_{t \rightarrow 0} \xi(t)/t^p = 0 .$$

Cette égalité se justifie par le développement :

$$\begin{aligned}
\exp(itx) &= \sum_{k=0}^p i^k \frac{t^k}{k!} x^k \\
&= \sum_{k>p} i^k \frac{t^k}{(k-p)!} x^k \frac{1}{(k-p+1) \cdots (k-1)k} \\
&= \sum_{k>p} i^k \frac{t^k}{(k-p)!} x^k \int_0^1 \int_0^{u_p} \int_0^{u_{p-1}} \cdots \int_0^{u_2} u_1^{k-p} du_1 du_2 \cdots du_p \\
&= i^p t^p x^p \int_0^1 \int_0^{u_p} \int_0^{u_{p-1}} \cdots \int_0^{u_2} \sum_{k>p} i^{k-p} \frac{t^{k-p}}{(k-p)!} x^{k-p} u_1^{k-p} du_1 du_2 \cdots du_p \\
&= i^p t^p x^p \int_0^1 \int_0^{u_p} \int_0^{u_{p-1}} \cdots \int_0^{u_2} (\exp(iu_1 t x) - 1) du_1 du_2 \cdots du_p,
\end{aligned}$$

où l'on a utilisé le théorème Fubini. On montre que l'espérance de ce dernier terme est  $o(t^p)$  par application du théorème de convergence dominée.  $\square$

REMARQUE.— Le résultat se généralise sans peine aux vecteurs aléatoires. 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

$$\frac{\partial^p \phi_X(t)}{\partial t_1 \cdots \partial t_p} = i^p \mathbb{E}(X_{t_1} X_{t_2} \cdots X_{t_p} \exp(i \langle t, X \rangle)) . \quad (6.4)$$

On en déduit aussi une méthode pour le calcul de moments à partir de l'expression de la fonction caractéristique.

### Application : moments de la loi gaussienne

Nous montrons que si  $X$  est une v.a. réelle gaussienne d'espérance  $m$  et de variance  $\sigma^2$  ( $\sigma^2 > 0$ ) i.e.,  $X \sim \mathcal{N}(m, \sigma^2)$  alors tous les moments impairs de  $X - m$  sont nuls et les moments pairs s'expriment à l'aide de  $\sigma^2$  : pour tout  $q \in \mathbb{N}$ ,

$$\mathbb{E}((X - m)^{2q+1}) = 0, \quad (6.5)$$

$$\mathbb{E}((X - m)^{2q}) = \sigma^{2q} (2q - 1)(2q - 3) \cdots 3 = \sigma^{2q} \frac{(2q)!}{2^q q!}. \quad (6.6)$$

*Démonstration.* On veut calculer  $\mathbb{E}((X - m)^q)$  pour tout  $q \in \mathbb{N}$ . Nous avons établi (exercice ??) que si  $X \sim \mathcal{N}(m, \sigma^2)$  alors  $\sigma^{-1}(X - m) \sim \mathcal{N}(0, 1)$ . Par suite, nous allons établir

$$\mathbb{E}(Y^{2q+1}) = 0, \quad \mathbb{E}(Y^{2q}) = (2q - 1)(2q - 3) \cdots 3 = \frac{(2q)!}{2^q q!}, \quad (6.7)$$

où  $Y \sim \mathcal{N}(0, 1)$ . On écrit

$$\mathbb{E}(\exp(itY)) = \mathbb{E}\left(\sum_{n \geq 0} \frac{(itY)^n}{n!}\right) = \mathbb{E}\left(\lim_{N \rightarrow +\infty} \sum_{k=0}^N \frac{(it)^k}{k!} Y^k\right).$$

Pour permuter limite et espérance, on applique le théorème de convergence dominée (théorème 3.14) 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,

$$\mathbb{E}(\exp(itY)) = \lim_{N \rightarrow +\infty} \sum_{k=0}^N \frac{(it)^k}{k!} \mathbb{E}(Y^k) = \sum_{k \geq 0} \frac{(it)^k}{k!} \mathbb{E}(Y^k) .$$

D'autre part, d'après le Tableau 6.1

$$\mathbb{E}(\exp(itY)) = \exp(-1/2 t^2) = \sum_{n \geq 0} \frac{(-1)^n t^{2n}}{2^n n!} .$$

On en déduit (6.7) par identification. □

Par convention, une loi gaussienne d'espérance  $m$  et de variance nulle est une masse de Dirac en  $m$  ( $m \in \mathbb{R}$ ). Si  $X \sim \delta_m$  alors tous ses moments centrés sont nuls et les égalités ci-dessus restent vraies.

Ainsi, on a établi que si  $X \sim \mathcal{N}(m, \sigma^2)$ , avec  $\sigma^2 \geq 0$ , on a (6.5) et (6.6).

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 $\mathcal{N}(m, \sigma^2)$ avec $\sigma^2 \geq 0$	$\exp(itm - 1/2 \sigma^2 t^2)$
Gamma $\Gamma(a, b)$	$\left(\frac{b}{b - it}\right)^a$

Table 6.1. – Quelques fonctions caractéristiques utiles.

## 6.4. Exercices

### Pour apprendre

EXERCICE 6.1.–

Etablir par le calcul l'expression des fonctions caractéristiques des lois usuelles données dans la table 6.1.

EXERCICE 6.2.–

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 6.3.–

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 6.4.–

La densité et la fonction caractéristique d'une loi gamma de paramètres  $(\alpha, \lambda) \in (0, +\infty)^2$  :

$$x \mapsto \frac{x^{\alpha-1} e^{-\lambda x} \lambda^\alpha}{\Gamma(\alpha)} \mathbf{1}_{[0, +\infty)}(x)$$

$$t \mapsto (1 - it/\lambda)^{-\alpha}$$

où  $\Gamma(\cdot)$  est la fonction gamma d'Euler. On rappelle que  $\Gamma(n+1) = n!$  pour  $n$  entier.

1. Qu'est-ce qu'une loi gamma de paramètres  $(1, \lambda)$  ?
2. Soient  $X_1, \dots, X_n$  iid de loi gamma de paramètres  $(\alpha, \lambda)$ . Quelle est la loi de  $S_n = X_1 + \dots + X_n$  ?

### Pour aller plus loin

EXERCICE 6.5.–

Soit  $X$  une v.a.r. de loi  $\mathbf{P}_X$  et de fonction caractéristique  $\phi_X$ . On veut démontrer la formule d'inversion (6.2). 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 (6.2) pour tout points  $a, b$  tels que  $\mathbf{P}_X(\{a\}) = \mathbf{P}_X(\{b\}) = 0$ .



# 7. Vecteurs gaussiens

On note  $\langle a, b \rangle$  le produit scalaire de deux vecteurs  $a, b$  de  $\mathbb{R}^d$ ; et on note  $A^T$  la transposée de la matrice  $A$ . **Par convention, les vecteurs sont des vecteurs-colonne.**

## 7.1. Préliminaires

La loi gaussienne (ou normale) de paramètres  $m, \sigma^2$  ( $\sigma \geq 0, m \in \mathbb{R}$ ) - 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

$$\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \frac{(x - m)^2}{\sigma^2}\right) . \quad (7.1)$$

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 6.1)

$$t \mapsto \exp(itm - \frac{1}{2} t^2 \sigma^2) . \quad (7.2)$$

## 7.2. Définition, propriétés

### 7.2.1. Définition

Dans la suite de ce chapitre,  $m \in \mathbb{R}^d$  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 4.2.5). Nous écrivons  $m = (m_1, \dots, m_d)$  et noterons  $\Gamma_{i,j}$  l'élément  $(i, j)$  de la matrice  $\Gamma$ .

Soit  $X$  un vecteur aléatoire à valeur dans  $\mathbb{R}^d$  défini sur  $(\Omega, \mathcal{F}, \mathbb{P})$  et possédant des moments d'ordre 2.

**Définition 7.1.**  $X$  est un *vecteur gaussien* (ou *variable gaussienne multivariée* ou *variable normale multivariée*) si et seulement si pour tout  $a \in \mathbb{R}^d$ , la loi de  $\langle a, X \rangle$  est une loi gaussienne (éventuellement de variance nulle).

### 7.2.2. Fonction caractéristique

Nous avons vu au chapitre 6 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 7.1.** 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  $t \in \mathbb{R}^d$ ,  $\langle t, X \rangle$  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 6.1.

Réciproquement, supposons b)). Identifions la loi de  $\langle t, X \rangle$ , pour tout  $t \in \mathbb{R}^d$ , en calculant la fonction caractéristique de cette v.a. à valeur dans  $\mathbb{R}$ . Soit  $y \in \mathbb{R}$  :

$$\begin{aligned} \phi_{\langle t, X \rangle}(y) &= \mathbf{E}(\exp(iy \langle t, X \rangle)) \\ &= \mathbf{E}(\exp(i\langle yt, X \rangle)) = \phi_X(yt) = \exp(i\langle yt, m \rangle - 1/2 (yt)^T \Gamma (yt)) \\ &= \exp(iy \langle t, m \rangle - 1/2 y^2 (t^T \Gamma t)) . \end{aligned}$$

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 6.1). Donc  $X$  est un vecteur gaussien. On peut identifier l'espérance  $\mathbb{E}(X_i)$  à  $m_i$  et la covariance  $\text{Cov}(X_i, X_j)$  à  $\Gamma_{i,j}$  en utilisant la formule (6.4).

□

Ce théorème, combiné au théorème 6.2, montre que la loi d'un vecteur gaussien est entièrement caractérisée par son espérance  $m$  et sa matrice de covariance  $\Gamma$ .

### 7.2.3. 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 4.17) donc, d'après le tableau 6.1, sa fonction caractéristique est donnée par  $\exp(-1/2 \|t\|^2)$ . D'après le théorème 7.1,  $X$  est un vecteur gaussien. □

Plus généralement, on peut obtenir un vecteur gaussien par concaténation de v.a. gaussiennes indépendantes.

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 (7.1)). La proposition suivante précise le lien entre les paramètres du vecteur gaussien et les paramètres de la loi gaussienne de chaque composante.

**Proposition 7.2.** 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  $t \in \mathbb{R}$  :

$$\phi_{X_k}(t) = \phi_X((0, \dots, 0, t, 0, \dots, 0)) = \exp(itm_k - 1/2 t^2 \Gamma_{k,k}) .$$

D'après le théorème 6.2, 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. (7.2)).  $\square$

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 7.5.

### 7.3. 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 7.5) et que le résultat est ici établi sous des hypothèses précises sur la loi jointe de cette famille de v.a.

**Théorème 7.3.** Soient  $(X_k)_{k \leq d}$  des v.a. réelles définies sur  $(\Omega, \mathcal{F}, \mathbb{P})$ . 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 7.1, la fonction caractéristique de  $X$  est

$$\phi_X(t) = \exp\left(i \sum_{k=1}^d t_k m_k\right) \exp\left(-1/2 \sum_{k=1}^d t_k^2 \sigma_k^2\right) = \prod_{k=1}^d \exp\left(it_k m_k - 1/2 t_k^2 \sigma_k^2\right) .$$

Or on sait que chaque composante  $X_k$  suit une loi  $\mathcal{N}(m_k, \sigma_k^2)$  (cf. Proposition 7.2). Donc  $\phi_X(t) = \prod_{k=1}^d \phi_{X_k}(t_k)$ . Ainsi, par le théorème 6.3, 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  $t \in \mathbb{R}^d$ ,  $\langle t, X \rangle$  est une v.a. gaussienne est établi en section 7.2.3.  $\square$

**Corollaire 7.4.** 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 7.5 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.

## 7.4. Stabilité par transformation affine

**Proposition 7.5.** Soient  $b \in \mathbb{R}^p$  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  $\mathbb{R}^p$ ), d'espérance  $Am + b$  et de matrice de covariance  $A\Gamma A^T$ .

*Démonstration.* Calculons la fonction caractéristique de  $AX + b$ . Soit  $t \in \mathbb{R}^p$ . On a

$$\phi_{AX+b}(t) = \exp(i\langle t, b \rangle) \phi_X(A^T t).$$

Par le Théorème 7.1, 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 7.1, on en déduit que  $AX + b$  est un vecteur gaussien d'espérance  $Am + b$  et de matrice de covariance  $A\Gamma A^T$ .  $\square$

EXEMPLE 27.— Si  $(X_1, \dots, X_d)$  est un vecteur gaussien, alors  $X_1 + \dots + X_d$  est une v.a. gaussienne par application du résultat précédent. En effet, la somme précédente s'écrit bien comme le produit scalaire de  $X$  et du vecteur  $(1, \dots, 1)$ , il s'agit bien d'une transformation affine de  $X$ . Pouvez-vous exprimer la moyenne et la variance ?

### Construction d'un vecteur gaussien

Le théorème 7.3 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 7.3 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 7.5 entraîne que  $m + \sqrt{\Gamma}Y$  a pour loi  $\mathcal{N}_d(m, \Gamma)$ .

## 7.5. Densité d'un vecteur gaussien

**Théorème 7.6.** Soit  $X \sim \mathcal{N}_d(m, \Gamma)$ . On suppose  $\Gamma$  inversible. Alors  $X$  admet une densité  $f_X(x)$  par rapport à  $\lambda_d$ , donnée par

$$\frac{1}{\sqrt{2\pi}^d} \frac{1}{\sqrt{\det(\Gamma)}} \exp(-1/2 (x - m)^T \Gamma^{-1} (x - m)) .$$

*Démonstration.* Posons  $Z = \sqrt{\Gamma}^{-1}(X - m)$ . Alors  $Z \sim \mathcal{N}_d(0, I_d)$ . Autrement dit, les composantes  $Z_1, \dots, Z_d$  du vecteur  $Z$  sont des v.a. gaussiennes centrées réduites  $\mathcal{N}(0, 1)$  indépendantes. La densité du vecteur  $Z$  est donc le produit des densités des  $Z_i$ , soit pour tout  $z \in \mathbb{R}^d$ ,

$$f_Z(z) = \prod_{i=1}^d \frac{1}{\sqrt{2\pi}} e^{-\frac{t_i^2}{2}} = \frac{1}{\sqrt{2\pi}^d} e^{-t^T t / 2} . \quad (7.3)$$

Puisque  $X = \sqrt{\Gamma}Z + m$ , la formule du changement de variable (voir corollaire 4.18) implique que  $X$  admet une densité, donnée par

$$f_X(x) = \frac{f_Z(\sqrt{\Gamma}^{-1}(x - m))}{\det \sqrt{\Gamma}} .$$

Il suffit de remplacer  $f_Z$  par son expression (7.3) pour obtenir la formule donnée dans le théorème.  $\square$

**Lorsque  $\Gamma$  est non-inversible.** La loi du vecteur gaussien n'admet pas de densité par rapport à  $\lambda_d$  (voir exercice 7.18).

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  $\sqrt{\Delta_{j,j}}$ .

## 7.6. Exercices

### Pour apprendre

#### EXERCICE 7.1.–

Soient  $X$  et  $Y$  deux variables aléatoires indépendantes de même loi normale  $\mathcal{N}(0, 1)$ . Soit  $\varepsilon$  une variable aléatoire indépendante de  $X$  et de  $Y$ , suivant la loi de Rademacher, c'est-à-dire :

$$\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}.$$

Lesquels des vecteurs suivants sont gaussiens ?

- |                         |                                       |
|-------------------------|---------------------------------------|
| 1. $(X, \varepsilon)$   | 5. $(\varepsilon X , \varepsilon Y )$ |
| 2. $(X, Y)$             | 6. $(X, X + Y)$                       |
| 3. $(X, \varepsilon X)$ | 7. $(X, X + \varepsilon Y)$           |
| 4. $(X, \varepsilon Y)$ | 8. $(X, \varepsilon X + Y)$           |

#### EXERCICE 7.2.–

Soient  $A \in \mathcal{M}_d(\mathbb{R})$  et  $\mathbf{b} \in \mathbb{R}^d$ . On se donne  $\mathbf{X} \leftrightarrow \mathcal{N}(\mathbf{m}, \Sigma)$  un vecteur gaussien d'espérance  $\mathbf{m}$  et de matrice de covariance  $\Sigma$  définie positive.

- Donner la loi de  $A\mathbf{X} + \mathbf{b}$
- En déduire la loi de  $\sqrt{\Sigma}^{-1}(\mathbf{X} - \mathbf{m})$ .

#### EXERCICE 7.3.–

Soit deux v.a. indépendantes  $X \sim \mathcal{N}(0, 1)$  et  $Y$  de loi  $dP_Y = \frac{1}{2}(\delta_{-1} + \delta_1)$ .

- Montrer que  $Z = YX$  est une v.a. gaussienne.
- Montrer que  $X$  et  $Z$  sont non corrélées.
- Si  $(X, Z)$  était un vecteur gaussien, quelle serait sa loi ?
- Calculer la loi de  $(X, Z)$ .
- Est-ce que  $(X, Z)$  est un vecteur gaussien ?
- Est-ce que  $X$  et  $Z$  sont indépendantes ?

#### EXERCICE 7.4.–

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 7.5.—

Soit  $X$  et  $Y$  deux gaussiennes centrées réduites indépendantes. Montrer que les v.a.  $X + Y$  et  $\sin(X - Y)$  sont indépendantes.

EXERCICE 7.6.—

(Loi du  $\chi^2$  - version changement de variable)

Soit  $d \in \mathbb{N}^*$  et soit  $\mathbf{X}$  un vecteur aléatoire suivant la loi  $\mathcal{N}(\mathbf{0}_R^d, I_d)$ . De plus, soit

$$\psi : \begin{array}{ccc} \mathbb{R}^d & \rightarrow & \mathbb{R}^d \\ \mathbf{x} = (x_1, x_2, \dots, x_d) & \mapsto & (\|\mathbf{x}\|^2, x_2, \dots, x_d). \end{array}$$

1. Donner l'espace d'arrivée de  $\psi$ . Est-elle injective ? Pouvez-vous y remédier ?
2. Donner la loi de  $Y = \|\mathbf{X}\|^2$  en utilisant  $\psi$ . On **admettra** que :

$$\forall R \in \mathbb{R}_+, \int_{\{y_1^2 + \dots + y_{d-1}^2 \leq R\}} \frac{1}{\sqrt{R - y_1^2 - \dots - y_{d-1}^2}} dy_1 \dots dy_{d-1} = \frac{\pi^{\frac{d}{2}} R^{\frac{d}{2}-1}}{\Gamma(\frac{d}{2})}.$$

EXERCICE 7.7.—

(Loi du  $\chi^2$  - version probabiliste)

Soient  $\lambda, a$  deux réels strictement positifs. On dit qu'une variable aléatoire suit la loi Gamma  $\Gamma(\lambda, a)$  lorsqu'elle admet la densité :

$$\forall x \in \mathbb{R}, \frac{\lambda^a x^{a-1} e^{-\lambda x}}{\Gamma(a)} \mathbf{1}_{\mathbb{R}_+^*}(x).$$

On admet<sup>2</sup> que si  $X$  et  $Y$  sont deux variables aléatoires indépendantes et suivant respectivement la loi  $\Gamma(\lambda, a)$  et  $\Gamma(\lambda, b)$ , avec  $a, b > 0$ , alors  $X + Y$  suit la loi  $\Gamma(\lambda, a + b)$ .

1. Soit  $Z \hookrightarrow \mathcal{N}(0, 1)$ . Quelle est la loi de  $Z^2$  ?
2. Soit  $\mathbf{X} \hookrightarrow \mathcal{N}(\mathbf{0}_{\mathbb{R}^d}, I_d)$ , avec  $d \in \mathbb{N}^*$ . Dédurre de la question précédente une densité de  $\|\mathbf{X}\|_2^2$ .

EXERCICE 7.8.—

Soit  $(X, Y)$  un couple de v.a.r. de densité

$$f_{(X,Y)}(x, y) = C \exp\left(-\frac{\{3x^2 + 2xy + 3y^2\}}{8}\right).$$

---

2. Cela se prouve *via* la fonction caractéristique (qu'il faut donc calculer!) ou le changement de variable contre-intuitif  $(x, y) \mapsto (x + y, \frac{x}{x+y})$

1. Montrer que  $(X, Y)$  est un vecteur gaussien, de moyenne  $\mathbf{m} = \mathbf{0}_{\mathbb{R}^2}$  et de matrice de covariance

$$\Sigma = \frac{1}{2} \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}$$

2. Déterminer la loi de  $X$ , de  $Y$ . Quelle est la valeur de  $\text{Cov}(X, Y)$ ? Les v.a.  $(X, Y)$  sont-elles indépendantes?
3. On pose  $U = X + Y$  et  $V = X - Y$ . Montrer que le couple  $(U, V)$  est gaussien, et caractériser sa loi. Les v.a.r.  $U$  et  $V$  sont-elles indépendantes?

**EXERCICE 7.9.** –

Soient  $X_1, X_2, X_3$  des v.a. iid de loi  $\mathcal{N}(0, 1)$ . On pose

$$U = X_1 - X_2 + X_3 \quad T_1 = X_1 + X_2 \quad T_2 = X_2 + X_3 \quad T_3 = X_1 - X_3.$$

1. Quelle est la loi de  $U$ ?
2. Montrer que  $U$  est indépendant du vecteur  $(T_1, T_2, T_3)$ , et indépendant de  $T_i$ ,  $i \in \{1, 2, 3\}$ .
3. On pose  $V = T_1^2 + T_2^2 + T_3^2$ .
  - a) Ecrire le vecteur  $T = (T_1, T_2, T_3)$  sous la forme  $A(X_1, X_2, X_3)^T$  pour une matrice  $A$  de taille  $3 \times 3$ .
  - b) Calculer les valeurs propres de  $A^T A$ .
  - c) En déduire la loi de  $V/3$ .
4. Quelle est la loi du couple  $(U, V)$ ?

**EXERCICE 7.10.** –

Soit  $X = (X_1, X_2)^T \sim \mathcal{N}(0, I_2)$  où  $I_2$  est l'identité de taille  $2 \times 2$ .

1. On pose  $U = 1 + X_1 + 2X_2$ . Quelle est la loi de  $U$ ?
2. Caractériser la loi du vecteur  $Y = AX$  où

$$A = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

3. Le vecteur  $Y$  admet-il une densité?

**EXERCICE 7.11.** –

**(Vecteur gaussien et espérance conditionnelle).** Soit  $(X_1, X_2, X_3)^T$  un vecteur gaussien de moyenne  $\mathbf{m} = (1, 0, -1)^T$  et de covariance

$$\Sigma = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}$$

1. Quel est la loi de  $(X_1, X_2)$ ?
2. Déterminer un réel  $\alpha$  tel que la v.a.  $Y = \alpha X_1 + X_2$  soit indépendante de  $X_1$ . Que vaut  $\mathbb{E}(Y)$ ? Que vaut sa variance?
3. En déduire  $\mathbb{E}(X_2|X_1)$ . Quelle est la loi conditionnelle de  $X_2$  sachant  $X_1$ ?

4. Déterminer un réel  $\beta$  tel que la v.a.  $Z = \beta X_1 + X_3$  soit indépendante de  $X_1$ . En déduire  $\mathbb{E}(X_3|X_1)$  et  $\mathbb{E}(X_3^2|X_1)$ .
5. Calculer  $\mathbb{E}(X_1^2 X_2 + X_3^2 X_1|X_1)$ .

EXERCICE 7.12 (Sphere hardening).—

Soit  $X_N$  un vecteur gaussien de  $\mathbb{R}^N$ , centré, réduit. Soit  $\|X_N\|$ , la norme euclidienne de  $X_N$  et  $X'_N = \|X_N\|/\sqrt{N}$ .

1. Calculer  $\mathbb{E}((X'_N)^2)$ .
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 7.13.—

Soit  $E$  un espace vectoriel de dimension  $d$ . On suppose qu'il se décompose en la somme directe orthogonale :

$$E = \bigoplus_{i=1}^p E_i,$$

avec  $p \geq 1$  et où les  $E_i$  sont des sous-espaces de  $E$ , respectivement de dimension  $d_i$ . Soit  $\mathbf{X}$  un vecteur gaussien de vecteur espérance  $\mathbf{m}$  et de matrice de covariance  $I_d$ . Alors il est clair qu'il existe  $\mathbf{X}_1, \dots, \mathbf{X}_p$ , avec  $\mathbf{X}_i \in E_i$ , tels que presque sûrement :

$$\mathbf{X} = \sum_{i=1}^p \mathbf{X}_i.$$

1. D'où provient cette décomposition ? Et pourquoi les  $\mathbf{X}_i$  sont-ils bien définis ?
2. Prouver que les  $\mathbf{X}_i$  sont des vecteurs gaussiens indépendants dont on donnera les paramètres.
3. (**Application**) Soit  $\mathbf{X} = (X_1, \dots, X_n)$  un vecteur gaussien de loi  $\mathcal{N}(\mathbf{m}, \sigma^2 I_n)$ , avec  $\sigma^2 > 0$ . On pose :

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \text{ et } \hat{s}_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Prouver que  $\bar{X}_n$  et  $\hat{s}_n^2$  sont indépendantes<sup>3</sup>.

- 
3. **Indice** : Considérer le vecteur  $\mathbf{1} := (1, \dots, 1)^T \in \mathbb{R}^n$ .

REMARQUE.— Le résultat de la question 2. porte le nom de théorème de Cochran (à prononcer "Kokran") et joue un rôle important en statistiques, en particulier en régression linéaire. La question 3. en est une illustration.

EXERCICE 7.14.—

Soit  $\mathbf{X} = (X_1, \dots, X_n)$  un vecteur gaussien standard de  $\mathbb{R}^n$  (les  $X_i$  sont i.i.d. de loi normale  $\mathcal{N}(0, 1)$ ).

1. On note  $\mathbf{O}(n)$  le groupe orthogonal de  $\mathbb{R}^n$ . Montrer que la loi de  $\mathbf{X}$  est invariante par l'action à gauche sur  $\mathbb{R}^n$  de  $\mathbf{O}(n)$ , c'est-à-dire que :

$$\forall O \in \mathbf{O}(n), O\mathbf{X} \stackrel{(\mathcal{L})}{=} \mathbf{X}.$$

2. On note  $\nu$  la loi de  $Y := \mathbf{X}/\|\mathbf{X}\|_2$ .

- a) Quel est le support de  $Y$  ?
- b) Montrer que  $\nu$  est invariante par l'action à gauche de  $\mathbf{O}(n)$  sur  $\mathbb{R}^n$ , *i.e.* que :

$$O\mathbf{Y} \stackrel{(\mathcal{L})}{=} \mathbf{Y}$$

3. On souhaite maintenant montrer que  $\|\mathbf{X}\|_2$  et  $\mathbf{X}/\|\mathbf{X}\|_2$  sont indépendantes.

- a) Soit  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction-test, disons mesurable bornée. On définit une nouvelle mesure (signée) sur l'hypersphère  $\mathbb{S}^{n-1}$  par :

$$\forall B \in \mathcal{B}(\mathbb{S}^{n-1}), \mu_\varphi(B) := \mathbb{E}\left[\varphi(\|\mathbf{X}\|_2)\mathbf{1}_B\left(\frac{\mathbf{X}}{\|\mathbf{X}\|_2}\right)\right].$$

Montrer que la mesure  $\mu_\varphi$  est invariante par l'action à gauche de  $\mathbf{O}(n)$  sur  $\mathbb{R}^n$  *i.e.*

$$\forall B \in \mathcal{B}(\mathbb{S}^{n-1}), \mu_\varphi(O^{-1}B) = \mu_\varphi(B).$$

- b) On admet<sup>4</sup> que toute mesure sur  $\mathbb{S}^{n-1}$  invariante par l'action à gauche de  $\mathbf{O}(n)$  est un multiple de  $\sigma^{n-1}$ , la probabilité uniforme sur l'hypersphère  $\mathbb{S}^{n-1}$ . En déduire que  $\|\mathbf{X}\|_2$  et  $\mathbf{X}/\|\mathbf{X}\|_2$  sont indépendantes.

EXERCICE 7.15.—

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

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \Sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$$

et

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

4. C'est une conséquence du *lemme de Christensen*. Voilà. Vous vous coucherez moins sots ce soir. Moi aussi d'ailleurs.

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 7.16.—

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  $s \in \mathbb{R}^n$ .

EXERCICE 7.17.—

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 7.18.—

Soit  $X$  un vecteur gaussien sur  $\mathbb{R}^d$  de moyenne  $m$  et de matrice de covariance  $\Gamma$ .

1. Montrer que  $\mathbb{P}(X - m \in \text{Im}(\Gamma)) = 1$  où  $\text{Im}(\Gamma)$  est l'image de la matrice  $\Gamma$ .
2. Montrer que si  $E \subset \mathbb{R}^d$  est n'importe quel sous-espace vectoriel de  $\mathbb{R}^d$  de dimension strictement inférieure à  $d$ , alors  $\lambda_d(E) = 0$ .
3. Dédire des deux questions précédentes que  $X$  n'admet pas de densité par rapport à  $\lambda_d$ .

EXERCICE 7.19 (Polynômes d'Hermite).—

Soit  $X$  une v.a.r. gaussienne centrée, réduite et  $\varphi(t, x) = \exp(tx)$ .

1. Trouver  $g(t)$  telle que  $g(t)\mathbb{E}(\varphi(t, X)) = 1$ .

2. On pose

$$\psi(t, x) = g(t)\varphi(t, x).$$

Montrer que

$$\mathbb{E}(\psi(t, X)\psi(s, X)) = \exp(\sigma^2 ts).$$

3. Montrer que

$$\psi(t, x) = \sum_{n=0}^{\infty} \left( \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{x^{n-2k}}{(n-2k)!} \frac{(-\sigma^2)^k}{2^k k!} \right) t^n.$$

4. On pose

$$P_n(x) = \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{x^{n-2k}}{(n-2k)!} \frac{(-\sigma^2)^k}{2^k k!}.$$

Montrer que

$$\mathbb{E}(P_n(X)P_m(X)) = \delta_{n,m}.$$

## 8. Convergences

Les preuves des grands théorèmes de ce chapitre seront vues en deuxième année (MACS203). Vue l'importance de ces résultats en probabilité, il est toutefois important que l'élève en connaisse les énoncés et sache les appliquer.

On fixe un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ .

### 8.1. Loi des grands nombres

**Définition 8.1.** On dit qu'une suite  $(X_n, n \geq 1)$  de v.a. sur  $\mathbb{R}^d$  converge  $\mathbb{P}$ -presque-sûrement (ou  $\mathbb{P}$ -presque-partout) vers une v.a.  $X$  lorsqu'il existe un ensemble  $A$  tel que  $\mathbb{P}(A^c) = 0$  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.

Puisque la composition par une fonction continue ne change pas la convergence d'une suite, nous avons le résultat immédiat suivant.

**Proposition 8.1.** Si  $f$  est une fonction continue et si  $(X_n)$  converge presque sûrement vers  $X$ , alors  $f(X_n)$  converge presque sûrement vers  $f(X)$ .

**Théorème 8.2** (Loi forte des grands nombres). Soit  $(X_n, n \geq 1)$  une suite de v.a. sur  $\mathbb{R}^d$  indépendantes, identiquement distribuées telles que  $\mathbb{E}(\|X_1\|) < \infty$  alors

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{n \rightarrow +\infty} \mathbb{E}(X_1), \mathbb{P} - \text{p.p.}$$

EXEMPLE 28.– On souhaite évaluer empiriquement la probabilité de succès d'un événement  $A$ . On réalise pour cela  $n$  expériences indépendantes. La probabilité *empirique* que  $A$  soit réalisé est donnée par

$$\hat{p}_n = \frac{\text{nombre de fois où } A \text{ est réalisé}}{n}.$$

En notant  $X_i$  la v.a. de Bernoulli égale à 1 si  $A$  est réalisé et zéro sinon, on obtient  $\hat{p}_n = \frac{1}{n} \sum_i X_i$ . Comme  $\mathbb{E}(X_1) = \mathbb{P}(A)$ , la loi des grands nombres implique que  $\hat{p}_n$  converge presque sûrement vers  $\mathbb{P}(A)$ . On dit que l'estimateur  $\hat{p}_n$  est *consistant*. La probabilité empirique se rapproche de la probabilité lorsque le nombre d'expériences augmente.

## 8.2. Théorème central limite

**Définition 8.2.** Soient  $(X_n), X$  des v.a. sur  $\mathbb{R}^d$  de fonctions de répartition  $F_{X_n}$  et  $F_X$ . On dit que  $X_n$  converge en loi vers  $X$  si

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

pour tout  $x$  point de continuité de  $F_X$ .

La convergence en loi est donc la convergence simple des fonctions de répartition (en tout point hormis les discontinuités). L'exemple suivant montre pourquoi, dans la définition, on a jugé utile d'exclure les points de discontinuité.

REMARQUE.— Soit  $X_n(\omega) = \frac{1}{n}$ . Cette suite converge en loi vers  $X = 0$ . En effet, la fonction de répartition de  $X$  est  $\mathbf{1}_{[0, +\infty)}$ . La fonction de répartition de  $X_n$  est  $\mathbf{1}_{[\frac{1}{n}, +\infty)}$ , elle converge simplement vers la fonction  $\mathbf{1}_{(0, +\infty)}$ . Cette limite n'est *pas* la fonction de répartition de  $X$ . Toutefois, on a bien  $\mathbf{1}_{[\frac{1}{n}, +\infty)}(x) \rightarrow \mathbf{1}_{[0, +\infty)}(x)$  et tout point  $x \neq 0$ , ce qui correspond bien à la définition de la convergence en loi.

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..

**Définition 8.3.** Pour un ensemble ouvert  $A \in \mathbb{R}^k$ , 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  $\mathbb{R}^k$ , la frontière au sens topologique correspond à la notion intuitive de bord.

Nous avons les caractérisations suivantes de la convergence en loi.

**Théorème 8.3.** Soient  $X, (X_n, n \geq 1)$  des v.a., à valeurs dans  $\mathbb{R}^d$ . Les propositions suivantes sont équivalentes.

- $X_n$  converge en loi vers  $X$ ,
- Pour toute fonction continue bornée  $f$  de  $\mathbb{R}^d$  dans  $\mathbb{R}$ ,

$$\mathbb{E}(f(X_n)) \xrightarrow{n \rightarrow +\infty} \mathbb{E}(f(X)),$$

- pour tout ensemble  $A \in \mathcal{B}(\mathbb{R}^d)$  tel que  $\mathbb{P}(X \in \partial A) = 0$ ,

$$\mathbb{P}(X_n \in A) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(X \in A),$$

- pour tout  $t \in \mathbb{R}^d$ ,

$$\mathbb{E}(e^{i\langle t, X_n \rangle}) \xrightarrow{n \rightarrow +\infty} \mathbb{E}(e^{i\langle t, X \rangle}).$$

Le plus souvent, on accède à la loi de  $X_n$  non par la fonction de répartition, mais par la densité (par rapport à la mesure de Lebesgue dans le cas continu, ou par rapport à la mesure de comptage dans le cas discret). On a alors le critère pratique suivant, qui dit en substance que la convergence en loi revient à la convergence des densités.

**Proposition 8.4.** Soit  $X_n$  une suite de v.a. de densité  $f_{X_n}$  par rapport à une mesure de référence  $\mu$ . Soit  $X$  une v.a. de densité  $f_X$  par rapport à  $\mu$ . On suppose que  $f_{X_n} \leq g$  avec  $\int g d\mu < \infty$  et que pour tout  $x$ ,

$$\lim_{n \rightarrow \infty} f_{X_n}(x) = f_X(x).$$

Alors  $X_n$  converge en loi vers  $X$ .

*Démonstration.* Application immédiate du théorème de convergence dominée.  $\square$

Par exemple, si  $X_n$  et  $X$  sont des v.a. discrètes (disons sur  $\mathbb{N}$ ), la convergence en loi de  $X_n$  vers  $X$  est équivalente à

$$\forall k \in \mathbb{N}, \lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k).$$

EXEMPLE 29.— Soit  $X_n$  une suite de v.a. de loi binomiale de paramètres  $(n, \frac{\nu}{n})$ , où  $\nu > 0$  :

$$\mathbb{P}(X_n = k) = \binom{n}{k} (\nu/n)^k (1 - \nu/n)^{n-k} \rightarrow \frac{\nu^k}{k!} e^{-\nu}.$$

Donc  $X_n$  converge en loi vers une loi de Poisson de paramètre  $\nu$ .

**Théorème 8.5.** La convergence presque sûre implique la convergence en loi mais la réciproque est fautive.

*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  $\mathbb{E}(\|f\|_\infty) < \infty$ ,

donc toutes les hypothèses du théorème de convergence dominée sont satisfaites, d'où

$$\mathbb{E}(f(X_n)) \xrightarrow{n \rightarrow +\infty} \mathbb{E}(f(X)).$$

D'après la 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.  $\square$

Comme pour la convergence presque-sûre, la composition par une fonction continue ne modifie pas la convergence en loi.

**Proposition 8.6.** Soit  $(X_n)$  une suite de v.a. sur  $\mathbb{R}^d$  convergeant en loi vers  $X$ . Soit  $f : \mathbb{R}^d \rightarrow \mathbb{R}^m$  une fonction continue. Alors  $f(X_n)$  converge en loi vers  $f(X)$ .

*Démonstration.* Soit  $h$  une fonction continue bornée quelconque. Comme  $h \circ f$  est continue, on a par la caractérisation de la convergence en loi,  $\mathbb{E}(h \circ f(X_n)) \rightarrow \mathbb{E}(h \circ f(X))$ . En posant  $Y_n = f(X_n)$  et  $Y = f(X)$ , cela se lit  $\mathbb{E}(h(Y_n)) \rightarrow \mathbb{E}(h(Y))$  pour tout  $h$  continue bornée. Donc  $Y_n$  converge en loi vers  $Y$ .  $\square$

**Théorème 8.7.** Soit  $(X_n, n \geq 1)$  une suite de v.a. sur  $\mathbb{R}^d$  indépendantes, identiquement distribuées telles que  $\mathbb{E}(\|X_1\|^2) < \infty$ . On suppose que les  $X_n$  sont *centrées* :  $\mathbb{E}(X_1) = 0$ . Soit  $\Gamma$  la matrice de covariance de  $X_1$ . Alors,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \Gamma), \text{ en loi}$$

EXEMPLE 30.– Considérons à nouveau l'estimateur  $\hat{p}_n$  de la probabilité d'un événement  $A$ , dans l'exemple 28. La loi des grands nombres nous dit que l'erreur d'estimation  $\hat{p}_n - \mathbb{P}(A)$  converge p.s. vers zéro. Le théorème central limite nous donne une information supplémentaire sur les fluctuations de cette erreur. Nous avons

$$\sqrt{n}(\hat{p}_n - \mathbb{P}(A)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mathbb{P}(A))$$

où  $X_i$  est définie dans l'exemple 28. La suite de v.a.  $X_n - \mathbb{P}(A)$  est iid centrée, de variance  $\mathbb{P}(A)(1 - \mathbb{P}(A))$ . Par le théorème central limite,

$$\sqrt{n}(\hat{p}_n - \mathbb{P}(A)) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \mathbb{P}(A)(1 - \mathbb{P}(A))), \text{ en loi.}$$

Cela signifie que dans le régime où  $n$  est grand, l'erreur d'estimation se comporte comme une v.a. gaussienne. On observe aussi que la vitesse de convergence est de l'ordre de  $\frac{1}{\sqrt{n}}$ . Enfin, la variance de l'erreur est la plus faible lorsque  $\mathbb{P}(A)$  est de l'ordre de 0.5, elle est plus grande lorsque  $\mathbb{P}(A)$  se rapproche de 0 ou de 1.

## 8.3. Exercices

### Pour apprendre

EXERCICE 8.1.—

Soit  $X_j \sim 0.5(\delta_{1/j} + \delta_{-1/j})$ . Montrer que  $(X_j)_j$  converge en loi. Soit  $F$  la fonction répartition de cette loi limite. Vérifier que  $F$  est continue sauf en  $t = 0$  et que  $\lim_j F_{X_j}(t) = F(t)$  en tout  $t \neq 0$ .

EXERCICE 8.2.—

Soient  $(a_j)_j$  et  $(p_j)_j$  des suites de réels strictement positifs tels que

$$a_j \geq 1 \quad 0 < p_j < 1 \quad \lim_j p_j = 0 .$$

Soit  $(X_j)_j$  une suite de v.a. indépendantes de loi

$$\mathbf{P}[X_j = 0] = 1 - p_j \quad \mathbf{P}[X_j = a_j] = p_j .$$

Montrer que  $X_j \xrightarrow{\mathbf{P}} 0$ ;  $X_j \xrightarrow{\text{p.s.}} 0$  ssi  $\sum_j p_j < +\infty$ ;  $X_j \xrightarrow{L^1} 0$  ssi  $\lim_j a_j p_j = 0$ ;  $X_j \xrightarrow{L^2} 0$  ssi  $\lim_j a_j^2 p_j = 0$ . En déduire des exemples de suites  $(a_j, p_j)_j$  pour lesquelles on a : convergence presque-sûre mais pas convergence dans  $L^1$ ; convergence dans  $L^1$  mais pas convergence presque-sûre; convergence dans  $L^1$  mais pas convergence dans  $L^2$ .

EXERCICE 8.3.—

Soit  $(X_n)_n$  des v.a. à valeur dans  $\mathbb{N}$  et  $X$  une v.a. à valeur dans  $\mathbb{N}$ . Montrer que  $X_n \xrightarrow{\mathcal{L}} X$  ssi pour tout  $k \in \mathbb{N}$ ,  $\lim_n \mathbf{P}[X_n = k] = \mathbf{P}[X = k]$ .

EXERCICE 8.4.—

On appelle *quantile d'ordre  $\alpha$*  d'une v.a.r.  $Z$  (à densité sur  $\mathbb{R}$ ), le réel  $c_\alpha$  tel que  $\mathbb{P}(Z \leq c) = \alpha$ . Le quantile d'ordre 0.5 est la *médiane*.

Une agence de voyage dispose de 100 places sur un vol Paris/ New-York. Pour tenir compte des éventuels désistements, elle décide d'accepter 120 réservations. La probabilité pour qu'un passager ayant réservé, se présente à l'embarquement est 0.8 et les passagers sont supposés indépendants.

1. Quelle est la loi de la v.a.  $S$  égale au nombre de passagers, qui, ayant réservé leur place par l'agence, se présenteront effectivement à l'embarquement ?
2. En utilisant le TCL, proposer une approximation de la forme  $\int_\alpha^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-0.5x^2) dx$  de la probabilité que le nombre de passagers présents à l'embarquement soit supérieur à 100.
3. Combien de réservations au maximum l'agence aurait-elle dû accepter pour que cette probabilité soit inférieure à 0.01 : on exprimera ce nombre en fonction d'un quantile (préciser l'ordre) de la loi  $\mathcal{N}(0, 1)$ .

### Pour s'entraîner

EXERCICE 8.5.—

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  $T_0 = 0$ .

- 8.5.a) Montrer que  $T_{i+1} - T_i$  suit une loi géométrique de paramètre  $1 - i/N$ .
- 8.5.b) Montrer que les variables aléatoires  $T_0, T_1 - T_0, \dots, T_N - T_{N-1}$  sont indépendantes dans leur ensemble.
- 8.5.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  $+\infty$ .
- 8.5.d) En utilisant l'inégalité de Bienaymé-Tchebychev, montrer que  $T_N/(N \log N)$  tend vers 1 en probabilité.

EXERCICE 8.6. –

On appelle loi de Cauchy la loi sur  $\mathbb{R}$  de densité

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

On rappelle la formule :  $\forall x > 0, \arctan(x) + \arctan\left(\frac{1}{x}\right) = \frac{\pi}{2}$ .

1. Démontrer que si  $Y$  et  $Z$  sont deux variables indépendantes gaussiennes centrées réduites (c'est à dire de variance 1) alors  $Y/Z$  suit la loi de Cauchy.
2. Exprimer la fonction de répartition correspondante.
3. Soient  $X_1, \dots, X_n$  des variables aléatoires indépendantes suivant la loi de Cauchy. Montrer que la variable

$$\frac{\pi}{n} \max(X_1, \dots, X_n)$$

converge en loi vers une variable aléatoire que l'on caractérisera.

EXERCICE 8.7. –

Soit  $(X_j)_j$  une famille de v.a. telles que pour tout  $j \geq 1$ ,

$$\mathbf{P}[X_j = k/j] = 1/j \quad k \in \{1, \dots, j\}.$$

Montrer que  $(X_j)_j$  converge en loi vers une v.a. que l'on précisera.

EXERCICE 8.8. –

On définit des intervalles de  $[0, 1]$  par :

$$I_{2^m+k} = \left[ \frac{k}{2^m}, \frac{k+1}{2^m} \right] \quad m \geq 0, k \in \{0, \dots, 2^m - 1\}.$$

Soit  $Y$  une v.a. uniforme sur  $[0, 1]$ . On définit la suite de v.a.  $(X_j)_j$  par

$$X_j = \begin{cases} 1 & \text{si } Y \in I_j \\ 0 & \text{sinon} \end{cases}.$$

Montrer que  $X_j \xrightarrow{\mathcal{L}} \delta_0$  mais ne converge pas presque-sûrement vers la v.a.  $X = 0$ .

EXERCICE 8.9.—

Soit  $\{U_n, V_n, n \geq 1\}$  des variables aléatoires i.i.d., définies sur le même espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ , de même loi  $\mathcal{U}([0, 1])$ .

On pose

$$X_n := \begin{cases} 4, & \text{si } U_n^2 + V_n^2 \leq 1, \\ 0, & \text{sinon.} \end{cases} \quad Z_n := \frac{1}{n}(X_1 + \dots + X_n) = \frac{4}{n} \sum_{k=1}^n \mathbf{1}_{U_k^2 + V_k^2 \leq 1}.$$

1. Quelle est la loi de  $X_n$  ?
2. Montrer que  $\{Z_n, n \geq 1\}$  converge p.s. vers  $\pi$ .
3. Expliquez comment estimer  $\pi$  à partir de la réalisation  $(u_1, v_1, \dots, u_n, v_n)$  du vecteur aléatoire  $(U_1, V_1, \dots, U_n, V_n)$ .
4. Soit  $\alpha \in ]0, 1[$ . Quelle est l'espérance et la variance de  $Z_n$  ? En utilisant l'inégalité de Tchebichev, montrer que pour tout  $n$ ,

$$\mathbb{P} \left( |Z_n - \pi| \leq \sqrt{\frac{\pi(4 - \pi)}{(1 - \alpha)n}} \right) \geq \alpha.$$

En déduire  $n_0$  tel que l'erreur d'approximation est inférieure à  $\epsilon$  avec une probabilité au moins égale à  $\alpha$ . A.N. lorsque  $\alpha = 0.95$  et  $\epsilon = 10^{-3}$ .

5. En utilisant l'approximation donnée par le Théorème Central Limite, trouver  $n_0$  tel que l'erreur d'approximation est inférieure à  $\epsilon$  avec une probabilité au moins égale à  $\alpha$ . Application numérique lorsque  $\alpha = 0.95$  et  $\epsilon = 10^{-3}$ .

### Pour aller plus loin

EXERCICE 8.10.— 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  $\mathbb{R}^+$ .
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

$$f_{\beta, \lambda, \mu}(x) = K x^\beta e^{-\lambda x^\mu} \mathbf{1}_{\mathbb{R}^+}(x).$$

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

$$\frac{1}{\sqrt{n}} (S_n - \bar{S}).$$

# A. Annexes

## A.1. Rappels sur les ensembles

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  $A \cap B = \emptyset$ .

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  $\mathbb{N}$ . Les ensembles  $\mathbb{N}$ ,  $\mathbb{N}^*$ ,  $\mathbb{Z}$  sont dénombrables. L'ensemble  $\mathbb{N} \times \mathbb{N}$  est dénombrable (voir la figure A.1) et par conséquent, l'ensemble  $\mathbb{Q}$  des rationnels est dénombrable. 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  $\cup_{i \in I} A_i = \Omega$ .

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.

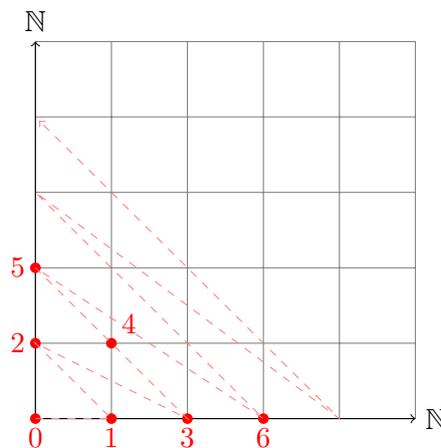


Figure A.1. – Bijection de  $\mathbb{N} \times \mathbb{N}$  avec  $\mathbb{N}$ .

## A.2. Rappels 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  $S < \infty$ .

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 :

$$\begin{aligned} \sum_{n \geq 0} n x^n &= x \sum_{n \geq 1} n x^{n-1} = \frac{x}{(1-x)^2} , \\ \sum_{n \geq 0} n^2 x^n &= x^2 \sum_{n \geq 2} n(n-1) x^{n-2} + x \sum_{n \geq 1} n x^{n-1} , \\ &= \frac{2x^2}{(1-x)^3} + \frac{x}{(1-x)^2} = \frac{x(x+1)}{(1-x)^3} . \end{aligned}$$

### A.3. Limite supérieure et limite inférieure

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  $\overline{\mathbb{R}}$  la droite réelle complétée :  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ .

**Définition A.3.** La limite inférieure et la limite supérieure d'une suite numérique  $(u_n, n \in \mathbb{N})$  sont les éléments de  $\overline{\mathbb{R}}$ , 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  $\overline{\mathbb{R}}$  : 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  $\ell \in \overline{\mathbb{R}}$  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 A.3.** 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 A.4.** Si  $\limsup_{n \rightarrow \infty} u_n = \liminf_{n \rightarrow \infty} u_n = \ell \in \overline{\mathbb{R}}$ , alors la suite  $(u_n)$  converge vers  $\ell$ .

### A.4. Fonctions convexes

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 A.4.** Une fonction d'un ensemble convexe  $C \subset X$  à valeurs  $\overline{\mathbb{R}}$  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.

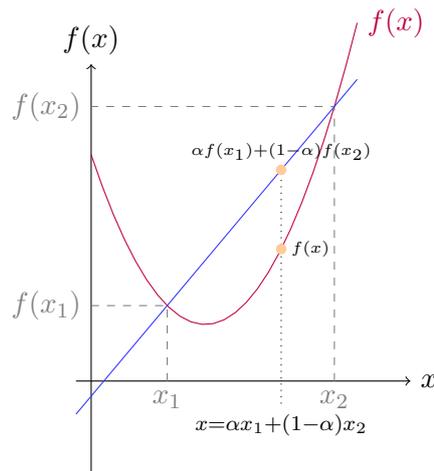


Figure A.2. – Illustration de la propriété de convexité.

EXEMPLE 31.– 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 ont de bonnes propriétés de régularité, en particulier, elles admettent en tout point une dérivée à gauche et une dérivée à droite.

## A.5. Théorème des classes monotones et conséquences\*

### A.5.1. Énoncé

**Définition A.5.** Un  $\lambda$ -système (ou classe monotone)  $\mathcal{L}$  est une classe de sous-ensembles de  $\Omega$  vérifiant :

A.5.i)  $\Omega \in \mathcal{L}$  ;

A.5.ii) Pour tout  $A \in \mathcal{L}$ ,  $A^c \in \mathcal{L}$  ;

A.5.iii) Pour toute suite  $(A_n)_{n \in \mathbb{N}^*}$  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. La propriété est donc plus facile à vérifier. 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.

**Théorème A.5.** Soit  $\mathcal{P}$  un  $\pi$ -système,  $\mathcal{L}$  un  $\lambda$ -système. Si  $\mathcal{P} \subset \mathcal{L}$  alors  $\sigma(\mathcal{P}) \subset \mathcal{L}$ .

*Démonstration.* A compléter. Voir [?, Théorème 1.1]. □

### A.5.2. Preuve du théorème 2.9

On rappelle l'énoncé du théorème 2.9. Soit  $\mu, \nu$  deux mesures sur  $(\Omega, \mathcal{F})$ . Soit  $\mathcal{P}$  un  $\pi$ -système engendrant  $\mathcal{F}$  i.e.,  $\mathcal{F} = \sigma(\mathcal{P})$ . Supposons que l'une des deux conditions suivantes est satisfaite :

1.  $\mu, \nu$  sont deux mesures de probabilité ;
2.  $\mu, \nu$  sont  $\sigma$ -finies sur  $\mathcal{P}$ .

Alors  $\mu = \nu$ .

Donnons d'abord la preuve pour des mesures de probabilité. Posons

$$\mathcal{L} = \{A \in \mathcal{F} : \mu(A) = \nu(A)\}.$$

On peut vérifier que  $\mathcal{L}$  est un  $\lambda$ -système. En effet,

- A.5.i) Par hypothèse  $\mu(\Omega) = \nu(\Omega)$  ;
- A.5.ii)  $\mu(A^c) = \mu(\Omega) - \mu(A) = \nu(\Omega) - \nu(A) = \nu(A^c)$ , donc  $A^c \in \mathcal{L}$  (on voit ici que le caractère fini des mesures intervient) ;
- A.5.iii) Si les  $(A_n, n \in \mathbb{N})$  sont deux à deux disjoints et appartiennent tous à  $\mathcal{L}$  alors

$$\mu\left(\bigcup_n A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \nu(A_n) = \nu\left(\bigcup_n A_n\right),$$

donc la réunion  $\cup_n A_n$  appartient à  $\mathcal{L}$ .

Nous avons par hypothèse que  $\mathcal{P} \subset \mathcal{L}$ . Par le théorème A.5,  $\sigma(\mathcal{P}) \subset \mathcal{L}$ . Ceci conclut la preuve.

Dans le cas où les mesures sont  $\sigma$ -finies sur  $\mathcal{P}$ , on choisit  $B \in \mathcal{P}$  tel que  $\mu(B) = \nu(B) < \infty$ . Il suffit d'adapter la preuve précédente. On pose

$$\mathcal{L}_B = \{A \in \mathcal{F} : \mu(A \cap B) = \nu(A \cap B)\}.$$

Comme précédemment,  $\mathcal{L}_B$  est un  $\lambda$ -système contenant  $\mathcal{P}$ , donc qui contient  $\sigma(\mathcal{P})$ . On se donne maintenant une suite  $(B_n)$  sur  $\mathcal{P}$ , telle que  $\Omega = \bigcup_n B_n$  et  $\mu(B_n) = \nu(B_n) < \infty$ . D'après ce qui précède, pour tout  $A \in \mathcal{F}$ , on a  $\mu(A \cap B_n) = \nu(A \cap B_n)$ . Dans le cas où  $B_n$  est une suite croissante (par exemple  $B_n = [-n, n]$  pour  $\Omega = \mathbb{R}$ ), il suffit de passer à la limite et on obtient  $\mu(A) = \nu(A)$ . Dans le cas général, il faut effectuer un « découpage » des  $B_n$  dans le même esprit que dans la preuve de la proposition 1.1, et on aboutit à la même conclusion (la preuve est laissée au lecteur).

**A.5.3. Complément de preuve du théorème 3.25**

Le théorème 3.25 a été démontré dans le corps du texte, à l'exception d'un point technique : si  $G \in \mathcal{E} \otimes \mathcal{F}$  est un élément de la tribu produit sur  $E \times F$ , et si  $\nu$  est une mesure  $\sigma$ -finie sur  $\mathcal{F}$ , la fonction

$$f_G : x \mapsto \int \mathbf{1}_G(x, y) d\nu(y)$$

est  $\mathcal{E}$ -mesurable. Nous démontrons ce point ci-dessous.

Nous supposons  $\nu$  finie (la généralisation au cas  $\sigma$ -fini emploie la même idée que dans le paragraphe précédent). Soit  $\mathcal{L}$  la classe d'ensembles  $G \in \mathcal{E} \otimes \mathcal{F}$  tels que la fonction  $f_G$  est  $\mathcal{E}$ -mesurable. On montre que  $\mathcal{L}$  est un  $\lambda$ -système. i) Comme  $f_{E \times F}$  est une fonction constante égale à  $\nu(F)$ ,  $E \times F \in \mathcal{L}$ . ii) Soit  $G$  tel que  $f_G$  est  $\mathcal{E}$ -mesurable. On a  $\nu(F) = \nu\{y : (x, y) \in G\} + \nu\{y : (x, y) \in G^c\}$  et donc, puisque  $\nu(F)$  est finie,  $f_G = \nu(F) - f_{G^c}$ . Ainsi,  $f_G$  est mesurable comme différence de deux fonctions mesurables. iii) Si  $(G_n)_n$  est une famille d'ensembles deux à deux disjoints,  $f_{\cup_n G_n} = \sum_n f_{G_n}$  est mesurable comme limite d'une suite de fonctions mesurables. Nous avons bien montré que  $\mathcal{L}$  est un  $\lambda$ -système. Pour tout  $A \in \mathcal{E}$  et  $B \in \mathcal{F}$ ,  $f_{A \times B}(x) = \nu(B)\mathbf{1}_A(x)$  est mesurable. Donc  $\mathcal{L}$  contient le  $\pi$ -système formé par la classe  $\mathcal{P}$  des ensembles de la forme  $A \times B$ . Par le théorème A.5,  $\mathcal{L}$  contient  $\sigma(\mathcal{P}) = \mathcal{E} \otimes \mathcal{F}$ . Ainsi,  $f_G$  est mesurable pour tout  $G \in \mathcal{E} \otimes \mathcal{F}$ .



## B. Correction des exercices

### Exercice 1.1 ▷

1. On cherche le nombre de parties de  $\llbracket 1; 25 \rrbracket$ . Il y en a  $2^{\#\llbracket 1; 25 \rrbracket} = 2^{25}$ . On peut sinon faire du dénombrement : On commence par choisir la taille  $k$  du tiercé, donc de la partie. Ensuite, on choisit quels chevaux, donc nombres, feront partie du tiercé : On choisit  $k$  chevaux parmi 25, il y a  $\binom{25}{k}$  possibilités. Ainsi on trouve un nombre total de tiercé de  $\sum_{i=0}^{25} \binom{25}{k} = 2^{25}$ .
2. Pour chaque position de lettre, on choisit la lettre parmi les 26 de l'alphabet : Il y a ainsi  $26^5$  possibilités.
3. On choisit, comme dans la question précédente, les cartes les unes après les autres. Cependant cette fois le tirage précédent impacte le suivant, car il reste alors moins de cartes parmi lesquelles choisir. Ainsi le nombre de mains est de  $52 \times 51 \times 50 \times 49 \times 48 = \frac{52!}{47!}$ .
4. Dénombrons d'abord les mains avec deux rois. Il faut pour cela choisir deux rois parmi les quatres, soit  $\binom{4}{2} = 12$ , puis choisir deux cartes quelconques dans le paquet sans les rois. On obtient  $12 \times 48 \times 47$ . Le nombre de mains de quatres cartes est par la question précédente  $52 \times 51 \times 50 \times 49$ . Ainsi la probabilité de tirer une main avec exactement deux rois est  $\frac{12 \times 48 \times 47}{52 \times 51 \times 50 \times 49}$ . A simplifier...

**Exercice 1.3 ▷** On peut s'attendre avec de tels degrés de fiabilité à ce qu'un individu positif au test ait une forte probabilité d'être malade. On va voir qu'en fait, ce n'est pas si évident que cela. Soit  $T$  l'ensemble des individus positifs au test et  $M$  l'ensemble des individus malades. On sait que

$$\mathbb{P}(M) = 0,01, \quad \mathbb{P}(T | M^c) = 0,02, \quad \mathbb{P}(T^c | M) = 0,01,$$

et on veut calculer

$$\mathbb{P}(M | T).$$

Revenons à la définition de la probabilité conditionnelle :

$$\mathbb{P}(M | T) = \frac{\mathbb{P}(M \cap T)}{\mathbb{P}(T)}.$$

Or

$$\mathbb{P}(M \cap T) = \mathbb{P}(T | M)\mathbb{P}(M)$$

et d'après la formule de probabilités totales (théorème 2.6)

$$\begin{aligned} \mathbb{P}(T) &= \mathbb{P}(T \cap M) + \mathbb{P}(T \cap M^c) \\ &= \mathbb{P}(T | M)\mathbb{P}(M) + \mathbb{P}(T | M^c)\mathbb{P}(M^c). \end{aligned}$$

On obtient donc

$$\mathbb{P}(M | T) = \frac{\mathbb{P}(T | M)\mathbb{P}(M)}{\mathbb{P}(T | M)\mathbb{P}(M) + \mathbb{P}(T | M^c)\mathbb{P}(M^c)} = \frac{0,99 \cdot 0,01}{0,99 \cdot 0,01 + 0,02 \cdot 0,99} = \frac{1}{3}.$$

La probabilité d'être malade n'est en fait que de 33%, contrairement à l'intuition qui laisse penser à un niveau bien plus élevé. Toutefois, cela veut quand même dire que l'on a 33 fois plus de « chances » d'être malade qu'en l'absence de l'information du test.

**Exercice 1.4** ▷

1.  $N$  est à valeur dans  $\mathbb{N}$ , et  $F$  également.

De plus, le nombre de filles est inférieur au nombre d'enfants, donc  $N \leq F$  presque sûrement.

Soit  $(n, m) \in \mathbb{N}^2$ ,  $m \leq n$ ,

$$\mathbb{P}(N = n, F = m) = \mathbb{P}(N = n) \times \mathbb{P}(F = m | N = n)$$

$$\mathbb{P}(N = n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

De plus la probabilité d'avoir  $m$  filles parmi  $n$  est la même que celle de tirer  $m$  fois face dans  $n$  lancers de pièce, car les expériences sont iid.

$$\text{Ainsi } \mathbb{P}(F = m | N = n) = \binom{n}{m} \left(\frac{1}{2}\right)^n$$

$$\text{De plus, } \binom{n}{m} = \frac{n!}{m!(n-m)!}$$

$$\text{D'où } \forall (n, m) \in \mathbb{N}^2, \text{ si } n \leq m, \mathbb{P}(N = n, F = m) = 0, \text{ sinon } \mathbb{P}(N = n, F = m) = e^{-\lambda} \left(\frac{\lambda}{2}\right)^n \frac{1}{m!(n-m)!}$$

2. Soit  $m \in \mathbb{N}$ .

$(N = n)_{n \in \mathbb{N}}$  est un système complet d'événements, donc  $\mathbb{P}(F = m) = \sum_{n=0}^{\infty} \mathbb{P}(F = m, N = n)$

Or s'il y a  $m$  filles, il ne peut y avoir moins d'enfants.

$$\begin{aligned} \text{Donc } \mathbb{P}(F = m) &= \sum_{n=m}^{\infty} \mathbb{P}(F = m, N = n) = \frac{e^{-\lambda}}{m!} \sum_{n=m}^{\infty} \left(\frac{\lambda}{2}\right)^n \frac{1}{(n-m)!} = \frac{e^{-\lambda}}{m!} \left(\frac{\lambda}{2}\right)^m \sum_{n=0}^{\infty} \left(\frac{\lambda}{2}\right)^n \frac{1}{n!} \\ &= \frac{e^{-\lambda}}{m!} \left(\frac{\lambda}{2}\right)^m e^{\lambda/2} = \frac{e^{-\lambda/2}}{m!} \left(\frac{\lambda}{2}\right)^m \end{aligned}$$

$M$  suit la même loi par symétrie du problème :  $\mathbb{P}(M = m) = \frac{e^{-\lambda/2}}{m!} \left(\frac{\lambda}{2}\right)^m$

3. On a simplement  $M + F = N$ . Soit  $(m, p) \in \mathbb{N}^2$ .  $\mathbb{P}(F = m, M = p) = \mathbb{P}(F + M = m + p, M = p) = \mathbb{P}(N = m + p, M = p)$

$$\text{Donc d'après q.1, } \mathbb{P}(F = m, M = p) = e^{-\lambda} \left(\frac{\lambda}{2}\right)^{m+p} \frac{1}{m!p!}$$

4. D'après les calculs précédents,  $\forall (m, p) \in \mathbb{N}^2$ ,  $\mathbb{P}(F = m, M = p) = \mathbb{P}(F = m)\mathbb{P}(M = p)$ .

Ainsi le nombre de femelles pondues par la tortue est indépendant du nombre de males.

**Exercice 1.5** ▷ D'une part,

$$\mathbb{E}(X) = 1/2$$

$$\mathbb{E}(Z) = 0$$

$$\mathbb{E}(Y) = \mathbb{E}(X)\mathbb{E}(Z) = 0$$

$$\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) = \mathbb{E}(ZX^2) = \mathbb{E}(Z)\mathbb{E}(X^2) = 0,$$

donc  $X$  et  $Y$  sont décorréliées. En revanche, on remarque que  $|Y| = |X|$  donc  $\mathbb{P}(X = 1, Y = 0) = 0$  alors que

$$\mathbb{P}(Y = 0) = \mathbb{P}(X = 0) = 1/2 \text{ donc } \mathbb{P}(X = 1)\mathbb{P}(Y = 0) = 1/4.$$

Les deux variables  $X$  et  $Y$  ne sont donc pas indépendantes.

---

**Exercice 1.6** ▷

- 1.6.a) Il faut prendre garde au fait que l'indice ne représente pas un ordre de sortie puisque les boules sont toutes sorties en même temps. L'indice identifie seulement la case dans laquelle une boule donnée est mise. Les boules sont indistinguables donc les cases aussi. Par conséquent, la numéro de chaque case n'a pas d'importance ; autrement dit, les vecteurs aléatoires  $(X_1, \dots, X_m)$  et  $(X_{\sigma(1)}, \dots, X_{\sigma(n)})$  ont la même loi pour tout  $\sigma$  permutation de  $\{1, \dots, n\}$ .
- 1.6.b) Pour une boule donnée, la probabilité qu'elle soit rouge est  $r/N$  donc  $\mathbb{P}(X_i = 1) = r/N$ . De même la probabilité qu'on ait 2 boules rouges est donnée par

$$\mathbb{P}(X_1 X_2 = 1) = \frac{r}{N} \frac{r-1}{N-1}.$$

- 1.6.c) Par conséquent

$$\mathbb{E}(X) = \sum_{i=1}^m \mathbb{E}(X_i) = \sum_{i=1}^m \frac{r}{N} = \frac{m}{N} r.$$

Comme  $X_i$  suit une loi de Bernoulli, on sait que  $\mathbb{E}(X_i^2) = r/N$ . D'autre part,

$$\begin{aligned} \mathbb{E}(X^2) &= \mathbb{E}\left(\left(\sum_{i=1}^m X_i\right)^2\right) \\ &= \sum_{i=1}^m \mathbb{E}(X_i)^2 + 2 \sum_{1 \leq i < j \leq m} \mathbb{E}(X_i X_j) \\ &= m \frac{r}{N} + 2 \frac{m(m-1)}{2} \frac{r}{N} \frac{r-1}{N-1}. \end{aligned}$$

Dans ces conditions,

$$\begin{aligned} \text{Var}(X) &= m \frac{r}{N} + 2 \frac{m(m-1)}{2} \frac{r}{N} \frac{r-1}{N-1} - \frac{m^2}{N^2} r^2 \\ &= m \frac{r}{N} \left(1 + \frac{(m-1)(r-1)}{N-1} - \frac{mr}{N}\right) = br \frac{m(N-m)}{N}. \end{aligned}$$

**Exercice 1.7** ▷ On a de manière évidente (voir exercice 3.12)

$$\mathbb{P}(A) = 1/2, \mathbb{P}(B) = 1/2, \mathbb{P}(C) = 1/2.$$

Par ailleurs,  $A$  et  $B$  sont indépendants puisqu'ils dépendent chacun d'un dé différent. D'autre part, si un dé est pair et la somme est paire alors automatiquement, les 2 dés sont pairs donc

$$\mathbb{P}(A \cap C) = \mathbb{P}(A \cap B) = 1/4 = \mathbb{P}(A)\mathbb{P}(C).$$

De même pour  $\mathbb{P}(B \cap C)$ , donc les événements  $A$ ,  $B$  et  $C$  sont 2 à 2 indépendants. Enfin,

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B) = 1/4 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = 1/8.$$

**Exercice 1.8** ▷

1.8.a) D'après la formule des probabilités totales théorème 2.6 et la définition de la probabilité conditionnelle (Définition 1.3)

$$\mathbb{P}(X = k) = \sum_{m=0}^N \mathbb{P}(X = k, M = m) = \sum_{m=0}^N \mathbb{P}(X = k | M = m) \mathbb{P}(M = m).$$

En développant les coefficients binomiaux, on obtien

$$\begin{aligned} \mathbb{P}(X = k) &= \\ &= \sum_{m=0}^N \frac{m! (N - m)! n! (N - n)! N!}{k! (m - k)! (n - k)! (N - m - n + k)! N! (N - m)! m!} \theta^m (1 - \theta)^{N - m} \\ &= \frac{n!}{k! (n - k)!} \sum_{m=0}^N \frac{(N - n)!}{(m - k)! (N - m - n + k)!} \theta^m (1 - \theta)^{N - m} \\ &= \frac{n!}{k! (n - k)!} \theta^k (1 - \theta)^{n - k} \sum_{m=0}^N \binom{N - n}{m - k} \theta^{m - k} (1 - \theta)^{N - m - n + k}. \end{aligned}$$

On reconnaît dans la somme (en  $m$ ) la somme des probabilités d'une loi binomiale de paramètres  $N - n$  et  $\theta$ . Cette somme vaut donc 1. Par conséquent,  $X$  suit une loi binomiale de paramètres  $n$  et  $\theta$ .

1.8.b) Par définition de la loi conditionnelle (Définition 1.3),

$$\begin{aligned} \mathbb{P}(M = m | X = k) &= \\ &= \frac{\mathbb{P}(M = m, X = k)}{\mathbb{P}(X = k)} \\ &= \frac{\mathbb{P}(X = k | M = m) \mathbb{P}(M = m)}{\mathbb{P}(X = k)} \\ &= \frac{\binom{m}{k} \binom{N - m}{n - k} \frac{k! (n - k)!}{n!} \theta^{-k} (1 - \theta)^{-(n - k)} \frac{N!}{(N - m)! m!} \theta^m (1 - \theta)^{N - m}}{\binom{N}{n}} \\ &= \frac{m! (N - m)! n! (N - n)! k! (n - k)! N!}{k! (m - k)! (n - k)! (N - m - n + k)! N! n! (N - m)! m!} \theta^{m - k} (1 - \theta)^{N - m - n + k} \\ &= \frac{(N - n)!}{(m - k)! (N - m - n + k)!} \theta^{m - k} (1 - \theta)^{N - m - n + k}. \end{aligned}$$

1.8.c) Pour  $k = 0$ , cela signifie que la loi de  $M$  conditionnellement à  $X = 0$  est une loi binomiale de paramètres  $N - n$  et  $\theta$ .

### Exercice 1.9 ▷

1.9.a) Dire que  $\mathbb{P}$  est la loi uniforme sur  $\{0, 1\}^n$  signifie que

$$\mathbb{P}(\{x_1, \dots, x_n\}) = 2^{-n}$$

pour tout  $(x_1, \dots, x_n) \in \{0, 1\}^n$ . L'événement  $(X_i = 1)$  s'écrit aussi

$$(X_i = 1) = \{(x_1, \dots, x_n) \in \{0, 1\}^n, x_i = 1\}.$$

Une seule coordonnée est fixée (en l'occurrence la  $i$ -ième) donc le cardinal de cet ensemble est  $2^{n-1}$ . Par définition de la probabilité uniforme sur un ensemble fini (cf. 1.2.3)

$$\mathbb{P}(X_i = 1) = \frac{\text{card}(X_i = 1)}{2^n} = \frac{2^{n-1}}{2^n} = \frac{1}{2}.$$

Par passage au complémentaire, on en déduit que  $\mathbb{P}(X_i = 0) = 1/2$ . Maintenant, soit  $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$  un ensemble d'indices (distincts). On doit prouver que pour tout  $(x_1, \dots, x_k) \in \{0, 1\}^k$ ,

$$\mathbb{P}(X_{i_1} = x_1, \dots, X_{i_k} = x_k) = \mathbb{P}(X_{i_1} = x_1) \dots \mathbb{P}(X_{i_k} = x_k).$$

Or dans l'ensemble  $(X_{i_1} = x_1, \dots, X_{i_k} = x_k)$  on a fixé  $k$  des  $n$  coordonnées donc son cardinal est  $2^{n-k}$ . Par conséquent,

$$\mathbb{P}(X_{i_1} = x_1, \dots, X_{i_k} = x_k) = \frac{2^{n-k}}{2^n} = 2^{-k} = \mathbb{P}(X_{i_1} = x_1) \dots \mathbb{P}(X_{i_k} = x_k),$$

puisque chacun des termes de ce produit vaut  $1/2$ .

- 1.9.b)** D'après sa définition, la longueur d'un mot est égale au nombre de ses bits à 1. En l'occurrence, cela revient à compter le nombre d'apparition de 1 dans une suite de  $n$  tirages iid où la probabilité d'apparition d'un 1 à chaque place est de  $1/2$ . Cela correspond à une distribution binomiale de paramètres  $n$  et  $1/2$  (cf. **1.2.3**) dont on sait que l'espérance vaut  $n/2$ .
- 1.9.c)** On sait aussi que la variance d'une loi binomiale  $B(n, p)$  est donnée par  $np(1-p)$ , soit ici  $n/4$ .
- 1.9.d)** Par définition de la distance de Hamming,

$$d(X, Y) = \sum_{i=1}^n \mathbf{1}_{X_i \neq Y_i}.$$

Les variables aléatoires  $Z_i = \mathbf{1}_{X_i \neq Y_i}$  sont des variables aléatoires de Bernoulli telles que (d'après la formule des probabilités totales cf. théorème **2.6**)

$$\begin{aligned} \mathbb{P}(Z_i = 0) &= \mathbb{P}(X_i = Y_i) = \mathbb{P}(X_i = Y_i, X_i = 0) + \mathbb{P}(X_i = Y_i, X_i = 1) \\ &= \mathbb{P}(X_i = 0, Y_i = 0) + \mathbb{P}(X_i = 1, Y_i = 1) = \mathbb{P}(X_i = 0)\mathbb{P}(Y_i = 0) + \mathbb{P}(X_i = 1)\mathbb{P}(Y_i = 1), \end{aligned}$$

puisque  $X_i$  et  $Y_i$  sont indépendantes. Enfin, comme  $X_i$  et  $Y_i$  sont des Bernoulli des paramètre  $1/2$ , on obtient

$$\mathbb{P}(Z_i = 0) = 1/2 \cdot 1/2 + 1/2 \cdot 1/2 = 1/2.$$

D'après le lemme ??, les  $Z_i$  sont indépendantes. On sait que la variance de la somme de variables aléatoires indépendantes est la somme de leur variance donc

$$\text{Var}(d(X, Y)) = \sum_{i=1}^n \text{Var}(Z_i) = n/4.$$

Comme  $\mathbb{E}(d(X, Y)) = \sum_{i=1}^n \mathbb{E}(Z_i) = n/2$ , on tire

$$\mathbb{E}(d(X, Y)^2) = \text{Var}(d(X, Y)) + \mathbb{E}(d(X, Y))^2 = n/4 + n^2/4 = \frac{n(n+1)}{4}.$$

**Exercice 1.11** ▷

1.11.a) Cet encadrement est connu sous le nom d'*inégalités de Bonferonni* et constitue un raffinement de la formule de Poincaré (ou principe d'inclusion-exclusion).

On remarque que  $X = 0$  si et seulement si  $X_i = 0$  pour tout  $i$ , en d'autres termes si et seulement si  $\prod_{i=1}^n (1 - X_i) = 1$  : par ailleurs, ce produit, s'il ne vaut pas 1 vaut 0. D'où

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

La quantité *aléatoire*  $\prod_{i=1}^n (1 - X_i)$  prend des valeurs du type  $\prod_{i=1}^n (1 - x_i)$  où chaque  $x_i$  vaut 0 ou 1. Chacune de ces valeurs est donc de la forme  $f(1)$  où

$$f : t \mapsto \prod_{i=1}^n (1 - tx_i).$$

Une des formules de Taylor nous dit que

$$f(1) = f(0) + f'(0) + f''(0) + \dots + f^{(r)}(0) + f^{(r+1)}(c) \quad (\text{B.1})$$

où  $c \in ]0, 1[$ . Or comme chaque  $x_i$  est dans  $[0, 1]$ , on constate que le signe de la quantité  $f^{(r+1)}(c)$  ne dépend que de la parité de  $r$ . Passer aux espérances dans (B.1) donne les inégalités demandées.

1.11.b) On a  $E[X_i] = (1 - p)^{n-1} \sim e^{-c}/n$ . D'où  $E[X] = nE[X_1] \sim e^{-c}$ .

1.11.c) On a  $F^{(k)} = \binom{n}{k} E[X_1 \cdots X_k]$ . De plus,

$$\begin{aligned} E[X_1 \cdots X_k] &= P[\text{les sommets } 1, \dots, k \text{ sont isolés}] \\ &= (1 - p)^{k(n-1) - \binom{k}{2}} \\ &= E[X_1]^k (1 - p)^{-\binom{k}{2}} \end{aligned}$$

Donc, pour  $k$  fixé,  $F^{(k)} \rightarrow \binom{n}{k} E[X_1]^k \rightarrow \frac{e^{-ck}}{k!}$ .

1.11.d) Il s'agit d'utiliser la question préliminaire : on se donne  $\varepsilon > 0$  et on choisit  $r$  tel que

$$\left| \sum_{k=0}^{2r} (-1)^k \frac{e^{-ck}}{k!} - e^{-e^{-c}} \right| < \frac{\varepsilon}{2}$$

puis un  $n_0$  tel que pour tout  $n > n_0$  et  $0 \leq k \leq 2r$

$$\left| F^{(k)} - \frac{e^{-ck}}{k!} \right| < \frac{\varepsilon}{2(2k+1)}.$$

On en déduit que pour tout  $n$  suffisamment grand

$$\mathbb{P}[X = 0] < e^{-e^{-c}} + \varepsilon.$$

On obtient une borne inférieure sur  $P[X = 0]$  de manière similaire. Comme les  $\varepsilon$  sont arbitraires, on en déduit le résultat.

*Commentaire : on pourrait montrer de la même manière que*

$$\mathbb{P}(X = j) \rightarrow e^{-e^{-c}} e^{-cj} / j!.$$

*La loi de  $X$  se rapproche d'une loi de Poisson, ce qui veut dire que les  $X_i$  se comportent de manière « de plus en plus indépendantes ».*

1.11.e) Il y a  $\binom{n}{2}$  paires de sommets. la probabilité qu'une paire de sommets donnée constitue une composante connexe vaut  $p(1-p)^{2(n-2)}$ . L'espérance du nombre de composantes connexes à deux sommets vaut donc

$$\binom{n}{2} p(1-p)^{2(n-2)} \sim \frac{p}{2} (ne^{-pn})^2 = \frac{p}{2} e^{-2c} \rightarrow 0$$

car  $p$  tend vers 0 quand  $n \rightarrow \infty$ .

1.11.f) On en déduit qu'avec probabilité tendant vers 1 le nombre de composantes connexes à  $t$  éléments avec  $2 \leq t \leq n/2$  tend vers 0.

1.11.g) Or  $G_{n,p}$  n'est pas connexe si et seulement s'il existe une composante connexe à  $t$  sommets pour  $1 \leq t \leq n/2$ . La probabilité d'être non connexe se comporte donc comme la probabilité d'avoir (au moins) un point isolé. Autrement dit, la probabilité que  $G_{n,p}$  soit connexe tend vers  $e^{-e^{-c}}$ . En particulier on en déduit que si  $p$  grandit moins vite que  $\ln n/n + c/n$  pour tout  $c$ , alors  $G_{n,p}$  n'est pas connexe avec probabilité tendant vers 1. Par contre si  $p$  grandit plus vite que  $\ln n/n + c/n$  pour tout  $c$ , alors  $G_{n,p}$  est connexe avec probabilité tendant vers 1.

### Exercice 1.12 ▷

1.12.a) Pour  $\theta > 0$ , la fonction  $(x \mapsto \exp(\theta x))$  est strictement croissante donc

$$(X \geq \eta) = (e^{\theta X} \geq e^{\theta \eta}).$$

1.12.b) D'après l'inégalité de Markov pour  $p = 1$  (cf. proposition 4.6),

$$\mathbb{P}(e^{\theta X} \geq e^{\theta \eta}) \leq e^{-\theta \eta} \mathbb{E}(e^{\theta X}).$$

1.12.c) Pour une loi de Poisson de paramètre  $\lambda$ ,

$$\mathbb{E}(e^{\theta X}) = \sum_{k=0}^{\infty} e^{\theta k} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^{\theta} \lambda)^k}{k!} = \exp(-\lambda + \lambda e^{\theta}).$$

1.12.d) Pour  $\eta$  fixé, on doit donc calculer

$$\operatorname{argmin}_{\theta > 0} \exp(-\theta \eta - \lambda + \lambda e^{\theta}) = \operatorname{argmin}_{\theta > 0} (-\theta \eta + \lambda e^{\theta}).$$

Comme

$$\frac{d}{d\theta} (-\theta \eta + \lambda e^{\theta}) = -\eta + \lambda e^{\theta},$$

le point critique est obtenu pour  $\theta_0 = \ln(\eta/\lambda)$ . La dérivée seconde étant positive, c'est bien un minimum.

1.12.e) On obtient donc

$$\ln \mathbb{P}(X \geq \eta) \leq -\theta_0 \eta + \eta - \lambda.$$

Pour  $\eta = K\lambda$ ,  $\theta_0 = \ln K$ , on a donc

$$\ln \mathbb{P}(X \geq \eta) \leq -\lambda(K \ln K - K + 1)$$

On doit résoudre l'équation

$$K \ln K - K + 1 = -\frac{\ln(0.001)}{\lambda}.$$

Avec le code SAGE suivant, on la résout pour  $\lambda = 5, 10, 15, 20, 25$  (la variable `lambda` est un mot réservé en SAGE...)

Listing B.1 – Calcul de  $K$

```
def f(x, l, epsilon):
    return x*log(x)-x+1+log(epsilon)/l
for i in range(5):
    l=(i+1)*5
    def g(x):
        return f(x, l, 0.001)
    print l, find_root(g, l, 4*l, rtol=0.0001)
```

On trouve les résultats suivants :

$\lambda$	5	10	15	20	25
K	3,08	2,39	2,10	1,94	1,83

On voit que si l'on doit stocker une loi de Poisson en mémoire, i.e. stocker les valeurs des probabilités  $\mathbb{P}(X = k)$ , en ne gardant que les valeurs de  $k$  jusqu'à trois fois la moyenne ( $\lambda$ ), on a une erreur inférieure à 0,1 pour cent. On peut évidemment raffiner en prenant des valeurs de seuil plus basses.

**Exercice 1.13** ▷

- 1.13.a) L'espace probabilisé est  $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$  muni de la probabilité uniforme. On note de façon naturelle  $D_1$  et  $D_2$  les deux composantes ou si l'on préfère  $D_1$  représente la valeur du dé numéro 1,  $D_2$  celle du dé numéro 2. On doit calculer  $\mathbb{P}(D_1 + D_2 = k)$  pour  $k \in \{2, \dots, 12\}$ . Par définition de la probabilité uniforme, pour  $k$  fixé, il faut calculer le nombre de couples de  $\Omega$  dont la somme fait  $k$ . La probabilité d'obtenir  $k$  s'obtient en divisant ce cardinal par le cardinal de  $\Omega$  qui est 36. Pour  $k = 2$ , un seul couple convient :  $(1, 1)$ . Pour  $k = 3$ , il y en a 2 :  $(1, 2)$  et  $(2, 1)$ . On notera qu'ici les dés sont distingués donc ces deux couples diffèrent. Au final, on obtient

$k$	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(D_1 + D_2 = k)$	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$5/36$	$4/36$	$3/36$	$2/36$	$1/36$

Par ailleurs,

$$\mathbb{E}(D_1 + D_2) = 2\mathbb{E}(D_1) = 2 \cdot (1 + 2 + 3 + 4 + 5 + 6)/6 = 7.$$

- 1.13.b) La v.a.r.  $\tau$  représente un premier temps de succès (cf. 1.2.3) si l'on considère que « succès » signifie l'obtention d'un 7. D'après la question précédente, la probabilité d'avoir 7 est de  $1/6$ , donc  $\tau$  suit une loi géométrique de paramètre  $1/6$ . D'après l'exercice ??, sa moyenne vaut 6.
- 1.13.c) Un initié connaît  $\tau$  donc s'arrête au coup d'avant.
- 1.13.d) Le gain de l'initié s'exprime comme

$$G = \sum_{i=1}^{\tau-1} S_i = \sum_{i=1}^{\tau} S_i - 7, \tag{B.2}$$

puisque par définition de  $\tau$ ,  $S_\tau = 7$ . Pour calculer son espérance, on ne peut pas utiliser la linéarité puisque le nombre de termes dans la somme est aléatoire. Il faut d'abord décomposer

l'espace en sous-ensembles où l'on connaît  $\tau$ . D'après la formule des probabilités totale (cf. théorème 2.6)

$$\begin{aligned}
& \mathbb{E} \left( \sum_{i=1}^{\tau} S_i \right) \\
&= \sum_{k=1}^{\infty} \mathbb{E} \left( \sum_{i=1}^{\tau} S_i \mathbf{1}_{\tau=k} \right) \\
&= \mathbb{E} (S_1 \mathbf{1}_{S_1=7}) + \sum_{k=2}^{\infty} \mathbb{E} \left( \sum_{i=1}^k S_i \mathbf{1}_{S_1 \neq 7, \dots, S_{k-1} \neq 7, S_k=7} \right) \\
&= \mathbb{E} (S_1 \mathbf{1}_{S_1=7}) \\
&+ \sum_{k=2}^{\infty} \left( \sum_{i=1}^{k-1} \mathbb{E} \left( (S_i \mathbf{1}_{S_i \neq 7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \mathbf{1}_{S_k=7} \right) + \mathbb{E} \left( (S_k \mathbf{1}_{S_k=7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \right) \right).
\end{aligned}$$

Dans chaque terme de la forme

$$\mathbb{E} \left( (S_i \mathbf{1}_{S_i \neq 7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \mathbf{1}_{S_k=7} \right),$$

on a un produit de fonctions de chacune des variables  $S_i$  donc en vertu de l'indépendance des v.a.r.  $S_k$  et du théorème 4.15, on peut séparer les espérances :

$$\begin{aligned}
& \mathbb{E} \left( (S_i \mathbf{1}_{S_i \neq 7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \mathbf{1}_{S_k=7} \right) \\
&= \mathbb{E} ((S_i \mathbf{1}_{S_i \neq 7})) \mathbb{P}(S_1 \neq 7) \dots \mathbb{P}(S_{k-1} \neq 7) \mathbb{P}(S_k = 7).
\end{aligned}$$

D'une part,

$$\mathbb{E} (S_k \mathbf{1}_{S_k=7}) = \mathbb{E} (7 \mathbf{1}_{S_k=7}) = 7 \mathbb{P}(S_k = 7) = \frac{7}{6}.$$

D'autre part, calculons maintenant  $\mathbb{E} (S_i \mathbf{1}_{S_i \neq 7})$ .

$$\mathbb{E} (S_i \mathbf{1}_{S_i \neq 7}) = \mathbb{E} (S_i) - \mathbb{E} (S_i \mathbf{1}_{S_i=7}) = 7 - 7 \mathbb{P}(S_i = 7) = 7 \left(1 - \frac{1}{6}\right) = \frac{35}{6}.$$

On en tire que

$$\mathbb{E} \left( (S_i \mathbf{1}_{S_i \neq 7}) \mathbf{1}_{S_1 \neq 7} \dots \mathbf{1}_{S_{k-1} \neq 7} \mathbf{1}_{S_k=7} \right) = \frac{35}{6} \left(\frac{5}{6}\right)^{k-2} \frac{1}{6}.$$

$$\begin{aligned}
\mathbb{E} \left( \sum_{i=1}^{\tau} S_i \right) &= \frac{7}{6} + \sum_{k=2}^{\infty} \left( \sum_{i=1}^{k-1} \left(\frac{5}{6}\right)^{k-2} \frac{1}{6} \frac{35}{6} + \left(\frac{5}{6}\right)^{k-1} \frac{7}{6} \right) \\
&= \frac{7}{6} + \frac{35}{36} \sum_{k=2}^{\infty} (k-1) \left(\frac{5}{6}\right)^{k-2} + \frac{7}{6} \sum_{k=2}^{\infty} \left(\frac{5}{6}\right)^{k-1} \\
&= \frac{7}{6} + \frac{35}{36} \sum_{k=1}^{\infty} k \left(\frac{5}{6}\right)^{k-1} + \frac{7}{6} \sum_{k=0}^{\infty} \left(\frac{5}{6}\right)^k.
\end{aligned}$$

On reconnaît dans les expressions de la forme

$$\sum_{k=1}^{\infty} k \rho^{k-1}$$

la dérivée de la série géométrique donc

$$\sum_{k=1}^{\infty} k\rho^{k-1} = \frac{d}{d\rho} \sum_{k=0}^{\infty} \rho^k = \frac{d}{d\rho} \frac{1}{1-\rho} = \frac{1}{(1-\rho)^2}.$$

On obtient donc

$$\mathbb{E} \left( \sum_{i=1}^{\tau} S_i \right) = \frac{7}{6} + \frac{35}{36} \frac{1}{(1-5/6)^2} + \frac{35}{36} \frac{1}{1-5/6} = \frac{7}{6} + 35 + \frac{35}{6} = 35 + 7.$$

Au vu de l'équation (B.2), on obtient  $\mathbb{E}(G) = 35$ .

1.13.e) Conditionnellement à  $(X_n = i)$ , les valeurs possibles de  $X_{n+1}$  sont 0 si  $i = 0$  et

$$\{0\} \cup \{i+1, i+2, \dots, i+6, 0, i+8, \dots, i+12\} = \{0\} \cup \{i+B\}$$

où  $B = \{2, 3, 4, 5, 6, 8, 9, 10, 11, 12\}$ , si  $i > 0$ . Au vu des rappels ou du corollaire ??,

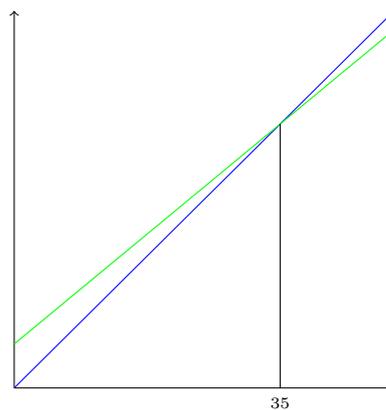
$$\begin{aligned} \mathbb{E}(X_{n+1} | X_n = i) &= \sum_{j \in i+B} j \mathbb{P}(X_{n+1} = j | X_n = i) \\ &= \sum_{k \in B} (i+k) \mathbb{P}(S_{n+1} = k | X_n = i) \\ &= \sum_{k \in B} (i+k) \mathbb{P}(S_{n+1} = k), \end{aligned}$$

car  $X_n$  qui est une fonction de  $S_1, \dots, S_n$  est indépendant de  $S_{n+1}$  d'après le lemme ??. On peut transformer cette quantité en remarquant que  $B$  contient toutes les valeurs possibles de  $S_{n+1}$  sauf 7

$$\mathbb{E}(X_{n+1} | X_n = i) = i\mathbb{P}(S_{n+1} \neq 7) + \mathbb{E}(S_{n+1} \mathbf{1}_{S_{n+1} \neq 7}) = \frac{5}{6}i + \frac{35}{6},$$

d'après les questions précédentes.

1.13.f) La droite d'équation  $y = 5/6x + 35/6$  croise la première bissectrice en  $(35, 35)$ . Avant elle est au dessus, après elle est en dessous. Cela signifie que si le score courant est inférieur (respectivement supérieur) à 35, on peut « espérer » s'enrichir (respectivement s'appauvrir). La stratégie consiste donc à jouer tant que l'on n'a pas atteint 35 et à s'arrêter dès que ce palier est atteint.

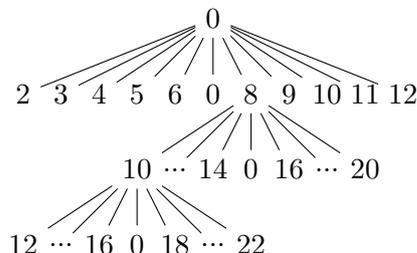


1.13.g) La méthode de la question 1.13.d) ne permet pas de calculer l'espérance du gain explicitement sous la stratégie optimale définie dans la question précédente. Nous présentons ici deux méthodes pour résoudre numériquement ce calcul.

- Par simulation (cf. ??). On simule un grand nombre de trajectoires du jeu sous cette stratégie : on joue jusqu'à perdre ou atteindre 35 ou plus. On fait la moyenne des gains obtenus et l'intervalle de confiance. Pour  $N = 1\,000\,000$  de tirages, on obtient comme intervalle de confiance

[14,1378267459162; 14,3732532540876].

- La deuxième solution consiste à utiliser l'arbre de toutes les trajectoires possibles.



Chaque branche est pondérée par sa probabilité : ainsi la branche de 0 vers 2 a un poids de  $1/36$ . On parcourt l'arbre de manière récursive jusqu'à atteindre toutes les feuilles de valeur supérieure ou égale à 35. Chaque chemin entre la racine et une feuille est pondéré par le produit des poids rencontrés sur le chemin et on fait la somme de tous ces poids sur toutes les feuilles. Evidemment, on n'examine pas les branches qui correspondent à un 7. Cela donne le résultat exact.

```
def score(init,proba,seuil,distrib):
    c=0.
    if (init>seuil):
        return init*proba
    else:
        for j,l in distrib.iteritems():
            c+=score(init+j,proba*l,seuil,distrib)
        return c
```

On notera qu'à travers le paramètre `distrib`, on peut changer la distribution des tirages (par exemple considérer un jeu à deux à 4 faces avec comme nombre interdit 5). L'appel à la fonction se fait donc par le code

```
distrib={2:1/36., 3:2/36., 4:3/36., 5:4/36., 6:5/36.,
        8:5/36., 9:4/36., 10:3/36., 11:2/36., 12:1/36.}
scoremoyen=score(0.,1.,35,distrib)
```

Ce qui donne comme résultat

Gain moyen = ??.

**Exercice 1.14** ▷ Considérons la fonction génératrice de la loi uniforme sur  $\{2, \dots, 12\}$ .

$$\mathbb{E}(s^U) = \sum_{k=2}^{12} s^k \frac{1}{11} = \frac{s^2}{11} \sum_{k=0}^{10} s^k = \frac{s^2}{11} \frac{s^{11} - 1}{s - 1}$$

C'est donc un polynôme de degré 12 donc les racines sont 0 (ordre 2) et les racines (simples) 11-ième de l'unité à l'exception de 1. Or la fonction génératrice de la loi de la somme, qui est

aussi un polynôme de degré 12, est le carré de la fonction génératrice de la loi d'un dé. Par conséquent, toutes ses racines sont au moins double. Comme un polynôme est déterminé par ses racines, on ne peut pas avoir égalité des fonctions génératrices.

**Exercice 2.1** ▷

1.  $\mathbf{1}_\Omega$  et  $\mathbf{1}_\emptyset$  valent respectivement 1 et 0.
2.  $\mathbf{1}_{\bar{A}} = 1 - \mathbf{1}_A$ ,  $\mathbf{1}_{A \cap B} = \mathbf{1}_A \mathbf{1}_B$ .
3. On raisonne par équivalence :  $\omega \in \cup_n A_n$  équivaut à  $\exists ! n, \omega \in A_n$ , ce qui équivaut à  $\sum_n \mathbf{1}_{A_n} = 1$ .

**Exercice 2.2** ▷ Solution de l'exercice 5

1.  $f^{-1}(\{1\}) = \{a\}$ ;  $f^{-1}(\{2\}) = \emptyset$ ;  $f^{-1}(\{3\}) = \{b, c\}$ ;  $f^{-1}(F) = E$
2. On va les montrer par double inclusion Soit  $x \in f^{-1}(A \cup B)$ ;  
 Donc,  $f(x) \in A \cup B$   
 donc  $f(x) \in A$  ou  $f(x) \in B$   
 donc  $x \in f^{-1}(A)$  ou  $x \in f^{-1}(B)$   
 D'où  $x \in f^{-1}(A) \cup f^{-1}(B)$   
 Et maintenant l'autre inclusion : soit  $x \in f^{-1}(A) \cup f^{-1}(B)$   
 donc  $x \in f^{-1}(A)$  ou  $x \in f^{-1}(B)$   
 donc  $f(x) \in A$  ou  $f(x) \in B$   
 donc  $f(x) \in A \cup B$   
 donc  $x \in f^{-1}(A \cup B)$

Pour l'intersection, on suit exactement le même raisonnement en remplaçant les "ou" par des "et".

Maintenant, montrons le résultat pour une union et intersection dénombrable. Pour ce faire, on va procéder par récurrence sur  $n \in \mathbb{N}$

- pour  $n = 1$  et  $n = 2$  : c'est évident.

- Soit  $n \in \mathbb{N}$ , supposons le résultat vrai au rang  $n$  :

On a :  $\cup_{i=1}^{n+1} A_i = \cup_{i=1}^n A_i \cup A_{n+1}$ , et on conclut grâce au cas  $n=2$ .

3.  $f^{-1}([0, \frac{1}{2}]) = [0, \theta] \cup [\pi - \theta, \pi]$  où  $\theta \in [0, \frac{\pi}{4}]$  tel que :  $\sin(\theta) = \frac{1}{2}$

**Exercice 2.3** ▷ Montrons que cette fonction de Dirac est une probabilité sur  $(\Omega, \mathcal{F})$ .

- $\delta_a(\emptyset) = 0$
- $\delta_a(\Omega) = 1$  car  $a \in \Omega$
- $\delta_a \geq 0$
- ( $\sigma$  - additivité) : Soient  $(A_n)_n$  des ensembles deux à deux disjoints de  $\Omega$ . Si  $a \in \cup_n A_n$  alors,  $\exists ! p \in \mathbb{N}$  tel que :  $a \in A_p$ , donc  $\delta_a(A_p) = 1$  et  $\forall n \neq p, \delta_a(A_n) = 0$ , donc  $\delta_a(\cup_n A_n) = 1 = \sum_n \delta_a(A_n)$ , et de même si  $a \notin \cup_n A_n$ , alors  $\forall n \in \mathbb{N}, a \notin A_n$ , alors  $\delta_a(\cup_n A_n) = 0 = \sum_n \delta_a(A_n)$ .

D'où elle est bien une probabilité.

**Exercice 2.4** ▷ On va montrer qu'elle est bien une mesure.

- Elle est bien positive, puisque tous les  $\alpha_i$  sont des réels positifs, et que  $\mu_i$  sont des mesures.
- On a :  $\mu(\emptyset) = \sum_{i \in I} \alpha_i \mu_i(\emptyset) = 0$
- Il est facile de voir qu'elle satisfait l'axiome de  $\sigma$ -additivité, car tous les  $\mu_i$  sont des mesures.

C'est en combinant cet exo avec le précédent qu'on construit deux mesures classiques : mesure de comptage (tous les  $\alpha_i$  sont égaux à 1 et les  $\mu_i$  sont des mesures de Dirac sur des points différents de  $\Omega$ ), et la mesure discrète : les  $\mu_i$  sont des mesures de Dirac.

Attention : Par contre, cette propriété reste vraie seulement si tous les  $\alpha_i$  sont positives : la différence entre deux mesures n'a aucun sens !

**Exercice 2.5** ▷

1. Montrons les deux inclusions. L'inclusion réciproque est évidente : Si  $\omega \in \bigcup_{i \geq 1}^{\infty} D_n$ , alors on dispose de  $N$  tel que  $\omega \in D_N$ . Alors  $\mu(\omega) > \frac{1}{N}$ , donc  $\omega \in D$   
D'où  $\bigcup_{i \geq 1}^{\infty} D_n \subset D$   
Montrons l'inclusion directe : Si  $\omega \in D$ , alors  $\mu(\omega) > 0$ , donc puisque  $(n \mapsto \frac{1}{n})$  tend vers 0, on dispose d'un rang  $N$  à partir duquel  $\frac{1}{N} < \mu(\omega)$ . Alors  $\omega \in D_N$ , et  $\omega \in \bigcup_{i \geq 1}^{\infty} D_n$ . Ainsi  $D = \bigcup_{i \geq 1}^{\infty} D_n$
2. Soit  $n \geq 1$ . Enumérons  $D_n = \{\omega_1, \omega_2, \dots\}$  éventuellement infini a priori.  
 $\bigcup_{\omega \in D_n} \{\omega\}$  est une union disjointe d'événements, et  $\mu(\bigcup_{\omega \in D_n} \{\omega\}) = \sum_{\omega \in D_n} \mu(\{\omega\}) \leq \frac{\#D_n}{n} \leq 1$  Ainsi le cardinal de  $D_n$  est fini, et même majoré par  $n$ .
3. Ici  $\Omega$  est  $\mathbb{R}$  car on travaille sur une fonction de répartition de  $\mu$ . Posons  $A$  l'ensemble des points de discontinuité de la fonction de répartition de  $\mu, F_\mu$ . Alors  $A = \{x \in \mathbb{R}, F_\mu(x^-) < F_\mu(x^+)\} = \{x \in \mathbb{R}, \mu(]-\infty, x[) < \mu(]-\infty, x])\} = \{x \in \mathbb{R}, \mu(\{x\}) \neq 0\} = B$  Ainsi  $A$  est l'union dénombrable d'ensembles finis donc dénombrables, donc  $A$  est dénombrable.

**Exercice 2.6** ▷

1. Notons  $F_1(t)$  la fonction de répartition de  $\mu_1$ . On a :

$$F_1(t) = \delta_2(]-\infty, t]) = \begin{cases} 1 & \text{si } t \geq 2, \\ 0 & \text{sinon.} \end{cases}$$

2. Notons  $F_2(t)$  la fonction de répartition de  $\mu_2$ . On a :

$$F_2(t) = \mu_2(]-\infty, t]) = \begin{cases} 0 & \text{si } t < -1, \\ 0,2 & \text{si } -1 \leq t < 0, \\ 0,6 & \text{si } 0 \leq t < 2, \\ 1 & \text{si } t \geq 2. \end{cases}$$

**Exercice 2.8** ▷

**Vrai** L'ensemble  $f^{-1}([2; 3.5])$  est l'ensemble des réels tels que

$$[\sqrt{|x|}] \in [2; 3.5] \iff 2 \leq \sqrt{|x|} < 4 \iff x \in [4; 16[ \cup ] - 16; -4].$$

**Vrai** Une fonction constante  $f$  ne prend qu'une valeur, que l'on note  $\alpha$ . Pour un mesurable  $H$  de l'espace d'arrivée, de deux choses l'une :  $\alpha$  appartient à  $H$  ou pas. Dans le premier cas,  $f^{-1}(H) = E$ , dans le deuxième,  $f^{-1}(H) = \emptyset$ . Comme toute tribu contient au moins ces deux ensembles (cf. définition 2.1),  $f$  est mesurable.

**Vrai** Par construction. Il suffit de regarder leur expression.

**Vrai** Si on appelle  $\alpha_1, \dots, \alpha_n$  les valeurs prises par  $f$  et  $A_i = f^{-1}(\{\alpha_i\})$ . On vérifie en épluchant tous les cas que

$$f(x) = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}(x).$$

C'est-à-dire que  $f$  est étagée.

**Faux** La mesure de Lebesgue d'un singleton est nulle. En effet,

$$\{x\} = \bigcap_{n=1}^{\infty} [x - 1/n, x + 1/n].$$

Or  $\lambda([x - 1/n, x + 1/n]) = 2/n$  et d'après la monotonie des mesures (cf. 2.7.d),

$$\lambda(\{x\}) = \lim_{n \rightarrow \infty} \lambda([x - 1/n, x + 1/n]) = 0.$$

**Faux** Comme

$$\bigcup_{x \uparrow x_0} ] - \infty, x[ = ] - \infty, x_0[,$$

la monotonie des mesures (cf. 2.7.c) assure seulement que

$$\lim_{x \uparrow x_0} \mu(] - \infty, x]) = \mu(] - \infty, x_0]).$$

Si le singleton  $\{x_0\}$  n'est pas de  $\mu$ -mesure nulle, ceci n'est pas égal à  $\mu(] - \infty, x_0])$ .

**Vrai** Compte-tenu de 2.7.d) et du fait que

$$\bigcap_{x \downarrow x_0} ] - \infty, x] = ] - \infty, x_0].$$

L'hypothèse  $\mu$  mesure de probabilité garantit que  $\mu(] - \infty, x]) < \infty$ .

**Exercice 2.9** ▷

1. Il n'y a aucun entier plus grand que n'importe quel autre donc  $\bigcap_n A_n = \emptyset$ .
2. Comme  $\mu(A_n) = +\infty$  et  $\mu(\emptyset) = 0$ , on a bien dissemblance des limites.
3. Il n'y a pas de contradiction avec 2.7.d) car l'on n'a pas  $\mu(A_1)$  fini.

**Exercice 2.11** ▷ Il suffit de vérifier les trois axiomes définissant une tribu. Ceci découle immédiatement des propriétés de l'image réciproque.

**Exercice 2.15** ▷

1. Soit  $x \in \mathbb{R}$  montrons que  $\mu_x : H \mapsto \lambda(x + H)$  est une mesure.

- Positivité : La positivité de  $\mu_x$  découle directement de celle de  $\lambda$
- Notons que  $x + \emptyset = \{x + y, y \in \emptyset\} = \emptyset$  en effet aucun  $y$  n'appartient à l'ensemble vide. Ainsi  $\mu_x(\emptyset) = \lambda(x + \emptyset) = \lambda(\emptyset) = 0$  (puisque  $\lambda$  est une mesure) On a donc vérifiée  $\mu_x(\emptyset) = 0$
- $\sigma$ -additivité : Pour montrer la  $\sigma$ -additivité commençons par établir que si  $(A_n)_{n \in \mathbb{N}}$  est une famille d'ensemble disjoint alors  $(x + A_n)_{n \in \mathbb{N}}$  l'est aussi. Supposons par l'absurde que  $\exists (i, j) \in \mathbb{N}, \exists y \in (x + A_i) \cap (x + A_j)$  alors  $y - x \in A_i \cap A_j$  or  $A_i \cap A_j = \emptyset$  c'est donc impossible. Ainsi les éléments de la famille  $(x + A_n)_{n \in \mathbb{N}}$  sont deux à deux disjoints. Par conséquent, Soit  $(A_n)_{n \in \mathbb{N}}$  une famille d'ensemble disjoint.

$$\begin{aligned} \mu_x\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= \lambda\left(\bigcup_{n \in \mathbb{N}} (x + A_n)\right) \\ &= \sum_{n \in \mathbb{N}} \lambda(x + A_n) \quad (\text{par } \sigma\text{-additivité de } \lambda) \\ &= \sum_{n \in \mathbb{N}} \mu_x(A_n) \end{aligned} \quad (\text{B.3})$$

2. Commençons par montrer que  $x + ]a; b] = ]a + x; b + x]$ , procédons par équivalence

Soit  $(a, b) \in \mathbb{R}^3$

Soit  $y \in x + ]a; b]$ ,  $\exists z \ y = x + z$  où  $a < z \leq b$  i.e.  $x + a < y \leq b + a$  i.e.  $y \in ]a + x; b + x]$

Ainsi  $\forall (x, a, b) \in \mathbb{R}^3, x + ]a; b] = ]a + x; b + x]$

Donc  $\mu_x(]a; b]) = \lambda(]a + x; b + x]) = (b + x) - (a + x) = b - a = \lambda(]a; b])$

3. Pour montrer  $\mu_x = \lambda$  on va utiliser le  $\pi - \lambda$  *theoreme*.

Pour cela on va montrer que  $P\{=]a; b], (a, b) \in \mathbb{R}^2\}$  est un  $\pi - \text{systeme}$ .

Soient deux intervalles réels  $]a; b]$  et  $]c; d]$  alors  $]a; b] \cap ]c; d] = ]\max(a, c); \min(b, d)]$  qui est aussi intervalle semi ouvert donc  $P$  est un  $\pi - \text{systeme}$

Or par la question précédente  $\lambda$  et  $\mu_x$  coïncident sur  $P$ , et elles sont  $\sigma$ -finies sur  $P$ , en effet,  $\bigcup_{n \geq 1} ]-n, n] = \mathbb{R}$  et  $\forall n \in \mathbb{N}, \mu_x(]-n, n]) = \lambda(]-n, n]) < \infty$ . Ainsi le  $\pi - \lambda$  *theoreme* affirme  $\mu_x = \lambda$ .

4. On a donc montré que  $\forall x \in \mathbb{R} \ \lambda = \mu_x$  i.e.  $\forall H \in B(\mathbb{R}), \forall x \in \mathbb{R} \ \lambda(H) = \lambda(x + H)$  i.e. la mesure de Lebesgue est stable par translation.

### Exercice 3.3 ▷

Montrons la contraposée. Soit  $A = \{f = \infty\}$ . Supposons que  $A$  ne soit pas de mesure nulle. Posons alors, pour tout entier  $n$ ,  $g_n$  la fonction positive étagée telle que  $g$  soit nulle en dehors de  $A$ , et égale à  $n$  sur  $A$ . Alors  $\forall n \in \mathbb{N}, \int f d\mu \geq \int g d\mu = n \int 1_A d\mu = n \times \mu(A) \rightarrow +\infty$ . Ainsi  $\int f d\mu = \infty$

### Exercice 3.8 ▷

1. Considérons les suite numériques  $(f_n(x))_n$  où  $\forall x \in \mathbb{R}, f_n(x) = (1 + \frac{x}{n})^n e^{-bx}$  montrons que celles ci sont croissantes. Pour cela nous allons utiliser l'inégalité arithméticogéométrique sur  $(1 + \frac{x}{n})^{\frac{n}{n+1}}$  (Voyons le comme  $\sqrt[n]{1 \times \frac{x}{n} \dots \times \frac{x}{n}}$ ) on a : Soit  $x \in \mathbb{R}$

$$(1 + \frac{x}{n})^{\frac{n}{n+1}} < \frac{1+n(1+\frac{x}{n})}{n+1} = 1 + \frac{x}{n+1}$$

$$\iff (1 + \frac{x}{n})^n < (1 + \frac{x}{n+1})^{n+1}$$

Ainsi  $(f_n(x))_n$  est croissante pour tout  $x$  donc puisque de plus les  $f_n$  sont positives et mesurables le théorème de convergence monotone permet d'écrire

$$\lim_{n \rightarrow \infty} \int_0^\infty (1 + \frac{x}{n})^n = \int_0^\infty \lim_{n \rightarrow \infty} (1 + \frac{x}{n})^n e^{-bx} dx = \int_0^\infty e^x e^{-bx} dx = \frac{1}{1-b} \text{ (car } b > 1)$$

2. On considère la suite de fonction  $(f_n)_n$  définie par  $\forall x \in \mathbb{R} f_n(x) = \frac{1}{x^n + e^x}$  on va majorer cette fonction pour appliquer le théorème de convergence dominée on a :

$$\forall x \in \mathbb{R}^+, \forall n \in \mathbb{N} |f_n(x)| < e^{-x} \text{ et } \int_0^\infty e^{-x} dx = 1 < \infty$$

Donc les fonctions sont bien dominées sur  $\mathbb{R}^+$  de sorte qu'en passant à la limite sous l'intégrale on obtient

$$\lim_{n \rightarrow \infty} \int_0^\infty \frac{dx}{x^n + e^x} = \int_0^1 \lim_{n \rightarrow \infty} \frac{dx}{x^n + e^x} + \int_1^\infty \lim_{n \rightarrow \infty} \frac{dx}{x^n + e^x} = 1 - e^{-1} + 0$$

**Exercice 3.11** ▷ Pour faire apparaître le  $e^{-x}$  on introduit le changement de variable  $t = x^n$ , ainsi

$$\int_1^\infty n e^{-x^n} dx = \int_1^\infty n e^{-t} \frac{1}{n t^{1-\frac{1}{n}}} dt = \int_1^\infty t^{\frac{1}{n}-1} e^{-t} dt$$

On pose ainsi  $\forall x \in ]1, \infty[ f_n(x) = x^{\frac{1}{n}-1} e^{-x}$

on a  $f_n \xrightarrow{cvs} \frac{e^{-x}}{x}$

et  $|f_n(x)| \leq t^2 e^{-t} = \phi(t) = o(\frac{1}{t^2})$  pour  $t \rightarrow +\infty$  donc  $\phi$  est  $\lambda$ -intégrable sur  $]1, \infty[$

D'après le théorème de convergence dominée

$$\lim_{n \rightarrow \infty} \int_1^\infty n e^{-x^n} dx = \lim_{n \rightarrow \infty} \int_1^\infty x^{\frac{1}{n}-1} e^{-x} dx = \int_1^\infty \lim_{n \rightarrow \infty} x^{\frac{1}{n}-1} e^{-x} dx = \int_1^\infty \frac{e^{-x}}{x} dx$$

**Exercice 3.12** ▷ 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 ;
- $g_n(x) = |f(x, t) - f(x, t_n)| \xrightarrow{n \rightarrow \infty} 0, \mu$ -p.p. ;
- $|g_n(x)| = |f(x, t) - f(x, t_n)| \leq 2G(x), \mu$ -p.p. ;
- $\int G d\mu < \infty$  ;

donc on peut appliquer le théorème de convergence dominée (théorème 3.14) 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.

**Exercice 3.16** ▷

1. La construction de la mesure d'un ensemble.
- 2.

$$\lambda^*(E) = \inf \left\{ \sum_{i=1}^{\infty} |I_i| : (I_i)_i \text{ intervalles, } E \subset \cup_{i=1}^{\infty} I_i \right\}.$$

3.  $E$  est mesurable si  $\lambda^*(E) + \lambda^*((a, b) \setminus E) = b - a$ .
4. i) Il est clair que  $\emptyset \in \mathcal{L}$ . ii) La définition de la mesurabilité implique trivialement que si  $E$  est mesurable, alors le complémentaire  $(a, b) \setminus E$  est aussi mesurable. iii) Soit  $E_1, E_2, \dots \in \mathcal{L}$ . L'affirmation de Lebesgue implique que  $\cup_k E_k \in \mathcal{L}$ .
5. Tout intervalle de  $(a, b)$  est mesurable au sens de Lebesgue. Si  $\mathcal{A}$  représente la classe des intervalles, on a donc  $\mathcal{A} \subset \mathcal{L}$ . Donc  $\sigma(\mathcal{A}) \subset \mathcal{L}$ .
6. Pour tout  $E \in \mathcal{L}$ , il existe  $E' \in \mathcal{B}((a, b))$  tel que  $E \subset E'$ , et  $E' \setminus E$  est de mesure nulle. Autrement dit, chaque lebesguien coïncide avec un borélien, à un ensemble de mesure nulle près. D'après J-P. Kahane, il est arrivé à E. Borel de dire que la contribution d'H. Lebesgue avait été d'apporter les ensembles de mesure nulle, ce qui aurait été une cause de fâcherie. Naturellement, la contribution va bien au-delà, puisque H. Lebesgue résout le problème de la construction de la mesure qui porte aujourd'hui son nom.
7. La  $\sigma$ -additivité.
8. C'est un espace mesuré.
9. D'après Lebesgue, la fonction  $f$  est dite sommable si  $f^{-1}((A, B]) \in \mathcal{L}$ , pour tout  $A, B$  dans  $[m, M]$ . D'après le lemme 2.14 du poly, cela revient à dire que  $f$  est mesurable, en prenant pour  $\mathcal{L}$  la tribu de départ, et pour tribu d'arrivée la tribu engendrée par les intervalles de la forme  $(A, B]$ . Cette dernière n'est autre que  $\mathcal{B}([m, M])$ , d'après le cours.
10.  $E_0 = y^{-1}(\{m\})$  et  $E_i = y^{-1}((m_{i-1}, m_i])$ .
11.  $I$  et  $J$  sont respectivement les intégrales des fonctions  $\varphi$  et  $\psi$  définies par :

$$\varphi = m_0 1_{E_0} + \sum_{i=1}^p m_i 1_{E_i} ; \quad \psi = m_0 1_{E_0} + \sum_{i=1}^p m_{i-1} 1_{E_i}$$

- 12.

$$\left| \int y d\lambda - I \right| = \left| \int (y - \varphi) d\lambda \right| \leq \int |y - \varphi| d\lambda.$$

Or  $|y - \varphi| \leq \max_i (m_i - m_{i-1})$ . On a donc l'inégalité voulue, avec  $C = b - a$ . Idem pour  $\psi$ .

13. Simple passage à la limite dans la question précédente.
14. Le théorème d'échange limite-intégrale, autrement dit, le théorème de convergence dominée

**Exercice 3.18** ▷

1. calculons

$$\int |f(x, y)| dy = \sin(x) \left[ -\frac{1}{x} e^{-xy} \right]_{y=0}^{y=\infty} = \frac{\sin(x)}{x}$$

or cette dernière expression est intégrable sur le segment  $[0, a]$  donc  $f$  est intégrable

2.  $f$  est integrable donc d'apres Fubini on peut integrer dans l'ordre qu'on souhaite.

Posons  $A = \int_0^a f(x, y) dx$

ainsi par IPP on trouve

$$A = \int_0^a e^{-xy} \sin(x) dx = 1 - \cos(a)e^{-ay} - y \int_0^a \cos(x)e^{-xy} dx$$

Grace a une deuxieme IPP on obtient l'equation

$$A = 1 - \cos(a)e^{-ay} - y(\sin(a)e^{-ay} + yA)$$

$$\text{D'ou } A = \frac{1}{1+y^2} - \cos(a) \frac{e^{-ay}}{1+y^2} - \sin(a) \frac{ye^{-ay}}{1+y^2}$$

Ainsi

$$\begin{aligned} \iint f(x, y) dy dx &= \iint f(x, y) dx dy \\ \int_0^a \frac{\sin(x)}{x} dx &= \int_0^\infty A dy \\ \int_0^a \frac{\sin(x)}{x} dx &= \frac{\pi}{2} - \cos a \int_0^\infty \frac{e^{-ay}}{1+y^2} dy - \sin a \int_0^\infty \frac{ye^{-ay}}{1+y^2} dy \end{aligned}$$

3.

$$\begin{aligned} \left| \int_0^a \frac{\sin(x)}{x} dx - \frac{\pi}{2} \right| &= \left| \cos a \int_0^\infty \frac{e^{-ay}}{1+y^2} dy + \sin a \int_0^\infty \frac{ye^{-ay}}{1+y^2} dy \right| \\ &\leq \int_0^\infty \frac{(1+y)e^{-ay}}{1+y^2} dy \text{ par inegalite triangulaire et } |\sin|, |\cos| \leq 1 \\ &\leq 2 \int_0^\infty e^{-ay} dy \text{ car } \frac{1+y}{1+y^2} \leq 2 \text{ (*)} \\ &\leq \frac{2}{a} \end{aligned}$$

Puis en faisant tendre  $a$  vers  $+\infty$  on obtient

$$\int_0^{+\infty} \frac{\sin(x)}{x} dx = \frac{\pi}{2}$$

Il s'agit de l'integrale de Dirichlet-Fresnel

Rm : Montrons (\*), le trinome  $2X^2 - X + 1$  est de discriminant negatif  $\Delta = -7$

$$\text{donc } \forall y \in \mathbb{R}^+ 2y^2 - y + 1 \geq 0 \Leftrightarrow \frac{1+y}{1+y^2} \leq 2$$

**Exercice 3.21** ▷

3.21.a) On doit vérifier les trois axiomes qui définissent une tribu.

- $T^{-1}(\emptyset) = \emptyset$  donc  $\emptyset \in \mathcal{I}$ .
- Rappelons que  $T^{-1}(A^c) = (T^{-1}(A))^c$  donc si  $A \in \mathcal{I}$ ,  $T^{-1}(A^c) = A^c$  soit  $A^c \in \mathcal{I}$ .
- De même, pour une famille dénombrable d'ensembles  $(A_n, n \geq 1)$ ,

$$T^{-1}\left(\bigcup_{n \geq 1} A_n\right) = \bigcup_{n \geq 1} T^{-1}(A_n).$$

Par conséquent, si les  $A_n$  appartiennent tous à  $\mathcal{I}$ ,

$$T^{-1}\left(\bigcup_{n \geq 1} A_n\right) = \bigcup_{n \geq 1} A_n \iff \bigcup_{n \geq 1} A_n \in \mathcal{I}.$$

**3.21.b)** Nous devons montrer que pour tout  $A \in \mathcal{B}(\mathbb{R})$ ,  $f^{-1}(A)$  appartient à  $\mathcal{I}$ . Ce qui revient à montrer que  $T^{-1} \circ f^{-1}(A) = f^{-1}(A)$ . Or par définition d'une fonction invariante  $f \circ T = f$  et par définition des images réciproques,  $(f \circ T)^{-1} = T^{-1} \circ f^{-1}$ .

**3.21.c)** Supposons que le système soit ergodique. Soit  $f$  une fonction invariante, pour tout  $\alpha \in \mathbb{R}$ ,

$$\mathbf{P}(f^{-1}(] - \infty, \alpha]) = 0 \text{ ou } 1.$$

D'autre part, la fonction

$$G_f : \alpha \mapsto \mathbf{P}(f^{-1}(] - \infty, \alpha])$$

est croissante et continue à droite. Donc il existe  $\alpha_0$  tel que

$$\forall \beta < \alpha, G_f(\beta) = 0 \text{ et } G_f(\beta) = 1, \text{ pour } \beta \geq \alpha.$$

Ceci signifie que  $\mathbf{P}(f^{-1}(\{\alpha_0\})) = 1$  donc que  $f = \alpha$   $\mathbf{P}$ -presque partout.

Réciproquement, supposons que toutes les fonctions invariantes  $f$  sont presque sûrement constantes. Si le système n'est pas ergodique, il existe un ensemble invariant de mesure non nulle :

$$\exists A \in \mathcal{E}, T^{-1}(A) = A \text{ et } \mathbf{P}(A) > 0.$$

Considérons alors la fonction  $f = \mathbf{1}_A$ . Elle est invariante par  $T$  puisque

$$f \circ T = \mathbf{1}_A \circ T = \mathbf{1}_{T^{-1}(A)} = \mathbf{1}_A = f,$$

et prend les valeurs 0 et 1 sur des ensembles de mesure non nulle, donc elle n'est pas constante, d'où une contradiction. Par conséquent, tous les ensembles  $T$ -invariants sont de mesure nulle.

**3.21.d)** Soit  $A$  un ensemble invariant. Appliquons (3.13) à  $f = g = \mathbf{1}_A$ . Comme  $f \circ T = f$ , il vient

$$\lim_{n \rightarrow \infty} \int_E \mathbf{1}_A^2 d\mathbf{P} = \left( \int_E \mathbf{1}_A d\mathbf{P} \right)^2,$$

soit  $\mathbf{P}(A) = \mathbf{P}(A)^2$  donc  $\mathbf{P}(A) = 0$  ou  $\mathbf{P}(A) = 1$ . Ce qui signifie que le système est ergodique.

**3.21.e)** Il suffit de le montrer pour  $n = 1$ , le résultat en découle par récurrence. Pour  $f = \mathbf{1}_A$ , c'est l'hypothèse d'invariance de  $\mathbf{P}$  par  $T$  qui permet d'écrire

$$\int_E \mathbf{1}_A \circ T d\mathbf{P} = \int_E \mathbf{1}_{T^{-1}(A)} d\mathbf{P} = \mathbf{P}(T^{-1}(A)) = \mathbf{P}(A) = \int_E \mathbf{1}_A d\mathbf{P}.$$

Par linéarité, l'équation (3.14) est vraie pour toutes les fonctions étagées positives puis par passage à la limite pour toutes les fonctions mesurables positives. Comme  $L^2(\mathbf{P}) \subset L^1(\mathbf{P})$ , toute fonction  $f$  de  $L^2$  s'écrit  $f = f^+ - f^-$  avec  $f^+$  et  $f^-$  mesurables, positives et d'intégrale finie. Par différence, l'équation (3.14) est vraie pour tous les éléments de  $L^2(\mathbf{P})$ .

**3.21.f)** Soit  $D$  l'ensemble sur lequel (3.13) est vérifiée. Soit  $f$  et  $g$  des éléments de  $L^2(\mathbf{P})$ . Pour tout  $\epsilon > 0$ , il existe  $f_\epsilon$  et  $g_\epsilon$  appartenant à  $D$  tels que  $\|f - f_\epsilon\|_2 < \epsilon$  et  $\|g - g_\epsilon\|_2 < \epsilon$ .

$$\begin{aligned} \int_E f \circ T^n g d\mathbf{P} &= \int_E (f \circ T^n - f_\epsilon \circ T^n) g d\mathbf{P} \\ &\quad + \int_E f_\epsilon \circ T^n (g - g_\epsilon) d\mathbf{P} + \int_E f_\epsilon \circ T^n g_\epsilon d\mathbf{P} := A_n + B_n + C_n. \end{aligned}$$

Par Cauchy-Schwarz et la question précédente,

$$|A_n| \leq \left( \int_E |(f - f_\epsilon) \circ T^n|^2 d\mathbf{P} \right)^{1/2} \|g\|_2 = \|f - f_n\|_2 \|g\|_2 \leq \epsilon \|g\|_2.$$

De même,

$$|B_n| \leq \|f \circ T^n\|_2 \|g - g_\epsilon\|_2 = \|f\|_2 \epsilon.$$

Enfin,

$$\begin{aligned} \left| C_n - \int_E f d\mathbf{P} \int_E g d\mathbf{P} \right| &\leq \left| \int_E (f_\epsilon - f) d\mathbf{P} \int_E g_\epsilon d\mathbf{P} \right| + \left| \int_E f d\mathbf{P} \int_E (g - g_\epsilon) d\mathbf{P} \right| \\ &\leq \|f - f_\epsilon\|_2 \|g_\epsilon\|_2 + \|f\|_2 \|g - g_\epsilon\|_2. \end{aligned}$$

D'après l'inégalité triangulaire,  $\|g_\epsilon\|_2 \leq \|g\|_2 + \epsilon$  donc

$$\left| C_n - \int_E f d\mathbf{P} \int_E g d\mathbf{P} \right| \leq \epsilon(\|g\|_2 + \epsilon + \|f\|_2).$$

Si l'on pose

$$I_n(f, g) = \int_E f \circ T^n g d\mathbf{P} - \int_E f d\mathbf{P} \int_E g d\mathbf{P},$$

on a donc montré que pour tout  $n \geq 1$ , qu'il existe  $c > 0$  tel que

$$|I_n(f, g) - I_n(f_\epsilon, g_\epsilon)| \leq c\epsilon.$$

On choisit  $n$  tel que  $|I_n(f_\epsilon, g_\epsilon)| \leq \epsilon$  pour  $m \geq n$ , et alors dans ces conditions,  $|I_m(f, g)| \leq (c + 1)\epsilon$ .

3.21.g) La fonction  $T_1$  est appelée *fonction tente* en raison de la forme de son graphe.

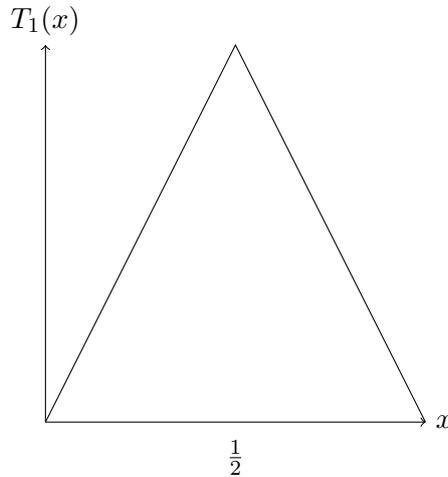


Figure B.1. – Fonction  $T_1$ .

Pour démontrer que  $\lambda$  est invariante par  $T_1$ , il suffit que l'on démontre que

$$\lambda(T_1^{-1}([a, b])) = \lambda([a, b]) = b - a,$$

pour tout  $a$  et  $b$ . Or

$$T_1^{-1}([a, b]) = [a/2, b/2] \cup [1 - b/2, 1 - a/2],$$

dont la mesure de Lebesgue est bien  $b - a$ .

3.21.h) Par définition,

$$e_k(T^n(x)) = \exp(2^n i \pi k x) \mathbf{1}_{[0, 1/2]}(x) + \exp(-2^n i \pi k x) \mathbf{1}_{[1/2, 1]}(x).$$

Par conséquent,

$$\int_0^1 e_k(T^n(x)) e_l(x) dx = \int_0^{1/2} e_{2^{n-1}k}(x) e_l(x) dx + \int_{1/2}^1 e_{-2^{n-1}k}(x) e_l(x) dx.$$

D'après le lemme de Riemman, ces deux intégrales tendent vers quand  $n$  tend vers l'infini. On a donc bien

$$\int_0^1 e_k(T^n(x)) e_l(x) dx = 0 = \int_0^1 e_l(x) dx \int_0^1 e_k(x) dx.$$

3.21.i) Pour identifier la mesure  $\mu$ , on procède en utilisant le théorème ??, soit  $\psi$  une fonction test, on a

$$\int_0^1 \psi(x) d\mu(x) = \int_0^1 \psi(T_1(x)) dx.$$

L'application  $\Theta$  est bijective de  $[0, 1]$  dans lui-même, on peut donc faire le changement de variable  $y = \Theta(x)$  donc

$$dy = \pi \sin(\pi x/2) \cos(\pi x/2) dx = \pi \sqrt{y} \sqrt{1-y} dx.$$

Par conséquent,

$$\int_0^1 \psi(T_1(x)) dx = \int_0^1 \psi(y) \frac{1}{\pi \sqrt{y} \sqrt{1-y}} dy.$$

D'où  $d\mu(y) = (\pi \sqrt{y(1-y)})^{-1} dy$ .

3.21.j) On a le diagramme suivant

$$\begin{array}{ccc} [0, 1[ & \xrightarrow{\Theta} & [0, 1[ \\ T_1 \downarrow & & \downarrow T \\ [0, 1[ & \xrightarrow{\Theta} & [0, 1[ \end{array}$$

Supposons que l'on ait démontré que

$$T \circ \Theta = \Theta \circ T_1.$$

Soit  $A$  invariant par  $T$ , i.e.  $T^{-1}(A) = A$ , et soit  $B = \Theta^{-1}(A)$ . On a alors

$$\begin{aligned} T_1^{-1}(B) &= T_1^{-1} \circ \Theta^{-1}(A) = (\Theta \circ T_1)^{-1}(A) = (T \circ \Theta)^{-1}(A) \\ &= \Theta^{-1}(T^{-1}(A)) = \Theta^{-1}(A) = B. \end{aligned}$$

Par conséquent,  $B$  est invariant par  $T_1$  donc  $\lambda(B) = 0$  ou  $\lambda(B) = 1$ . Comme  $\mu$  est la mesure image de  $\lambda$  par  $\Theta$ ,

$$\lambda(B) = \lambda(\Theta^{-1}(A)) = \mu(A),$$

on a bien  $\mu(A) = 0$  ou  $\mu(A) = 1$ , soit  $(E, T, \mu)$  ergodique. En vertu de théorème de Birkhoff, on a bien

$$\lim_{N \rightarrow \infty} \sum_{n=1}^N \mathbf{1}_{[a,b]}(x_n) = \int_a^b \frac{1}{\pi \sqrt{x(1-x)}} dx.$$

**Exercice 4.1** ▷ On utilise le thm de transfert (Thm. 4.3). Pour toute  $\psi$  continue bornée, par changement de variable  $y = -x$

$$\mathbb{E}(\psi(-X)) = \int \psi(-x)f(x)dx = \int \psi(y)f(-y)dy.$$

Cela induit bien que  $f_{-X}(x) = f(-x)$ .

**Exercice 4.4** ▷ Pour commencer on rappelle la relation suivante, valable pour tous les ensembles  $A, B$  tels que  $A \subset B$  :

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A).$$

La preuve repose essentiellement sur la relation :

$$B = B \cup A = (B \setminus A) \uplus A,$$

puis par un passage aux probabilités.

1. On en déduit ainsi :

—

$$\begin{aligned} \mathbb{P}(a < X \leq b) &= \mathbb{P}(\{X \in ]-\infty, b]\} \setminus \{X \in ]-\infty, a]\}) \\ &= \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) \\ &= F_X(b) - F_X(a). \end{aligned}$$

—

$$\begin{aligned} \mathbb{P}(a < X < b) &= \mathbb{P}(\{X \in ]-\infty, b[ \} \setminus \{X \in ]-\infty, a]\}) \\ &= \mathbb{P}(X < b) - \mathbb{P}(X \leq a) \\ &= \mathbb{P}(\{X \leq b\} \setminus \{X = b\}) - \mathbb{P}(X \leq a) \\ &= \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) - \mathbb{P}(X = b) \\ &= F_X(b) - F_X(a) - \mathbb{P}(X = b). \end{aligned}$$

—

$$\begin{aligned} \mathbb{P}(a \geq X < b) &= \mathbb{P}(\{X \in ]-\infty, b[ \} \setminus \{X \in ]-\infty, a]\}) \\ &= \mathbb{P}(X < b) - \mathbb{P}(X < a) \\ &= F_X(b) - F_X(a) + \mathbb{P}(X = a) - \mathbb{P}(X = b). \end{aligned}$$

En particulier, quand  $X$  est à densité, ou plus généralement quand sa fonction de répartition est continue à gauche (donc continue, car une fonction de répartition est toujours continue à droite) en  $a$  et en  $b$ , toutes ces quantités sont égales.

2. Comme  $X$  a même loi que  $(b-a)U + a$ , où  $U$  suit une loi uniforme sur  $[0, 1]$ , le support de  $X$  est bien  $X(\Omega) = (b-a)[0, 1] + a = [a, b]$ . Donc la fonction de répartition de  $X$  est nulle sur  $] -\infty, a[$  et égale à 1 sur  $]b, +\infty[$ . Pour  $x \in [a, b]$ , on a :

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}((b-a)U + a \leq x) = \mathbb{P}(U \leq \frac{x-a}{b-a}) = \frac{x-a}{b-a},$$

car cette dernière quantité est bien comprise entre 0 et 1 vu l'hypothèse sur  $x$ . De plus on inverse pas le sens des inégalités car  $b-a$  est strictement positif. Finalement :

$$\forall x \in \mathbb{R}, F_X = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } x \in [a, b] \\ 1 & \text{si } x > b. \end{cases}$$

### Exercice 4.5 ▷

- Tout d'abord on donne le support de  $Y = aX + b$ , c'est-à-dire l'image par  $Y$  de  $\Omega$ , ou encore de façon plus intuitive (ce n'est pas une définition formelle) les valeurs que peut prendre  $Y$  "avec probabilité strictement positive"<sup>1</sup>. Comme on suppose que  $X(\Omega) = \mathbb{R}$ , on a si  $a \neq 0$  que  $Y(\Omega) = \mathbb{R}$ . Sinon  $a = 0$  et  $Y(\Omega) = \{b\}$ . Calculer le support d'une variable aléatoire permet de déterminer les "valeurs intéressantes" en lesquelles calculer sa fonction de répartition, et son éventuelle densité. Cela évite aussi de nombreux pièges en révélant les distinctions de cas à faire dès le départ (ici sur la valeur de  $a$ ).
- Lorsqu'on travaille sur une fonction de répartition, on manipule essentiellement une probabilité. Puisque deux événements égaux ont même probabilité<sup>2</sup>, il est intéressant dans ce contexte de manipuler directement des événements plutôt que d'utiliser des méthodes plus analytiques (comme le changement de variable, vu plus loin dans ce cours). Concrètement, cela revient à exprimer l'évènement  $\{aX + b \leq x\}$  comme un évènement portant seulement sur  $X$ , dont on suppose connaître la loi, et donc la fonction de répartition.

Ici on trouve :

$$\{aX + b \leq x\} = \left\{ X \leq \frac{x-a}{b} \right\}.$$

Si  $a > 0$ , on en déduit :

$$F_{aX+b}(x) = \mathbb{P}(aX + b \leq x) = \mathbb{P}\left(X \leq \frac{x-a}{b}\right) = F_X\left(\frac{x-a}{b}\right).$$

Si  $a < 0$ , alors  $aX + b \leq x \Leftrightarrow X \geq \frac{x-a}{b}$ , et donc :

$$F_{aX+b}(x) = \mathbb{P}(aX + b \leq x) = \mathbb{P}\left(X \geq \frac{x-a}{b}\right) = 1 - \mathbb{P}\left(X \leq \frac{x-a}{b}\right) = 1 - F_X\left(\frac{x-a}{b}\right).$$

1. Façon de parler, une variable aléatoire à densité (disons  $f_X$ ) ne prend aucune valeur avec une probabilité strictement positive par exemple. Dans ce cas particulier, le support est  $X(\Omega) = \overline{f_X^{-1}(\mathbb{R}_+^*)}$ .

2. La réciproque étant archi-fausse.

Enfin si  $a = 0$ ,

$$F_{aX+b}(x) = \mathbb{P}(b \leq x) = \mathbf{1}_{\{b \leq x\}}.$$

### Exercice 4.6 ▷

1. Par définition :

$$F_U(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

2. On utilise de nouveau l'équivalence :

$$X \leq x \Leftrightarrow F^{-1}(U) \leq x \Leftrightarrow U \leq F(x),$$

en utilisant que  $F$  est croissante pour garder l'inégalité dans le même sens. On en déduit que :

$$\begin{aligned} F_X(x) &= \mathbb{P}(F^{-1}(U) \leq x) \\ &= \mathbb{P}(U \leq F(x)) \\ &= F(x), \end{aligned}$$

puisque  $F$  est une fonction de répartition, et donc  $F(x) \in [0, 1]$  quelque soit  $x \in \mathbb{R}$ . En d'autres termes,  $X$  suit la loi donnée par la fonction de répartition  $F$ .

3. Ce résultat fournit une méthode générale pour simuler une loi de probabilité donnée, pourvu que sa fonction de répartition soit au moins bijective (une fonction de répartition est **toujours** croissante). Pour cela, on simule  $U$  suivant une loi uniforme (grâce au générateur de nombre aléatoires de Python, disons *via* la fonction `numpy.random.rand()`). Puis on calcule  $X := F^{-1}(U)$ . Par le résultat précédent,  $X$  est une réalisation de la loi cible.

REMARQUE.— En pratique, à part des cas simples comme ceux ci-dessous, cette méthode<sup>3</sup> a surtout un intérêt théorique : elle montre qu'on peut simuler n'importe quelle loi de probabilité, y compris de "gros" objets aléatoires (matrices aléatoires, suites aléatoires, processus stochastiques, *etc.*) uniquement en générant un seul nombre aléatoire. Dans la pratique  $F$  est rarement explicite (penser à la fonction de répartition d'une loi gaussienne!), sans parler de sa bijection réciproque (ou pseudo-inverse<sup>4</sup> lorsqu'on a des hypothèses plus faibles qu'ici). Des méthodes plus efficaces seront vues dans le cours sur les méthodes de Monte-Carlo.

---

3. Modulo une généralisation à un cadre plus large que le nôtre.

4. Définie par  $F^{-1}(t) := \inf\{x \in \mathbb{R} \mid F(x) \geq t\}$ .

4. La fonction de répartition d'une loi exponentielle  $\mathcal{E}(\lambda)$  est strictement croissante et continue sur  $\mathbb{R}^+$ , et son inverse est explicite :

$$\forall u \in ]0, 1[, F^{-1}(u) = -\frac{1}{\lambda} \log(1 - u).$$

De ce qui précède, on en déduit que  $X := -\frac{1}{\lambda} \log(1 - U)$  est une réalisation de la loi exponentielle  $\mathcal{E}(\lambda)$  si  $U \hookrightarrow \mathcal{U}([0, 1])$ .

### Exercice 4.7 ▷

Commençons par calculer le support de  $Y$ . On a que  $Y(\Omega) = \exp(X(\Omega)) = \exp(\mathbb{R}) = \mathbb{R}_+$ . Donc les densités des  $Y$  sont nulles presque partout sur  $\mathbb{R}_-$ . Pour calculer  $f_Y(y)$  avec  $y \in \mathbb{R}_+^*$ , on peut utiliser la méthode du changement de variable ou calculer la fonction de répartition puis la dériver en ses points de continuité. Donnons les deux approches :

- **(Fonction de répartition)** Par croissance et bijectivité<sup>5</sup> de la fonction  $\log$  sur  $\mathbb{R}_+^*$ , on a les égalités suivantes :

$$\begin{aligned} \mathbb{P}(Y \leq y) &= \mathbb{P}(\exp(X) \leq y) \\ &= \mathbb{P}(X \leq \log y) \\ &= F_X(\log y). \end{aligned}$$

En dérivant par rapport à  $y \in \mathbb{R}_+^*$ , on trouve finalement :

$$\forall y \in \mathbb{R}_+^*, f_Y(y) = \frac{1}{y} f_X(\log y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\log y - \mu)^2}{2\sigma^2}}.$$

- **(Changement de variable)** La méthode est toujours la même : pour calculer une densité de  $Y = g(X)$ , où  $g$  et une densité  $f_X$  de  $X$  sont connues, on se donne  $\varphi : \mathbb{R} \mapsto \mathbb{R}$  une fonction continue bornée (ou borélienne positive, ou continue à support compact, *etc.*<sup>6</sup>), on "calcule"  $\mathbb{E}[\varphi(Y)] = \mathbb{E}[\varphi(g(X))]$  en faisant un changement de variable pour remplacer  $g(x)$  par  $x$ . La fonction "à côté" de  $\varphi(x)$  constitue une densité de  $Y$ . Elle s'écrit formellement comme le produit du Jacobien de la transformation  $y = g(x)$  et de  $f_X(g^{-1}(y))$ . Dans notre contexte,  $g$  est la fonction exponentielle et son Jacobien est la fonction  $y \mapsto 1/y$  :

$$\begin{aligned} \mathbb{E}[\varphi(Y)] &= \mathbb{E}[\varphi(\exp(X))] \\ &= \int_{\mathbb{R}} \varphi(e^x) f_X(x) dx \\ &= \int_{\mathbb{R}_+^*} \varphi(y) \frac{1}{y} f_X(\log y) dy. \end{aligned}$$

5. Ces deux arguments sont essentiels !

6. Essentiellement il faut que  $\varphi$  soit mesurable, telle que  $\mathbb{E}[\varphi(Y)]$  ait toujours un sens, éventuellement dans  $\overline{\mathbb{R}_+}$  et que la classe de fonctions dans laquelle on prend  $\varphi$  soit assez "large".

Et comme précédemment, on retrouve bien :

$$\forall y \in \mathbb{R}d+^*, f_Y(y) = \frac{1}{y} f_X(\log y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\log y - \mu)^2}{2\sigma^2}}.$$

Cette méthode est très puissante et permet de gérer des changements de lois qu'il serait compliqué de réaliser avec les fonctions de répartition. C'est particulièrement vrai dès que plusieurs variables aléatoires sont présentes dans le changement de loi (voir les exercices suivants). En contrepartie on est essentiellement limité à des transformations *via* des  $C^1$ -difféomorphismes (par exemple il semble compliqué de calculer la loi de  $\max(X, Y)$  quand  $X, Y$  sont i.i.d., alors que ce calcul est simplissime avec les fonctions de répartition). Enfin, les changements de variable ne permettent pas de calculer directement la fonction de répartition, seulement une densité, ce qui suppose au préalable que les variables en jeu admettent des densités.

REMARQUE.— — La loi de  $Y$  porte le nom de **loi log-normale**, car  $\log Y$  suit une loi normale (par définition!). Elle joue un rôle important dans le modèle de Black-Scholes, modèle  $\hat{O}$  combien fondamental en finance malgré ses innombrables défauts.

- La loi log-normale fournit aussi un contre-exemple au **problème des moments** : la connaissance seule des moments<sup>7</sup>  $\mathbb{E}[Y^n]$  de  $Y$  ne permet pas de retrouver la loi de  $Y$ . Cela reste un cas pathologique : dans la plupart des applications, les lois de probabilités rencontrées sont caractérisées par leurs moments, par exemple *via* leur fonction caractéristique qui est souvent de classe  $C^\infty$ .

### Exercice 4.8 ▷

Une densité de probabilité intègre toujours à 1, donc on doit avoir :

$$\int_{\mathbb{R}} f(x) dx = c \int_{\mathbb{R}} \frac{1}{1+x^2} dx = 1.$$

La fonction  $x \mapsto (1+x^2)^{-1}$  admet  $\arctan$  comme primitive, elle même de limite égale à  $\pi/2$  en  $+\infty$  (et  $-\pi/2$  en  $-\infty$ ). Donc on trouve :

$$c = \frac{\pi}{2} + \frac{\pi}{2} = \pi.$$

Pour calculer la loi de  $Y = 1/X$ , le plus simple est d'utiliser la méthode du changement de variable, en faisant  $y = 1/x$ . Cette transformation n'est pas un  $C^1$ -difféomorphisme à proprement dit, mais son seul point de discontinuité est en 0, et les singletons sont de mesure de Lebesgue nulle. On trouve donc :

---

7. Que l'on peut facilement expliciter : pouvez-vous les calculer ?

$$\begin{aligned}
\mathbb{E}\left[\varphi\left(\frac{1}{X}\right)\right] &= \int_{\mathbb{R}} \varphi\left(\frac{1}{x}\right) f(x) dx \\
&= \int_{\mathbb{R}} \varphi(x) \left|\frac{1}{x^2}\right| f\left(\frac{1}{x}\right) dx \\
&= \int_{\mathbb{R}} \varphi(x) \frac{1}{\pi} \frac{1}{x^2} \frac{1}{1 + \frac{1}{x^2}} dx \\
&= \int_{\mathbb{R}} \varphi(x) f(x) dx
\end{aligned}$$

Et oh surprise, on en déduit que  $1/X$  a même loi que  $X$  ! (ce qui bien sûr ne veut pas dire qu'elles prennent les mêmes valeurs presque sûrement, seulement qu'elles ont même densité, fonction de répartition, *etc.*)

REMARQUE.— La loi de Cauchy fait partie de la famille des **lois stables**. Ce sont des distributions de probabilités stables pour la convolution, modulo une mise-à-l'échelle appropriée. Par exemple la moyenne (arithmétique) de deux variables aléatoires indépendantes suivant une loi de Cauchy suit encore une loi de Cauchy. Par récurrence, c'est encore vrai pour la moyenne de  $n$  variables aléatoires indépendantes de même loi de Cauchy. En particulier, quand  $n$  tend vers l'infini, cette moyenne empirique garde la même loi et donc reste aléatoire ; elle ne tend pas vers une constante comme le suggérerait la loi des grands nombres : la loi de Cauchy n'admet en effet pas d'espérance !

#### Exercice 4.10 ▷

1. La méthode de la fonction de répartition est aussi simple que celle de la fonction muette ici, et donne en prime la fonction de répartition. Le support de  $Y(\Omega)$  est  $\mathbb{R}_+$ , donc on a déjà que pour  $y < 0$ ,  $F_Y(y) = 0$ . Pour  $y > 0$ , puisque  $X$  ne prend que des valeurs positives  $\mathbb{P}$ -p.s., on trouve :

$$\begin{aligned}
F_Y &= \mathbb{P}(X^2 \leq y) \\
&= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\
&= \mathbb{P}(X \leq \sqrt{y}) \\
&= 1 - e^{-\lambda\sqrt{y}}.
\end{aligned}$$

On pourrait dériver cette fonction en  $y$  pour trouver une densité de  $Y$ , mais ce n'est pas explicitement demandé par l'énoncé (la fonction de répartition caractérise la loi d'une variable aléatoire).

REMARQUE.— La loi obtenue est une loi de Weibull de paramètre  $1/2$ . Cette famille de lois joue un rôle important dans l'étude d'un maximum(/minimum) de variables aléatoires indépendantes (c'est une **loi d'extremum généralisée**). Elle intervient aussi en analyse de fiabilité, aux côtés de la loi de Fréchet et de la loi exponentielle.

2. On pourrait réécrire :

$$Y = X\mathbf{1}_{\{X \leq 1\}} + 2X\mathbf{1}_{\{X > 1\}} = X(\mathbf{1}_{\{X \leq 1\}} + 2\mathbf{1}_{\{X > 1\}}) = g(X),$$

avec  $g = x \mapsto x(\mathbf{1}_{\{x \leq 1\}} + 2\mathbf{1}_{\{x > 1\}})$ . La méthode de la fonction de répartition est plus naturelle ici (et donne la fonction de répartition de  $Y$ ), mais on va plutôt utiliser celle de la fonction muette afin de montrer comment gérer la discontinuité en 1 de  $g$ , qui n'est donc pas un  $\mathcal{C}^1$ -difféomorphisme. On a en séparant les intégrales en 1 :

$$\begin{aligned} \mathbb{E}[\varphi(X)] &= \int_{[0,1]} \varphi(x)\lambda e^{-\lambda x} dx + \int_{[1,+\infty[} \varphi(2x)\lambda e^{-\lambda x} dx \\ &= \int_{[0,1]} \varphi(x)\lambda e^{-\lambda x} dx + \int_{[1/2,+\infty[} \varphi(x)\lambda \frac{1}{2} e^{-\frac{\lambda}{2}x} dx \\ &= \int_{[0,+\infty[} \varphi(x) \left[ \lambda e^{-\lambda x} \mathbf{1}_{\{x \leq 1\}} + \frac{1}{2} e^{-\frac{\lambda}{2}x} \mathbf{1}_{\{x > 1/2\}} \right] dx, \end{aligned}$$

et donc :

$$f_Y(y) = \lambda e^{-\lambda y} \mathbf{1}_{\{y \leq 1\}} + \frac{1}{2} e^{-\frac{\lambda}{2}y} \mathbf{1}_{\{y > 1/2\}}$$

Remarquer que je ne suis pas cohérent sur le sens des crochets (d'un côté on intègre sur  $[1/2, +\infty[$ , de l'autre il y a une inégalité stricte dans l'indicatrice  $\mathbf{1}_{\{y > 1/2\}}$ ). Cela ne change rien au fait que cette densité est bien *une* densité de  $Y$  : changer la valeur d'une densité de probabilité en un nombre (par exemple) dénombrable de points ne change pas la loi associés, car ces points forment un ensemble négligeable pour la mesure de Lebesgue.

3.

**Exercice 4.11**  $\triangleright$  Par hypothèse,  $F_X = F_{2X}$ , et donc

$$F_{2X}(x) = \mathbb{P}(2X \leq x) = \mathbb{P}(X \leq x/2) = F_X(x/2).$$

Donc  $F_X(x) = F_X(x/2)$ , ou encore  $F_X(2x) = F_X(x)$ . On en déduit immédiatement  $F(2^n x) = F(x)$  quelque soit  $x \in \mathbb{R}$ . Mais  $F_X$  admet 1 comme limite en  $+\infty$  et 0 en  $-\infty$  car c'est une fonction de répartition. Par conséquent :

$$\forall x > 0, F_X(x) = \lim_{n \rightarrow \infty} F_X(2^n x) = 1,$$

$$\forall x < 0, F_X(x) = \lim_{n \rightarrow \infty} F_X(2^n x) = 0.$$

Par continuité à droite de  $F_X$ , on en déduit que  $F_X(0) = 1$ . Ainsi  $F_X$  est la fonction de répartition de la loi de Dirac en 0 et  $X \stackrel{\mathcal{L}}{=} 0$ . On peut avoir une intuition de ce résultat en

supposant que  $X$  admet un moment d'ordre 2 et en remarquant que l'hypothèse implique  $\mathbb{E}[X] = \mathbb{V}(X) = 0$ .

### Exercice 4.13 ▷

1. La loi de Cauchy n'admet pas d'espérance car la fonction  $x \mapsto x/(1+x^2)$  n'est pas intégrable sur  $[1, +\infty[$ . En effet :

$$\frac{|x|}{1+x^2} \underset{x \rightarrow \infty}{\sim} \frac{1}{|x|}.$$

Il est donc faux d'utiliser un argument de parité pour conclure que cette espérance vaudrait 0 : cela ne fonctionne que si on est assuré de la convergence de l'intégrale auparavant. C'est par exemple le cas dans la question suivante.

2. Plus généralement  $X \hookrightarrow \mathcal{N}(0, 1) \Rightarrow \mathbb{E}[X^{2k+1}] = 0$  quelque soit  $k \in \mathbb{N}$  car  $\mathbb{E}[|X^{2k+1}|]$  est finie (grâce aux comparaisons habituelles entre fonction polynômes et exponentielles), et la fonction  $x \mapsto x^{2k+1}e^{-x^2/2}$  est impaire et intégrée sur un intervalle centré en 0. Et les moments d'ordre pair ? Ils peuvent être calculés sous une forme explicite eux aussi, mais cela demande un peu plus de travail. Posons :

$$\forall n \in \mathbb{N}, u_n := \mathbb{E}[X^{2n}].$$

On sait déjà que  $u_1 = 1$  (c'est le second paramètre de la loi normale centrée réduite  $\mathcal{N}(0, 1)$  !). Comme d'habitude avec ce type de suite définie par une intégrale, on essaye d'obtenir une relation de récurrence *via* une intégration par parties, ce qu'on peut faire directement. Afin de se la ramener un peu cependant, on va utiliser un résultat plus général sur la loi normale<sup>8</sup> :

$$X \hookrightarrow \mathcal{N}(0, 1) \Leftrightarrow \forall f \in \mathcal{C}_b^1(\mathbb{R}), \mathbb{E}[f'(X)] = \mathbb{E}[Xf(X)].$$

Ce résultat est connu sous le nom de *lemme de Stein* et a des applications pour le calcul de vitesse de convergence de certains théorèmes limites sur les suites de variables aléatoires (*via* la *méthode de Stein*). Nous n'aurons besoin que du sens gauche-droite de cette équivalence, qui s'avère être le plus facile à prouver (le second nécessite d'avoir vu les fonctions caractéristiques et sera à votre portée à ce moment-là) :

$$\begin{aligned} \mathbb{E}[f'(X)] &= \int_{\mathbb{R}} f'(x) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \\ &= \left[ f(x) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \right]_{-\infty}^{+\infty} + \int_{\mathbb{R}} xf(x) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \\ &= 0 + \int_{\mathbb{R}} xf(x) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \\ &= \mathbb{E}[Xf(X)]. \end{aligned}$$

On remarque que ce résultat reste valable pour les fonctions  $f$  de classe  $\mathcal{C}^1$ , telles que  $\mathbb{E}[f'(X)]$  et  $\mathbb{E}[Xf(X)]$  existent. On peut donc l'appliquer à la fonction (non bornée !)  $x \mapsto x^{2n+1}$  pour avoir :

---

8. Dans ce qui suit  $\mathcal{C}_b^1(\mathbb{R})$  est l'ensemble des fonctions de classe  $\mathcal{C}^1$  et **bornées** sur  $\mathbb{R}$ .

$$u_n = \mathbb{E}[X^{2n}] = \frac{1}{2n+1} \mathbb{E}[X^{2n+2}] = \frac{1}{2n+1} u_{n+1},$$

c'est-à-dire  $u_{n+1} = (2n+1)u_n$ . On reconnaît alors le produit des  $n-1$  premiers nombres impairs :

$$\begin{aligned} u_n &= \frac{u_n}{u_1} \\ &= \prod_{k=1}^{n-1} \frac{u_{k+1}}{u_k} \\ &= \prod_{k=1}^{n-1} (2k+1) \\ &= \frac{\left(\prod_{k=1}^{n-1} (2k+1)\right) \left(\prod_{k=1}^n 2k\right)}{\prod_{k=1}^n 2k} \\ &= \frac{\prod_{k=1}^{2n} k}{\prod_{k=1}^n 2k} \\ &= \frac{(2n)!}{2^n n!}. \end{aligned}$$

3. Un calcul d'intégrale immédiat donne :

$$X \hookrightarrow \mathcal{U}([a, b]) \Rightarrow \mathbb{E}[X] = \frac{a+b}{2}.$$

En d'autres termes, la "moyenne" d'une loi uniforme sur  $[a, b]$  est la moyenne de  $a$  et de  $b$ .

#### Exercice 4.14 ▷

1. Le charcutier ne peut pas vendre plus de choucroute qu'il n'en a fabriqué, donc le profit n'est **pas**  $P = pX - cx$  ! On trouve la bonne formule en supposant qu'il vend exactement ce qu'il a produit puis en corrigeant selon la demande qu'il a reçue durant la journée :

$$P = (p-c)x - p(x-X)_+,$$

en notant  $x_+ = \max(0, x) = 0 \vee x$  la partie positive de  $x$ . En effet si le nombre de clients est supérieur à la quantité de choucroute prévue  $x$ , le charcutier vend tout son stock de la journée et réalise un chiffre d'affaire de  $(p-c)x$  (on retire au prix le coût de production pour donner le bénéfice net). Si en revanche la demande est insuffisante pour écouler la quantité de choucroute préparée, alors on corrige le profit maximal possible  $((p-c)x)$  avec la différence invendue, d'où le terme  $-p(x-X)_+$ . On remarque que si la demande est plus forte que prévue *i.e.*  $x \leq X$ , alors on retrouve  $P = pX - cx$ .

2. Par linéarité de l'espérance, les quantités en jeu étant supposées exister (il suffit que  $X$  soit intégrable pour  $(X-x)_+$  le soit aussi par exemple), on a :

$$\mathbb{E}[P] = (p-c)x - p\mathbb{E}[(X-x)_+].$$

On voit aisément que presque sûrement on a :

$$\begin{aligned}
\int_0^x \mathbf{1}_{\{X \leq t\}} dt &= \left[ \int_0^x \mathbf{1}_{\{X \leq t\}} dt \right] \mathbf{1}_{\{X \leq x\}} \\
&= \left[ \int_0^x \mathbf{1}_{\{X \leq t\}} dt \right] \mathbf{1}_{\{X \leq x\}} \\
&= \left[ \int_{t \geq X} 1 dt \right] \mathbf{1}_{\{X \leq x\}} \\
&= (x - X) \mathbf{1}_{\{X \leq x\}} \\
&= (x - X)_+.
\end{aligned}$$

Par Fubini-Tonelli (la fonction  $(\omega, t) \mapsto \mathbf{1}_{\{X(\omega) \leq t\}}$  est mesurable et positive presque sûrement), on peut alors écrire<sup>9</sup> :

$$\begin{aligned}
\mathbb{E} \left[ \int_0^x \mathbf{1}_{\{X \leq t\}} dt \right] &= \int_0^x \mathbb{E}[\mathbf{1}_{\{X \leq t\}}] dt \\
&= \int_0^x \mathbb{P}(X \leq t) dt \\
&= \int_0^x F_X(t) dt.
\end{aligned}$$

En recollant les morceaux, on trouve :

$$\mathbb{E}[P] = (p - c)x - p \int_0^x F_X(t) dt.$$

Donc  $\alpha = p - c$  et  $\beta = p$ .

- On cherche les points critiques de  $\varphi : x \mapsto \varphi(x) = \mathbb{E}[P]$  sur  $\mathbb{R}_+$  (la quantité de choucroute produite  $x$  est positive!). Cette fonction est dérivable car  $F_X$  est supposée continue, et donc son intégrale de 0 à  $x$  est  $\mathcal{C}^1$ . On étudie le sens de variation de  $\varphi$  :

$$\varphi'(x) \geq 0 \Leftrightarrow (p - c) - pF_X(x) \geq 0,$$

ou encore  $x \geq F^{-1}(\frac{p-c}{p})$  en supposant que l'équation  $F(x) = (p - c)/p$  admet une unique solution). Remarquons que  $(p - c)/p \in ]0, 1[$  puisque  $c \in ]0, p[$ , donc sans hypothèse autre que le continuité de  $F_X$ , on a toujours une solution au moins à cette équation. Finalement la quantité à produire qui maximise le chiffre d'affaire moyen est  $x = F^{-1}(\frac{p-c}{p})$ .

- Il s'agit de trouver  $a$  tel que :

$$\mathbb{P}(P \geq a) = \mathbb{P}((x - X)_+ \leq (p - c)x - a) = 0.95.$$

Dit autrement, il s'agit d'exprimer  $F_{(x-X)_+}$  en fonction de  $F_X$  (et de  $c$  et de  $p$ ). Par un conditionnement qui doit être évident maintenant, pour  $t \geq 0$  :

---

9. Rappel : pour un événement  $A$  quelconque,  $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$ .

$$\begin{aligned}
 F_{(x-X)_+}(t) &= \mathbb{P}((x-X)_+ \leq t) \\
 &= \mathbb{P}((x-X)_+ \leq t, X \leq x) + \mathbb{P}((x-X)_+ \leq t, X > x) \\
 &= \mathbb{P}(x-X \leq t, X \leq x) + \mathbb{P}(0 \leq t, X > x) \\
 &= \mathbb{P}(x-t \leq X \leq x) + \mathbb{P}(X > x) \\
 &= F_X(x) - F_X(x-t) + (1 - F_X(x)) \\
 &= 1 - F_X(x-t).
 \end{aligned}$$

On remarque que si  $x \leq t$ , alors cette dernière quantité vaut 1 ( $X$  est à valeurs positives). C'est cohérent : dans ce cas  $(x-X)_+ \leq (t-X)_+ \leq t$ . Finalement :

$$\mathbb{P}(P \geq a) = 1 - F_X((p-c)x - a - t),$$

et la valeur  $a$  recherchée est  $a = (p-c)x - t - F_X^{-1}(0.05)$ . Cela s'interprète comme le profit minimal réalisé avec probabilité 0.95 **une fois que la quantité de choucroute à produire  $x$  est déterminée.**

5. Par ce qui précède, on a

$$x = -\frac{1}{\lambda} \log\left(\frac{p-c}{p}\right),$$

et :

$$a = -\frac{p-c}{\lambda} \log\left(\frac{p-c}{p}\right) - t + \frac{1}{\lambda} \log(0.95).$$

**Exercice 4.18** ▷ on a

$$\begin{aligned}
 \mathbb{E}(Z) &= \mathbb{E}(YB) + \mathbb{E}(X(1-B)) \\
 &= \mathbb{E}(Y)\mathbb{E}(B) + \mathbb{E}(X)(1 - \mathbb{E}(B)) \text{ par indépendance}
 \end{aligned}$$

or on sait que pour une v.a  $X$  qui suit une loi exponentielle de paramètre  $\lambda$ ,  $\mathbb{E}(X) = \frac{1}{\lambda}$  donc

$$\begin{aligned}
 \mathbb{E}(Z) &= \frac{2}{\lambda} \frac{3}{4} + \frac{1}{\lambda} \left(1 - \frac{3}{4}\right) \\
 &= \frac{7}{4\lambda} \geq \frac{1}{\lambda} = \mathbb{E}(X)
 \end{aligned}$$

Donc le traitement allonge bien l'esperance de vie des patients

**Exercice 4.19** ▷

1. Par definition d'une probabilité, on sait que :

$$1 = \mathbb{P}((X, Y) \in \mathbb{R}^2) = \iint \alpha 1_C(x, y) dx dy = \alpha \lambda_2(C) = \alpha \pi$$

$$\text{d'ou } \alpha = \frac{1}{\pi}$$

2. On a

$$f_X(x) = \int f_{(X,Y)}(x,y)dy = \int \frac{1}{\pi} \mathbf{1}_C(x,y)dy \stackrel{*}{=} \frac{1}{\pi} \int \mathbf{1}_{[-\sqrt{1-x^2}, \sqrt{1-x^2}]}(y) \mathbf{1}_{[-1,1]}(x)dy$$

enfin on obtient  $f_X(x) = \frac{2}{\pi} \sqrt{1-x^2} \mathbf{1}_{[-1,1]}(x)$

Montrons \*, en effet

$$(x,y) \in C \Leftrightarrow x^2 + y^2 \leq 1 \Leftrightarrow |y| \leq \sqrt{1-x^2} \Leftrightarrow y \in [-\sqrt{1-x^2}, \sqrt{1-x^2}], x \in [-1, 1]$$

$x$  et  $y$  jouent un rôle symétrique, on en déduit que  $f_Y(y) = \frac{2}{\pi} \sqrt{1-y^2} \mathbf{1}_{[-1,1]}(y)$

3. On remarque que  $f_{(X,Y)}(x,y) \neq f_X(x)f_Y(y)$ , donc  $X$  et  $Y$  ne sont pas indépendantes.

**Exercice 4.21** ▷

1.

$$\text{On a } T = \begin{cases} 0 & \text{si } U \in [0, v] \\ r + v - U & \text{sinon} \end{cases}$$

2.  $U$  suit une loi uniforme donc

$$\mathbb{P}(T = 0) = \mathbb{P}(U \in [0, v]) = \int_0^v \frac{\mathbf{1}_{[0,r+v]}(x)}{r+v} dx = \frac{v}{r+v}$$

3. on écrit astucieusement

$$h(T) = h(T) \mathbf{1}_{U \in [0,v]} + h(T) \mathbf{1}_{U \in [v,v+r]}$$

ainsi on obtient

$$\mathbb{E}(h(T)) = \mathbb{E}(h(0) \mathbf{1}_{U \in [0,v]}) + \mathbb{E}(h(r+v-U) \mathbf{1}_{U \in [v,v+r]}) \text{ d'après 1.}$$

$$= h(0) \mathbb{P}(U \in [0, v]) + \int h(r+v-u) \mathbf{1}_{u \in [v,v+r]} f_U(u)$$

$$= h(0) \frac{v}{v+r} + \frac{1}{v+r} \int_0^{v+r} h(r+v-u) du$$

$$= h(0) \frac{v}{v+r} + \frac{1}{v+r} \int_0^r h(u) du \text{ par changement de variable}$$

4.  $P_U = \frac{v}{v+r} \delta_0 + \frac{1}{v+r} \lambda_{[0,v]}$

**Exercice 4.23** ▷

1. on a  $T = 2\pi - |V - U|$

2. Calculons  $\mathbb{E}(|U - V|)$

Grâce à une astuce qu'on retrouve souvent en probabilité, on écrit

$$\begin{aligned} \mathbb{E}(|U - V|) &= \mathbb{E}(\mathbf{1}_{U \geq V} \cdot (U - V)) + \mathbb{E}(\mathbf{1}_{U < V} (V - U)) \\ &= 2\mathbb{E}(\mathbf{1}_{U > V} (U - V)) \text{ par symétrie de } U \text{ et } V \\ &= 2\mathbb{E}(\mathbf{1}_{\Delta}(U, V) \cdot (U - V)) \text{ avec } \Delta = (u, v), u > v \\ &= \frac{2}{(2\pi)^2} \iint \mathbf{1}_{\Delta}(u, v) \cdot (u - v) \mathbf{1}_{[0,2\pi]}(u) \mathbf{1}_{[0,2\pi]}(v) dudv \\ &= \frac{1}{2\pi^2} \int_0^{2\pi} \int_0^u (u - v) dv du \\ &= \frac{2\pi}{3} \end{aligned}$$

puis enfin

$$\mathbb{E}(T) = 2\pi - \mathbb{E}(|U - V|) = \frac{4\pi}{3}$$

ce resultat est *a priori* surprenant dans la mesure où la réponse à laquelle on s'attend est  $\pi$

3. coming soon

**Exercice 4.24** ▷

1. Par indépendance des variables à densité  $U$  et  $V$ ,  $(U, V)$  est un vecteur aléatoire à densité, de densité :

$$f_{(U,V)}(u, v) = (f_U \otimes f_V)(u, v) = f_U(u)f_V(v) = \mathbf{1}_{[0,1]}(u)\mathbf{1}_{[0,1]}(v) = \mathbf{1}_{[0,1]^2}((u, v)).$$

La notation tenseur utilisée à la première égalité n'est pas au programme. Plus généralement, si  $f$  et  $g$  sont deux fonctions de  $\mathbb{R}$  (par exemple) dans  $\mathbb{R}$  (ça c'est important), alors on définit :

$$(f \otimes g)(x, y) := f(x)g(y).$$

L'analogie avec les mesures produits est évidente !

2. Comme dans l'exercice précédent, on passe par les fonctions de queue plutôt que par la méthode de la fonction muette. Le support de  $U = \min(X, Y)$  est toujours  $[0, 1]$ . On trouve que :

$$\forall t \in [0, 1], \mathbb{P}(\min(X, Y) \leq t) = 1 - \mathbb{P}(\min(X, Y) \geq t) = 1 - \mathbb{P}(X \leq t)\mathbb{P}(Y \leq t) = 1 - t^2,$$

et  $F_U$  est nulle sur  $]-\infty, 0[$ , égale à 1 sur  $[1, +\infty[$ .

3. Remarquons déjà que le support de  $(X, Y)$  est :

$$(X, Y)(\Omega) = E := \{(x, y) \in [0, 1] \times [0, 1], x \leq y\}.$$

Plutôt que la fonction de répartition de  $(X, Y)$ , on va d'abord calculer  $\forall (x, y) \in E$  :

$$\begin{aligned} \mathbb{P}(X \geq x, Y \leq y) &= \mathbb{P}(x \leq U \leq y, x \leq V \leq y) \\ &= \mathbb{P}(x \leq U \leq y)\mathbb{P}(x \leq V \leq y) \\ &= (y - x)^2 \end{aligned}$$

Or, puisque  $\mathbb{P}(X \leq x, Y \leq y) = \mathbb{E}[\mathbf{1}_{\{X \leq x\}}\mathbf{1}_{\{Y \leq y\}}] = \mathbb{E}[(1 - \mathbf{1}_{\{X > x\}})\mathbf{1}_{\{Y \leq y\}}] = \mathbb{P}(Y \leq y) - \mathbb{P}(X > x, Y \leq y)$ , on a :

$$F_{(X,Y)}(x, y) = \mathbb{P}(Y \leq y) - \mathbb{P}(X > x, Y \leq y) = y^2 - (y - x)^2.$$

Cette fonction est nulle dès que l'une des deux variables est plus petite que 0 ou que  $x > y$ , et égale à 1 quand **les deux** variables sont plus grandes que 1 **et** vérifient  $x \leq y$ .

4. Par le théorème de Schwarz, puisque  $F_{(X,Y)}$  de classe  $\mathcal{C}^2$  sur  $\mathbb{R}^2$ , on peut calculer une densité de  $f_{(X,Y)}$  en dérivant une fois par rapport à  $x$ , et une fois par rapport à  $y$ , dans n'importe quel ordre. On trouve :

$$f_{(X,Y)}(x, y) = 2\mathbf{1}_E((x, y)).$$

Au passage on en déduit sans plus de calculs que l'aire de  $E$  est  $1/2$ . Pourquoi ?

---

**Exercice 4.25** ▷

Vue la formulation de la question, il est plus appropriée ici d'utiliser la méthode de la fonction de répartition. Le support de  $X/Y$  est toujours  $\mathbb{R}_+^*$ , donc on se restreint aux points  $t > 0$ . Par la théorème de transfert :

$$\mathbb{P}(X/Y \leq t) = \mathbb{E}[\mathbf{1}_{X/Y \leq t}] = \int_{\mathbb{R}^2} \mathbf{1}_{\{x/y \leq t\}} d\mathbb{P}_{(X,Y)}(x, y).$$

Or  $X$  et  $Y$  sont indépendantes, donc leur loi de couple vérifie :

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y.$$

De plus toutes les quantités en jeu sont positives et les lois de probabilité de  $X$  et de  $Y$  sont finies (car  $\mathbb{P}_X(\Omega) = \mathbb{P}_Y(\Omega) = 1$ !), donc  $\sigma$ -finies, et le théorème de Fubini-Tonelli permet d'intégrer dans n'importe quel ordre (par rapport à  $x$  en premier, puis  $y$ , ou l'inverse) :

$$\begin{aligned} \mathbb{P}(X/Y \leq t) &= \int_{\mathbb{R}^2} \mathbf{1}_{\{x/y \leq t\}} d(\mathbb{P}_X \otimes \mathbb{P}_Y)(x, y) \\ &= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} \mathbf{1}_{\{x/y \leq t\}} d\mathbb{P}_Y(y) \right) d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} \mathbf{1}_{\{y \geq x/t\}} d\mathbb{P}_Y(y) \right) d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}} \mathbb{P}(Y \geq x/t) d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}} e^{-x/t} d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}} e^{-x/t} e^{-x} dx \\ &= \frac{1}{1 + \frac{1}{t}} \\ &= \frac{t}{t + 1} \\ &= 1 - \frac{1}{t + 1}. \end{aligned}$$

En dérivant, on obtient :

$$f_{X/Y}(t) = \frac{1}{(t + 1)^2}$$

REMARQUE.— — Cette loi appartient à la famille des lois de Pareto, très utilisés pour modéliser les inégalités de distribution des richesses ("loi du 80-20", ou principe de Pareto : 20% des individus détiennent 80% des richesses). On remarque que sa fonction de queue

$(\bar{F}_{X/Y} := 1 - F_{X/Y})$  décroît vers 0 en l'infini à vitesse polynomiale, bien plus lentement que les fonctions de queue de la loi exponentielle ou de la loi normale par exemple : la loi de Pareto est **à queue lourde**, *i.e.* elle peut prendre des valeurs très élevées avec une probabilité non négligeable<sup>10</sup>.

- On présente une autre méthode de calcul bien plus simple et reposant sur les espérances conditionnelles (que vous verrez sous peu) :

$$\begin{aligned}\mathbb{P}(X/Y \leq t) &= \mathbb{E}[\mathbb{P}(X/Y \leq t \mid X)] \\ &= \mathbb{E}[\mathbb{P}(Y \geq X/t \mid X)] \\ &= \mathbb{E}[e^{-X/t}] \\ &= 1 - \frac{1}{t+1},\end{aligned}$$

en reconnaissant la transformée de Laplace d'une loi exponentielle en  $1/t$ .

- Les techniques de calcul utilisées dans cet exercice sont réemployées dans l'exercice sur la méthode du rejet. Elles se simplifieront grandement une fois que vous aurez vu l'espérance conditionnelle.

### Exercice 4.26 ▷

1. Par croissance de l'intégrale, l'hypothèse de l'énoncé implique :

$$1 = \int_{\mathbb{R}} f(x) dx \leq c \int_{\mathbb{R}} g(x) dx = c,$$

puisque  $f, g$  sont des densités de probabilité, donc intègrent à 1.

2. On rappelle l'identité fondamentale, valable pour tout événement  $A$  :

$$\mathbb{P}(A) = \mathbb{E}[\mathbf{1}_A].$$

On peut donc écrire, *via* le théorème de transfert par rapport à la mesure de probabilité  $\mathbb{P}_{(U,Y)} = \mathbb{P}_U \otimes \mathbb{P}_Y$  ( $U$  et  $Y$  sont indépendantes) :

---

10. Une définition rigoureuse de cette notion existe.

$$\begin{aligned}
\mathbb{P}\left(U \leq \frac{f(Y)}{cg(Y)}\right) &= \mathbb{E}\left[\mathbf{1}_{\left\{U \leq \frac{f(Y)}{cg(Y)}\right\}}\right] \\
&= \int_{\mathbb{R} \times [0,1]} \mathbf{1}_{\left\{u \leq \frac{f(y)}{cg(y)}\right\}} d(\mathbb{P}_U \otimes \mathbb{P}_Y)(u, y) \\
&= \int_{\mathbb{R}} \int_{[0,1]} \mathbf{1}_{\left\{u \leq \frac{f(y)}{cg(y)}\right\}} d\mathbb{P}_U(u) d\mathbb{P}_Y(y) \\
&= \int_{\mathbb{R}} \mathbb{P}\left(U \leq \frac{f(y)}{cg(y)}\right) d\mathbb{P}_Y(y) \\
&= \int_{\mathbb{R}} \mathbb{P}\left(U \leq \frac{f(y)}{cg(y)}\right) g(y) dy \\
&= \int_{\mathbb{R}} \frac{f(y)}{c} dy \\
&= \frac{1}{c},
\end{aligned}$$

car la fonction de la répartition de la loi uniforme est la fonction  $x \mapsto x$  sur  $[0, 1]$  et  $f(y)/(cg(y)) \in [0, 1]$  par hypothèse. Comme  $c \geq 1$ , son inverse est bien compris entre 0 et 1, ce qui est le minimum syndical pour une probabilité !

REMARQUE.— Ce type de calcul revient à manipuler des espérances conditionnelles sans le dire (au lieu de calculer une espérance par rapport à deux variables aléatoires en même temps, on calcule deux espérances successivement, pour la première en supposant que l'autre variable aléatoire est connue. Avec des notations que vous verrez prochainement, on aurait pu écrire :

$$\int_{\mathbb{R}} \left( \int_{[0,1]} \mathbf{1}_{\left\{u \leq \frac{f(y)}{cg(y)}\right\}} d\mathbb{P}_U(u) \right) d\mathbb{P}_Y(y) = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\left\{U \leq \frac{f(Y)}{cg(Y)}\right\}} \mid Y\right]\right]$$

Les espérances conditionnelles rendent beaucoup de calculs d'espérances/probabilités bien plus simples et conceptuels, au lieu d'obscurcir les idées avec des intégrales dans tous les sens.

3. Le même type de raisonnement qu'à la question précédente donne :

$$\begin{aligned}
\mathbb{P}\left(U \leq \frac{f(Y)}{cg(Y)}, Y \leq t\right) &= \int_{\mathbb{R}} \int_{[0,1]} \mathbf{1}_{\left\{u \leq \frac{f(y)}{cg(y)}\right\}} \mathbf{1}_{\{y \leq t\}} d\mathbb{P}_U(u) d\mathbb{P}_Y(y) \\
&= \int_{-\infty}^t \int_{[0,1]} \mathbf{1}_{\left\{u \leq \frac{f(y)}{cg(y)}\right\}} d\mathbb{P}_U(u) d\mathbb{P}_Y(y) \\
&= \int_{-\infty}^t \mathbb{P}\left(U \leq \frac{f(y)}{cg(y)}\right) g(y) dy \\
&= \int_{-\infty}^t \frac{f(y)}{c} dy \\
&= \frac{1}{c} \mathbb{P}(Y \leq t) \\
&= \mathbb{P}\left(U \leq \frac{f(Y)}{cg(Y)}\right) \mathbb{P}(Y \leq t).
\end{aligned}$$

4. Les  $B_k$  sont des variables aléatoires indépendantes (car les  $U_k$  et les  $Y_k$  le sont) de même loi de Bernoulli  $\mathcal{B}(1/c)$  d'après la question 1. (on rappelle que le paramètre d'une loi de Bernoulli est la probabilité qu'elle prenne la valeur 1).
5. Déjà le support de  $T$  est  $T(\Omega) = \mathbb{N}^*$ . En effet  $T$  est un infimum sur un ensemble d'indices inclus dans  $\mathbb{N}^* \cup \{\infty\}$ . De plus  $T < \infty$  p.s. car  $1/c \neq 0$  et les événements  $B_k$  sont indépendants, tous de probabilité de réussite  $1/c$ . Enfin un calcul classique montre que :

$$\forall k \in \mathbb{N}^*, \mathbb{P}(T = k) = \frac{1}{c} \left(1 - \frac{1}{c}\right)^{k-1},$$

c'est-à-dire que  $T \hookrightarrow \mathcal{G}(1/c)$ .

6. La logique est toujours la même : dès que l'on manipule des variables aléatoires avec deux "couches d'aléa" ou plus, on conditionne par l'un d'entre elles et on se sert des informations obtenues pour effectuer une partie des calculs. Ensuite on termine en sommant le résultat sur toutes les possibilités associées aux premier aléa (on "déconditionne"). Ici le plus naturel est de conditionner par rapport à  $T$ , qui est une variable discrète. On peut donc utiliser la formule des probabilités totales et calculer la fonction de répartition de  $Y_T$ . Soit  $t \in \mathbb{R}$ , on a :

$$\begin{aligned} \mathbb{E}[\mathbf{1}_{\{Y_T \leq t\}}] &= \sum_{k=1}^{\infty} \mathbb{E}[\mathbf{1}_{\{Y_T \leq t\}} \mathbf{1}_{\{T=k\}}] \\ &= \sum_{k=1}^{\infty} \mathbb{E}[\mathbf{1}_{\{Y_k \leq t\}} \mathbf{1}_{\{T=k\}}] \\ &= \sum_{k=1}^{\infty} \mathbb{E}[\mathbf{1}_{\{Y_k \leq t\}} \mathbf{1}_{\{\bigcap_{i=1}^{k-1} \{B_i=0\} \cap \{B_k=1\}\}}] \\ &= \sum_{k=1}^{\infty} \mathbb{P}\left(\bigcap_{i=1}^{k-1} \{B_i = 0\}\right) \mathbb{E}[\mathbf{1}_{\{Y_k \leq t\}} \mathbf{1}_{\{B_k=1\}}] \\ &= \sum_{k=1}^{\infty} \mathbb{P}(T \geq k) \mathbb{E}[\mathbf{1}_{\{Y_k \leq t\}} \mathbf{1}_{\{B_k=1\}}] \\ &= \sum_{k=1}^{\infty} \left(1 - \frac{1}{c}\right)^{k-1} \mathbb{E}[\mathbf{1}_{\{Y_k \leq t\}} \mathbf{1}_{\{B_k=1\}}], \end{aligned}$$

en effet  $Y_k$  ne dépend de  $\{T = k\}$  qu'à travers  $\{B_k = 1\}$ . Mais on sait par la question 3. que  $\mathbb{E}[\mathbf{1}_{\{Y_k \leq t\}} \mathbf{1}_{\{B_k=1\}}] = \mathbb{P}(B_k = 1) \mathbb{P}(X \leq t)$ . D'où finalement :

$$\begin{aligned} \mathbb{P}(Y_T \leq t) &= \sum_{k=1}^{\infty} \left(1 - \frac{1}{c}\right)^{k-1} \frac{1}{c} \mathbb{P}(X \leq t) \\ &= \mathbb{P}(X \leq t) \sum_{k=1}^{\infty} \left(1 - \frac{1}{c}\right)^{k-1} \frac{1}{c} \\ &= \mathbb{P}(X \leq t). \end{aligned}$$

Et donc  $X' = Y_T$  admet même densité  $f$ , donc même loi, que  $X$  car elles ont la même fonction de répartition.

7. On propose l'algorithme suivant :

Input: Une densité cible  $f$ , une densité proxy  $g$  selon laquelle on sait simuler  
 Output: Une variable aléatoire  $X$  de densité  $f$

```

k = 1
Simuler  $Y_k$  selon  $g$  et  $U_k$  de loi uniforme sur  $[0,1]$ 

Tant que  $U_k > f(Y_k)/(cg(Y_k))$ , faire:

    k = k+1
    Simuler  $Y_k$  selon  $g$  et  $U_k$  de loi uniforme sur  $[0,1]$ 

Retourner  $Y_k$ 
  
```

8. L'auteur concède sa nullité en matière de calcul de complexité, même pour les plus élémentaires :)

REMARQUE.— Cette méthode permet de simuler de nombreuses lois de probabilités sans forcément connaître leur fonction de répartition, ou même leur densité : il suffit de pouvoir la majorer par une constante  $c$  fois la densité d'une loi qu'on sait simuler. Ici réside la principale difficulté dans l'application de cette méthode, à savoir trouver une densité  $g$  accessible à la simulation et telle que  $c$  soit la plus proche possible de 1. Sinon  $1/c$  devient petit et  $T$  peut prendre des valeurs très grandes. En d'autres termes, on va devoir simuler beaucoup de variables aléatoires  $U_i$  et  $Y_i$  afin de trouver un  $Y_k$  satisfaisant  $U_k \leq f(Y_k)/(cg(Y_k))$ . Cela constitue un gaspillage de ressources que certaines versions de la méthode du rejet essaient d'ailleurs de pallier (voir le livre de Luc Devroye cité plus bas).

**Exercice 4.29** ▷

1. Soit  $A \in M_n(\mathbb{C})$ ,  $b \in M_{n,1}(\mathbb{K})$  on a :

$$\mathbb{E}(AX + b) = \begin{pmatrix} \sum_{k=0}^n \mathbb{E}(a_{1,k}X_k + b_1) \\ \dots \\ \sum_{k=0}^n \mathbb{E}(a_{n,k}X_k + b_n) \end{pmatrix} = \begin{pmatrix} \sum_{k=0}^n a_{1,k}\mathbb{E}(X_k) + \mathbb{E}(b_1) \\ \dots \\ \sum_{k=0}^n a_{n,k}\mathbb{E}(X_k) + \mathbb{E}(b_n) \end{pmatrix}$$

Soit :

$$\mathbb{E}(AX + b) = \begin{pmatrix} \sum_{k=0}^n a_{1,k}\mathbb{E}(X_k) \\ \dots \\ \sum_{k=0}^n a_{n,k}\mathbb{E}(X_k) \end{pmatrix} + \begin{pmatrix} \mathbb{E}(b_1) \\ \dots \\ \mathbb{E}(b_n) \end{pmatrix} = A\mathbb{E}(X) + b$$

2. On a  $\forall i \in \{1 \dots n\} (Cov(X))_{i,i} = Cov(X_i, X_i) = V(X_i)$

3. Si  $\forall (i, j) \in \mathbb{N}^2 X_i$  et  $X_j$  sont indépendantes alors  $\forall i, j$  tel que  $i \neq j$  on a

$$Cov(X_i, X_j) = 0 \text{ donc } Cov(X) \text{ est diagonale}$$

4. Calculons  $\mathbb{E}(X_c X_c^T)$  :

$$\mathbb{E}(X_c X_c^T) = \mathbb{E} \begin{pmatrix} X_1 - \mathbb{E}(X_1) \\ \dots \\ X_d - \mathbb{E}(X_d) \end{pmatrix} (X_1 - \mathbb{E}(X_1) \dots X_d - \mathbb{E}(X_d))$$

Donc :

$$\mathbb{E}(X_c X_c^T) = \mathbb{E} \begin{pmatrix} (X_1 - \mathbb{E}(X_1))^2 & \dots & (X_1 - \mathbb{E}(X_1))(X_d - \mathbb{E}(X_d)) \\ (X_2 - \mathbb{E}(X_2))(X_1 - \mathbb{E}(X_1)) & \dots & (X_2 - \mathbb{E}(X_2))(X_d - \mathbb{E}(X_d)) \\ \dots & \dots & \dots \\ (X_d - \mathbb{E}(X_d))(X_1 - \mathbb{E}(X_1)) & \dots & (X_d - \mathbb{E}(X_d))^2 \end{pmatrix}$$

Ainsi

$$\mathbb{E}(X_c X_c^T) = (Cov(X_i, X_j))_{(i,j) \in \mathbb{N}^2} = Cov(X)$$

**Exercice 4.30** ▷

1. On applique la méthode de la variable muette. Soit  $\varphi$  une fonction-test. Par indépendance de  $X$  et de  $Y$ , leur loi jointe vérifie (par définition)  $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y$  :

$$\begin{aligned} \mathbb{E}[\varphi(Z)] &= \mathbb{E}[\varphi(X + Y)] \\ &= \int_{\mathbb{R} \times \mathbb{R}} \varphi(x + y) d(\mathbb{P}_X \otimes \mathbb{P}_Y)(x, y) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x + y) d\mathbb{P}_X(x) d\mathbb{P}_Y(y) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x + y) f_X(x) f_Y(y) dx dy \\ &= \int_{\mathbb{R}} f_Y(y) \left( \int_{\mathbb{R}} \varphi(x + y) f_X(x) dx \right) dy \\ &= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} \varphi(z) f_X(y - z) f_Y(y) dz \right) dy \\ &= \int_{\mathbb{R}} \varphi(z) \left( \int_{\mathbb{R}} f_X(y - z) f_Y(y) dy \right) dz, \end{aligned}$$

en faisant le changement de variable  $z = x + y$  et en spammant le théorème de Fubini(-Tonelli ou -Lebesgue, selon les hypothèses sur  $\varphi$ ). On en déduit qu'une densité de  $Z = X + Y$  est :

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(z - y) f_Y(y) dy.$$

2. Non ! Si on prend  $X$  une variable à densité quelconque et  $Y = -X$ , on trouve  $X + Y = 0$ , qui n'est certainement pas à densité ! (la mesure de Dirac en 0 est une loi de probabilité discrète donc elle n'est pas absolument continue par rapport à la mesure de Lebesgue)
3. Le piège débile de la question est que  $\sigma_X$  et  $\sigma_Y$  peuvent être nuls, auquel cas  $X = Y = 0$  presque sûrement et  $Z$  n'est pas à densité. Si un des écart-types est nul, disons  $\sigma_Y$ , alors  $Z$  a même loi que  $X + 0 = X$ , et donc est admet  $f_X$  comme densité. Supposons maintenant que  $\sigma_X$  et  $\sigma_Y$  soient tous deux strictements positifs. Alors par la question 1.,  $Z$  est bien à

densité, comme somme de telles variables aléatoires (indépendantes). Il s'agit maintenant de calculer l'intégrale obtenue à la fin de la première question :

$$\begin{aligned}
 f_{X+Y}(z) &= \int_{\mathbb{R}} f_X(x) f_Y(z-x) dx \\
 &= \frac{1}{2\pi\sigma_X\sigma_Y} \int_{\mathbb{R}} e^{-\frac{x^2}{2\sigma_X^2}} e^{-\frac{(z-x)^2}{2\sigma_Y^2}} dx \\
 &= \frac{1}{2\pi\sigma_X\sigma_Y} \int_{\mathbb{R}} e^{-\frac{x^2}{2\sigma_X^2} - \frac{z^2+x^2-2zx}{2\sigma_Y^2}} dx \\
 &= \frac{1}{2\pi\sigma_X\sigma_Y} \int_{\mathbb{R}} e^{-\left(\frac{1}{2\sigma_X^2} + \frac{1}{2\sigma_Y^2}\right)x^2 + \frac{z}{\sigma_Y^2}x - \frac{z^2}{2\sigma_Y^2}} dx \\
 &= \frac{1}{2\pi\sigma_X\sigma_Y} \int_{\mathbb{R}} e^{-a\left(x + \frac{b}{2a}\right)^2 + \frac{\Delta}{4a}} dx \\
 &= \frac{1}{2\pi\sigma_X\sigma_Y} e^{\frac{\Delta}{4a}} \int_{\mathbb{R}} e^{-ax^2} dx \\
 &= \frac{1}{2\pi\sigma_X\sigma_Y} e^{\frac{\Delta}{4a}} \sqrt{\frac{\pi}{a}} \\
 &= \frac{1}{2\sigma_X\sigma_Y\sqrt{a\pi}} e^{\frac{\Delta}{4a}},
 \end{aligned}$$

avec bien sûr :

$$\begin{aligned}
 a &= \frac{1}{2\sigma_X^2} + \frac{1}{2\sigma_Y^2} = \frac{\sigma_X^2 + \sigma_Y^2}{2\sigma_X^2\sigma_Y^2}, \\
 b &= -\frac{z}{\sigma_Y^2} \text{ et } c = \frac{z^2}{2\sigma_Y^2}.
 \end{aligned}$$

Et donc :

$$\Delta = \frac{z^2}{\sigma_Y^4} - 4 \frac{z^2}{2\sigma_Y^2} \frac{\sigma_X^2 + \sigma_Y^2}{2\sigma_X^2\sigma_Y^2} = -\frac{z^2}{\sigma_X^2\sigma_Y^2},$$

d'où :

$$\frac{\Delta}{4a} = -\frac{z^2}{2(\sigma_X^2 + \sigma_Y^2)}.$$

De sorte que :

$$f_{X+Y}(z) = \frac{1}{\sqrt{2(\sigma_X^2 + \sigma_Y^2)}\pi} e^{-\frac{z^2}{2(\sigma_X^2 + \sigma_Y^2)}}.$$

Ainsi  $X + Y \hookrightarrow \mathcal{N}(0, \sigma_X^2 + \sigma_Y^2)$ .

REMARQUE.— — Ce résultat fondamental sur les lois normales, fruit d'un calcul un brin laborieux, sera considérablement généralisé dans le chapitre sur les vecteurs gaussiens, et la preuve sera bien plus rapide grâce aux fonctions caractéristiques. Ce

sera l'occasion de voir une nouvelle fois l'analyse de Fourier simplifier les calculs de convolution, cette fois d'un point de vue probabiliste.

- La convolution de deux variables aléatoires peut se définir dans un cadre plus général que celui de  $\mathbb{R}$  ou même des espaces vectoriels. Les groupes de Lie par exemple sont dotés de l'équivalent d'une mesure de Lebesgue (appelée **mesure de Haar**) qui est invariante par translation (à gauche ou à droite, pas toujours les deux à la fois!). La multiplication du groupe remplace alors l'addition. Un bon cadre pour ce type de problème est les groupes de Lie compact, où la mesure de Haar est normalisable et bi-invariante, et peut donc définir une probabilité uniforme. Par exemple on peut définir une loi uniforme sur  $\mathbf{O}(n)$ .

**Exercice 4.31** ▷

1. Comme  $X$  et  $Y$  sont indépendantes, et admettent chacune une densité, leur couple admet aussi une densité donné par le produit des densités individuelles :

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y) = \mathbf{1}_{[0,1]}(x)\mathbf{1}_{[0,1]}(y) = \mathbf{1}_{[0,1]^2}((x, y)).$$

2. Vue la tête du changement de loi, utiliser la fonction de répartition semble très mal indiqué ici. La méthode de la variable muette est un bien meilleur choix. Posons :

$$g : ]0, 1[ \times ]0, 1[ \rightarrow \mathbb{R}^2 \\ (x, y) \mapsto (\sqrt{-2 \log x} \cos(2\pi y), \sqrt{-2 \log x} \sin(2\pi y))$$

Il est important que  $g$  prenne ses valeurs dans l'ouvert  $]0, 1[ \times ]0, 1[$ , plutôt que dans le fermé  $[0, 1] \times [0, 1]$ , ne serait-ce que pour exclure les valeurs problématiques comme 0. Les ensembles de la forme  $\{a\} \times ]0, 1[$  étant de mesure de Lebesgue  $\lambda_2$  nulle, cela ne change rien au problème. Ainsi,  $g$  est bien définie sur son ensemble de définition, et donc en  $(X, Y)$  presque sûrement. Il faut maintenant prouver que  $g$  est un  $\mathcal{C}^1$ -difféomorphisme, et en fait, expliciter son inverse (dont on a besoin pour faire le calcul proprement dit). Il s'agit donc de résoudre en  $(x, y)$  :

$$\begin{cases} \sqrt{-2 \log x} \cos(2\pi y) = u \\ \sqrt{-2 \log x} \sin(2\pi y) = v \end{cases}$$

En passant aux carrés et en sommant, on trouve que  $-2 \log x = u^2 + v^2$ , et donc :

$$x = e^{-\frac{u^2+v^2}{2}}.$$

Et en quotientant la seconde ligne du système par la première, on trouve que  $\tan(2\pi y) = v/u$ , soit encore :

$$y = \arctan\left(\frac{v}{u}\right),$$

en passant les problèmes de division par zéro car les points problématiques forment un ensemble de mesure de Lebesgue nulle... Les lecteurs attentifs auront remarqué que ces calculs sont essentiellement ceux du passage en coordonnées polaires. Le Jacobien associé à  $g$  vaut :

$$J(u, v) = \begin{vmatrix} \frac{-\cos(2\pi y)}{x\sqrt{-2\log x}} & -2\pi\sqrt{-2\log x}\sin(2\pi y) \\ \frac{-\sin(2\pi y)}{x\sqrt{-2\log x}} & 2\pi\sqrt{-2\log x}\cos(2\pi y) \end{vmatrix} = -\frac{2\pi}{x} = -2\pi e^{\frac{u^2+v^2}{2}}.$$

Ainsi on trouve, pour  $\varphi : \mathbb{R}^2 \mapsto \mathbb{R}$  une fonction test (continue bornée par exemple), en n'oubliant de faire apparaître la valeur absolue du Jacobien (qui est l'inverse du déterminant calculé précédemment, exprimée dans les nouvelles coordonnées) :

$$\begin{aligned} \mathbb{E}[\varphi(U, V)] &= \int_{\mathbb{R}^2} \varphi\left(\sqrt{-2\log x}\cos(2\pi y), \sqrt{-2\log x}\sin(2\pi y)\right) f_{(X,Y)}(x, y) d\lambda_2(x, y) \\ &= \int_{]0,1[^2} \varphi\left(\sqrt{-2\log x}\cos(2\pi y), \sqrt{-2\log x}\sin(2\pi y)\right) dx dy \\ &= \int_{\mathbb{R}^2} \varphi(u, v) \frac{1}{2\pi} e^{-\frac{u^2+v^2}{2}} du dv \end{aligned}$$

Donc  $(U, V)$  admet comme densité la fonction  $(u, v) \mapsto \exp(-(u^2 + v^2)/2)$ , c'est-à-dire la densité d'un couple de variables aléatoires indépendantes suivant chacune la loi normale centrée réduite  $\mathcal{N}(0, 1)$ . En particulier  $U$  et  $V$  sont indépendantes et de même loi normale centrée réduite.

REMARQUE.— Tout l'intérêt de cet exercice, au-delà de vous entraîner à pratiquer les changements de variable en dimension supérieure, est de fournir un algorithme permettant de simuler deux variables aléatoires suivant la loi normale  $\mathcal{N}(0, 1)$ , indépendantes l'une de l'autre. Et ce uniquement en partant de deux uniformes indépendantes. La fonction de répartition de la loi normale n'étant pas explicite, il aurait été impossible de l'inverser pour appliquer la méthode d'inversion vue dans les exercices précédents, d'où l'intérêt de ce résultat. D'autres méthodes existent qui lui sont préférées, car ne faisant pas appel aux fonctions trigonométriques et au logarithme, qui sont coûteuses à estimer (voir la grande référence en la matière, à savoir le livre de Luc Devroye : *Non-Uniform Random Variate Generation*)

### Exercice 4.32 ▷

1. En prenant  $b = 1$ , on retrouve la densité d'une loi exponentielle  $\mathcal{E}(\lambda)$ . Ce lien entre loi exponentielle et loi Gamma est important dans beaucoup d'applications liées aux télécommunications (voir par exemple la formule (et la loi) d'Erlang).
2. Utilisons la méthode de la variable muette. Soit  $\varphi$  une fonction-test. On a :

$$\begin{aligned}
 \mathbb{E}[\varphi(Z^2)] &= \int_{\mathbb{R}} \varphi(z^2) f_Z(z) dz \\
 &= \int_{\mathbb{R}_+} \varphi(z^2) f_Z(z) dz + \int_{\mathbb{R}_-} \varphi(z^2) f_Z(z) dz \\
 &= \int_0^\infty \varphi(u) \frac{1}{2\sqrt{u}} f_Z(\sqrt{u}) du + \int_{+\infty}^0 \varphi(u) f_Z(-\sqrt{u}) \frac{1}{-2\sqrt{u}} du \\
 &= \int_{\mathbb{R}_+} \varphi(u) \frac{1}{2\sqrt{u}} f_Z(\sqrt{u}) du + \int_{\mathbb{R}_+} \varphi(u) f_Z(-\sqrt{u}) \frac{1}{2\sqrt{u}} du \\
 &= \int_{\mathbb{R}_+} \varphi(u) \frac{1}{\sqrt{u}} f_Z(\sqrt{u}) du \\
 &= \int_{\mathbb{R}_+} \varphi(u) \frac{1}{\sqrt{u}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u}{2}} du \\
 &= \int_{\mathbb{R}_+} \varphi(u) \frac{u^{\frac{1}{2}-1} e^{-\frac{u}{2}}}{\sqrt{2}\Gamma(1/2)} du
 \end{aligned}$$

La séparation en deux intégrales est nécessaire puisque  $x \mapsto x^2$  n'est pas un  $\mathcal{C}^1$ -difféomorphisme sur  $\mathbb{R}$  (elle n'est pas injective!). Elle l'est en revanche sur  $\mathbb{R}_+$  et  $\mathbb{R}_-$ . On retrouve une loi Gamma  $\Gamma(1/2, 1/2)$ , qui est un cas particulier de loi du Chi 2, notée <sup>11</sup>, plus précisément la loi du Chi 2 à un degré de liberté  $\chi^2(1)$ . Il s'agit là encore d'une loi très importante dans les applications, en particulier en statistiques.

3. La méthode de la variable muette est la plus indiquée ici. On pose :

$$\begin{aligned}
 g : \mathbb{R}_+^* \times \mathbb{R}_+^* &\rightarrow \mathbb{R}_+^* \times ]0, 1[ \\
 (x, y) &\mapsto \left(x + y, \frac{x}{x+y}\right)
 \end{aligned}$$

Normalement on devrait prendre comme ensemble d'arrivée le support de  $(X, Y)$ , c'est-à-dire  $\mathbb{R}_+ \times \mathbb{R}_+$ , mais pour faire des changements de variable, il vaut mieux travailler sur des ouverts. Ici on remarque que les ensembles  $\{0\} \times \mathbb{R}_+^*$  et  $\mathbb{R}_+^* \times \{0\}$  sont de mesure de Lebesgue  $\lambda_2$  nulle. Donc cela ne pose pas de problème de retirer l'origine dans les ensembles de départ. Prouvons que  $g$  est bien un  $\mathcal{C}^1$ -difféomorphisme. Soient  $(u, v) \in \mathbb{R}_+^* \times ]0, 1[$ , on a :

$$\begin{cases} x + y = u \\ \frac{x}{x+y} = v \end{cases} \iff \begin{cases} x + y = u \\ x = uv \end{cases} \iff \begin{cases} y = u(1 - v) \\ x = uv \end{cases}$$

Or la fonction

$$\begin{aligned}
 g^{-1} : \mathbb{R}_+^* \times ]0, 1[ &\rightarrow \mathbb{R}_+^* \times \mathbb{R}_+^* \\
 (u, v) &\mapsto (uv, u(1 - v))
 \end{aligned}$$

est de classe  $\mathcal{C}^1$  sur son ensemble de définition, de même que  $g$ . Donc  $g$  est bien un  $\mathcal{C}^1$ -difféomorphisme. Calculons son Jacobien :

$$J(u, v) = \begin{vmatrix} v & u \\ 1 - v & -u \end{vmatrix} = -u.$$

---

11. Prononcer "Ki deux".

Finalement en appliquant ce changement de variable, et en utilisant que  $X$  et  $Y$  sont indépendantes pour scinder leur densité jointe, on trouve :

$$\begin{aligned}
\mathbb{E}\left[\varphi\left(X + Y, \frac{X}{X + Y}\right)\right] &= \int_{\mathbb{R}_+^* \times \mathbb{R}_+^*} \varphi\left(x + y, \frac{x}{x + y}\right) f_{(X,Y)}(x, y) dx dy \\
&= \int_{\mathbb{R}_+^* \times \mathbb{R}_+^*} \varphi\left(x + y, \frac{x}{x + y}\right) f_X(x) f_Y(y) dx dy \\
&= \int_{\mathbb{R}_+^* \times ]0,1[} \varphi(u, v) u f_X(uv) f_Y(u(1 - v)) du dv \\
&= \int_{\mathbb{R}_+^* \times ]0,1[} \varphi(u, v) u \frac{\lambda^a (uv)^{a-1} e^{-\lambda uv}}{\Gamma(a)} \frac{\lambda^b (u(1 - v))^{b-1} e^{-\lambda u(1-v)}}{\Gamma(b)} du dv \\
&= \int_{\mathbb{R}_+^* \times ]0,1[} \varphi(u, v) \frac{v^{a-1} (1 - v)^{b-1}}{\Gamma(a)\Gamma(b)} \lambda^{a+b} u^{a+b-1} e^{-\lambda u} du dv \\
&= \int_{\mathbb{R}_+^* \times ]0,1[} \varphi(u, v) \tilde{f}_U(u) \tilde{f}_V(v) du dv.
\end{aligned}$$

Ainsi  $(U, V)$  admet une densité qui se factorise comme produit de deux fonctions (positives !) à variables séparées. Donc  $U$  et  $V$  sont indépendantes. Il reste à calculer les constantes de normalisation. On sait que  $(u, v) \mapsto \tilde{f}_U(u) \tilde{f}_V(v)$  est une densité de probabilité. De plus on sait aussi que  $\int_{\mathbb{R}_+^*} \tilde{f}_V(v) dv = \Gamma(a + b)$ . Par conséquent :

$$1 = \int_{\mathbb{R}_+^* \times ]0,1[} \tilde{f}_U(u) \tilde{f}_V(v) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \int_{]0,1[} v^{a-1} (1 - v)^{b-1} dv.$$

Finalement

$$\begin{aligned}
f_U(u) &= \frac{\lambda^{a+b} u^{a+b-1} e^{-\lambda u}}{\Gamma(a + b)} \mathbf{1}_{\mathbb{R}_+^*}(u) \\
f_V(v) &= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)} v^{a-1} (1 - v)^{b-1} \mathbf{1}_{]0,1[}(v).
\end{aligned}$$

REMARQUE.— La loi de  $V$  s'appelle

4. Puisque  $f_V$  est une densité de probabilité, elle doit intégrer à 1 sur  $]0,1[$ . D'où :

$$B(a, b) = \int_{]0,1[} v^{a-1} (1 - v)^{b-1} dv = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}.$$

REMARQUE.— — Le loi Bêta est utilisée en statistique bayésienne pour estimer le paramètre  $p$  de lois comme la loi binomiale ou géométrique. On suppose au préalable que  $p$  est aléatoire, suivant la loi Bêta (de support effectivement égal à  $[0, 1]$ ) puis on utilise la formule de Bayes pour calculer la loi de  $p$  sachant les observations. La loi Bêta est donc utilisée comme **loi a priori** sur  $p$ .

— La fonction  $\Gamma$  généralise la factorielle aux nombres réels strictement positifs, donc la formule que l'on vient de démontrer permet de dire que la fonction  $B$  généralise les coefficients binomiaux aux réels strictements positifs. En effet, pour  $(a, b) \in \mathbb{N}^2$ , on a :

$$B(a+1, b+1) = \frac{a!b!}{(a+b+1)!} = \frac{1}{(a+b+1)} \frac{1}{\binom{a+b}{a}}.$$

**Exercice 4.33** ▷

1. Les calculs sont exactement les mêmes que dans l'exercice précédent, et en fait, en utilisant la question 1. de ce même exercice, on sait que  $X_1^2$  et  $X_2^2$  suivent toutes deux la loi  $\Gamma(1/2, 1/2)$ . Ainsi le résultat de la question 3., valable car  $X_1^2$  et  $X_2^2$  sont indépendantes, donne que :

$$(U, V) \hookrightarrow \beta(1/2, 1/2) \otimes \Gamma(1/2, 1),$$

où  $\beta(a, b)$  est la loi Bêta (voir exercice précédent) de densité :

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)} \mathbf{1}_{[0,1]}(x).$$

En particulier  $U$  et  $V$  sont indépendantes.

2. En explicitant les densités pour les paramètres de la question précédente, on trouve :

$$f_U(u) = \frac{1}{\pi\sqrt{u(1-u)}} \mathbf{1}_{[0,1]}(u),$$

$$f_V(v) = \frac{1}{2} e^{-\frac{1}{2}v} \mathbf{1}_{\mathbb{R}_+^*}(v).$$

On reconnaît pour la première densité la loi de l'arcsinus (hors-programme mais importante pour l'étude des marches aléatoires). La seconde est simplement une loi exponentielle de paramètre  $1/2$ .

REMARQUE.– Rappelons qu'une densité de probabilité n'est définie qu'à "un presque partout" près. En d'autres termes, on peut poser également  $f(0) = f(1) = 0$  ou  $+\infty$  sans que cela ne change rien à la loi de  $U$ . Pour une densité de probabilité, seule compte son comportement une fois intégrée contre la mesure de Lebesgue. Changer un nombre infini (discret) de ses valeurs n'a aucun impact sur son comportement dans une intégrale contre la mesure de Lebesgue (c'est pour cela que la régularité des densités de probabilité ne sera jamais abordée : elles peuvent être discontinues ou de classe  $\mathcal{C}^\infty$  et définir la même loi!).

3. Suivant l'indice de l'énoncé, on remarque que :

$$U = \frac{X_1^2}{X_1^2 + X_2^2} = \frac{X_1^2}{X_2^2} \frac{X_2^2}{X_1^2 + X_2^2} = \frac{1}{Z} \frac{X_2^2}{X_1^2 + X_2^2} = \frac{1}{Z}(1 - U),$$

et donc :

$$Z = \frac{1}{U} - 1.$$

On doit donc calculer la loi de  $Z := 1/U - 1$ . On remarque au préalable que  $Z(\Omega) = \mathbb{R}_+$ . Par la méthode de la fonction muette, il vient <sup>12</sup> :

---

12. Dans la suite on intègre sur l'ouvert  $]0, 1[$  pour éviter les problèmes. La mesure de Lebesgue ne chargeant pas les singletons, en particulier  $\{0\}$  et  $\{1\}$ , cela ne change pas la valeur de l'intégrale.

$$\begin{aligned}\mathbb{E}\left[\varphi\left(\frac{1}{U}-1\right)\right] &= \int_{]0,1[} \varphi\left(\frac{1}{u}-1\right) f_U(u) du \\ &= \int_{\mathbb{R}_+} \varphi(z) \frac{1}{(z+1)^2} f_U\left(\frac{1}{z+1}\right) dz,\end{aligned}$$

de sorte qu'après simplifications :

$$f_Z(z) = \frac{1}{\pi(w+1)\sqrt{w}} \mathbf{1}_{\mathbb{R}_+}(w).$$

On reconnaît (ou pas!) une loi **Bêta prime** de paramètres  $\alpha = \beta = 1/2$ . Aucune idée de à quoi elle sert celle-là!

### Exercice 5.11 ▷

- 1 C'est la fonction qui à tout  $t$  associe  $1 - (t_0/t)^\alpha$  si  $t \geq t_0$  et 0 sinon.
- 2 La fonction de répartition est de classe  $C^1$  par morceaux. Elle admet donc une densité, qui coïncide avec sa dérivée presque partout, soit  $f_T(t) = \alpha t_0^\alpha / t^{\alpha+1} \mathbf{1}_{t \geq t_0}$ .
- 3 D'après l'énoncé, c'est la densité uniforme sur  $[0, t]$ , soit  $\frac{1}{t} \mathbf{1}_{[0,t]}$ .
- 4 On a  $\mathbb{E}(X_1|T=t) = \frac{1}{t} \int_0^t x dx = \frac{t}{2}$ . Ainsi,

$$\mathbb{E}(X_1) = \mathbb{E}(\mathbb{E}(X_1|T)) = \mathbb{E}\left(\frac{T}{2}\right) = \frac{\alpha t_0^\alpha}{2} \int \frac{dt}{t^\alpha}$$

donc  $\mathbb{E}(X_1) = +\infty$  lorsque  $\alpha \leq 1$  et dans le cas contraire,

$$\mathbb{E}(X_1) = \frac{\alpha t_0}{2(\alpha - 1)}.$$

- 5 Les variables aléatoires  $X_1, \dots, X_n$  étant indépendantes conditionnellement à  $\{T=t\}$ , c'est le produit des densités, soit :

$$f_{X|T=t}(x) = \frac{1}{t^n} \mathbf{1}_{[0,t]^n}(x).$$

- 6 On calcule pour  $x_1, \dots, x_n \geq 0$ ,

$$f_{T|X=x}(t) = \frac{f_{T,X}(t,x)}{f_X(x)}.$$

La densité jointe est donnée par

$$\begin{aligned}f_{T,X}(t,x) &= f_T(t) \prod_{i=1}^n f_{X_i|T=t}(x_i) \\ &= \frac{f_T(t)}{t^n} \prod_{i=1}^n \mathbf{1}_{[0,t]}(x_i) \\ &= \frac{f_T(t)}{t^n} \mathbf{1}_{[0,t]}(m(x)) \\ &= \frac{\alpha t_0^\alpha}{t^{n+\alpha+1}} \mathbf{1}_{[m(x) \vee t_0, +\infty[}(t)\end{aligned}$$

Donc à  $x$  fixé, il existe une constante  $\beta(x)$  ne dépendant pas de  $t$  telle que

$$f_{T|X=x}(t) = \frac{\beta(x)}{t^{n+\alpha+1}} \mathbf{1}_{[m(x) \vee t_0, +\infty[}(t)$$

Ainsi  $T|X = x$  suit une loi de Pareto de paramètres  $(m(x) \vee t_0, \alpha + n)$ .

7 Après calcul,  $\mathbb{E}(T|X = x) = \frac{\alpha+n}{\alpha+n-1}(m(x) \vee t_0)$ .

**Exercice 5.13** ▷

1. D'après l'indépendance des variables aléatoires en jeu,

$$\mathbf{P}(X \in \prod_{i=1}^n ]-\infty, \alpha_i]) = \prod_{i=1}^n \mathbf{P}(X_i \in ]-\infty, \alpha_i]). \quad (\text{B.4})$$

D'après la formule des probabilités totales (cf. théorème 2.6),

$$\begin{aligned} & \mathbf{P}(X' \in \prod_{i=1}^n ]-\infty, \alpha_i]) \\ &= \sum_{j=1}^n \mathbf{P}(X' \in \prod_{i=1}^n ]-\infty, \alpha_i], U = j) \\ &= \sum_{j=1}^n \mathbf{P}(\dots, X_j \in ]-\infty, \alpha_j], Y_j \in ]-\infty, \alpha_j], X_{j+1} \in ]-\infty, \alpha_{j+1}], \dots, U = j) \quad (\text{B.5}) \\ &= \sum_{j=1}^n \prod_{i \neq j} \mathbf{P}(\dots, X_i \in ]-\infty, \alpha_i]) \mathbf{P}(Y_j \in ]-\infty, \alpha_j]) \mathbf{P}(U = j), \end{aligned}$$

où la dernière ligne provient de l'indépendance de  $(X_i, i \neq j, Y_j, U)$ . Comme  $X_j$  et  $Y_j$  ont même loi,

$$\mathbf{P}(Y_j \in ]-\infty, \alpha_j]) = \mathbf{P}(X_j \in ]-\infty, \alpha_j]).$$

En reportant dans (B.5), on obtient

$$\mathbf{P}(X' \in \prod_{i=1}^n ]-\infty, \alpha_i]) = \sum_{j=1}^n \frac{1}{n} \prod_{i=1}^n \mathbf{P}(X_i \in ]-\infty, \alpha_i]) = \prod_{i=1}^n \mathbf{P}(X_i \in ]-\infty, \alpha_i]),$$

puisque aucun des membres de la somme ne dépend de  $n$ . Par comparaison avec (B.4), on voit que

$$\mathbf{P}(X' \in \prod_{i=1}^n ]-\infty, \alpha_i]) = \mathbf{P}(X \in \prod_{i=1}^n ]-\infty, \alpha_i]).$$

En vertu du théorème ??, cela signifie que  $X$  et  $X'$  ont même loi.

2. Considérons l'application

$$\begin{aligned} T : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (x_1, \dots, x_n) &\longmapsto \sum_{i=1}^n x_i. \end{aligned}$$

Comme  $W = T(X)$  et  $W' = T(X')$ , on déduit de la question précédente que  $W$  et  $W'$  ont même loi.

3.  $W_n$  suit une loi binomiale de paramètres  $n$  et  $p$  (cf. 1.2.3).
4. De même,  $W_{n-1}$  suit une loi binomiale de paramètres  $n-1$  et  $p$ . D'une part,  $W_n = X_1 + \sum_{j \neq 1} X_j$ . D'autre part, les variables  $X_1$  et  $\sum_{j \neq 1} X_j$  sont indépendantes. Enfin,

$$\sum_{j \neq 1} X_j = T(X_2, \dots, X_n)$$

donc  $\tilde{W}_{n-1} := \sum_{j \neq 1} X_j$  a la même loi que  $W_{n-1}$ . On a bien montré que

$$W_n = X_1 + \tilde{W}_{n-1}$$

avec  $\tilde{W}_{n-1} \perp X_1$  et  $\tilde{W}_{n-1} \stackrel{\text{loi}}{=} W_{n-1}$ .

5. Par définition de la probabilité conditionnelle (définition 1.3),

$$\begin{aligned} \mathbf{P}(X_1 = 1 \mid W_n = m) &= \frac{\mathbf{P}(X_1 = 1, W_n = m)}{\mathbf{P}(W_n = m)} \\ &= \frac{\mathbf{P}(X_1 = 1, X_1 + \tilde{W}_{n-1} = m)}{\mathbf{P}(W_n = m)} \\ &= \frac{\mathbf{P}(X_1 = 1, \tilde{W}_{n-1} = m - 1)}{\mathbf{P}(W_n = m)} \\ &= \mathbf{P}(X_1 = 1) \frac{\mathbf{P}(\tilde{W}_{n-1} = m - 1)}{\mathbf{P}(W_n = m)}, \end{aligned}$$

où la dernière ligne vient de l'indépendance de  $X_1$  et de  $\tilde{W}_{n-1}$ . Par ailleurs, les lois de  $\tilde{W}_{n-1}$  et  $W_n$  étant connues, on obtient

$$\begin{aligned} \mathbf{P}(X_1 = 1 \mid W_n = m) &= p \frac{\binom{n-1}{m-1} p^{m-1} (1-p)^{n-1-(m-1)}}{\binom{n}{m} p^m (1-p)^{n-m}} \\ &= \frac{(n-1)! m! (n-m)!}{(m-1)! (n-1-(m-1))! n!} \\ &= \frac{m}{n}. \end{aligned}$$

Ce résultat n'est pas une surprise. Il signifie que si l'on a eu  $m$  succès dans une série de  $n$  tentatives, la probabilité d'avoir eu un succès à la première tentative est de  $m/n$ . On peut aisément se convaincre (ou refaire les calculs) que ce résultat reste vrai pour tout indice, i.e.

$$\mathbf{P}(X_i = 1 \mid W = m) = \frac{m}{n}, \text{ pour tout } i \in \{1, \dots, n\}.$$

Une probabilité conditionnelle étant une mesure de probabilité, on en déduit aussi que

$$\mathbf{P}(X_i = 0 \mid W = m) = 1 - \mathbf{P}(X_i = 1 \mid W = m) = 1 - \frac{m}{n}.$$

6. Comme une probabilité conditionnelle est une mesure de probabilité, la formule des

probabilités totales s'applique (cf. théorème 2.6). Par conséquent,

$$\begin{aligned}
 \mathbf{P}(W_n - W'_n = \epsilon \mid W_n = m) &= \mathbf{P}(X_U - Y_U = \epsilon \mid W_n = m) \\
 &= \sum_{j=1}^n \mathbf{P}(X_U - Y_U = \epsilon, U = j \mid W_n = m) \\
 &= \sum_{j=1}^n \mathbf{P}(X_j = Y_j + \epsilon, U = j \mid W_n = m) \\
 &= \frac{1}{\mathbf{P}(W_n = m)} \sum_{j=1}^n \mathbf{P}(X_j = Y_j + \epsilon, U = j, W_n = m) \\
 &= \frac{1}{\mathbf{P}(W_n = m)} \sum_{j=1}^n \mathbf{P}(X_j = Y_j + \epsilon, W_n = m) \mathbf{P}(U = j).
 \end{aligned} \tag{B.6}$$

Il nous reste donc à calculer

$$\mathbf{P}(X_j = Y_j + \epsilon, W_n = m).$$

Posons  $Z_j^\epsilon = Y_j + \epsilon$  dont on remarque que c'est une v.a.r. indépendante de  $X_j$  et  $W_n$ . Toujours d'après la formule des probabilités totales (cf. théorème 2.6),

$$\begin{aligned}
 \mathbf{P}(X_j = Y_j + \epsilon, W_n = m) &= \\
 &= \mathbf{P}(X_j = 0, W_n = m, Z_j^\epsilon = 0) + \mathbf{P}(X_j = 1, W_n = m, Z_j^\epsilon = 1).
 \end{aligned}$$

Comme  $Z_j^\epsilon$  est indépendante de  $(X_j, W_n)$ , on obtient

$$\begin{aligned}
 \mathbf{P}(X_j = Y_j + \epsilon, W_n = m) &= \mathbf{P}(X_j = 0, W_n = m) \mathbf{P}(Z_j^\epsilon = 0) + \mathbf{P}(X_j = 1, W_n = m) \mathbf{P}(Z_j^\epsilon = 1) \\
 &= \mathbf{P}(X_j = 0 \mid W_n = m) \mathbf{P}(Z_j^\epsilon = 0) \mathbf{P}(W_n = m) \\
 &\quad + \mathbf{P}(X_j = 1 \mid W_n = m) \mathbf{P}(W_n = m) \mathbf{P}(Z_j^\epsilon = 1) \mathbf{P}(W_n = m)
 \end{aligned} \tag{B.7}$$

d'après la question précédente. Maintenant, d'après la définition de  $Z_j^\epsilon$ ,

$$\begin{aligned}
 \mathbf{P}(Z_j^{-1} = 1) &= \mathbf{P}(Y_j = 2) = 0; \\
 \mathbf{P}(Z_j^{-1} = 0) &= \mathbf{P}(Y_j = 1) = p \\
 \\ \\
 \mathbf{P}(Z_j^0 = 1) &= \mathbf{P}(Y_j = 1) = p; \\
 \mathbf{P}(Z_j^0 = 0) &= \mathbf{P}(Y_j = 0) = 1 - p \\
 \\ \\
 \mathbf{P}(Z_j^1 = 1) &= \mathbf{P}(Y_j = 0) = 1 - p; \\
 \mathbf{P}(Z_j^1 = 0) &= \mathbf{P}(Y_j = -1) = 0.
 \end{aligned} \tag{B.8}$$

On constate donc que  $\mathbf{P}(Z_j^\epsilon = 1)$  ne dépend pas de l'indice  $j$ . Par conséquent, en reportant (B.7) dans (B.6), il vient

$$\mathbf{P}(W_n - W'_n = \epsilon \mid W_n = m) = \left(1 - \frac{m}{n}\right) \mathbf{P}(Z_1^\epsilon = 0) + \frac{m}{n} \mathbf{P}(Z_1^\epsilon = 1).$$

Compte-tenu de (B.8),

$$\begin{aligned}\mathbf{P}(W_n - W'_n = -1 \mid W_n = m) &= p\left(1 - \frac{m}{n}\right), \\ \mathbf{P}(W_n - W'_n = 0 \mid W_n = m) &= (1 - p)\left(1 - \frac{m}{n}\right) + p\frac{m}{n}, \\ \mathbf{P}(W_n - W'_n = 1 \mid W_n = m) &= (1 - p)\frac{m}{n}.\end{aligned}$$

7. On a

$$\begin{aligned}\mathbb{E}(W_n - W'_n \mid W_n = m) &= \sum_{\epsilon=\pm 1,0} \epsilon \mathbf{P}(W_n - W'_n = \epsilon \mid W_n = m) \\ &= (+1)\mathbf{P}(W_n - W'_n = 1 \mid W_n = m) + (-1)\mathbf{P}(W_n - W'_n = -1 \mid W_n = m) \\ &= \frac{m}{n} - p.\end{aligned}$$

### Exercice 5.14 ▷

- 1 On sait d'après le cours qu'avec une telle matrice de covariance,  $X \sim \mathcal{N}(0, 1)$ .
- 2 Par définition de la matrice de covariance,  $\text{cov}(X, Y) = \rho$ . Comme  $(X, Y)$  est un vecteur gaussien, elles sont indépendantes ssi  $\rho = 0$ .
- 3 Le polynôme caractéristique est

$$(X - 1)^2 - \rho^2 = (X - 1 - \rho)(X - 1 + \rho).$$

Les deux valeurs propres sont donc  $\rho_1 = 1 + \rho$  et  $\rho_2 = 1 - \rho$ . La matrice est définie positive si et seulement ces deux valeurs sont strictement positives, ce qui implique  $\rho \in ]-1, 1[$ .

- 4 On sait qu'il existe  $A$  orthogonale telle que

$$\Gamma = {}^t O \begin{pmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{pmatrix} O.$$

On construit  $\Gamma^{1/2}$  en prenant

$$\Gamma^{1/2} = {}^t O \begin{pmatrix} \sqrt{\rho_1} & 0 \\ 0 & \sqrt{\rho_2} \end{pmatrix} O.$$

Les résultats demandés s'en déduisent aisément. Par exemple,

$$\|(\Gamma^{1/2})^{-1}z\|^2 = (\Gamma^{1/2})^{-1}z \cdot (\Gamma^{1/2})^{-1}z = (\Gamma^{1/2})^{-1}(\Gamma^{1/2})^{-1}z \cdot z$$

puisque  $(\Gamma^{1/2})^{-1}$  est symétrique. D'autre part,

$$\begin{aligned}(\Gamma^{1/2})^{-1}(\Gamma^{1/2})^{-1} &= {}^t O \begin{pmatrix} \sqrt{\rho_1}^{-1} & 0 \\ 0 & \sqrt{\rho_2}^{-1} \end{pmatrix} O {}^t O \begin{pmatrix} \sqrt{\rho_1}^{-1} & 0 \\ 0 & \sqrt{\rho_2}^{-1} \end{pmatrix} O \\ &= {}^t O \begin{pmatrix} \rho_1^{-1} & 0 \\ 0 & \rho_2^{-1} \end{pmatrix} O = \Gamma^{-1}.\end{aligned}$$

5 On sait que  $(X, Y)$  a la même loi que  $\Gamma^{1/2}(U, V)$ . On peut donc considérer que

$$\begin{aligned} X &= \gamma_{11}U + \gamma_{12}V \\ Y &= \gamma_{12}U + \gamma_{22}V. \end{aligned}$$

Le changement de variables

$$\begin{aligned} x &= \gamma_{11}u + \gamma_{12}v \\ y &= \gamma_{12}u + \gamma_{22}v \end{aligned}$$

est bijectif puisque  $\Gamma^{1/2}$  est bijective. Par ailleurs, la jacobienne d'une application linéaire est elle-même (ce qui se vérifie immédiatement par le calcul) :

$$J_{\Gamma^{1/2}}(u, v) = \Gamma^{1/2}$$

donc le jacobien est  $\det \Gamma^{1/2} = \sqrt{\det \Gamma}$  d'après la question précédente. De plus, comme  $U$  et  $V$  sont indépendantes,

$$f_{(U,V)}(u, v) = f_U(u)f_V(v) = \frac{1}{2\pi}e^{-u^2/2}e^{-v^2/2}.$$

Au final,

$$\begin{aligned} \mathbf{E}[\phi(X, Y)] &= \iint \psi(\gamma_{11}u + \gamma_{12}v, \gamma_{12}u + \gamma_{22}v) e^{-u^2/2} e^{-v^2/2} \frac{du dv}{2\pi} \\ &= \iint \psi(\Gamma^{1/2} \begin{pmatrix} u \\ v \end{pmatrix}) \exp(-\frac{1}{2} \|\begin{pmatrix} u \\ v \end{pmatrix}\|^2) \frac{du dv}{2\pi} \\ &= \iint \psi(x, y) \exp(-\frac{1}{2} \|(\Gamma^{1/2})^{-1}(x, y)\|^2) \frac{1}{\sqrt{\det \Gamma}} \frac{dx dy}{2\pi} \\ &= \iint \psi(x, y) \exp(-\frac{1}{2} \Gamma^{-1}(x, y) \cdot (x, y)) \frac{1}{\sqrt{\det \Gamma}} \frac{dx dy}{2\pi}. \end{aligned}$$

La densité du couple  $(X, Y)$  est donc donnée par

$$\frac{1}{2\pi\sqrt{\det \Gamma}} \exp(-\frac{1}{2} \Gamma^{-1}(x, y) \cdot (x, y)).$$

6 Les formules classiques donnent

$$\Gamma^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} := \begin{pmatrix} \delta & -\zeta \\ -\zeta & \delta \end{pmatrix}$$

avec  $\delta = 1/(1-\rho^2)$  et  $\zeta = \rho\delta$ . La densité jointe du couple est donc

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi\sqrt{\det \Gamma}} \exp(-\frac{1}{2}(\delta x^2 + \delta y^2 - 2\zeta xy)).$$

On sait que la loi conditionnelle de  $Y$  sachant  $X = x$  a pour densité

$$\begin{aligned} y \mapsto \frac{f_{(X,Y)}(x, y)}{f_X(x)} &= \frac{\sqrt{2\pi}}{2\pi\sqrt{\det \Gamma}} \exp(-\frac{1}{2}(\delta x^2 + \delta y^2 - 2\zeta xy - x^2)) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\det \Gamma}} \exp(-\frac{1}{2}(\delta - 1)x^2) \exp(-\frac{1}{2}(\delta y^2 - 2\zeta xy)) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\det \Gamma}} \exp(-\frac{1}{2}(\delta - 1 - \frac{\zeta^2}{\delta})x^2) \exp(-\frac{\delta}{2}(y - \frac{\zeta}{\delta}x)^2) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp(-\frac{\delta}{2}(y - \frac{\zeta}{\delta}x)^2) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp(-\frac{1}{2(1-\rho^2)}(y - \rho x)^2). \end{aligned}$$

La loi conditionnelle de  $Y$  sachant  $X = x$  est donc une loi gaussienne de moyenne  $\rho x$  et de variance  $1/\delta = 1 - \rho^2$ .

**Exercice 6.1** ▷ Les calculs sont immédiats, à l'exception des deux dernières.

Etablissons la formule pour une loi normale  $\mathcal{N}(0, 1)$ . Soit  $Y \sim \mathcal{N}(0, 1)$ . Les conditions de dérivation sous l'intégrale sont vérifiées (voir exercice 3.13) 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.

Le résultat se généralise facilement à une gaussienne  $\mathcal{N}(m, \sigma^2)$ . On peut soit reproduire les calculs, soit plus simplement remarquer que la densité  $\mathcal{N}(m, \sigma^2)$  est la densité de la v.a.  $m + \sigma Y$ , avec  $Y$  de loi normale centrée réduite. Il suffit alors d'exprimer la fonction caractéristique de  $m + \sigma Y$  en fonction de celle de  $Y$  (immédiat) et de remplacer  $\phi_Y$  par l'expression trouvée ci-dessus.

Soit  $X \sim \Gamma(a, b)$ . Par application du théorème de dérivation sous l'intégrale (voir exercice 3.13) 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.

**Exercice 7.1** ▷

Les vecteurs 2., 4., 6. et 7. sont gaussiens, pas les autres.

— Le 2. est un vecteur gaussien standard sur  $\mathbb{R}^2$  d'après le cours, et le 4. s'écrit aussi :

$$\begin{pmatrix} X \\ X + Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Le 4. est gaussien car  $X$  et  $\varepsilon Y$  sont deux variables aléatoires indépendantes (par le lemme des coalitions) et suivant chacune la loi normale  $\mathcal{N}(0, 1)$ . En effet en conditionnant par rapport à  $\varepsilon$ , on trouve<sup>13</sup> :

$$\forall x \in \mathbb{R}, \mathbb{P}(\varepsilon Y \leq x) = \mathbb{E}[\mathbb{P}(\varepsilon Y \leq x | \varepsilon)] = \mathbb{E}[\mathbb{P}(Y \leq x)] = \mathbb{P}(Y \leq x),$$

et donc  $Y$  et  $\varepsilon Y$  ont même loi. On a utilisé que  $Y$  admet une densité paire, donc  $-Y$  a même loi que  $Y$  (on dit que  $Y$  est une variable aléatoire **symétrique**). L'indépendance

13. Notation :  $\mathbb{P}(A|\varepsilon) := \mathbb{E}[\mathbf{1}_A|\varepsilon]$ , pour  $A$  est un événement.

de  $Y$  et de  $\varepsilon$  est fondamentale : si on a  $\varepsilon = \text{sgn}(Y)$ , le signe de  $Y$ , alors on peut prouver que  $\varepsilon$  est encore une variable aléatoire suivant la loi de Rademacher ; mais  $\varepsilon Y = |Y|$ , qui ne suit certainement pas la loi normale ! Si cette preuve ne vous convainc pas (il le faudra pourtant si vous continuez à faire des probabilités/statistiques), vous pouvez revenir à la formule des probabilités totales et utiliser l'identité :

$$\forall x \in \mathbb{R}, F_Y(-x) = 1 - F_Y(x),$$

valable **dès que  $Y$  est une variable aléatoire symétrique.** En particulier  $(X, \varepsilon Y)$  est un vecteur gaussien standard sur  $\mathbb{R}^2$ . Enfin  $(X, X + \varepsilon Y)$  est aussi gaussien, pour la même raison que  $(X, X + Y)$  l'est (transformation linéaire du vecteur gaussien  $(X, \varepsilon Y)$ ).

- $(X, \varepsilon)$  n'est pas gaussien car son support est  $\mathbb{R} \times \{-1/2, 1/2\}$ . Or le support d'un vecteur gaussien ne peut être qu'un espace affine (dont le terme source est donné par le vecteur espérance).  $(X, \varepsilon X)$  n'est pas gaussien, bien que ses coordonnées suivent chacune la loi normale  $\mathcal{N}(0, 1)$ . En effet  $X + \varepsilon X = (1 + \varepsilon)X$  n'est ni une variable aléatoire suivant une loi de Dirac, ni une variable aléatoire à densité (elle vaut 0 avec probabilité  $1/2 \neq 0$ , donc admet un atome!). En particulier on a exhibé une combinaison linéaire des coordonnées de  $(X, \varepsilon X)$  qui ne suit pas une loi normale.  $(\varepsilon|X|, \varepsilon|Y|)$  ne peut pas non plus être un vecteur gaussien car son support n'est pas un espace affine, mais :

$$(\varepsilon|X|, \varepsilon|Y|)(\Omega) = \mathbb{R}_-^2 \cup \mathbb{R}_+^2,$$

puisque le signe de ses coordonnées est le même. Si l'on avait remplacé  $\varepsilon Y$  par  $\varepsilon' Y$ , avec  $\varepsilon'$  de même loi que  $\varepsilon$  et indépendante de  $X, Y$  et de  $\varepsilon$ , alors on aurait eu un vecteur gaussien<sup>14</sup> ! Enfin, supposons que  $(X, \varepsilon X + Y)$  soit gaussien. Comme :

$$\text{Cov}(X, \varepsilon X + Y) = \mathbb{E}[X(\varepsilon X + Y)] = 0,$$

on aurait que  $X$  et  $\varepsilon X + Y$  sont indépendantes. Pourtant :

$$\mathbb{E}[X^2(\varepsilon X + Y)^2] = \mathbb{E}[X^4] + \mathbb{E}[X^2 Y^2] + 0 = 4 \neq 2 = \mathbb{E}[X^2] \mathbb{E}[(\varepsilon X + Y)^2].$$

Donc notre hypothèse de départ est absurde !

### Exercice 7.2 ▷

1. Par hypothèse, on sait que  $\mathbf{X}$  est un vecteur gaussien, donc toute combinaison linéaire de ses sous-vecteurs redonne un vecteur gaussien. Or la loi d'un vecteur gaussien est déterminée uniquement par son vecteur espérance et sa matrice de covariance. Par linéarité de l'espérance, on a clairement :

$$\mathbb{E}[A\mathbf{X} + \mathbf{b}] = A\mathbb{E}[\mathbf{X}] + \mathbf{b} = A\mathbf{m} + \mathbf{b}.$$

De plus le cours donne :

$$\text{Cov}(A\mathbf{X} + \mathbf{b}) = A\text{Cov}(\mathbf{X})A^T = A\Sigma A^T.$$

---

14. Pourquoi ?

D'où finalement :

$$A\mathbf{X} + \mathbf{b} \hookrightarrow \mathcal{N}(A\mathbf{m} + \mathbf{b}, A\Sigma A^T).$$

2. La question précédente appliquée à :

$$A = \sqrt{\Sigma}^{-1}, \text{ et}$$

$$\mathbf{b} = -\sqrt{\Sigma}^{-1} \mathbf{m}$$

donne immédiatement que :

$$\sqrt{\Sigma}^{-1}(\mathbf{X} - \mathbf{m}) \hookrightarrow \mathcal{N}(\mathbf{0}_{\mathbb{R}^d}, I_d).$$

REMARQUE.— Aussi simple que puisse paraître cet exercice, la seconde question a cependant une conséquence pratique très utile : pour simuler un vecteur gaussien de matrice de covariance  $\Sigma$  dans  $\mathbb{R}^d$ , il suffit de simuler  $d$  variables aléatoires i.i.d. suivant la loi normale  $\mathcal{N}(0, 1)$  et de calculer la racine carrée de l'inverse de  $\Sigma$ . En d'autres mots, tous les problèmes de simulation se ramènent uniquement à des problèmes d'algèbre linéaire numérique, des problèmes qui sont pour l'essentiel bien compris maintenant !

### Exercice 7.6 ▷

1. L'espace d'arrivée de  $\psi$  est  $\{(y_1, y_2, \dots, y_d) \mid y_1 \geq 0 \text{ et } y_2^2 + \dots + y_d^2 \leq y_1\}$ .
2. L'idée va être d'appliquer la méthode de la variable muette. On se donne donc  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction-test, disons positive mesurable. Comme suggéré dans l'énoncé, on pose  $(y_1, y_2, \dots, y_d) = (\|\mathbf{x}\|^2, x_2, \dots, x_d)$ . Tel quel, ce changement de variable n'est pas injectif, car on ne peut pas retrouver le signe de  $x_1$  juste en connaissant la norme de  $\mathbf{x}$  et les autres valeurs  $x_2, \dots, x_d$ . Qu'à cela ne tienne, on écrit<sup>15</sup> :

$$\begin{aligned} \mathbb{E}[\varphi(Y)] &= \int_{\mathbb{R}^d} \varphi(\|\mathbf{x}\|^2) \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|\mathbf{x}\|^2}{2}} d\mathbf{x} \\ &= \int_{x_1 \in \mathbb{R}_+} \left( \int_{\mathbb{R}^{d-1}} \varphi(\|\mathbf{x}\|) e^{-\frac{\|\mathbf{x}\|^2}{2}} dx_2 \dots dx_d \right) dx_1 + \int_{x_1 \in \mathbb{R}_-} \left( \int_{\mathbb{R}^{d-1}} \varphi(\|\mathbf{x}\|) e^{-\frac{\|\mathbf{x}\|^2}{2}} dx_2 \dots dx_d \right) dx_1. \end{aligned}$$

On fait donc deux changements de variable au lieu d'un selon le signe de  $x_1$ , mais dans le fond cette distinction de cas ne va pas changer grand chose. Déjà on trouve lorsque  $x_1 \in \mathbb{R}_+$  :

$$\begin{cases} x_1 &= \sqrt{y_1 - y_2^2 - \dots - y_d^2} \\ x_2 &= y_2 \\ &\dots \\ x_d &= y_d. \end{cases}$$

15. Avec Fubini-Tonelli tout est permis !

On trouve la même chose lorsque  $x_1 \leq 0$ , à ceci près que  $x_1 = -\sqrt{y_1 - y_2^2 - \dots - y_d^2}$ . La matrice Jacobienne du changement de variable vaut :

$$J(x_1, x_2, \dots, x_d) = \begin{pmatrix} 2x_1 & 2x_2 & \dots & 2x_d \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Le déterminant de cette matrice triangulaire est donc  $2x_1 = \pm 2\sqrt{y_1 - y_2^2 - \dots - y_d^2}$  dans les nouvelles coordonnées. On trouve ainsi, en n'oubliant pas la valeur absolue dans le Jacobien, qui "tue" le signe résultant de la distinction de cas, et de changer l'espace d'arrivée sur lequel on intègre (qu'on connaît grâce à la première question) :

$$\begin{aligned} \mathbb{E}[\varphi(Y)] &= \int_{\mathbb{R}^d} \varphi(\|\mathbf{x}\|^2) \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|\mathbf{x}\|^2}{2}} d\mathbf{x} \\ &= \frac{1}{(2\pi)^{d/2}} \int_{x_1 \in \mathbb{R}_+} \left( \int_{\mathbb{R}^{d-1}} \varphi(\|\mathbf{x}\|) e^{-\frac{\|\mathbf{x}\|^2}{2}} dx_2 \dots dx_d \right) dx_1 \\ &\quad + \frac{1}{(2\pi)^{d/2}} \int_{x_1 \in \mathbb{R}_-} \left( \int_{\mathbb{R}^{d-1}} \varphi(\|\mathbf{x}\|) e^{-\frac{\|\mathbf{x}\|^2}{2}} dx_2 \dots dx_d \right) dx_1 \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}_+} \int_{\{y_1^2 + \dots + y_{d-1}^2 \leq y_1\}} \varphi(y_1) \frac{1}{\sqrt{y_1 - y_2^2 - \dots - y_{d-1}^2}} e^{-\frac{y_1}{2}} dy_2 \dots dy_{d-1} dy_1 \\ &= \frac{1}{(2\pi)^{d/2}} \frac{\pi^{\frac{d}{2}-1}}{\Gamma(\frac{d}{2})} \int_{\mathbb{R}_+} \varphi(y_1) y_1^{\frac{d}{2}-1} e^{-\frac{y_1}{2}} dy_1 \\ &= \frac{1}{2^{d/2} \Gamma(\frac{d}{2})} \int_{\mathbb{R}_+} \varphi(y_1) y_1^{\frac{d}{2}-1} e^{-\frac{y_1}{2}} dy_1, \end{aligned}$$

grâce au résultat admis dans l'énoncé. Finalement,  $\|\mathbf{X}\|^2$  admet pour densité :

$$f(y) = \frac{1}{2^{\frac{d}{2}} \Gamma(\frac{d}{2})} y^{\frac{d}{2}-1} e^{-\frac{y}{2}} \mathbf{1}_{\mathbb{R}_+}(y).$$

C'est la loi du Chi-2 à  $d$  degrés de liberté, qu'on note  $\chi^2(d)$  (prononcer "Ki deux") et qui est d'une très grande importance en statistiques.

**Exercice 7.7** ▷

1. Utilisons la méthode de la variable muette. Soit  $\varphi$  une fonction-test. On a :

$$\begin{aligned}
\mathbb{E}[\varphi(Z^2)] &= \int_{\mathbb{R}} \varphi(z^2) f_Z(z) dz \\
&= \int_{\mathbb{R}_+} \varphi(z^2) f_Z(z) dz + \int_{\mathbb{R}_-} \varphi(z^2) f_Z(z) dz \\
&= \int_0^\infty \varphi(u) \frac{1}{2\sqrt{u}} f_Z(\sqrt{u}) du + \int_{+\infty}^0 \varphi(u) f_Z(-\sqrt{u}) \frac{1}{-2\sqrt{u}} du \\
&= \int_{\mathbb{R}_+} \varphi(u) \frac{1}{2\sqrt{u}} f_Z(\sqrt{u}) du + \int_{\mathbb{R}_+} \varphi(u) f_Z(-\sqrt{u}) \frac{1}{2\sqrt{u}} du \\
&= \int_{\mathbb{R}_+} \varphi(u) \frac{1}{\sqrt{u}} f_Z(\sqrt{u}) du \\
&= \int_{\mathbb{R}_+} \varphi(u) \frac{1}{\sqrt{u}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u}{2}} du \\
&= \int_{\mathbb{R}_+} \varphi(u) \frac{u^{\frac{1}{2}-1} e^{-\frac{u}{2}}}{\sqrt{2}\Gamma(1/2)} du
\end{aligned}$$

La séparation en deux intégrales est nécessaire puisque  $x \mapsto x^2$  n'est pas un  $\mathcal{C}^1$ -difféomorphisme sur  $\mathbb{R}$  (elle n'est pas injective!). Elle l'est en revanche sur  $\mathbb{R}_+$  et  $\mathbb{R}_-$ . On retrouve une loi Gamma  $\Gamma(1/2, 1/2)$ .

2. Le résultat admis dans l'énoncé (et étendue par récurrence à une somme de  $n$  variables aléatoires indépendantes) nous donne que  $\|\mathbf{X}\|_2^2 \hookrightarrow \Gamma(1/2, n/2)$ . Ainsi  $\|\mathbf{X}\|_2^2 = \sum_{i=1}^d X_i^2$  admet la densité :

$$f(x) = \frac{1}{2^{\frac{d}{2}} \Gamma(\frac{d}{2})} x^{\frac{d}{2}-1} e^{-\frac{x}{2}} \mathbf{1}_{\mathbb{R}_+}(x).$$

### Exercice 7.8 ▷

1. La densité d'un vecteur aléatoire caractérise sa loi, exactement comme dans le cas univarié. De plus l'énoncé suggère de montrer que  $(X, Y)$  est un vecteur gaussien en utilisant sa densité, la seule information disponible concernant sa loi. D'après le cours, il suffit de reconnaître dans la quantité en exponentielle un produit scalaire faisant intervenir  $\mathbf{m}$  et  $\Sigma$ . On vérifie donc que :

$$\Sigma^{-1} = \frac{1}{4} \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix},$$

puis que :

$$\forall \mathbf{z} = (x, y) \in \mathbb{R}^2, \frac{1}{2} \langle \Sigma^{-1} \mathbf{z}, \mathbf{z} \rangle = \frac{1}{8} (3x^2 + 2xy + 3y^2).$$

Il reste à calculer la constante de normalisation  $C$ . Toujours d'après le cours :

$$C = \frac{1}{\sqrt{(2\pi)^2 \det(\Sigma)}} = \frac{1}{2\pi\sqrt{2}}.$$

2. Le cours nous indique que  $X$  et  $Y$ , en tant que sous-vecteurs d'un vecteur gaussien, suivent elles-mêmes des lois normales. En lisant le vecteurs d'espérance de  $(X, Y)$  ainsi que sa matrice de covariance, on trouve :

$$X \hookrightarrow \mathcal{N}(0, 3/2),$$

$$Y \hookrightarrow \mathcal{N}(0, 3/2).$$

Donc  $X$  et  $Y$  ont même loi. Pour autant elles ne sont pas indépendantes, puisque leur covariance est non nulle :

$$\text{Cov}(X, Y) = -1/2 \neq 0.$$

Rappelons qu'être décorréelées est une condition **nécessaire**, mais non suffisante, pour que deux variables aléatoires soient indépendantes. Elle devient suffisante lorsque ces deux variables forment un vecteur gaussien.

3. La logique pour ce type de question sera toujours la même : on part d'un vecteur gaussien (ici  $(X, Y)$ ) puis on lui applique une transformation linéaire et on demande la loi du nouveau vecteur. Le plus simple est de tout écrire avec des matrices ; cela évitera pas mal d'erreurs. On pose :

$$M = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Il est clair que  $M(X, Y)^T$  est encore un vecteur gaussien (comme transformation linéaire d'un autre vecteur gaussien), et il reste à donner son vecteur d'espérance et sa matrice de covariance. On trouve :

$$\mathbf{m}_{M(X, Y)^T} = M\mathbf{m} = \mathbf{0}_{\mathbb{R}^2},$$

$$\Sigma_{M(X, Y)^T} = M\Sigma M^T = \begin{pmatrix} 4 & 0 \\ 0 & 8 \end{pmatrix},$$

qui est une matrice diagonale. Donc  $X + Y$  et  $X - Y$  sont indépendantes, alors que  $X$  et  $Y$  ne l'étaient pas.

### Exercice 7.9 ▷

1.  $U$  est somme de trois variables aléatoires indépendantes suivant la loi normale  $\mathcal{N}(0, 1)$  (rappelons que la loi normale est symétrique par rapport à 0, d'où  $-X_2 \hookrightarrow \mathcal{N}(0, 1)$ ), et donc  $U \hookrightarrow \mathcal{N}(0, 3)$ .
2. On va utiliser les propriétés des vecteurs gaussiens. Mais pour cela il nous faut d'abord un vecteur gaussien "de base" pour commencer ! Ici il s'agit de  $\mathbf{X} := (X_1, X_2, X_3)$ , qui suit la loi normale centrée sur  $\mathbb{R}^3$ , de matrice de covariance  $I_3$ . On peut alors affirmer que  $(U, T_1, T_2, T_3)^T$  est aussi un vecteur gaussien, car :

$$\begin{pmatrix} U \\ T_1 \\ T_2 \\ T_3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} =: M\mathbf{X}.$$

Et donc par le cours  $(U, T_1, T_2, T_3)^T \hookrightarrow \mathcal{N}(\mathbf{0}_{\mathbb{R}^4}, \Sigma)$ , où :

$$\Sigma = MI_3M^T = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 1 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 1 & -1 & 2 \end{pmatrix}.$$

En particulier, les coefficients extra-diagonaux de la première ligne sont nuls. Puisque  $(U, T_1, T_2, T_3)^T$  est un vecteur gaussien, cela suffit pour dire que  $U$  est indépendante de  $(T_1, T_2, T_3)$ , et donc de  $T_i$  pour tout  $i \in \llbracket 1, 3 \rrbracket$ .

- a) Il suffit d'extraire de la matrice  $M$  la sous-matrice associée au sous-vecteur  $(T_1, T_2, T_3)^T$  de  $(U, T_1, T_2, T_3)^T$  (les sceptiques pourront retrouver  $A$  directement pour s'en convaincre) :

$$\begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} =: A\mathbf{X}.$$

- b) On trouve :

$$AA^T = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}.$$

Un calcul taupinal donne comme polynôme caractéristique  $P(X) = -Y^3 + 3Y - 2$ , avec  $Y = X - 2$ . Une racine évidente (toujours la même !) est  $Y = 1$ , soit  $X = 3$ , qui est une racine double. La dernière racine est  $X = 0$ , et donc :

$$\text{Sp}(A^T A) = \{3, 3, 0\}.$$

- c) Tout d'abord on remarque que :

$$V = \|(U, T_1, T_2, T_3)^T\|^2 = \|A\mathbf{X}\|^2 = \langle A^T A\mathbf{X}, \mathbf{X} \rangle.$$

Or  $A^T A$  est une matrice symétrique réelle, donc diagonalisable **en base orthonormée** :

$$\exists O \in \mathbf{O}(n), A^T A = ODO^T, \text{ avec } D := \text{diag}(3, 3, 0).$$

Ainsi  $V = \langle DO^T \mathbf{X}, O^T \mathbf{X} \rangle$ . Mais on sait que  $\mathbf{X}$  est un vecteur gaussien standard, donc de loi invariante par rotation. Ainsi

$$U^T \mathbf{X} \stackrel{\mathcal{L}}{=} \mathbf{X}.$$

Par conséquent, la loi de  $V$  est la même que celle de  $\langle D\mathbf{X}, \mathbf{X} \rangle = 3X_1^2 + 3X_2^2$ . Ainsi  $V/3$  suit la loi du Chi-2 à deux degrés de liberté  $\chi^2(2)$  (voir les exercices précédents).

3. On sait que  $U$  et le **vecteur**<sup>16</sup>  $(T_1, T_2, T_3)$  sont indépendants, donc  $U$  est indépendante de toute fonction de ce vecteur, donc de sa norme en particulier. De plus on connaît déjà les lois marginales de  $(U, V)$ . D'où :

$$\mathcal{L}((U, V)) = \mathcal{N}(0, 3) \otimes 3\chi^2(2),$$

où  $3\chi^2(2)$  signifie simplement "la loi d'une variable aléatoire ayant la loi du Chi-2 et multipliée par 3".

**Exercice 7.10** ▷

1. Par indépendance de  $X_1$  et de  $X_2$  (ils forment un vecteur gaussien de matrice de covariance diagonale), on a que :

$$U \hookrightarrow \mathcal{N}(1, 3).$$

2.  $\mathbf{X}$  est un vecteur gaussien, donc  $A\mathbf{X}$  aussi. Il reste à calculer son vecteur espérance et sa matrice de covariance. On trouve grâce au cours :

$$Y \hookrightarrow \mathcal{N}(\mathbf{0}_{\mathbb{R}^2}, \Sigma),$$

avec  $\Sigma = 2A$ .

3. Cette dernière matrice n'est pas inversible, donc ne peut être définie positive. Par le cours,  $Y$  n'admet pas de densité pour la mesure de Lebesgue (son support est  $\text{Im}(\sqrt{A})$ , qui est un hyperplan de  $\mathbb{R}^2$ , donc de mesure de Lebesgue nulle).

**Exercice 7.11** ▷

1. Un sous-vecteur d'un vecteur gaussien est encore gaussien, et ses paramètres se lisent dans les paramètres d'origine directement : le vecteur espérance de  $(X_1, X_2)$  correspond aux deux premières coordonnées de  $\mathbf{m}$ . De même la matrice de covariance associée est la matrice extraite de  $\Sigma$  de coefficients  $((\text{Cov}(X_i, X_j))_{i,j \in [1,2]}$ . Ainsi :

$$(X_1, X_2) \hookrightarrow \mathcal{N}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\right).$$

2. Le vecteur  $(X_1, Y)^T = (X_1, \alpha X_1 + X_2)^T$  est gaussien car ses coordonnées sont des combinaisons linéaires des coordonnées d'un autre vecteur gaussien. Donc il suffit de trouver  $\alpha$  tel que  $\text{Cov}(X_1, Y) = 0$ . Un calcul facile donne

$$\alpha = -\frac{1}{2}.$$

Par linéarité de l'espérance, on a :

$$\mathbb{E}[Y] = -\frac{1}{2}.$$

Enfin, en utilisant l'identité  $\mathbb{V}(\alpha X_1 + X_2) = \alpha^2 \mathbb{V}(X_1) + \mathbb{V}(X_2)$ , on trouve :

$$\mathbb{V}(Y) = \frac{1}{2} + 2 = \frac{5}{2}.$$

---

16. Il n'aurait pas suffi que  $U$  soit indépendante des  $T_i$  séparément !

3. Ce type de calcul est classique et a des applications importantes en régression linéaire. L'astuce qui suit est à retenir au passage. On trouve grâce à la propriété caractéristique de  $\alpha$  :

$$\mathbb{E}[X_2|X_1] = \mathbb{E}[X_2 + \alpha X_1|X_1] - \alpha X_1 = \mathbb{E}[X_2 + \alpha X_1] - \alpha X_1 = -\frac{1}{2} + \frac{1}{2}X_1 = \frac{1}{2}(X_1 - 1).$$

On a fait apparaître  $Y$  en ajoutant-retranchant  $\alpha X_1$ , car ce terme est mesurable par rapport à  $\sigma(X_1)$ , donc "sort" de l'espérance conditionnelle sans problème, et parce que  $Y$  est indépendante de  $X_1$ , donc  $\mathbb{E}[Y|X_1] = \mathbb{E}[Y]$ .

4. Même astuce : on commence par dire pourquoi  $(X_1, Z)$  est un vecteur gaussien (sinon les arguments qui suivent n'ont pas de raison d'être corrects en général). Puis on cherche  $\beta$  tel que  $X_1$  et  $Z$  soient indépendantes, *i.e.* en faisant en sorte que  $\text{Cov}(X_1, Z) = 0$ . On trouve :

$$\beta = \frac{1}{2}.$$

On en déduit  $\mathbb{E}[X_3|X_1] = \mathbb{E}[Z|X_1] - \beta X_1 = \frac{1}{2}X_1$  et :

$$\begin{aligned} \mathbb{E}[X_3^2|X_1] &= \mathbb{E}[(X_3 + \beta X_1 - \beta X_1^2)|X_1] \\ &= \mathbb{E}[(Z - \beta X_1^2)|X_1] \\ &= \mathbb{E}[Z^2|X_1] + \beta^2 \mathbb{E}[X_1^2|X_1] - 2\beta \mathbb{E}[Z X_1|X_1] \\ &= \mathbb{E}[Z^2] + \beta^2 X_1^2 - 2\beta X_1 \mathbb{E}[Z] \\ &= \mathbb{E}[Z^2] + \frac{1}{4} X_1^2 - X_1 \mathbb{E}[Z] \\ &= \left[ \frac{1}{4}(2+1) + (2+1) - 2 \cdot \frac{1}{2} \right] + \frac{1}{4} X_1^2 + \frac{1}{2} X_1 \\ &= \frac{11}{4} + \frac{1}{2} X_1 + \frac{1}{4} X_1^2. \end{aligned}$$

5. Par linéarité de l'espérance conditionnelle, et en utilisant les résultats des questions précédentes, on trouve :

$$\begin{aligned} \mathbb{E}[X_1^2 X_2 + X_3^2 X_1|X_1] &= X_1^2 \mathbb{E}[X_2|X_1] + X_1 \mathbb{E}[X_3^2|X_1] \\ &= \frac{1}{2} X_1^2 (X_1 - 1) + X_1 \left( \frac{11}{4} + \frac{1}{2} X_1 + \frac{1}{4} X_1^2 \right) \\ &= \frac{3}{4} X_1^3 + \frac{11}{4} X_1. \end{aligned}$$

### Exercice 7.13 ▷

1. La décomposition est une conséquence du fait que la somme des  $E_i$  redonne  $E$  puisqu'on a alors par définition :

$$\forall \omega \in \Omega, \exists \mathbf{X}_1(\omega), \dots, \mathbf{X}_p(\omega) \text{ tels que } \mathbf{X}(\omega) = \sum_{i=1}^p \mathbf{X}_i(\omega),$$

et où chaque  $\mathbf{X}_i(\omega)$  appartient à  $E_i$  respectivement. Le fait que ces applications (de  $\Omega$  dans  $E_i$ ) soient bien définies résulte de l'unicité de la décomposition précédente, puisque les  $E_i$  sont en somme directe (sinon on aurait pu avoir plusieurs valeurs possibles pour un même  $\mathbf{X}_i$ ).

2. Remarquons que  $\mathbf{X}_i = \pi_i \mathbf{X}$ , où  $\pi_i$  est le projecteur orthogonal sur  $E_i$ . On trouve donc par la premier exercice que :

$$\mathbf{X}_i \hookrightarrow \mathcal{N}(\pi_i \mathbf{m}, \pi_i).$$

La matrice de covariance est  $\pi_i \pi_i^T = \pi_i \pi_i = \pi_i$ , car un projecteur orthogonal est auto-adjoint. Pour prouver l'indépendance des  $\mathbf{X}_i$ , on se donne une BON  $(\mathbf{e}_{i,1}, \dots, \mathbf{e}_{i,d_i})$  de  $E_i$ . Dans cette base, la  $k$ -ème coordonnée de  $\mathbf{X}_i$  s'écrit :

$$\forall k \in \llbracket 1, d_i \rrbracket, X_{i,k} = \langle \mathbf{X}, \mathbf{e}_{i,k} \rangle \mathbf{e}_{i,k}.$$

Par conséquent, pour  $i' \neq i$  et  $k' \in \llbracket 1, d_{i'} \rrbracket$ , on a :

$$\text{Cov}(\langle \mathbf{X}, \mathbf{e}_{i,k} \rangle \mathbf{e}_{i,k}, \langle \mathbf{X}, \mathbf{e}_{i',k'} \rangle \mathbf{e}_{i',k'}) = \langle \mathbf{X}, \mathbf{e}_{i,k} \rangle \langle \mathbf{X}, \mathbf{e}_{i',k'} \rangle \langle \mathbf{e}_{i,k}, \mathbf{e}_{i',k'} \rangle = 0.$$

En conséquence, comme  $(\mathbf{X}_i, \mathbf{X}_{i'})$  est un vecteur gaussien (il ne faut pas oublier de le dire, sinon l'argument qui suit n'a aucun raison d'être vrai en général!), on en déduit que  $\mathbf{X}_i$  et  $\mathbf{X}_{i'}$  sont indépendantes. Le vecteur  $(\mathbf{X}_1, \dots, \mathbf{X}_p)$  étant lui aussi gaussien, on en déduit le résultat en général.

3. On note  $F = \text{Vect}(\mathbf{1})$ . Par le théorème de Cochran, les projetés orthogonaux  $\pi_F \mathbf{X}$  et  $\pi_{F^\perp} \mathbf{X}$  sont indépendants. Or :

$$\pi_F \mathbf{X} = \langle \mathbf{X}, \mathbf{e}_1 \rangle \mathbf{e}_1 = \overline{X_n}$$

$$\pi_{F^\perp} \mathbf{X} = \mathbf{X} - \pi_F \mathbf{X} = \begin{pmatrix} X_1 - \overline{X_n} \\ \vdots \\ X_n - \overline{X_n} \end{pmatrix}.$$

Le lemme des coalitions permet de conclure à l'indépendance de  $\overline{X_n}$  et de  $\widehat{s}_n^2$ .

### Exercice 7.14 ▷

1. La formulation de l'énoncé est abstraite, mais sa résolution est simple : on sait que  $O\mathbf{X}$  est encore un vecteur gaussien, comme transformation linéaire d'un autre vecteur gaussien (à savoir  $\mathbf{X}$ ). Il reste à calculer son vecteur espérance et sa matrice de covariance. Le premier vaut :

$$O\mathbf{0}_{\mathbb{R}^n} = \mathbf{0}_{\mathbb{R}^n}.$$

Et la matrice de covariance est :

$$\Sigma_{O\mathbf{X}} = O I_n O^T = I_n.$$

D'où l'invariance en loi demandée.

2. a) Par définition, le support de  $Y$  est la sphère unité de  $\mathbb{R}^d$ .  
 b) Il suffit d'écrire :

$$O\mathbf{Y} = \frac{O\mathbf{X}}{\|\mathbf{X}\|_2} = \frac{O\mathbf{X}}{\|O\mathbf{X}\|_2} \stackrel{(\mathcal{L})}{=} \frac{\mathbf{X}}{\|\mathbf{X}\|_2} = \mathbf{Y}.$$

Au passage, faites bien attention à l'endroit où apparaît le signe "égal en loi" ! Toutes les autres égalités sont des égalités presque sûres en revanche.

3. a) Comme  $\|O\mathbf{X}\|_2 = \|\mathbf{X}\|_2$  et

$$\mathbf{1}_{\{\mathbf{Y} \in O^{-1}B\}} = \mathbf{1}_{\{O\mathbf{Y} \in B\}} = \mathbf{1}_{\left\{ \frac{O\mathbf{X}}{\|O\mathbf{X}\|_2} \in B \right\}},$$

et  $O\mathbf{X}$  a même loi que  $\mathbf{X}$ , on trouve bien que  $\mu_\varphi$  est invariante par translation du groupe de rotation  $\mathbf{O}(n)$ .

- b) Grâce à la question précédente, on sait qu'il existe une constante  $C_\varphi$  telle que :

$$\forall B \in \mathcal{B}or(\mathbb{S}^{n-1}), \mathbb{E}\left[\varphi(\|\mathbf{X}\|_2) \mathbf{1}_B\left(\frac{\mathbf{X}}{\|\mathbf{X}\|_2}\right)\right] = C_\varphi \sigma^{n-1}(B).$$

Prenons  $\varphi = 1$ , qui est bien une fonction mesurable bornée (!). On trouve alors que :

$$\nu(B) = \mathbb{P}\left(\frac{\mathbf{X}}{\|\mathbf{X}\|_2} \in B\right) = C_1 \sigma^{n-1}(B).$$

Pour  $B = \mathbb{S}^{n-1}$ , on trouve que  $C_1 = 1$ . Donc  $\nu$ , la loi d'un vecteur gaussien standard renormalisé, est la loi de probabilité uniforme sur l'hypersphère  $\mathbb{S}^{n-1}$ . En particulier :

$$\forall B \in \mathcal{B}or(\mathbb{S}^{n-1}), \mathbb{E}\left[\varphi(\|\mathbf{X}\|_2) \mathbf{1}_B\left(\frac{\mathbf{X}}{\|\mathbf{X}\|_2}\right)\right] = C_\varphi \mathbb{E}\left[\mathbf{1}_B\left(\frac{\mathbf{X}}{\|\mathbf{X}\|_2}\right)\right].$$

En prenant de nouveau  $B = \mathbb{S}^{n-1}$ , on trouve  $C_\varphi = \mathbb{E}[\varphi(\|\mathbf{X}\|_2)]$ , et donc :

$$\forall B \in \mathcal{B}or(\mathbb{S}^{n-1}), \mathbb{E}\left[\varphi(\|\mathbf{X}\|_2) \mathbf{1}_B\left(\frac{\mathbf{X}}{\|\mathbf{X}\|_2}\right)\right] = \mathbb{E}[\varphi(\|\mathbf{X}\|_2)] \mathbb{E}\left[\mathbf{1}_B\left(\frac{\mathbf{X}}{\|\mathbf{X}\|_2}\right)\right].$$

On en déduit l'indépendance entre  $\mathbf{Y}$  et  $\|\mathbf{X}\|_2$  en prenant  $\varphi = \mathbf{1}_A$ , avec  $A \in \mathcal{B}or(\mathbb{R}_+)$ .

### Exercice 8.5 ▷

- 8.5.a) La variable aléatoire  $T_{i+1} - T_i$  représente le nombre de tablettes nécessaires pour obtenir une nouvelle image sachant que l'on en a déjà  $i$ . Soit  $S_i$  l'ensemble des  $i$  images obtenues, à chaque nouvelle tablette, on trouve une image de cet ensemble avec probabilité  $|S_i|/N = i/N$ . Le nombre de tablettes nécessaires pour en obtenir une nouvelle suit donc un loi géométrique (cf. 1.2.3) : on a un succès lorsque l'image obtenue n'est pas dans  $S_i$ , ce qui se produit avec probabilité  $1 - i/N$ . Au final,

$$\mathbb{P}(T_{i+1} - T_i = k) = \left(\frac{i}{N}\right)^{k-1} \left(1 - \frac{i}{N}\right).$$

8.5.b) Notons  $(X_k, k \geq 1)$  le numéro de l'image obtenue dans la tablette  $k$ . Par hypothèse, les v.a.r.  $X_k$  sont indépendantes (et de loi uniforme de sur  $\{1, \dots, N\}$ ). L'événement  $(T_1 - T_0 = i_1, \dots, T_k - T_{k-1} = i_k)$  dépend des variables aléatoires  $X_1, \dots, X_{i_1+\dots+i_k}$ . Conditionnellement à cet événement, l'événement  $(T_{k+1} - T_k = i_{k+1})$  dépend des variables aléatoires  $X_{i_1+\dots+i_{k+1}}$  et suivantes. Ces deux ensembles de variables aléatoires sont disjoints donc en vertu du lemme ??, il y a indépendance conditionnelle :

$$\begin{aligned} \mathbb{P}(T_{k+1} - T_k = i_{k+1} \mid T_1 - T_0 = i_1, \dots, T_k - T_{k-1} = i_k) \\ = \mathbb{P}(T_{k+1} - T_k = i_{k+1})\mathbb{P}(T_1 - T_0 = i_1, \dots, T_k - T_{k-1} = i_k). \end{aligned}$$

Par récurrence, on en déduit alors que

$$\mathbb{P}(T_1 - T_0 = i_1, \dots, T_N - T_{N-1} = i_N) = \prod_{k=0}^{N-1} \mathbb{P}(T_{k+1} - T_k = i_{k+1}).$$

Ce qui signifie que les v.a.r.  $T_0, T_1 - T_0, \dots, T_N - T_{N-1}$  sont indépendantes dans leur ensemble.

8.5.c) On a la relation évidente

$$T_N = (T_N - T_{N-1}) + (T_{N-1} - T_{N-2}) + \dots + (T_1 - T_0) + T_0.$$

Compte-tenu de la linéarité de l'espérance et de l'exercice ?? où l'on a calculé la moyenne et la variance d'une loi géométrique, on a pour

$$\mathbb{E}(T_{k+1} - T_k) = \frac{1}{1 - k/N} = \frac{N}{N - k},$$

d'où

$$\mathbb{E}(T_N) = N + \frac{N}{2} + \frac{N}{3} + \dots + \frac{N}{N-1} = N \sum_{k=1}^{N-1} \frac{1}{k}.$$

Il est alors supposé connu que

$$\sum_{k=1}^{N-1} \frac{1}{k} \sim \ln(N)$$

pour  $N$  tendant vers l'infini. D'où l'équivalent de  $\mathbb{E}(T_N)$  qui est  $N \ln(N)$ . Pour la variance, d'après l'exercice ??,

$$\text{Var}(T_{k+1} - T_k) = \frac{1 - (1 - k/N)}{(1 - k/N)^2} = \frac{kN^2}{N(N - k)^2} = N \frac{k}{(N - k)^2}.$$

Comme les v.a.r. sont indépendantes, la variance de la somme est la somme des variance (cf. proposition 4.10) donc

$$\text{Var}(T_N) = N \sum_{k=1}^{N-1} \frac{k}{(N - k)^2} \leq N^2 \sum_{k=1}^{N-1} \frac{1}{k^2},$$

en majorant  $k$  par  $N$  et en réindexant. Comme la série de terme général  $1/k^2$  converge, on en déduit que  $\text{Var}(T_N) = O(N^2)$ .

8.5.d) Pour pouvoir appliquer l'inégalité de Bienaymé-Tchebychev, il suffit de vérifier que

$$\frac{\text{Var}(T_N)}{\mathbb{E}(T_N)^2} \xrightarrow{N \rightarrow \infty} 0.$$

Or d'après la question précédente,

$$\frac{\text{Var}(T_N)}{\mathbb{E}(T_N)^2} = O\left(\frac{N^2}{N^2 \ln(N)^2}\right) = O\left(\frac{1}{\ln(N)^2}\right) \xrightarrow{N \rightarrow \infty} 0.$$