

# **Aléatoire**

Josselin Garnier, Sylvie Méléard, Nizar Touzi

Département de Mathématiques Appliquées  
Ecole Polytechnique

Mars 2020



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Avant-propos	8
1.2	Phénomènes aléatoires	9
1.3	Deux idées majeures et incontournables	10
1.3.1	La loi des grands nombres	10
1.3.2	Conditionnement et indépendance	10
1.4	Les variables aléatoires	11
1.4.1	Loi d'une variable aléatoire	11
1.4.2	Simulation de variables aléatoires	11
1.5	Historique	12
<b>2</b>	<b>Espace de probabilité</b>	<b>15</b>
2.1	Le langage des probabilités	15
2.1.1	Information issue d'une expérience aléatoire	15
2.1.2	Probabilité - Premières propriétés	18
2.2	Espace fondamental fini ou dénombrable	23
2.2.1	Caractérisation	23
2.2.2	Modèles d'urnes et calcul combinatoire	25
2.3	Conditionnement et indépendance	29
2.3.1	Probabilités conditionnelles	29
2.3.2	Indépendance	32
2.3.3	Le lemme de Borel-Cantelli	35
2.4	Rappels et compléments	36
2.4.1	Rappels sur les ensembles	36
2.4.2	Modélisation ensembliste des événements	37
2.4.3	Les ensembles dénombrables	38
2.4.4	Quelques résultats utiles sur les séries	38
2.5	Exercices sur le chapitre 2	40
<b>3</b>	<b>Espace fini ou dénombrable</b>	<b>43</b>
3.1	Variables aléatoires discrètes	43
3.2	Espérance des v.a. discrètes	45

3.2.1	Définition . . . . .	45
3.2.2	Propriétés de l'espérance des v.a. discrètes . . . . .	47
3.2.3	Variance et écart-type . . . . .	48
3.2.4	Moments d'une variable aléatoire . . . . .	49
3.3	Fonction génératrice d'une v.a. entière . . . . .	50
3.4	Variables aléatoires discrètes usuelles . . . . .	52
3.4.1	Variable aléatoire de Bernoulli . . . . .	52
3.4.2	Variable aléatoire binomiale . . . . .	52
3.4.3	Probabilité de succès et variable aléatoire géométrique . . . . .	54
3.4.4	Variable aléatoire de Poisson . . . . .	55
3.5	Lois conditionnelles et indépendance . . . . .	56
3.5.1	Lois marginales . . . . .	56
3.5.2	Lois conditionnelles . . . . .	57
3.5.3	Espérance conditionnelle . . . . .	57
3.5.4	Variables aléatoires indépendantes . . . . .	59
3.5.5	Somme de v.a. indépendantes . . . . .	61
3.6	Exercices sur le chapitre 3 . . . . .	63
<b>4</b>	<b>Espérance mathématique de variables aléatoires réelles</b> . . . . .	<b>67</b>
4.1	Probabilité uniforme et mesure de Lebesgue . . . . .	67
4.2	Variables aléatoires dans $\mathbb{R}^d$ . . . . .	69
4.3	Espérance des v.a. réelles . . . . .	71
4.3.1	Espérance de v.a. positives . . . . .	71
4.3.2	Espérance de v.a. réelles . . . . .	74
4.4	Variables aléatoires de carré intégrable . . . . .	76
4.4.1	Variance et Covariance . . . . .	76
4.4.2	Approximation linéaire . . . . .	78
4.5	Des inégalités fameuses . . . . .	79
4.5.1	Inégalité de Bienaymé-Chebyshev . . . . .	79
4.5.2	Inégalité de Cauchy-Schwarz . . . . .	80
4.5.3	Inégalité de Jensen . . . . .	81
4.6	Loi d'un vecteur aléatoire . . . . .	82
4.6.1	Propriété de transfert . . . . .	82
4.6.2	Fonction de répartition . . . . .	83
4.6.3	Variables aléatoires indépendantes . . . . .	85
4.7	Premiers éléments de simulation aléatoire . . . . .	88
4.7.1	Génération de la loi uniforme . . . . .	88
4.7.2	Simulation par inversion de la fonction de répartition . . . . .	89
4.8	Exercices sur le chapitre 4 . . . . .	89
<b>5</b>	<b>Variables aléatoires à densité</b> . . . . .	<b>91</b>
5.1	Notion de densité . . . . .	91
5.2	Exemples classiques de v.a. réelles à densité . . . . .	95
5.3	Vecteurs aléatoires à densité . . . . .	102

5.3.1	Densités marginales . . . . .	103
5.3.2	Densités conditionnelles . . . . .	104
5.3.3	Indépendance de variables aléatoires à densité . . . . .	106
5.4	Recherche de densité . . . . .	106
5.5	Simulation de suites de variables aléatoires indépendantes . . . . .	111
5.5.1	Inversion de la fonction de répartition . . . . .	111
5.5.2	Méthode du rejet . . . . .	112
5.6	Exercices sur le chapitre 5 . . . . .	114
<b>6</b>	<b>Convergences et loi des grands nombres</b>	<b>119</b>
6.1	Convergences de v.a. . . . .	120
6.2	La loi des grands nombres . . . . .	125
6.3	Méthode de Monte-Carlo . . . . .	129
6.4	Exercices sur le chapitre 6 . . . . .	130
<b>7</b>	<b>Fonctions caractéristiques et convergence en loi</b>	<b>133</b>
7.1	La fonction caractéristique . . . . .	133
7.1.1	Définition et premières propriétés . . . . .	133
7.1.2	Exemples . . . . .	135
7.1.3	Propriété fondamentale . . . . .	137
7.1.4	Somme de vecteurs aléatoires indépendants . . . . .	139
7.1.5	Fonction caractéristique et moments . . . . .	140
7.2	Vecteurs gaussiens . . . . .	141
7.3	Convergence en loi . . . . .	144
7.4	Le théorème de la limite centrale . . . . .	148
7.5	Exercices sur le chapitre 7 . . . . .	152
<b>8</b>	<b>Statistique : Estimation</b>	<b>155</b>
8.1	Introduction à l'estimation . . . . .	156
8.2	Qualités d'un estimateur . . . . .	158
8.3	Estimateurs empiriques . . . . .	161
8.4	Méthode de substitution . . . . .	166
8.5	Méthode des moments . . . . .	166
8.6	Maximum de vraisemblance . . . . .	168
8.7	Exercices sur le chapitre 8 . . . . .	172
<b>9</b>	<b>Statistique : Intervalle de confiance</b>	<b>175</b>
9.1	Intervalle de confiance et estimation . . . . .	175
9.2	Intervalles exacts pour le modèle gaussien . . . . .	178
9.2.1	Estimation par intervalle de confiance de la moyenne . . . . .	180
9.2.2	Estimation par intervalle de confiance de la variance . . . . .	183
9.3	Résultats asymptotiques . . . . .	186
9.3.1	Intervalles de confiance asymptotiques . . . . .	186
9.3.2	Sondages . . . . .	187

9.3.3	Calcul d'intégrale par la méthode de Monte-Carlo . . . . .	189
9.4	Exercices sur le chapitre 9 . . . . .	191
<b>10</b>	<b>Statistique : Test</b>	<b>193</b>
10.1	Tests et erreurs . . . . .	194
10.2	Modèle gaussien . . . . .	197
10.2.1	Tests pour la moyenne $\mu$ . . . . .	197
10.2.2	Tests pour la variance $\sigma^2$ . . . . .	198
10.3	Test du $\chi^2$ (test du chi-deux) . . . . .	199
10.3.1	Test d'adéquation à une loi . . . . .	199
10.3.2	Test d'adéquation à une famille de lois . . . . .	203
10.3.3	Test d'indépendance . . . . .	206
10.4	Exercices sur le chapitre 10 . . . . .	207
<b>11</b>	<b>Corrections des exercices</b>	<b>209</b>
11.1	Corrigés des exercices du chapitre 2 . . . . .	209
11.2	Corrigés des exercices du chapitre 3 . . . . .	214
11.3	Corrigés des exercices du chapitre 4 . . . . .	219
11.4	Corrigés des exercices du chapitre 5 . . . . .	219
11.5	Corrigés des exercices du chapitre 6 . . . . .	225
11.6	Corrigés des exercices du chapitre 7 . . . . .	228
11.7	Corrigés des exercices du chapitre 8 . . . . .	232
11.8	Corrigés des exercices du chapitre 9 . . . . .	235
11.9	Corrigés des exercices du chapitre 10 . . . . .	240
<b>12</b>	<b>Textes et corrigés d'examens</b>	<b>241</b>
	<b>Bibliographie</b>	<b>261</b>
	<b>Index</b>	<b>263</b>

# Chapitre 1

## Introduction

*A quoi tu penses ?*

*Je pense que,*

*si en ouvrant un dictionnaire au hasard, on tombait sur le mot hasard, ce serait un miracle, alors que si on tombait sur le mot miracle, ce serait un hasard.*

*H. Le Tellier, Les amnésiques n'ont rien vécu d'inoubliable.*

Il peut paraître irréaliste et prétentieux de vouloir, de par sa nature même, **quantifier le hasard**. C'est pourtant ce qui a conduit à la notion de **Probabilité**. Nous allons dans ce livre introduire ce concept mathématique, dont la puissance permettra de **modéliser** d'innombrables situations où le hasard intervient, dépassant ainsi largement le cadre restreint des jeux de dés et tirages de cartes. La modélisation probabiliste est fondamentale dans tous les domaines d'applications, qu'ils soient issus des sciences dures ou des sciences humaines, de la physique (physique quantique, physique des particules), de la climatologie, de la biologie (mutations du génôme), de l'écologie (variabilité des comportements individuels ou variations environnementales), de l'informatique et des réseaux de télécommunications, du traitement du signal et de la parole, de la médecine (imagerie médicale), de l'économie, l'assurance, la finance (marchés boursiers), ou de la sociologie.

## 1.1 Avant-propos

Le mot **Hasard** est un mot d'origine arabe : *az-zahr*, le dé. Il est apparu en français pour signifier tout d'abord un jeu de dés, puis plus généralement un événement non prévisible, et par extension le mode d'apparition de ce type d'événement.

Dans la vie quotidienne, chacun est familier avec le mot et même le concept de probabilité : probabilité qu'il pleuve la semaine suivante, probabilité d'avoir une fille aux yeux bleus, probabilité de gagner au loto ou celle d'être dans la bonne file au supermarché. Les assurances fixent le contrat d'assurance-vie d'un individu de 20 ans, grâce à une estimation de sa probabilité de survie à 80 ans. Dans de nombreux domaines, les probabilités interviennent : les entreprises cherchent à calculer le besoin probable de leurs produits dans le futur, les médecins cherchent à connaître les probabilités de succès de différents protocoles de soin, les compagnies pharmaceutiques doivent estimer les probabilités d'apparitions d'effets secondaires pour leurs médicaments. Un exemple récent et spectaculaire est celui de l'utilisation des probabilités en économie, et en particulier en finance. Nous pouvons citer également d'autres domaines d'applications extrêmement importants et en pleine expansion, aussi variés que le calcul de structures, la théorie du signal, l'optimisation et le contrôle des systèmes, l'imagerie médicale, la génomique et la théorie de l'évolution.

Les probabilités sont en lien étroit avec la vie quotidienne. A ce titre, elles s'appuient sur un passage du concret à l'abstrait : la modélisation mathématique. En effet, la première difficulté face à un problème concret va être de transformer cette réalité physique en un modèle mathématique abstrait qu'il est possible d'étudier et sur lequel des calculs peuvent être menés. Il est alors possible de fabriquer des expérimentations fictives du problème concret sur ordinateur, que l'on appelle des simulations numériques, obtenues à partir du modèle mathématique. Ces simulations sont utilisées, soit à des fins descriptives, soit à des fins numériques.

Pour pouvoir modéliser les innombrables situations, de natures très différentes, où le hasard intervient, un cadre très général d'étude est nécessaire. Ce cadre abstrait a été défini rigoureusement par Andreï Kolmogorov en 1933 (donc très récemment), sous le nom de modèle probabiliste. Sa définition a nécessité préalablement le développement de théories d'analyse importantes telles le calcul intégral et la théorie de la mesure.

C'est ce grand écart entre l'apparente simplicité de certains problèmes probabilistes concrets, et l'abstraction que nécessite leur résolution, qui peut rendre le monde de l'aléatoire difficile ou inquiétant, mais c'est aussi ce qui en fait un domaine mathématique fascinant et palpitant.



## 1.2 Phénomènes aléatoires

Le but de ce cours est d'introduire les notions de base de la théorie des probabilités, et surtout de permettre d'acquérir le raisonnement probabiliste. Cette théorie des probabilités ne peut se construire axiomatiquement qu'en utilisant la théorie de la mesure et de l'intégration, ce qui en constitue une des difficultés principales. Nous n'en donnerons dans ce texte que les éléments nécessaires à sa bonne compréhension, sans exiger de prérequis dans ce domaine. (Mais nous remarquerons que la théorie des probabilités constitue un très bel exemple d'application de la théorie de l'intégration).

L'objet de la théorie des probabilités est l'analyse mathématique de phénomènes dans lesquels le hasard intervient. Ces phénomènes sont appelés des **phénomènes aléatoires**.

**Définition 1.1** Un phénomène est dit aléatoire si, reproduit maintes fois dans des conditions identiques, il se déroule chaque fois différemment de telle sorte que le résultat de l'expérience change d'une fois sur l'autre de manière imprévisible.

Nous pouvons donner des exemples variés de tels phénomènes :

- Jeu de Pile ou Face
- Jeu de lancer de dés

Dans ces deux exemples, la différence entre les résultats, si l'on réitère l'expérience, peut être liée à l'impulsion initiale communiquée au dé, à la rugosité de la table, aux vibrations du plancher... Le hasard est l'illustration de la méconnaissance des conditions initiales, car la pièce ou le dé ont des trajectoires parfaitement définies par la mécanique classique.

- Durée de vie d'une ampoule électrique
- Temps de passage d'un bus
- Nombre de voitures passant une borne de péage
- Promenade d'un ivrogne : un pas en avant, deux pas en arrière...
- Position d'un impact sur une cible, dans un jeu de fléchettes
- Evolution du prix d'un actif financier au cours du temps
- Mutations dans le génôme.

Ces exemples présentent comme point commun des variations liées à la présence de facteurs extérieurs, influant sur le résultat de l'expérience, et que l'on ne sait pas contrôler. De nombreux effets physiques fonctionnent ainsi, et chaque phénomène déterministe est inévitablement accompagné d'écarts aléatoires. Dans certains cas, il est possible de négliger les éléments aléatoires et de remplacer le phénomène réel par un schéma simplifié, en sélectionnant pour ce faire les paramètres les plus importants (comme par exemple en mécanique classique). Mais cette approximation n'est pas toujours possible et il est souvent fondamental de pouvoir quantifier les écarts aléatoires.

Dans d'autres domaines, tels la physique quantique, l'aléatoire fait intrinsèquement partie de la théorie, et certaines mesures ne peuvent être connues qu'aléatoirement dans un ensemble de résultats possibles.

## 1.3 Deux idées majeures et incontournables

Deux idées majeures illustrent la théorie des probabilités et son extrême richesse : la loi des grands nombres et le conditionnement (lié à la notion d'indépendance). Ces deux notions formeront l'ossature de ce cours et méritent d'être assimilées en profondeur.

### 1.3.1 La loi des grands nombres

La notion de hasard, ou d'aléatoire, est souvent liée à la méconnaissance de paramètres intervenant dans une expérience, ou à la trop grande multitude de ceux-ci. Néanmoins, bien que ces comportements aléatoires soient a priori sujets à des variations imprévisibles, nous serons capables de donner des renseignements sur ce type de phénomènes. L'idée majeure est que ces informations seront données par la répétition de l'expérience.

En effet, l'observation d'un grand nombre de répétitions d'un même phénomène aléatoire permet d'y déceler généralement des lois régissant les résultats, tout à fait déterminées, stables. Par exemple, pour toute pièce non truquée d'un jeu de Pile ou Face, et quelque soit l'endroit où se déroule le jeu, 1000 lancers de la pièce donneront environ 50% de piles et 50% de faces. De même, l'étude de la répartition des tailles d'un groupe d'individus, et quel que soit l'échantillon pris dans ce groupe, montre qu'il y aura toujours une courbe des répartitions de même type. Il va être ainsi possible de prévoir la fréquence d'apparition de chaque résultat, la valeur moyenne de ces résultats et les oscillations autour de cette valeur moyenne.

C'est cette stabilité confirmée par l'expérience qui s'appelle **Loi des grands nombres**, et qui légitime l'utilisation d'un modèle mathématique.

### 1.3.2 Conditionnement et indépendance

La construction d'un modèle probabiliste repose sur l'information connue *a priori* sur l'expérience aléatoire. Ce modèle permet de quantifier les probabilités de réalisation de certains résultats de l'expérience. Il est fondamental de remarquer que si l'information change, les probabilités de réalisation changent. Par exemple, la chance de choisir au

hasard un homme de plus de 100 kilos parmi 1000 hommes de la population française est plus grande si le groupe est composé d'hommes de plus de 1,80m que si le groupe est composé d'hommes de moins de 1,65m. La richesse du modèle probabiliste que nous allons construire réside dans le fait que si l'information change par rapport au modèle initial, les nouvelles chances de réalisation pourront être calculées. Ce raisonnement lié à l'information a priori se résume en théorie des Probabilités par le mot **conditionnement**. Quand l'information donnée a priori sur un phénomène aléatoire n'a aucune influence sur la réalisation d'un autre phénomène, par exemple deux tours successifs de roulette dans un casino, ces phénomènes aléatoires sont dits indépendants. Cette hypothèse d'indépendance sera fondamentale dans toute la théorie, et simplifiera de nombreux calculs.

## 1.4 Les variables aléatoires

### 1.4.1 Loi d'une variable aléatoire

Nous allons dans ce livre étudier des fonctions qui dépendent du résultat de l'expérience aléatoire sous-jacente. Elles sont appelées **variables aléatoires**, car leurs valeurs varient en fonction du hasard. Plutôt que de chercher les antécédents de chaque valeur possible de la fonction, nous allons nous intéresser à la chance de réalisation de l'ensemble des antécédents qui permettent à la fonction d'être égale à une de ces valeurs ou d'appartenir à un ensemble de ces valeurs. C'est cela que nous appellerons la loi de la variable aléatoire. Cette notion de loi d'une variable aléatoire est à la base du raisonnement probabiliste moderne.

### 1.4.2 Simulation de variables aléatoires

La simulation consiste en une expérimentation fictive sur machine d'un phénomène modélisé. Elle permet de visualiser une expérience aléatoire, de calculer des quantités numériques et de vérifier certains résultats théoriques.

La méthode de simulation probabiliste la plus célèbre est la méthode de Monte-Carlo, du nom du quartier où se trouve le casino de Monaco. Elle consiste à effectuer certains calculs (calculs d'intégrales notamment) par de nombreuses simulations numériques de réalisations indépendantes de variables aléatoires de loi donnée. Ce procédé est fondé sur la loi des grands nombres qui en assure la convergence. Mais, pour obtenir une précision acceptable, nous verrons qu'il faut accomplir une grande quantité de simulations, ce qui explique que la méthode n'a pu se développer de manière significative que depuis l'introduction d'ordinateurs performants.

L'outil de base est un générateur de nombres au hasard qui simule une variable aléatoire de loi uniforme. La plupart des langages de programmation et des logiciels mathématiques en possèdent un :

- la méthode `Math.random` en JAVA,
- la fonction `rand` sous MATLAB ou SCILAB,
- la bibliothèque `numpy.random` en PYTHON.

Ainsi, par exemple, l'application répétée de la fonction `rand` fournit une suite de nombres indépendants les uns des autres et uniformément répartis sur  $[0, 1]$ . Nous verrons comment, à partir de ce générateur, nous pouvons simuler de nombreux types de loi.

## 1.5 Historique

La notion de modèle abstrait commun à des expériences variées a mis beaucoup de temps à émerger. Le hasard étant par nature pour nos ancêtres une représentation du divin, il a fallu, pour définir la notion de probabilité, attendre une certaine maturité de la pensée. Il y a très peu d'écrits anciens concernant le calcul des probabilités. Au 4ème siècle, l'existence d'une science des jeux de dés apparaît dans le Mahabharata (célèbre ouvrage indien), de même que ses rapports étroits avec une évaluation de type sondage (cf. Hacking). Mais les premières références publiées sur les chances de gagner au jeu datent de Cardan (1501-1576) dans son livre *De Ludo Alea*. Des calculs de probabilité apparaissent aussi dans les œuvres de Kepler (1571-1630) et de Galilée (1564-1642). Le calcul probabiliste se développe au cours du 17ème siècle, motivé en particulier par l'engouement frénétique pour les jeux de hasard à cette époque. Le sujet commence réellement à être rigoureusement développé par Pascal (1623-1662) et Fermat (1601-1665) vers 1654, comme un calcul combinatoire, à partir de paradoxes issus de ces jeux (les paradoxes du Chevalier de Méré que l'on verra au Chapitre 2). Dès 1657, Huyghens (1629-1695) rédige un mémoire amplifiant sensiblement les résultats de Pascal et Fermat, et son travail reste jusqu'à la fin du 17ème siècle l'exposé le plus profond de calcul des Probabilités. Bernoulli (1654-1705) établit la loi des grands nombres sous sa forme la plus simple, résultat fondamental qu'il dit avoir médité vingt ans. Vers la fin du 17ème siècle, une autre impulsion au calcul des probabilités vient d'Angleterre et de Hollande, motivée par des problèmes d'assurance (Halley (1656-1742), De Witt (1625-1672)). En effet, l'évaluation des populations (par exemple : tables de mortalité et rentes viagères) devient une discipline essentielle à la gouvernance moderne des états.

La théorie des probabilités se construit dans la modélisation d'une réalité qui n'est pas forcément (pas souvent) de nature physique. Pascal la croit utilisable en théologie. Le célèbre Pari de Pascal montre que croire en Dieu est une solution statistiquement plus avantageuse, en supposant au préalable que les deux hypothèses d'existence ou non de Dieu ont la même probabilité. Leibniz (1646-1716), et plus tard Laplace (1749-1827),

Poisson (1781-1840) (*Recherches sur la probabilité des jugements en matière criminelle et matière civile*), l'appliquent aux controverses juridiques. Les probabilités sont un outil privilégié de modélisation des comportements humains, comme en témoigne l'intérêt récurrent des philosophes pour leurs fondements.

De Moivre (1667-1754) et Euler (1707-1803) développent les idées de Pascal et Fermat, Bayes (1671-1746) introduit la notion de probabilité conditionnelle (probabilité a priori), mais faute d'outils mathématiques puissants, il faut pour développer plus avant la théorie, attendre Laplace (1749-1827). Celui-ci donne une application magistrale du calcul différentiel et intégral à la théorie des probabilités dans son très important *Traité analytique des probabilités* (en 1812). Laplace formule le postulat du déterminisme universel. Cette intelligence est un idéal, un horizon, que notre science ne nous permet pas d'atteindre. Le calcul des probabilités est imaginé comme un outil permettant de pallier cette faiblesse. Laplace permet à la discipline de dépasser définitivement sa première phase combinatoire. Il met en avant le rôle de la loi normale et démontre une version du théorème de la limite centrale. Gauss (1777-1855) développe intensément la théorie. Dans les pays anglo-saxons se développe également l'outil statistique, avec l'étude des données et l'analyse prédictive à partir de ces données. Le mot "statistique" vient du mot "état", et cette science a été, depuis cette époque, un outil puissant pour les organismes de décisions. Elle se développe en utilisant le support d'un modèle probabiliste.

Le développement des probabilités grâce aux méthodes d'analyse occupe le 19ème siècle et le début du 20ème siècle, fondé en particulier sur les travaux de Borel (1871-1956) et de Lebesgue (1875-1941) sur la théorie de la mesure. Les avancées au 19ème siècle de la physique statistique (Maxwell (1831-1879), Boltzmann (1844-1906)) apportent un nouveau point de vue qui dépasse les idées rationalistes de Laplace et permet d'envisager que le hasard est une réalité objective indépendante de nos connaissances, conformément aux idées du philosophe Cournot (1801-1877) qui le premier affirme que le hasard et le déterminisme sont compatibles entre eux. Le principe d'incertitude d'Heisenberg montrera ultérieurement (1927) l'impossibilité de connaître avec une infinie précision la position et la vitesse d'une particule ; on ne peut les connaître qu'à l'aide d'une loi de probabilité.

Sous l'incitation de problèmes de physique statistique, mais aussi de démographie, commence à se dégager, vers la fin du 19ème siècle, la notion fondamentale de fonction aléatoire, destinée à rendre compte d'un phénomène aléatoire qui évolue au cours du temps. Les probabilités entrent à cette époque dans une nouvelle phase de développement. Dès 1875, Galton (1822-1911) et Watson (1827-1903) étudient l'évolution du nombre d'individus d'une population au cours de ses générations successives, mettant en évidence un exemple de processus aléatoire qui sera introduit dans toute sa généralité par Markov (1856-1922). Einstein (1879-1955) vers 1905 s'intéresse à la notion de mouvement Brownien. Brown avait déjà observé le mouvement d'une particule de pollen sur la surface de l'eau, heurtée de toutes parts par des molécules

d'eau ; ce mouvement paraît totalement désordonné. En fait, Bachelier (1870-1946) avait lui aussi introduit le mouvement brownien en 1900 pour modéliser la dynamique d'un cours boursier. Ce processus aléatoire, évoluant de manière apparemment totalement erratique, s'est avéré être un outil fondamental de modélisation probabiliste dès lors que l'on s'intéresse à un phénomène aléatoire évoluant continûment au cours du temps.

La période moderne, caractérisée par l'étude systématique des processus aléatoires, débute vers 1930. Dans les *Fondements de la Théorie des Probabilités* que Kolmogorov (1903-1987) publie en 1933 apparaît l'axiomatique rigoureuse, fondée sur la théorie de la mesure et de l'intégrale de Lebesgue et qui sera universellement adoptée ensuite. L'expression mathématique donnée ainsi aux concepts confère à ceux-ci une clarté et une maniabilité beaucoup plus grandes, et cette axiomatique s'est révélée indispensable dans l'étude de tous les modèles dynamiques. Après le travail fondamental de Kolmogorov, Lévy (1886-1971), donne le ton pour les probabilités modernes par son travail sur les processus stochastiques, ainsi que sur les fonctions caractéristiques et les théorèmes limites. Mentionnons ici le rôle essentiel joué par les écoles russes et japonaises et notamment par Itô (prix Gauss 2006), qui définit une notion d'intégrale par rapport au mouvement brownien et, grâce à elle, conduit à la création d'un calcul intégral, appelé calcul stochastique, pour certaines familles de processus stochastiques. Ces résultats avaient été, en partie et de manière totalement indépendante, découverts par le mathématicien français Doebelin pendant la deuxième guerre mondiale. Celui-ci sentant sa fin proche (il est mort en 1940 dans les Ardennes) envoya ses trouvailles sous forme d'un « pli cacheté » à l'Académie des Sciences. Ce pli a été découvert et ouvert il y a seulement quelques années et a suscité une grande émotion.

De nos jours, l'Ecole française de Probabilités est très active. La première Médaille Fields décernée à un probabiliste a été attribuée à Wendelin Werner en 2006. Les probabilités se développent de plus en plus, alimentées en particulier de manière essentielle par la physique, le développement des réseaux de télécommunications, la finance, et plus récemment, par la biologie et la médecine. Elles permettent de construire des modèles mathématiques, qui peuvent être validés par les données suivant la théorie statistique, et fournissent également des possibilités d'expérimentations fictives dans de multiples domaines d'applications.

## Chapitre 2

# Espace de probabilité

*On ne peut guère donner une définition satisfaisante de la probabilité. La définition complète de la probabilité est donc une sorte de pétition de principe.*

*Henri Poincaré (1854-1912) - Calcul des Probabilités.*

### 2.1 Le langage des probabilités

#### 2.1.1 Information issue d'une expérience aléatoire

L'ensemble de tous les résultats possibles d'une expérience aléatoire est appelé **espace fondamental** et sera noté  $\Omega$ . Ses éléments, notés classiquement  $\omega \in \Omega$ , sont aussi appelés **événements élémentaires**.

Un **événement** est un sous-ensemble de  $\Omega$  dont on peut dire au vu de l'expérience s'il est réalisé ou non. On notera  $\mathcal{P}(\Omega)$  l'ensemble de toutes les parties de  $\Omega$ .

L'espace fondamental  $\Omega$  peut prendre diverses formes. En voici quelques exemples.

##### Exemple 2.1

1. Deux Lancers d'une pièce à Pile ou Face :  $\Omega = \{PP, PF, FP, FF\}$ . Le sous-ensemble  $\{PP, PF, FP\}$  est l'événement "obtenir au moins un P".
2. Lancer simultané de deux pièces identiques à Pile ou Face :  $\Omega = \{PP, PF, FF\}$ , puisque l'expérience aléatoire ne permet pas l'accès à l'ordre d'obtention de P et F. Le sous-ensemble  $\{PP, PF\}$  est l'événement "obtenir au moins un P".

3. Lancer d'un dé :  $\Omega = \{1, 2, 3, 4, 5, 6\}$ . Le sous-ensemble  $\{2, 4, 6\}$  est l'événement "obtenir un chiffre pair".
4. Nombre de passages de véhicules à une borne de péage pendant une journée :  $\Omega = \mathbb{N}$ . Le sous-ensemble  $\{10^4, \dots, 10^5\}$  est l'événement "le nombre de passages de véhicule à cette borne de péage est compris entre  $10^4$  et  $10^5$ ".
5. Durée de vie d'une ampoule électrique :  $\Omega = [0, +\infty[$ . Le sous-ensemble  $[0, 2500]$  est l'événement "la durée de vie de l'ampoule est inférieure ou égale à 2500 heures".
6. Envoi d'une fléchette sur une cible circulaire de 30 cm de diamètre. L'expérience consiste à décrire le point d'impact de la fléchette dans un repère orthonormé de centre le centre de la cible :  $\Omega = \{(x, y) \in \mathbb{R}^2, \sqrt{x^2 + y^2} \leq 15\}$ . Le sous-ensemble  $\{x^2 + y^2 \leq 2\}$  est l'événement "le point d'impact de la fléchette est à 2 cm du centre de la cible".
7. Temps de passage des véhicules à une borne de péage :  $\Omega = (\mathbb{R}_+)^{\mathbb{N}}$ .
8. Prix d'un actif financier sur un intervalle de temps  $[t_1, t_2]$  :  $\Omega = C([t_1, t_2], \mathbb{R}_+)$ , ensemble des fonctions continues de  $[t_1, t_2]$  dans  $\mathbb{R}_+$ . Le sous-ensemble  $\{\omega \in C([t_1, t_2], \mathbb{R}_+) : \sup_{t \in [t_1, t_2]} \omega(t) \leq \alpha\}$  est l'événement "le prix de l'actif ne dépasse pas le seuil  $\alpha$  pendant l'intervalle de temps  $[t_1, t_2]$ ".
9. Vitesse d'une molécule dans un gaz raréfié sur un intervalle de temps  $[t_1, t_2]$  :  $\Omega = \mathcal{D}([t_1, t_2], \mathbb{R}^3)$ , l'ensemble des fonctions continues à droite et avec limites à gauche de  $[t_1, t_2]$  dans  $\mathbb{R}^3$ .

Ainsi, l'espace fondamental peut être fini (exemples 1 à 3), infini dénombrable (exemple 4), ou infini non dénombrable (exemples 5 à 9). Il peut être dépourvu d'une structure topologique naturelle, comme dans l'exemple 1, ou en avoir une plus ou moins riche. Cela permet de réaliser la richesse de la théorie qu'il faut mettre en place afin d'englober tous ces cas.

Le modèle abstrait que nous allons construire est fondé sur la modélisation de l'information qui est définie mathématiquement par le choix de la famille des événements (sous-ensembles de  $\Omega$ ) qui peuvent être discernés par l'expérience aléatoire. La définition suivante permet de mettre en place cette notion en suivant l'intuition de l'information révélée par une telle expérience aléatoire.

**Définition 2.2** Une classe  $\mathcal{A} \subset \mathcal{P}(\Omega)$  est appelée **tribu** ou  $\sigma$ -algèbre si :

- (A1)  $\Omega \in \mathcal{A}$ .
- (A2)  $\mathcal{A}$  est stable par passage au complémentaire :  $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ .
- (A3)  $\mathcal{A}$  est stable par réunion dénombrable, i.e. si  $(A_n)_{n \in \mathbb{N}}$  est une suite d'éléments de  $\mathcal{A}$ , alors  $\cup_{n \in \mathbb{N}} A_n$  est dans  $\mathcal{A}$ .

Notons que (A1) et (A2) impliquent qu'une tribu  $\mathcal{A}$  contient nécessairement  $\emptyset = \Omega^c$



et que la combinaison de (A2) et (A3) implique que  $\mathcal{A}$  est stable par intersection dénombrable.

Notons également que (A3) n'implique pas que  $\mathcal{A}$  soit stable par réunion ou intersection infinie non dénombrable. Cependant en considérant le cas particulier  $A_0 = A$  et  $A_n = B$  pour  $n \geq 1$  dans (A3), on obtient la condition plus faible suivante.

**Définition 2.3** Une classe  $\mathcal{A} \subset \mathcal{P}(\Omega)$  est dite additive si elle vérifie :

(A3)<sub>f</sub>  $\mathcal{A}$  est stable par union finie : si  $A, B \in \mathcal{A}$ , alors  $A \cup B \in \mathcal{A}$ .

Si  $\Omega$  est fini, alors  $\mathcal{A} \subset \mathcal{P}(\Omega)$  est finie, et (A3) est équivalente à (A3)<sub>f</sub>.

**Remarque fondamentale :** Dans la modélisation de notre phénomène aléatoire, la tribu représente un ensemble de parties de  $\Omega$  (parties composées de certains résultats de l'expérience) dont on va pouvoir **mesurer la chance de réalisation**. C'est pour un élément  $A$  de cette tribu que nous allons être capable de définir sa probabilité de réalisation  $\mathbb{P}(A)$ , tandis que  $\mathbb{P}(A)$  n'aura pas de sens dès lors que  $A$  n'appartient pas à la tribu  $\mathcal{A}$ .

**Exemple 2.4** (i)  $\mathcal{A} = \{\emptyset, \Omega\}$  est la tribu grossière, ou triviale : c'est la plus petite tribu de  $\Omega$ .

(ii) L'ensemble  $\mathcal{P}(\Omega)$  des parties de  $\Omega$  est une tribu sur  $\Omega$ . C'est celle que nous choisissons systématiquement si  $\Omega$  est un ensemble fini ou dénombrable. Cependant, pour des raisons fondamentales que nous indiquerons ultérieurement, cette tribu sera trop grande dès que  $\Omega$  est infini non dénombrable pour que l'on puisse définir la probabilité de tous ses éléments de manière non triviale.

En dehors des cas très simples, il est souvent impossible de lister les éléments d'une algèbre ou d'une  $\sigma$ -algèbre. Il est alors commode de les caractériser par des sous-ensembles "assez riches". D'après la définition d'une tribu, on voit que toute intersection (même infinie non dénombrable) de tribus est une tribu. Comme  $\mathcal{P}(\Omega)$  est une tribu, on peut définir la notion suivante.

**Définition 2.5** Soit  $C \subset \mathcal{P}(\Omega)$ . La tribu engendrée par  $C$  est la plus petite tribu contenant  $C$  et est donnée par

$$\sigma(C) = \bigcap_{\mathcal{A} \text{ tribu, } C \subset \mathcal{A} \subset \mathcal{P}(\Omega)} \mathcal{A}.$$

**Exemple 2.6** (i) La tribu engendrée par un ensemble  $A \subset \Omega$  est  $\{\emptyset, A, A^c, \Omega\}$ .

(ii) Si  $(A_i)_{i \in I}$  est une partition finie ou dénombrable de  $\Omega$  (i.e. les  $A_i$  sont deux-à-deux disjoints et leur réunion est  $\Omega$ ), la tribu engendrée par  $\{A_i, i \in I\}$  est l'ensemble des réunions  $B_J = \cup_{i \in J} A_i$ , où  $J$  décrit la classe de toutes les parties de  $I$ .

**Définition 2.7** Si  $\Omega$  est un espace topologique<sup>1</sup>, on appelle **tribu borélienne** la tribu engendrée ses ouverts.

À titre d'exemple, la tribu borélienne de  $\mathbb{R}$  est la tribu engendrée par la classe des intervalles ouverts de  $\mathbb{R}$ . Le résultat suivant montre qu'on peut se restreindre aux intervalles (ouverts ou fermés) non bornés à gauche (ou à droite). La démonstration est un très bon exercice de maniement des tribus :

**Proposition 2.8** *La tribu borélienne de  $\mathbb{R}$  est la tribu engendrée par les intervalles de la forme  $] - \infty, a]$  pour  $a \in \mathbb{Q}$ .*

**Preuve.** Rappelons que toute tribu est stable par passage au complémentaire, par réunion ou intersection dénombrable. Puisque  $] - \infty, a]$  est le complémentaire de l'intervalle ouvert  $]a, +\infty[$ , ce dernier appartient à la tribu borélienne, et donc la tribu  $\mathcal{C}$  engendrée par ces intervalles est incluse dans la tribu borélienne. Réciproquement, soit  $]x, y[$  un intervalle ouvert de  $\mathbb{R}$ . Soit  $(x_n)_n$  une suite de rationnels décroissant vers  $x$  et  $(y_n)_n$  une suite de rationnels croissant strictement vers  $y$ . On a :

$$]x, y[ = \bigcup_n (]-\infty, y_n] \cap ]-\infty, x_n]^c).$$

Nous en déduisons que tout intervalle ouvert appartient à  $\mathcal{C}$ , d'où le résultat.  $\square$

## 2.1.2 Probabilité - Premières propriétés

Nous cherchons à définir, pour un événement  $A \in \mathcal{A}$ , la vraisemblance accordée a priori à  $A$  (avant le résultat de l'expérience). Nous voulons donc associer à chaque événement  $A$  un nombre  $\mathbb{P}(A)$  compris entre 0 et 1, qui représente la chance que cet événement soit réalisé à la suite de l'expérience.

Pour justifier notre définition d'une probabilité, nous allons faire appel à notre intuition et discuter la signification usuelle de ce qu'est la probabilité d'un événement. Considérons un événement  $A$  pouvant se produire lors d'une certaine expérience aléatoire (par exemple,  $A = \ll \text{obtenir Pile} \gg$  lors du lancer d'une pièce). Supposons que l'on puisse répéter un grand nombre  $n$  de fois cette expérience aléatoire. Notons  $n(A)$  le nombre de fois où l'événement  $A$  se produit. La fréquence de réalisation de  $A$ ,

$$f_n(A) = \frac{n(A)}{n},$$

---

1. Un espace topologique est muni d'une famille d'ouverts contenant l'ensemble vide, stable par union quelconque, et stable par intersection finie.

est elle-même aléatoire. Mais notre expérience courante tend à nous faire penser que, lorsque le nombre  $n$  de répétitions de l'expérience augmente,  $f_n(A)$  se stabilise autour d'une valeur limite déterministe (on peut penser que  $f_n(A)$  tend vers  $1/2$  dans le cas où  $A$  est l'événement « obtenir Pile » lors du lancer d'une pièce non biaisée). Cette limite est notre idée intuitive de la probabilité  $\mathbb{P}(A)$  de l'événement  $A$ . L'approche intuitive et naturelle consiste donc à définir  $\mathbb{P}(A)$  comme étant la limite quand  $n$  tend vers l'infini des fréquences de réalisation  $f_n(A)$  :

$$\mathbb{P}(A) = \text{limite de } f_n(A) \text{ quand } n \uparrow +\infty. \quad (2.1)$$

Nous donnerons ultérieurement une justification et un sens précis à cette limite, grâce à **la loi des grands nombres**, qui est un des théorèmes fondamentaux de la théorie, justifiant toute la construction mathématique. Des propriétés évidentes vérifiées par les fréquences de réalisation sont les suivantes :

$$f_n(A) \in [0, 1] ; f_n(\Omega) = 1 ;$$

$$\text{Si } A \text{ et } B \text{ sont disjoints, alors } f_n(A \cup B) = f_n(A) + f_n(B) ;$$

Ceci motive la définition générale 2.9 d'une probabilité. On dira qu'une suite d'événement  $(A_n)_n$  sont 2à2 disjoints si  $A_i \cap A_j = \emptyset$  pour tous  $i \neq j$

**Définition 2.9** Soit  $\mathcal{A}$  une tribu sur  $\Omega$ . Une probabilité sur  $(\Omega, \mathcal{A})$  est une application

$$\mathbb{P} : \mathcal{A} \longrightarrow [0, 1]$$

vérifiant les deux propriétés suivantes :

- Masse totale unitaire :  $\mathbb{P}(\Omega) = 1$ , (2.2)

- $\sigma$ -additivité :  $\mathbb{P}(\cup_n A_n) = \sum_n \mathbb{P}(A_n)$ , pour tout  $(A_n)_n \subset \mathcal{A}$ , 2à2 disjoints. (2.3)

Remarque : La probabilité  $\mathbb{P}$  (dite aussi mesure de probabilité) est une mesure abstraite de masse 1 sur l'espace mesurable  $(\Omega, \mathcal{A})$ . Le cadre dans lequel nous travaillons est mathématiquement développé par la théorie de la mesure, mais nous n'invoquerons aucun résultat général de cette théorie.

La propriété de  $\sigma$ -additivité (2.3) est plus forte que la propriété

- additivité :  $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$ , pour tous  $A_1, A_2 \in \mathcal{A}$  disjoints. (2.4)

Pour le voir, nous commençons par appliquer (2.3) avec  $A_n = \emptyset$  pour tout  $n \in \mathbb{N}$  : si  $a = \mathbb{P}(\emptyset)$ , nous obtenons  $\sum_n a = a$ , ce qui entraîne  $a = 0$ . Ensuite, si  $A, B \in \mathcal{A}$  sont disjoints, nous appliquons (2.3) avec  $A_0 = A, A_1 = B$  et  $A_n = \emptyset$  pour tout  $n \geq 2$ , ce qui donne  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) + \sum_{n \geq 2} \mathbb{P}(\emptyset) = \mathbb{P}(A) + \mathbb{P}(B)$ , d'où (2.4).

Notons cependant que les propriétés de  $\sigma$ -additivité (2.3) et d'additivité (2.4) sont équivalentes dans le cadre d'un espace fondamental  $\Omega$  fini.

**Corollaire 2.10** Une probabilité vérifie de plus :

$$\bullet \quad \mathbb{P}(\emptyset) = 0 \quad (2.5)$$

$$\bullet \quad \mathbb{P}(A) + \mathbb{P}(A^c) = 1, \quad (2.6)$$

$$\bullet \quad \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i), \text{ si les } A_i \text{ sont 2\`a2\`a disjoints,} \quad (2.7)$$

$$\bullet \quad \mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B), \quad (2.8)$$

$$\bullet \quad \mathbb{P}(A) \leq \mathbb{P}(B) \quad \text{si } A \subset B. \quad (2.9)$$

**Preuve.** La propriété (2.5) se montre en appliquant (2.4) avec  $A = B = \emptyset$ , et (2.6) s'obtient de la même façon avec  $A$  et  $A^c$ . Pour prouver (2.8), nous décomposons l'ensemble  $A$  en l'union des deux ensembles disjoints  $A \cap B$  et son complémentaire  $A \setminus B = A \cap B^c$ , et de même pour  $B$ , comme cela est représenté dans la figure 2.1. L'inégalité (2.9) se déduit de (2.4) avec  $B = A \cup (B \setminus A)$ .  $\square$

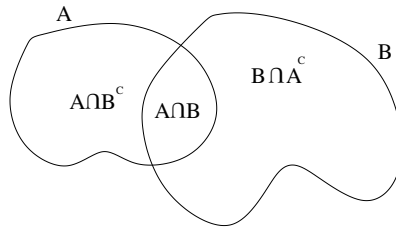


FIGURE 2.1 –  $A$ ,  $B$ ,  $A \cap B^c$ ,  $A \cap B$ ,  $B \cap A^c$  et  $A \cup B$

**Remarque 2.11 (Pourquoi l'additivité ne suffit-elle pas?)** Nous allons nous en rendre compte à travers d'un jeu à Pile ou Face. Si nous jouons  $n$  fois, l'espace  $\Omega$  naturel est l'ensemble  $\{P, F\}^n$  (ensemble des mots de  $n$  lettres avec un alphabet à deux lettres  $P$  et  $F$ ). C'est un ensemble fini de cardinal  $2^n$ . Si la pièce n'est pas truquée, les probabilités de chaque tirage sont égales et nous nous retrouvons dans le cadre de la combinatoire (voir exemple 2.18). Ainsi,

$$\mathbb{P}_n(A) = \frac{\text{card}(A)}{2^n} \quad \text{pour tout } A \subset \Omega.$$

Supposons maintenant que le jeu se poursuive indéfiniment. L'espace fondamental devient  $\Omega = \{P, F\}^{\mathbb{N}^*}$ , c'est-à-dire l'ensemble des mots de longueur infinie, avec le même alphabet à deux lettres  $P$  et  $F$ . C'est un ensemble infini. Essayons d'évaluer la probabilité  $\mathbb{P}(A)$  de l'événement  $A =$  "on ne tire jamais Pile". Soit  $A_n =$  "on ne tire aucun Pile lors des  $n$  premiers tirages". D'après l'expression de  $\mathbb{P}_n(A)$  ci-dessus, nous avons  $\mathbb{P}(A_n) = \mathbb{P}_n(A_n) = 2^{-n}$ . Remarquons que  $A$  est la limite

naturelle des ensembles  $A_n$ , au sens où les  $A_n$  sont décroissants (i.e.  $A_{n+1} \subset A_n$ ) et où  $A = \bigcap_n A_n$ . Il est alors naturel d'écrire que

$$\mathbb{P}(A) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 0,$$

ce qui ne peut pas être justifié par la propriété d'additivité (2.4). La proposition 2.12 dit que la  $\sigma$ -additivité permet exactement de justifier ce passage à la limite.

Nous utilisons les notations suivantes. Une suite  $(A_n)_n \subset \mathcal{P}(\Omega)$  est dite décroissante si  $A_{n+1} \subset A_n$  pour tout  $n$ . Pour une telle suite, on définit  $A = \bigcap_n A_n$ , et on écrit  $A_n \downarrow A$  ou  $\lim_n \downarrow A_n = A$ . De même, la suite  $(A_n)_n$  est croissante si  $A_n \subset A_{n+1}$  pour tout  $n$ , on définit  $A = \bigcup_n A_n$  et on écrit  $A_n \uparrow A$  ou  $\lim_n \uparrow A_n = A$ .

**Proposition 2.12** Soit  $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$  une application vérifiant (2.2) et (2.4). Alors, il y a équivalence entre :

- (i) La propriété de  $\sigma$ -additivité (2.3).
- (ii) Pour toute suite  $(A_n)_n$  **croissante**, on a  $\mathbb{P}(\lim_n \uparrow A_n) = \lim_n \uparrow \mathbb{P}(A_n)$ .
- (iii) Pour toute suite  $(A_n)_n$  **décroissante**, on a  $\mathbb{P}(\lim_n \downarrow A_n) = \lim_n \downarrow \mathbb{P}(A_n)$ .

En particulier que si  $(A_n)_n$  est une suite croissante ou décroissante d'événements, la suite  $(\mathbb{P}(A_n))_n$  admet une limite quand  $n$  tend vers l'infini.

**Preuve.** Etant donné (2.6), on a (ii)  $\Leftrightarrow$  (iii). Montrons que (i)  $\Leftrightarrow$  (ii).

Supposons d'abord (ii). Considérons une suite  $(A_n)_n$  d'éléments de  $\mathcal{A}$  deux-à-deux disjoints, et posons  $B_n = \bigcup_{p \leq n} A_p$  et  $B = \bigcup_n A_n$ . Comme  $\mathbb{P}$  vérifie (2.4), elle vérifie (2.7) et  $\mathbb{P}(B_n) = \sum_{p \leq n} \mathbb{P}(A_p)$  croît vers  $\sum_n \mathbb{P}(A_n)$  et aussi vers  $\mathbb{P}(B)$  par (ii). Nous avons donc (i).

Supposons maintenant (i). Soit  $A_n \in \mathcal{A}$  pour  $n \geq 0$ , avec  $A_n \uparrow A$ . Soit aussi  $B_0 = A_0$ , et définissons par récurrence  $B_n = A_n \setminus B_{n-1}$ , pour  $n \geq 1$ . Comme  $\bigcup_n B_n = A$  et comme les  $B_n$  sont deux-à-deux disjoints, nous avons

$$\mathbb{P}(A) = \sum_n \mathbb{P}(B_n) = \lim_n \sum_{p=0}^n \mathbb{P}(B_p) = \lim_n \mathbb{P}(A_n),$$

la dernière égalité provenant de (2.7). Nous obtenons donc le résultat.  $\square$

La propriété (2.3) donne la probabilité de la réunion  $\bigcup_n A_n$  en fonction des probabilités  $\mathbb{P}(A_n)$ , lorsque les événements  $A_n$  sont deux-à-deux disjoints. Si ce n'est pas le cas, nous avons tout de même la majoration suivante, très utile en pratique :

**Proposition 2.13** Soit  $\mathbb{P}$  une probabilité, et soit  $(A_n)_{n \in I}$  une famille finie ou dénombrable d'événements. On a alors

$$\mathbb{P}\left(\bigcup_{n \in I} A_n\right) \leq \sum_{n \in I} \mathbb{P}(A_n).$$

**Preuve.** a) Supposons d’abord l’ensemble  $I$  fini, et montrons par récurrence que

$$\mathbf{P}_k : \mathbb{P}(A_1 \cup \dots \cup A_k) \leq \mathbb{P}(A_1) + \dots + \mathbb{P}(A_k), \quad \text{pour tout } k \in \mathbb{N}. \quad (2.10)$$

Cette propriété est évidente pour  $k = 1$ . Supposons la propriété vraie pour  $k - 1$ , avec  $k \geq 2$ , et posons  $B = A_1 \cup \dots \cup A_{k-1}$  et  $C = B \cup A_k$ . En vertu de (2.8), nous avons  $\mathbb{P}(C) + \mathbb{P}(B \cap A_k) = \mathbb{P}(B) + \mathbb{P}(A_k)$ , donc  $\mathbb{P}(C) \leq \mathbb{P}(B) + \mathbb{P}(A_k)$ , et nous en déduisons immédiatement  $\mathbf{P}_k$ .

b) Considérons maintenant le cas où  $I$  est dénombrable. Nous pouvons supposer sans restriction que  $I = \mathbb{N}^*$ . Posons  $B_n = \bigcup_{i=1}^n A_i$ , qui croît vers l’ensemble  $C = \bigcup_{n \in I} A_n$ . D’après (a), nous avons  $\mathbb{P}(B_n) \leq \sum_{i=1}^n \mathbb{P}(A_i)$ . Mais le membre de gauche ci-dessus croît vers  $\mathbb{P}(C)$  en vertu de la proposition précédente, tandis que le membre de droite croît vers  $\sum_{n \in I} \mathbb{P}(A_n) \in [0, +\infty]$ . En passant à la limite, nous obtenons donc le résultat voulu.  $\square$

Nous avons pu ainsi construire dans toute sa généralité un espace abstrait  $\Omega$  (non-vide), une tribu sur cet espace et définir la notion de probabilité sur cette tribu. C’est cette idée géniale qu’a introduite Kolmogorov en 1933 : avoir mis au cœur des probabilités un objet nouveau, une mesure de probabilité, et ne pas s’intéresser aux causes de l’expérience génériquement représentées par  $\omega \in \Omega$ .

**Définition 2.14** On appelle le triplet  $(\Omega, \mathcal{A}, \mathbb{P})$  un **espace de probabilité**. C’est un espace mesuré au sens de la théorie de la mesure.

**La modélisation probabiliste consiste donc à décrire une expérience aléatoire par la donnée d’un espace de probabilité.**

Une question fondamentale va ainsi être de décrire et caractériser les mesures de probabilité définies pour des espaces de probabilité de plus en plus gros :  $\mathbb{N}, \mathbb{Q}, \mathbb{R}, \mathbb{R}^n, C([t_1, t_2], \mathbb{R})$ . Le dernier exemple, qui traite de trajectoires aléatoires, ne pourra pas être abordé dans ce cours de base.

Remarquons que l’on peut construire de nombreuses probabilités distinctes sur le même espace  $(\Omega, \mathcal{A})$ . Nous verrons beaucoup d’exemples ultérieurement, mais nous pouvons nous en convaincre rapidement dans le cadre du jeu de Pile ou Face : suivant que la pièce est truquée ou non truquée, la probabilité de faire Pile est  $1/2$  ou  $p \in [0, 1]$ .

La définition suivante introduit une notion de “vrai ou faux” qui dépend de la probabilité choisie sur l’espace fondamental.

**Définition 2.15** Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité.

- (i) Un événement de probabilité nulle est dit négligeable.
- (ii) Une propriété est vraie  $\mathbb{P}$ -presque-sûrement (en abrégé  $\mathbb{P}$ -p.s.), si l’ensemble des  $\omega \in \Omega$

pour lesquels elle est vraie est de probabilité égale à 1, ou en d'autres termes, l'ensemble des  $\omega$  pour lesquels la propriété est fautive est négligeable.

## 2.2 Espace fondamental fini ou dénombrable

### 2.2.1 Caractérisation

Si  $\Omega$  est au plus dénombrable, les singletons  $\{\omega\}$ ,  $\omega \in \Omega$ , forment une partition de  $\Omega$ , nous avons alors la caractérisation suivante des probabilités sur  $\Omega$ .

**Proposition 2.16** *Supposons que l'espace fondamental  $\Omega$  est au plus dénombrable.*

(i) *Pour une probabilité  $\mathbb{P}$  sur  $\Omega$ , on a*

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) \quad \text{pour tout } A \in \mathcal{P}(\Omega),$$

*et  $\mathbb{P}$  est ainsi caractérisée par ses valeurs sur les singletons  $\{p_\omega = \mathbb{P}(\{\omega\}), \omega \in \Omega\}$ .*

(ii) *Soit  $(p_\omega)_{\omega \in \Omega} \in \mathbb{R}^\Omega$ . Alors il existe une unique probabilité  $\mathbb{P}$  telle pour tout  $\omega \in \Omega$ ,  $p_\omega = \mathbb{P}(\{\omega\})$ , si et seulement si*

$$0 \leq p_\omega \leq 1 \quad \text{et} \quad \sum_{\omega \in \Omega} p_\omega = 1.$$

**Preuve.** Commençons par le cas  $\Omega$  fini. (i) Pour une probabilité  $\mathbb{P}$  sur  $\Omega$ , le résultat découle de (2.7) du fait que tout  $A \in \mathcal{P}(\Omega)$  est réunion disjointe (et finie) des singletons  $\{\omega\}$ , pour les  $\omega \in A$ .

(ii) La condition nécessaire découle de (i) appliqué à  $A = \Omega$  et de  $\mathbb{P}(\Omega) = 1$ . Inversement, considérons  $n$  nombres  $(p_i)_{1 \leq i \leq n}$  vérifiant la condition du (ii). Nous posons  $\mathbb{P}(\{\omega_i\}) = p_i$  et pour tout  $A \subset \Omega$ , nous définissons  $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$  et nous vérifions immédiatement que  $\mathbb{P}$  est une probabilité au sens de la définition 2.9.

Si  $\Omega$  est infini dénombrable, on suit le même argument, si ce n'est que pour prouver que  $\mathbb{P}$  définie par  $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$  vérifie (2.3), il faut utiliser la propriété de sommation par paquets pour les séries. (Voir Section 2.4.4 ci-après.)  $\square$

**Exemple 2.17 (Loi de Bernoulli de paramètre  $p \in [0, 1]$ )** *L'espace  $\Omega$  est :*

$$\Omega = \{\omega_1, \omega_2\} \quad \text{et} \quad p_{\omega_1} = p ; p_{\omega_2} = 1 - p.$$

Cette probabilité modélise en particulier la chance pour une pièce de tomber sur Pile dans un jeu de Pile ou Face. Dans ce cas,  $\Omega = \{P, F\}$  peut être assimilé à  $\{0, 1\}$ .

Si la pièce est équilibrée, alors  $p = 1/2$ . Mais cette probabilité peut aussi modéliser la probabilité de réalisation d'un des résultats, pour toute expérience aléatoire avec deux résultats possibles (mon premier enfant sera-t-il une fille ou un garçon?).

**Exemple 2.18 (Loi uniforme.)** Nous dirons que la probabilité  $\mathbb{P}$  sur l'espace fini  $\Omega$  est uniforme si  $p_\omega = \mathbb{P}(\{\omega\})$  ne dépend pas de  $\omega$ . Nous avons donc :

$$p_{\omega_i} = \frac{1}{n} \quad \text{pour tout } i = 1, \dots, n,$$

où  $n = \text{card}(\Omega)$  désigne le cardinal de  $\Omega$ , c'est-à-dire son nombre d'éléments. Nous déduisons de la proposition 2.16(i) que

$$\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}, \quad (2.11)$$

de sorte que le calcul des probabilités se ramène à des dénombrements : nous sommes dans le cadre du **calcul combinatoire**.

Remarquons que sur un espace fini donné  $\Omega$ , il existe une et une seule probabilité uniforme. Cette probabilité décrit mathématiquement l'expression intuitive de "au hasard" (tirage au hasard d'une carte, lancer au hasard d'un dé, choix au hasard d'un échantillon dans une population).

**Exemple 2.19 (Loi de Poisson de paramètre  $\theta$ ).** On définit

$$p_n = e^{-\theta} \frac{\theta^n}{n!} \quad \text{pour tout } n \in \mathbb{N}.$$

On a  $p_n \in [0, 1]$  et  $\sum_n p_n = e^{-\theta} \sum_n \frac{\theta^n}{n!} = 1$ . La suite  $(p_n)_n$  définit une probabilité sur  $\mathbb{N}$ . Cette loi fut introduite par Siméon Denis Poisson, dans son ouvrage "Recherches sur la probabilité des jugements en matière criminelle et en matière civile" (1837).

**Exemple 2.20 (Mesure de Dirac - Loi discrète)**

1) Considérons un espace mesurable  $(\Omega, \mathcal{A})$  arbitraire (avec  $\Omega$  non-vide) et un point  $\omega_0$  fixé dans  $\Omega$ . On suppose que  $\{\omega_0\} \in \mathcal{A}$ . Alors la mesure de Dirac en  $\omega_0$  est :

$$\delta_{\omega_0}(A) = 1 \text{ si } \omega_0 \in A ; \delta_{\omega_0}(A) = 0 \text{ si } \omega_0 \notin A \text{ pour tout } A \in \mathcal{A}. \quad (2.12)$$

Une propriété est alors vraie  $\delta_{\omega_0}$ -presque-sûrement si elle est satisfaite par  $\omega_0$ , et l'ensemble  $\Omega \setminus \{\omega_0\}$  est un ensemble  $\delta_{\omega_0}$ -négligeable.

2) Soit  $\mathbb{P}$  une probabilité définie sur  $\Omega = \{\omega_n, n \in \mathbb{N}\}$ , et  $p_n = \mathbb{P}(\{\omega_n\})$ . Alors

$$\mathbb{P} = \sum_{n \in \mathbb{N}} p_n \delta_{\omega_n} \quad \text{et} \quad \mathbb{P}(A) = \sum_{n \in \mathbb{N}} p_n \delta_{\omega_n}(A) \quad \text{pour tout } A \in \mathcal{A}.$$



### 2.2.2 Modèles d’urnes et calcul combinatoire

Dans les calculs de probabilités uniformes sur des ensembles finis, il est fondamental de faire très attention à bien préciser l’espace de probabilité sous-jacent. Cette remarque prend toute son ampleur dans ce paragraphe, où nous allons développer différents “modèles d’urnes” que l’on peut également voir comme des modèles de prélèvement d’échantillons dans une population au cours d’un sondage. Ces modèles interviennent aussi en contrôle de fabrication, ou dans de multiples autres situations. Si le lecteur n’est pas inspiré par les couleurs des boules d’une urne, il pourra transcrire l’analyse suivante dans le cadre des opinions politiques dans la population française ou celui du niveau de perfection d’un objet dans une chaîne de fabrication.

Le modèle général est le suivant : une urne contient  $N$  boules de  $k$  couleurs différentes, réparties en  $N_1$  boules de couleur 1,  $N_2$  boules de couleur 2,  $\dots$ ,  $N_k$  boules de couleur  $k$ . Nous appelons  $p_i = N_i/N$  la proportion de boules de couleur  $i$  dans l’urne. Tirons au hasard  $n$  boules de cette urne,  $2 \leq n \leq N$ , et intéressons-nous à deux événements :  
 -  $A$  = “la couleur des deux premières boules tirées est 1”,  
 -  $B$  = “on obtient  $n_1$  boules de couleur 1,  $n_2$  boules de couleur 2,  $\dots$ ,  $n_k$  boules de couleur  $k$ ”, avec  $n_i$  des entiers tels que  $n_1 + n_2 + \dots + n_k = n$ .  
 On remarque que l’événement  $A$  dépend de l’ordre de tirages, mais pas l’événement  $B$ .

Nous allons considérer trois façons de tirer les boules au hasard : tirage avec remise, tirage sans remise, tirage simultané. Nous verrons que chaque tirage donnera lieu à un calcul de probabilité et à des résultats différents.

Le problème du choix du tirage de l’échantillon se pose sans cesse dès que l’on souhaite récolter des données statistiques.

Pour  $k$  et  $n$  deux entiers tels que  $k \leq n$ , nous allons souvent utiliser, dans la suite, le nombre de parties  $\binom{n}{k}$  à  $k$  éléments dans un ensemble à  $n$  éléments, qui vaut :

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

**Tirage exhaustif ou simultané - La loi hypergéométrique.** Nous tirons toutes les boules en même temps. L’ensemble  $\Omega$  est alors l’ensemble de toutes les parties possibles de  $n$  éléments distincts, et le nombre de résultats possibles est  $\binom{N}{n}$ . Dans ce cadre,  $A$  n’est pas un événement, car on ne peut pas dire si  $A$  est réalisé en connaissant seulement le résultat du tirage simultané, donc on ne peut donc pas évaluer sa probabilité.

$B$  est un événement et le nombre de cas favorables donnant la bonne répartition des couleurs pour l’événement  $B$  est alors égal à  $\binom{N_1}{n_1} \cdots \binom{N_k}{n_k}$ . La probabilité de  $B$

recherchée vaut donc

$$P_B^{(1)} = \frac{\binom{N_1}{n_1} \cdots \binom{N_k}{n_k}}{\binom{N}{n}}. \quad (2.13)$$

Cette distribution s'appelle la distribution polygénométrique. Dans le cas de deux couleurs, on obtient  $P_B^{(1)} = \hat{P}_{n_1, n-n_1}$  avec

$$\hat{P}_{n_1, n-n_1} = \frac{\binom{N_1}{n_1} \binom{N-N_1}{n-n_1}}{\binom{N}{n}},$$

qui est appelée distribution (ou loi) hypergénométrique.

**Exemple 2.21** Dans une fabrication en série, nous savons que parmi  $N$  pièces usinées,  $M$  sont à mettre au rebut, et si nous choisissons au hasard et simultanément un échantillon de  $n$  pièces, la probabilité pour que cet échantillon contienne  $k$  pièces défectueuses sera  $\frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$ .

**Tirage avec remise - La loi binomiale.** Les tirages sont successifs. Nous remplaçons la boule tirée dans l'urne avant le tirage suivant. Nous pouvons donc tirer plusieurs fois la même boule. L'ensemble  $\Omega$  est alors l'ensemble de tous les  $n$ -uplets d'éléments de l'urne. Toutes les répétitions étant possibles,  $\text{card}(\Omega) = N^n$ . Nous munissons  $\Omega$  de sa probabilité uniforme.

La probabilité de l'événement  $A$  est facile à calculer, c'est la probabilité de tirer une des  $N_1$  boules de couleur 1 parmi les  $N$  boules de l'urne deux fois de suite, donc  $A$  a pour probabilité  $P_A^{(2)} = (N_1/N)^2 = p_1^2$ .

Pour évaluer la probabilité de l'événement  $B$ , on commence par dire que le nombre de façons de déterminer les places des  $k$  couleurs parmi  $n$  est égal au nombre de façons de partager  $n$  en  $k$  parties de tailles  $n_i$ , à savoir  $\frac{n!}{n_1!n_2! \cdots n_k!}$ . Une fois la place des couleurs choisie, nous avons  $N_i$  possibilités pour chaque boule de couleur  $i$ . Le nombre de  $n$ -uplets de répartition  $n_1, n_2, \dots, n_k$  est alors égal à  $\frac{n!}{n_1!n_2! \cdots n_k!} N_1^{n_1} \cdots N_k^{n_k}$ . Nous avons donc finalement que la probabilité de l'événement  $B$  est :

$$P_B^{(2)} = \frac{n!}{n_1!n_2! \cdots n_k!} \frac{N_1^{n_1} \cdots N_k^{n_k}}{N^n}. \quad (2.14)$$

Cette probabilité est appelée une distribution multinomiale. Dans le cas particulier où  $k = 2$ ,  $p_1 = p = N_1/N$  et  $p_2 = 1 - p$ , on a  $P_B^{(2)} = P_{n_1, n-n_1}$  avec

$$P_{n_1, n-n_1} = \binom{n}{n_1} p^{n_1} (1-p)^{n-n_1} \quad (2.15)$$

La probabilité définie par  $P_{n_1, n-n_1}$  est appelée loi binomiale de paramètres  $n$  et  $p$ . (Elle fait intervenir les coefficients du binôme, d'où son nom). Attention, les paramètres ne

sont pas interchangeables :  $n \in \mathbb{N}$  et  $p \in [0, 1]$ . Notons également que cette probabilité est définie sur l'espace  $\Omega = \{0, 1, 2, \dots, n\}$  comportant  $n + 1$  éléments.

**Remarque 2.22** On rappelle la formule du binôme  $(a + b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}$  pour tous  $a, b \in \mathbb{R}$ . En prenant  $a = p$  et  $b = 1 - p$ , on a alors  $\sum_{i=0}^n \binom{n}{i} p^i (1 - p)^{n-i} = 1$ .

**Tirage sans remise.** Nous tirons maintenant successivement les boules de l'urne, mais sans les replacer dans l'urne après tirage. L'ensemble  $\Omega$  est alors l'ensemble des suites de  $n$  éléments distincts parmi  $N$  et le nombre de cas possibles sera  $N(N - 1) \cdots (N - n + 1) = A_N^n$ . Nous munissons  $\Omega$  de sa probabilité uniforme.

La probabilité de l'événement  $A$  est la probabilité de tirer une des  $N_1$  boules de couleur 1 parmi les  $N$  de l'urne lors du premier tirage, puis de tirer une des  $N_1 - 1$  boules de couleur 1 parmi  $N - 1$  boules restantes de l'urne, donc  $A$  a pour probabilité  $P_A^{(3)} = \frac{N_1}{N} \frac{N_1 - 1}{N - 1}$ .

En raisonnant comme dans le cas avec remise pour évaluer la probabilité de  $B$ , nous pouvons montrer que le nombre de cas favorables vaut  $\frac{n!}{n_1! n_2! \cdots n_k!} A_{N_1}^{n_1} \cdots A_{N_k}^{n_k}$ , ce qui finalement donne pour probabilité la même que celle du cas de tirage simultané :  $P_B^{(3)} = P_B^{(1)}$ .

Ainsi, il y a équivalence du tirage sans remise et du tirage simultané pour les événements qui ne dépendent pas de l'ordre des individus dans le tirage.

**Remarque 2.23** Cas d'une urne dont le nombre de boules est infini. Nous nous plaçons dans les hypothèses du tirage simultané, avec 2 couleurs, en supposant que  $N$  et  $N_1$  tendent vers l'infini de telle manière que  $\frac{N_1}{N}$  converge vers  $p \in ]0, 1[$ . Il est facile de montrer qu'alors,

$$\hat{P}_{n_1, n - n_1} = \frac{\binom{N_1}{n_1} \binom{N - N_1}{n - n_1}}{\binom{N}{n}} \text{ converge vers } P_{n_1, n - n_1} = \binom{n}{n_1} p^{n_1} (1 - p)^{n - n_1}.$$

Ce résultat est intuitivement évident car si le nombre de boules devient infini, les tirages de boules avec ou sans remise deviennent presque équivalents : on a une chance très infime de tomber deux fois sur la même boule.

Dans la dernière remarque, nous avons obtenu la loi binomiale comme limite de lois hypergéométriques. Cette convergence d'une suite de probabilités vers une probabilité donnée sera développée au chapitre 7.3 (convergence en loi).

**EXERCICE 2.1** Les yeux bandés, vous manipulez 7 fiches où sont écrites les lettres E, E, T, B, R, L, I. Quelle est la probabilité que vous écriviez le mot

LIBERTE

*Solution :*  $\frac{2}{7!} = \frac{1}{2520}$ .

**EXERCICE 2.2** On tire au hasard 4 cartes d'un jeu de 52 cartes. Quelle est la probabilité pour que, parmi ces 4 cartes, il y ait exactement 2 rois ?

*Solution :* L'hypothèse *au hasard* amène à modéliser cette expérience comme un tirage uniforme dans un certain ensemble  $\Omega$  qu'il faut préciser. Ici, on prend pour  $\Omega$  la classe des parties à 4 éléments de l'ensemble de 52 cartes. Le cardinal de  $\Omega$  est  $\binom{52}{4}$  et  $\mathbb{P}$  est la probabilité uniforme sur  $\Omega$ . Les résultats favorables sont les tirages qui contiennent exactement 2 rois, à savoir 2 rois et 2 cartes parmi les 48 cartes autres que des rois. Ainsi, la probabilité cherchée vaut  $\frac{\binom{4}{2}\binom{48}{2}}{\binom{52}{4}}$ .

**EXERCICE 2.3** On lance trois dés parfaitement équilibrés. Montrer que la probabilité que la somme des points amenés dépasse dix strictement est égale à la probabilité que cette somme ne dépasse pas dix. (d'où un jeu parfaitement équitable...)

*Solution :* L'ensemble  $\Omega$  est ici l'ensemble des suites  $(a_1, a_2, a_3)$  de 3 nombres compris entre 1 et 6, muni de la probabilité  $\mathbb{P}$  uniforme. Remarquons que

$$a_1 + a_2 + a_3 > 10 \Leftrightarrow (7 - a_1) + (7 - a_2) + (7 - a_3) \leq 10.$$

Ainsi, si  $A$  désigne l'événement "la somme des points obtenus est strictement supérieure à 10", nous remarquons que l'application  $(a_1, a_2, a_3) \mapsto (7 - a_1, 7 - a_2, 7 - a_3)$  est une bijection de  $A$  sur  $A^c$ . Les événements  $A$  et  $A^c$  ont donc même cardinal, et donc même probabilité de réalisation.

**EXERCICE 2.4** (*Problème du chevalier de Méré*) Ce personnage marquant de la cour de Louis XIV qui "avait très bon esprit, mais n'était pas très bon géomètre" (cf. lettre de Pascal à Fermat du 29 juillet 1654) était un joueur impénitent, toujours à la recherche de règles cachées lui permettant d'avoir un avantage sur ses adversaires. Voici deux de ses règles.

- (1) Il est avantageux de parier sur l'apparition d'au moins un 6 en lançant un dé 4 fois de suite.
- (2) Il est avantageux de parier sur l'apparition d'au moins un double-six en lançant deux dés 24 fois de suite.

Calculer les probabilités des événements ci-dessus et juger l'opportunité de ces règles.

*Solution :* La première règle est bonne puisque la probabilité de l'événement vaut

$$1 - \left(\frac{5}{6}\right)^4 \simeq 0,5177 > \frac{1}{2}.$$

La différence avec  $\frac{1}{2}$  est faible, mais apte à fournir à long terme des gains assurés : le chevalier devait jouer souvent selon la règle 1...

La deuxième règle est mauvaise, puisque la probabilité de l'événement cherché vaut :

$$1 - \left(\frac{35}{36}\right)^{24} \simeq 0,4914 < \frac{1}{2}.$$

Le Chevalier était donc moins heureux avec cette règle qu'avec la précédente. En fait, il s'était laissé abuser par un soi-disant argument d'homothétie : en lançant un dé, il y a 6 résultats possibles, en lançant deux dés, il y en a  $6^2 = 36$ , soit 6 fois plus. Comme il est avantageux de parier sur l'apparition d'au moins un 6 en lançant un dé 4 fois de suite, il doit être avantageux de parier sur l'apparition d'un double-six en lançant deux dés  $4 \times 6 = 24$  fois de suite. Paradoxe !

## 2.3 Conditionnement et indépendance

### 2.3.1 Probabilités conditionnelles

La notion de conditionnement est l'une des plus fructueuses de la théorie des probabilités, et la différencie fondamentalement de la théorie de la mesure. L'idée de base permettant la compréhension de cette notion est la suivante : **une information supplémentaire concernant l'expérience modifie la vraisemblance que l'on accorde à l'événement étudié.**

Par exemple (et il serait bien de méditer sur cet exemple très simple), cherchons, pour un lancer de deux dés, la probabilité de l'événement "la somme est supérieure ou égale à 10". Elle vaut  $\frac{1}{6}$  sans information supplémentaire,  $\frac{1}{2}$  si l'on sait que le résultat d'un des dés est 6, 0 si l'on sait a priori que le résultat d'un des dés est 2. Pour obtenir ces résultats, nous avons dans chaque cas calculé le rapport du nombre de résultats favorables sur le nombre de cas possibles. Nous remarquons qu'il est indispensable de bien définir l'espace de probabilité lié à l'expérience munie de l'information a priori. Remarquons également que l'information a priori a changé la valeur de la probabilité de l'événement.

L'approche intuitive pour formaliser cette notion consiste à revenir à la notion de fréquence empirique. La fréquence de réalisation de l'événement  $A$  sachant que l'événement  $B$  est réalisé, sur  $n$  expériences, est égale au nombre de réalisations de  $A$  parmi celles pour lesquelles  $B$  est réalisé. Elle vaut donc

$$\frac{n_{A \cap B}}{n_B} = \frac{f_n(A \cap B)}{f_n(B)},$$

et en faisant tendre  $n$  vers l'infini, nous aboutissons à la définition suivante.

**Définition 2.24** Soit  $A$  et  $B$  deux événements, avec  $\mathbb{P}(B) > 0$ . **La probabilité conditionnelle de  $A$  sachant  $B$**  est le nombre

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (2.16)$$

**Proposition 2.25** (i) Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité et  $B \in \mathcal{A}$  de probabilité  $\mathbb{P}(B) > 0$ . Alors l'application de  $A \in \mathcal{A} \mapsto \mathbb{P}(A|B) \in [0, 1]$  définit une nouvelle probabilité sur  $\Omega$ , appelée probabilité conditionnelle sachant  $B$ .

(ii) Si  $\mathbb{P}(A) > 0$  et  $\mathbb{P}(B) > 0$ , nous avons  $\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A)$ .

**Preuve.** Il est clair que  $0 \leq \mathbb{P}(A|B) \leq 1$ . Par ailleurs, les propriétés (2.2) et (2.3) pour  $\mathbb{P}(\cdot|B)$  proviennent des mêmes propriétés pour  $\mathbb{P}$  et des remarques suivantes :  $\Omega \cap B = B$ , et  $(\cup_n A_n) \cap B = \cup_n (A_n \cap B)$ . De plus, si  $A$  et  $C$  sont disjoints, il en est de même de  $A \cap B$  et  $C \cap B$ . L'assertion (ii) est évidente.  $\square$

**Proposition 2.26 Formule des probabilités composées.** Si  $A_1, \dots, A_n$  sont des événements de  $\Omega$  tels que  $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$ , alors

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1}). \quad (2.17)$$

**Preuve.** On procède par récurrence. Si  $n = 2$ , les formules (2.16) et (2.17) sont les mêmes. Supposons que (2.17) soit vraie pour  $n - 1$ , et soit  $B = A_1 \cap \dots \cap A_{n-1}$ . D'après (2.16), on a  $\mathbb{P}(B \cap A_n) = \mathbb{P}(B)\mathbb{P}(A_n|B)$ . En remplaçant  $\mathbb{P}(B)$  par sa valeur dans (2.17) avec  $n - 1$ , nous obtenons (2.17) pour  $n$ .  $\square$

**Proposition 2.27 Formule des probabilités totales.** Soit  $(B_i)_{i \in I}$  une partition finie ou dénombrable d'événements de  $\Omega$ , telle que  $\mathbb{P}(B_i) > 0$  pour chaque  $i \in I$ . Pour tout  $A \in \mathcal{A}$ , on a alors

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap B_i) = \sum_{i \in I} \mathbb{P}(A|B_i) \mathbb{P}(B_i).$$

**Preuve.** On a  $A = \cup_{i \in I} (A \cap B_i)$ , avec les ensembles  $(A \cap B_i)$  deux-à-deux disjoints. Comme par ailleurs  $\mathbb{P}(A \cap B_i) = \mathbb{P}(A|B_i)\mathbb{P}(B_i)$ , il suffit d'appliquer (2.3).  $\square$

**Théorème 2.28 Formule de Bayes.** Sous les mêmes hypothèses que la proposition 2.27, et si  $\mathbb{P}(A) > 0$ ,

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j \in I} \mathbb{P}(A|B_j)\mathbb{P}(B_j)} \quad \text{pour tout } i \in I.$$

**Preuve.** Le dénominateur de de la dernière expression vaut  $\mathbb{P}(A)$  d’après la proposition 2.27, tandis que (2.16) implique

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

□

**EXERCICE 2.5** Un individu est tiré au hasard dans une population où l’on trouve une proportion  $10^{-4}$  de séropositifs. On lui fait passer un test de détection de la séropositivité. Par ailleurs, des expérimentations antérieures ont permis de savoir que la probabilité d’avoir un résultat positif lors de l’application du test sur un individu séropositif est égale à 0,99 (c’est la sensibilité du test); la probabilité d’avoir un résultat positif lors de l’application du test sur un individu séronégatif est 0,001 ( $0,999 = 1 - 0,001$  est la spécificité du test). Sachant que le test donne un résultat positif, quelle est la probabilité pour que l’individu soit effectivement séropositif?

*Solution :* Considérons les événements  $A$  “l’individu est séropositif”, et  $B$  “le test de détection donne un résultat positif”. Les données fournissent  $\mathbb{P}(A) = 10^{-4}$  d’où  $\mathbb{P}(A^c) = 0,9999$ ,  $\mathbb{P}(B|A) = 0,99$  et  $\mathbb{P}(B|A^c) = 0,001$ . Nous trouvons alors

$$\begin{aligned} \mathbb{P}(A|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A) \mathbb{P}(A)}{\mathbb{P}(B|A) \mathbb{P}(A) + \mathbb{P}(B|A^c) \mathbb{P}(A^c)} \\ &= \frac{0,99 \times 10^{-4}}{0,99 \times 10^{-4} + 0,001 \times 0,9999} \simeq 0,09. \end{aligned}$$

Remarquons que contrairement à l’intuition, cette probabilité est petite.

**EXERCICE 2.6** On classe les gérants de portefeuilles en deux catégories, les bien informés et les autres. Lorsqu’un gérant bien informé achète une valeur boursière pour son client, on peut montrer par une étude préalable que la probabilité que le cours de cette valeur monte est de 0,8. Si le gérant est mal informé, la probabilité que le cours descende est de 0,6. On sait par ailleurs que si l’on choisit au hasard un gérant de portefeuille, il y a une chance sur 10 que celui-ci soit un gérant bien informé. Un client choisit au hasard un gérant dans l’annuaire, et lui demande d’acheter une valeur. Sachant que le cours de cette valeur est monté, cherchons la probabilité pour que le gérant soit mal informé.

*Solution :* Notons  $M$  l’événement “la valeur monte” et  $I$  l’événement “le gérant est bien informé”. Par la formule des probabilités totales, la probabilité que la valeur monte vaut

$$\mathbb{P}(M) = \mathbb{P}(M|I) \mathbb{P}(I) + \mathbb{P}(M|I^c) \mathbb{P}(I^c) = 0,8 \times 0,1 + 0,4 \times 0,9 = 0,44.$$

La formule de Bayes donne alors

$$\mathbb{P}(I^c|M) = \frac{\mathbb{P}(M|I^c)\mathbb{P}(I^c)}{\mathbb{P}(M)} = \frac{0,4 \times 0,9}{0,44} = 0,818.$$

### 2.3.2 Indépendance

La notion d'indépendance est absolument fondamentale en probabilités et nous verrons par la suite toutes ses implications dans la modélisation de l'aléatoire.

#### Événements indépendants

Intuitivement, deux événements  $A$  et  $B$  sont indépendants si le fait de savoir que  $A$  est réalisé ne donne aucune information sur la réalisation de  $B$  et réciproquement.

Supposons que la réalisation de l'événement  $B$  n'ait aucune influence sur la réalisation de  $A$ . Alors, après  $n$  expériences, la fréquence empirique de réalisation de  $A$  sera approximativement la même, que l'on sache ou non que  $B$  est réalisé. Ainsi donc,  $f_n(A|B) = \frac{f_n(A \cap B)}{f_n(B)}$  doit être approximativement égal à  $f_n(A)$  (Le conditionnement ne change pas l'information que l'on a sur l'expérience). Par passage à la limite sur le nombre d'expériences, nous en déduisons les définitions suivantes.

Si  $B$  est un événement de probabilité strictement positive,  $A$  sera dit **indépendant** de  $B$  si  $\mathbb{P}(A|B) = \mathbb{P}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ . On remarque que cette formule se symétrise et la notion d'indépendance se définit finalement comme suit.

**Définition 2.29** Deux événements  $A$  et  $B$  sont **indépendants** si et seulement si

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

La probabilité de voir  $A$  réalisé ne dépend pas de la réalisation de  $B$ , et réciproquement.

**Remarque 2.30** 1) Cette notion est liée au choix de la probabilité  $\mathbb{P}$  et n'est pas une notion ensembliste. Cela n'a en particulier rien à voir avec le fait que  $A$  et  $B$  soient disjoints ou non. (Cf. Exemple 2.31 ci-dessous).

2) Si  $\mathbb{P}(A) > 0$  et  $\mathbb{P}(B) > 0$ , alors

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

**Exemple 2.31** 1. On lance 3 fois un dé. Si  $A_i$  est un événement qui ne dépend que du  $i$ ème lancer, alors  $A_1, A_2, A_3$  sont indépendants.



2. On tire une carte au hasard dans un jeu de 52 cartes. On considère les événements  $A = \{\text{la carte est une dame}\}$ ;  $B = \{\text{la carte est un cœur}\}$ . Il est facile de voir que  $\mathbb{P}(A) = \frac{4}{52}$ ,  $\mathbb{P}(B) = \frac{13}{52}$ , et  $\mathbb{P}(A \cap B) = \mathbb{P}(\text{la carte est la dame de cœur}) = \frac{1}{52} = \mathbb{P}(A)\mathbb{P}(B)$ . Ainsi, les événements  $A$  et  $B$  sont indépendants pour la probabilité uniforme  $\mathbb{P}$ .
3. Supposons maintenant que le jeu de cartes soit trafiqué. Soit  $\mathbb{Q}$  la nouvelle probabilité correspondant au tirage de cartes. Supposons que

$$\mathbb{Q}(\text{valet de pique}) = \frac{1}{2}, \quad \mathbb{Q}(\text{autre carte}) = \frac{1}{2} \times \frac{1}{51} = \frac{1}{102}.$$

Alors

$$\mathbb{Q}(A \cap B) = \frac{1}{102} \neq \mathbb{Q}(A)\mathbb{Q}(B) = \frac{2}{51} \times \frac{13}{102}.$$

Les événements  $A$  et  $B$  ne sont pas indépendants sous la probabilité  $\mathbb{Q}$ .

Nous laissons en exercice (très simple à vérifier) la démonstration de la proposition suivante, dont le résultat est tout-à-fait intuitif.

**Proposition 2.32** *Si les événements  $A$  et  $B$  sont indépendants, alors il en est de même de  $A$  et  $B^c$ ,  $A^c$  et  $B$ ,  $A^c$  et  $B^c$ .*

La notion d'indépendance se généralise à une suite finie ou infinie d'événements de la manière suivante.

**Définition 2.33** Une suite  $(A_n)_n$  d'événements de  $(\Omega, \mathcal{A}, P)$  est dite indépendante si pour toute suite finie  $(i_1, \dots, i_k)$  d'entiers deux-à-deux distincts :

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_k}).$$

Cette définition est délicate. Par exemple, pour que la suite  $(A, B, C)$  soit indépendante, la propriété doit être vérifiée pour toutes les intersections de deux événements et l'intersection des trois événements. Il ne suffit pas d'avoir  $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ . Par exemple, prenons un lancer de dé avec  $A = \{1, 2, 3\}$ ,  $B = \{2, 4, 6\}$  et  $C = \{1, 2, 4, 5\}$ . Nous avons  $\mathbb{P}(A) = \frac{1}{2}$ ,  $\mathbb{P}(B) = \frac{1}{2}$ ,  $\mathbb{P}(C) = \frac{2}{3}$ . Ainsi, nous avons bien  $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$  mais  $\mathbb{P}(A \cap B) \neq \mathbb{P}(A)\mathbb{P}(B)$ .

Il ne suffit pas non plus que les événements soient indépendants deux-à-deux. On joue 2 fois à Pile ou Face et on considère les événements  $A = \{\text{Face au premier lancer}\}$ ,  $B = \{\text{Face au deuxième lancer}\}$  et  $C = \{\text{les deux tirages donnent le même résultat}\}$ . On vérifie que ces événements sont deux-à-deux indépendants mais que la probabilité de leur intersection n'est pas égale au produit des probabilités.

### Expériences aléatoires indépendantes et espace de probabilité produit

Soit  $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$  une suite d'espaces de probabilité modélisant des expériences aléatoires. Nous souhaiterions construire un espace de probabilité rendant compte de toutes ces expériences indépendantes les unes des autres.

Si nous avons uniquement deux espaces  $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$  et  $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$ , nous prendrons  $\Omega = \Omega_1 \times \Omega_2$ , que nous munirons de la tribu produit  $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ . Cette tribu produit de  $\mathcal{A}_1$  et  $\mathcal{A}_2$  est définie comme étant la tribu engendrée par les pavés  $A_1 \times A_2$ ,  $A_1 \in \mathcal{A}_1$ ,  $A_2 \in \mathcal{A}_2$ , (voir définition 2.5). Nous définissons  $\mathbb{P}$  sur les pavés de  $\mathcal{A}$  par  $\mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)$ . On peut montrer que cela suffit pour caractériser une probabilité sur  $\mathcal{A}$ , que l'on appelle probabilité produit, notée  $\mathbb{P}_1 \otimes \mathbb{P}_2$ .

Nous pouvons généraliser cette notion d'espace de probabilité produit, et considérer le produit (dénombrable) cartésien  $\Omega = \prod_n \Omega_n$ ,  $\mathcal{A} = \otimes_n \mathcal{A}_n$  où  $\otimes_n \mathcal{A}_n$  désigne la plus petite tribu de  $\Omega$  engendrée par les produits cartésiens finis d'éléments des tribus coordonnées, donc contenant tous les ensembles de la forme

$$A_1 \times A_2 \times \cdots \times A_k \times \Omega_{k+1} \times \Omega_{k+2} \times \cdots, \quad A_i \in \mathcal{A}_i, \quad k = 1, 2, 3, \dots$$

Il est possible de montrer par un théorème général de théorie de la mesure qu'il existe une unique probabilité  $\mathbb{P}$  sur  $(\Omega, \mathcal{A})$  qui vérifie

$$\mathbb{P}(A_1 \times A_2 \times \cdots \times A_k \times \Omega_{k+1} \times \Omega_{k+2} \times \cdots) = \prod_{i=1}^k \mathbb{P}_i(A_i)$$

pour tous  $k = 1, 2, \dots$  et  $A_i \in \mathcal{A}_i$ . Cette probabilité rend ainsi indépendantes les expériences aléatoires correspondant à chaque espace  $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ .

En particulier, en prenant tous les espaces coordonnées  $(\Omega_j, \mathcal{A}_j)$  égaux, cela nous permettra de modéliser la même expérience répétée une infinité (dénombrable) de fois, de manière indépendante et dans les mêmes conditions.

**Exemple 2.34** *Considérons les lancers successifs et indépendants d'une même pièce de monnaie, telle que la probabilité de tirer Pile soit égale à  $p \in ]0, 1[$ . Soient  $F_n$  l'événement "Face au  $n$ -ième lancer" et  $P_n$  l'événement "Pile au  $n$ -ième lancer". Soient  $T_k$  l'événement "le premier Pile est obtenu au  $k$ -ième lancer" et  $A$  l'événement "on n'obtient jamais de Pile". Alors, par indépendance des lancers, nous avons*

$$\begin{aligned} \mathbb{P}(T_k) &= \mathbb{P}(F_1 \cap F_2 \cap \cdots \cap F_{k-1} \cap P_k) \\ &= \mathbb{P}(\text{Face})^{k-1} \mathbb{P}(\text{Pile}) = (1-p)^{k-1} p \quad \text{pour tout } k \in \mathbb{N}^*. \end{aligned}$$

Remarquons que  $\{T_k, k \geq 1, A\}$  est une partition (dénombrable) de  $\Omega$ , donc

$$\mathbb{P}(A) = 1 - \sum_{k \geq 1} \mathbb{P}(T = k) = 1 - 1 = 0.$$

### 2.3.3 Le lemme de Borel-Cantelli

Nous allons maintenant voir un théorème fameux, dans lequel intervient fondamentalement la notion d'indépendance, et qui sera très important, en particulier pour les théorèmes de convergence (cf. Chapitre 6).

Pour une suite  $(A_n)_n$  d'événements de  $\mathcal{A}$ , on définit :

$$\limsup_n A_n = \bigcap_p \bigcup_{n \geq p} A_n.$$

Comme  $\mathcal{A}$  est stable par union et intersection dénombrables, on a  $\limsup_n A_n \in \mathcal{A}$ . Remarquons que

$$\begin{aligned} \omega \in \limsup_n A_n &\Leftrightarrow \forall p, \exists n \geq p, \text{ tel que } \omega \in A_n \\ &\Leftrightarrow \omega \text{ appartient à une infinité de } A_n, \\ \text{et que } \omega \notin \limsup_n A_n &\Leftrightarrow \omega \text{ appartient à au plus un nombre fini de } A_n. \end{aligned}$$

**Théorème 2.35** Soit  $(A_n)_n$  une suite d'événements de  $\mathcal{A}$ .

(i) Si la série  $\sum_n \mathbb{P}(A_n) < +\infty$ , alors  $\mathbb{P}(\limsup_n A_n) = 0$ , c'est-à-dire que  $\mathbb{P}$ -presque sûrement, il y a au plus un nombre fini de  $A_n$  qui sont réalisés.

(ii) Si de plus la suite  $(A_n)_n$  est indépendante, alors

$$\sum_n \mathbb{P}(A_n) = +\infty \implies \mathbb{P}(\limsup_n A_n) = 1.$$

Dans ce cas,  $\mathbb{P}$ -presque sûrement, une infinité de  $A_n$  sont réalisés.

Il est clair que cette dernière propriété n'est plus vraie dans le cas où la suite n'est pas indépendante. Il suffit pour s'en convaincre de prendre tous les  $A_n$  égaux à un même événement  $A$  de probabilité  $\mathbb{P}(A) \in ]0, 1[$ .

La première partie de ce lemme est un outil précieux pour démontrer qu'une propriété est vraie  $\mathbb{P}$ -presque sûrement. Nous en verrons un exemple dans la preuve de la loi forte des grands nombres donnée dans la section 6.2. La deuxième partie du lemme caractérise entièrement, dans le cas indépendant, le fait que  $\mathbb{P}(\limsup_n A_n)$  vaut 0 ou 1 suivant la convergence ou la divergence de la série de terme général  $\mathbb{P}(A_n)$ .

**Preuve.** Remarquons tout d'abord que

$$\mathbb{P}(\limsup_n A_n) = \lim_p \downarrow \mathbb{P}(\bigcup_{n \geq p} A_n) \leq \lim_p \downarrow \sum_{n \geq p} \mathbb{P}(A_n),$$

où  $\lim \downarrow$  désigne la limite d'une suite décroissante. Si la série  $\sum_n \mathbb{P}(A_n)$  est convergente, le reste de cette série tend vers 0 et la dernière inégalité implique que  $\mathbb{P}(\limsup_n A_n) = 0$ .

Supposons maintenant que les  $A_n$  soient indépendants et que la série  $\sum_n \mathbb{P}(A_n)$  diverge. Soit  $m$  un nombre entier. Nous avons

$$\begin{aligned} \mathbb{P}(\cup_{i=p}^m A_i) &= 1 - \mathbb{P}(\cap_{i=p}^m A_i^c) = 1 - \prod_{i=p}^m \mathbb{P}(A_i^c) \quad \text{grâce à l'indépendance} \\ &= 1 - \prod_{i=p}^m (1 - \mathbb{P}(A_i)) \geq 1 - e^{-\sum_{i=p}^m \mathbb{P}(A_i)} \end{aligned}$$

grâce à l'inégalité  $1 - x \leq e^{-x}$  pour  $x \geq 0$ . Ainsi,

$$\mathbb{P}(\cup_{i=p}^\infty A_i) \geq 1 - e^{-\sum_{i=p}^\infty \mathbb{P}(A_i)} = 1$$

et l'on conclut finalement que pour tout  $p$ ,  $\mathbb{P}(\cup_{i=p}^\infty A_i) = 1$ , ce qui implique finalement que  $\mathbb{P}(\limsup_n A_n) = 1$ .  $\square$

**Application :** Considérons une suite de parties indépendantes de Pile ou Face, la probabilité d'apparition d'un Pile étant égale à  $p \in ]0, 1[$ . Soit  $A$  un "mot" de longueur  $l$  choisi *a priori*, c'est-à-dire une suite de  $l$  lettres prises dans l'alphabet  $\{P, F\}$ . Désignons par  $A_1$  l'événement consistant en le fait que le mot se réalise dans les  $l$  premières parties, par  $A_2$  l'événement consistant en le fait que le mot se réalise dans les  $l$  parties suivantes, etc. Les événements  $A_1, A_2, \dots$ , sont indépendants et pour tout  $n \geq 1$ , nous avons  $\mathbb{P}(A_n) = \mathbb{P}(A_1) > 0$ , d'où  $\sum_n \mathbb{P}(A_n) = +\infty$ . Il résulte du lemme de Borel-Cantelli (deuxième assertion), qu'avec une probabilité égale à 1, le mot  $A$  se réalise une infinité de fois au cours du jeu. Le même raisonnement montre que si un singe tape au hasard sur une machine à écrire, alors, avec une probabilité égale à 1, le mot ABRACADABRA se réalisera une infinité de fois au cours de la frappe. C'est vrai pour n'importe quel texte, donc il tapera aussi *une infinité de fois* le livre "A LA RECHERCHE DU TEMPS PERDU".

## 2.4 Rappels et compléments

### 2.4.1 Rappels sur les ensembles

Considérons un ensemble  $\Omega$  non-vidé, c'est-à-dire une collection d'objets appelés éléments de  $\Omega$ , ou points de  $\Omega$ . L'appartenance d'un point  $\omega$  à l'ensemble  $\Omega$  est notée  $\omega \in \Omega$ , et  $\omega \notin \Omega$  signifie que le point  $\omega$  n'appartient pas à  $\Omega$ .

Une partie  $A$  de  $\Omega$  est aussi un ensemble, appelé sous-ensemble de  $\Omega$ . Dans ce cas,  $A$  est dit inclus dans  $\Omega$ , ce qui s'écrit  $A \subset \Omega$ . Rappelons les opérations élémentaires sur les parties d'un ensemble.

**Intersection** :  $A \cap B$  est l'intersection des ensembles  $A$  et  $B$ , c'est-à-dire l'ensemble des points appartenant à la fois à  $A$  et à  $B$ .

**Réunion** :  $A \cup B$  est la réunion des ensembles  $A$  et  $B$ , c'est-à-dire l'ensemble des points appartenant à au moins l'un des deux ensembles.

**Ensemble vide** : C'est l'ensemble ne contenant aucun point. Il est noté  $\emptyset$ .

**Ensembles disjoints** : Les ensembles  $A$  et  $B$  sont dits disjoints si  $A \cap B = \emptyset$ .

**Complémentaire** : Si  $A \in \Omega$ , son complémentaire (dans  $\Omega$ ) est l'ensemble des points de  $\Omega$  n'appartenant pas à  $A$ . Il est noté  $A^c$  ou parfois  $\Omega \setminus A$ . Les ensembles  $A$  et  $A^c$  sont disjoints.

**Différence** : Si  $A$  et  $B$  sont deux sous-ensembles de  $\Omega$ ,  $A \setminus B$  désigne l'ensemble des points qui sont dans  $A$  mais pas dans  $B$ . Ainsi  $A \setminus B = A \cap B^c$ .

La réunion et l'intersection sont des opérations commutatives et associatives. Nous avons  $A \cup B = B \cup A$  et  $A \cap B = B \cap A$ , et aussi  $A \cup (B \cup C) = (A \cup B) \cup C$  et  $A \cap (B \cap C) = (A \cap B) \cap C$ , ensembles que nous notons naturellement  $A \cup B \cup C$  et  $A \cap B \cap C$ .

Plus généralement, pour une famille  $(A_i)_{i \in I}$  d'ensembles, indexée par un ensemble quelconque  $I$ ,  $\cup_{i \in I} A_i$  désigne **la réunion de cette famille**, i.e. l'ensemble des points appartenant à au moins l'un des  $A_i$ . De même,  $\cap_{i \in I} A_i$  désigne **l'intersection de cette famille**, i.e. l'ensemble des points appartenant à tous les  $A_i$ . Dans ces deux cas, l'ordre d'indexation des  $A_i$  n'a pas d'importance.

Une **partition** de  $\Omega$  est une famille  $(A_i)_{i \in I}$  telle que les ensembles  $A_i$  soient disjoints deux-à-deux ( $A_i \cap A_j = \emptyset, \forall i, j, i \neq j$ ), et que  $\cup_{i \in I} A_i = \Omega$ .

## 2.4.2 Modélisation ensembliste des événements

Les événements étant des ensembles, (rappelons-nous qu'une partie de  $\Omega$  décrit un sous-ensemble de résultats possibles de l'expérience), nous pourrons effectuer les opérations ensemblistes précédemment décrites, avec l'interprétation suivante. Si  $A$  et  $B$  sont deux événements, alors :

- **NON** : la réalisation de l'événement contraire à  $A$  est représentée par  $A^c$  : le résultat de l'expérience n'appartient pas à  $A$ .
- **ET** : l'événement “ $A$  et  $B$  sont réalisés” est représenté par  $A \cap B$  : le résultat de l'expérience se trouve à la fois dans  $A$  et dans  $B$ .

- **OU** : l'événement “ $A$  ou  $B$  sont réalisés” est représenté par l'événement  $A \cup B$  : le résultat de l'expérience se trouve dans  $A$  ou dans  $B$ .
- **IMPLICATION** : le fait que la réalisation de l'événement  $A$  entraîne la réalisation de  $B$  se traduit par  $A \subset B$ .
- **INCOMPATIBILITE** : si  $A \cap B = \emptyset$ ,  $A$  et  $B$  sont dits incompatibles. Un résultat de l'expérience ne peut être à la fois dans  $A$  et dans  $B$ .
- **TOUJOURS VRAI** : l'événement  $\Omega$  est l'événement certain (tous les résultats de l'expérience prennent leurs valeurs dans  $\Omega$ ).
- **IMPOSSIBLE** :  $\emptyset$  est l'événement impossible.

### 2.4.3 Les ensembles dénombrables

Un ensemble  $E$  est dénombrable s'il est en bijection avec  $\mathbb{N}$ , c'est-à-dire si ses points peuvent être énumérés en une suite  $(x_n)_{n \in \mathbb{N}}$ . C'est le cas de l'ensemble  $\mathbb{N}$  lui-même, de  $\mathbb{Z}$ , de  $\mathbb{Q}$ , des entiers pairs ou de toute suite strictement croissante d'entiers. Ce n'est pas le cas de  $\mathbb{R}$ , ni des intervalles  $[a, b]$  lorsque  $a < b$ , ni de  $\{0, 1\}^{\mathbb{N}}$ , ni de  $\mathcal{P}(\mathbb{N})$ .<sup>2</sup>

Enonçons quelques propriétés des ensembles dénombrables.

- Tout ensemble dénombrable est infini. La réciproque est fautive, comme nous l'avons vu ci-dessus.
- Toute partie d'un ensemble dénombrable est elle-même finie ou dénombrable.
- La réunion d'une famille finie ou dénombrable d'ensembles eux-mêmes finis ou dénombrables est un ensemble fini ou dénombrable.
- Si  $A$  n'est ni fini, ni dénombrable, il en est de même de  $A \setminus B$ , pour tout  $B \subset A$  qui est fini ou dénombrable.

### 2.4.4 Quelques résultats utiles sur les séries

Nous rappelons les résultats essentiels sur les séries, qui seront d'usage constant dans l'étude des variables aléatoires sur un espace dénombrable.

Soit  $(u_n)_{n \geq 1}$  une suite numérique, et  $S_n = u_1 + \dots + u_n$ ,  $n \geq 1$ , la suite des sommes partielles.

---

2. Il n'existe pas de bijection entre  $\mathbb{N}$  dans  $\mathcal{P}(\mathbb{N})$ . En effet, pour toute fonction  $f : \mathbb{N} \rightarrow \mathcal{P}(\mathbb{N})$ ,  $E = \{n \in \mathbb{N} : n \notin f(n)\}$  n'a pas d'antécédent : si  $f(n_0) = E$ , alors soit  $n_0 \in E$  et donc  $n_0 \notin f(n_0) = E$ , contradiction, soit  $n_0 \notin E$  et donc  $n_0 \in f(n_0) = E$ , encore une contradiction !

**S1** La série  $\sum_n u_n$  est dite *convergente* si  $S_n$  converge vers une limite *finie*  $S$ , notée aussi  $S = \sum_n u_n$ . C'est la “somme” de la série.

**S2** Si la série  $\sum_n u_n$  converge, la suite  $(u_n)_{n \geq 1}$  tend vers 0. La réciproque est **fausse** : on peut avoir  $u_n \rightarrow 0$  sans que la série  $\sum_n u_n$  converge (Prendre par exemple  $u_n = \frac{1}{n}$ ).

**S3** La série  $\sum_n u_n$  est dite *absolument convergente* si la série  $\sum_n |u_n|$  converge.

**S4** Si  $u_n \geq 0$  pour tout  $n$ , alors la suite  $S_n$  est croissante, donc elle tend toujours vers une limite  $S$  éventuellement infinie. On écrit encore  $S = \sum_n u_n$ , bien que la série converge au sens de (S1) si et seulement si  $S < \infty$ .

En général l'ordre dans lequel on considère les termes d'une série est important. Il existe en effet de nombreux exemples de suites  $(u_n)_{n \geq 1}$  et de bijections  $v$  de  $\mathbb{N}^*$  dans lui-même pour lesquels  $\sum_n u_n$  converge et  $\sum_n u_{v(n)}$  diverge, ou converge vers une somme différente. Cela étant, il existe deux cas importants où l'ordre des termes n'a pas d'importance :

**S5** Si  $u_n \geq 0$  pour tout  $n$ , la somme  $\sum_n u_n$  ne change pas si l'on change l'ordre de sommation. Rappelons rapidement la démonstration de cette propriété, qui est fondamentale pour les probabilités : soit  $v$  une bijection de  $\mathbb{N}^*$  dans lui-même,  $S_n = u_1 + \dots + u_n$  et  $S'_n = u_{v(1)} + \dots + u_{v(n)}$  ; les suites  $(S_n)$  et  $(S'_n)$  sont croissantes. Notons  $S$  et  $S'$  leur limites respectives (dans  $\mathbb{R}_+ \cup \{+\infty\}$ ). Pour tout  $n$  il existe un entier  $m(n)$  tel que  $v(i) \leq m(n)$  dès que  $i \leq n$ . Comme  $u_i \geq 0$ , clairement  $S'_n \leq S_{m(n)} \leq S$  et donc en passant à la limite nous obtenons  $S' \leq S$ . Nous montrons de même que  $S \leq S'$ , et donc  $S = S'$ .

**S6** Lorsque les  $u_n$  sont des réels de signe quelconque et que la série est absolument convergente, nous pouvons modifier de manière arbitraire l'ordre des termes sans changer la propriété d'être absolument convergente, ni la somme de la série.

**S7** Si  $u_n \geq 0$ , il est possible de “sommer par paquets”. Cela signifie la chose suivante : soit  $(A_i)_{i \in I}$  une partition finie ou dénombrable de  $\mathbb{N}^*$ . Pour chaque  $i \in I$ , posons  $v_i = \sum_{n \in A_i} u_n$ . Si  $A_i$  est fini, c'est une somme ordinaire ; sinon  $v_i$  est elle-même la somme d'une série à termes positifs. Nous avons alors  $\sum_n u_n = \sum_{i \in I} v_i$ , qui est de nouveau la somme d'une série à termes positifs si  $I = \mathbb{N}^*$ . La démonstration de ce résultat est analogue à celle de (S5) ci-dessus.

**S8** Si la série  $\sum_n u_n$  est absolument convergente, la propriété (S7) est encore vraie.

**S9** Théorème de Fubini pour les séries. Soit  $(a_{mn})_{m,n \in \mathbb{N}}$  une série double telle que la série de terme général  $\sum_m |a_{mn}|$  converge. Alors les séries  $\sum_n \sum_m a_{mn}$  et  $\sum_m \sum_n a_{mn}$  sont convergentes et de même somme.

## 2.5 Exercices sur le chapitre 2

### EXERCICE 2.7

1) Parmi  $n$  personnes en présence ( $n \leq 365$ ), quelle est la probabilité pour qu'au moins deux personnes soient nées le même jour ? (On conviendra de ne pas prendre en compte les personnes nées le 29 février). Que vaut cette probabilité pour  $n = 4$ ,  $n = 16$ ,  $n = 22$ ,  $n = 40$ ,  $n = 64$  ?

2) Déterminer  $n_{min}$  pour que la probabilité qu'au moins deux personnes soient nées le même jour soit supérieure à 0,5. On pourra utiliser la formule de Stirling  $m! \sim_{m \rightarrow \infty} \sqrt{2\pi m} m^{m+\frac{1}{2}} e^{-m}$ .

**EXERCICE 2.8** Montrer la formule de Poincaré :

$$\mathbb{P}(\cup_{m=1}^n A_m) = \sum_{k=1}^n (-1)^{k-1} p_k, \quad \text{où } p_k = \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}).$$

**EXERCICE 2.9** Un facteur possède  $n$  lettres adressées à  $n$  destinataires distincts. Il est totalement ivre et poste au hasard une lettre par boîte.

- 1) Quelle est la probabilité d'obtenir la bonne répartition ?
- 2) Quelle est la probabilité qu'une lettre au moins arrive à la bonne adresse ?
- 3) Quelle est la probabilité qu'aucune lettre n'arrive à la bonne destination ?
- 4) Quel est le nombre  $d_n$  de manières différentes de poster les lettres de telle sorte qu'aucune n'arrive à destination ?

**EXERCICE 2.10** Soit  $a \in ]0, 1[$ . Montrer que la suite de nombres définie par  $p_n = (1-a)a^{n-1}$  caractérise une probabilité sur  $\mathbb{N}^*$ .

**EXERCICE 2.11** M. et Mme Barbétipoil ont deux enfants, garçons ou filles, les 4 configurations sont équiprobables. Quelle est la probabilité que les deux enfants Barbétipoil soient des filles,

- sans autre information,
- sachant que l'aînée est une fille,
- sachant que l'un des deux enfants est une fille.

**EXERCICE 2.12** Modèle de Hardy-Weinberg. Les caractères héréditaires dans certains organismes, tels que les humains, sont portés par des paires de gènes. Dans le cas le plus simple, chaque gène peut prendre deux formes appelées allèles,  $A$  et  $a$ . Ces allèles se trouvent dans une population parentale avec les proportions  $p$  et  $q$ . Comme  $Aa$  et  $aA$  ne sont pas discernables, il y a 3 génotypes possibles,  $AA$ ,  $aa$ ,  $Aa$ . Nous supposerons que la reproduction peut avoir lieu entre deux individus quelconques de la



population, indépendamment des gènes considérés. Chaque parent transmet un gène de son génotype de façon équiprobable, les deux gènes ainsi obtenus constituant le génotype du descendant.

Calculer la probabilité des différents génotypes dans la génération suivante. Montrer que la proportion de chacun des allèles reste la même dans la deuxième génération.

**EXERCICE 2.13** L'hémophilie est transmise par la mère. La reine porte le gène de l'hémophilie avec une probabilité de 0,5. Si elle est porteuse, chaque prince aura une chance sur deux de souffrir de cette maladie. La reine a eu 3 fils non hémophiles. Quelle est la probabilité qu'elle soit porteuse du gène ? S'il naît un quatrième prince, avec quelle probabilité sera-t-il hémophile ?

**EXERCICE 2.14** Le jeu des 3 portes est un jeu télévisé populaire (Let's make a deal) diffusé dans les années 70 aux USA. Trois portes A, B, C sont fermées. Derrière l'une d'elle il y a une Ferrari, derrière les autres une chèvre.

- Le joueur choisit une porte, disons A.
- Le présentateur, qui sait où se trouve la voiture, l'informe alors qu'elle n'est pas derrière la porte B et lui offre la possibilité de réviser son choix (i.e. de choisir la porte C).

Le joueur a-t-il intérêt à réviser son choix ?

**EXERCICE 2.15** Un problème simple de démographie. Soit  $a \in ]0, 1[$ . Soit  $p_k$ , la probabilité qu'une famille ait  $k$  enfants. Nous supposons que

$$p_0 = p_1 = a ; \quad p_k = (1 - 2a)2^{-(k-1)}, \quad \forall k \geq 2,$$

et que  $\mathbb{P}(\text{Fille}) = \mathbb{P}(\text{Garçon}) = \frac{1}{2}$ .

On pose :  $E_n$  : "la famille a  $n$  enfants",  $F_n$  : " $n$  filles",  $G_n$  : " $n$  garçons".

- 1) Donner la probabilité qu'une famille ayant deux filles aient deux enfants seulement ?
- 2) Donner la probabilité qu'une famille ait deux garçons sachant qu'elle a deux filles ?

**EXERCICE 2.16** On jette indéfiniment une pièce de monnaie, la probabilité d'apparition d'un Face étant égale à  $p$ . Soit  $A_k$ , pour  $k$  entier, l'événement selon lequel au moins " $k$  Faces" consécutifs apparaissent au cours des lancers numérotés  $2^k, 2^k + 1, \dots, 2^{k+1} - 1$ .

Montrer que  $\mathbb{P}(\limsup_k A_k)$  vaut 0 ou 1 selon que  $p < \frac{1}{2}$  ou que  $p \geq \frac{1}{2}$ . Comment interprétez-vous ce résultat ?

**EXERCICE 2.17** Montrer qu'il n'existe pas de probabilité  $\mathbb{P}$  sur  $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$  telle que  $\mathbb{P}(k\mathbb{N}) = \frac{1}{k}$  pour tout entier strictement positif  $k$ , où  $k\mathbb{N}$  désigne l'ensemble des entiers multiples de  $k$ .



## Chapitre 3

# Variables aléatoires sur un espace fini ou dénombrable

*I have deeply regretted that I did not proceed at least to understand something of the great leading principles of mathematics; for men thus endowed seem to have an extra sense.*

Charles Darwin, *Uses and abuses of Mathematics in Biology*

Dans ce chapitre et le suivant, nous allons introduire la notion de variable aléatoire et en développer une étude plus systématique. La grande nouveauté va être de comprendre que ce n'est pas la variable aléatoire en tant que fonction précise de l'aléa qui nous intéresse, mais sa loi, c'est-à-dire la description de son "comportement probable".

**Dans tout ce chapitre, l'espace fondamental  $\Omega$  est fini ou dénombrable.**

### 3.1 Variables aléatoires discrètes

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité fini ou dénombrable muni de la tribu  $\mathcal{A} = \mathcal{P}(\Omega)$ . Toute application  $X$  définie sur  $\Omega$  :

$$X : \Omega \longrightarrow \mathcal{X}$$

est appelée variable aléatoire (v.a.). Comme  $\Omega$  est fini ou dénombrable, l'ensemble d'arrivée (quitte à le remplacer par  $X(\Omega)$  sans perte de généralité) est fini ou dénombrable et de la forme  $\mathcal{X} = \{x_i, i \in I\}$ , pour un certain sous-ensemble  $I \subset \mathbb{N}$ .

**Attention**, cette définition de v.a. n'est valable que dans le cadre d'un espace  $\Omega$  **fini ou dénombrable**. On donnera une définition générale de v.a. dans le cas d'un espace arbitraire dans le chapitre suivant.

La loi d'une variable aléatoire  $X$  est définie par la probabilité induite sur  $\mathcal{X}$ , muni de la tribu  $\mathcal{P}(\mathcal{X})$ , et le résultat suivant est une conséquence immédiate de la proposition 2.16.

**Proposition 3.1** *La loi d'une v.a.  $X$  à valeurs dans un espace au plus dénombrable  $\mathcal{X}$  est caractérisée par*

$$\{(x_i, p_i^X), x_i \in \mathcal{X}\},$$

avec  $p_i^X = \mathbb{P}_X(\{x_i\}) = \mathbb{P}(\{\omega, X(\omega) = x_i\}) = \mathbb{P}(X = x_i) = \sum_{\omega: X(\omega)=x_i} p_\omega$ .

**Exemple 3.2** *Une v.a. de loi uniforme sur  $\{1, \dots, n\}$  a pour loi  $\{(k, \frac{1}{n})\}_{1 \leq k \leq n}$ .*

La représentation graphique d'une loi de variable discrète en utilisant un diagramme "en bâtons" est très parlante. Les valeurs  $x_i$  sont placées en abscisse et les images  $p_i^X$  en ordonnée, comme dans la figure 3.1.

**Exemple 3.3** *Considérons le lancer d'un dé. L'espace fondamental est  $\Omega = \{1, \dots, 6\}$  muni de la probabilité uniforme. Le résultat du lancer d'un dé est la variable aléatoire  $X(\omega) = \omega$ , à valeurs dans  $\mathcal{X} = \{1, \dots, 6\}$ , et on a donc :*

$$p_1^X = \frac{1}{6}, \quad p_2^X = \frac{1}{6}, \quad \dots, \quad p_6^X = \frac{1}{6}.$$

On représente cette loi par un diagramme en bâtons sur la figure 3.1a.

**Exemple 3.4** *Considérons le lancer de deux dés. L'espace fondamental est  $\Omega = \{1, \dots, 6\}^2$  muni de la probabilité uniforme. Le résultat total est la somme des deux lancers,  $S(\omega) = \omega_1 + \omega_2$ , où  $\omega = (\omega_1, \omega_2)$ . Cette v.a. prend ses valeurs dans l'ensemble des entiers  $\mathcal{X} = \{2, \dots, 12\}$ , et on obtient par des opérations de dénombrement élémentaires :*

$$\begin{aligned} p_2^S &= \frac{1}{36}, & p_3^S &= \frac{2}{36}, & \dots, & p_7^S &= \frac{6}{36}, \\ p_8^S &= \frac{5}{36}, & \dots, & p_{11}^S &= \frac{2}{36}, & p_{12}^S &= \frac{1}{36}. \end{aligned}$$

On représente cette loi par un diagramme en bâtons sur la figure 3.1b.

La représentation graphique ne donne pas d'information quantitative sur la loi. Dans les paragraphes suivants, nous allons définir des nombres réels qui vont résumer en un certain sens le comportement de la variable aléatoire.

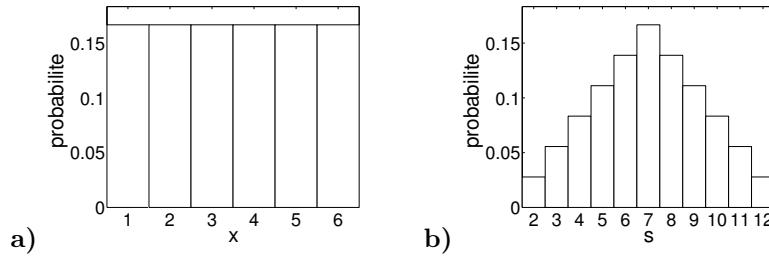


FIGURE 3.1 – Diagramme en bâtons de la loi d'un lancer de dé (a) et de la loi de la somme de deux lancers de dés (b).

## 3.2 Espérance des v.a. discrètes

Nous supposons ici que  $\mathcal{X}$  est un sous-ensemble fini ou dénombrable de  $\mathbb{R}$ .

### 3.2.1 Définition

**Motivation :** Répétons  $n$  fois une expérience aléatoire, et notons  $X_1, \dots, X_n$  les valeurs successives prises par  $X$ . Pour avoir une idée du comportement de la variable  $X$ , il est naturel de considérer leur moyenne arithmétique  $M_n = \frac{1}{n}(X_1 + \dots + X_n)$ . (Pensez à vos propres calculs de moyennes en tant qu'élèves.) En regroupant suivant les différents résultats possibles  $\omega$  d'une expérience, nous obtenons

$$M_n = \sum_{\omega \in \Omega} f_n(\{\omega\}) X(\omega),$$

où  $f_n(\{\omega\})$  est la fréquence de réalisation du singleton  $\{\omega\}$  au cours des  $n$  expériences. Remarquons que la dernière somme est finie puisqu'il ne peut exister qu'un nombre fini de  $f_n(\omega)$  non nuls. Nous voulons faire tendre  $n$  vers l'infini. Si la propriété (2.1) intuitée au Chapitre 2 est vraie, c'est-à-dire si  $f_n(\{\omega\}) \rightarrow p_\omega$ , et si dans l'expression ci-dessus on peut intervertir la somme et la limite (ce qui est en particulier vrai si  $\Omega$  est fini), alors la suite  $(M_n)_n$  tend vers  $\sum_{\omega \in \Omega} p_\omega X(\omega)$ . Nous justifierons cette assertion plus loin, dans l'un des théorèmes les plus importants de la théorie des probabilités, appelé la loi des grands nombres.

**Définition 3.5** Soit  $X$  une variable aléatoire réelle définie sur l'espace fini ou dénombrable  $\Omega$  à valeurs dans  $\mathcal{X}$  (i.e. une application de  $\Omega$  dans  $\mathcal{X} \subset \mathbb{R}$ ). Son espérance

(appelée aussi parfois moyenne) est

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} p_{\omega} X(\omega), \quad (3.1)$$

**pourvu que la somme  $\mathbb{E}(|X|) = \sum_{\omega \in \Omega} p_{\omega} |X(\omega)|$  soit finie.**

(Rappelons que, en vertu de (S6) (section 2.4.4), comme la série de terme général  $p_{\omega} X(\omega)$  est absolument convergente, la somme  $\mathbb{E}(X)$  de la série ne dépend pas de la manière dont les  $\omega$  sont ordonnés.)

L'espérance mathématique  $\mathbb{E}(X)$  est un réel qui donne une valeur moyenne résumant la v.a.  $X$ . C'est l'un des concepts les plus importants de la théorie des probabilités. La dénomination d'espérance pour cette quantité fait référence aux problèmes de jeux et d'espérance de gain. Cette terminologie imagée a été introduite par Pascal.

**Théorème 3.6** *Pour une variable aléatoire  $X$  satisfaisant  $\mathbb{E}(|X|) < +\infty$ , on a :*

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} p_{\omega} X(\omega) = \sum_{x_i \in \mathcal{X}} x_i \mathbb{P}(X = x_i) = \sum_{x_i \in \mathcal{X}} x_i p_i^X. \quad (3.2)$$

*En particulier, nous remarquons que l'espérance de  $X$  ne dépend que de la loi de  $X$ .*

**Preuve.** Puisque  $\sum_{\omega \in \Omega} p_{\omega} |X(\omega)| = \mathbb{E}(|X|) < +\infty$ , la sommation par paquets est justifiée par (S8) (section 2.4.4), et on obtient

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} p_{\omega} X(\omega) = \sum_{x_i \in \mathcal{X}} \sum_{\omega: X(\omega)=x_i} p_{\omega} x_i = \sum_{x_i \in \mathcal{X}} x_i \mathbb{P}(\{\omega: X(\omega)=x_i\}) = \sum_{x_i \in \mathcal{X}} x_i p_i^X.$$

□

**Analogie avec une notion de mécanique :** Le concept d'espérance est à rapprocher de la notion de *centre de gravité* d'un groupe de masses au sens de la mécanique. Considérons en effet une variable  $X$  de loi de probabilité  $\{(x_i, p_i^X), i \geq 1\}$ . On montre que si les masses  $p_i^X, i \geq 1$  sont réparties sur une barre sans poids aux abscisses  $x_i, i \geq 1$ , le centre de gravité, c'est-à-dire le point sur lequel la barre pourra être posée et rester en équilibre, est d'abscisse  $\mathbb{E}(X)$ . En effet, il suffit d'établir que la somme des moments des forces gravitationnelles par rapport au point d'abscisse  $\mathbb{E}(X)$  est nulle :

$$0 = \sum_{x_i \in \mathcal{X}} (x_i - \mathbb{E}(X)) p_i^X,$$

ce qui est immédiat à vérifier à partir de la définition de l'espérance mathématique.

**Exemple 3.7** Un nombre est choisi au hasard entre 1 et 10, et nous devons deviner ce nombre en posant des questions auxquelles il ne sera répondu que par oui ou par non. Calculons l'espérance du nombre  $N$  de questions nécessaires dans les cas suivants :

- La  $i$ ème question est du type “Est-ce  $i$  ?”,  $i$  étant égal à  $1, 2, \dots, 10$ .

Soit  $A_k$  l'événement : “le nombre  $k \in \{1, \dots, 10\}$  a été choisi”. Comme les  $A_k$  forment une partition de l'espace fondamental, on a

$$\mathbb{P}(N = k) = \sum_{k'=1}^{10} \mathbb{P}(N = k | A_{k'}) \mathbb{P}(A_{k'}) = \mathbb{P}(N = k | A_k) \mathbb{P}(A_k) = \frac{1}{10}$$

et

$$\mathbb{E}(N) = \sum_{k=1}^{10} k \mathbb{P}(N = k) = \frac{11}{2}.$$

- Avec chaque question, nous éliminer à peu près la moitié des réponses possibles, avec le protocole suivant : est-ce  $\leq 5$ ,  $\leq 2$  (resp.  $\leq 7$ ),  $\leq 4$  (resp.  $\leq 9$ ). Alors

$$\mathbb{E}(N) = 3 \times \frac{6}{10} + 4 \times \frac{4}{10} = \frac{17}{5}.$$

### 3.2.2 Propriétés de l'espérance des v.a. discrètes

**Définition 3.8** Une v.a.  $X$  est dite intégrable si  $\sum_{\omega \in \Omega} p_{\omega} |X(\omega)| = \mathbb{E}(|X|) < +\infty$ . L'ensemble des v.a. intégrables est noté  $L^1$ , il dépend de  $\Omega$  et de  $\mathbb{P}$ .

Les propriétés suivantes sont immédiates :

#### Proposition 3.9

- (i)  $L^1$  est un espace vectoriel, et l'espérance est linéaire sur  $L^1$  :

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y) \quad \text{pour tous } X, Y \in L^1, \text{ et } a, b \in \mathbb{R}.$$

- (ii)  $X \in L^1 \iff |X| \in L^1$ , et dans ce cas,  $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$ .

- (iii) L'espérance est positive : si  $X \geq 0$  alors  $\mathbb{E}(X) \geq 0$ .

- (iv) Pour  $X, Y \in L^1$ ,  $X \leq Y$  alors  $\mathbb{E}(X) \leq \mathbb{E}(Y)$ .

- (v)  $L^1$  contient toutes les v.a. bornées. ( $X$  est bornée si  $\sup_{\omega \in \Omega} |X(\omega)| < \infty$ ).

- (vi) L'espérance d'une variable constante est égale à cette constante : si  $X(\omega) = a$  pour tout  $\omega \in \Omega$ , alors  $\mathbb{E}(X) = a$ .

- (vii) Si  $\Omega$  est fini,  $L^1$  contient toutes les v.a. réelles.

**Exemple 3.10** Pour tout événement  $A$ , on définit

$$\mathbf{1}_A(\omega) = 1 \quad \text{si } \omega \in A \quad \text{et} \quad \mathbf{1}_A(\omega) = 0 \quad \text{si } \omega \notin A.$$

Cette v.a. est appelée fonction indicatrice de  $A$ , et nous avons :

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

### 3.2.3 Variance et écart-type

On note par  $L^2$  l'ensemble des v.a. réelles  $X$  dont le carré  $X^2$  est intégrable. Nous dirons dans ce cas que  $X$  **est de carré intégrable**.

**Proposition 3.11**  $L^2$  est un sous-espace vectoriel de  $L^1$ , et si  $X \in L^2$  on a

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|) \leq \sqrt{\mathbb{E}(X^2)}. \quad (3.3)$$

**Preuve.** Soient  $X, Y \in L^2$  et  $a \in \mathbb{R}$ . Comme  $(aX + Y)^2 \leq 2a^2X^2 + 2Y^2$ , nous déduisons de la proposition 3.9(i) que  $aX + Y \in L^2$ . Ainsi,  $L^2$  est un espace vectoriel. L'inclusion  $L^2 \subset L^1$  découle de  $|X| \leq 1 + X^2$  et de la linéarité de la proposition 3.9(i). La première inégalité de (3.3) est celle de la proposition 3.9(ii). Pour la seconde, nous pouvons nous limiter au cas où  $X$  est positive. Soit alors  $a = \mathbb{E}(X)$  et  $Y = X - a$ . D'après la proposition 3.9(i) il vient

$$\mathbb{E}(Y^2) = \mathbb{E}(X^2) - 2a\mathbb{E}(X) + a^2 = \mathbb{E}(X^2) - a^2,$$

et  $\mathbb{E}(Y^2) \geq 0$  par la proposition 3.9(iii). Donc  $a^2 \leq \mathbb{E}(X^2)$ . □

**Définition 3.12** Si  $X \in L^2$ , sa **variance** est définie par

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \sum_{x_i \in \mathcal{X}} (x_i - \mathbb{E}(X))^2 p_i^X.$$

$\text{Var}(X)$  est positive, et  $\sigma_X = \sqrt{\text{Var}(X)}$  s'appelle **l'écart-type** de  $X$ .

L'écart-type est une grandeur qui mesure la moyenne (en un certain sens) de l'écart des valeurs de  $X$  à sa moyenne, d'où son nom. En développant le carré  $(X - \mathbb{E}(X))^2$  comme dans la preuve ci-dessus, nous obtenons la formule de Huygens :

$$\sigma_X^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (3.4)$$

**Exemple 3.13 (Loto)** On considère tout d'abord la forme classique du loto, qui a été utilisée de 1976 à 2008. Le joueur coche 6 numéros sur une grille en comportant 49, dans le cas d'un bulletin simple. Les 6 numéros gagnants sont déterminés par tirage



au sort. L'espace fondamental est ici l'ensemble des parties à 6 éléments (« tirage ») de  $\{1, \dots, 49\}$  muni de la probabilité uniforme. Son cardinal est  $\text{Card}(\Omega) = \binom{49}{6} = 13\,983\,816$ . Notant  $N$  le nombre de numéros gagnants figurant parmi les numéros cochés par le joueur (on rappelle que la grille du joueur comporte 6 numéros cochés et 43 non-cochés), l'événement  $\{N = n\}$  ( $n \in \{0, \dots, 6\}$ ) est réalisé si le tirage produit  $n$  numéros cochés et  $6 - n$  numéros non-cochés. La loi de la v.a.  $N$  est donc :

$$\mathbb{P}(N = n) = \frac{\binom{6}{n} \binom{43}{6-n}}{\binom{49}{6}}, \quad n \in \{0, \dots, 6\},$$

et certaines valeurs numériques sont données dans le tableau ci-dessous. Au premier tirage du 10 mai 2006, on recevait pour une mise de 0,3 Euros (soit une grille) le gain  $G = g(N)$  suivant :

$n$ numéros gagnants	gain $g(n)$	probabilité $\mathbb{P}(N = n)$
6	789 177,00 Euros	$7,2 \cdot 10^{-8}$
5	1 813,80 Euros	$1,8 \cdot 10^{-5}$
4	30,70 Euros	$9,7 \cdot 10^{-4}$
3	2,70 Euros	$1,8 \cdot 10^{-2}$

Le gain moyen est  $\mathbb{E}[G] = 2,70 \times 1,8 \cdot 10^{-2} + 30,70 \times 9,7 \cdot 10^{-4} + \dots \simeq 0,168$  Euros, tandis que l'écart-type  $\sigma(G) = (\mathbb{E}[G^2] - \mathbb{E}[G]^2)^{1/2} \simeq 212$  Euros. On voit que le jeu est défavorable au joueur, dont le bénéfice moyen est  $\mathbb{E}[G] - 0,3 = -0,132$  Euros, et la grande valeur de  $\sigma$  vient de ce que parfois (mais très rarement) ça rapporte gros.

La forme moderne du loto, mise en place à partir de 2008, est légèrement différente : le joueur coche 5 numéros sur une grille en comportant 49 et un « numéro chance » sur une autre grille en comportant 10. Donc la probabilité qu'un joueur trouve la bonne combinaison est :

$$p = \frac{1}{\binom{49}{5}} \frac{1}{10} = \frac{1}{19\,068\,840} \simeq 5,2 \cdot 10^{-8},$$

car il y a  $\binom{49}{5}$  choix possibles des cinq premiers numéros, et 10 choix possibles du « numéro chance ». Notez qu'on a moins de chance de gagner le gros lot avec la nouvelle version qu'avec la version classique.

### 3.2.4 Moments d'une variable aléatoire

Soit  $f$  une fonction de  $\mathcal{X}$  dans  $\mathbb{R}$ . La v.a. réelle  $f(X)$  prend un nombre fini ou dénombrable de valeurs. Nous pouvons aussi considérer  $f$  comme une v.a. définie sur l'espace de probabilité  $(\mathcal{X}, \mathcal{P}(\mathcal{X}), \mathbb{P}_X)$ . Le résultat suivant, appelé propriété de transfert, montre la cohérence de la notion d'espérance.

**Théorème 3.14** *Pour  $f : \mathcal{X} \rightarrow \mathbb{R}$ , la v.a.  $f(X)$  définie sur  $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$  est intégrable si et seulement si la v.a.  $f$  sur  $(\mathcal{X}, \mathbb{P}_X)$  l'est également. Dans ce cas, les espérances de ces deux v.a. sont égales, et*

$$\mathbb{E}(f(X)) = \sum_{\omega \in \Omega} f(X(\omega)) p_\omega = \sum_{x_i \in \mathcal{X}} f(x_i) \mathbb{P}(X = x_i). \quad (3.5)$$

**Preuve.** Comme nous pouvons sommer par paquets par (S7) (section 2.4.4) dans une série à termes positifs, nous voyons comme pour (3.2) que les deux expressions de droite de (3.5) sont égales si nous remplaçons  $f$  par  $|f|$ ; elles sont donc finies simultanément. D'après la définition de l'espace  $L^1$ , nous avons donc  $f(X) \in L^1(\Omega, \mathbb{P}) \Leftrightarrow f \in L^1(\mathcal{X}, \mathbb{P}_X)$ .

Si ces propriétés sont réalisées, en utilisant cette fois (S8) (section 2.4.4), nous voyons de la même manière que les deux expressions de droite de (3.5) sont aussi égales pour  $f$ , ce qui, compte-tenu de (3.1), achève la démonstration.  $\square$

**Définition 3.15** Soit  $p \in \mathbb{N}^*$ . On dit que la v.a.  $X$  admet un moment d'ordre  $p$  si  $X^p \in L^1$ , et en utilisant le théorème précédent :

$$\mathbb{E}(X^p) = \sum_{x_i \in \mathcal{X}} x_i^p \mathbb{P}(X = x_i).$$

### 3.3 Fonction génératrice d'une v.a. entière

La loi d'une v.a. entière (à valeurs dans  $\mathbb{N}$ )  $X$  est caractérisée par les nombres  $p_n = p_n^X = \mathbb{P}(X = n)$ . Le but de ce paragraphe est de montrer qu'une telle loi de probabilité peut également être caractérisée par une fonction, appelée fonction génératrice, définie sur  $[0, 1]$  et indéfiniment dérivable sur  $[0, 1[$ . Les dérivées auront leur interprétation en termes de moments de la variable aléatoire.

**Définition 3.16** La fonction génératrice  $G_X$  d'une v.a. entière  $X$  est la fonction définie sur l'intervalle  $[0, 1]$  par la formule suivante :

$$G_X(s) = \mathbb{E}(s^X) = \sum_{n=0}^{\infty} p_n s^n, \quad \text{pour tout } s \in [0, 1]. \quad (3.6)$$

La fonction génératrice ne dépend que de la loi de  $X$ , c'est-à-dire de  $(p_n)_n$ .

**Proposition 3.17** *La fonction génératrice  $G_X$  est continue sur  $[0, 1]$  et indéfiniment dérivable sur  $[0, 1[$ ; elle caractérise la loi de  $X$ .*

**Preuve.** La fonction  $G_X$  est la somme d'une série entière qui converge absolument au point 1, puisque  $\sum_n p_n = 1$ . Les propriétés de continuité et de dérivabilité en découlent. Comme la dérivée  $n^{\text{ième}}$  en 0 est  $G_X^{(n)}(0) = p_n n!$ , la fonction  $G_X$  caractérise les  $p_n$ , donc la loi de  $X$ .  $\square$

**Proposition 3.18** *Une v.a. entière  $X$  est intégrable si et seulement si  $G_X$  est dérivable à gauche au point 1, et dans ce cas  $\mathbb{E}(X) = G'_X(1)$ .*

**Preuve.** Rappelons d'abord un résultat sur les séries : si les fonctions  $s \mapsto u_n(s)$  sont croissantes et positives sur  $[0, 1[$ , on a :

$$\lim_{s \uparrow 1} \sum_{n \geq 0} u_n(s) = \sum_{n \geq 0} \lim_{s \uparrow 1} u_n(s). \quad (3.7)$$

Si  $s < 1$ , nous avons

$$\frac{G_X(s) - G_X(1)}{s - 1} = \sum_{n \geq 0} p_n \frac{s^n - 1}{s - 1} = \sum_{n \geq 0} p_n (1 + s + \cdots + s^{n-1}),$$

et les fonctions  $u_n(s) = p_n (1 + s + \cdots + s^{n-1})$  sont croissantes et positives, avec  $\lim_{s \uparrow 1} u_n(s) = n p_n$ . Le résultat découle alors de (3.7).  $\square$

Plus généralement, la même démonstration prouve que

**Proposition 3.19** *La v.a.  $X(X - 1) \cdots (X - p)$  est intégrable (et donc  $X$  admet un moment d'ordre  $p + 1$ ), si et seulement si  $G_X$  est  $p + 1$  fois dérivable à gauche au point 1, et nous avons alors*

$$\mathbb{E}(X(X - 1) \cdots (X - p)) = G_X^{(p+1)}(1). \quad (3.8)$$

En particulier,  $\mathbb{E}(X(X - 1)) = G_X''(1)$ , d'où

$$\text{Var}(X) = G_X''(1) - (G_X'(1))^2 + G_X'(1).$$

Pour retrouver cette formule, nous pouvons dériver formellement la série (3.6) terme à terme  $p + 1$  fois au point  $s = 1$ , ce qui donne

$$G_X^{(p+1)}(1) = \sum_n p_n n(n - 1) \cdots (n - p),$$

et le membre de droite ci-dessus est égal au membre de gauche de (3.8) lorsque ce dernier existe, d'après (3.5). Cela revient à dériver formellement  $p + 1$  fois les deux membres de l'égalité  $G_X(s) = \mathbb{E}(s^X)$ , en échangeant les dérivées et le signe espérance.

**Remarque 3.20** Pour calculer les moments d'une variable aléatoire entière (espérance, variance,...), il peut être plus simple d'utiliser les dérivées de la fonction génératrice plutôt qu'un calcul direct, comme nous le verrons dans le paragraphe ci-dessous.

## 3.4 Variables aléatoires discrètes usuelles

### 3.4.1 Variable aléatoire de Bernoulli

Nous lançons une pièce  $n$  fois. Nous associons 1 à Pile et 0 à Face. L'espace des résultats de l'expérience sera donc

$$\Omega = \{0, 1\}^n.$$

Nous supposons que les lancers sont indépendants les uns des autres. Par ailleurs, la pièce peut être truquée, ce qui nous conduit à supposer que la probabilité de Pile vaut  $p \in ]0, 1[$ . Pour une pièce équilibrée, nous prendrons  $p = \frac{1}{2}$ .

Pour tout  $k \in \{1, \dots, n\}$ , appelons  $X_k$  le résultat du  $k$ -ième lancer.  $X_k$  peut prendre les deux valeurs 0 et 1, et

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

Nous remarquons que chaque variable  $X_k$  a la même loi, prenant les deux valeurs 1 et 0 avec respectivement les probabilités  $p$  et  $1 - p$ .

**Définition 3.21**  $X$  est une **variable de Bernoulli** de paramètre  $p$  si  $X$  prend ses valeurs dans  $\{0, 1\}$  avec  $\mathbb{P}(X = 1) = p$  et  $\mathbb{P}(X = 0) = 1 - p$ . Sa loi est **la loi de Bernoulli de paramètre  $p$ , que l'on note  $\mathcal{B}(p)$** .

L'espérance, la variance et la fonction génératrice d'une variable aléatoire de Bernoulli de paramètre  $p$  valent respectivement

$$\mathbb{E}(X) = p, \quad \text{Var}(X) = p(1 - p) \quad \text{et} \quad G_X(s) = 1 - p + ps. \quad (3.9)$$

Nous avons en effet  $\mathbb{E}(X) = p \times 1 + 0 \times (1 - p) = p$ ,  $\text{Var}(X) = p - p^2 = p(1 - p)$ , et le calcul de  $G_X$  est tout aussi immédiat.

### 3.4.2 Variable aléatoire binomiale

Le nombre de Piles obtenus sur les  $n$  lancers est donné par :

$$S_n = \sum_{i=1}^n X_i.$$

La v.a.  $S_n$  peut prendre toutes les valeurs entières entre 0 et  $n$ . Calculons sa loi. Pour tout  $k \in \{0, \dots, n\}$ , nous cherchons

$$\mathbb{P}_{S_n}(\{k\}) = \mathbb{P}(S_n = k) = \mathbb{P}\left(\{\omega \in \Omega, \sum_{i=1}^n X_i(\omega) = k\}\right).$$

Cela veut dire que les  $n$  lancers ont donné  $k$  Piles et  $n - k$  Faces. Il faut tenir compte des places des Piles dans la suite de résultats obtenus. Il y a  $\binom{n}{k}$  possibilités d'obtenir les  $k$  Piles parmi les  $n$  lancers. Si nous fixons une répartition précise, (par exemple les  $k$  Piles sont les  $k$  premiers), et comme les lancers sont indépendants, la probabilité de cette répartition est égale à  $p^k(1-p)^{n-k}$ . Ainsi, nous obtenons que

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{pour tout } k = 0, \dots, n.$$

**Définition 3.22**  $X$  est une **variable binomiale** de paramètres  $n$  et  $p$  si  $X$  prend ses valeurs dans  $\{0, \dots, n\}$  et si pour  $k \in \{0, \dots, n\}$ ,

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

Sa loi est une **loi binomiale de paramètres  $n$  et  $p$** , que l'on note  $\mathcal{B}(n, p)$ .

Nous avons obtenu cette loi dans nos modèles d'urnes, par un raisonnement intuitif. Elle correspond au choix d'un tirage avec remise. Remarquons que  $\mathcal{B}(1, p) = \mathcal{B}(p)$  est la loi de Bernoulli de paramètre  $p$ .

Pour une variable binomiale  $X$  de loi  $\mathcal{B}(n, p)$ , on a

$$\mathbb{E}(X) = np, \quad \text{Var}(X) = np(1-p) \quad \text{et} \quad G_X(s) = (1-p+ps)^n. \quad (3.10)$$

En effet, on calcule que  $G_X(s) = \sum_{k=0}^n \binom{n}{k} p^k s^k (1-p)^{n-k} = (1-p+ps)^n$ . En dérivant  $G_X$  et en considérant la dérivée de  $G_X$  en  $s = 1$ , nous obtenons que  $\mathbb{E}(X) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = G'_X(1) = pn$ . Si nous dérivons deux fois en 1 ce polynôme  $G_X$ , nous obtenons  $G''_X(1) = \sum_{k=1}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} = \mathbb{E}(X(X-1))$ , et par suite, puisque  $\mathbb{E}(X) = np$ , nous en déduisons que  $\text{Var}(X) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X^2) = np(1-p)$ .

**Exemple 3.23** *Aux jeux olympiques de Vancouver (2010), 86 médailles d'or ont été mises en jeu. Nous faisons l'hypothèse que le nombre de médailles remportées par pays est proportionnel à sa population. Soit  $X$  le nombre de médailles prévues pour la France.  $X$  va suivre une loi binomiale  $\mathcal{B}(86, p)$ , où*

$$p = \frac{\text{population France}}{\text{population monde}} \approx \frac{60 \times 10^6}{6000 \times 10^6} = 0,01.$$

Ainsi  $\mathbb{E}(X) = 86 \times 0,01 = 0,86$ . La probabilité pour que le nombre de médailles soit inférieur à 3 est  $\mathbb{P}(X \leq 3) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3)$ , avec

$$\mathbb{P}(X = k) = \binom{86}{k} (0,01)^k (0,99)^{86-k},$$

d'où  $\mathbb{P}(X \leq 3) = 0,9889$  (La France avait en effet remporté 2 médailles d'or).

### 3.4.3 Probabilité de succès et variable aléatoire géométrique

Toujours sur ce jeu de  $n$  lancers de Pile (P) ou Face (F), intéressons-nous au premier temps où nous allons obtenir un P. Pour ce faire, définissons la variable aléatoire  $T$  par

$$T(\omega) = \inf\{k \in \{1, \dots, n\}, X_k(\omega) = P\},$$

avec la convention  $\inf \emptyset = +\infty$ . La v.a.  $T$  est à valeurs dans  $\{1, 2, \dots, n, +\infty\}$  et

$$\mathbb{P}(T = k) = \mathbb{P}(X_1 = F, \dots, X_{k-1} = F, X_k = P) = (1-p)^{k-1}p, \quad 1 \leq k \leq n,$$

car nous avons supposé que nos lancers étaient indépendants. De même,

$$\mathbb{P}(T = +\infty) = \mathbb{P}(X_1 = F, \dots, X_n = F) = (1-p)^n.$$

Si nous faisons (heuristiquement) tendre le nombre de lancers  $n$  vers l'infini, la loi de la variable aléatoire  $T$  de succès du premier Pile est alors une probabilité définie par

$$\mathbb{P}(T = k) = p(1-p)^{k-1}, \quad \text{pour tout } k \in \mathbb{N}^*.$$

**Définition 3.24**  $X$  est une **variable géométrique** de paramètre  $p \in ]0, 1[$  si  $X$  est une variable aléatoire à valeurs dans  $\mathbb{N}^*$  telle que  $\forall k \in \mathbb{N}^*$ ,

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

Pour une variable aléatoire  $X$  de loi géométrique de paramètre  $p$ , on a :

$$\mathbb{E}(X) = \frac{1}{p}, \quad \text{Var}(X) = \frac{1-p}{p^2} \quad \text{et} \quad G_X(s) = \frac{ps}{1-(1-p)s}.$$

En effet, on calcule directement  $G_X(s) = \sum_{k \geq 1} p(1-p)^{k-1} s^k = \frac{ps}{1-(1-p)s}$ , et on déduit  $\mathbb{E}(X) = G'_X(1) = \frac{p}{(1-s+ps)^2} \Big|_{s=1} = \frac{1}{p}$ . Le calcul de la variance est similaire.

Ainsi, si on répète des épreuves indépendantes de Bernoulli avec même probabilité de succès  $p$  (obtention d'un Pile par exemple), jusqu'à l'obtention du premier succès, le nombre moyen des répétitions nécessaires est  $1/p$ . Il faut donc s'attendre à lancer 6 fois un dé équilibré avant d'obtenir le premier 1, en moyenne.

### 3.4.4 Variable aléatoire de Poisson

**Définition 3.25**  $X$  est une **v.a. de Poisson** de paramètre  $\theta > 0$  si  $X$  prend ses valeurs dans  $\mathbb{N}$  et si sa loi est caractérisée pour tout  $k \in \mathbb{N}$  par

$$\mathbb{P}(X = k) = e^{-\theta} \frac{\theta^k}{k!}. \quad (3.11)$$

Sa loi est une loi de Poisson de paramètre  $\theta > 0$ , que l'on note  $\mathcal{P}(\theta)$ .

Pour une v.a. de Poisson  $X$  de paramètre  $\theta$ , on a

$$\mathbb{E}(X) = \theta, \quad \text{Var}(X) = \theta \quad \text{et} \quad G_X(s) = e^{\theta(s-1)}. \quad (3.12)$$

En effet, on calcule directement que  $\mathbb{E}(X) = \sum_{k=0}^{\infty} k e^{-\theta} \frac{\theta^k}{k!} = \theta$ , puis  $\mathbb{E}(X(X-1)) = e^{-\theta} \sum_{k=2}^{\infty} k(k-1) \frac{\theta^k}{k!} = \theta^2$ , d'où  $\text{Var}(X) = \theta$ , vu que  $\mathbb{E}(X) = \theta$ . Le calcul de  $G_X$  se fait de la même façon, et on retrouve par dérivations successives au point 1 les expressions de  $\mathbb{E}(X) = \theta$  et  $\mathbb{E}(X(X-1)) = \theta^2$ .

**Exemple 3.26** *Supposons que le nombre d'erreurs  $Y$  par page d'un livre suit une loi de Poisson de paramètre  $\frac{1}{2}$ . Calculons la probabilité qu'il y ait au moins une erreur dans une page donnée :*

$$\mathbb{P}(Y \geq 1) = 1 - \mathbb{P}(Y = 0) = 1 - e^{-\frac{1}{2}} \approx 0,39.$$

#### La loi de Poisson comme limite de lois binomiales.

Considérons, pour  $n \in \mathbb{N}^*$  et  $a_n \in [0, 1]$ , la probabilité sur  $\mathbb{N}$  définie par

$$p_j(a_n, n) = \mathbf{1}_{\{j \leq n\}} \binom{n}{j} (a_n)^j (1 - a_n)^{n-j}, \quad j \in \mathbb{N}.$$

Ainsi,  $(p_j(a_n, n))_{j \in \mathbb{N}}$  est l'extension naturelle sur  $\mathbb{N}$  de la loi binomiale  $\mathcal{B}(n, a_n)$  de paramètre  $a_n$  et de taille  $n$ . Supposons alors que la suite  $(a_n)_n$  tende vers 0 quand  $n$  tend vers l'infini, de telle sorte que  $na_n \rightarrow \theta \in \mathbb{R}_+^*$ . En développant les combinaisons  $\binom{n}{j}$ , il est facile de vérifier que pour tout  $j \in \mathbb{N}$ ,

$$p_j(a_n, n) \xrightarrow{n \rightarrow \infty} p_j = e^{-\theta} \frac{\theta^j}{j!}.$$

La loi de Poisson modélise donc la probabilité du nombre d'apparitions d'un événement rare ( $a_n \sim \frac{\theta}{n}$  est petit) dans une grande suite d'événements ( $n$  grand). Ce résultat est très utile pour les calculs numériques, dans le cas où l'on souhaite

modéliser les occurrences d'un événement rare. Il permet de remplacer la loi binomiale par la loi de Poisson, ce qui conduit à des calculs beaucoup plus simples.

On peut citer beaucoup d'exemples de variables aléatoires qui peuvent être modélisées par une loi de Poisson (parce que l'on approxime ainsi une loi binomiale) :

- le nombre de centenaires dans une communauté humaine,
- le nombre de clients entrant dans un bureau de poste en l'espace d'un jour,
- le nombre de bactéries dans un volume de liquide fixé,
- le nombre de particules émises par un gramme de matière radioactive,
- le nombre d'objets défectueux trouvés pendant un contrôle de qualité.

Reprenons l'exemple 3.23. La variable aléatoire  $X$  peut être approchée par une variable aléatoire  $Z$  de loi de Poisson  $\mathcal{P}(0,86)$ . Comparons les probabilités :

$k$	$\mathbb{P}(X = k)$	$\mathbb{P}(Z = k)$
0	0,4213	0,4231
1	0,3660	0,3639
2	0,1571	0,1564
3	0,0444	0,0448

L'approximation est très bonne.

## 3.5 Lois conditionnelles et variables indépendantes

### 3.5.1 Lois marginales

Les notions de probabilités conditionnelles et d'événements aléatoires indépendants ont été introduites au chapitre 2. Dans ce paragraphe, nous considérons deux v.a.  $X$  et  $Y$  définies sur le même espace  $\Omega$  fini ou dénombrable, muni de la probabilité  $\mathbb{P}$ . Nous supposons que  $X$  et  $Y$  sont à valeurs respectivement dans  $\mathcal{X}$  et  $\mathcal{Y}$ , que nous pouvons supposer eux-mêmes finis ou dénombrables. Nous posons  $p_i^X = \mathbb{P}(X = x_i)$  pour  $x_i \in \mathcal{X}$ , et  $p_j^Y = \mathbb{P}(Y = y_j)$  pour  $y_j \in \mathcal{Y}$ .

La connaissance des deux lois  $\mathbb{P}^X$  et  $\mathbb{P}^Y$  ne donne pas d'information sur les liens qui peuvent unir les comportements aléatoires de  $X$  et de  $Y$ .

Il est plus intéressant de considérer le couple  $\mathbf{Z} = (X, Y)$  comme une variable aléatoire discrète à valeurs dans le produit cartésien  $\mathcal{X} \times \mathcal{Y}$ . Notons  $\mathbb{P}_{\mathbf{Z}} = (p_k^{\mathbf{Z}}, \mathbf{z}_k \in \mathcal{X} \times \mathcal{Y})$  la loi du vecteur aléatoire  $\mathbf{Z}$ . Alors

$$p_k^{\mathbf{Z}} = \mathbb{P}(\mathbf{Z} = \mathbf{z}_k) = \mathbb{P}(X = x_i \text{ et } Y = y_j) \text{ pour tous } \mathbf{z}_k = (x_i, y_j) \in \mathcal{X} \times \mathcal{Y}.$$



**Définition 3.27** Les lois  $\mathbb{P}^X$  et  $\mathbb{P}^Y$  de  $X$  et  $Y$  s'appellent **les lois marginales du couple  $(X, Y)$  de variables aléatoires  $X$  et  $Y$** .

L'ensemble  $\{X = x_i\}$  est la réunion (finie ou dénombrable) des ensembles deux-à-deux disjoints  $\{X = x_i, Y = y_j\} = \{\mathbf{Z} = (x_i, y_j)\}$ ,  $y_j \in \mathcal{Y}$ . La propriété de  $\sigma$ -additivité de la probabilité permet d'exprimer les lois marginales à partir de la loi jointe :

$$\mathbb{P}(X = x_i) = \sum_{y_k \in \mathcal{Y}} \mathbb{P}(\mathbf{Z} = (x_i, y_k)) \quad \text{pour tout } x_i \in \mathcal{X}, \quad (3.13)$$

$$\mathbb{P}(Y = y_j) = \sum_{x_k \in \mathcal{X}} \mathbb{P}(\mathbf{Z} = (x_k, y_j)) \quad \text{pour tout } y_j \in \mathcal{Y}. \quad (3.14)$$

### 3.5.2 Lois conditionnelles

Nous lançons en même temps un dé rouge et un dé bleu. Soient  $X$  le résultat du dé rouge et  $Y$  le résultat de la somme des deux dés. Il est clair que la connaissance de la valeur de  $X$  va influencer sur les valeurs possibles que peut prendre  $Y$  et sur sa loi. Par exemple, si  $X = 3$ , alors  $Y$  ne pourra prendre que des valeurs supérieures ou égales à 4, ce qui n'est pas le cas si  $X = 1$ . Il est donc naturel de s'intéresser, pour chaque valeur fixée  $x_i$  de  $X$ , à la loi de  $Y$  avec l'information a priori que  $X = x_i$ .

**Définition 3.28** Soient  $X$  et  $Y$  deux variables aléatoires définies sur le même espace de probabilité, à valeurs dans  $\mathcal{X}$  et  $\mathcal{Y}$ . Soit  $x_i \in \mathcal{X}$  tel que  $p_i^X = \mathbb{P}(X = x_i) > 0$ . On appelle **loi conditionnelle** de  $Y$  sachant  $X = x_i$  la probabilité sur  $\mathcal{Y}$  définie par

$$p_j^{Y|X=x_i} = \mathbb{P}(Y = y_j | X = x_i) = \frac{\mathbb{P}(\mathbf{Z} = (x_i, y_j))}{\mathbb{P}(X = x_i)}, \quad \forall y_j \in \mathcal{Y}. \quad (3.15)$$

Cette définition est cohérente avec (2.16). La loi conditionnelle de  $Y$  sachant  $X = x_i$  est caractérisée par  $\{(y_j, p_j^{Y|X=x_i}), y_j \in \mathcal{Y}\}$ . Ces lois conditionnelles sont a priori différentes pour chaque valeur de  $x_i$  et différentes de la loi marginale  $\mathbb{P}_Y$ . L'équivalence suivante découle immédiatement de (3.13)-(3.14) et la définition 3.28.

**Proposition 3.29** Il est équivalent de connaître les  $(p_k^Z : \mathbf{z}_k \in \mathcal{X} \times \mathcal{Y})$ , et les  $(p_i^X : x_i \in \mathcal{X})$  et  $(p_j^{Y|X=x_i} : y_j \in \mathcal{Y})$  pour  $x_i \in \mathcal{X}$  tels que  $p_i^X > 0$ .

### 3.5.3 Espérance conditionnelle

Pour  $i$  fixé tel que  $p_i^X > 0$ , la loi conditionnelle de  $Y$  sachant  $X = x_i$  donnée par  $\{(y_j, p_j^{Y|X=x_i}), y_j \in \mathcal{Y}\}$  définit une probabilité. Nous pouvons lui associer son

espérance, sa variance ou plus généralement ses moments, dès qu'ils existent. Commençons par l'espérance conditionnelle.

**Définition 3.30** Soit  $Y$  une v.a. intégrable. L'espérance conditionnelle de  $Y$  sachant  $\{X = x_i\}$  est l'espérance de la loi conditionnelle de  $Y$  sachant  $\{X = x_i\}$  :

$$\mathbb{E}(Y|X = x_i) = \sum_j y_j p_j^{Y|X=x_i} = \sum_j y_j \mathbb{P}(Y = y_j|X = x_i). \quad (3.16)$$

**Remarque 3.31** La série définie en (3.16) est bien convergente. En effet, nous pouvons utiliser l'observation suivante

$$\begin{aligned} \sum_i \mathbb{E}(|Y| |X = x_i) \mathbb{P}(X = x_i) &= \sum_i \sum_j |y_j| \mathbb{P}(X = x_i, Y = y_j) \\ &= \sum_j |y_j| \sum_i \mathbb{P}(X = x_i, Y = y_j) = \mathbb{E}(|Y|) < \infty, \end{aligned}$$

d'après le théorème de Fubini (voir (S9), section 2.4.4) et l'expression de la loi marginale (3.14) de  $Y$ . On en déduit que  $\mathbb{E}(|Y| |X = x_i) < \infty$  pour tout  $x_i$  tel que  $\mathbb{P}(X = x_i) > 0$ .

Cette espérance conditionnelle sachant  $\{X = x_i\}$  est une fonction de  $x_i$ , que nous noterons  $\psi(x_i)$ . Elle n'est définie que sur les valeurs possibles  $x_i$  de  $X$ . Nous pouvons alors plutôt considérer la fonction  $\psi(X)$  elle-même.

**Définition 3.32** L'espérance conditionnelle de  $Y$  sachant  $X$  est la v.a.

$$\mathbb{E}(Y|X) = \psi(X), \quad \text{avec} \quad \psi(x) = \begin{cases} \mathbb{E}(Y|X = x) & \text{si } \mathbb{P}(X = x) > 0 \\ 0 & \text{si } \mathbb{P}(X = x) = 0 \end{cases} \quad (3.17)$$

**ATTENTION :** Contrairement à l'espérance qui est un nombre réel, l'espérance conditionnelle de  $Y$  sachant  $X$  est une variable aléatoire, dépendant de l'aléa "à travers" la variable aléatoire  $X$ .

L'espérance conditionnelle étant définie comme l'espérance de la loi conditionnelle, elle hérite des propriétés usuelles de l'espérance :

- a)  $\mathbb{E}(aY + bZ | X) = a \mathbb{E}(Y | X) + b \mathbb{E}(Z | X)$
- b)  $\mathbb{E}(Y | X) \geq 0$  si  $Y \geq 0$
- c)  $\mathbb{E}(1 | X) = 1$ .

De plus,

$$\mathbb{E}(Y g(X) | X) = g(X) \mathbb{E}(Y | X) \quad (3.18)$$

est une généralisation de l'égalité a) ci-dessus, au cas où  $a = g(X)$ , qui doit être considéré "comme une constante" dans le calcul de l'espérance conditionnelle sachant  $X$ . ( $X$  est fixée comme une donnée connue *a priori*.)

**Théorème 3.33** Si  $Y$  est intégrable, alors  $\mathbb{E}(Y | X)$  est intégrable, et

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

**Preuve.** Nous avons

$$\begin{aligned} \mathbb{E}(\psi(X)) &= \sum_i \psi(x_i) p_X(x_i) = \sum_{i,j} y_j p^{Y|X=x_i}(y_j) p_X(x_i) \\ &= \sum_{i,j} y_j p_{X,Y}(x_i, y_j) = \sum_j y_j p_Y(y_j) = \mathbb{E}(Y). \end{aligned}$$

Nous avons utilisé ici le théorème de Fubini pour les séries. La justification de l'intégrabilité de  $\psi(X)$  et du théorème de Fubini sont montrées en utilisant le même calcul que ci-dessus où on a remplacé  $y_j$  par  $|y_j|$ .  $\square$

Ce résultat permet de calculer  $\mathbb{E}(Y)$  en conditionnant par une variable auxiliaire  $X$  :

$$\mathbb{E}(Y) = \sum_i \mathbb{E}(Y | X = x_i) \mathbb{P}(X = x_i)$$

Il généralise la formule des probabilités totales de la proposition 2.27, qui correspond ici à  $Y = \mathbf{1}_A$  et  $B_i = \{X = x_i\}$ .

**EXERCICE 3.1** Le nombre  $N$  de voitures passant devant une station d'essence en un jour suit la loi de Poisson de paramètre  $\lambda > 0$ . Chaque voiture décide de s'arrêter à la station avec probabilité  $p$  indépendamment des autres. On note  $K$  le nombre de voitures qui s'arrêtent à la station. Trouver  $\mathbb{E}(K)$ .

*Solution :* Nous avons  $p_N(n) = \frac{\lambda^n}{n!} e^{-\lambda}$  et  $p^{K|N=n}(k) = \binom{n}{k} p^k (1-p)^{n-k}$ . D'où  $\mathbb{E}(K | N = n) = \sum_k k p^{K|N=n}(k) = np$ , soit  $\mathbb{E}(K | N) = pN$ . D'après le théorème 3.33,  $\mathbb{E}(K) = \mathbb{E}(\mathbb{E}(K | N)) = \mathbb{E}(Np) = p\mathbb{E}(N) = p\lambda$ , puisque  $\lambda = \mathbb{E}(N)$ .

### 3.5.4 Variables aléatoires indépendantes

**Remarque :** Les formules (3.13) et (3.14) expriment les lois marginales  $\mathbb{P}_X$  et  $\mathbb{P}_Y$  de  $X$  et de  $Y$  en fonction de la loi  $\mathbb{P}_Z$  de  $Z$ . L'exemple suivant illustre que la connaissance des lois marginales ne suffit pas, en général, à retrouver la loi de  $Z = (X, Y)$ .

**Exemple 3.34** Considérons une variable aléatoire  $Z$  qui vaut  $(1, 1)$  et  $(-1, -1)$  avec probabilité  $\frac{p}{2}$ , et  $(1, -1)$  et  $(-1, 1)$  avec probabilité  $\frac{1-p}{2}$ , où  $p \in ]0, 1[$ . Alors, les variables

aléatoires  $X$  et  $Y$  prennent les valeurs 1 et  $-1$  avec probabilité  $\frac{1}{2}$ , et leur loi ne dépend donc pas du paramètre  $p \in ]0, 1[$  choisi. Pour le voir, nous avons calculé par exemple

$$\mathbb{P}(X = 1) = \mathbb{P}(\mathbf{Z} = (1, 1)) + \mathbb{P}(\mathbf{Z} = (1, -1)) = \frac{1}{2}.$$

Il convient alors d'étudier le cas où l'information sur  $X$  ne change rien à la loi de  $Y$ , généralisant ainsi la notion d'indépendance introduite pour les événements.

**Définition 3.35** Les v.a.  $X$  et  $Y$  sont dites **indépendantes** si pour tous événements  $A \subset \mathcal{X}$ ,  $B \subset \mathcal{Y}$  elles vérifient

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B). \quad (3.19)$$

Rappelons que  $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(X \in A \text{ et } Y \in B)$ .

**Proposition 3.36** Il y a équivalence entre :

- (i) Les v.a.  $X$  et  $Y$  sont indépendantes, de lois respectives  $\mathbb{P}_X$  et  $\mathbb{P}_Y$ .
- (ii) On a  $\mathbb{P}(\mathbf{Z} = (x_i, y_j)) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) = p_i^X p_j^Y$  pour tous  $x_i \in \mathcal{X}$ ,  $y_j \in \mathcal{Y}$ .
- (iii) On a  $p_j^{Y|X=x_i} = p_j^Y$  pour tout  $y_j \in \mathcal{Y}$  et tout  $x_i \in \mathcal{X}$  tel que  $p_i^X > 0$ .

**Remarque 3.37** (iii) signifie que la loi conditionnelle de  $Y$  sachant  $X = x_i$  est égale à la loi marginale de  $Y$ , ce qui correspond bien à l'idée intuitive d'indépendance. Bien entendu, comme la définition 3.35 de l'indépendance est symétrique en  $X$  et  $Y$ , nous pouvons ci-dessus échanger les v.a.  $X$  et  $Y$ .

**Preuve.** Pour obtenir (i)  $\Rightarrow$  (ii) il suffit de prendre  $A = \{x_i\}$  et  $B = \{y_j\}$  dans (3.19). Inversement, supposons (ii). En sommant par paquets dans la série à termes positifs, nous obtenons pour  $A \subset \mathcal{X}$  et  $B \subset \mathcal{Y}$  :

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \mathbb{P}(\mathbf{Z} \in A \times B) = \sum_{(x_i, y_j) \in A \times B} \mathbb{P}(\mathbf{Z} = (x_i, y_j)) \\ &= \sum_{x_i \in A} \sum_{y_j \in B} p_i^X p_j^Y = \sum_{x_i \in A} p_i^X \sum_{y_j \in B} p_j^Y = \mathbb{P}(X \in A)\mathbb{P}(Y \in B), \end{aligned}$$

et (i) s'en déduit. Enfin, l'équivalence (ii)  $\Leftrightarrow$  (iii) provient de la définition 3.28.  $\square$

**Proposition 3.38** Supposons les v.a.  $X$  et  $Y$  indépendantes. Soient  $f$  et  $g$  deux fonctions réelles sur  $\mathcal{X}$  et  $\mathcal{Y}$  respectivement, telles que  $f(X) \in L^1$  et  $g(Y) \in L^1$ . Alors le produit  $f(X)g(Y)$  est aussi intégrable et vérifie

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X)) \mathbb{E}(g(Y)). \quad (3.20)$$

**Preuve.** Exactement comme dans la démonstration précédente, nous pouvons écrire

$$\begin{aligned} \sum_{(x_i, y_j) \in \mathcal{X} \times \mathcal{Y}} |f(x_i) g(y_j)| p_{(x_i, y_j)}^Z &= \sum_{x_i \in \mathcal{X}, y_j \in \mathcal{Y}} |f(x_i) g(y_j)| p_i^X p_j^Y \\ &= \left( \sum_{x_i \in \mathcal{X}} |f(x_i)| p_i^X \right) \left( \sum_{y_j \in \mathcal{Y}} |g(y_j)| p_j^Y \right), \end{aligned}$$

qui est fini par hypothèse. Par suite, la variable aléatoire  $f(X)g(Y)$  est intégrable. En utilisant alors (S8) (section 2.4.4), la même démonstration montre que les égalités ci-dessus sont également vérifiées en enlevant les valeurs absolues, ce qui donne bien (3.20).  $\square$

**Exemple 3.39** *Considérons  $n$  v.a. de Bernoulli  $(X_i)_{i=1}^n$  indépendantes et de paramètre  $p \in ]0, 1[$ . Soient  $x_i \in \{0, 1\}$ , pour  $i \in \{1, \dots, n\}$ . La probabilité que la suite  $(X_1, \dots, X_n)$  soit égale à  $(x_1, \dots, x_n)$ , vaut*

$$\mathbb{P}(X_i = x_i, 1 \leq i \leq n) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}, \quad (3.21)$$

et nous retrouvons les résultats donnés au chapitre 2 (modèles d'urnes).

**Exemple 3.40** *On admet que le nombre de clients dans un bureau de poste pendant une journée est une v.a. de Poisson de paramètre  $\lambda > 0$ . Soit  $p$  la probabilité qu'une personne entrant dans le bureau de poste soit une femme. Dans ce cas, le nombre de femmes  $X$  et celui des hommes  $Y$  parmi les clients quotidiens sont des v.a. indépendantes, et suivent des lois de Poisson de paramètres respectifs  $\lambda p$  et  $\lambda(1-p)$ . Pour s'en assurer, le lecteur pourra utiliser les calculs de l'exemple 3.1.*

### 3.5.5 Somme de v.a. indépendantes

Les résultats suivants concernent la loi d'une somme de variables aléatoires indépendantes. Ce sont des résultats très utiles dans la pratique.

**Proposition 3.41** *Supposons que  $X$  et  $Y$  sont deux v.a. à valeurs dans  $\mathbb{Z}$  et notons  $Z = (X, Y)$ . Alors*

$$\mathbb{P}(X + Y = i) = \sum_{j \in \mathbb{Z}} \mathbb{P}(Z = (j, i - j)) = \sum_{j \in \mathbb{Z}} \mathbb{P}(Z = (i - j, j)). \quad (3.22)$$

En particulier si  $X$  et  $Y$  sont indépendantes, on a

$$\mathbb{P}(X + Y = i) = \sum_{j \in \mathbb{Z}} \mathbb{P}(X = j) \mathbb{P}(Y = i - j) = \sum_{j \in \mathbb{Z}} \mathbb{P}(X = i - j) \mathbb{P}(Y = j). \quad (3.23)$$

La loi de  $X + Y$  est donc obtenue grâce à un calcul de convolution discrète.

**Preuve.** (3.23) découle immédiatement de (3.22) et de (ii) de la proposition 3.36. Pour (3.22), il suffit d'appliquer (2.3) et le fait que  $\{X+Y = i\}$  est la réunion des ensembles deux-à-deux disjoints  $\{X = j, Y = i - j\}$  pour  $j \in \mathbb{Z}$ , et aussi des  $\{X = i - j, Y = j\}$  pour  $j \in \mathbb{Z}$ .  $\square$

**Remarque 3.42** Ces formules peuvent se généraliser à la somme de  $n$  variables aléatoires indépendantes. En particulier, (3.21) entraîne que la somme  $S = X_1 + \dots + X_n$  de  $n$  variables aléatoires indépendantes de loi de Bernoulli de paramètre  $p$  suit une loi binomiale  $\mathcal{B}(n, p)$ .

Ainsi, la loi binomiale  $\mathcal{B}(n, p)$  peut être interprétée comme la loi de la somme de  $n$  variables aléatoires indépendantes, de loi de Bernoulli de paramètre  $p$ .

**Proposition 3.43** Supposons les v.a.  $X$  et  $Y$  indépendantes, à valeurs dans  $\mathbb{N}$ , et  $U = X + Y$ . Notons  $G_X$ ,  $G_Y$  et  $G_U$  les fonctions génératrices de  $X$ ,  $Y$  et  $U$ . Nous avons alors pour tout  $s \in [0, 1]$

$$G_U(s) = G_X(s) G_Y(s). \quad (3.24)$$

**Preuve.** Il suffit de remarquer que pour  $s \in [0, 1]$ ,  $G_U(s) = \mathbb{E}(s^U) = \mathbb{E}(s^{X+Y})$  et  $G_X(s) = \mathbb{E}(s^X)$  et  $G_Y(s) = \mathbb{E}(s^Y)$ , et d'appliquer (3.20).  $\square$

Ce résultat permet d'identifier dans certains cas très facilement la loi d'une somme de variables aléatoires.

**Exemple 3.44** Soient  $X$  et  $Y$  des v.a. indépendantes de loi  $\mathcal{B}(n, p)$  et  $\mathcal{B}(m, p)$  (avec le même paramètre  $p$ ). D'après (3.10),  $U = X + Y$  vérifie

$$G_U(s) = (1 - p + ps)^n (1 - p + ps)^m = (1 - p + ps)^{n+m}.$$

En appliquant encore (3.10) et la proposition 3.17, nous en déduisons que  $X + Y$  suit la loi binomiale  $\mathcal{B}(n + m, p)$ .

**Exemple 3.45** Soient  $X$  et  $Y$  des v.a. indépendantes de loi de Poisson de paramètres respectifs  $\theta$  et  $\zeta$ . D'après (3.12),  $U = X + Y$  vérifie

$$G_U(s) = e^{\theta(s-1)} e^{\zeta(s-1)} = e^{(\theta+\zeta)(s-1)},$$

de sorte que  $X + Y$  suit la loi de Poisson de paramètre  $\theta + \zeta$ .

### 3.6 Exercices sur le chapitre 3

**EXERCICE 3.2** Un sauteur tente de franchir des hauteurs successives numérotées  $1, \dots, n, \dots$ . On suppose que les sauts sont indépendants les uns des autres, et que  $\mathbb{P}(\text{le sauteur réussit son } n\text{-ième saut}) = \frac{1}{n}$ .

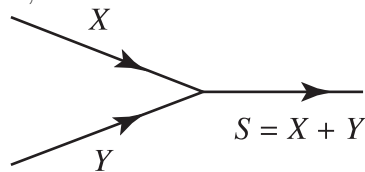
Soit  $X$  le dernier saut réussi. Quelle est la loi de  $X$ ? Calculer  $\mathbb{E}(X)$ ,  $\text{Var}(X)$ .

**EXERCICE 3.3** Si  $X$  est à valeurs dans  $\mathbb{N}$ , montrer que  $\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k)$ .

**EXERCICE 3.4** Le nombre d'accidents  $N$  en une semaine dans une usine est aléatoire d'espérance  $m$  et de variance  $\sigma^2$ . Le nombre d'individus  $X$  blessés dans un accident est aléatoire, d'espérance  $\mu$  et de variance  $\tau^2$ . Tous ces événements sont supposés indépendants.

Donner la fonction génératrice du nombre  $Y$  d'individus blessés par semaine, en fonction des fonctions génératrices de  $N$  et de  $X$ . En déduire la valeur des espérance et variance de  $Y$ , en fonction de  $m$ ,  $\sigma^2$ ,  $\mu$  et  $\tau^2$ .

**EXERCICE 3.5** On étudie le flux de véhicules durant une tranche horaire donnée à un raccordement de routes, décrit dans le dessin ci-dessous.



On note  $X$  (respectivement  $Y$ ), le nombre de véhicules empruntant la première (respectivement la deuxième) branche, et donc  $S = X + Y$  véhicules empruntent l'auto-route après le raccordement.  $X$  et  $Y$  sont modélisées par des variables aléatoires de loi de Poisson de paramètres respectifs  $\lambda > 0$  et  $\mu > 0$ . Les variables aléatoires  $X$  et  $Y$  sont supposées indépendantes.

Déterminer la loi de  $S$  et l'espérance conditionnelle de  $X$  sachant  $S$ .

**EXERCICE 3.6** Soit  $Z$  le nombre d'enfants d'une famille;  $X$  le nombre de filles et  $Y$  le nombre de garçons. Nous supposons que la probabilité qu'une famille ainsi choisie possède  $k$  enfants dont  $n$  filles, est donnée par :

$$p_{k,n} = \mathbb{P}(Z = k; X = n) = \frac{e^{-2} 2^k (0,52)^n (0,48)^{k-n}}{n!(k-n)!} \mathbf{1}_{\{0 \leq n \leq k\}}.$$

- 1) Montrer que  $Z$  et  $X$  ne sont pas indépendantes mais que  $Y$  et  $X$  le sont.
- 2) Donner la loi conditionnelle de  $X$  sachant  $Z = k$ . En déduire l'espérance conditionnelle de  $X$  sachant  $Z$ .

**EXERCICE 3.7** Considérons un jeu de Pile ou Face, qui associe 1 à Pile et 0 à Face. On appelle  $X_n$  le résultat du  $n$ -ième lancer et on suppose que

$$\mathbb{P}(X_n = 1) = p, \quad \text{où } p \in ]0, 1[.$$

- 1) Les événements  $A_n = \{X_{n-1} \neq X_n\}$ , pour  $n \geq 2$ , sont-ils indépendants? Discuter selon  $p$ .
- 2) On introduit la variable aléatoire  $T$  définie par

$$T(\omega) = \inf\{n, X_{n-1}(\omega) \neq X_n(\omega)\}$$

si cet ensemble est non vide et  $T(\omega) = +\infty$  sinon.

- a - Calculer  $\mathbb{P}(T = n)$ .
- b - Montrer que  $\mathbb{P}(T < +\infty) = 1$ .
- 3) On désigne par  $X_T$  la variable aléatoire définie par  $X_T(\omega) = X_{T(\omega)}(\omega)$ .  
Quand a-t-on  $\mathbb{P}((X_T, X_{T+1}) = (1, 0)) = \mathbb{P}((X_T, X_{T+1}) = (0, 1))$ ?

**EXERCICE 3.8** On note  $\mathcal{P}$  l'ensemble des nombres premiers différents de 1. On sait que tout  $x \in \mathbb{N}^*$  s'écrit de manière unique comme produit de puissances entières de nombres premiers de  $\mathcal{P}$ . Pour tout nombre entier  $p$ , il existe donc une fonction

$$U_p : \mathbb{N}^* \rightarrow \mathbb{N} \quad \text{telle que } \forall x \in \mathbb{N}^*, \quad x = \prod_{p \in \mathcal{P}} p^{U_p(x)}.$$

Soit  $\mathbb{Q}$  la probabilité sur  $\mathbb{N}^*$  définie par  $\mathbb{Q}(\{x\}) = \frac{c}{x^2}$ , ( $0 < c < \infty$ ).

- 1) Trouver la loi de  $U_p$ , pour chaque  $p \in \mathcal{P}$ .
- 2) Calculer  $\mathbb{Q}(U_p \geq n)$ , pour  $n \in \mathbb{N}$ .
- 3) Montrer que pour  $\mathbb{Q}$ , les variables  $(U_p)_p$  sont indépendantes.
- 4) Calculer la fonction génératrice de la v.a.  $U_p$ , ainsi que son espérance et sa variance.

**EXERCICE 3.9** Probabilités de Gibbs sur un système fini.

*Rappel :* Considérons  $\psi$  une fonction strictement convexe, et pour  $i \in \{1, \dots, n\}$ , des réels  $x_i$  et des réels positifs  $a_i$  dont l'un au moins est non nul. Alors

$$\psi \left( \frac{\sum_i a_i x_i}{\sum_i a_i} \right) \leq \frac{\sum_i a_i \psi(x_i)}{\sum_i a_i},$$

avec égalité si et seulement si tous les  $x_i$  sont égaux.

- 1) Soit un espace de probabilité fini  $\Omega = \{\omega_i, i = 1, \dots, n\}$  de cardinal  $n$ . On notera  $p$  une probabilité  $p = \{p_i, i = 1, \dots, n\}$  sur  $\Omega$  et on définit son entropie :

$$H(p) := - \sum_{i=1}^n p(\omega_i) \ln p(\omega_i) = - \sum_{i=1}^n p_i \ln p_i.$$



- 1-a) Vérifier que  $\phi : [0, 1] \rightarrow [0, +\infty[$ ,  $\phi(t) = -t \log t$  est strictement concave.  
 1-b) Montrer que  $H(p) = 0$  si et seulement si  $p$  est une mesure de Dirac sur  $\Omega$ .  
 1-c) Montrer que  $H(p) \leq \ln |\Omega| = \ln n$ .  
 2) Considérons une variable aléatoire réelle  $U$ , supposée non constante, et pour  $p$  une probabilité sur  $\Omega$ , nous noterons  $\langle U \rangle_p$  son espérance et  $\text{Var}_p(U)$  sa variance. Nous appellerons fonction de partition associée à l'énergie  $U$  la fonction

$$Z(\beta) = \sum_{i=1}^n e^{-\beta U(\omega_i)}, \quad \beta \in \mathbb{R}.$$

La probabilité de Gibbs associée est définie pour chaque  $\omega_i$  par  $\mu_\beta(\omega_i) := \frac{e^{-\beta U(\omega_i)}}{Z(\beta)}$ .

2-a) Que vaut  $\ln Z(0)$  ?

2-b) Montrer que quand  $\beta$  tend vers  $+\infty$ ,  $\mu_\beta$  devient une probabilité uniforme sur l'ensemble  $\Omega_{\min} := \{\omega; U(\omega) = \min_{\Omega} U\}$ , et que lorsque  $\beta \rightarrow -\infty$ ,  $\mu_\beta$  devient une probabilité uniforme sur  $\Omega_{\max} := \{\omega; U(\omega) = \max_{\Omega} U\}$ . En déduire que

$$\lim_{\beta \rightarrow +\infty} \langle U \rangle_{\mu_\beta} = \min_{\Omega} U \quad ; \quad \lim_{\beta \rightarrow -\infty} \langle U \rangle_{\mu_\beta} = \max_{\Omega} U. \quad (3.25)$$

2-c) Montrer que l'application définie sur  $\mathbb{R}$  par  $\beta \mapsto \ln Z(\beta)$  est de classe  $C^\infty$  et que

$$(\ln Z)'(\beta) = -\langle U \rangle_{\mu_\beta} \quad ; \quad (\ln Z)''(\beta) = \text{Var}_{\mu_\beta}(U).$$

En déduire que la fonction  $\beta \mapsto \ln Z(\beta)$  est strictement convexe.

3) Notre but est de rechercher une probabilité  $\mu$  sur  $\Omega$  qui maximise l'entropie  $p \rightarrow H(p)$  et telle que  $\langle U \rangle_\mu = E$ , où  $E \in ]\min_{\Omega} U, \max_{\Omega} U[$  est donné.

On admet que, par un théorème de Lagrange,  $\mu$  est un extremum de la fonction

$$p = (p_i; i \in \{1, \dots, n\}) \mapsto F(p) := H(p) - \beta (\langle U \rangle_p - E) - \lambda \sum_{i=1}^n p_i$$

où  $\beta$  est un paramètre à déterminer par la contrainte  $\langle U \rangle_\mu = E$  et  $\lambda$  un paramètre à déterminer par la contrainte que  $\mu$  est une probabilité. Ces paramètres sont les "multiplicateurs" de Lagrange.

3-a) Montrer qu'il existe un unique  $\beta = \beta(E) \in \mathbb{R}$  tel que  $\langle U \rangle_{\mu_\beta} = E$ .

3-b) Supposons que  $\mu_\beta$  maximise l'entropie. Montrer qu'alors son entropie vaut

$$H(\mu_\beta) = \ln Z(\beta) + \beta E.$$

3-c) Montrer que la fonction  $p \mapsto H(p) - \beta \langle U \rangle_p - \ln Z(\beta)$  est négative ou nulle, et qu'elle est nulle si et seulement si  $p = \mu_\beta$ .

En déduire que  $\mu_\beta$  est bien un maximum et que c'est en fait l'unique maximum.



## Chapitre 4

# Espérance mathématique de variables aléatoires réelles

*Je crois au hasard avec une obstination constante ; c'est même cela qui fait que lorsque je vois, je vois comme personne d'autre...*

*Nicolas de Stael.*

### 4.1 Probabilité uniforme et mesure de Lebesgue

Dans ce paragraphe, nous découvrons les difficultés importantes que l'on rencontre dans le cas de l'espace fondamental  $\Omega = [0, 1]^d$  muni de sa tribu Borélienne  $\mathcal{B}_{[0,1]^d}$ . Le cas unidimensionnel  $[0, 1]$  est l'exemple le plus simple d'un ensemble infini non dénombrable. Nous allons voir en particulier qu'il n'est pas possible de construire une probabilité uniforme qui s'étend sur  $\mathcal{P}([0, 1])$ , justifiant ainsi le besoin crucial d'introduire des tribus plus petites que  $\mathcal{P}([0, 1])$ .

Pour définir une probabilité uniforme  $\lambda : \mathcal{B}_{[0,1]} \rightarrow [0, 1]$ , il est naturel de commencer par le sous-ensemble  $\mathcal{A}_0 \subset \mathcal{B}_{[0,1]}$  constitué des parties  $A \subset ]0, 1]$  de la forme

$$A = \cup_{1 \leq i \leq n} ]a_i, b_i] \quad \text{pour } n \in \mathbb{N} \text{ et } 0 \leq a_1 \leq b_1 \leq \dots \leq a_n \leq b_n \leq 1. \quad (4.1)$$

Le candidat naturel pour une probabilité uniforme se doit d'être défini sur  $\mathcal{A}_0$  par :

$$\lambda_0(A) := \sum_{i=1}^n (b_i - a_i). \quad \text{pour tout } A \in \mathcal{A}_0 \text{ de la forme (4.1).} \quad (4.2)$$

L'objectif est maintenant d'obtenir une probabilité  $\lambda$  sur  $\mathcal{B}_{[0,1]}$  qui soit une extension de  $\lambda_0$ , c'est-à-dire  $\lambda$  vérifiant les propriétés de la définition 2.9 et  $\lambda(A) = \lambda_0(A)$  pour tout  $A \in \mathcal{A}_0$ .

**Théorème 4.1 (admis)** *Il existe une unique probabilité  $\lambda$  sur  $([0,1], \mathcal{B}_{[0,1]})$  qui soit une extension de l'application  $\lambda_0$  de (4.2).*

Cette probabilité est appelée **mesure de Lebesgue** sur  $[0,1]$  et a la propriété de ne pas charger les points, c'est-à-dire  $\lambda(\{x\}) = 0$  pour tout  $x \in [0,1]$ . Le résultat précédent repose sur l'application d'un résultat difficile de la théorie de la mesure dont nous rappelons l'énoncé dans le cadre d'une probabilité.

**Théorème 4.2 (Caratheodory, admis)** *Soient*

- $\mathcal{A}_0$  une algèbre sur  $\Omega$  (c'est-à-dire vérifiant (A1)-(A2) et (A3)<sub>f</sub> des définitions 2.2 et 2.3),
- et  $\mu_0 : \mathcal{A}_0 \rightarrow [0,1]$  une fonction vérifiant les conditions de  $\sigma$ -additivité et de masse totale unitaire de la définition 2.9.

*Alors il existe une unique probabilité  $\mu$  sur  $\mathcal{A} = \sigma(\mathcal{A}_0)$  telle que  $\mu = \mu_0$  sur  $\mathcal{A}_0$ .*

L'ensemble de cette construction s'étend au cas multidimensionnel.

**Théorème 4.3 (admis)** *Pour tout entier  $d \geq 1$  fixé, il existe une unique probabilité  $\lambda$  définie sur  $\mathcal{B}_{[0,1]^d}$  qui coïncide avec le volume sur les pavés :*

$$\lambda(\prod_{i=1}^d ]a_i, b_i]) = \prod_{i=1}^d (b_i - a_i), \quad \text{pour tous } 0 \leq a_i < b_i \leq 1, \quad i = 1, \dots, d.$$

*Cette probabilité  $\lambda$  s'appelle la **mesure de Lebesgue** sur  $[0,1]^d$ .*

La probabilité uniforme sur un ensemble borélien borné  $V$  de  $\mathbb{R}^d$  d'intérieur non vide se définit comme suit. La bornitude de  $V$  assure l'existence d'une constante  $r > 0$  tel que  $rV \subset [0,1]^d$ . On peut alors définir la probabilité uniforme sur  $V$  par :

$$\mathbb{P}_V(A) = \frac{\lambda(rA)}{\lambda(rV)} \quad \text{pour tout } A \in \mathcal{B}_V.$$

(On peut vérifier que le résultat ne dépend pas du choix de  $r$ ). La probabilité que le résultat tombe dans une partie  $A$  de  $V$  est proportionnelle au volume de cette partie. Si  $n = 1$  et  $A$  est un intervalle, le volume est la longueur de l'intervalle. Le cas de la loi uniforme sur  $[a, b]$  est donc le cas particulier où  $V = [a, b]$ .

Enfin, cette construction s'étend sur  $\mathbb{R}^d$  et permet de définir une mesure (de masse infinie) sur l'espace  $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$  qui coïncide avec le volume sur les pavés.

**Théorème 4.4 (admis)** *Pour tout entier  $d \geq 1$  fixé, il existe une unique mesure  $\lambda$  sur  $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ , appelée **mesure de Lebesgue** sur  $\mathbb{R}^d$ , telle que*

$$\lambda(\Pi_{i=1}^d ]a_i, b_i]) = \prod_{i=1}^d (b_i - a_i), \quad \text{pour tous } a_i < b_i, \quad i = 1, \dots, d.$$

*La mesure  $\lambda$  est invariante par translation : pour tout  $B \in \mathcal{B}_{\mathbb{R}^d}$  et  $\mathbf{b} \in \mathbb{R}^d$ , on a  $\lambda(\mathbf{b} + B) = \lambda(B)$ .*

**Pour aller plus loin...**

**Remarque 4.5** *Il n'est pas possible d'étendre la probabilité  $\lambda$  à  $\mathcal{P}([0, 1])$ . En particulier,  $\mathcal{B}_{[0,1]} \subsetneq \mathcal{P}([0, 1])$ . Nous allons en effet définir un ensemble (dit de Vitali) qui ne peut être mesuré par la probabilité  $\lambda$  définie dans le théorème 4.1.*

- *On introduit la relation d'équivalence sur  $[0, 1]$  :  $x \sim y$  si  $x - y \in \mathbb{Q}$ . On note  $(A_i)_{i \in I}$  les classes d'équivalence.*
- *En invoquant l'axiome du choix (car  $I$  non dénombrable), on identifie chaque  $A_i$  avec  $x_i \in A_i$ , et on note  $B := (x_i)_{i \in I}$ .*
- *On introduit  $B_n := r_n + B$ , où  $(r_n)_{n \in \mathbb{N}} = \mathbb{Q} \cap [-1, 1]$ . Alors*
  - *Les  $(B_n)_n$  sont deux à deux disjoints. En effet, si  $x \in B_n \cap B_p$ , alors il existe  $i_n$  et  $i_p$  tels que  $x = r_n + x_{i_n} = r_p + x_{i_p}$  ce qui implique que  $x_{i_n} \sim x_{i_p}$ , donc  $x_{i_n} = x_{i_p}$  et  $n = p$ .*
  - *Pour tout  $x \in [0, 1]$ , il existe  $i$  tel que  $x \in A_i$ , donc  $x = x_i + r$  pour un certain rationnel  $r \in \mathbb{Q} \cap [-1, 1]$ . Donc*

$$[0, 1] \subset \cup_n B_n \subset [-1, 2].$$

- *On fait un raisonnement par l'absurde. Supposons que  $B$  est mesurable. Alors  $\lambda(B_n) = \lambda(B)$  pour tout  $n$  car la mesure de Lebesgue  $\lambda$  est invariante par translation, et d'une part*

$$1 = \lambda([0, 1]) \leq \sum_{n \geq 1} \lambda(B_n) = \sum_{n \geq 1} \lambda(B),$$

*ce qui implique que  $\lambda(B) > 0$ , et d'autre part*

$$3 = \lambda([-1, 2]) \geq \sum_{n \geq 1} \lambda(B_n) = \sum_{n \geq 1} \lambda(B),$$

*ce qui implique que  $\lambda(B) = 0$ . Il y a contradiction, ce qui montre que  $B$  n'est pas mesurable.*

## 4.2 Variables aléatoires dans $\mathbb{R}^d$

Les v.a. que nous considérons maintenant peuvent être à valeurs dans  $\mathbb{R}^d$  tout entier (ou, dans ce paragraphe, dans n'importe quel espace topologique). Comme dans

le chapitre précédent, nous souhaitons pouvoir évaluer les chances de réalisation d'événements du type  $X^{-1}(B) = \{X \in B\}$  pour tout  $B \in \mathcal{B}_{\mathbb{R}^d}$ . Cela nous conduit à la définition suivante.

**Définition 4.6** Soit  $\Omega$  l'espace fondamental muni de la tribu  $\mathcal{A}$  des événements. On dit que  $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$  est une fonction mesurable ou une **variable aléatoire** si  $X^{-1}(B) \in \mathcal{A}$  pour tout  $B \in \mathcal{B}_{\mathbb{R}^d}$ .

Dans le cas où  $\Omega$  est fini ou dénombrable et muni de la tribu  $\mathcal{A} = \mathcal{P}(\Omega)$ , toute application de  $(\Omega, \mathcal{A})$  dans  $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$  est une variable aléatoire, comme on l'a défini dans le chapitre précédent.

**Exemple 4.7** (i) Soit  $A \subset \Omega$ . Alors  $X = \mathbf{1}_A$  (l'indicatrice de  $A$ ) est une v.a. si et seulement si  $A \in \mathcal{A}$  ( $A$  mesurable). En effet pour  $B \in \mathcal{B}_{\mathbb{R}}$ , on a  $X^{-1}(B)$  est égal à  $\emptyset$ ,  $A$ ,  $A^c$  ou  $\Omega$  selon que  $B \cap \{0, 1\}$  est égal à  $\emptyset$ ,  $\{1\}$ ,  $\{0\}$  ou  $\{0, 1\}$ .  
(ii) Une fonction continue  $f : (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}) \rightarrow (\mathbb{R}^{d'}, \mathcal{B}_{\mathbb{R}^{d'}})$  est mesurable. En effet, l'image réciproque de tout ouvert de  $\mathbb{R}^{d'}$  est un ouvert de  $\mathbb{R}^d$ , donc un élément de  $\mathcal{B}_{\mathbb{R}^d}$ .

Plus généralement, nous avons le résultat suivant :

**Proposition 4.8** (i) Soient  $X$  est une variable aléatoire à valeurs dans  $\mathbb{R}^d$  et  $f : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$  mesurable (par rapport aux tribus Boréliennes). Alors  $f(X)$  est une variable aléatoire.

(ii) Soit  $X : \Omega \rightarrow \mathbb{R}$ . Alors  $X$  est une v.a. ssi  $X^{-1}(] - \infty, a]) \in \mathcal{A}$  pour tout  $a \in \mathbb{R}$  ssi  $X^{-1}([a, \infty[) \in \mathcal{A}$  pour tout  $a \in \mathbb{R}$ .

(iii) Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de v.a. réelles (à valeurs dans  $\mathbb{R}$ ). Alors

$$\inf_{n \geq 1} X_n, \quad \sup_{n \geq 1} X_n, \quad \liminf_n X_n = \sup_{n \geq 1} \inf_{p \geq n} X_n \quad \text{et} \quad \limsup_n X_n = \inf_{n \geq 1} \sup_{p \geq n} X_n$$

sont des variables aléatoires. En particulier, si la suite  $(X_n(\omega))_n$  est convergente pour tout  $\omega \in \Omega$ , alors  $\lim_n X_n$  est une variable aléatoire.

**Preuve.** Pour le (i), il suffit de remarquer que pour tout  $B \in \mathcal{B}_{\mathbb{R}^{d'}}$ , on a  $f^{-1}(B) \in \mathcal{B}_{\mathbb{R}^d}$  d'après la mesurabilité de  $f$  et  $X^{-1}(B) = X^{-1}(f^{-1}(B)) \in \mathcal{A}$  d'après la mesurabilité de  $X$ .

Pour le (ii), notons  $\mathcal{R} = \{B \in \mathcal{B}_{\mathbb{R}} : X^{-1}(B) \in \mathcal{A}\}$ . Comme l'image réciproque  $X^{-1}$  commute avec la réunion, l'intersection et le passage au complémentaire, il est clair que  $\mathcal{R}$  est une tribu contenue dans  $\mathcal{B}_{\mathbb{R}}$ . Alors, si  $\mathcal{R}$  contient les intervalles  $] - \infty, a]$  pour tout  $a \in \mathbb{R}$ , on déduit de la proposition 2.8 que  $\mathcal{R} = \mathcal{B}_{\mathbb{R}}$ . On utilise le même argument pour montrer la deuxième équivalence.

Enfin, pour (iii), il suffit de remarquer que pour tout  $a \in \mathbb{R}$ , on a  $\{\sup_n X_n \leq a\} = \bigcap_n \{X_n \leq a\} \in \mathcal{A}$ , et on applique le résultat de (ii). On utilise un argument similaire pour  $\inf_n X_n$ . En appliquant ces deux résultats successivement, on déduit la mesurabilité de  $\liminf_n X$  et  $\limsup_n X$ . Enfin, si  $(X_n(\omega))_n$  est convergente pour tout  $\omega \in \Omega$ , on a  $\lim_n X_n = \liminf_n X = \limsup_n X$  est une variable aléatoire.  $\square$

### 4.3 Espérance des v.a. réelles

Dans cette partie, nous généralisons la notion d'espérance introduite au Chapitre 3 à toutes les v.a. réelles "pas trop grandes" (en un certain sens). L'idée naturelle est de se ramener au chapitre précédent en approchant  $X$  par une suite de v.a. **prenant un nombre fini de valeurs**.

**Remarque fondamentale :** Il est important de comprendre ici que l'on connaît bien l'espace d'arrivée d'une v.a. réelle ( $\mathbb{R}$ ) mais qu'on connaît mal son espace de départ (l'espace abstrait  $\Omega$ ). L'idée majeure, qui est à la base de la théorie de l'intégration de Lebesgue, est d'approcher nos v.a. par une suite obtenue par **découpage de l'espace d'arrivée**, et non pas de l'espace de départ (comme dans le cas de l'intégrale de Riemann).

Dans ce paragraphe, nous introduisons la notion d'**intégrale de Lebesgue** sur un espace abstrait muni d'une mesure de probabilité  $\mathbb{P}$ . Cependant, tous les arguments sont en fait valable dans le cadre plus général d'un espace  $(\Omega, \mathcal{A})$  muni d'une mesure.

#### 4.3.1 Espérance de v.a. positives

**Définition 4.9** Une v.a.  $Y$  est dite **étagée** si  $Y(\Omega)$  est fini, c'est-à-dire qu'elle prend un nombre fini de valeurs  $Y(\Omega) = \{y_1, \dots, y_n\}$  et on a

$$Y = \sum_{i=1}^n y_i \mathbf{1}_{A_i}, \quad \text{avec } A_i = \{Y = y_i\}.$$

On note  $\mathcal{E}_+$  l'ensemble des v.a. étagées positives.

Notons également  $L^0$  l'ensemble des v.a. réelles, et  $L_+^0$  le sous-ensemble des v.a. réelles positives. D'après (3.1), l'espérance d'une v.a. étagée est donnée par

$$\mathbb{E}(Y) = \sum_{y \in Y(\Omega)} y \mathbb{P}(Y = y) \quad \text{pour tout } Y \in \mathcal{E}_+. \quad (4.3)$$

Ici, il est commode d'autoriser la valeur  $+\infty$  pour  $Y$ , et on utilisera les règles de calcul  $0 \times \infty = \infty \times 0 = 0$ . Nous étendons à présent la définition de l'espérance à l'ensemble  $L_+^0$  des v.a. positives.

**Définition 4.10** Pour  $X \in L_+^0$ , l'espérance de  $X$  par rapport à  $\mathbb{P}$  est définie par

$$\mathbb{E}(X) = \sup \{ \mathbb{E}(Y) : Y \in \mathcal{E}_+ \text{ et } Y \leq X \}.$$

L'ensemble  $\{Y \in \mathcal{E}_+ : Y \leq X\}$ , dont la borne supérieure définit l'espérance, contient la fonction nulle. Notant par  $[\cdot]$  la partie entière (le plus grand minorant entier), on peut aussi construire des éléments non triviaux en introduisant la fonction

$$\alpha_n(x) = n \wedge 2^{-n} \lfloor 2^n x \rfloor, \quad x \in \mathbb{R}.$$

En effet, pour tout  $X \in L^0$  :

$$Y_n = \alpha_n(X), n \in \mathbb{N}^*, \text{ définit une suite croissante de } \mathcal{E}_+ \text{ qui converge vers } X. \quad (4.4)$$

Une illustration de l'approximation de  $X$  par la suite de v.a. étagées  $Y_n = \alpha_n(X)$  est donnée par la figure 4.3.1. Nous verrons un peu plus loin (Remarque 4.14) que cette approximation permet de donner une définition équivalente de l'espérance mathématique.

La définition de l'espérance implique immédiatement que

$$\mathbb{E}(cX) = c\mathbb{E}(X) \quad \text{pour tous } c \in \mathbb{R}_+ \text{ et } X \in L_+^0, \quad (4.5)$$

ainsi que la propriété de monotonie suivante.

**Lemme 4.11** Soient  $X_1, X_2 \in L_+^0$ . Si  $X_1 \leq X_2$ , alors  $0 \leq \mathbb{E}(X_1) \leq \mathbb{E}(X_2)$ . De plus  $\mathbb{E}(X_1) = 0$  si et seulement si  $X_1 = 0$   $\mathbb{P}$ -p.s.

**Preuve.** Pour la première partie, il suffit de remarquer que  $\{Y \in \mathcal{E}_+ : Y \leq X_1\} \subset \{Y \in \mathcal{E}_+ : Y \leq X_2\}$ . Pour la deuxième partie de l'énoncé, rappelons que  $\mathbb{P}(X > 0) = \lim \uparrow \mathbb{P}(X > n^{-1})$  d'après la proposition 2.12. Si  $\mathbb{P}(X > 0) > 0$ , on a  $\mathbb{P}(X > n^{-1}) > 0$  pour  $n$  assez grand. Alors  $X \geq Y = n^{-1} \mathbf{1}_{\{X > n^{-1}\}} \in \mathcal{E}_+$ , et on déduit de la définition de l'intégrale que  $\mathbb{E}(X) \geq \mathbb{E}(Y) = n^{-1} \mathbb{P}(X > n^{-1}) > 0$ .  $\square$

Le résultat à la base de la théorie de l'intégration est l'extension suivante de la propriété de convergence monotone de  $\mathbb{P}$  énoncée dans la proposition 2.12.

**Théorème 4.12 (convergence monotone)** Soit  $(X_n)_n \subset L_+^0$  une suite croissante  $\mathbb{P}$ -p.s., i.e. pour tout  $n \geq 1$ ,  $X_n \leq X_{n+1}$ ,  $\mathbb{P}$ -p.s. Alors

$$\mathbb{E}(\lim \uparrow X_n) = \lim \uparrow \mathbb{E}(X_n).$$



**Preuve.** On procède en deux étapes.

1) On commence par supposer que la suite est monotone  $X_n \leq X_{n+1}$  sur  $\Omega$  (et non seulement p.s. c'est à dire sur une partie de mesure pleine de  $\Omega$ ).

On note  $X := \lim \uparrow X_n$ . D'après le lemme 4.11, la suite des espérances  $(\mathbb{E}(X_n))_n$  hérite la croissance de la suite  $(X_n)_n$  et est majorée par  $\mathbb{E}(X)$ . Ceci montre l'inégalité  $\lim \uparrow \mathbb{E}(X_n) \leq \mathbb{E}(\lim \uparrow X_n)$ .<sup>1</sup>

Pour l'inégalité inverse, montrons que  $\lim \uparrow \mathbb{E}(X_n) \geq \mathbb{E}(Y)$  pour toute v.a. de  $\mathcal{E}_+$ ,  $Y = \sum_{i=1}^k a_i \mathbf{1}_{A_i} \in \mathcal{E}_+ \leq X$ . Pour  $c \in [0, 1[$ , on déduit du lemme 4.11 et de (4.5) que :

$$\mathbb{E}(X_n) \geq \mathbb{E}(X_n \mathbf{1}_{\{X_n \geq cY\}}) \geq c \mathbb{E}(Y \mathbf{1}_{\{X_n \geq cY\}}) = c \sum_{i=1}^k a_i \mathbb{P}(A_i \cap \{X_n \geq ca_i\}).$$

En utilisant la propriété de convergence monotone des mesures d'ensembles énoncée dans la proposition 2.12 (i), on obtient alors :

$$\lim \uparrow \mathbb{E}(X_n) \geq c \sum_{i=1}^k a_i \mathbb{P}(A_i) = c \mathbb{E}(Y) \longrightarrow \mathbb{E}(Y) \text{ quand } c \rightarrow 1.$$

2) Dans le reste de la preuve, on veut passer de la monotonie de la suite  $(X_n)_n$  sur  $\Omega$  à la monotonie  $\mathbb{P}$ -p.s. Pour cela, introduisons  $\Omega_0 = \{\omega \in \Omega : (X_n(\omega))_n \text{ croissante}\}$  et la suite croissante (sur  $\Omega$ )  $\tilde{X}_n := X_n \mathbf{1}_{\Omega_0}$ . La première étape de cette preuve s'applique à la suite  $(\tilde{X}_n)$ , alors il suffit de montrer que  $\mathbb{E}(\tilde{X}_n) = \mathbb{E}(X_n)$ . Comme  $\tilde{X}_n \leq X_n$ , l'inégalité  $\mathbb{E}(\tilde{X}_n) \leq \mathbb{E}(X_n)$  découle du lemme 4.11. Pour tout  $\varepsilon > 0$ , il existe  $Y_n^\varepsilon \in \mathcal{E}_+$  tel que  $Y_n^\varepsilon \leq X_n$  et  $\mathbb{E}(Y_n^\varepsilon) \geq \mathbb{E}(X_n) - \varepsilon$ . Notons  $\tilde{Y}_n^\varepsilon = Y_n^\varepsilon \mathbf{1}_{\Omega_0}$ . Alors  $\tilde{Y}_n^\varepsilon \in \mathcal{E}_+$  et vérifie  $\tilde{Y}_n^\varepsilon \leq \tilde{X}_n$ . Alors,  $\mathbb{E}(\tilde{Y}_n^\varepsilon) \leq \mathbb{E}(\tilde{X}_n)$ . Comme  $\mathbb{E}(\tilde{Y}_n^\varepsilon) = \mathbb{E}(Y_n^\varepsilon)$ , on déduit que  $\mathbb{E}(X_n) \leq \mathbb{E}(Y_n^\varepsilon) + \varepsilon \leq \mathbb{E}(\tilde{X}_n) + \varepsilon \searrow \mathbb{E}(\tilde{X}_n)$  pour  $\varepsilon \searrow 0$ .  $\square$

**Remarque 4.13** Par le même argument que l'étape 2 ci-dessus (approximation par les fonctions étagées (4.4) et utilisation du théorème de convergence monotone), on montre facilement que :

(i) Pour  $X_1, X_2 \in L_+^0$  telles que  $X_1 = X_2$   $\mathbb{P}$ -p.s., on a  $\mathbb{E}(X_1) = \mathbb{E}(X_2)$ . En particulier, pour toute constante  $a \in \mathbb{R}_+$ ,

$$\text{Si } X = a \text{ } \mathbb{P}\text{-p.s. alors } \mathbb{E}[X] = a.$$

(ii) Pour  $X_1, X_2 \in L_+^0$ , on a  $\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)$ .

**Remarque 4.14 (Définition équivalente de l'espérance)** Une conséquence directe du théorème de convergence monotone 4.12 (n'utilisant en fait que l'étape 1 de la preuve) est que l'espérance peut être définie de manière équivalente par :

$$\mathbb{E}(X) = \lim \uparrow \mathbb{E}(\alpha_n(X)) \quad \text{pour tout } X \in L_+^0,$$

1. Pour une première lecture, la première étape de cette preuve sera suffisante pour l'ensemble des résultats de ce chapitre.

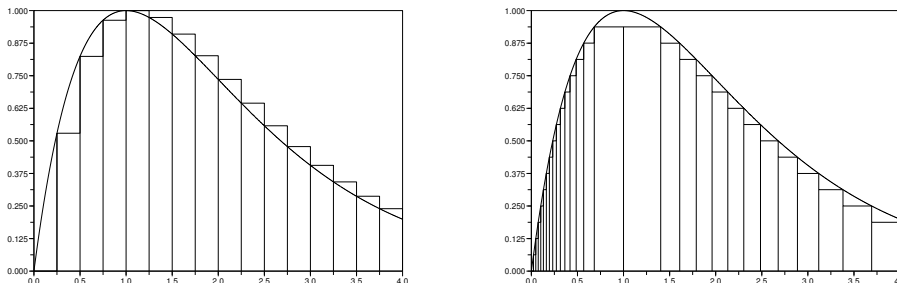


FIGURE 4.1 – Découpage “à la Riemann” (à gauche) ou “à la Lebesgue” (à droite), en 16 intervalles de même longueur de l’abscisse ou de l’ordonnée, respectivement.

où  $\alpha_n(X)$  est l’approximation de  $X$  introduite dans (4.4) qui approxime  $X$  par un découpage de l’espace d’arrivée. On aurait pu en fait utiliser n’importe quelle approximation monotone, la limite est toujours la même, à savoir l’espérance.

### 4.3.2 Espérance de v.a. réelles

Pour une v.a.  $X \in L^0$ , on introduit les v.a.  $X^+ := \max\{X, 0\}$  et  $X^- := \max\{-X, 0\}$  si bien que  $|X| = X^+ + X^-$  et  $X = X^+ - X^-$ .

**Définition 4.15** Une v.a.  $X \in L^0$  est dite  $\mathbb{P}$ -intégrable si  $\mathbb{E}(|X|) = \mathbb{E}(X^+) + \mathbb{E}(X^-) < \infty$ . Dans ce cas, son espérance est définie par

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) \quad \text{notée aussi} \quad \int_{\Omega} X(\omega) \mathbb{P}(d\omega).$$

On note par  $L^1$  l’ensemble des v.a.  $\mathbb{P}$ -intégrables (ou  $L^1(\Omega, \mathcal{A}, \mathbb{P})$  s’il convient de préciser l’espace de probabilité sous-jacent).

On voit immédiatement que  $L^1$  est un espace vectoriel et que

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y) \quad \text{pour tous } X, Y \in L^1 \text{ et } a, b \in \mathbb{R}. \quad (4.6)$$

Ceci est une conséquence immédiate de la validité de (4.5) et de la remarque 4.13 pour les v.a. positives.

**Remarque 4.16** Les fonctions Riemann intégrables sont Lebesgue intégrables. La réciproque n’est pas vraie (par exemple, la fonction  $f = \mathbf{1}_{\mathbb{Q} \cap [0,1]}$  est Lebesgue-intégrable,

mais n'est pas Riemann-intégrable). Montrons ceci dans  $\Omega = [0, 1]$  muni de la mesure de Lebesgue  $\lambda$ . Soit  $f$  une fonction Riemann intégrable bornée sur  $\Omega$  d'intégrale (au sens de Riemann)  $\int_0^1 f(x)dx$ . Alors  $f$  est Lebesgue intégrable et  $\mathbb{E}(f) = \int_0^1 f(x)dx$ . Si  $f$  est une fonction en escalier, ce résultat est trivial. Pour une fonction Riemann intégrable bornée  $f$  arbitraire, on peut trouver deux suites de fonctions en escalier  $(g_n)_n$  et  $(h_n)_n$  respectivement croissante et décroissante telles que  $g_n \leq f \leq h_n$  et

$$\inf_n \int_0^1 (g_n - h_n)(x)dx = \lim_{n \rightarrow \infty} \int_0^1 (g_n - h_n)(x)dx = 0.$$

Sans perte de généralité, on peut supposer  $h_n \leq 2 \max |f|$ . Les fonctions  $f_* := \sup_n g_n$  et  $f^* := \inf_n h_n$  sont boréliennes, et on a  $f_* \leq f \leq f^*$ . D'après théorème 4.12 de convergence monotone :

$$0 \leq \mathbb{E}(f^* - f_*) = \mathbb{E}(\inf_n (h_n - g_n)) \leq \inf_n \mathbb{E}(h_n - g_n) = 0,$$

et par suite  $f = f^* = f_*$ . Enfin,  $\mathbb{E}(f_*) = \lim \uparrow \mathbb{E}(g_n) = \lim \uparrow \int_0^1 g_n(x)dx = \int_0^1 f(x)dx$ .

**Remarque 4.17** Les outils développés tout au long de ce paragraphe ne sont pas spécifiques aux probabilités, et s'écrivent exactement de la même manière pour les mesures, c'est-à-dire les fonctions  $\sigma$ -additives de  $\mathcal{A}$  dans  $\mathbb{R}_+$  (sans condition de masse totale unitaire). Par exemple, les mêmes arguments (fonctions étagées, puis convergence monotone pour les fonction positives, puis décomposition en partie négative et positive) permettent de définir l'intégrale par rapport à la mesure de Lebesgue (introduite dans le théorème 4.4) :

$$\int_{\mathbb{R}^d} f(\mathbf{x})\lambda(d\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{x})dx_1 \dots dx_d \quad \text{pour tout } f \in L_+^0 \cup L^1(\lambda),$$

et le théorème 4.12 est valable dans ce cadre.

### Pour aller plus loin...

On peut lever une source d'ambiguïté concernant l'espérance d'une variable aléatoire réelle  $X$ . Pour cela, utilisons la notation plus précise  $L^1(\mathcal{A})$  qui met en évidence la dépendance par rapport à la tribu  $\mathcal{A}$ . Pour  $X \in L^1(\mathcal{A})$  et une partie  $A \in \mathcal{A}$ , on peut définir l'espérance de  $X\mathbf{1}_A$ , soit en intégrant la restriction  $X_A = X|_A$  par rapport à la restriction  $\mathbb{P}_A$  de  $\mathbb{P}$  à l'espace mesurable  $(A, \mathcal{A}_A)$ , où  $\mathcal{A}_A$  est la  $\sigma$ -algèbre définie par  $\mathcal{A}_A = \mathcal{P}(A) \cap \mathcal{A}$ .

**Proposition 4.18** Pour tout  $X \in L^1(\mathcal{A})$  et  $A \in \mathcal{A}$ , on a  $\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}_A(X_A)$ .

**Preuve.** Tout d'abord, cette propriété est vraie pour les fonctions  $X = \mathbf{1}_B$ ,  $B \in \mathcal{A}$ , puisque dans ce cas  $\mathbb{E}(\mathbf{1}_B\mathbf{1}_A) = \mathbb{P}(A \cap B) = \mathbb{E}_A(\mathbf{1}_B|_A)$ . Par linéarité, cette égalité

reste vraie pour les fonctions étagées, puis par convergence monotone pour les fonctions mesurables positives. Enfin, pour  $X \in L^1(\mathcal{A}, \mathbb{P})$ , on décompose  $X = X^+ - X^-$ , et on obtient le résultat voulu en appliquant l'égalité à  $X^+$  et  $X^-$ .  $\square$

## 4.4 Variables aléatoires de carré intégrable

### 4.4.1 Variance et Covariance

Outre l'espace  $L^1$ , nous définissons aussi, comme au Chapitre 3, l'espace  $L^2$  des variables aléatoires  $X$  dont le carré  $X^2$  est dans  $L^1$ .

**Définition 4.19** Une v.a. réelle  $X$  est de carré intégrable si la variable aléatoire  $X^2$  est intégrable, c'est-à-dire si son espérance  $\mathbb{E}(X^2)$  est finie.

**Proposition 4.20**  $L^2$  est un sous-espace vectoriel de  $L^1$ , et si  $X \in L^2$ ,

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

La preuve est identique à celle de la proposition 3.11.

**Définition 4.21** La **variance** d'une v.a. réelle  $X \in L^2$ , est  $\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$ , notée aussi  $\sigma_X^2$ , ou  $\sigma^2$  s'il n'y a pas d'ambiguïté. Sa racine carrée positive  $\sigma_X = \sqrt{\text{Var}(X)}$  s'appelle l'**écart-type** de  $X$ .

De même que l'espérance a été comparée au centre de gravité d'un ensemble de masses, la variance peut être rapprochée du concept mécanique de moment d'inertie (par rapport à l'espérance).

**Remarque 4.22** Soit  $X \in L^2$ .

(i) En utilisant la linéarité de l'espérance, nous obtenons encore la formule de Huygens : "la variance est égale à la moyenne des carrés moins le carré de la moyenne"

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (4.7)$$

(ii)  $\text{Var}(X) = 0$  si et seulement si  $X = \mathbb{E}(X)$ ,  $\mathbb{P}$ -p.s. La condition nécessaire découle du lemme 4.11. La condition suffisante est évidente.

Remarquons que si  $X$  et  $Y$  sont dans  $L^2$ , la variable aléatoire  $XY$  est dans  $L^1$ . En effet, il suffit d'écrire  $|XY| \leq \frac{1}{2}(X^2 + Y^2)$ . Nous pouvons alors définir la notion suivante.

**Définition 4.23** La covariance de deux v.a.  $X, Y \in L^2$  est :

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))). \quad (4.8)$$

Si  $\text{Var}(X)\text{Var}(Y) > 0$  (i.e.  $X, Y$  ne sont pas constantes,  $\mathbb{P}$ -p.s.), le **coefficient de corrélation** est le nombre

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}. \quad (4.9)$$

Par linéarité de l'espérance, nous obtenons

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y), \quad (4.10)$$

qui est extension de la formule de Huygens ci-dessus car  $\text{Var}(X) = \text{Cov}(X, X)$ . On obtient aussi que la covariance est une forme bilinéaire sur l'espace vectoriel des v.a. de carré intégrable, et nous avons

$$\text{Cov}(aX + b, a'Y + b') = aa'\text{Cov}(X, Y) \quad \text{pour tous } a, a', b, b' \in \mathbb{R}.$$

En particulier,

$$\text{Var}(aX + b) = a^2 \text{Var}(X), \quad (4.11)$$

et nous déduisons pour des v.a. non constantes  $\mathbb{P}$ -p.s. que  $\rho(X, Y) = \rho(aX + b, a'Y + b')$  lorsque  $aa' > 0$ .

Enfin, si  $X_1, \dots, X_n \in L^2$ , alors  $\sum_{i=1}^n X_i \in L^2$  et

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j). \quad (4.12)$$

**Remarque 4.24** Pour une v.a.  $X \in L^2$  non constante  $\mathbb{P}$ -p.s.

$$\frac{X - \mathbb{E}(X)}{\sigma_X} \quad \text{est d'espérance nulle et d'écart-type 1.}$$

Nous dirons que cette v.a. est **centrée et réduite**.

La notion de variance s'étend aux vecteurs aléatoires comme suit. Dans la suite  $|\cdot|$  désigne la norme Euclidienne dans  $\mathbb{R}^n$ ,  $\top$  l'opérateur de transposition et  $\cdot$  le produit scalaire correspondant.

**Définition 4.25**  $\mathbf{X} = (X_1, \dots, X_n)^\top : \Omega \longrightarrow \mathbb{R}^n$  est appelé **vecteur aléatoire** si ses composantes  $X_i$  sont des v.a. réelles. Un vecteur aléatoire sera toujours considéré sous forme de vecteur colonne dans  $\mathbb{R}^n$ .

(i) On dit que  $\mathbf{X}$  est intégrable, et on note  $\mathbf{X} \in L^1$ , si  $|\mathbf{X}| \in L^1$  (ou de manière équivalente  $X_i \in L^1$  pour tout  $i = 1, \dots, n$ ). Son espérance est donnée par le vecteur espérance  $\mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))^\top$ .

(ii) On dit que  $\mathbf{X}$  est de carré intégrable, et on note  $\mathbf{X} \in L^2$ , si  $|\mathbf{X}|^2 \in L^1$  (ou de manière équivalente  $X_i \in L^2$  pour tout  $i = 1, \dots, n$ ). Dans ce cas, sa **matrice de variance** ou de **covariance** est la matrice de taille  $n \times n$  symétrique  $\text{Var}(\mathbf{X}) = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq n}$ .

(iii) Soit  $\mathbf{Y}$  un deuxième vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ . Si  $\mathbf{X}$  et  $\mathbf{Y}$  sont de carré intégrable, leur covariance est définie par

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}((\mathbf{X} - \mathbb{E}\mathbf{X})(\mathbf{Y} - \mathbb{E}\mathbf{Y})^\top),$$

C'est une matrice réelle de taille  $n \times d$  de composantes  $\text{Cov}(\mathbf{X}, \mathbf{Y})_{ij} = \text{Cov}(X_i, Y_j)$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq d$ . En particulier,  $\text{Cov}(\mathbf{X}, \mathbf{X}) = \text{Var}(\mathbf{X})$  et  $\text{Cov}(\mathbf{Y}, \mathbf{X}) = \text{Cov}(\mathbf{X}, \mathbf{Y})^\top$ .

**Proposition 4.26** Pour un vecteur aléatoire  $\mathbf{X}$  dans  $\mathbb{R}^n$ , la matrice  $\text{Var}(\mathbf{X})$  est symétrique positive, et  $\text{Var}(\mathbf{A}\mathbf{X}) = \mathbf{A}\text{Var}(\mathbf{X})\mathbf{A}^\top$  pour toute matrice réelle  $\mathbf{A}$  de taille  $m \times n$ .

**Preuve.** La symétrie est évidente. En notant  $c_{i,j} := \text{Cov}(X_i, X_j)$ , on vérifie par un calcul direct que pour tout  $\mathbf{a} = (a_1, \dots, a_n)^\top \in \mathbb{R}^n$ , on a  $\mathbf{a}^\top \text{Var}(\mathbf{X}) \mathbf{a} = \sum_{i,j=1}^n a_i a_j c_{i,j} = \text{Var}(\sum_{i=1}^n a_i X_i) = \text{Var}(\mathbf{a} \cdot \mathbf{X}) \geq 0$ , ce qui prouve que la matrice  $\text{Var}(\mathbf{X})$  est positive. La formule pour  $\text{Var}(\mathbf{A}\mathbf{X})$  est obtenue par un calcul direct.  $\square$

## 4.4.2 Approximation linéaire

Pour des v.a. réelles  $X, Y \in L^2$ , nous souhaitons trouver la meilleure approximation de  $Y$  par une fonction affine de  $X$  de la forme  $aX + b$ , au sens des moindres carrés : il s'agit de déterminer les constantes  $a$  et  $b$  telles que  $\mathbb{E}(|Y - (aX + b)|^2)$  soit minimale. Grâce à la linéarité de l'espérance, on vérifie immédiatement que  $\mathbb{E}(|Y - (aX + b)|^2)$  est une fonction quadratique du couple  $(a, b) \in \mathbb{R}^2$  et qu'elle est convexe. Les minimiseurs sont donc caractérisés par la condition du premier ordre (annulation des dérivées partielles) qui fournit l'unique solution :

$$\hat{a} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \rho(X, Y) \frac{\sigma_Y}{\sigma_X} \quad \text{et} \quad \hat{b} = \mathbb{E}(Y) - a\mathbb{E}(X).$$

La meilleure approximation linéaire de  $Y$  basée sur  $X$  au sens de la distance quadratique moyenne est donc

$$\hat{Y} = \mathbb{E}(Y) + \rho(X, Y) \frac{\sigma_Y}{\sigma_X} (X - \mathbb{E}(X)).$$

Le carré moyen de l'erreur vaut alors

$$\mathbb{E}(|Y - \widehat{Y}|^2) = \sigma_Y^2 + \rho^2(X, Y)\sigma_Y^2 - 2\rho^2(X, Y)\sigma_Y^2 = \sigma_Y^2(1 - \rho^2(X, Y)).$$

Nous constatons que lorsque  $\rho(X, Y)$  est voisin de  $+1$  ou  $-1$ , le carré moyen de l'erreur est presque nul : plus  $|\rho(X, Y)|$  est proche de 1, meilleure est l'approximation.

Le résultat suivant étend la régression linéaire au cas multidimensionnel.

**Proposition 4.27 (Régression linéaire)** *Soient  $Y$  et  $\mathbf{X} = (X_1, \dots, X_d)$  des variables aléatoires dans  $L^2$  telles que la matrice  $\text{Var}(\mathbf{X})$  est inversible. Alors, le problème de minimisation des moindres carrés  $\min_{(\mathbf{a}, b) \in \mathbb{R}^d \times \mathbb{R}} \mathbb{E}(|Y - \mathbf{a} \cdot \mathbf{X} - b|^2)$  admet un unique minimiseur  $(\widehat{\mathbf{a}}, \widehat{b})$  donné par*

$$\widehat{\mathbf{a}} = \text{Var}(\mathbf{X})^{-1} \text{Cov}(\mathbf{X}, Y) \quad \text{et} \quad \widehat{b} = \mathbb{E}(Y) - \widehat{\mathbf{a}} \cdot \mathbb{E}(\mathbf{X}).$$

La meilleure approximation linéaire de  $Y$  par  $\mathbf{X}$  est

$$Y^{\text{reg}} := \mathbb{E}(Y) + \text{Var}(\mathbf{X})^{-1} \text{Cov}(\mathbf{X}, Y) \cdot (\mathbf{X} - \mathbb{E}(\mathbf{X})).$$

**Définition 4.28** L'équation  $y = \widehat{\mathbf{a}} \cdot x + \widehat{b}$ , où  $\widehat{\mathbf{a}}$  et  $\widehat{b}$  sont donnés par la proposition 4.27, définit **la droite des moindres carrés (ou de régression) de  $Y$  sur  $\mathbf{X}$** .

## 4.5 Des inégalités fameuses

### 4.5.1 Inégalité de Bienaymé-Chebyshev

**Théorème 4.29** *Soit  $p \geq 1$  et  $a > 0$ . Nous avons les inégalités suivantes*

- *Inégalité de Markov : soit  $X \in L^p$ , alors*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|^p)}{a^p}. \quad (4.13)$$

- *Inégalité de Bienaymé-Chebyshev : soit  $X \in L^2$ ,*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}. \quad (4.14)$$

**Preuve.** On a  $|X|^p \geq a^p \mathbf{1}_{[a, \infty[}(|X|)$ , donc

$$\mathbb{E}(|X|^p) \geq a^p \mathbb{E}(\mathbf{1}_{[a, \infty[}(|X|)) = a^p \mathbb{P}(|X| \geq a),$$

ce qui donne la première formule (inégalité de Markov). La seconde découle de la première appliquée à  $X - \mathbb{E}(X)$  et  $p = 2$ .  $\square$

**Remarque 4.30** *L'inégalité de Bienaymé-Chebyshev est très utile dans la pratique, comme nous le verrons par la suite. Elle permet de mesurer la probabilité des grands écarts entre  $X$  et sa moyenne. Par exemple, avec  $a = 10\sigma_X$ , il en résulte qu'il est improbable qu'une v.a.  $X$  dévie de  $\mathbb{E}(X)$  de plus de 10 fois son écart-type (probabilité inférieure à 0,01). Cette inégalité, tout à fait générale, n'est cependant pas très précise, et surestime très souvent en pratique le membre de gauche de (4.14), comme nous allons nous en convaincre ci-dessous.*

**Exemple 4.31** *Posons  $m = \mathbb{E}(X)$  et  $\sigma^2 = \text{Var}(X)$ . Alors, pour tout  $a > 0$ ,*

$$\mathbb{P}(|X - m| \geq a) \leq \frac{\sigma^2}{a^2},$$

soit encore  $\mathbb{P}(|X - m| \geq a\sigma) \leq \frac{1}{a^2}$ , d'où

$$\mathbb{P}(|X - m| < a\sigma) = \mathbb{P}(X \in ]m - a\sigma, m + a\sigma]) \geq 1 - \frac{1}{a^2}.$$

En particulier, pour  $a = 2$  et  $a = 3$ , nous avons respectivement :

$$\mathbb{P}(X \in ]m - 2\sigma, m + 2\sigma]) \geq 1 - \frac{1}{4} \approx 0,75; \quad \mathbb{P}(X \in ]m - 3\sigma, m + 3\sigma]) \geq 1 - \frac{1}{9} \approx 0,88.$$

Ces approximations sont universelles mais très grossières. Par exemple, si  $X$  suit une loi de Bernoulli de paramètre  $\frac{1}{2}$ , alors  $m = \frac{1}{2}$  et  $\sigma = \frac{1}{2}$  et on obtient précisément  $\mathbb{P}(X \in ]m - 2\sigma, m + 2\sigma]) = \mathbb{P}(X \in ]-\frac{1}{2}, \frac{3}{2}]) = 1$ .

## 4.5.2 Inégalité de Cauchy-Schwarz

**Proposition 4.32** *Soient  $X$  et  $Y$  deux variables aléatoires de carré intégrable, alors  $XY \in L^1$ , et on a l'inégalité de Cauchy-Schwarz :*

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}. \quad (4.15)$$

Il y a égalité dans (4.15) si et seulement si  $X$  et  $Y$  sont  $\mathbb{P}$ -p.s. proportionnelles.

**Preuve.** Comme  $|XY| \leq \frac{1}{2}(X^2 + Y^2)$ , nous avons  $XY \in L^1$  dès que  $X, Y \in L^2$ . De plus, pour tout  $x \in \mathbb{R}$ , nous avons d'après la linéarité et la positivité de l'espérance

$$x^2\mathbb{E}(X^2) + 2x\mathbb{E}(XY) + \mathbb{E}(Y^2) = E[(xX + Y)^2] \geq 0.$$

Mais ceci n'est possible que si ce trinôme en  $x$  a au plus une seule racine réelle. Son discriminant doit donc être négatif ou nul, ce qui donne immédiatement (4.15).



Le discriminant est nul si et seulement si il y a une racine double  $x_0$  et dans ce cas,  $Y(\omega) = -x_0X(\omega)$  pour presque tout  $\omega$ .  $\square$

Nous déduisons en particulier de l'inégalité de Cauchy-Schwarz, que le coefficient de corrélation défini par (4.9) vérifie

$$-1 \leq \rho(X, Y) \leq 1. \quad (4.16)$$

### 4.5.3 Inégalité de Jensen

**Théorème 4.33** Soient  $X$  une v.a. réelle intégrable et  $f : \mathbb{R} \rightarrow \mathbb{R}$  convexe.

- (i) Si  $f(X)$  est intégrable, alors on a  $\mathbb{E}(f(X)) \geq f(\mathbb{E}(X))$ .
- (ii)  $f(X)^- = -\min(f(X), 0)$  est intégrable et, en définissant  $\mathbb{E}(f(X)) = \mathbb{E}(f(X)^+) - \mathbb{E}(f(X)^-) \in \mathbb{R} \cup \{+\infty\}$ , l'inégalité du (i) reste vraie.

**Preuve.** Il suffit de montrer la seconde assertion, puisque la première en découle. Puisque  $f$  est convexe, elle est continue et  $f(X)$  est mesurable. Par ailleurs, la convexité de  $f$  implique que pour tout  $a \in \mathbb{R}$ , il existe  $\lambda_a \in \mathbb{R}$  tel que

$$f(x) \geq f(a) + \lambda_a(x - a) \quad \text{pour tout } x \in \mathbb{R}.$$

Appliquant cette inégalité à  $x = X(\omega)$ , on obtient d'une part que  $f(X)^- \leq (f(a) + \lambda_a(X - a))^- \in L^1$  du fait que  $X \in L^1$ , et d'autre part que  $f(X) \geq f(a) + \lambda_a(X - a)$ . L'inégalité voulue est obtenue en fixant  $a = \mathbb{E}(X)$  en prenant l'espérance.  $\square$

**Exemple 4.34** Si  $X$  est une v.a. centrée ( $\mathbb{E}(X) = 0$ ), alors

- par convexité de la fonction  $x \mapsto (x - K)^+$  pour tout  $\lambda \in \mathbb{R}$ , on a  $\mathbb{E}(e^{\lambda X}) \geq 1$ .
- par convexité de la fonction  $x \mapsto (x - K)^+$ , pour un paramètre  $K \in \mathbb{R}$ , on a  $\mathbb{E}[(x + X - K)^+] \geq (x - K)^+$  pour tout  $x \in \mathbb{R}$ .

**Exemple 4.35** Un investisseur a le choix : soit il place son capital dans une affaire risquée rapportant une somme aléatoire  $X$  d'espérance  $m$ , soit il le place en titres sans risque qui rapporteront  $m$  avec probabilité 1. Il va prendre sa décision de manière à maximiser l'espérance de  $u(R)$ , où  $R$  est son bénéfice et  $u$  sa fonction de préférence. L'inégalité de Jensen nous montre que si  $u$  est une fonction concave,  $\mathbb{E}(u(X)) \leq u(m)$ , ce qui rend le placement sûr préférable. Si par contre  $u$  est convexe, le placement risqué est meilleur puisque  $\mathbb{E}(u(X)) \geq u(m)$ .

## 4.6 Loi d'un vecteur aléatoire

**Définition 4.36** La loi  $\mathbb{P}_{\mathbf{X}}$  d'un vecteur aléatoire  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^d$  est la probabilité définie sur  $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$  par :

$$\mathbb{P}_{\mathbf{X}}(B) = \mathbb{P}(\mathbf{X} \in B) \text{ pour tout borélien } B \in \mathcal{B}_{\mathbb{R}^d}.$$

En théorie de la mesure,  $\mathbb{P}_{\mathbf{X}}$  est appelée mesure image de  $\mathbb{P}$  par  $\mathbf{X}$ .

Nous étudions dans ce paragraphe diverses caractérisations de cette probabilité.

### 4.6.1 Propriété de transfert

Le résultat suivant est une extension du théorème 3.14. Nous considérons une vecteur aléatoire  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^d$  et une fonction mesurable  $h$  de  $\mathbb{R}^d$  dans  $\mathbb{R}$ . Alors  $Y = h(\mathbf{X})$  est une v.a. réelle sur  $\Omega$ .

Nous dirons que  $h$  est intégrable si  $h$  appartient à  $L^1(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}, \mathbb{P}_{\mathbf{X}})$ .

**Proposition 4.37** Soient  $\mathbf{X}$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ .

(i) Pour toute fonction mesurable  $h : \mathbb{R}^d \rightarrow \mathbb{R}$ , on a que  $h$  est  $\mathbb{P}_{\mathbf{X}}$ -intégrable si et seulement si  $h(\mathbf{X})$  est  $\mathbb{P}$ -intégrable, et dans ce cas :

$$\mathbb{E}(h(\mathbf{X})) = \int_{\Omega} h(\mathbf{X}(\omega))\mathbb{P}(d\omega) = \int_{\mathbb{R}^d} h(\mathbf{x})\mathbb{P}_{\mathbf{X}}(d\mathbf{x}).$$

(ii) Soit  $\mu$  une probabilité sur  $\mathbb{R}^d$  telle que ,

$$\mathbb{E}(h(\mathbf{X})) = \int_{\mathbb{R}^d} h(\mathbf{x})\mu(d\mathbf{x}), \text{ pour toute fonction } h \text{ continue bornée,}$$

alors  $\mathbb{P}_{\mathbf{X}} = \mu$ .

**Preuve.** (i) Tout d'abord, le résultat est évident lorsque  $h$  ne prend qu'un nombre fini de valeurs, grâce à (4.3) et au fait que  $\mathbb{P}(\mathbf{X}^{-1}(B)) = \mathbb{P}_{\mathbf{X}}(B)$ , par définition de la loi de  $\mathbf{X}$ . Si  $h$  est une fonction positive, nous suivons la recette de la remarque 4.14 : nous l'approchons par la suite de fonctions étagées  $\alpha_n \circ h$  définie dans (4.4) et nous déduisons le résultat voulu grâce au théorème 4.12 de convergence monotone, les trois membres étant simultanément finis ou infinis. Le résultat général s'en déduit par décomposition de  $h = h^+ - h^-$ .

(ii) Pour tout  $\mathbf{x} \in \mathbb{R}^d$ , on introduit la fonction  $H^{\mathbf{x}} = \mathbf{1}_{]-\infty, \mathbf{x}]}$  de  $\mathbb{R}^d$  dans  $\mathbb{R}$ , avec

la convention  $] - \infty, \mathbf{x}] = ] - \infty, x_1] \times \cdots \times ] - \infty, x_d]$ . On pose pour tout  $n \geq 1$ ,  $\phi_n^{\mathbf{x}}(\mathbf{y}) = \prod_{i=1}^d (1 - n(y_i - x_i + \frac{1}{n}))^+$ ,  $\mathbf{y} \in \mathbb{R}^d$ . Alors  $\phi_n^{\mathbf{x}}$  est une fonction continue bornée, et  $(\phi_n^{\mathbf{x}})_n$  est une suite croissante qui converge simplement vers  $H^{\mathbf{x}}$ . Par hypothèse, nous avons que  $\mathbb{E}(\phi_n^{\mathbf{x}}(\mathbf{X})) = \int_{\mathbb{R}^d} \phi_n^{\mathbf{x}}(\mathbf{y}) \mu(d\mathbf{y})$ , ce qui implique que  $\mathbb{P}_{\mathbf{X}}(] - \infty, \mathbf{x}]) = \mathbb{E}(H^{\mathbf{x}}(\mathbf{X})) = \int_{\mathbb{R}^d} H^{\mathbf{x}}(\mathbf{y}) \mu(d\mathbf{y}) = \mu(] - \infty, \mathbf{x}])$  grâce au théorème de convergence monotone. Comme  $\mathbf{x} \in \mathbb{R}^d$  est arbitraire, ceci montre que  $\mathbb{P}_{\mathbf{X}}$  et  $\mu$  coïncident sur tous les pavés de la forme  $] - \infty, \mathbf{x}]$ , et en utilisant les propriétés de  $\sigma$ -additivité, on voit alors  $\mathbb{P}_{\mathbf{X}}$  et  $\mu$  coïncident sur tous les pavés de  $\mathbb{R}^d$ . Le théorème 4.2 de Carathéodory permet de conclure que  $\mathbb{P}_{\mathbf{X}}$  et  $\mu$  coïncident sur  $\mathcal{B}_{\mathbb{R}^d}$ .  $\square$

Ainsi, la loi de  $\mathbf{X}$  est caractérisée par son action  $h \mapsto \int h(\mathbf{x}) \mathbb{P}_{\mathbf{X}}(d\mathbf{x}) = \mathbb{E}(h(\mathbf{X}))$  sur toutes les fonctions mesurables intégrables  $h$ , et permet en effet de calculer  $\mathbb{E}(h(\mathbf{X}))$  par un calcul d'intégrale (au sens de Lebesgue) sur  $\mathbb{R}^d$ . La situation se simplifie encore plus si  $\mathbf{X}$  est une v.a. discrète, chargeant une suite de points  $(\mathbf{x}_n)_n$  avec probabilité  $p_n$ , et nous retrouvons que  $\int_{\mathbb{R}^d} h(\mathbf{x}) \mathbb{P}_{\mathbf{X}}(d\mathbf{x}) = \sum_n h(\mathbf{x}_n) p_n$ .

## 4.6.2 Fonction de répartition

Pour  $\mathbf{x} \in \mathbb{R}^d$ , on note  $] - \infty, \mathbf{x}] = ] - \infty, x_1] \times \cdots \times ] - \infty, x_d]$ .

**Définition 4.38** La fonction de répartition d'un vecteur aléatoire  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^d$  est :

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}_{\mathbf{X}}(] - \infty, \mathbf{x}]) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d), \text{ pour tout } \mathbf{x} \in \mathbb{R}^d. \quad (4.17)$$

**Proposition 4.39** La fonction de répartition  $F_{\mathbf{X}}$  d'un vecteur aléatoire  $\mathbf{X}$  caractérise sa loi  $\mathbb{P}_{\mathbf{X}}$ .

**Preuve.** Ce résultat est en fait déjà prouvé dans la deuxième moitié de la preuve de la proposition 4.37 (ii). Pour une meilleure compréhension, nous la reprenons ici de manière plus détaillée en dimension  $d = 1$ . D'après (4.17), nous avons  $\mathbb{P}_{\mathbf{X}}(]x, y]) = F_{\mathbf{X}}(y) - F_{\mathbf{X}}(x)$  pour tous  $x < y$ . Par suite, si  $B = \cup_{i=1}^n ]x_i, y_i]$ , avec  $x_i < y_i < x_{i+1}$ , nous avons  $\mathbb{P}_{\mathbf{X}}(B) = \sum_{i=1}^n \mathbb{P}_{\mathbf{X}}(]x_i, y_i]) = \sum_{i=1}^n (F_{\mathbf{X}}(y_i) - F_{\mathbf{X}}(x_i))$ , car les intervalles sont disjoints. Enfin,  $\mathbb{P}_{\mathbf{X}}(]x, +\infty]) = 1 - F_{\mathbf{X}}(x)$ .

Ainsi  $F_{\mathbf{X}}$  caractérise la restriction de  $\mathbb{P}_{\mathbf{X}}$  à l'ensemble de toutes les réunions finies d'intervalles disjoints de la forme  $]x, y]$  ou  $]x, +\infty[$ . Cet ensemble contient  $\mathbb{R}$ ,  $\emptyset$  et est stable par passage au complémentaire et par réunion finie, c'est donc une algèbre. Le théorème 4.2 de Carathéodory permet de conclure que  $F_{\mathbf{X}}$  caractérise  $\mathbb{P}_{\mathbf{X}}$ .  $\square$

La notion de fonction de répartition est communément utilisée pour les v.a. réelles car il est en général difficile de les caractériser dans le cas multi-dimensionnel.

**Exemple 4.40 Loi uniforme sur  $[0, 1]$ .** On considère sur  $([0, 1], \mathcal{B}_{[0,1]}, \lambda)$  la variable aléatoire  $X(\omega) = \omega$  pour tout  $\omega \in [0, 1]$ . Alors,  $\mathbb{P}_X = \lambda : X$  est de loi uniforme sur  $[0, 1]$ . Sa fonction de répartition est donnée par  $F_X(x) = x^+ \wedge 1$ , l'identité sur  $[0, 1]$ , nulle sur  $\mathbb{R}_-$  et égale à 1 sur  $[1, \infty[$ .

En général, une fonction de répartition sur  $\mathbb{R}$  n'est pas continue, comme illustré par l'exemple d'un variable aléatoire constante.

**Exemple 4.41** Si  $X = 0$  (presque-sûrement), la variable aléatoire constante nulle. Alors  $\mathbb{P}_X = \delta_0$  est la mesure de Dirac en 0, voir Exemple 2.20. La fonction de répartition est la fonction de Heaviside en 0 :  $H^0(x) = \mathbf{1}_{[0, +\infty)}(x)$  pour tout  $x \in \mathbb{R}$ .

**Proposition 4.42** (i) La fonction de répartition  $F_X$  d'une v.a. réelle  $X$  vérifie :

- (F1)  $F_X$  est croissante,
- (F2)  $F_X$  est continue à droite,
- (F3)  $\lim_{x \downarrow -\infty} F_X(x) = 0$  et  $\lim_{x \uparrow +\infty} F_X(x) = 1$ .

(ii) Soit  $F$  est une fonction réelle vérifiant les conditions (F1)-(F2)-(F3). Alors, il existe une unique probabilité  $\mu$  sur  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  dont  $F$  est la fonction de répartition :  $F(x) = \mu(] - \infty, x])$ . De plus, il existe une v.a.  $X$  définie sur l'espace de probabilité  $([0, 1], \mathcal{B}_{[0,1]}, \lambda)$  telle que  $F = F_X$ .

Remarquons qu'on ne peut pas, en général, définir la probabilité  $\mu$  de (ii) sur la tribu  $\mathcal{P}(\mathbb{R})$  de toutes les parties de  $\mathbb{R}$ , comme illustré par la remarque 4.5 dans l'exemple de la fonction de répartition de la loi uniforme de l'exemple 4.40. En fait, la preuve de (ii) ci-dessous montre que les "probabilités discrètes" du chapitre précédent sont les seules à pouvoir être définies sur la tribu  $\mathcal{P}(\mathbb{R})$ .

**Preuve.** (i) La croissance de  $F_X$  est immédiate. Puis, pour une suite  $x_n$  qui décroît vers  $x$ , on a  $] - \infty, x_n]$  décroît vers  $] - \infty, x]$  et donc  $F(x_n)$  décroît vers  $F(x)$  d'après la proposition 2.12. Enfin, on obtient les limites dans (F3) par le même argument du fait que  $] - \infty, x]$  décroît vers  $\emptyset$  (resp. croît vers  $\mathbb{R}$ ) lorsque  $x$  décroît vers  $-\infty$  (resp. croît vers  $+\infty$ ).

(ii) Sur l'espace  $([0, 1], \mathcal{B}_{[0,1]}, \lambda)$ , avec  $\lambda$  la mesure de Lebesgue, la variable aléatoire

$$\underline{X}(\omega) := \inf\{u \in \mathbb{R} : F(u) \geq \omega\}, \quad \omega \in [0, 1],$$

admet  $F$  pour fonction de répartition. Pour montrer cette assertion pour  $\underline{X}$ , il suffit de montrer que  $\{\omega \leq F(c)\} = \{\underline{X}(\omega) \leq c\}$ , puisqu'alors  $F(c) = \lambda(\{\underline{X} \leq c\}) = F_{\underline{X}}(c)$ . L'inclusion  $\{\omega \leq F(c)\} \subset \{\underline{X}(\omega) \leq c\}$  découle de la définition. Pour l'inclusion inverse, on observe que  $F(\underline{X}(\omega)) \geq \omega$ . En effet, si ce n'était pas le cas, on déduirait de la continuité à droite de  $F$  que  $F(\underline{X}(\omega) + \varepsilon) < \omega$  pour  $\varepsilon > 0$  assez petit, impliquant l'absurdité  $\underline{X}(\omega) + \varepsilon \leq \underline{X}(\omega)$  ! Avec cette observation et la croissance de  $F$ , on voit que  $\underline{X}(\omega) \leq c$  implique  $\omega \leq F(\underline{X}(\omega)) \leq F(c)$ , qui à son tour implique  $\omega \leq F(c)$ .  $\square$

**Remarque 4.43** Dans le cadre de la dernière preuve du (ii), la variable aléatoire  $\bar{X}(\omega) := \inf\{u : F(u) > \omega\}$  admet aussi  $F$  pour fonction de répartition. En effet, nous avons par définition que  $\omega < F(c)$  implique  $\bar{X}(\omega) \leq c$ . Mais  $\bar{X}(\omega) \leq c$  implique  $\underline{X}(\omega) \leq c$  puisque  $\underline{X} \leq \bar{X}$ . On en déduit que  $F(c) \leq \mathbb{P}[\bar{X} \leq c] \leq \mathbb{P}[\underline{X} \leq c] = F(c)$ .

Analysons enfin les points de continuité d'une fonction de répartition. Comme  $F$  est croissante, elle admet une limite à gauche en chaque point notée  $F(x-)$ . Comme  $]-\infty, y[ = \lim_n \uparrow ]-\infty, y_n]$  pour  $y_n \uparrow y$ , on a :

$$\begin{aligned} \text{pour } x < y, \quad \mathbb{P}_X(]x, y]) &= \mathbb{P}(x < X \leq y) = F_X(y) - F_X(x), \\ \mathbb{P}_X(]x, y[) &= \mathbb{P}(x < X < y) = F_X(y-) - F_X(x), \\ \mathbb{P}_X([x, y]) &= \mathbb{P}(x \leq X \leq y) = F_X(y) - F_X(x-), \\ \mathbb{P}_X([x, y[) &= \mathbb{P}(x \leq X < y) = F_X(y-) - F_X(x-). \end{aligned}$$

En particulier, les sauts de  $F$  sont caractérisés par  $\mathbb{P}_X(\{x\}) = F(x) - F(x-)$ .

**Proposition 4.44** Pour une v.a. réelle  $X$  et  $x \in \mathbb{R}$ , on a  $F_X$  continue en  $x$  si et seulement si  $\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x) = 0$ .

**Exemple 4.45 (Loi d'une v.a. à valeurs dans  $\mathcal{X} \subset \mathbb{R}$  dénombrable)** La loi  $\mathbb{P}_X$  de  $X$  est caractérisée pour tout  $x_i \in \mathcal{X}$  par  $p_i = \mathbb{P}_X(\{x_i\}) = \mathbb{P}(X = x_i)$ . Sa fonction de répartition est donnée par

$$F_X(x) = \sum_{x_i \in \mathcal{X}: x_i \leq x} p_i, \quad \text{pour tout } x \in \mathbb{R},$$

avec la convention qu'une somme "vide" vaut 0. La fonction  $F_X$  est en escalier, avec des sauts d'amplitude  $p_i$  en tout point  $x_i \in \mathcal{X}$ . Il s'agit d'une fonction **purement discontinue**, au sens où elle est complètement caractérisée par ses sauts  $\Delta F_X(x) = F_X(x) - F_X(x-)$ , via la formule  $F_X(x) = \sum_{y \leq x} \Delta F_X(y)$ .

Notons que l'ensemble  $\mathcal{X}$  du dernier exemple, quoiqu'au plus dénombrable, peut être dense dans  $\mathbb{R}$ . Par exemple il peut être égal à l'ensemble des rationnels  $\mathbb{Q}$ . Dans ce cas, si  $q_i > 0$  pour tout  $x_i \in \mathbb{Q}$ , la fonction  $F_X$  nous donne un exemple de fonction discontinue en tout nombre rationnel, et continue partout ailleurs.

### 4.6.3 Variables aléatoires indépendantes

**Définition 4.46** Soient  $\mathbf{X}$  et  $\mathbf{Y}$  deux vecteurs aléatoires à valeurs dans  $\mathbb{R}^d$  et  $\mathbb{R}^m$ , respectivement. On dit que  $\mathbf{X}$  et  $\mathbf{Y}$  sont **indépendants** si

$$\mathbb{P}(\mathbf{X} \in A, \mathbf{Y} \in B) = \mathbb{P}(\mathbf{X} \in A) \mathbb{P}(\mathbf{Y} \in B) \quad \text{pour tous } A \in \mathcal{B}_{\mathbb{R}^d} \text{ et } B \in \mathcal{B}_{\mathbb{R}^m}.$$

Plus généralement, les vecteurs aléatoires  $\mathbf{X}_1, \dots, \mathbf{X}_n$ , à valeurs dans  $\mathbb{R}^{d_1}, \dots, \mathbb{R}^{d_n}$ , sont **indépendants** (ou, “mutuellement indépendants”) si

$$\mathbb{P}(\mathbf{X}_1 \in A_1, \dots, \mathbf{X}_n \in A_n) = \prod_{i=1}^n \mathbb{P}(\mathbf{X}_i \in A_i) \quad \text{pour tous } A_i \in \mathcal{B}_{\mathbb{R}^{d_i}}, 1 \leq i \leq n.$$

**Proposition 4.47** Soient  $\mathbf{X}$  et  $\mathbf{Y}$  deux vecteurs aléatoires indépendants à valeurs dans  $\mathbb{R}^m$  et  $\mathbb{R}^n$ , respectivement. Alors pour toutes fonctions mesurables  $\mathbf{g} : \mathbb{R}^m \rightarrow \mathbb{R}^{m'}$  et  $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^{n'}$ , les vecteurs aléatoires  $\mathbf{g}(\mathbf{X})$  et  $\mathbf{h}(\mathbf{Y})$  sont aussi indépendants. Si de plus  $\mathbf{g}(\mathbf{X})$  et  $\mathbf{h}(\mathbf{Y})$  sont intégrables, alors  $\mathbf{g}(\mathbf{X})\mathbf{h}(\mathbf{Y})^\top$  (à valeurs dans les matrices réelles de taille  $m' \times n'$ ) est aussi intégrable, et

$$\mathbb{E}(\mathbf{g}(\mathbf{X})\mathbf{h}(\mathbf{Y})^\top) = \mathbb{E}(\mathbf{g}(\mathbf{X}))\mathbb{E}(\mathbf{h}(\mathbf{Y}))^\top. \quad (4.18)$$

Si de plus  $\mathbf{X}, \mathbf{Y} \in L^2$ , alors  $\text{Cov}(\mathbf{X}, \mathbf{Y}) = 0$ .

**Preuve.** La première assertion est évidente par définition même de l’indépendance. Pour montrer (4.18), il suffit de considérer le cas  $m' = n' = 1$ . Remarquons d’abord qu’elle se réduit à la définition 4.46 si  $g$  et  $h$  sont des fonctions indicatrices. Par linéarité, on obtient (4.18) lorsque  $g$  et  $h$  sont des fonctions étagées dans  $\mathcal{E}_+$ , puis pour les fonctions mesurables positives grâce au théorème 4.12 de convergence monotone. Enfin, si  $g$  et  $h$  sont de signe quelconque, (4.18) est vérifiée pour  $|g|$  et  $|h|$ , donc  $g(\mathbf{X})h(\mathbf{Y})$  hérite l’intégrabilité de  $g(\mathbf{X})$  et  $h(\mathbf{Y})$ , et en décomposant  $g = g^+ - g^-$  et  $h = h^+ - h^-$ , on obtient (4.18) en développant le produit. Enfin, si  $\mathbf{X}, \mathbf{Y} \in L^2$ , on obtient  $\text{Cov}(\mathbf{X}, \mathbf{Y}) = 0$  en appliquant (4.18) à  $\mathbf{g} = Id_{\mathbb{R}^m}$  et  $\mathbf{h} = Id_{\mathbb{R}^n}$ .  $\square$

Pour des v.a. réelles non constantes,  $\mathbb{P}$ -p.s. nous savons que  $|\rho(X, Y)| \leq 1$  par (4.16). Si  $X$  et  $Y$  sont indépendantes, alors  $\rho(X, Y) = 0$ . Si les v.a.  $X$  et  $Y$  sont indépendantes, le résultat précédent dit que  $\rho(X, Y) = 0$ ; au contraire si  $|\rho(X, Y)|$  est proche de 1, nous avons déjà remarqué que les variables aléatoires sont “fortement dépendantes”, d’où le nom de “coefficient de corrélation”. Attention cependant,  $\rho(X, Y) = 0$  n’implique pas que  $X$  et  $Y$  soient indépendantes.

**Corollaire 4.48** Soient  $X_j$ ,  $1 \leq j \leq n$ , des variables aléatoires réelles indépendantes et de carré intégrable. Alors  $\text{Cov}(X_i, X_j) = 0$  pour  $i \neq j$  et on déduit de (4.12) que :

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i). \quad (4.19)$$

**Définition 4.49** La suite  $(\mathbf{X}_n)_{n \in \mathbb{N}^*}$  de vecteurs aléatoires est dite **indépendante** si pour tout  $n$ , la famille finie  $\mathbf{X}_1, \dots, \mathbf{X}_n$  est indépendante.

Le résultat suivant est immédiat à vérifier.

**Proposition 4.50** *L'indépendance de la suite  $(\mathbf{X}_n)_n$  entraîne celle de*

- 1) *toute sous-suite  $(\mathbf{X}_{i_k})_k$ ,*
- 2) *toute suite de vecteurs issus de  $\mathbf{X}_n$ ,*
- 3) *toute suite  $(f_n(\mathbf{X}_n))_n$ , où les fonctions  $f_n$  sont des fonctions mesurables.*

**Exemple 4.51** *Nous considérons l'ensemble  $\Omega = [0, 1[$  muni de la tribu borélienne restreinte à cet ensemble, et de la mesure de Lebesgue. A chaque réel  $\omega$ , nous associons son développement dyadique (unique si l'on impose que les  $\omega_i$  ne soient pas tous égaux à 1 à partir d'un certain rang) :*

$$\omega = \sum_{i \geq 1} \frac{\omega_i}{2^i}, \quad \omega_i \in \{0, 1\}.$$

L'application  $X_i : \Omega \rightarrow \{0, 1\}$ , qui à  $\omega$  associe  $X_i(\omega) = \omega_i$  est une v.a. sur  $\Omega$ . En effet,

$$\{X_i = x_i\} = \bigcup_{x_1, \dots, x_{i-1} \in \{0, 1\}} \left[ \sum_{j=1}^i \frac{x_j}{2^j}, \sum_{j=1}^i \frac{x_j}{2^j} + \frac{1}{2^i} \right], \text{ pour } x_i \in \{0, 1\}, 1 \leq i \leq n,$$

qui est bien un élément de la tribu borélienne de  $\Omega = [0, 1[$ , et

$$\mathbb{P}(\{X_i = x_i\}) = \frac{1}{2^i} \sum_{x_1, \dots, x_{i-1} = 0}^1 1 = \frac{1}{2}.$$

Montrons l'indépendance des variables aléatoires  $(X_i)_{1 \leq i \leq n}$ . Nous avons

$$\bigcap_{1 \leq i \leq n} \{X_i = x_i\} = \left[ \sum_{i=1}^n \frac{x_i}{2^i}, \sum_{i=1}^n \frac{x_i}{2^i} + \frac{1}{2^n} \right], \text{ et donc } \mathbb{P} \left( \bigcap_{1 \leq i \leq n} \{X_i = x_i\} \right) = \frac{1}{2^n},$$

qui est la mesure de Lebesgue de l'intervalle ci-dessus. Cela démontre que les variables aléatoires  $X_i$  sont indépendantes et de loi de Bernoulli de paramètre  $\frac{1}{2}$ . Ainsi, nous venons de construire sur  $\Omega = [0, 1[$  une suite de v.a. telle que toute sous-suite finie est constituée de variables aléatoires indépendantes. C'est donc une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre  $\frac{1}{2}$ .

Le résultat suivant répond dans un cadre général au problème de la construction d'une suite de variables aléatoires indépendantes de lois données.

**Théorème 4.52 (admis)** *Soit  $\mu_n$  une probabilité sur  $\mathbb{R}^{d_n}$ ,  $d_n \in \mathbb{N}^*$ , pour tout  $n \in \mathbb{N}$ . Alors, il existe un espace  $\Omega$  muni d'une tribu  $\mathcal{A}$  et d'une probabilité  $\mathbb{P}$  sur lequel on peut définir une suite  $(\mathbf{X}_n)_n$  de vecteurs aléatoires indépendants, avec  $\mathbf{X}_n$  de loi  $\mu_n$  pour tout  $n \geq 0$ .*

## 4.7 Premiers éléments de simulation aléatoire

### 4.7.1 Génération de la loi uniforme

La loi d'une v.a. uniforme est donnée par la mesure de Lebesgue  $\lambda$  sur  $([0, 1], \mathcal{B}_{[0,1]})$ . Les réalisations d'une telle v.a. ont autant de chance d'être situées au voisinage de chaque point de  $]0, 1[$ .

Il n'est pas facile<sup>2</sup> de trouver un "bon" générateur de nombres au hasard. Les générateurs les plus simples sont fondés sur des procédés purement déterministes par une récurrence de la forme  $x_{n+1} = \phi(x_n)$  sur un intervalle d'entier  $I = [0, m[ \cap \mathbb{N}$ , pour une application  $\phi$  de  $I$  dans  $I$ . Les valeurs normalisées  $u_n = x_n/m$  sont comprises entre 0 et 1. La suite  $(u_n)$  est périodique de période  $p \leq m$ . La séquence  $(u_0, u_1, \dots, u_{p/10})$  devrait avoir les caractéristiques statistiques d'une suite de tirages indépendants de variables aléatoires de loi uniforme sur  $[0, 1[$  (il est conseillé de ne pas dépasser le dixième de la période). Ces générateurs n'ont donc en fait rien d'aléatoire et c'est pourquoi on les qualifie aussi de pseudo-aléatoires.

Sans être les plus récents, les générateurs congruentiels linéaires sont très utilisés. Ils produisent une suite d'entiers

$$x_{n+1} = ax_n + b \quad \text{modulo } m ,$$

à partir d'une valeur initiale  $x_0$ . Bien entendu, les entiers  $m, a$  et  $b$  doivent être soigneusement choisis (on exige au moins que la période  $p$  soit maximale, c'est-à-dire égale à  $m$ ).

Exemples de tels générateurs :

- la fonction `rand` de SCILAB utilise les valeurs  $m = 2^{31}$ ,  $a = 843314861$  et  $b = 453816693$  ;
- la méthode `Math.random` de JAVA utilise les valeurs  $m = 2^{48}$ ,  $a = 25214903917$  et  $b = 11$ .
- Par contre, la fonction `grand` de SCILAB est basée sur un générateur plus performant et plus moderne : le Mersenne twister<sup>3</sup> de période colossale  $2^{19937} - 1$ . C'est le même générateur que celui utilisé par la bibliothèque Numpy en PYTHON.

2. PARK & MILLER, Random number generators, good ones are hard to find, Commun. ACM (1988), **31**, p. 1192-1201.

3. [http://fr.wikipedia.org/wiki/Mersenne\\_Twister](http://fr.wikipedia.org/wiki/Mersenne_Twister)



### 4.7.2 Simulation par inversion de la fonction de répartition

Le résultat suivant donne une méthode de simulation d'une v.a. réelle  $X$  de loi quelconque à partir d'une v.a. uniforme  $U$  sur  $[0, 1]$ , générée comme ci-dessus. Ce résultat a en fait été démontré dans la preuve de la proposition 4.42 (iii) et repose sur l'utilisation de l'inverse continue à gauche d'une fonction de répartition

$$G(y) := \inf\{x \in \mathbb{R} : F(x) \geq y\}, \quad y \in [0, 1],$$

qui coïncide avec  $F^{-1}$  si  $F$  est inversible.

**Proposition 4.53** *Soit  $U$  une variable aléatoire de loi uniforme sur  $[0, 1]$ , et  $F$  une fonction de répartition. Alors la v.a.  $X = G(U)$  admet  $F$  pour fonction de répartition.*

**Exemple 4.54 Simulation d'une variable aléatoire discrète.** Soit  $\mathbb{P} = (p_1, p_2, \dots)$  une probabilité sur un ensemble dénombrable  $\{x_1, x_2, \dots\}$  avec  $x_1 < x_2 < \dots$ . La fonction de répartition associée est  $F(x) = \sum_{i: x_i \leq x} p_i$ . C'est une fonction en escalier, de même que  $G$  donnée par  $G(u) = x_i$  lorsque  $\sum_{j < i} p_j < u \leq \sum_{j \leq i} p_j$ . La proposition conduit alors à l'expression naturelle

$$X = x_i \quad \text{si} \quad \sum_{j < i} p_j < U \leq \sum_{j < i} p_j + p_i, \quad i \geq 1$$

qui définit une v.a.  $X$  de loi  $\mathbb{P}$ .

Ainsi, une variable aléatoire de Bernoulli  $X$  de paramètre  $p$  sera simulée ainsi :

- Si  $U \in [0, 1 - p]$ , nous posons  $X = 0$ ,
- Si  $U \in ]1 - p, 1]$ , nous posons  $X = 1$ .

**Remarque 4.55** *Le choix de la valeur de  $X$  lorsque  $U = 1 - p$  importe peu, car la probabilité que  $U = 1 - p$  est nulle du fait que la fonction de répartition de  $U$  n'a pas de saut, voir exemple 4.40 et la proposition 4.44.*

## 4.8 Exercices sur le chapitre 4

**EXERCICE 4.1** Soit  $X$  une v.a. de carré intégrable. Montrer que  $\mathbb{E}((X - a)^2)$  atteint son minimum lorsque  $a = \mathbb{E}(X)$ .



## Chapitre 5

# Variables aléatoires à densité

*Random numbers should not be generated with a method chosen at random*

*Donald Knuth.*

### 5.1 Notion de densité

Dans cette section, nous allons nous servir de la mesure de Lebesgue  $\lambda$  sur  $\mathbb{R}^d$  introduite dans le théorème 4.4, ainsi que l'intégrale par rapport à cette mesure introduite dans la remarque 4.17. Nous allons commencer par introduire la notion de fonction de densité qui permet de décrire une large classe de probabilités à partir de la mesure de Lebesgue  $\lambda$  sur  $\mathbb{R}^d$ .

**Définition 5.1** Soit  $f : (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  une fonction mesurable. On dit que  $f$  est une densité si  $f \geq 0$  et  $\int_{\mathbb{R}^d} f d\lambda = 1$ . Dans ce cas, on définit la fonction  $\mathbb{P}^f$  de  $\mathcal{B}_{\mathbb{R}^d} \rightarrow \mathbb{R}$  :

$$\mathbb{P}^f(A) = \int_{\mathbb{R}^d} f \mathbf{1}_A d\lambda = \int_A f(x) dx \quad \text{pour tout } A \in \mathcal{B}_{\mathbb{R}^d}.$$

On vérifie que  $\mathbb{P}^f$  définit une probabilité sur  $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ . En effet :

- $\mathbb{P}^f$  est de masse totale unitaire :  $\mathbb{P}^f(\mathbb{R}^d) = \int f \mathbf{1}_{\mathbb{R}^d} d\lambda = \int f d\lambda = 1$ ,
- $\mathbb{P}^f$  est  $\sigma$ -additive : pour une suite  $(A_n)_n \subset \mathcal{B}_{\mathbb{R}^d}$ , avec les  $A_n$  deux à deux disjoints, on a  $\mathbb{P}^f(\cup_n A_n) = \int f \mathbf{1}_{\cup_n A_n} d\lambda = \int f \sum_n \mathbf{1}_{A_n} d\lambda = \sum_n \int f \mathbf{1}_{A_n} d\lambda = \sum_n \mathbb{P}^f(A_n)$ , grâce à la proposition 2.12 (ii) et la positivité de  $f$ .

**Définition 5.2** (i) On dit que  $f$  est la densité de  $\mathbb{P}^f$  par rapport à  $\lambda$ , et on note  $\mathbb{P}^f = f \cdot \lambda$  et  $\mathbb{E}^f$  l'espérance mathématique sous  $\mathbb{P}^f$ .  
(ii) Une probabilité  $\mathbb{P}$  sur  $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$  est dite **à densité** s'il existe une densité  $f$  sur  $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$  telle que  $\mathbb{P} = f \cdot \lambda$ .  
(iii) Une v.a.  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^d$  est dite à densité si sa loi  $\mathbb{P}_{\mathbf{X}}$  est à densité.

**Théorème 5.3** Soient  $\mathbf{X}$  une v.a. dans  $\mathbb{R}^d$  et à densité  $f$  et  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  mesurable.  
(i) Si  $\phi$  est positive, on a  $\mathbb{E}(\phi(\mathbf{X})) = \int_{\mathbb{R}^d} (f\phi)(\mathbf{x}) d\mathbf{x}$ .  
(ii)  $\phi(\mathbf{X}) \in L^1$  si et seulement si  $f\phi \in L^1(\lambda)$ , et alors  $\mathbb{E}(\phi(\mathbf{X})) = \int_{\mathbb{R}^d} (f\phi)(\mathbf{x}) d\mathbf{x}$ .

**Preuve.** Si  $\phi = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$  est une fonction étagée, on a  $\mathbb{E}(\phi(\mathbf{X})) = \sum_{i=1}^n a_i \mathbb{P}_{\mathbf{X}}(A_i) = \sum_{i=1}^n a_i \int f \mathbf{1}_{A_i} d\lambda = \int f\phi d\lambda$ . Puis, pour  $\phi \in L^1_+$  on utilise l'approximation par les v.a. étagées  $Y_n = \alpha_n(\phi(\mathbf{X}))$  comme dans (4.4) pour laquelle nous avons  $\mathbb{E}^f(Y_n) = \int f Y_n d\lambda$ , et on conclut par le théorème 4.12 de convergence monotone que  $\mathbb{E}(\phi(\mathbf{X})) = \int f\phi d\lambda$ , ce qui termine la démonstration de (i). Enfin pour obtenir (ii), on procède à la décomposition  $\phi = \phi^+ - \phi^-$ .  $\square$

Pour les v.a.  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^d$ , on s'intéresse plus particulièrement aux cas où sa loi  $\mathbb{P}_{\mathbf{X}}$  est à densité.

**Proposition 5.4** Si une v.a.  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^d$  est à densité  $f$ , alors  
(i) sa fonction de répartition  $F_{\mathbf{X}}$  vérifie

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{]-\infty, \mathbf{x}] } f(\mathbf{y}) d\mathbf{y} = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(y_1, \dots, y_d) dy_1 \dots dy_d \quad (5.1)$$

pour tout  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ ,

(ii) Une fonction mesurable  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  est  $\mathbb{P}_{\mathbf{X}}$ -intégrable si et seulement si  $\int |\phi(\mathbf{x})| f(\mathbf{x}) d\mathbf{x} < \infty$  (i.e.  $\phi f$  est Lebesgue-intégrable), et on a

$$\mathbb{E}(\phi(\mathbf{X})) = \int_{\mathbb{R}^d} \phi(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} (\phi f)(x_1, \dots, x_d) dx_1 \dots dx_d \quad (5.2)$$

**Preuve.** L'application du résultat du théorème 5.3 (i) à la loi  $\mathbb{P}_{\mathbf{X}}$  pour  $Y = \mathbf{1}_{]-\infty, \mathbf{x}] }(\mathbf{X})$  donne  $\mathbb{E}(Y) = F_{\mathbf{X}}(\mathbf{x}) = \int_{]-\infty, \mathbf{x}] } f(\mathbf{x}') d\mathbf{x}'$ , soit (5.1). De même, (5.2) est une ré-écriture du théorème 5.3 (ii) à la v.a.  $Y = \phi(\mathbf{X})$ .  $\square$

La proposition suivante se démontre immédiatement en utilisant des résultats bien connus d'analyse.

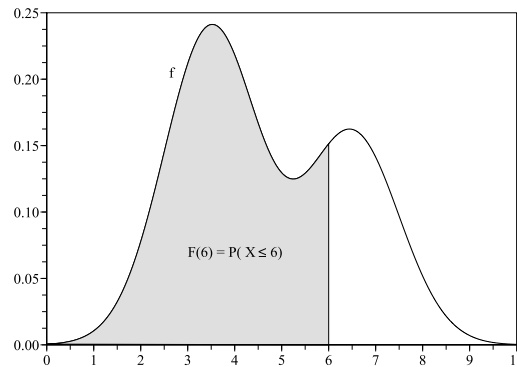


FIGURE 5.1 – Graphe de la densité  $f$  et de la fonction de répartition  $F$  en dimension 1.  $F(x)$  est la surface grise dans la Figure 5.1, délimitée par l'axe  $y = 0$ , le graphe de  $f$ , et la droite verticale d'abscisse  $x$  (ici,  $x = 6$ ).

**Proposition 5.5** Soit  $X$  une v.a. réelle de fonction de répartition  $F_X$ .

(i) Si  $X$  est à densité  $f$ , alors  $F_X$  est continue, de sorte que  $\mathbb{P}(X = x) = 0$  pour tout  $x \in \mathbb{R}$ ;  $F_X$  est dérivable en tout point  $x$  où  $f$  est continue et  $F'_X(x) = f(x)$ .

(ii) Réciproquement, si  $F_X$  est dérivable, ou seulement continue partout et dérivable par morceaux, alors  $X$  admet la densité  $F'(x) = f(x)$ .

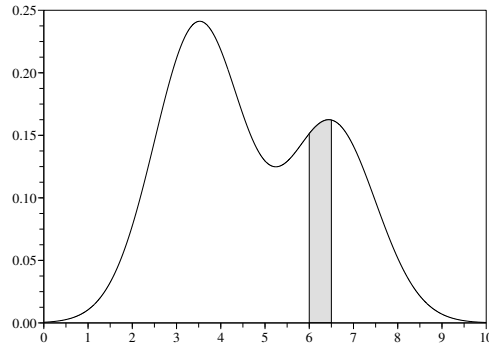
Ainsi pour une v.a. réelle  $X$  à densité continue, celle-ci a l'interprétation (en un point de continuité  $x$ )

$$f(x) \sim \frac{F(x + \Delta x) - F(x)}{\Delta x} = \frac{\mathbb{P}_X([x, x + \Delta x])}{\Delta x}. \quad (5.3)$$

La dernière proposition donne les expressions suivantes pour l'espérance et la variance d'une v.a. réelle à densité  $f$  (intégrable ou de carré intégrable) :

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{+\infty} x f(x) dx, \\ \text{Var}(X) &= \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f(x) dx = \int_{-\infty}^{+\infty} x^2 f(x) dx - \left( \int_{-\infty}^{+\infty} x f(x) dx \right)^2. \end{aligned}$$

**Remarque 5.6** Notons que la réciproque (ii) peut être améliorée (jusqu'à l'obtention d'une équivalence) en utilisant des outils de dérivation plus avancés qui n'ont pas encore été abordés dans votre scolarité. Si en effet  $F_X$  vérifie (5.1) pour  $d = 1$ , elle vérifie la même propriété en remplaçant  $f$  par  $g = f + \mathbf{1}_{\mathbb{Q}}$ ; alors  $g$  est encore une densité de  $\mathbb{P}_X$ . D'une manière générale, une densité est définie à un ensemble de mesure de Lebesgue nulle près.

FIGURE 5.2 –  $f(6) \sim (F(6,5) - F(6))/0,5$ 

**Exemple 5.7** La durée de fonctionnement d'un ordinateur avant sa première panne est une v.a. positive de densité donnée par

$$f(x) = \frac{1}{100} e^{-\frac{x}{100}} \mathbf{1}_{\mathbb{R}_+}(x), \quad x \in \mathbb{R}.$$

La probabilité que  $X$  soit comprise entre 50 et 150 heures est

$$\mathbb{P}(X \in [50, 100]) = \int_{50}^{100} \frac{1}{100} e^{-\frac{x}{100}} dx = \frac{\sqrt{e} - 1}{e} \approx 0,24.$$

La probabilité que l'ordinateur fonctionne moins de 100 heures est :

$$\mathbb{P}(X \leq 100) = \int_0^{100} \frac{1}{100} e^{-\frac{x}{100}} dx = \frac{\sqrt{e} - 1}{e} \approx 0,63.$$

**Remarque 5.8** Il existe des v.a. qui n'ont pas de densité : c'est le cas des variables aléatoires discrètes. Il existe des cas "mixtes" : soit  $f$  une fonction positive intégrable et d'intégrale strictement positive, et soit  $I \subset \mathbb{R}$  une partie finie ou dénombrable non vide, et des  $q_i > 0$ ,  $i \in I$ , tels que  $\int_{-\infty}^{+\infty} f(x)dx + \sum_{i \in I} q_i = 1$ ; alors, on peut définir la fonction de répartition

$$F(x) = \int_{-\infty}^x f(y)dy + \sum_{i \in I: i \leq x} q_i, \quad x \in \mathbb{R},$$

la probabilité  $\mathbb{P}_X$  associée n'admet pas de densité et n'est pas non plus discrète.

**Exemple 5.9** Soient  $X$  une v.a. de loi de densité  $f$  et  $a \in \mathbb{R}$  fixé. Considérons la v.a.  $Y$  définie par  $Y = \max(a, X)$ . La fonction de répartition de  $Y$  est donnée par :

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(a \leq y, X \leq y) = \mathbb{P}(X \leq y) \mathbf{1}_{[a, \infty[}(y) = F_X(y) \mathbf{1}_{[a, \infty[}(y).$$

Si  $F_X(a) > 0$ , alors  $F_Y$  n'est pas continue au point  $a$  et  $Y$  n'est pas à densité. Par contre, on peut exprimer la loi de  $Y$  sous forme d'une masse au point  $a$  et d'une partie à densité à droite de  $a$ .

## 5.2 Exemples classiques de v.a. réelles à densité

**Variable aléatoire uniforme.** Soient  $a < b$  deux réels,  $X$  est uniforme sur  $[a, b]$  si sa loi  $\mathbb{P}_X$  est de densité

$$f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x), \quad x \in \mathbb{R}.$$

En vertu de la proposition 5.5, nous pourrions aussi bien choisir  $f(a) = 0$  ou  $f(b) = 0$ . Au vu de l'interprétation (5.3), le fait que  $f$  soit constante sur  $[a, b]$  exprime que si nous simulons une variable aléatoire selon la probabilité  $\mathbb{P}_X$ , nous avons “autant de chances” de tomber au voisinage de chaque point de l'intervalle  $[a, b]$ . Cela explique le nom “uniforme”. Remarquons aussi que  $\mathbb{P}_X(\{x\}) = 0$  pour tout  $x$  (comme pour toutes les probabilités avec densité). Nous avons donc une probabilité **nette** de tomber exactement en un point  $x$  fixé à l'avance.

On vérifie immédiatement que  $X$  est de carré intégrable, et on calcule :

$$\mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2} \quad ; \quad \text{Var}(X) = \frac{(b-a)^2}{12}. \quad (5.4)$$

Avec la même chance de tirer “un quelconque point” de l'intervalle  $[a, b]$ , il est naturel d'obtenir comme espérance le milieu du segment.

**EXERCICE 5.1** A partir de 7h, les bus passent toutes les 15 minutes à un arrêt donné. Un usager se présente entre 7h et 7h30 à cet arrêt, l'heure exacte de son arrivée étant une variable uniforme sur cette période. Trouver la probabilité qu'il doive attendre moins de 5 minutes, puis plus de 10 minutes.

*Solution :* Soit  $X$  le nombre de minutes s'écoulant entre 7h et l'arrivée de l'usager. Alors  $X$  est uniforme sur  $[0, 30]$ . L'attente est inférieure à 5 mns si l'usager arrive entre 7h10 et 7h15 ou entre 7h25 et 7h30. La probabilité d'attendre moins de 5 mins est donc  $\mathbb{P}(10 < X < 15) + \mathbb{P}(25 < X < 30) = \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = \frac{1}{3}$ . La probabilité d'attendre plus de 10 mins vaut  $\mathbb{P}(0 < X < 5) + \mathbb{P}(15 < X < 20) = \frac{1}{3}$ .

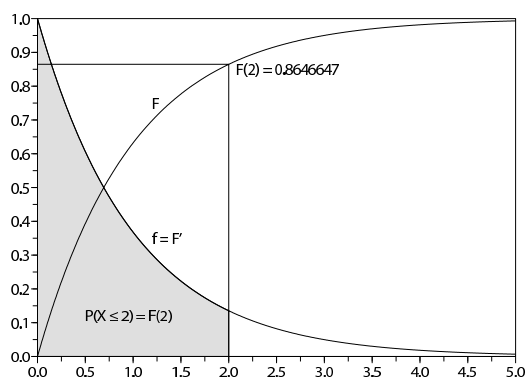


FIGURE 5.3 – Graphe de la densité  $f(x) = e^{-x}$  et de la fonction de répartition associée  $F(x) = 1 - e^{-x}$ .

**Variable aléatoire exponentielle.** On dit que  $X$  suit la loi exponentielle de paramètre  $\lambda > 0$ , que l'on note  $\mathcal{E}(\lambda)$ , si  $\mathbb{P}_X$  admet la loi de densité

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x), \quad x \in \mathbb{R}. \quad (5.5)$$

On calcule alors sa fonction de répartition, ainsi que son espérance et sa variance :

$$F(x) = (1 - e^{-\lambda x})^+, \quad x \in \mathbb{R}, \quad \mathbb{E}(X) = \frac{1}{\lambda}, \quad \text{et} \quad \text{Var}(X) = \frac{1}{\lambda^2}. \quad (5.6)$$

En effet, on calcule par par intégrations par parties successives que  $\mathbb{E}(X) = \int_0^\infty x \lambda e^{-\lambda x} dx$  et  $\mathbb{E}(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx$ .

Dans la pratique, la loi exponentielle modélise souvent une durée de vie ou le temps d'attente avant l'arrivée d'un événement spécifique. Par exemple la durée de vie d'une bactérie, la durée d'une conversation téléphonique ou le temps qui nous sépare du prochain tremblement de terre peuvent être considérées comme des v.a. de loi exponentielle.

**EXERCICE 5.2** Supposons que la durée d'une conversation téléphonique mesurée en minutes suit une loi  $\mathcal{E}(\lambda)$  de paramètre  $\lambda = \frac{1}{10}$  (exprimé en minute<sup>-1</sup>). Vous arrivez à une cabine téléphonique et quelqu'un entre juste devant vous. Avec quelle probabilité devez-vous attendre plus de 10 minutes ? entre 10 et 20 minutes ?

*Solution :*  $X$  désigne la durée de la conversation de la personne précédente. Alors, on calcule que  $\mathbb{P}(X > 10) = \int_{10}^{+\infty} \frac{1}{10} e^{-\frac{x}{10}} dx \approx 0.368$  et  $\mathbb{P}(10 < X < 20) \approx 0,233$ .

La loi exponentielle possède la propriété importante dite de non vieillissement, on dit aussi



qu'elle est **sans mémoire** au sens que :

$$\mathbb{P}_X(X > s) > 0 \text{ et } \mathbb{P}_X(X > t + s | X > t) = \mathbb{P}(X > s) \text{ pour tous } s, t \in \mathbb{R}. \quad (5.7)$$

**Proposition 5.10** *La loi exponentielle est l'unique loi vérifiant (5.7).*

**Preuve.** Soit  $G(t) = \mathbb{P}(X > t) = 1 - F_X(t)$ . D'après (2.16), la propriété de l'énoncé équivaut à dire que  $G(t+s) = G(t)G(s)$  pour tous  $s, t > 0$ . Comme  $G$  est décroissante et continue à droite et tend vers 0 à l'infini, cela revient aussi à dire que c'est une exponentielle négative, de la forme  $G(t) = e^{-\lambda t}$  pour un  $\lambda > 0$  [Poser  $\alpha = G(1)$ , vérifier par récurrence que  $G(n) = \alpha^n$  pour tout  $n$  entier, puis que  $G(q) = \alpha^q$  pour tout  $q$  rationnel, puis conclure que  $G(t) = \alpha^t$  pour tout  $t \in \mathbb{R}_+$  par la continuité à droite]. Le résultat s'obtient alors en comparant à (5.6) et en utilisant le fait qu'une fonction de répartition caractérise la loi à laquelle elle est associée.  $\square$

Représentons-nous  $X$  comme la durée de vie d'un certain instrument. La propriété ci-dessus signifie que si l'instrument fonctionne après  $t$  heures de service, la loi de sa durée de vie à partir de là est la même que la loi de la durée de vie de l'appareil neuf. L'appareil fonctionne sans mémoire du temps d'usage déjà écoulé.

Enfin, on peut simuler une loi exponentielle en utilisant la méthode d'inversion de la fonction de répartition du paragraphe 4.7.2. En effet, on calcule immédiatement que

$$F^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u) \quad u \in ]0, 1[.$$

Soit  $U$  est une variable aléatoire uniforme, alors les variables aléatoires

$$X' = -\frac{1}{\lambda} \ln(1 - U) \quad \text{et} \quad X = -\frac{1}{\lambda} \ln U$$

suivent une loi exponentielle de paramètre  $\lambda$  (car  $U$  et  $1 - U$  ont même loi).

**Loi gamma.** Une variable aléatoire  $X$  suit une loi gamma de paramètre d'échelle  $\theta$  et d'indice  $\alpha$ , que l'on note  $\Gamma(\alpha, \theta)$ , si sa loi est à densité :

$$f(x) = \frac{1}{\Gamma(\alpha)} \theta^\alpha x^{\alpha-1} e^{-\theta x} \mathbf{1}_{\mathbb{R}_+}(x), \quad x \in \mathbb{R},$$

où la fonction gamma est définie par :

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad \text{pour tout } \alpha > 0. \quad (5.8)$$

Une intégration par parties montre que  $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ , et on a de manière évidente  $\Gamma(1) = 1$ . Il s'ensuit que  $\Gamma(n + 1) = n!$  pour tout entier  $n \geq 0$ , avec la convention que  $0! = 1$ . Enfin, le changement de variable  $x \mapsto \theta x$  montre que la fonction  $f$  ci-dessus est d'intégrale égale à 1, et est donc bien une densité.

Remarquons que  $\Gamma(1, \theta) = \mathcal{E}(\theta)$ . Pour une loi  $\Gamma(\alpha, \theta)$ , on calcule aussi que :

$$\mathbb{E}(X) = \frac{\alpha}{\theta}, \quad \sigma_X^2 = \frac{\alpha}{\theta^2} \quad \text{et} \quad \mathbb{E}(X^\beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} \frac{1}{\theta^\beta}.$$

En revanche si  $\beta \leq -\alpha$ , nous avons  $\mathbb{E}(X^\beta) = +\infty$ .

Pour  $\alpha = n \in \mathbb{N}$ , la loi gamma représente la loi du temps d'attente avant la  $n$ -ième occurrence d'événements indépendants de lois  $\mathcal{E}(\theta)$  (voir Exemple 7.9).

**Variables aléatoires normales (ou variables gaussiennes).** Une **variable aléatoire normale centrée réduite** est une variable aléatoire  $X$  de loi notée  $\mathcal{N}(0, 1)$  de densité

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}. \quad (5.9)$$

Pour vérifier que  $I = \int_{-\infty}^{+\infty} f(x)dx = 1$ , nous remarquons que

$$I^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x)f(y)dxdy = \frac{1}{2\pi} \int_0^{2\pi} d\theta [e^{-\rho^2/2}]_0^\infty = \frac{1}{2\pi} \int_0^{2\pi} d\theta = 1,$$

en passant en coordonnées polaires dans l'intégrale double. On calcule alors que

$$\mathbb{E}(X) = 0 \quad \text{et} \quad \text{Var}(X) = 1,$$

justifiant les qualificatifs "centrée" et "réduite".

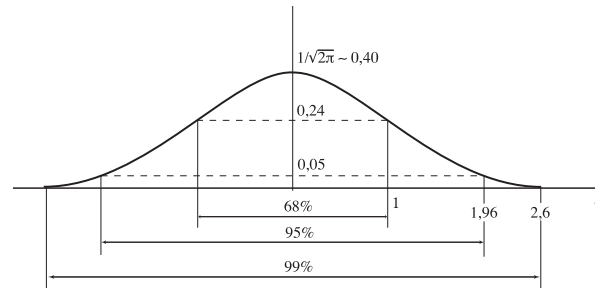


FIGURE 5.4 – La courbe de Gauss  $f$  et les proportions d'aire sous la courbe

**Définition 5.11** On dit que  $X$  est une variable aléatoire de loi normale  $\mathcal{N}(m, \sigma^2)$ , pour  $m \in \mathbb{R}$  et  $\sigma^2 > 0$ , si cette variable a la loi de densité

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right). \quad (5.10)$$

Nous pouvons vérifier que  $f$  est une densité en nous ramenant par le changement de variable  $x \mapsto \frac{x-m}{\sigma}$  à la fonction (5.9). Le même changement de variable permet de voir que  $m$  et  $\sigma^2$  sont l'espérance et la variance d'une v.a. de loi  $\mathcal{N}(m, \sigma^2)$ .

La distribution normale fut introduite par De Moivre en 1733. Celui-ci l'utilisa pour approximer une v.a. binomiale  $\mathcal{B}(n, p)$  pour  $n$  grand. Ce résultat fut ensuite progressivement généralisé par Laplace et d'autres confrères pour devenir le théorème connu sous le nom de théorème de la limite centrale, qui sera démontré dans un chapitre ultérieur. Ce théorème est l'un des plus importants de la théorie des probabilités et prouve que de très nombreux phénomènes aléatoires suivent approximativement une loi normale. Nous pouvons citer à titre d'exemple la taille d'un individu choisi au hasard, les composantes de la vitesse d'une molécule de gaz ou l'erreur de mesure d'une quantité physique.

**Pour les calculs sur la loi normale.** Il est difficile de faire des calculs avec la loi normale car la densité de la loi normale centrée réduite n'admet pas de primitive explicite. Aussi des tables numériques ont-elles été construites pour permettre aux utilisateurs d'obtenir très rapidement des valeurs numériques.

La table 5.1 donne les valeurs de  $\Phi(x) = \mathbb{P}(X \leq x)$ , lorsque  $X$  suit la loi gaussienne centrée réduite. Bien sûr, les logiciels classiques de statistique permettent d'obtenir les valeurs données par la table. Par exemple, pour une variable  $X$  de loi  $\mathcal{N}(0, 1)$ , et en utilisant la symétrie de la densité et la table numérique ci-dessous, nous avons l'approximation

$$\mathbb{P}(|X| < 2) = 2\mathbb{P}(0 < X < 2) = 2\left(\mathbb{P}(X < 2) - \frac{1}{2}\right) \approx 2(0,9772 - 0,5) = 0,9544.$$

**EXERCICE 5.3** Lors d'un procès en attribution de paternité, un expert témoigne que la durée de grossesse, en jours, est de loi approximativement normale de paramètres  $m = 270$  et  $\sigma^2 = 100$ . L'un des pères possibles est en mesure de prouver son absence du pays pendant une période s'étendant entre le 290-ième et le 240-ième jour précédant l'accouchement. Quelle est la probabilité que la conception de l'enfant ait pu avoir lieu pendant la présence de cet homme au pays ?

*Solution :* Comme  $\frac{X-m}{\sigma} = \frac{X-270}{10}$  suit une loi  $\mathcal{N}(0, 1)$ . D'après la table 5.1.

$$\begin{aligned} \mathbb{P}(X > 290 \text{ ou } X < 240) &= \mathbb{P}\left(\frac{X-270}{10} > 2\right) + \mathbb{P}\left(\frac{X-270}{10} < -3\right) \\ &= 1 - \Phi(2) + 1 - \Phi(3) \approx 0,02411. \end{aligned}$$

**Variable aléatoire de Cauchy.** Une variable aléatoire réelle  $X$  suit une **loi de Cauchy** si elle admet la densité

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R},$$

$x$	0,00	0,010	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
$x$	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
$\Phi(x)$	0,99865	0,99903	0,99931	0,99952	0,99966	0,99977	0,999841	0,999928	0,999968	0,999997

TABLE 5.1 – Valeurs de la fonction de répartition  $\Phi(x) = \int_{-\infty}^x f(y)dy$  de la loi normale  $\mathcal{N}(0,1)$ . Si on cherche la valeur de  $\Phi(x = k \times 0,1 + j \times 0,01)$ , il faut regarder dans le gros tableau supérieur, la valeur située à l'intersection de la ligne  $k + 1$  et de la colonne  $j + 1$ . Le tableau donne ainsi toutes les valeurs de  $\Phi(x)$  pour  $x$  allant de 0 à 2,99 par pas de 0,01. Le petit tableau inférieur donne les valeurs de  $\Phi(x)$  pour  $x$  allant de 3 à 4,5 par pas variable. Enfin, si on souhaite trouver une valeur de  $\Phi(x)$  pour un  $x$  négatif, alors on utilise le fait que  $\Phi(x) = 1 - \Phi(-x)$ .

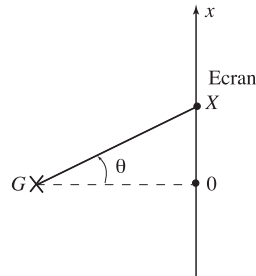


FIGURE 5.5 – Gyrophare

(on vérifie immédiatement que  $\int_{\mathbb{R}} f(x)dx = 1$ ). Il s'agit de l'exemple classique d'une v.a. qui n'a pas d'espérance. En effet, l'intégrale généralisée  $\int_{-\infty}^{+\infty} \frac{|x|}{\pi(1+x^2)} dx$  diverge car  $f(x) \sim_{x \rightarrow \infty} \frac{1}{|x|}$ . Une variable de Cauchy n'est donc pas intégrable.

L'exemple suivant illustre l'apparition naturelle de cette loi.

**EXERCICE 5.4** Un gyrophare  $G$  envoie un flash lumineux dans une direction aléatoire uniforme d'angle  $\theta$ . Chercher la loi de l'abscisse  $X$  du point d'impact du rayon lumineux sur un écran plan infini situé à distance 1 de  $G$ , comme indiqué sur la figure 5.5.

*Solution* : L'angle  $\theta$  est une v.a. de densité  $q(\theta) = \frac{1}{\pi} \mathbf{1}_{]-\frac{\pi}{2}, \frac{\pi}{2}[}(\theta)$ , variable de loi uniforme sur  $]-\frac{\pi}{2}, \frac{\pi}{2}[$ . L'abscisse  $X$  est donnée par  $X = \tan \theta$ , c'est donc une v.a. de fonction de répartition donnée pour  $x \in \mathbb{R}$  par

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\theta \leq \arctan x) = \int_{-\infty}^{\arctan x} q(\theta) d\theta = \frac{1}{\pi} \arctan x + \frac{1}{2}.$$

La fonction  $F$  est de classe  $C^1$  et sa dérivée donne précisément la densité de la loi de Cauchy. Par conséquent, la v.a.  $X$  a la loi de densité  $f$ .

**Remarque 5.12** Dans les exemples que nous venons de voir, il est intéressant de remarquer que les densités sont paramétrées par des nombres réels (un ou deux dans nos exemples), paramètres liés très directement aux valeurs de l'espérance et de la variance de la variable. C'est très important en Statistique. En effet, si nous savons a priori que la loi de  $X$  appartient à une certaine classe de lois (lois exponentielles, lois normales), nous pourrions trouver laquelle en estimant ses paramètres en fonction des observations de  $X$ .

### 5.3 Vecteurs aléatoires à densité

Commençons par l'extension de la loi normale au cadre multidimensionnel. Une étude plus systématique sera faite dans un chapitre ultérieur.

**Exemple 5.13 Vecteur gaussien  $n$ -dimensionnel non-dégénéré.** Soient  $\mathbf{m} \in \mathbb{R}^d$  et  $\mathbf{C}$  une matrice réelle de taille  $d \times d$ , symétrique définie positive (c'est-à-dire  $\mathbf{x} \cdot \mathbf{C}\mathbf{x} > 0$  pour tout  $\mathbf{x} \in \mathbb{R}^d \setminus \{0_{\mathbb{R}^d}\}$ ). Le vecteur aléatoire  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^d$  est gaussien de paramètres  $\mathbf{m}$  et  $\mathbf{C}$  si sa densité s'écrit

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \sqrt{\det(\mathbf{C})}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}) \cdot \mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})}, \quad \mathbf{x} \in \mathbb{R}^d.$$

On a alors  $\mathbb{E}(\mathbf{X}) = \mathbf{m}$  et  $\text{Var}(\mathbf{X}) = \mathbf{C}$ . On note cette loi  $\mathcal{N}(\mathbf{m}, \mathbf{C})$ .

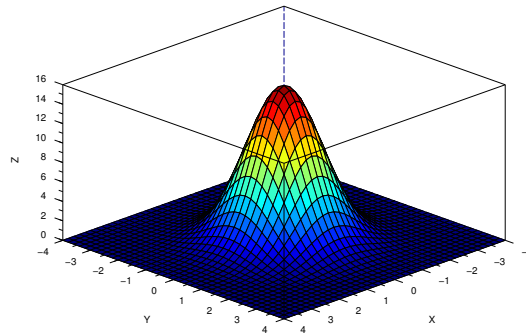


FIGURE 5.6 – Densité gaussienne en dimension  $d = 2$ .

Nous avons vu dans la proposition 5.4 que la loi d'un vecteur aléatoire à densité fait intervenir des intégrales par rapport à la mesure de Lebesgue dans  $\mathbb{R}^d$ , soit des intégrales multiples. Nous commençons donc par rappeler le théorème de Fubini, essentiel dans les calculs d'intégrales multiples. Pour simplifier la présentation, nous considérons le cas  $d = 2$ , l'extension à une dimension quelconque est immédiate.

**Théorème 5.14 (Tonelli, Fubini)** Soit  $f : (\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  mesurable. Si  $f$  est positive, ou  $\int_{\mathbb{R}^2} |f(x, y)| dx dy < \infty$ , i.e.  $f \in L_+^0 \cup L^1(\lambda)$ , alors les fonctions

$$x \mapsto \int_{\mathbb{R}} f(x, y) dy \quad \text{et} \quad y \mapsto \int_{\mathbb{R}} f(x, y) dx$$

sont mesurables (par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ ), et

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} f(x, y) dy \right) dx = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} f(x, y) dx \right) dy.$$

Remarquons l’analogie avec le théorème de Fubini (S9) pour les séries. En fait, la théorie de l’intégration de Lebesgue permet de voir ces résultats comme issus du même théorème général, dont découle aussi le résultat “mixte” suivant.

**Théorème 5.15** *Pour une suite de fonctions mesurables  $(f_k)_k$  de  $\mathbb{R}^d$  dans  $\mathbb{R}$  telle que  $\sum_k \int_{\mathbb{R}^d} |f_k(\mathbf{x})| d\mathbf{x} < +\infty$ ,*

$$\sum_k \int_{\mathbb{R}^d} f_k(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^d} \left( \sum_k f_k(\mathbf{x}) \right) d\mathbf{x}.$$

### 5.3.1 Densités marginales

Nous allons nous limiter dans la suite de ce paragraphe au cas où  $d = 2$ , pour simplifier les notations. Nous considérons un vecteur aléatoire  $\mathbf{Z}$  à valeurs dans  $\mathbb{R}^2$  de loi à densité, et nous notons  $X$  et  $Y$  ses deux composantes :  $\mathbf{Z} = (X, Y)$ . Comme dans le cas discret, la loi des marginales  $X$  et  $Y$  peut s’obtenir à partir de la loi de  $\mathbf{Z}$ . Ce résultat se généralise sans peine à une dimension supérieure.

**Proposition 5.16** *Soit  $\mathbf{Z} = (X, Y)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^2$  et à densité  $f$ . Alors  $X$  et  $Y$  admettent les densités  $f_X$  et  $f_Y$  :*

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy, \quad x \in \mathbb{R} \quad \text{et} \quad f_Y(y) = \int_{\mathbb{R}} f(x, y) dx, \quad y \in \mathbb{R}. \quad (5.11)$$

Les fonctions  $f_X$  et  $f_Y$  s’appellent les **densités marginales** de  $f$ .

**Preuve.** On utilise (5.2) avec  $\phi(u, v) = \mathbf{1}_{u \leq x}$ , pour  $x, u, v \in \mathbb{R}$  :

$$\mathbb{P}(X \leq x) = \mathbb{E}(\phi(X, Y)) = \int_{-\infty}^x du \left( \int_{-\infty}^{+\infty} f(u, v) dv \right) = \int_{-\infty}^x f_X(u) du,$$

où  $f_X$  est définie par (5.11). Ceci montre que  $f_X$  est la densité de  $X$  et on procède de même pour  $Y$ .  $\square$

L’exemple suivant illustre que la réciproque de cette proposition est fautive, en général : le fait que  $X$  et  $Y$  soient à densité ne garantit pas que le vecteur aléatoire  $\mathbf{Z} = (X, Y)$  soit à densité. En particulier, il faut faire attention au fait que dans le cas général, la densité d’un couple de v.a. à densité n’est pas le produit des densités.

**Exemple 5.17** On considère le cas  $X = Y$  et on note  $\Delta = \{(x, x) : x \in \mathbb{R}\}$  la diagonale de  $\mathbb{R}^2$ . Alors  $\mathbb{P}_{\mathbf{Z}}(\Delta) = 1$ , tandis que si  $\mathbf{Z}$  était à densité  $f$ , nous aurions  $\mathbb{P}_{\mathbf{Z}}(\Delta) = \mathbb{E}(\mathbf{1}_{\Delta}(\mathbf{Z})) = \int_{\mathbb{R}^2} \mathbf{1}_{\Delta}(z) f(z) dz = 0$  car  $\Delta$  a un volume nul dans  $\mathbb{R}^2$ .

**Exemple 5.18** On lance une fléchette sur une cible circulaire de rayon unité et on décide de n'observer que les lancés qui atteignent la cible. Le joueur est suffisamment loin de la cible pour que le point d'impact  $M$  de la fléchette soit supposé uniformément distribué sur la cible. Les coordonnées cartésiennes de  $M \in D = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 \leq 1\}$  constituent un couple de v.a. de densité uniforme sur le disque :

$$f_{(X,Y)}(x, y) = \frac{1}{\pi} \mathbf{1}_{\{x^2+y^2 \leq 1\}}, \quad (x, y) \in \mathbb{R}^2.$$

L'abscisse  $X$  est distribuée selon la densité marginale

$$f_X(x) = \int f_{(X,Y)}(x, y) dy = \frac{2}{\pi} (1 - x^2)^{\frac{1}{2}} \mathbf{1}_{[-1,1]}(x), \quad x \in \mathbb{R}.$$

La loi de  $Y$  a la même densité.

### 5.3.2 Densités conditionnelles

Pour tout  $x \in \mathbb{R}$  tel que  $f_X(x) > 0$ , on introduit la fonction

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}, \quad y \in \mathbb{R}. \quad (5.12)$$

Il est clair que  $f_{Y|X=x} \geq 0$  et  $\int_{\mathbb{R}} f_{Y|X=x}(y) dy = 1$ . Il s'agit donc d'une densité.

**Définition 5.19** La fonction  $f_{Y|X=x}$  est appelée densité conditionnelle de  $Y$  sachant  $X = x$ . Elle induit pour toute fonction  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ , mesurable positive ou intégrable (cf. remarque 5.20 (ii) ci-dessous), l'espérance

$$\mathbb{E}(h(X, Y)|X = x) = \int_{\mathbb{R}} h(x, y) f_{Y|X=x}(y) dy \quad \text{pour tout } x \in \mathbb{R} \text{ tel que } f_X(x) > 0.$$

L'espérance conditionnelle de  $h(X, Y)$  sachant  $X$  est la v.a. définie par :

$$\mathbb{E}(h(X, Y)|X) = \psi(X), \quad \text{avec } \psi(x) = \mathbb{E}(h(X, Y)|X = x) \mathbf{1}_{\{f_X(x) > 0\}}.$$

L'interprétation de (5.12) est la suivante : la fonction  $f_{Y|X=x}$  est la densité de la "loi conditionnelle de  $Y$  sachant que  $X = x$ ". Bien sûr, nous avons  $\mathbb{P}(X = x) = 0$  puisque  $X$  admet une densité, donc la phrase ci-dessus n'a pas réellement de sens,



mais elle se justifie heuristiquement ainsi :  $\Delta x$  et  $\Delta y$  étant de “petits” accroissements des variables  $x$  et  $y$ , nous avons comme en (5.3), et lorsque  $f$  est continue :

$$\begin{aligned} f_X(x)\Delta x &\approx \mathbb{P}(x \leq X \leq x + \Delta x), \\ f(x, y)\Delta x\Delta y &\approx \mathbb{P}(x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y). \end{aligned}$$

Par suite

$$\begin{aligned} f_{Y|X=x}(y)\Delta y &\approx \frac{\mathbb{P}(x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y)}{\mathbb{P}(x \leq X \leq x + \Delta x)} \\ &\approx \mathbb{P}(y \leq Y \leq y + \Delta y | x \leq X \leq x + \Delta x). \end{aligned}$$

**Remarque 5.20** (i) La fonction  $\psi(x)$  est défini de manière arbitraire sur  $B = \{u, f_X(u) = 0\}$ , ici nous avons choisi la valeur nulle. Remarquons que  $\mathbb{P}(X \in B) = \int_B f_X(u)du = 0$ . Donc la définition 5.19 définit l’espérance conditionnelle  $\psi(X) = \mathbb{E}(Y | X)$   $\mathbb{P}_X$ -p.s. c’est-à-dire avec probabilité pleine sous la loi de  $X$ .

(ii) Comme  $\mathbb{E}(\mathbb{E}(|Y| | X)) = \int_{\mathbb{R}} (\int_{\mathbb{R}} |y| \frac{f_{X,Y}(x,y)}{f_X(x)} dy) f_X(x) dx = \mathbb{E}(|Y|)$  grâce au théorème 5.14 de Fubini, l’espérance conditionnelle de  $Y$  sachant  $X$  est bien définie dès que  $Y$  est intégrable.

(iii) Les rôles de  $X$  et de  $Y$  peuvent être inversés dans tous les résultats.

En remplaçant les sommes par des intégrales, nous pouvons adapter les preuves de la Section 3.5.3 et obtenir dans le cadre des lois à densité les mêmes propriétés de l’espérance conditionnelle résumées ci-dessous.

**Proposition 5.21** Soient  $X$  et  $Y$  deux variables aléatoires réelles.

(i) Si  $Y$  est intégrable, alors  $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = \int_{\mathbb{R}} \mathbb{E}(Y|X = x) f_X(x) dx$ , et pour toute fonction mesurable  $h$  positive ou intégrable sur  $\mathbb{R}^2$ , on a  $\mathbb{E}(h(X, Y)) = \mathbb{E}(\mathbb{E}(h(X, Y) | X))$ .

(ii) Si  $\mathbf{Y}$  est un vecteur aléatoire intégrable à valeurs dans  $\mathbb{R}^2$ , alors  $\mathbb{E}(\mathbf{a} \cdot \mathbf{Y} | X) = \mathbf{a} \cdot \mathbb{E}(\mathbf{Y} | X)$  pour tout  $\mathbf{a} \in \mathbb{R}^2$  (où  $\mathbb{E}(\mathbf{Y} | X)$  est le vecteur d’espérances conditionnelles).

(iii)  $\mathbb{E}(1|X) = 1$ , et si  $Y \geq 0$ , alors  $\mathbb{E}(Y | X) \geq 0$ .

(iv) Si  $g$  est mesurable et  $g(X)Y$  positive ou intégrable, alors  $\mathbb{E}(Yg(X)|X) = g(X)\mathbb{E}(Y|X)$ .

**Exemple 5.22** Soient  $X$  et  $Y$  de densité jointe  $f_{X,Y}(x, y) = \frac{1}{x} \mathbf{1}_T(x, y)$  où  $T$  est le triangle  $T = \{0 < y < x < 1\}$ . La densité de  $X$  est donnée par  $f_X(x) = \int f_{X,Y}(x, y) dy = \mathbf{1}_{]0,1[}(x)$  et pour  $x \in ]0, 1[$ ,

$$f_{Y|X=x}(y) = \frac{1}{x} \mathbf{1}_{]0,x[}(y).$$

Ainsi,  $X$  est uniformément distribuée sur  $]0, 1[$ , et la loi de  $Y$  sachant  $X = x$  est uniforme sur  $]0, x[$  ( $0 < x < 1$ ). Pour un tel  $x$ , l'espérance conditionnelle  $\mathbb{E}(Y | X = x)$  est donnée par le milieu  $x/2$  de l'intervalle portant cette loi uniforme, et nous obtenons  $\mathbb{E}(Y | X) = X/2$ .

### 5.3.3 Indépendance de variables aléatoires à densité

Dans la suite de ce paragraphe, nous considérons un couple  $(X, Y)$  de variables aléatoires réelles. Le résultat suivant est un analogue de l'équivalence (i)  $\iff$  (ii) dans la proposition 3.36.

**Proposition 5.23** *Supposons que  $X$  et  $Y$  soient à densité  $f_X$  et  $f_Y$ . Alors  $X$  et  $Y$  sont indépendantes si et seulement si le couple  $\mathbf{Z} = (X, Y)$  admet la densité suivante :*

$$f_{\mathbf{Z}}(x, y) = f_X(x) f_Y(y) \quad \text{pour tous } (x, y) \in \mathbb{R}^2. \quad (5.13)$$

**Preuve.** Par définition de l'indépendance, on a

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \mathbb{P}(Y \leq y) = \int_{-\infty}^x f_X(u) du \int_{-\infty}^y f_Y(v) dv,$$

ce qui montre que  $\mathbb{P}_{\mathbf{Z}}$  est à densité  $f_{\mathbf{Z}}$  donnée par (5.13). Réciproquement, le même calcul montre que (5.13) implique que  $\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \mathbb{P}(Y \leq y)$  pour tous  $x, y \in \mathbb{R}$ . En utilisant le même argument que dans la preuve de la proposition 4.42, ceci entraîne que  $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B)$  pour tous boréliens  $A$  et  $B$ , d'où l'indépendance de  $X$  et  $Y$ .  $\square$

## 5.4 Recherche de densité

Dans ce paragraphe, nous abordons le problème important suivant. Soit  $\mathbf{X}$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ , admettant la densité  $f_{\mathbf{X}}$ . Soit  $\mathbf{g}$  une fonction mesurable de  $\mathbb{R}^d$  dans  $\mathbb{R}^{d'}$ , de sorte que  $\mathbf{Y} = \mathbf{g}(\mathbf{X})$  soit aussi un vecteur aléatoire.

1. Est-ce que  $\mathbf{Y}$  admet une densité ?
2. Si oui, comment la calculer ?

Concernant la première question, la réponse est : en général non ! pensez par exemple au cas où  $d = d' = 1$ , la fonction  $g$  est constante,  $g(x) = a$  pour tout  $x \in \mathbb{R}$ , alors la loi de  $Y$  est la masse de Dirac en  $a$ , qui n'a pas de densité. On peut aussi revenir à l'exemple 5.9 qui illustre un cas mixte d'une v.a. à valeurs dans  $\mathbb{R}$  où  $g(x) = \max(x, a)$ ,  $x \in \mathbb{R}$ , place une masse au point  $a$ .

Continuons alors en supposant que la v.a. a une densité. Afin de répondre à la deuxième question, nous allons utiliser la caractérisation de la loi d'une v.a. obtenue dans la proposition 4.37, qui dans le cas présent de recherche de densité se réécrit :

**Proposition 5.24** *Soit  $\mathbf{x}$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ . S'il existe une fonction  $f$  telle que  $\mathbb{E}(h(\mathbf{X})) = \int_{\mathbb{R}^d} h(\mathbf{x})f(\mathbf{x})d\mathbf{x}$ , pour toute fonction continue bornée  $h$ , alors la loi de  $\mathbf{X}$  admet la densité  $f$ .*

On recherche alors une telle fonction de densité en écrivant pour toute fonction  $h$  continue bornée :

$$\mathbb{E}(h(\mathbf{Y})) = \mathbb{E}(h \circ \mathbf{g}(\mathbf{X})) = \int_{\mathbb{R}^d} h \circ \mathbf{g}(\mathbf{x})f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad (5.14)$$

qu'on voudrait mettre sous la forme

$$\int_{\mathbb{R}^d} h(\mathbf{y})f(\mathbf{y}) d\mathbf{y}, \quad (5.15)$$

en faisant le changement de variable  $\mathbf{y} = \mathbf{g}(\mathbf{x})$  dans cette intégrale. Dans le cas unidimensionnel  $d = 1$ , ceci nécessite que  $g$  soit dérivable et bijective "par morceaux", et il faut faire très attention aux domaines où  $g$  est croissante ou décroissante. Plutôt qu'exposer une théorie générale, donnons des exemples.

**Exemple 5.25** *Soit  $Y = aX + b$ , où  $a \neq 0$  et  $b$  sont des constantes. Le changement de variable  $y = ax + b$  dans (5.14) donne  $\mathbb{E}(h(Y)) = \int_{\mathbb{R}} h(ax + b) f_X(x) dx = \int_{\mathbb{R}} h(y) f_X\left(\frac{y-b}{a}\right) \frac{1}{|a|} dy$  (on peut considérer séparément les cas  $a > 0$  et  $a < 0$ ). On peut ainsi conclure  $Y$  est une v.a. de densité*

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \frac{1}{|a|}, \quad y \in \mathbb{R}.$$

- Si  $X$  suit la loi normale  $\mathcal{N}(m, \sigma^2)$ , alors  $\frac{X-m}{\sigma}$  suit la loi  $\mathcal{N}(0, 1)$ .
- Si  $X$  suit la loi normale  $\mathcal{N}(0, 1)$ , alors  $aX + b$  suit la loi  $\mathcal{N}(b, a^2)$ .
- Si  $X$  suit la loi uniforme sur  $[\alpha, \beta]$ , alors  $aX + b$  suit la loi uniforme sur  $[a\alpha + b, a\beta + b]$ .
- Si  $X$  suit la loi  $\Gamma(\alpha, \theta)$ , alors  $aX$  suit la loi  $\Gamma(\alpha, \theta/a)$ .

**Exemple 5.26** *Soit  $Y = X^2$ . La fonction  $x \mapsto x^2$  est décroissante sur  $\mathbb{R}_-$  et croissante sur  $\mathbb{R}_+$ . Le changement de variable  $y = x^2$  donne alors*

$$\begin{aligned} \mathbb{E}(h(Y)) &= \int_{-\infty}^0 h(x^2) f_X(x) dx + \int_0^{+\infty} h(x^2) f_X(x) dx \\ &= \int_0^{+\infty} h(y) f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}} dy + \int_0^{+\infty} h(y) f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} dy, \end{aligned}$$

et nous trouvons donc que  $Y$  est à densité

$$f_Y(y) = (f_X(-\sqrt{y}) + f_X(\sqrt{y})) \frac{1}{2\sqrt{y}} \mathbf{1}_{]0,+\infty[}(y). \quad (5.16)$$

**Exemple 5.27** Soit  $X$  une v.a. de loi  $\mathcal{N}(0,1)$ . La variable aléatoire  $X^2$  suit la loi  $\Gamma(1/2, 1/2)$  (également appelée loi du  $\text{Chi}^2$  à un degré de liberté, et noté  $\mathcal{X}^2(1)$ ), car  $Y = X^2$  est à densité (5.16) :

$$f_Y(y) = \exp\left(-\frac{y}{2}\right) \frac{1}{\sqrt{y}} \mathbf{1}_{]0,+\infty[}(y).$$

Dans le cas des vecteurs aléatoires, l'idée est la même : il convient d'appliquer la formule de changement de variable dans l'intégrale que nous allons maintenant rappeler.

**Théorème 5.28** Soit  $A$  un ouvert de  $\mathbb{R}^d$ , et  $\psi : A \rightarrow B$  une bijection de classe  $C^1$ . Alors,  $\int_B \phi(\mathbf{y}) d\mathbf{y} = \int_A \phi \circ \psi(\mathbf{x}) |\det(\mathbf{J}_\psi(\mathbf{x}))| d\mathbf{x}$ , pour  $\phi$  positive ou intégrable.

Ici, nous rappelons que

- $\mathbf{J}_\psi(\mathbf{x})$  est la matrice Jacobienne de  $\psi$ , c'est la matrice de taille  $d \times d$  de composantes  $J_{\psi_{i,j}}(\mathbf{x}) = \frac{\partial \psi_i}{\partial x_j}(\mathbf{x})$ ,  $1 \leq i, j \leq d$ ,
- $\det(\mathbf{J}_\psi(\mathbf{x}))$  est appelé déterminant Jacobien, et intervient par sa valeur absolue (car, contrairement à la dimension 1, "il n'y a plus d'orientation" dans le calcul de l'intégrale).

Comme première conséquence de la formule de changement de variable dans l'intégrale, on voit qu'il faut se restreindre au cas où  $\mathbf{X}$  et  $\mathbf{Y}$  sont de même dimension. En effet, si  $\mathbf{Y}$  est de dimension plus grande que  $\mathbf{X}$ , alors  $\mathbf{Y} = \mathbf{g}(\mathbf{X})$  n'a pas de densité, et si  $\mathbf{Y}$  est de dimension inférieure à celle de  $\mathbf{X}$ , alors il convient de "compléter"  $\mathbf{Y}$  artificiellement avec autant de composantes que nécessaires afin de se retrouver avec la même nombre de composantes que  $\mathbf{X}$  (voir la remarque 5.32 ci-dessous).

**Proposition 5.29** Soit  $\mathbf{X}$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$  de densité  $f_{\mathbf{X}}$  nulle en dehors d'un ouvert  $A \subset \mathbb{R}^d$ . Soit  $\mathbf{g} : A \rightarrow B$  une bijection telle que  $\mathbf{g}^{-1}$  est de classe  $C^1$ . Alors  $\mathbf{Y}$  est un vecteur aléatoire de densité :

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \mathbf{1}_B(\mathbf{y}) f_{\mathbf{X}} \circ \mathbf{g}^{-1}(\mathbf{y}) |\det \mathbf{J}_{\mathbf{g}^{-1}}(\mathbf{y})| \\ &= \mathbf{1}_B(\mathbf{y}) f_{\mathbf{X}} \circ \mathbf{g}^{-1}(\mathbf{y}) \frac{1}{|\det \mathbf{J}_{\mathbf{g}}(\mathbf{g}^{-1}(\mathbf{y}))|}, \quad \mathbf{y} \in \mathbb{R}^d. \end{aligned}$$

**Preuve.** Sous ces hypothèses, nous pouvons continuer le calcul dans (5.14) en utilisant la formule de changement de variable  $\mathbf{x} = \psi(\mathbf{y}) = \mathbf{g}^{-1}(\mathbf{y})$  du théorème 5.28 :

$$\mathbb{E}(h(\mathbf{Y})) = \int_A h \circ g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_B h(\mathbf{y}) f_{\mathbf{X}} \circ \mathbf{g}^{-1}(\mathbf{y}) d\mathbf{y},$$

et on conclut grâce à la proposition 5.24 et au fait que  $|\det \mathbf{J}_{\mathbf{g}^{-1}}(\mathbf{y})| = \frac{1}{|\det \mathbf{J}_{\mathbf{g}}(\mathbf{g}^{-1}(\mathbf{y}))|}$ .  
□

**Exemple 5.30 (Coordonnées polaires.)** Soit  $\mathbf{X} = (U, V)$  un vecteur aléatoire de  $\mathbb{R}^2$ , et  $\mathbf{Y} = (R, \Theta)$  ses coordonnées polaires. La transformation  $\mathbf{g}$  est un difféomorphisme de  $A = \mathbb{R}^2 \setminus \{0\}$  dans  $B = ]0, \infty[ \times ]0, 2\pi]$ , et son inverse  $\mathbf{g}^{-1}$  s'écrit facilement :  $u = r \cos \theta$ ,  $v = r \sin \theta$ . Alors le déterminant jacobien  $\det \mathbf{J}_{\mathbf{g}^{-1}}(r, \theta) = r$  et on obtient par la proposition 5.29 que :

$$f_{\mathbf{Y}}(r, \theta) = r f_{\mathbf{X}}(r \cos \theta, r \sin \theta) \mathbf{1}_B(r, \theta), \quad (r, \theta) \in \mathbb{R}^2.$$

Par exemple si  $U$  et  $V$  sont indépendantes et de loi  $\mathcal{N}(0, 1)$ , (5.13) entraîne que  $f_{\mathbf{X}}(u, v) = \frac{1}{2\pi} \exp\left(-\frac{u^2+v^2}{2}\right)$ , et

$$f_{\mathbf{Y}}(r, \theta) = \frac{1}{2\pi} r e^{-r^2/2} \mathbf{1}_{]0, \infty[}(r) \mathbf{1}_{]0, 2\pi]}(\theta), \quad (r, \theta) \in \mathbb{R}^2.$$

En particulier les v.a.  $R$  et  $\Theta$  sont indépendantes, la première suit la loi de densité  $r e^{-r^2/2} \mathbf{1}_{]0, \infty[}(r)$ , et la seconde est uniforme sur  $[0, 2\pi]$ .

**Remarque 5.31** Supposons qu'il existe une partition finie  $(A_i)_{1 \leq i \leq d}$  de  $A = \{\mathbf{x} > 0\}$ , telle que  $\mathbf{g} : A_i \rightarrow B_i = \mathbf{g}(A_i)$  est une bijection dont l'inverse est de classe  $C^1$ . Dans ce cas, on découpe l'intégrale selon les  $A_i$ , on applique la proposition 5.29 à chaque morceau, et on obtient en sommant :

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{i=1}^d \mathbf{1}_{B_i}(\mathbf{y}) f_{\mathbf{X}} \circ \mathbf{g}^{-1}(\mathbf{y}) \frac{1}{|\det \mathbf{J}_{\mathbf{g}}(\mathbf{g}^{-1}(\mathbf{y}))|}, \quad \mathbf{y} \in \mathbb{R}^d,$$

où  $\mathbf{g}^{-1}$  est définie sur chaque  $B_i$  comme image réciproque de la restriction de  $\mathbf{g}$  à  $A_i$ .

**Remarque 5.32** Supposons que  $\mathbf{Y}$  a moins de composantes que  $\mathbf{X}$ , i.e.  $d' < d$ . Dans ce cas, on considère un vecteur  $\bar{\mathbf{Y}} = (\mathbf{Y}, \mathbf{Y}')$  complétant  $\mathbf{Y}$  afin de construire une application  $\bar{\mathbf{g}} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  dont les  $d'$  premières composantes coïncident avec les composantes de  $\mathbf{g}$ , et pour laquelle la proposition 5.29 s'applique. Nous obtenons ainsi la densité  $f_{\bar{\mathbf{Y}}}$  de  $\bar{\mathbf{Y}} = \bar{\mathbf{g}}(\mathbf{X})$  et nous appliquons l'extension évidente de (5.11) :

$$f_{\mathbf{Y}}(\mathbf{y}) = \int_{\mathbb{R}^{d-d'}} f_{\bar{\mathbf{Y}}}(\mathbf{y}, \mathbf{y}') d\mathbf{y}'.$$

**Exemple 5.33 (Loi bêta)** Soit  $\mathbf{X} = (U, V)$  un vecteur aléatoire de  $\mathbb{R}^2$ , avec  $U$  et  $V$  indépendantes de lois  $\Gamma(\alpha, \theta)$  et  $\Gamma(\beta, \theta)$ . Alors, en notant  $A = ]0, \infty[^2$ , la densité du vecteur  $\mathbf{X}$  est

$$f_{\mathbf{X}}(u, v) = \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1} v^{\beta-1} e^{-\theta(u+v)} \mathbf{1}_A(u, v), \quad (u, v) \in \mathbb{R}^2.$$

Pour déterminer la densité de  $Y = \frac{U}{U+V}$ , nous introduisons  $\overline{\mathbf{Y}} = (Y, Y')$ , avec  $Y' = U + V$ , ce qui correspond à une transformation  $\mathbf{g}(u, v) = \left(\frac{u}{u+v}, u+v\right)$ . Cette application est bijective de  $A = ]0, \infty[^2$  dans  $B = ]0, 1[ \times ]0, \infty[$ , et nous avons  $\mathbf{g}^{-1}(y, y') = (yy', y'(1-y))$ , qui a pour déterminant jacobien  $y'$ . La proposition 5.29 donne :

$$f_{\overline{\mathbf{Y}}}(y, y') = \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} y'^{\alpha+\beta-1} y^{\alpha-1} (1-y)^{\beta-1} e^{-\theta y'} \mathbf{1}_B(y, y'), \quad (y, y') \in \mathbb{R}^2,$$

et on déduit la densité de la première marginale par (5.11) :

$$\begin{aligned} f_Y(y) &= \int_0^\infty f_{\overline{\mathbf{Y}}}(y, y') dy' \\ &= \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1} \mathbf{1}_{]0,1[}(y) \int_0^\infty y'^{\alpha+\beta-1} e^{-\theta y'} dy' \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1} \mathbf{1}_{]0,1[}(y), \end{aligned} \quad (5.17)$$

en utilisant (5.8). On appelle **loi bêta** de paramètres  $\alpha$  et  $\beta$  la loi admettant cette densité. Nous obtenons aussi facilement la densité de  $Y'$ . En effet, on voit que  $f_{\overline{\mathbf{Y}}}(y, y')$  est le produit de  $f_Y(y)$  par la fonction

$$f_{Y'}(y') = \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha+\beta)} y'^{\alpha+\beta-1} e^{-\theta y'} \mathbf{1}_{]0, \infty[}(y'),$$

qui d'après (5.11) est la densité de la v.a.  $Y'$ . Ceci montre que  $Y'$  suit la loi  $\Gamma(\alpha+\beta, \theta)$ . On retrouvera ce résultat de manière plus simple (grâce à la fonction caractéristique) dans l'Exemple 7.9.

Nous avons en fait démontré le résultat suivant.

**Proposition 5.34** Si  $U$  et  $V$  sont indépendantes et de lois respectives  $\Gamma(\alpha, \theta)$  et  $\Gamma(\beta, \theta)$ , alors  $U + V$  suit la loi  $\Gamma(\alpha + \beta, \theta)$  et est indépendante de  $\frac{U}{U+V}$  qui suit une loi bêta de paramètres  $\alpha$  et  $\beta$ .

En utilisant la même méthode on aussi le résultat suivant.

**Proposition 5.35 (Produit de convolution)** *Si  $U$  et  $V$  sont deux variables aléatoires réelles indépendantes de densités respectives  $f_U$  et  $f_V$ , alors  $Z = U + V$  admet la densité*

$$f_Z(z) = \int_{-\infty}^{+\infty} f_U(u)f_V(z-u)du = \int_{-\infty}^{+\infty} f_U(z-v)f_V(v)dv.$$

La fonction  $f_Z$  est appelée le **produit de convolution** des deux fonctions  $f_U$  et  $f_V$ .

**Preuve.** Si  $U$  et  $V$  sont indépendantes de densités respectives  $f_U$  et  $f_V$ , nous pouvons de la même manière trouver la densité de la somme  $Z = U + V$ . Là encore, nous “complétons”  $Z$  en le couple  $\mathbf{T} = (U, Z)$  (par exemple), correspondant à la bijection  $\mathbf{g}(u, v) = (u, u+v)$  sur  $\mathbb{R}^2$ , dont le jacobien est 1. Par suite la proposition 5.29 permet de conclure.  $\square$

**Exemple 5.36** *La somme de deux variables aléatoires indépendantes, de lois normales  $\mathcal{N}(m, \sigma^2)$  et  $\mathcal{N}(\mu, \tau^2)$  est une variable aléatoire de loi normale  $\mathcal{N}(m + \mu, \sigma^2 + \tau^2)$ . Nous verrons au chapitre 6 une autre preuve de ce résultat grâce à la fonction caractéristique.*

## 5.5 Simulation de suites de variables aléatoires indépendantes

Un générateur aléatoire nous permet d’obtenir une suite  $(X_n)_n$  potentiellement infinie, de v.a. indépendantes et de loi uniforme sur  $[0, 1]$ . Nous voulons construire une suite  $(Y_n)_n$  de vecteur aléatoires indépendants, à valeurs dans  $\mathbb{R}^d$ , de loi donnée notée  $\mathbb{P}_{\mathbf{Y}}$ .

### 5.5.1 Inversion de la fonction de répartition

Considérons le cas  $d = 1$  et généralisons la méthode introduite à la section refsect :simull.

Supposons que l’on connaisse la fonction de répartition  $F$  des variables  $Y_i$ . A partir d’une suite de v.a. uniformes sur  $[0, 1]$  et indépendantes, nous pouvons simuler une suite de v.a. réelles indépendantes de fonction de répartition  $F$ . Comme au paragraphe 4.7.2, nous définissons la fonction “inverse” continue à gauche de  $F$  par :

$$G(x) = \inf\{y \in \mathbb{R} : F(y) \geq x\}, \quad \forall x \in ]0, 1[. \quad (5.18)$$

**Proposition 5.37** *La suite  $(Y_n)_n$  définie par  $Y_n = G(X_n)$  est une suite de variables aléatoires indépendantes et identiquement distribuées de loi  $\mathbb{P}_Y$ .*

**Preuve.** L'indépendance et le fait que les  $Y_n$  suivent la même loi sont évidents. Le calcul de la loi a été fait à la section 4.7.  $\square$

## 5.5.2 Méthode du rejet

Cette méthode s'applique lorsque la probabilité  $\mathbb{P}_Y$  admet une densité  $f$ , et lorsqu'on connaît une autre probabilité  $\nu$ , telle que

- il est possible de simuler des v.a. de loi  $\nu$  (par exemple par la méthode précédente);
- $\nu$  admet une densité  $g$  telle que

$$f(\mathbf{x}) \leq ag(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^d, \quad (5.19)$$

pour une constante  $a$  connue (nécessairement  $a \geq 1$ , et même  $a > 1$  si  $\mathbb{P}_Y \neq \nu$ , puisque  $f$  et  $g$  sont deux fonctions positives ayant la même intégrale 1).

Un cas particulier est le cas où  $f$  est une densité continue à support compact (donc bornée).

Nous supposons aussi que nous disposons, d'une part de la suite  $(X_n)_n$  ci-dessus (constituée de v.a. indépendantes de loi uniforme sur  $[0, 1]$ ), et d'autre part d'une suite  $(Z_n)_n$  potentiellement infinie de v.a. indépendantes de loi  $\nu$  et indépendantes des  $(X_n)_n$ . Nous posons alors

$$N = \inf\{n \in \mathbb{N}^*; f(Z_n) > aX_n g(Z_n)\}, \quad (5.20)$$

avec la convention  $\inf \emptyset = +\infty$ , et

$$Y = \begin{cases} Z_n & \text{si } N = n, \\ 0_{\mathbb{R}^d} & \text{si } N = +\infty. \end{cases} \quad (5.21)$$

**Proposition 5.38** *La v.a.  $N$  suit une loi géométrique de paramètre  $\frac{1}{a}$  (et donc d'espérance  $a$ ). Le vecteur aléatoire  $Y$  suit la loi  $\mathbb{P}_Y$ . Les variables aléatoires  $N$  et  $Y$  sont indépendantes.*

**Preuve.** Notons  $A_n = \{f(Z_n) > aX_n g(Z_n)\}$ . Les événements  $A_n$  sont indépendants et ont même probabilité. Posons  $\mathbb{P}(A_n) = \alpha$  (nous calculerons  $\alpha$  plus tard), de sorte



que  $\mathbb{P}(N > n) = (1 - \alpha)^n$ . Pour toute fonction  $h$  continue bornée, nous avons

$$\begin{aligned}\mathbb{E}(h(\mathbf{Y})) &= \sum_{n=1}^{\infty} \mathbb{E}(h(\mathbf{Z}_n)\mathbf{1}_{N=n}) + \mathbb{E}(h(0)\mathbf{1}_{N=+\infty}) \\ &= \sum_{n=1}^{\infty} \mathbb{E}(h(\mathbf{Z}_n)\mathbf{1}_{A_n}\mathbf{1}_{\{N>n-1\}}) + h(0)\mathbb{P}(N = +\infty).\end{aligned}$$

Les v.a.  $h(\mathbf{Z}_n)\mathbf{1}_{A_n}$  d'une part et  $\mathbf{1}_{\{N>n-1\}}$  d'autre part sont indépendantes, car  $\{N > n - 1\} = \cap_{k=1}^{n-1} A_k^c$ , donc

$$\mathbb{E}(h(\mathbf{Z}_n)\mathbf{1}_{A_n}\mathbf{1}_{\{N>n-1\}}) = \mathbb{E}(h(\mathbf{Z}_n)\mathbf{1}_{A_n})(1 - \alpha)^{n-1}. \quad (5.22)$$

L'espérance  $\mathbb{E}(h(\mathbf{Z}_n)\mathbf{1}_{A_n})$  vaut

$$\int_{\mathbb{R}^d} h(\mathbf{z}) \left( \int_0^1 \mathbf{1}_{\{f(\mathbf{z}) > axg(\mathbf{z})\}} dx \right) g(\mathbf{z}) d\mathbf{z} = \int_{\mathbb{R}^d} h(\mathbf{z}) \frac{f(\mathbf{z})}{ag(\mathbf{z})} g(\mathbf{z}) d\mathbf{z} = \frac{1}{a} \int_{\mathbb{R}^d} h(\mathbf{z}) f(\mathbf{z}) d\mathbf{z}.$$

En particulier  $\alpha$  est égal à cette expression lorsque  $h = 1$ , donc  $\alpha = 1/a$ , puisque  $f$  est une densité. Ainsi,  $0 < \alpha < 1$  et comme  $\mathbb{P}(N > n) = (1 - \alpha)^n$ , ceci montre que la v.a.  $N$  est géométrique de paramètre  $\alpha$ , donc en particulier  $\mathbb{P}(N = +\infty) = 0$ .

En remplaçant  $\mathbb{E}(h(\mathbf{Z}_n)\mathbf{1}_{A_n})$  par sa valeur dans (5.22), nous obtenons que  $\mathbb{E}(h(\mathbf{Y})) = \int h(\mathbf{z}) f(\mathbf{z}) d\mathbf{z}$ , ce qui montre que  $\mathbf{Y}$  est de loi  $\mathbb{P}_{\mathbf{Y}}$ .

Enfin, l'indépendance s'obtient en vérifiant que

$$\mathbb{E}(h(\mathbf{Y})\mathbf{1}_{N=n}) = \alpha(1 - \alpha)^{n-1} \int_{\mathbb{R}^d} h(\mathbf{z}) f(\mathbf{z}) d\mathbf{z} = \mathbb{E}(h(\mathbf{Y}))\mathbb{P}(N = n),$$

donc  $\mathbb{E}(h(\mathbf{Y})g(N)) = \mathbb{E}(h(\mathbf{Y}))\mathbb{E}(g(N))$  pour toute fonction  $g$  bornée.  $\square$

Nous avons ainsi obtenu une v.a. de loi  $\mathbb{P}_{\mathbf{Y}}$ . Pour obtenir une suite de telles variables aléatoires indépendantes, il faut répéter la même procédure.

Nous pouvons comparer les deux méthodes :

- La première est très simple à mettre en œuvre, si l'on connaît explicitement la fonction  $G = F^{-1}$ , ce qui est assez rare dans la pratique.
- La seconde nécessite la connaissance de  $f$ ,  $g$  et  $a$ , et aussi le fait que l'on sache préalablement simuler selon la loi  $\nu$ . L'idée est par exemple d'utiliser la première méthode pour cette loi. Les conditions sont assez souvent remplies, mais cette deuxième méthode est malheureusement parfois longue à mettre en œuvre (sa "longueur" est proportionnelle à  $N$ ).

**Un cas particulier :** Supposons que la densité  $f$  est à support dans le compact  $[b, c]$  et est bornée par une constante  $C$ . Alors la constante  $a$  de l'énoncé général peut-être remplacée par  $a = C(c - b)$ . Nous pouvons adapter la proposition 5.38 et montrer que si  $(U_k)_k$  et  $(V_k)_k$  sont des suites de variables aléatoires indépendantes et de loi respectivement uniforme sur le rectangle  $[b, c]$  et sur le rectangle  $[0, C]$ , alors la variable aléatoire

$$Y = U_N, \quad \text{avec } N = \inf\{k \geq 1 : V_k \leq f(U_k)\},$$

définit une variable aléatoire de densité  $f$ .

Nous allons en déduire l'algorithme suivant :

Tirer  $(U, V)$  de lois uniformes sur  $[b, c]$  et sur  $[0, C]$ , jusqu'à ce que  $V \leq f(U)$ . Poser alors  $X = U$ .

D'après la proposition 5.38, la v.a.  $X$  ainsi obtenue a pour densité  $f$ . Nous rejetons les couples  $(U, V)$  tels que  $V > f(U)$ . Cet algorithme s'appelle l'**algorithme du rejet**.

Remarquons que la loi de  $N$  est une loi géométrique de paramètre  $\frac{1}{C(c-b)}$ , d'espérance  $\mathbb{E}(N) = C(c - b)$ . Comme le nombre d'appels au générateur pseudo-aléatoire est  $2N$ , l'algorithme sera efficace si la majoration de la densité  $f$  par la constante  $C$  sur le support  $[b, c]$  est bien ajustée.

## 5.6 Exercices sur le chapitre 5

**EXERCICE 5.5** Soit  $X \geq 0$  une variable aléatoire à densité, et  $g$  une fonction positive croissante de classe  $C^1$ , telle que  $g(0) = 0$ . Montrer que

$$\mathbb{E}(g(X)) = \int_0^{+\infty} g'(t) \mathbb{P}(X > t) dt.$$

On pourra en particulier voir ce que devient cette formule dans le cas où  $g(x) = x$ .

**EXERCICE 5.6** 1) Un poste à incendie doit être installé le long d'une route forestière de longueur  $A$ ,  $A < \infty$ .

Si les incendies se déclarent en des points uniformément répartis sur  $(0, A)$ , ou doit-on placer ce poste de façon à minimiser l'espérance de la distance entre le poste et l'incendie ?

2) Supposons que la route soit infiniment longue, s'étendant de 0 à l'infini. Si la distance entre un incendie et l'origine est exponentiellement distribuée avec un paramètre  $\lambda$ , où doit-on alors installer le poste à incendie ?

**EXERCICE 5.7** La durée de fonctionnement d'une lampe suit une loi de densité

$$f(t) = \frac{1}{16} t e^{-\frac{t}{4}} \mathbf{1}_{t \geq 0}$$

(l'unité de temps est l'année).

- 1) Vérifier que  $f$  est une densité de probabilité.
- 2) Calculer l'espérance et la variance de cette durée de fonctionnement.
- 3) Quelle est la probabilité que la lampe fonctionne pendant 6 ans ?

**EXERCICE 5.8** Soit  $X$  une variable aléatoire de densité  $f_X$ . Considérons la variable aléatoire

$$Y = c e^{-\alpha X} \mathbf{1}_{\{X > 0\}}, \quad (\alpha > 0, c > 0).$$

Etudier l'existence d'une densité pour  $Y$  et donner sa valeur en fonction de  $f_X$ .

**EXERCICE 5.9** Soit  $X$  une variable aléatoire de loi  $\mathcal{N}(0, 1)$ .

- 1) Calculer  $\mathbb{E}(e^{\lambda X})$  pour  $\lambda \in \mathbb{R}$ .
- 2) En déduire que pour  $a \geq 0$ , on a  $\mathbb{P}(X \geq a) \leq e^{-\frac{a^2}{2}}$ .
- 3) Utiliser la forme intégrale de  $\mathbb{P}(X \geq a)$  pour obtenir l'encadrement

$$\left( \frac{1}{a\sqrt{2\pi}} - \frac{1}{a^3\sqrt{2\pi}} \right) e^{-\frac{a^2}{2}} \leq \mathbb{P}(X \geq a) \leq \frac{1}{a\sqrt{2\pi}} e^{-\frac{a^2}{2}}.$$

**EXERCICE 5.10** Considérons deux paramètres  $a > 0$  et  $\alpha > 0$ . On dira que  $X$  suit une loi de Pareto de paramètres  $(a, \alpha)$  si  $X$  est une variable aléatoire de densité  $f$  définie par

$$f(x) = 0 \text{ si } x < a$$

$$f(x) = \frac{\alpha}{a} \left( \frac{a}{x} \right)^{\alpha+1} \text{ si } x \geq a.$$

Cette variable aléatoire décrit par exemple la répartition des richesses.

- 1) Montrer que  $f$  est bien une densité de probabilité. Allure du graphe de  $f$  : on pourra prendre  $a = 1$  et  $\alpha = 3$ .
- 2) Calculer  $\mathbb{P}(X > x)$ . Allure de la fonction de répartition pour les paramètres ci-dessus.
- 3) Soit  $y > 0$  fixé. Calculer  $\lim_{x \rightarrow \infty} \mathbb{P}(X > x + y | X > x)$ . Qu'en conclure ? Cette propriété est-elle vraie pour une variable exponentielle ?
- 4) Montrer que  $X$  admet une espérance si et seulement si  $\alpha > 1$ . Calculer  $\mathbb{E}(X)$  dans ce cas.
- 5) Montrer que  $X$  admet une variance si et seulement si  $\alpha > 2$ . Calculer  $\text{Var}(X)$  dans ce cas.

**EXERCICE 5.11** Soit  $(X, Y)$  un couple aléatoire de densité  $f(x, y) = \frac{1}{2\pi}|x|e^{-\frac{x^2(1+y^2)}{2}}$ .

- 1) Calculer la loi de  $X$  et la loi de  $Y$ . Est-ce que  $X$  et  $Y$  sont indépendantes ?
- 2) Est-ce que  $X$  et  $XY$  sont indépendantes ? Calculer la loi de  $XY$ .
- 3) Quelle est la loi de  $X(1 + Y)$  ?

**EXERCICE 5.12** Soient  $X$  et  $Y$  deux v.a. réelles. On suppose que la densité conditionnelle de  $X$  sachant  $Y = y$  est la densité  $\mathbf{1}_{\mathbb{R}_+}(x)y^2xe^{-xy}$  et que la loi de  $Y$  est de densité  $\frac{1}{y^2}\mathbf{1}_{[1,+\infty[}(y)$ .  
On pose  $T = XY$ .

- 1) Trouver la loi du couple  $(T, Y)$ . Qu'en déduit-on ?
- 2) Trouver la loi conditionnelle de  $Y$  sachant  $X = x$ .
- 3) Calculer  $\mathbb{E}(Y|X)$ .

**EXERCICE 5.13** Soit  $Y$  une v.a. de loi uniforme sur  $[0, 1]$  et  $X$  à valeurs dans  $\mathbb{N}$ , avec  $X$  et  $Y$  indépendantes. On considère  $Z = XY$ .  
Trouver la loi de  $Z$  et donner une condition pour que la loi de  $Z$  admette une densité par rapport à la mesure de Lebesgue.

**EXERCICE 5.14** Soit  $\alpha, \beta > 0$ . On considère deux v.a. :  $X$  à valeurs dans  $\mathbb{N}$  et  $Y$  à valeurs dans  $\mathbb{R}_+$ . La loi du couple est donnée, pour  $n \in \mathbb{N}$  et  $y > 0$ , par

$$\mathbb{P}(X = n, Y \leq y) = \beta \int_0^y e^{-(\alpha+\beta)z} \frac{(\alpha z)^n}{n!} dz.$$

Donner les lois respectives de  $X$  et de  $Y$ .

**EXERCICE 5.15** Soient  $X_1, \dots, X_n$  des v.a. indépendantes de loi uniforme sur  $[0, 1]$ . On pose :

$$U_1 = X_1, \quad U_2 = X_1 X_2, \quad \dots, \quad U_n = X_1 \cdots X_n.$$

- 1) Chercher la loi du  $n$ -uplet  $(U_1, \dots, U_n)$ .
- 2) Chercher la loi conditionnelle de  $U_n$  sachant  $U_{n-1} = u$ .

**EXERCICE 5.16** *Cet exercice donne une méthode (usuelle) pour simuler deux variables aléatoires indépendantes et de loi normale.*

Soient  $X$  et  $Y$  deux variables aléatoires indépendantes de loi  $\mathcal{N}(0, 1)$ . On considère le couple  $(R, \Theta)$  de variables aléatoires à valeurs  $\mathbb{R}_+ \times [0, 2\pi[$ , obtenu en exprimant  $(X, Y)$  en coordonnées polaires.

- 1) Calculer la loi de  $(R, \Theta)$  et celle de  $(R^2, \Theta)$ .
- 2) En déduire le procédé pratique suivant de simulation de variables aléatoires

indépendantes et de loi normale : si  $U$  et  $V$  sont deux variables aléatoires indépendantes uniformes sur  $[0, 1]$ , alors  $X = \sqrt{-2 \ln U} \cos(2\pi V)$  et  $Y = \sqrt{-2 \ln U} \sin(2\pi V)$  sont deux variables aléatoires indépendantes de loi  $\mathcal{N}(0, 1)$ .

3) Quelle est la loi de  $\frac{Y}{X}$  ?

**EXERCICE 5.17** Soient  $Z_1, \dots, Z_n$  des variables aléatoires réelles, indépendantes et de même loi, de fonction de répartition  $F$ . On pose

$$X = \min_{1 \leq k \leq n} Z_k, \quad Y = \max_{1 \leq k \leq n} Z_k.$$

1) Calculer la fonction de répartition jointe  $F_{X,Y}$  en fonction de  $F$ . Calculer les fonctions de répartition  $F_X$  et  $F_Y$ .

2) Dans le cas où la loi des  $Z_k$  est de densité  $f$ , calculer les lois de  $X$ , de  $Y$  et de  $(X, Y)$ .

3) Nous supposons désormais que les  $Z_k$  ont une loi uniforme sur  $[a, b]$  pour  $a < b$  dans  $\mathbb{R}$ .

Expliciter les lois de  $X$ , de  $Y$  et de  $(X, Y)$ . Calculer  $\mathbb{E}(X)$  et  $\mathbb{E}(Y)$ ,  $\text{Var}(X)$  et  $\text{Var}(Y)$ ,  $\text{Cov}(X, Y)$  et  $\rho(X, Y)$ .

**EXERCICE 5.18** Soient  $X$  et  $Y$  deux variables aléatoires exponentielles indépendantes de paramètres  $\lambda$  et  $\mu$ . On pose  $M = \min(X, Y)$  et  $D = |X - Y| = \max(X, Y) - \min(X, Y)$ .

1) Calculer  $\mathbb{P}(M > a, D > b, X > Y)$  pour  $a, b \geq 0$ .

2) En déduire  $\mathbb{P}(X > Y)$ ,  $\mathbb{P}(X < Y)$ , la loi de  $M$ , la loi de  $D$  conditionnellement à  $X < Y$ , la loi de  $D$  conditionnellement à  $X > Y$ .

3) Montrer que  $M$  et  $\{X < Y\}$  sont indépendants.

4) Soient, pour  $1 \leq i \leq n$ , des variables aléatoires  $X_i$  indépendantes de lois exponentielles de paramètre  $\lambda_i$ . Que dire de la loi de  $\min_{1 \leq i \leq n} X_i$  et de l'événement  $\{X_k = \min_{1 \leq i \leq n} X_i\}$  ?



## Chapitre 6

# Convergences et loi des grands nombres

*Un coup de dés jamais n'abolira le hasard*

*Stéphane Mallarmé.*

Nous allons présenter dans ce chapitre l'un des résultats essentiels de la théorie des probabilités, qui va justifier toute la théorie que nous avons construite à partir de l'approche heuristique du Chapitre 2. Ce résultat montre rigoureusement que, quand le nombre de répétitions de l'expérience tend vers l'infini, la fréquence de réalisation d'un événement converge vers la probabilité de réalisation de cet événement. Ainsi, notre modèle est bien cohérent avec l'intuition. Ce résultat, appelé **Loi des grands nombres**, a d'autres portées fondamentales. Philosophique tout d'abord, puisqu'il permet de voir le monde déterministe comme la limite macroscopique d'une accumulation de phénomènes élémentaires microscopiques aléatoires. Portée numérique aussi, car nous verrons que ce théorème est à l'origine de méthodes de calcul numérique appelées **Méthodes de Monte-Carlo**, qui sont extrêmement puissantes et robustes. Elles sont par exemple très utilisées en Physique ou en Mathématiques Financières.

Considérons un espace fondamental  $\Omega$  (muni de la tribu  $\mathcal{A}$  et de la probabilité  $\mathbb{P}$ ). Nous voulons étudier la répartition des valeurs d'une v.a.  $X$  de loi  $\mathbb{P}_X$ , réalisées au cours d'une succession de  $n$  expériences aléatoires indépendantes. Par exemple, nous interviewons  $n$  personnes choisies au hasard et nous leur demandons si elles aiment les brocolis. Ici, la réponse sera 0 ou 1 et la v.a. associée  $X$  sera une variable de loi de Bernoulli  $\mathbb{P}_X$ .

Nous allons modéliser les résultats possibles de  $X$  au cours des  $n$  expériences par une suite  $X_1, \dots, X_n$  de v.a. indépendantes et de même loi  $\mathbb{P}_X$ . Nous nous intéressons au comportement aléatoire de cette suite de résultats et en particulier à leur moyenne empirique, quand le nombre  $n$  tend vers l'infini ( $n$  est le nombre d'expériences analogues réalisées, par exemple la taille de l'échantillon dans un sondage). La question est donc : comment définir

$$\lim_{n \rightarrow +\infty} \frac{X_1 + \dots + X_n}{n},$$

sachant que chaque  $X_i$  est une fonction de  $\omega$  ?

Pour ce faire nous allons, de manière générale, définir les notions de **convergence de variables aléatoires** et voir que plusieurs définitions différentes sont possibles, non équivalentes, ce qui enrichit mais complique aussi la description des comportements asymptotiques.

## 6.1 Convergences de v.a.

Dans ce paragraphe, nous allons étudier des modes de convergence impliquant la proximité des v.a. elles-mêmes, contrairement au cas de la convergence en loi qui sera étudiée ultérieurement.

Nous considérons une suite  $(\mathbf{X}_n)_{n \geq 1}$  de vecteurs aléatoires, définis sur un même espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$ , et à valeurs dans  $\mathbb{R}^d$ . Nous considérons également sur le même espace un vecteur "limite"  $\mathbf{X}$ . On notera  $|\cdot|$  la valeur absolue dans  $\mathbb{R}$  ou la norme dans  $\mathbb{R}^d$ .

**Définition 6.1** a) La suite  $(\mathbf{X}_n)_n$  converge **presque sûrement** vers  $\mathbf{X}$ , ce qui s'écrit  $\mathbf{X}_n \rightarrow \mathbf{X}$  p.s., s'il existe un ensemble  $N \in \mathcal{A}$  de probabilité nulle tel que

$$\mathbf{X}_n(\omega) \rightarrow \mathbf{X}(\omega) \quad \text{quand } n \rightarrow \infty \quad \text{pour tout } \omega \notin N.$$

b) La suite  $(\mathbf{X}_n)_n$  converge **en probabilité** vers  $\mathbf{X}$ , ce qui s'écrit  $\mathbf{X}_n \xrightarrow{\mathbb{P}} \mathbf{X}$ , si pour tout  $\varepsilon > 0$  on a

$$\mathbb{P}(|\mathbf{X}_n - \mathbf{X}| \geq \varepsilon) \rightarrow 0 \quad \text{quand } n \rightarrow \infty. \quad (6.1)$$

c) La suite  $(\mathbf{X}_n)_n$  converge **en moyenne** vers  $\mathbf{X}$ , ce qui s'écrit  $\mathbf{X}_n \xrightarrow{L^1} \mathbf{X}$ , si  $\mathbf{X}_n$  et  $\mathbf{X}$  sont dans  $L^1$  et si

$$\mathbb{E}(|\mathbf{X}_n - \mathbf{X}|) \rightarrow 0 \quad \text{quand } n \rightarrow \infty. \quad (6.2)$$

**Remarque 6.2** La convergence p.s. est la plus proche de la convergence simple des fonctions. Mais ici, nous permettons à certains  $\omega$  de ne pas vérifier  $\mathbf{X}_n(\omega) \rightarrow \mathbf{X}(\omega)$ , si toutefois la probabilité de réalisation de l'ensemble de ces  $\omega$  est nulle.



Ces convergences ne sont pas équivalentes, comme le montre l'exemple suivant.

**Exemple 6.3** (i) Soit  $(X_n)_n$  une suite de v.a. de Bernoulli  $\mathcal{B}(\frac{1}{n})$ , i.e.  $\mathbb{P}(X_n = 1) = \frac{1}{n} = 1 - \mathbb{P}(X_n = 0)$ . Alors

$$X_n \longrightarrow 0 \quad \text{en probabilité et en moyenne.}$$

En effet  $\mathbb{P}(|X_n| > \varepsilon) = \frac{1}{n} \longrightarrow 0$  pour tout  $\varepsilon \in ]0, 1[$ , prouvant la convergence en probabilité, et  $\mathbb{E}|X_n - 0| = \mathbb{E}(X_n) = \frac{1}{n} \longrightarrow 0$ , prouvant la convergence en moyenne.

(ii) Pour la suite  $Y_n = n^2 X_n$  dont la loi est caractérisée par  $\mathbb{P}(Y_n = n^2) = 1 - \mathbb{P}(Y_n = 0) = \frac{1}{n}$ , pour tout  $n \geq 1$ , nous avons par les mêmes calculs :

$$Y_n \longrightarrow 0 \quad \text{en probabilité, mais } Y_n \not\rightarrow 0 \quad \text{en moyenne,}$$

du fait que  $\mathbb{E}(|Y_n - 0|) = \mathbb{E}(Y_n) = n \longrightarrow \infty$  ! (en fait,  $(Y_n)_n$  ne converge vers aucune autre limite finie).

**Exemple 6.4** Soit  $U$  une v.a. uniforme sur  $[0, 1]$ . Alors, la suite de v.a. de loi  $\mathcal{B}(\frac{1}{n})$  :

$$Z_n = \mathbf{1}_{\{U \leq \frac{1}{n}\}} \longrightarrow 0, \quad p.s.$$

En effet, considérant l'ensemble négligeable  $N = \{U = 0\}$ , on a que pour tout  $\omega \in N^c$ , il existe  $n_0$  tel que  $U(\omega) > \frac{1}{n_0}$ , impliquant que  $Z_n(\omega) = 0$  pour tout  $n \geq n_0$ .

Nous allons maintenant étudier les liens entre ces différentes convergences. Commençons par le résultat le plus simple.

**Théorème 6.5** La convergence en moyenne entraîne la convergence en probabilité.

**Preuve.** Pour des vecteurs aléatoires intégrables  $\mathbf{X}_n$  et  $\mathbf{X}$ , il suffit de remarquer que l'inégalité de Markov (4.13) donne pour tout  $\varepsilon > 0$ ,

$$\mathbb{P}(|\mathbf{X}_n - \mathbf{X}| \geq \varepsilon) \leq \frac{\mathbb{E}(|\mathbf{X}_n - \mathbf{X}|)}{\varepsilon}.$$

□

Nous avons vu dans l'exemple 6.3 que la réciproque n'est pas toujours vraie.

L'un des résultats les plus importants de la théorie de l'intégration relie la convergence en moyenne et la convergence presque-sûre. C'est ce résultat qui, en grande partie, fait la supériorité de l'intégrale de Lebesgue par rapport à celle de Riemann. Commençons

par un autre résultat qui est également très utile et qui est une conséquence simple du théorème 4.12 de convergence monotone. Pour une suite réelle  $(u_n)_n$ , on rappelle la notation

$$\liminf_n u_n = \sup_{n \geq 1} \inf_{k \geq n} u_k \quad \text{et} \quad \limsup_n u_n = \inf_{n \geq 1} \sup_{k \geq n} u_k$$

On rappelle que  $\liminf_n u_n \leq \limsup_n u_n$  avec égalité si et seulement si la limite existe, et dans ce cas  $\lim_n u_n = \liminf_n u_n = \limsup_n u_n$ .

**Lemme 6.6** (Fatou) *Pour une suite de v.a. positifs  $(X_n)_n$ , on a  $\mathbb{E}(\liminf_n X_n) \leq \liminf_n \mathbb{E}(X_n)$ .*

**Preuve.** D'après la monotonie de l'espérance,  $\inf_{k \geq n} \mathbb{E}(X_k) \geq \mathbb{E}(\inf_{k \geq n} X_k)$  pour tout  $n \geq 1$ . Comme la suite  $(\inf_{k \geq n} X_k)_{n \geq 1}$  est croissante  $\mathbb{P}$ -p.s., on obtient le résultat par application du théorème 4.12 de convergence monotone.  $\square$

**Théorème 6.7** (de Lebesgue, ou de convergence dominée) *Si la suite de vecteurs aléatoires  $\mathbf{X}_n$  converge presque-sûrement sur  $\Omega$  vers une limite  $\mathbf{X}$  et si*

$$\forall n, |\mathbf{X}_n| \leq Z \text{ avec } Z \in L^1,$$

alors  $\mathbf{X}_n$  et  $\mathbf{X}$  sont intégrables et on a

$$\mathbb{E}(|\mathbf{X}_n - \mathbf{X}|) \xrightarrow{n \rightarrow \infty} 0.$$

En particulier,  $\mathbb{E}(\mathbf{X}_n) \xrightarrow{n \rightarrow \infty} \mathbb{E}(\mathbf{X})$ .

**Preuve.** On note  $Y_n = |\mathbf{X}_n - \mathbf{X}|$  et  $Y^* = \sup_n Y_n$ . Par le lemme de Fatou,  $\mathbb{E}(|\mathbf{X}|) \leq \liminf_n \mathbb{E}(|\mathbf{X}_n|) \leq \mathbb{E}(Z) < \infty$ . Alors  $\mathbf{X} \in L^1$  et  $|Y^*| \leq |\mathbf{X}| + Z \in L^1$ . Comme la v.a.  $Y^* - Y_n$  est positive et que  $Y_n \rightarrow 0$ ,  $\mathbb{P}$ -p.s., on obtient par le lemme de Fatou que  $\liminf_n \mathbb{E}(Y^* - Y_n) \geq \mathbb{E}(Y^*)$ . Simplifiant par  $\mathbb{E}(Y^*)$ , ceci implique que  $0 \geq \limsup_n \mathbb{E}(Y_n) \geq \liminf_n \mathbb{E}(Y_n) \geq 0$  du fait de la positivité de  $Y_n$ .  $\square$

**Attention :** la réciproque est fautive. Dans l'exemple 6.3, supposons que les variables aléatoires réelles  $(X_n)_n$  soient indépendantes. Puisque  $\mathbb{P}(|X_n| > \varepsilon) = \frac{1}{n}$ , la série de terme général  $\mathbb{P}(|X_n| > \varepsilon) = \frac{1}{n}$  est divergente. Par ailleurs les événements  $A_n = \{X_n > \varepsilon\}$  sont indépendants. Par le théorème 2.35 de Borel-Cantelli, nous en déduisons que pour presque tout  $\omega$ , une infinité de  $X_n(\omega)$  seront supérieurs à  $\varepsilon$ . Ainsi, la suite ne peut pas converger vers 0. Ainsi, la suite  $(X_n)_n$  est bornée par 1, elle tend en moyenne vers 0 mais pas presque-sûrement.

En revanche, considérons maintenant la suite  $(V_n)_n$  avec, pour  $a > 1$

$$\mathbb{P}(V_n = 1) = \frac{1}{n^a} = 1 - \mathbb{P}(V_n = 0).$$

Puisque  $a > 1$ , la série de terme général  $\mathbb{P}(V_n \geq \varepsilon)$  converge, et donc toujours par le théorème de Borel-Cantelli, nous savons que pour presque tout  $\omega$ , un nombre fini au plus de  $V_n(\omega)$  seront supérieurs à  $\varepsilon$ . On a donc  $V_n \rightarrow 0$  p.s.

Nous pouvons donc voir à travers ces exemples que ces notions sont délicates.

**Remarque 6.8** *L'hypothèse de domination est nécessaire. Considérons une suite de variables aléatoires réelles  $(T_n)_n$  telle que*

$$\mathbb{P}(T_n = n^2) = \frac{1}{n\sqrt{n}} = 1 - \mathbb{P}(T_n = 0).$$

*Par un argument similaire à l'argument précédent, nous pouvons montrer que la suite  $(T_n)_n$  converge presque-sûrement vers 0. En revanche,  $\mathbb{E}(T_n) = \sqrt{n}$ , et la suite  $(T_n)_n$  ne peut pas converger en moyenne.*

Nous explorons maintenant d'autres relations entre ces convergences.

**Proposition 6.9** *La convergence presque-sûre entraîne la convergence en probabilité.*

**Preuve.** Soit  $A_{n,\varepsilon} = \{\omega, |\mathbf{X}_n(\omega) - \mathbf{X}(\omega)| \geq \varepsilon\}$ . Supposons que  $\mathbf{X}_n \xrightarrow{n \rightarrow \infty} \mathbf{X}$  presque-sûrement. Soit  $N$  l'ensemble de probabilité nulle en dehors duquel on a  $\mathbf{X}_n(\omega) \rightarrow \mathbf{X}(\omega)$ . Si  $\omega \notin N$ , alors  $\omega \notin A_{n,\varepsilon}$  pour tout  $n \geq n_0$ , où  $n_0$  dépend de  $\omega$  et de  $\varepsilon$ , ce qui implique que les v.a.  $Y_{n,\varepsilon} = \mathbf{1}_{N^c \cap A_{n,\varepsilon}}$  tendent simplement vers 0 quand  $n \rightarrow \infty$ . Comme nous avons aussi  $0 \leq Y_{n,\varepsilon} \leq 1$ , le théorème de convergence dominée (Théorème 6.7) entraîne que  $\mathbb{E}(Y_{n,\varepsilon}) \xrightarrow{n \rightarrow \infty} 0$ . Mais

$$\mathbb{P}(A_{n,\varepsilon}) \leq \mathbb{P}(N^c \cap A_{n,\varepsilon}) + \mathbb{P}(N) = \mathbb{P}(N^c \cap A_{n,\varepsilon}) = \mathbb{E}(Y_{n,\varepsilon}) \xrightarrow{n \rightarrow \infty} 0,$$

et nous en déduisons (6.1). □

La convergence en probabilité n'entraîne pas la convergence en moyenne, comme nous l'avons vu dans l'exemple 6.3. Si les  $X_n$  ne sont "pas trop grands", il y a cependant équivalence entre les deux modes de convergence. En voici un exemple.

**Proposition 6.10** *S'il existe une constante  $a$  telle que  $|\mathbf{X}_n| \leq a$  presque-sûrement, il y a équivalence entre  $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$  et  $\mathbf{X}_n \xrightarrow{L^1} \mathbf{X}$ .*

**Preuve.** Etant donné le théorème 6.5, il suffit de montrer que la convergence en probabilité implique la convergence en moyenne, lorsque  $|\mathbf{X}_n| \leq a$ .

Comme  $|\mathbf{X}_n| \leq a$ , nous avons  $\{|\mathbf{X}| > a + \varepsilon\} \subset A_{n,\varepsilon}$  (avec la même notation que précédemment), et donc  $\mathbb{P}(|\mathbf{X}| > a + \varepsilon) \leq \mathbb{P}(A_{n,\varepsilon})$ . En faisant tendre  $n$  vers  $+\infty$ , nous en déduisons que  $\mathbb{P}(|\mathbf{X}| > a + \varepsilon) = 0$ . Ceci est vrai pour tout  $\varepsilon$ , et donc

$$\mathbb{P}(|\mathbf{X}| > a) = 0. \quad (6.3)$$

Comme  $|\mathbf{X}_n| \leq a$ , nous avons aussi

$$|\mathbf{X}_n - \mathbf{X}| \leq \varepsilon \mathbf{1}_{A_{n,\varepsilon}^c} + (|\mathbf{X}_n| + |\mathbf{X}|) \mathbf{1}_{A_{n,\varepsilon}} \leq \varepsilon + 2a \mathbf{1}_{A_{n,\varepsilon}}$$

sur l'ensemble  $\{|\mathbf{X}| \leq a\}$ , qui est de probabilité 1. Donc il vient

$$\mathbb{E}(|\mathbf{X}_n - \mathbf{X}|) \leq \varepsilon + 2a\mathbb{P}(A_{n,\varepsilon}).$$

Il s'en suit (par (6.1)) que  $\limsup_n \mathbb{E}(|\mathbf{X}_n - \mathbf{X}|) \leq \varepsilon$ , et comme  $\varepsilon$  est arbitrairement proche de 0, nous en déduisons (6.2).  $\square$

Les rapports entre convergence presque-sûre et convergence en probabilité sont plus subtils. La première de ces deux convergences est plus forte que la seconde d'après la proposition 6.9, mais "à peine plus", comme le montre le résultat suivant.

**Proposition 6.11** *Si  $\mathbf{X}_n \rightarrow \mathbf{X}$  en probabilité, alors il existe une sous-suite  $(n_k)$  telle que  $\mathbf{X}_{n_k} \rightarrow \mathbf{X}$  p.s.*

**Preuve.** Comme la suite  $(\mathbf{X}_n)_n$  converge en probabilité vers  $\mathbf{X}$ , nous pouvons définir une sous-suite de la manière suivante. Posons  $n_1 = 1$ , et soit

$$n_j = \inf \left\{ n > n_{j-1}; \mathbb{P} \left( |\mathbf{X}_r - \mathbf{X}_s| > \frac{1}{2^j} \right) < \frac{1}{3^j}, \text{ pour } r, s \geq n \right\}.$$

Il résulte alors de :  $\sum_j \mathbb{P}(|\mathbf{X}_{n_{j+1}} - \mathbf{X}_{n_j}| > \frac{1}{2^j}) < \sum_j \frac{1}{3^j} < \infty$  que la suite  $(\mathbf{X}_{n_j})_{j \geq 1}$  converge presque-sûrement. En effet, c'est une conséquence du lemme de Borel-Cantelli (Théorème 2.35) appliqué aux ensembles  $A_j = \{|\mathbf{X}_{n_{j+1}} - \mathbf{X}_{n_j}| > \frac{1}{2^j}\}$ .  $\square$

**Exemple 6.12** *Soient  $\Omega = \mathbb{R}$  muni de sa tribu borélienne et  $\mathbb{P}$  la probabilité uniforme sur  $[0, 1]$ . Soit  $X_n = \mathbf{1}_{A_n}$ , où  $A_n$  est un intervalle de  $[0, 1]$  de longueur  $1/n$ . Ainsi,  $\mathbb{E}(X_n) = 1/n$ , et la suite  $X_n$  tend vers  $X = 0$  en moyenne, et donc en probabilité. Supposons que les  $A_n$  soient placés bout-à-bout, en recommençant en 0 chaque fois qu'on arrive au point 1. Il est clair que l'on parcourt indéfiniment l'intervalle  $[0, 1]$  (car la série de terme général  $1/n$  diverge). Ainsi la suite numérique  $X_n(\omega)$  ne converge pour aucun  $\omega$ , et on n'a pas  $X_n \rightarrow X$  presque-sûrement; cependant comme la série  $\sum_n 1/n^2$  converge, il s'en suit que  $X_{n^2} \rightarrow X = 0$  presque-sûrement. Nous avons donc la convergence presque-sûre de la sous-suite  $(X_{n^2})_n$ .*

**Proposition 6.13** *Soit  $f$  une fonction continue de  $\mathbb{R}^d$  dans  $\mathbb{R}$ .*

(a) *Si  $\mathbf{X}_n \rightarrow \mathbf{X}$  p.s. alors  $f(\mathbf{X}_n) \rightarrow f(\mathbf{X})$  p.s.*

(b) *Si  $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$ , alors  $f(\mathbf{X}_n) \xrightarrow{P} f(\mathbf{X})$ .*

**Preuve.** (a) est évident. Pour (b) remarquons d’abord que si  $K > 0$  et  $\varepsilon > 0$ ,

$$\{|f(\mathbf{X}_n) - f(\mathbf{X})| \geq \varepsilon\} \subset \{|\mathbf{X}| > K\} \cup \{|\mathbf{X}| \leq K, |f(\mathbf{X}_n) - f(\mathbf{X})| \geq \varepsilon\}. \quad (6.4)$$

La fonction  $f$  est uniformément continue sur  $\{\mathbf{x} : |\mathbf{x}| \leq 2K\}$ , donc il existe  $\eta > 0$  tel que  $|\mathbf{x} - \mathbf{y}| < \eta$  et  $|\mathbf{x}| \leq 2K$ ,  $|\mathbf{y}| \leq 2K$  impliquent que  $|f(\mathbf{x}) - f(\mathbf{y})| < \varepsilon$ . Ainsi, en choisissant  $\eta$  suffisamment petit,  $|\mathbf{x}| \leq K$  entraîne  $|\mathbf{y}| \leq 2K$ , et (6.4) implique :

$$\{|f(\mathbf{X}_n) - f(\mathbf{X})| \geq \varepsilon\} \subset \{|\mathbf{X}| > K\} \cup \{|\mathbf{X}_n - \mathbf{X}| \geq \eta\},$$

$$\mathbb{P}(|f(\mathbf{X}_n) - f(\mathbf{X})| \geq \varepsilon) \leq \mathbb{P}(|\mathbf{X}| > K) + \mathbb{P}(|\mathbf{X}_n - \mathbf{X}| \geq \eta).$$

D’après l’hypothèse il vient

$$\limsup_n \mathbb{P}(|f(\mathbf{X}_n) - f(\mathbf{X})| \geq \varepsilon) \leq \mathbb{P}(|\mathbf{X}| > K). \quad (6.5)$$

Enfin  $\lim_{K \rightarrow \infty} \mathbb{P}(|\mathbf{X}| > K) = 0$  (nous le montrons grâce au théorème de convergence dominée) et donc dans (6.5) la  $\limsup$  est nulle. Nous obtenons ainsi le résultat.  $\square$

## 6.2 La loi des grands nombres

Reprenons la modélisation présentée dans l’introduction de ce chapitre. Nous considérons une suite  $(X_n)_{n \geq 1}$  de variables aléatoires réelles **indépendantes et de même loi**, que l’on note iid (pour “indépendantes et identiquement distribuées”). Notre objectif est de montrer que la moyenne empirique définie comme la moyenne des  $n$  premières variables aléatoires

$$M_n = \frac{1}{n}(X_1 + \dots + X_n),$$

converge vers l’espérance des variables  $X_n$  lorsque cette dernière existe (comme les  $X_n$  ont même loi, cette espérance est la même pour tout  $n$ ). Il s’agit là, répétons-le, d’un des résultats essentiels de toute la théorie des probabilités, connu sous le nom de **loi des grands nombres**.

**Théorème 6.14** *Soit  $(X_n)_n$  une suite de v.a. iid, intégrables. Alors*

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{p.s.} \mathbb{E}[X_1] \text{ p.s. et en moyenne (et donc en probabilité).}$$

**Remarque 6.15** (i) *Le résultat sur la convergence en probabilité est appelé **loi faible des grands nombres**. Si  $X_n$  est de carré intégrable, de moyenne  $m$  et de variance  $\sigma^2 = \text{Var}(X_n)$ , sa preuve est immédiate : il suffit de remarquer que  $\mathbb{E}(M_n) = m$  et*

$$\mathbb{E}(|M_n - m|) \leq \sqrt{\mathbb{E}((M_n - m)^2)} = \sqrt{\text{Var}(M_n)} = \frac{\sigma}{\sqrt{n}} \rightarrow 0, \quad \text{quand } n \rightarrow \infty,$$

d'où  $M_n \rightarrow m$  en moyenne, et par suite en probabilité d'après le théorème 6.5. Nous avons en fait obtenu la convergence en moyenne quadratique  $\mathbb{E}((M_n - m)^2) \xrightarrow{n \rightarrow \infty} 0$ .

(ii) *La convergence presque-sure de la moyenne empirique vers la moyenne est appelé **loi forte des grands nombres**. Il s'agit d'un résultat plus fort donnant une information sur la trajectoire  $n \rightarrow X_n(\omega)$ , pour presque tout  $\omega$ .*

Remarquons aussi que l'indépendance des v.a.  $X_n$  ne peut être relâchée simplement. Prenons par exemple toutes les  $X_n$  égales à une même variable aléatoire  $X$ . Alors  $M_n = X$ , qui ne convergera vers  $m$  que si  $X$  est constante, égale à  $m$ .

**Preuve.** Cette preuve est due à Randal Douc. Remarquons d'abord que l'on peut toujours se ramener au cas  $\mathbb{E}(X_1) = 0$ , quitte à considérer les variables centrées. Notons  $S_n = \sum_{i=1}^n Y_i$ , où  $Y_i = X_i + c$  pour une constante  $c > 0$ . Alors les  $Y_i$  sont iid, intégrables, et  $\mathbb{E}(Y_i) = c > 0$ . Nous allons montrer que

$$L^* = \inf_{n \geq 1} S_n > -\infty, \quad \mathbb{P}\text{-p.s.}, \quad (6.6)$$

ce qui suffira pour obtenir la loi forte des grands nombres. En effet, puisque  $S_k \geq L^*$  pour tout  $k \geq 1$ , on a pour tout  $n \geq 1$  :  $\inf_{k \geq n} \frac{1}{k} S_k \geq \inf_{k \geq n} \frac{1}{k} L^* = 0$   $\mathbb{P}$ -p.s., car  $L^*$  est fini  $\mathbb{P}$ -p.s.. Donc on a :

$$c + \liminf_n M_n = \liminf_n \frac{S_n}{n} = \sup_{n \geq 1} \inf_{k \geq n} \frac{S_k}{k} \geq 0, \quad \mathbb{P}\text{-p.s.}$$

En appliquant le même raisonnement à  $Y_i = -X_i + c$ , on obtient de même  $0 \leq c + \liminf_n (-M_n) = c - \limsup_n M_n$ ,  $\mathbb{P}$ -p.s. D'où

$$-c \leq \liminf_n M_n \leq \limsup_n M_n \leq c, \quad \mathbb{P}\text{-p.s.}$$

et en envoyant  $c$  vers 0, on obtient que  $\liminf_n M_n = \limsup_n M_n = 0$ ,  $\mathbb{P}$ -p.s., ce qui est exactement le résultat voulu.

Montrons à présent (6.6) ou, de manière équivalente,  $\mathbb{P}(A(Y)) = 0$ , où  $Y = (Y_n)_{n \geq 1}$  et  $A(Y) = \{L^* = -\infty\}$ . Notant aussi  $L_n(Y) = \inf_{k \leq n} S_k$  et  $\theta(Y) = (Y_n)_{n \geq 2}$ , on a :

$$L_n(Y) = Y_1 + \inf_{k \leq n} (S_k - Y_1) = Y_1 + \min \{0, L_{n-1}(\theta(Y))\} \geq Y_1 + \min \{0, L_n(\theta(Y))\}.$$

Comme  $A(Y) = A(\theta(Y))$  et  $L^- = -\min(0, L)$ , en multipliant par  $\mathbf{1}_{A(Y)}$  et en prenant l'espérance, on obtient :

$$\begin{aligned} \mathbb{E}(\mathbf{1}_{A(Y)}Y_1) &\leq \mathbb{E}(\mathbf{1}_{A(Y)}L_n(Y)) + \mathbb{E}(\mathbf{1}_{A(Y)}L_n(\theta(Y))^-) \\ &= \mathbb{E}(\mathbf{1}_{A(Y)}L_n(Y)) + \mathbb{E}(\mathbf{1}_{A(\theta(Y))}L_n(\theta(Y))^-) \\ &= \mathbb{E}(\mathbf{1}_{A(Y)}L_n(Y)) + \mathbb{E}(\mathbf{1}_{A(Y)}L_n(Y)^-) = \mathbb{E}(\mathbf{1}_{A(Y)}L_n(Y)^+), \end{aligned}$$

puisque  $Y$  et  $\theta(Y)$  ont la même loi du fait que les  $Y_i$  sont iid. Ceci implique que  $\mathbb{E}(\mathbf{1}_{A(Y)}Y_1) \leq \lim_n \mathbb{E}(\mathbf{1}_{A(Y)}L_n(Y)^+) = 0$ , d'après le théorème de convergence dominée puisque  $0 \leq \mathbf{1}_{A(Y)}L_n(Y)^+ \leq Y_1^+ \in L^1$  pour tout  $n \geq 1$ . Utilisant de nouveau l'égalité  $A(Y) = A(\theta(Y))$ , on obtient alors

$$0 \geq \mathbb{E}(\mathbf{1}_{A(Y)}Y_1) = \mathbb{E}(\mathbf{1}_{A(\theta(Y))}Y_1) = \mathbb{E}(\mathbf{1}_{A(\theta(Y))})\mathbb{E}(Y_1) = \mathbb{P}(A(Y))\mathbb{E}(Y_1),$$

par indépendance des  $Y_i$ . Comme  $\mathbb{E}(Y_1) = c > 0$ , ceci montre que  $\mathbb{P}(A(Y)) = 0$ .  $\square$

Revenons à "l'approche par les fréquences" du Chapitre 2. Soit un événement  $A$ . Nous répétons l'expérience, et nous notons  $X_n$  la v.a. qui vaut 1 si  $A$  est réalisé au cours de la  $n^{\text{ième}}$  expérience et 0 sinon. La fréquence de réalisation de  $A$  au cours des  $n$  premières expériences est alors

$$f_n(A) = \frac{1}{n}(X_1 + \dots + X_n) = M_n.$$

Par ailleurs, les  $X_i$  sont indépendantes et de même loi et  $\mathbb{E}(X_i) = \mathbb{P}(X_i = 1) = \mathbb{P}(A)$ . Donc la loi forte des grands nombres implique que  $f_n(A)$  converge vers  $\mathbb{P}(A)$  presque-sûrement. Nous obtenons ainsi une justification *a posteriori* de l'approche par les fréquences, qui, sans en démontrer de manière rigoureuse la validité (c'est évidemment impossible), montre au moins que cette approche est compatible avec la théorie qui a été fondée dessus.

En outre, la loi des grands nombres nous indique aussi dans quel sens il convient de prendre la convergence dans (2.1), à savoir au sens presque-sûr. Il faut remarquer que dans les théorèmes précédents, et donc aussi dans l'approche par les fréquences, on **ne peut pas** avoir convergence de  $M_n(\omega)$  vers  $m$  pour tout  $\omega$ . Nous allons voir dans l'exemple suivant qu'il existe un ensemble négligeable sur lequel la convergence n'est pas réalisée.

**Exemple 6.16** *Considérons un jeu de Pile ou Face infini, c'est-à-dire une suite  $X_n$  de v.a. ne prenant que les valeurs 0 et 1. Nous définissons cette suite sur l'espace  $\Omega = \{0, 1\}^{\mathbb{N}^*}$ , i.e. un point  $\omega$  est une suite numérique  $x_1, \dots, x_n, \dots$  de 0 et de 1. Chaque suite est en principe possible. Soit alors  $\mathbb{P}_p$  une probabilité sous laquelle les  $X_n$  sont indépendantes et de même loi de Bernoulli de paramètre  $p \in ]0, 1[$ . (Il est possible de construire une telle loi.) La loi des grands nombres nous dit que pour toute suite*

$(x_n)_n$  en dehors d'un ensemble de probabilité nulle, la moyenne  $\frac{1}{n}(x_1 + \dots + x_n)$  tend vers le nombre  $p$  quand  $n$  tend vers l'infini. Il existe évidemment beaucoup de suites ne vérifiant pas cette propriété, par exemple  $x_n = 0$  pour tout  $n$ . Si nous considérons maintenant la probabilité  $\mathbb{P}_q$  définie de la même manière pour  $q \neq p$ , l'ensemble de probabilité 1 pour  $\mathbb{P}_p$  des suites  $(x_n)_n$  qui tendent vers  $p$ , devient sous  $\mathbb{P}_q$  un ensemble de probabilité nulle. Remarquons également que la probabilité de chaque suite particulière est nulle (elle vaut  $\lim_{k \rightarrow +\infty} p^{k_1}(1-p)^{k_2}$ ,  $k_1 + k_2 = k$ ), ce qui veut dire que  $\mathbb{P}$  n'est pas une somme de mesures de Dirac.

Cet exemple montre que lorsque l'on étudie la convergence des variables aléatoires, il est **indispensable** d'introduire la convergence presque-sûre, puisqu'on n'a généralement pas la convergence simple (i.e. pour tout  $\omega$ ).

### A titre de complément...

Voici une autre démonstration de la loi forte des grands nombres sous l'hypothèse (plus forte) que les variables iid  $X_i$  sont de carré intégrable. L'argument suivant est basé sur le lemme de Borel-Cantelli. On note  $\sigma^2 = \text{Var}(X)$  et on se ramène au cas où  $\mathbb{E}(X) = 0$ , quitte à considérer les variables centrées. On procède en deux étapes :

1) Montrons d'abord que  $M_{n^2} \rightarrow 0$ ,  $\mathbb{P}$ -p.s. Notant  $A_{n,q} = \{|M_{n^2}| \geq \frac{1}{q}\}$ ,  $q \in \mathbb{N}^*$ , on voit d'après la remarque 6.15 et l'inégalité de Bienaymé-Chebyshev (4.14) que

$$\mathbb{P}(A_{n,q}) \leq \frac{\sigma^2 q^2}{n^2}, \quad \text{et par suite} \quad \sum_{n \geq 1} \mathbb{P}(A_{n,q}) < \infty.$$

Le lemme de Borel-Cantelli (Théorème 2.35) assure alors que les événements  $C_q = \limsup_n A_{n,q}$  et (par conséquent)  $N = \cup_{q \geq 1} C_q$  sont négligeables. Maintenant,  $\omega \notin N$  si et seulement si  $\omega \notin C_q$  pour tout  $q \geq 1$ , c'est à dire que pour tout  $q \geq 1$  il existe  $n$  assez grand tel que  $M_{k^2}(\omega) \leq \frac{1}{q}$  dès que  $k \geq n$ . En d'autres termes,  $M_{n^2}(\omega) \rightarrow 0$  pour tout  $\omega \notin N$ , ce qui est bien le résultat voulu.

2) Montrons maintenant que la suite  $(M_n)_n$  tend presque-sûrement vers 0. Pour tout entier  $n$ , notons  $p_n$  l'entier tel que  $p_n^2 \leq n < (p_n + 1)^2$ . Comme les v.a.  $X_i$  sont iid,

$$\mathbb{E}\left(\left(M_n - \frac{p(n)^2}{n} M_{p_n^2}\right)^2\right) = \mathbb{E}\left(\left(\frac{1}{n} \sum_{p=p_n^2+1}^n X_p\right)^2\right) = \frac{n - p_n^2}{n^2} \sigma^2 \leq \frac{2\sqrt{n} + 1}{n^2} \sigma^2,$$

puisque  $n - p_n^2 \leq (1 + p_n)^2 - p_n^2 = 1 + 2p_n \leq 1 + 2\sqrt{n}$ . Une nouvelle application de l'inégalité de Bienaymé-Chebyshev donne alors

$$\mathbb{P}(B_{n,q}) \leq \frac{2\sqrt{n} + 1}{n^2} \frac{\sigma^2}{q^2}, \quad \text{où} \quad B_{n,q} = \left| M_n - \frac{p_n^2}{n} M_{p_n^2} \right| \geq q, \quad q \in \mathbb{N}^*.$$



Comme la série  $\sum_n \frac{2\sqrt{n}+1}{n^2}$  converge, le même raisonnement que la première étape montre que  $M_n - \frac{p_n^2}{n} M_{p_n^2} \rightarrow 0$  quand  $n \rightarrow \infty$ ,  $\mathbb{P}$ -p.s. Avec le résultat de la première étape, et le fait que  $p_n^2/n \rightarrow 1$ , ceci implique que  $M_n \rightarrow 0$  quand  $n \rightarrow \infty$ ,  $\mathbb{P}$ -p.s.

### 6.3 Méthode de Monte-Carlo

Montrons comment nous pouvons appliquer la loi des grands nombres au calcul d'intégrales. Pour cela, énonçons tout d'abord un corollaire immédiat de la loi des grands nombres.

**Corollaire 6.17** *Soient  $(X_n)_n$  une suite de v.a. iid de loi uniforme sur  $[0, 1]$ , et  $f$  une fonction mesurable bornée sur  $[0, 1]$ . Alors*

$$\frac{f(X_1) + \cdots + f(X_n)}{n} \rightarrow \int_0^1 f(x)dx, \quad \text{quand } n \rightarrow \infty, \quad p.s.$$

**Preuve.** Nous appliquons la loi des grands nombres aux v.a.  $f(X_i)$  qui vérifient bien toutes les hypothèses voulues puisque  $f$  est bornée. Nous avons alors

$$\lim_n \frac{f(X_1) + \cdots + f(X_n)}{n} = \mathbb{E}(f(X)) = \int_0^1 f(x)dx.$$

□

En choisissant des v.a. de loi uniforme sur  $[a, b]$ , nous pouvons obtenir de même une approximation d'une intégrale définie sur l'intervalle  $[a, b]$ .

Ce résultat se généralise à toutes les dimensions.

Nous voulons calculer l'intégrale  $I = \int_A f(\mathbf{x})d\mathbf{x}$ , où  $f$  est une fonction mesurable bornée de  $\mathbb{R}^d$  dans  $\mathbb{R}$  et  $A$  est le cube  $\{\mathbf{x} = (x_1, \dots, x_d) : |x_i| \leq \alpha \ \forall i\}$  de  $\mathbb{R}^d$ . Pour calculer  $I$ , nous pouvons simuler une suite  $\mathbf{X}_1, \dots, \mathbf{X}_n$  de v.a. indépendantes et de loi uniforme sur  $A$ . Cela revient à dire que si chaque  $\mathbf{X}_n$  admet les composantes  $X_{n,j}$  ( $1 \leq j \leq d$ ), les v.a.  $(X_{n,j} : n \geq 1, 1 \leq j \leq d)$  sont indépendantes et uniformes sur  $[-\alpha, \alpha]$ . Une suite de valeurs approchées de  $I$  est alors

$$I_n = \frac{(2\alpha)^d}{n} (f(\mathbf{X}_1) + \cdots + f(\mathbf{X}_n)). \quad (6.7)$$

En effet la loi uniforme sur  $A$  admet la densité  $g(\mathbf{x}) = \frac{1}{(2\alpha)^d} 1_A(\mathbf{x})$ , donc l'espérance des  $f(\mathbf{X}_i)$  est égale à  $\frac{I}{(2\alpha)^d}$ , et il s'ensuit que  $I_n$  converge vers  $I$  par la loi des grands nombres.

L'inconvénient de cette méthode est que  $I_n$  est une approximation "aléatoire" de  $I$ , donc on a un peu de peine à contrôler l'erreur  $I_n - I$ . Toutefois, le deuxième théorème fondamental de ce cours, à savoir le théorème de la limite centrale qui sera l'objet du chapitre suivant, va donner un contrôle de cette erreur comme on le montre dans la section 9.3.

Un avantage de cette méthode est qu'elle reste valable si la fonction  $f$  est très irrégulière (alors que les méthodes déterministes de type "méthode du trapèze" ne se justifient que si la fonction  $f$  est continue). En outre, à précision donnée, elle est peu sensible à la dimension  $d$ , le temps de calcul étant proportionnel à  $d$ , alors que les méthodes déterministes ne sont possibles, du point de vue du temps de calcul, que pour  $d$  petit, (disons  $d \leq 3$ ), puisque ce temps de calcul est environ proportionnel à une constante à la puissance  $d$ . Dans notre cas, tirer une variable  $X$  de loi uniforme sur  $A$  revient à tirer ses  $d$  composantes, chacune selon la loi uniforme sur  $[0, 1]$ . Nous verrons également que la vitesse de convergence de  $I_n$  vers  $I$  ne dépend pas non plus de la dimension.

Pour toutes ces raisons, les algorithmes obtenus par méthodes de Monte-Carlo sont extrêmement utilisés dans toutes les situations nécessitant des temps de calcul très courts ou en grande dimension.

## 6.4 Exercices sur le chapitre 6

**EXERCICE 6.1** Soit  $(u_n)_n$  une suite de nombres réels tels que pour tout  $n \geq 1$ , on ait  $0 < u_n \leq 1$ . Soit  $(X_n)_n$  une suite de v.a. indépendantes telles que

$$\mathbb{P}\left(X_n = \frac{1}{u_n}\right) = u_n; \quad \mathbb{P}(X_n = 0) = 1 - u_n \quad \text{pour tout } n \geq 1.$$

- 1) Calculer  $\mathbb{E}(X_n)$ .
- 2) On suppose que  $\sum_{n \geq 1} u_n < +\infty$ . En déduire que  $X_n \xrightarrow{\text{P}} 0$ . Montrer en fait que  $X_n \xrightarrow{\text{P.S.}} 0$  quand  $n$  tend vers l'infini.
- 3) Montrer que la suite  $\frac{X_1 + \dots + X_n}{n} \xrightarrow{\text{P.S.}} 0$  quand  $n$  tend vers l'infini. Ce résultat est-il en contradiction avec la loi des grands nombres ?

**EXERCICE 6.2** Montrer que la loi forte des grands nombres reste vraie pour des variables aléatoires indépendantes positives de même loi, d'espérance commune égale à  $+\infty$ , c'est-à-dire que l'on a  $\frac{X_1 + \dots + X_n}{n} \xrightarrow{\text{P.S.}} +\infty$ .

**EXERCICE 6.3** Soit  $(X_n)_n$  une suite de variables aléatoires indépendantes, de même loi de Bernoulli

$$\mathbb{P}(X_n = 1) = x ; \mathbb{P}(X_n = 0) = 1 - x, \quad \text{où } x \in ]0, 1[.$$

1) Montrer que pour toute fonction continue de  $[0, 1]$  dans  $\mathbb{R}$ ,

$$\mathbb{E} \left( f \left( \frac{X_1 + \dots + X_n}{n} \right) \right) \xrightarrow{n \rightarrow \infty} f(x).$$

En déduire qu'il existe une suite de polynômes qui convergent simplement vers  $f$ . On les appelle polynômes de Bernstein.

2) Montrer qu'en fait, la convergence est uniforme.

**EXERCICE 6.4** *Inégalité de Chernov.* Soit  $(X_n)_n$  une suite de v.a. réelles indépendantes de loi  $\mathcal{N}(m, \sigma^2)$ .

1) On pose pour  $u \in \mathbb{R}$  :  $M(u) = \mathbb{E}(e^{u(X_1 - m)})$ . Calculer  $M(u)$ .

2) On pose  $S_n = X_1 + \dots + X_n$ . Montrer pour tout  $a \in \mathbb{R}$  que

$$\begin{aligned} \mathbb{P}(S_n - nm \geq a) &\leq e^{-ua} (M(u))^n, \quad \text{pour tout } u \in \mathbb{R}_+, \\ \mathbb{P}(S_n - nm \leq a) &\leq e^{-ua} (M(u))^n \quad \text{pour tout } u \in \mathbb{R}_-. \end{aligned}$$

3) Soit  $Y_n = \frac{S_n}{n}$ . Montrer que  $\forall \varepsilon > 0$ ,

$$\mathbb{P}(|Y_n - m| \geq \varepsilon) \leq 2 \exp \left( \frac{-n\varepsilon^2}{2\sigma^2} \right).$$

Cette inégalité est appelée inégalité de Chernov.

4) On suppose que  $m = 1$  et que  $\sigma^2 = 10$ .

Avec  $\alpha = 0,95$  et  $\varepsilon = 0,05$ , quelle taille d'échantillon doit-on choisir pour obtenir

$$\mathbb{P}(|Y_n - m| \leq \varepsilon) \geq \alpha,$$

- En utilisant l'inégalité de Bienaymé-Chebyshev ?
- En utilisant l'inégalité de Chernov ?

**EXERCICE 6.5 (Placement risqué)** Un particulier place une somme  $S_0$  sur un placement risqué. L'évolution de son placement à des échéances fixes  $n = 1, 2, \dots$  est donnée par  $S_n = (1 + R_n)S_{n-1}$ , où les intérêts aléatoires  $R_n$  forment une suite indépendante et de même loi à valeurs dans  $] -1, \infty[$  et d'espérance finie. Que pouvez-vous dire de l'évolution du placement  $S_n$  pour de grands  $n$ ? Que se passe-t-il si  $\mathbb{E}(R_1) = 0$  ?

**EXERCICE 6.6** Soient  $(X_n)_n$  une suite de variables aléatoires indépendantes de même loi, intégrables, et  $f$  une fonction continue bornée sur  $\mathbb{R}$ . On pose  $\mathbb{E}(X_1) = a$ .

- 1) Montrer que  $f\left(\frac{X_1+\dots+X_n}{n}\right) \rightarrow f(a)$  p.s.
- 2) Montrer que  $\mathbb{E}\left(f\left(\frac{X_1+\dots+X_n}{n}\right)\right) \xrightarrow{n \rightarrow \infty} f(a)$ .
- 3) Soit  $g \in \mathcal{C}([0, 1])$ . Calculer  $\lim_n I_n$ , où  $I_n = \int_{[0,1]^n} g\left(\frac{x_1+\dots+x_n}{n}\right) dx_1 \cdots dx_n$ .
- 4) En utilisant des variables aléatoires de loi  $\mathcal{E}\left(\frac{1}{a}\right)$ , montrer que

$$f(a) = \lim_{n \rightarrow \infty} \frac{(-1)^{n-1}}{(n-1)!} \left(\frac{n}{a}\right)^n F^{(n-1)}\left(\frac{n}{a}\right)$$

où  $F(t) = \int_0^\infty f(x)e^{-tx}dx$ . On pourra montrer que la somme de  $n$  variables aléatoires indépendantes de loi exponentielle de paramètre  $\alpha$  suit une loi Gamma de paramètres  $(n, \alpha)$ , c'est-à-dire de densité  $x \mapsto \frac{\alpha^n}{(n-1)!}x^{n-1}e^{-\alpha x}$  pour  $x > 0$ .

## Chapitre 7

# Fonctions caractéristiques, convergence en loi et théorème de la limite centrale

*Rien ne m'est sûr que la chose incertaine ; Obscur, fors ce qui est tout évident ;  
Doute ne fais, fors en chose certaine ; Science tiens à soudain accident.*

*François Villon - Ballade du concours de Blois*

### 7.1 La fonction caractéristique

Dans ce paragraphe, nous introduisons un outil important en calcul des probabilités : il s'agit de ce que l'on appelle la **fonction caractéristique** d'une variable aléatoire, et qui dans d'autres branches des mathématiques s'appelle aussi la **transformée de Fourier**.

#### 7.1.1 Définition et premières propriétés

Nous noterons  $\langle \mathbf{x}, \mathbf{y} \rangle$  le produit scalaire de deux vecteurs de  $\mathbb{R}^n$ . Si  $\mathbf{u} \in \mathbb{R}^n$ , la fonction (complexe)  $\mathbf{x} \mapsto e^{i \langle \mathbf{u}, \mathbf{x} \rangle}$  est continue, de module 1. Donc si  $\mathbf{X}$  est un vecteur aléatoire à valeurs dans  $\mathbb{R}^n$ , nous pouvons considérer  $e^{i \langle \mathbf{u}, \mathbf{X} \rangle}$  comme une variable aléatoire à

valeurs complexes. Ses parties réelle  $Y = \cos(\langle \mathbf{u}, \mathbf{X} \rangle)$  et imaginaire  $Z = \sin(\langle \mathbf{u}, \mathbf{X} \rangle)$  sont des v.a. réelles. Ces v.a. réelles sont de plus bornées par 1, donc elles admettent une espérance. Il est alors naturel d'écrire que l'espérance de  $e^{i\langle \mathbf{u}, \mathbf{X} \rangle}$  est

$$\mathbb{E}(e^{i\langle \mathbf{u}, \mathbf{X} \rangle}) = \mathbb{E}(Y) + i\mathbb{E}(Z) = \mathbb{E}(\cos(\langle \mathbf{u}, \mathbf{X} \rangle)) + i\mathbb{E}(\sin(\langle \mathbf{u}, \mathbf{X} \rangle)).$$

**Définition 7.1** Si  $\mathbf{X}$  est un vecteur aléatoire à valeurs dans  $\mathbb{R}^n$ , sa **fonction caractéristique** est la fonction  $\phi_{\mathbf{X}}$  de  $\mathbb{R}^n$  dans  $\mathbb{C}$  définie par

$$\phi_{\mathbf{X}}(\mathbf{u}) = \mathbb{E}\left(e^{i\langle \mathbf{u}, \mathbf{X} \rangle}\right). \quad (7.1)$$

Si  $X$  est à valeurs réelles,  $\phi_X$  est définie de  $\mathbb{R}$  dans  $\mathbb{C}$  et vaut

$$\forall u \in \mathbb{R}, \quad \phi_X(u) = \mathbb{E}(e^{iuX}). \quad (7.2)$$

Remarquons que la fonction caractéristique ne dépend en fait **que de la loi**  $\mathbb{P}_{\mathbf{X}}$  de  $\mathbf{X}$  : c'est la "transformée de Fourier" de la loi  $\mathbb{P}_{\mathbf{X}}$ .

Nous verrons que cette fonction porte bien son nom, au sens où elle caractérise la loi  $\mathbb{P}_{\mathbf{X}}$ . C'est une notion qui, de ce point de vue, généralise la fonction génératrice que nous avons vue au Chapitre 2. Elle vérifie

$$\phi_X(u) = G_X(e^{iu}),$$

pour une variables aléatoire  $X$  à valeurs dans  $\mathbb{N}$ .

**Proposition 7.2** La fonction  $\phi_{\mathbf{X}}$  est de module inférieur ou égal à 1, continue, avec

$$\phi_{\mathbf{X}}(0) = 1; \quad \phi_{\mathbf{X}}(-\mathbf{u}) = \overline{\phi_{\mathbf{X}}(\mathbf{u})}.$$

**Preuve.**  $|z|$  désigne le module d'un nombre complexe  $z$ .

Comme  $\mathbb{E}(Y)^2 \leq \mathbb{E}(Y^2)$  pour toute variable aléatoire réelle  $Y$ , nous avons :

$$|\phi_{\mathbf{X}}(\mathbf{u})|^2 = (\mathbb{E}(\cos(\langle \mathbf{u}, \mathbf{X} \rangle)))^2 + (\mathbb{E}(\sin(\langle \mathbf{u}, \mathbf{X} \rangle)))^2 \leq \mathbb{E}(\cos^2(\langle \mathbf{u}, \mathbf{X} \rangle) + \sin^2(\langle \mathbf{u}, \mathbf{X} \rangle)),$$

et donc  $|\phi_{\mathbf{X}}(\mathbf{u})| \leq 1$ .

Pour montrer la continuité, considérons une suite  $\mathbf{u}_p \rightarrow_{p \rightarrow \infty} \mathbf{u}$ . Il y a convergence simple de  $e^{i\langle \mathbf{u}_p, \mathbf{X} \rangle}$  vers  $e^{i\langle \mathbf{u}, \mathbf{X} \rangle}$ . Comme ces variables aléatoires sont de module borné par 1, nous pouvons appliquer le théorème de convergence dominée (Théorème 6.7), et obtenir que  $\phi_{\mathbf{X}}(\mathbf{u}_p) \xrightarrow{p \rightarrow \infty} \phi_{\mathbf{X}}(\mathbf{u})$ . Ainsi, la fonction  $\phi_{\mathbf{X}}$  est continue.  $\square$

**Proposition 7.3** Si  $\mathbf{X}$  est un vecteur aléatoire à valeurs dans  $\mathbb{R}^n$ , si  $\mathbf{a} \in \mathbb{R}^m$  et si  $\mathbf{A}$  est une matrice de taille  $m \times n$ , nous avons :

$$\phi_{\mathbf{a} + \mathbf{A}\mathbf{X}}(\mathbf{u}) = e^{i\langle \mathbf{u}, \mathbf{a} \rangle} \phi_{\mathbf{X}}(\mathbf{A}^t \mathbf{u}), \quad \forall \mathbf{u} \in \mathbb{R}^m, \quad (7.3)$$

où  $\mathbf{A}^t$  désigne la transposée de la matrice  $\mathbf{A}$ .

**Preuve.** Nous avons  $e^{i\langle \mathbf{u}, \mathbf{a} + \mathbf{A}\mathbf{X} \rangle} = e^{i\langle \mathbf{u}, \mathbf{a} \rangle} e^{i\langle \mathbf{A}^t \mathbf{u}, \mathbf{X} \rangle}$ . En effet,  $\langle \mathbf{u}, \mathbf{A}\mathbf{X} \rangle = \langle \mathbf{A}^t \mathbf{u}, \mathbf{X} \rangle$ . Il suffit alors de prendre les espérances pour obtenir le résultat.  $\square$

### 7.1.2 Exemples

1)  $X$  suit une loi binomiale  $\mathcal{B}(n, p)$

$$\begin{aligned} \mathbb{E}(e^{iuX}) &= \sum_{k=0}^n e^{iuk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n \binom{n}{k} (e^{iu}p)^k (1-p)^{n-k} \\ &= (e^{iu}p + 1-p)^n. \end{aligned}$$

Ainsi,

$$\phi_X(u) = (pe^{iu} + 1-p)^n. \quad (7.4)$$

2)  $X$  suit une loi de Poisson de paramètre  $\theta$

$$\mathbb{E}(e^{iuX}) = \sum_{k \geq 0} e^{iuk} \frac{\theta^k}{k!} e^{-\theta} = e^{\theta e^{iu}} e^{-\theta} = e^{\theta(e^{iu}-1)}.$$

Ainsi,

$$\phi_X(u) = e^{\theta(e^{iu}-1)}. \quad (7.5)$$

3)  $X$  suit une loi uniforme sur  $[a, b]$

$$\phi_X(u) = \frac{1}{b-a} \int_a^b e^{iux} dx = \frac{1}{b-a} \frac{1}{iu} [e^{iux}]_a^b = \frac{e^{iub} - e^{iua}}{iu(b-a)}.$$

Ainsi, pour une loi uniforme sur  $[-a, a]$ ,  $a > 0$ ,

$$\phi_X(u) = \frac{\sin ua}{ua}.$$

4)  $X$  suit une loi exponentielle de paramètre  $\lambda > 0$

$$\phi_X(u) = \lambda \int_0^{+\infty} e^{-\lambda x} e^{iux} dx = \lambda \int_0^{+\infty} e^{(iu-\lambda)x} dx = \frac{\lambda}{iu-\lambda} [e^{(iu-\lambda)x}]_0^{+\infty} = \frac{\lambda}{\lambda-iu}.$$

Ainsi,

$$\phi_X(u) = \frac{\lambda}{\lambda-iu}.$$

5)  $X$  suit une loi  $\Gamma(\alpha, \theta)$

$$\phi_X(u) = \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} e^{iux} x^{\alpha-1} e^{-\theta x} dx = \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} e^{(iu-\theta)x} x^{\alpha-1} dx = \frac{\theta^\alpha}{(\theta-iu)^\alpha}.$$

Ainsi,

$$\phi_X(u) = \frac{\theta^\alpha}{(\theta-iu)^\alpha}. \quad (7.6)$$

6)  $X$  suit une loi normale  $\mathcal{N}(0, 1)$

Appelons  $g$  la densité normale (ou gaussienne) définie par  $g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ . Nous constatons tout d'abord que pour tout réel  $s$ ,

$$\int_{-\infty}^{+\infty} e^{sx} g(x) dx = e^{s^2/2}, \quad (7.7)$$

puisque  $g(x) e^{sx} = g(x-s) e^{s^2/2}$ . Nous voulons montrer que (7.7) reste vraie pour  $s$  complexe. En développant en série entière de  $s$  les deux membres de cette égalité, nous pouvons facilement justifier que

$$\int_{-\infty}^{+\infty} e^{sx} g(x) dx = \sum_{n=0}^{\infty} \frac{s^n}{n!} \int_{-\infty}^{+\infty} x^n g(x) dx \quad ; \quad e^{s^2/2} = \sum_{n=0}^{\infty} \frac{s^{2n}}{2^n n!}.$$

Par identification des coefficients de  $s^n$ , nous en déduisons les moments de  $g$ ,

$$\int_{-\infty}^{+\infty} x^{2n+1} g(x) dx = 0 \quad ; \quad \int_{-\infty}^{+\infty} x^{2n} g(x) dx = \frac{(2n)!}{2^n n!}.$$

Ce résultat peut aussi s'obtenir par intégration par parties et un calcul direct des intégrales. Le rayon de convergence de la série entière étant infini, nous déduisons de ces résultats que si  $s$  est complexe, les développements de Taylor des deux membres de (7.7) sont encore égaux, et donc que

$$\phi_X(u) = e^{-u^2/2}. \quad (7.8)$$



**Remarque 7.4** *Au cours de la preuve précédente, nous avons obtenu les moments d'une variable aléatoire  $X$  de loi normale centrée réduite,*

$$\mathbb{E}(X^{2n}) = \frac{(2n)!}{2^n n!} \quad ; \quad \mathbb{E}(X^{2n+1}) = 0.$$

Par exemple,  $\mathbb{E}(X^4) = 3$ .

7)  $X$  suit une loi normale  $\mathcal{N}(m, \sigma^2)$

Dans ce cas, la v.a.  $X$  s'écrit  $X = m + \sigma Y$ , où  $Y$  suit la loi  $\mathcal{N}(0, 1)$ . D'après (7.3) et (7.8), sa fonction caractéristique vaut alors

$$\phi_X(u) = e^{i u m - u^2 \sigma^2 / 2}. \quad (7.9)$$

### 7.1.3 Propriété fondamentale

L'intérêt majeur de la fonction caractéristique réside dans le fait qu'elle caractérise la loi de la variable aléatoire.

**Théorème 7.5** *La fonction caractéristique  $\phi_{\mathbf{X}}$  caractérise la loi du vecteur aléatoire  $\mathbf{X}$ . Ainsi, si deux vecteurs aléatoires  $\mathbf{X}$  et  $\mathbf{Y}$  ont même fonction caractéristique, ils ont même loi.*

**Preuve.** La preuve montre que la transformée de Fourier d'une probabilité sur  $\mathbb{R}^n$  caractérise cette probabilité. Introduisons, pour  $\sigma > 0$ , les fonctions

$$f_\sigma(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-|\mathbf{x}|^2 / 2\sigma^2} \quad \text{et} \quad \hat{f}_\sigma(\mathbf{u}) = e^{-|\mathbf{u}|^2 \sigma^2 / 2}.$$

Alors  $f_\sigma(\mathbf{x})$  est la densité d'un vecteur aléatoire composé de  $n$  coordonnées indépendantes et de loi  $\mathcal{N}(0, \sigma^2)$ . Nous avons

$$\int_{\mathbb{R}^n} f_\sigma(\mathbf{x}) e^{i\langle \mathbf{u}, \mathbf{x} \rangle} d\mathbf{x} = \int_{\mathbb{R}^n} \prod_{j=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-x_j^2}{2\sigma^2} + i u_j x_j\right) dx_1 \cdots dx_n = \prod_{j=1}^n e^{-u_j^2 \sigma^2 / 2} = \hat{f}_\sigma(\mathbf{u}),$$

par le théorème de Fubini.

Remarquons ainsi que  $f_\sigma(\mathbf{u}-\mathbf{v}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \hat{f}_\sigma\left(\frac{\mathbf{u}-\mathbf{v}}{\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{n/2}} \int_{\mathbb{R}^n} f_\sigma(\mathbf{x}) e^{i\langle \frac{\mathbf{u}-\mathbf{v}}{\sigma^2}, \mathbf{x} \rangle} d\mathbf{x}$ .

Considérons deux vecteurs aléatoires  $\mathbf{X}$  et  $\mathbf{Y}$  de même fonction caractéristique. Montrons qu'ils ont même loi :  $\mathbb{P}_{\mathbf{X}} = \mathbb{P}_{\mathbf{Y}}$ . Nous avons (en utilisant le théorème de Fubini) que

$$\begin{aligned} \int_{\mathbb{R}^n} f_{\sigma}(\mathbf{u} - \mathbf{v}) \mathbb{P}_{\mathbf{X}}(d\mathbf{u}) &= \int_{\mathbb{R}^n} \frac{1}{(2\pi\sigma^2)^{n/2}} \left( \int_{\mathbb{R}^n} f_{\sigma}(\mathbf{x}) e^{i\langle \frac{\mathbf{u}-\mathbf{v}}{\sigma^2}, \mathbf{x} \rangle} d\mathbf{x} \right) \mathbb{P}_{\mathbf{X}}(d\mathbf{u}) \\ &= \int_{\mathbb{R}^n} f_{\sigma}(\mathbf{x}) \frac{1}{(2\pi\sigma^2)^{n/2}} \phi_{\mathbf{X}}\left(\frac{\mathbf{x}}{\sigma^2}\right) e^{i\langle \frac{-\mathbf{v}}{\sigma^2}, \mathbf{x} \rangle} d\mathbf{x}, \end{aligned}$$

et la même égalité reste vraie pour  $\mathbb{P}_{\mathbf{Y}}$ . Nous en déduisons que

$$\int_{\mathbb{R}^n} g(\mathbf{x}) \mathbb{P}_{\mathbf{X}}(d\mathbf{x}) = \int_{\mathbb{R}^n} g(\mathbf{x}) \mathbb{P}_{\mathbf{Y}}(d\mathbf{x}), \quad (7.10)$$

pour toute fonction  $g$  de l'espace vectoriel des fonctions engendrées par  $\mathbf{u} \mapsto f_{\sigma}(\mathbf{u} - \mathbf{v})$ . Nous pouvons appliquer le théorème de Stone-Weierstrass pour montrer que cet ensemble est dense dans l'ensemble des fonctions continues sur  $\mathbb{R}^n$  qui tendent vers 0 à l'infini, muni de la convergence uniforme. L'égalité (7.10) reste vraie pour de telles fonctions. Un argument de théorie de la mesure montre que cela suffit pour en déduire que  $\mathbb{P}_{\mathbf{X}} = \mathbb{P}_{\mathbf{Y}}$ .  $\square$

**Corollaire 7.6** Soit  $\mathbf{X} = (X_1, \dots, X_n)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^n$ . Les composantes  $X_i$  sont indépendantes si et seulement si pour tous  $u_1, \dots, u_n \in \mathbb{R}$

$$\phi_{\mathbf{X}}(u_1, \dots, u_n) = \prod_{j=1}^n \phi_{X_j}(u_j), \quad (7.11)$$

où  $\phi_{\mathbf{X}}$  désigne la fonction caractéristique du vecteur aléatoire  $\mathbf{X}$ , et  $\phi_{X_j}$  celle de la composante  $X_j$ , pour chaque  $j$ .

**Preuve.** Nous avons  $\langle \mathbf{u}, \mathbf{X} \rangle = \sum_{j=1}^n u_j X_j$ . Si les  $X_i$  sont indépendantes et comme  $e^{i\langle \mathbf{u}, \mathbf{X} \rangle} = \prod_j e^{i u_j X_j}$ , nous obtenons immédiatement (7.11), en utilisant (4.18).

Supposons inversement qu'on ait (7.11). Nous pouvons alors construire des variables aléatoires  $X'_j$  indépendantes, telles que  $X'_j$  et  $X_j$  aient mêmes lois pour tout  $j$  et donc telles que  $\phi_{X'_j} = \phi_{X_j}$ . Si  $\mathbf{X}' = (X'_1, \dots, X'_n)$ , alors, en utilisant la condition nécessaire et (7.11), nous en déduisons que  $\phi_{\mathbf{X}'} = \phi_{\mathbf{X}}$ . Ainsi,  $\mathbf{X}$  et  $\mathbf{X}'$  ont même loi, ce qui entraîne que pour tous boréliens  $A_j$ ,

$$\mathbb{P}\left(\bigcap_j \{X_j \in A_j\}\right) = \mathbb{P}\left(\bigcap_j \{X'_j \in A_j\}\right) = \prod_j \mathbb{P}(X'_j \in A_j) = \prod_j \mathbb{P}(X_j \in A_j),$$

d'où l'indépendance cherchée.  $\square$

**La transformée de Laplace :** Lorsque  $X$  est une v.a. à valeurs dans  $\mathbb{R}_+$ , nous pouvons définir sa transformée de Laplace par

$$\psi_X(\lambda) = \mathbb{E}(e^{-\lambda X}), \quad \lambda \in \mathbb{R}_+. \quad (7.12)$$

C'est une fonction définie sur  $\mathbb{R}_+$ , indéfiniment dérivable sur  $]0, \infty[$ , et qui satisfait formellement  $\psi_X(\lambda) = \phi_X(i\lambda)$ . Ainsi, il n'est pas étonnant que la transformée de Laplace ait des propriétés analogues à celles de la fonction caractéristique. En particulier, elle caractérise la loi  $\mathbb{P}_X$ .

Si de plus  $X$  est une v.a. à valeurs dans  $\mathbb{N}$ , de fonction génératrice  $G_X$ , alors  $\psi_X(\lambda) = G_X(e^{-\lambda})$ .

**Exemple 7.7 Transformée de Laplace d'une loi gamma.** Soit  $X$  une variable aléatoire de loi  $\Gamma(\alpha, \theta)$ . Alors,

$$\psi_X(\lambda) = \frac{1}{\Gamma(\alpha)} \theta^\alpha \int_0^{+\infty} e^{-\lambda x} x^{\alpha-1} e^{-\theta x} dx = \frac{1}{\Gamma(\alpha)} \theta^\alpha \int_0^{+\infty} e^{-(\lambda+\theta)x} x^{\alpha-1} dx = \frac{\theta^\alpha}{(\lambda+\theta)^\alpha}. \quad (7.13)$$

En particulier la transformée de Laplace d'une v.a. de loi exponentielle de paramètre  $\theta$ , prise en  $\lambda$ , vaut  $\frac{\theta}{\lambda+\theta}$ .

#### 7.1.4 Somme de vecteurs aléatoires indépendants

**Proposition 7.8** Si  $\mathbf{X}$  et  $\mathbf{Y}$  sont deux vecteurs aléatoires indépendants à valeurs dans  $\mathbb{R}^n$ , la fonction caractéristique de la somme  $\mathbf{X} + \mathbf{Y}$  est donnée par

$$\phi_{\mathbf{X}+\mathbf{Y}} = \phi_{\mathbf{X}} \phi_{\mathbf{Y}}. \quad (7.14)$$

**Preuve.** Comme  $e^{i\langle \mathbf{u}, \mathbf{X}+\mathbf{Y} \rangle} = e^{i\langle \mathbf{u}, \mathbf{X} \rangle} e^{i\langle \mathbf{u}, \mathbf{Y} \rangle}$ , il suffit d'appliquer (4.18).  $\square$

**Exemple 7.9** Soient  $X, Y$  deux v.a. réelles indépendantes, et  $Z = X + Y$  :

- 1) si  $X$  suit loi normale  $\mathcal{N}(m, \sigma^2)$  et  $Y$  suit une loi  $\mathcal{N}(m', (\sigma')^2)$ , alors  $Z$  suit une loi  $\mathcal{N}(m + m', \sigma^2 + (\sigma')^2)$ ;

Cela découle de (7.9) et (7.14).

- 2) si  $X$  et  $Y$  suivent des lois de Poisson de paramètres  $\theta$  et  $\theta'$ , alors  $Z$  suit une loi de Poisson de paramètre  $\theta + \theta'$ ;

Cela découle de (7.5) et (7.14).

- 3) si  $X$  suit une loi binomiale  $\mathcal{B}(n, p)$  et  $Y$  suit une loi  $\mathcal{B}(m, p)$ , alors  $Z$  suit une loi  $\mathcal{B}(n + m, p)$ ;

Cela découle de (7.4) et (7.14).

- 4) si  $X$  suit une loi  $\Gamma(\alpha, \theta)$  et  $Y$  suit une loi  $\Gamma(\alpha', \theta)$ , alors  $Z$  suit une loi  $\Gamma(\alpha + \alpha', \theta)$ ;

Cela découle de (7.6) et (7.14).

### 7.1.5 Fonction caractéristique et moments

**Proposition 7.10** Soit  $\mathbf{X}$  un vecteur aléatoire de  $\mathbb{R}^n$ . Si la v.a.  $|\mathbf{X}|^m$  (où  $|\mathbf{x}|$  désigne la norme euclidienne du vecteur  $\mathbf{x}$ ) est dans  $L^1$  pour un entier  $m$ , la fonction  $\phi_{\mathbf{X}}$  est  $m$  fois continûment différentiable sur  $\mathbb{R}^n$ , et pour tout choix des indices  $i_1, \dots, i_m$ ,

$$\frac{\partial^m}{\partial u_{i_1} \partial u_{i_2} \cdots \partial u_{i_m}} \phi_{\mathbf{X}}(\mathbf{u}) = i^m \mathbb{E} \left( e^{i \langle \mathbf{u}, \mathbf{X} \rangle} X_{i_1} X_{i_2} \cdots X_{i_m} \right) \quad (7.15)$$

(les  $X_j$  sont les composantes de  $\mathbf{X}$ ).

En prenant  $\mathbf{u} = 0$  ci-dessus, cette formule permet de calculer  $\mathbb{E}(X_{i_1} X_{i_2} \cdots X_{i_m})$  en fonction des dérivées à l'origine de  $\phi_{\mathbf{X}}$ . Par exemple, si  $X$  est à valeurs réelles et est intégrable (respectivement de carré intégrable), nous avons

$$\mathbb{E}(X) = -i \phi'_X(0), \quad (\text{resp. } \mathbb{E}(X^2) = -\phi''_X(0)). \quad (7.16)$$

**Preuve.** Prouvons le résultat quand  $m = 1$ , le cas général se montrant de la même manière, par récurrence sur  $m$ . Soit  $\mathbf{v}_j = (0, \dots, 0, 1, 0, \dots, 0)$  le  $j^{\text{ième}}$  vecteur de base

de  $\mathbb{R}^n$ . Nous avons

$$\frac{\phi_{\mathbf{X}}(\mathbf{u} + t\mathbf{v}_j) - \phi_{\mathbf{X}}(\mathbf{u})}{t} = \mathbb{E} \left( e^{i\langle \mathbf{u}, \mathbf{X} \rangle} \frac{e^{itX_j} - 1}{t} \right). \quad (7.17)$$

Soit  $t_p$  une suite de réels tendant vers 0. Les v.a.  $(e^{it_p X_j} - 1)/t_p$  convergent simplement vers  $iX_j$ , en restant bornées en module par la v.a.  $|X_j|$  (car  $|e^{itx} - 1| = 2|\sin(tx/2)| \leq |tx|$ ), qui par hypothèse est intégrable. Donc par le théorème 6.7 (de Lebesgue), nous en déduisons que (7.17) converge vers  $i\mathbb{E}(e^{i\langle \mathbf{u}, \mathbf{X} \rangle} X_j)$  quand  $t$  tend vers 0. Nous en déduisons que la première dérivée partielle de  $\phi_{\mathbf{X}}$  par rapport à  $u_j$  existe et est donnée par la formule (7.15). Enfin, nous pouvons montrer comme dans la Proposition 7.2 que cette dérivée est continue.  $\square$

## 7.2 Vecteurs gaussiens

**Définition 7.11** Un vecteur aléatoire  $\mathbf{X} = (X_1, \dots, X_n)$  à valeurs dans  $\mathbb{R}^n$  est appelé un **vecteur gaussien** si toute combinaison linéaire  $\sum_{j=1}^n a_j X_j = \langle \mathbf{a}, \mathbf{X} \rangle$ , pour  $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$ , suit une loi normale (avec la convention que la masse de Dirac au point  $m$  est la “loi normale”  $\mathcal{N}(m, 0)$ ).

Cela entraîne bien entendu que chaque composante  $X_j$  suit elle-même une loi normale.

**Exemple 7.12** Si les  $X_i$  sont des v.a. normales indépendantes, le vecteur  $\mathbf{X}$  est gaussien (cela découle de l'exemple 7.9 (1)).

**Contre-exemple** Si les composantes  $X_i$  sont de loi normale mais pas indépendantes, il se peut que  $\mathbf{X}$  ne soit pas un vecteur gaussien. Prenons par exemple  $X_1$  de loi  $\mathcal{N}(0, 1)$ , et

$$X_2 = \begin{cases} X_1 & \text{si } |X_1| \leq 1 \\ -X_1 & \text{sinon.} \end{cases}$$

Alors  $X_2$  suit également la loi  $\mathcal{N}(0, 1)$ , mais  $\mathbf{X} = (X_1, X_2)$  n'est pas un vecteur gaussien, puisque  $0 < \mathbb{P}(X_1 + X_2 = 0) < 1$  (donc  $X_1 + X_2$  ne suit pas une loi normale).

**Théorème 7.13**  $\mathbf{X}$  est un vecteur gaussien si et seulement si sa fonction caractéristique s'écrit

$$\phi_{\mathbf{X}}(\mathbf{u}) = e^{i\langle \mathbf{u}, \mathbf{m} \rangle - \frac{1}{2} \langle \mathbf{u}, \mathbf{C} \mathbf{u} \rangle}, \quad (7.18)$$

où  $\mathbf{m} \in \mathbb{R}^n$  et  $\mathbf{C}$  est une matrice de taille  $n \times n$  symétrique positive ; dans ce cas,  $\mathbf{m} = \mathbb{E}(\mathbf{X})$  (i.e.  $m_j = \mathbb{E}(X_j)$  pour chaque  $j$ ), et  $\mathbf{C}$  est la matrice de covariance de  $\mathbf{X}$ .

**Preuve.** a) Condition suffisante : supposons (7.18). Pour toute combinaison linéaire  $Y = \sum_j a_j X_j = \langle \mathbf{a}, \mathbf{X} \rangle$ , et pour  $v \in \mathbb{R}$ , nous avons

$$\phi_Y(v) = \phi_{\mathbf{X}}(v\mathbf{a}) = e^{iv \langle \mathbf{m}, \mathbf{a} \rangle - \frac{v^2}{2} \langle \mathbf{a}, \mathbf{C}\mathbf{a} \rangle},$$

donc  $Y$  suit la loi  $\mathcal{N}(\langle \mathbf{a}, \mathbf{m} \rangle, \langle \mathbf{a}, \mathbf{C}\mathbf{a} \rangle)$ .

b) Condition nécessaire : Soit  $\mathbf{C}$  la matrice de covariance de  $\mathbf{X}$  et  $\mathbf{m}$  son vecteur moyenne. Notons que ces quantités existent, car chaque composante  $X_j$  étant de loi normale, elle est de carré intégrable. Si  $Y = \langle \mathbf{a}, \mathbf{X} \rangle = \sum_{j=1}^n a_j X_j$  avec  $\mathbf{a} \in \mathbb{R}^n$ , un calcul simple montre que

$$\mathbb{E}(Y) = \langle \mathbf{a}, \mathbf{m} \rangle, \quad \text{Var}(Y) = \langle \mathbf{a}, \mathbf{C}\mathbf{a} \rangle.$$

Par hypothèse,  $Y$  suit une loi normale, donc vu ce qui précède, sa fonction caractéristique est

$$\phi_Y(v) = e^{iv \langle \mathbf{a}, \mathbf{m} \rangle - \frac{v^2}{2} \langle \mathbf{a}, \mathbf{C}\mathbf{a} \rangle}.$$

Mais  $\phi_Y(1) = \phi_{\langle \mathbf{a}, \mathbf{X} \rangle}(1) = \mathbb{E}(e^{i \langle \mathbf{a}, \mathbf{X} \rangle}) = \phi_{\mathbf{X}}(\mathbf{a})$ , d'où (7.18). □

**Corollaire 7.14** *Si  $\mathbf{X}$  est un vecteur gaussien, ses composantes sont indépendantes si et seulement si sa matrice de covariance est diagonale.*

**Attention :** ce résultat peut être faux si  $X$  n'est pas gaussien (prendre le contre-exemple précédent).

**Preuve.** Il suffit de combiner (7.18) et le corollaire 7.6. □

**Proposition 7.15** *Soit  $\mathbf{X}$  un vecteur gaussien à valeurs dans  $\mathbb{R}^n$ , de moyenne  $\mathbf{m}$ . Il existe des variables aléatoires réelles indépendantes  $Y_1, \dots, Y_n$  de lois normales  $\mathcal{N}(0, \lambda_j)$  avec  $\lambda_j \geq 0$  (si  $\lambda_j = 0$  on convient que  $Y_j = 0$ ) et une matrice orthogonale  $\mathbf{A}$  telles que  $\mathbf{X} = \mathbf{m} + \mathbf{A}\mathbf{Y}$ , où  $\mathbf{Y} = (Y_1, \dots, Y_n)$ .*

**Preuve.** Comme  $\mathbf{C}$  est une matrice symétrique positive (cf. Proposition 4.26), il existe une matrice orthogonale  $\mathbf{A}$  et une matrice diagonale  $\mathbf{L}$  dont les éléments diagonaux

vérifient  $\lambda_j \geq 0$ , et telle que la matrice de covariance de  $\mathbf{X}$  s'écrive  $\mathbf{C} = \mathbf{A}\mathbf{L}\mathbf{A}^t$ . Soit  $\mathbf{Y} = \mathbf{A}^t(\mathbf{X} - \mathbf{m})$ . Alors  $\mathbf{Y}$  est un vecteur gaussien de covariance  $\mathbf{C}' = \mathbf{A}^t\mathbf{C}\mathbf{A} = \mathbf{L}$  et de moyenne nulle. Les composantes  $Y_j$  de  $\mathbf{Y}$  répondent à la question.  $\square$

**Proposition 7.16** *Le vecteur gaussien  $\mathbf{X}$  admet une densité sur  $\mathbb{R}^n$  si et seulement si sa matrice de covariance  $\mathbf{C}$  est non-dégénérée (ou inversible, ou de valeurs propres toutes strictement positives). Dans ce cas cette densité est donnée par*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\mathbf{C})}} e^{-\frac{1}{2} \langle \mathbf{x} - \mathbf{m}, \mathbf{C}^{-1}(\mathbf{x} - \mathbf{m}) \rangle}. \quad (7.19)$$

**Preuve.** Reprenons la preuve de la proposition précédente : les  $\lambda_j$  qui y paraissent sont les valeurs propres de  $\mathbf{C}$ .

Si  $\lambda_j > 0$  pour tout  $j$ , le vecteur aléatoire  $\mathbf{Y}$  admet la densité suivante sur  $\mathbb{R}^n$  :

$$f_{\mathbf{Y}}(\mathbf{y}) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi\lambda_j}} e^{-y_j^2/(2\lambda_j)}.$$

Comme  $\mathbf{X} = \mathbf{m} + \mathbf{A}\mathbf{Y}$ , nous en déduisons que  $\mathbf{X}$  admet la densité donnée par (7.19).

Si au contraire  $\mathbf{C}$  n'est pas inversible, il existe  $\mathbf{a} \in \mathbb{R}^n$  tel que  $\mathbf{a} \neq 0$  et  $\mathbf{C}\mathbf{a} = 0$ . La v.a. réelle  $Z = \langle \mathbf{X}, \mathbf{a} \rangle$  a pour variance  $\langle \mathbf{a}, \mathbf{C}\mathbf{a} \rangle = 0$  et pour moyenne  $z = \langle \mathbf{m}, \mathbf{a} \rangle$ , donc  $\mathbb{P}(Z = z) = 1$ . Ainsi, avec une probabilité 1, le vecteur  $\mathbf{X}$  est dans un hyperplan  $H$  orthogonal à  $\mathbf{a}$ , i.e.  $\mathbb{P}(\mathbf{X} \in H) = 1$ . Or, si  $\mathbf{X}$  admettait la densité  $f$ , nous aurions  $\mathbb{P}(\mathbf{X} \in H) = \int_H f(\mathbf{x}) d\mathbf{x}$ , donc  $\mathbb{P}(\mathbf{X} \in H) = 0$  puisque le "volume" de l'hyperplan  $H$  est nul.  $\square$

**Définition 7.17** On appelle v.a. de  $\chi^2$  (chi-deux) à  $d$  degrés de liberté toute v.a. positive qui a même loi que la somme de  $d$  carrés de variables aléatoires de loi normale  $N(0, 1)$  indépendantes.

L'intérêt de cette loi provient essentiellement de son utilisation massive en statistique (voir chapitres 9-10), puisqu'elle permet de construire un test (appelé test du chi-deux), fondamental pour "savoir" (statistiquement parlant) si une v.a. suit une certaine loi donnée a priori ou si deux v.a. sont indépendantes.

Grâce à l'exemple 5.27 et la proposition 5.34, nous obtenons immédiatement que la loi du  $\chi^2$  à  $d$  degrés de liberté est la loi  $\Gamma(d/2, 1/2)$  de densité définie sur  $\mathbb{R}_+$  par

$$\frac{1}{2^{d/2}\Gamma(d/2)} \exp(-y/2) y^{d/2-1}. \quad (7.20)$$

En utilisant la définition, nous obtenons immédiatement que si  $Y$  suit une loi de  $\chi^2$  à  $d$  degrés de liberté, alors

$$\mathbb{E}(Y) = d \quad , \quad \text{Var}(Y) = 2d. \quad (7.21)$$

### 7.3 Convergence en loi

Nous allons introduire maintenant une nouvelle notion de convergence de suites de variables aléatoires. (Nous renvoyons au Chapitre 5 pour les définitions des différentes notions de convergence déjà introduites.)

La convergence en loi définie dans ce paragraphe va concerner les lois des variables aléatoires. Elle signifiera que les lois sont asymptotiquement “proches”, sans que les variables aléatoires elles-mêmes le soient nécessairement. Nous verrons également que toutes les convergences introduites au Chapitre 6 entraînent la convergence en loi.

Considérons des vecteurs aléatoires  $\mathbf{X}_n$  et  $\mathbf{X}$ , tous à valeurs dans le même espace  $\mathbb{R}^d$ , mais pouvant éventuellement être définis sur des espaces de probabilité différents.

**Définition 7.18** Nous dirons que la suite  $(\mathbf{X}_n)_n$  **converge en loi** vers  $\mathbf{X}$ , et nous écrirons  $\mathbf{X}_n \xrightarrow{\mathcal{L}} \mathbf{X}$ , si pour toute fonction  $f$  **continue bornée** sur  $\mathbb{R}^d$ ,

$$\mathbb{E}(f(\mathbf{X}_n)) \xrightarrow{n \rightarrow \infty} \mathbb{E}(f(\mathbf{X})).$$

**Exemple 7.19** *Un cas très simple est celui où toutes les v.a.  $X_n$  ont un nombre fini de valeurs  $\{x_i, 1 \leq i \leq N\}$ . Alors, la suite  $(X_n)_n$  converge en loi vers  $X$  si et seulement si*

$$\lim_n \mathbb{P}(X_n = x_i) = \mathbb{P}(X = x_i) \quad \forall 1 \leq i \leq N.$$

*Il suffit d'écrire*

$$\mathbb{E}(f(X_n)) = \sum_{i=1}^N f(x_i) \mathbb{P}(X_n = x_i).$$

*Dans l'exemple ci-dessus,  $N$  est fini et fixé. Mais nous avons une définition analogue (en faisant tendre  $N$  vers l'infini), si les variables aléatoires ont un nombre dénombrable de valeurs. Au paragraphe 3.4.4, nous avons montré la convergence en loi d'une suite de variables aléatoires binomiales vers une v.a. de Poisson, pour un bon choix des paramètres.*



**Exemple 7.20** Soient  $(X_n)_n$  et  $X$  des variables aléatoires de lois respectives  $\mathcal{N}(0, \sigma_n^2)$  et  $\mathcal{N}(0, \sigma^2)$ , pas forcément définies sur le même espace de probabilité. Nous supposons que la suite de réels positifs  $(\sigma_n)_n$  converge vers  $\sigma > 0$ , quand  $n$  tend vers l'infini. Alors la suite  $(X_n)_n$  converge en loi vers  $X$ . En effet, soit  $f$  une fonction continue bornée sur  $\mathbb{R}$ . Nous avons

$$\begin{aligned} \mathbb{E}(f(X_n)) &= \int f(y) \frac{1}{\sqrt{2\pi}\sigma_n} e^{-y^2/2\sigma_n^2} dy \\ &\xrightarrow{n \rightarrow \infty} \int f(y) \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2} dy = \mathbb{E}(f(X)). \end{aligned}$$

Cette convergence est justifiée par le théorème de convergence dominée.

La convergence en loi est plus faible que la convergence en probabilité.

**Proposition 7.21** Si  $X_n \xrightarrow{P}_{n \rightarrow \infty} X$ , alors  $X_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} X$ .

Ainsi, les convergences en moyenne et presque-sûre entraînent également la convergence en loi, par la proposition 6.9.

**Preuve.** Supposons que la suite  $(X_n)_n$  converge en probabilité vers  $X$ . Soit  $f$  une fonction continue bornée sur  $\mathbb{R}^d$ . D'après la proposition 6.13, la suite  $(f(X_n))_n$  converge en probabilité vers  $f(X)$ , et comme  $f$  est bornée, la proposition 6.10 entraîne que  $f(X_n)$  converge aussi en moyenne vers  $f(X)$ . En particulier,  $(\mathbb{E}(f(X_n)))_n$  converge vers  $\mathbb{E}(f(X))$ .  $\square$

**Proposition 7.22** Soient  $X_n$  et  $X$  des v.a. réelles de fonctions de répartition respectives  $F_n$  et  $F$ . Pour que  $X_n \xrightarrow{\mathcal{L}} X$  il faut et il suffit que  $F_n(x) \xrightarrow{n \rightarrow \infty} F(x)$  pour tout  $x$  en lequel  $F$  est continue.

Notons que puisque la fonction  $F$  est continue à droite et croissante, l'ensemble des points où  $F$  est continue est l'ensemble  $D = \{x : F(x-) = F(x)\}$ , et son complémentaire est au plus dénombrable. Ainsi,  $D$  est dense dans  $\mathbb{R}$ .

**Preuve.** a) Supposons d'abord que  $X_n \xrightarrow{\mathcal{L}} X$ . Soit  $a$  avec  $F(a-) = F(a)$ . Pour tout  $p \in \mathbb{N}^*$  et tout  $b \in \mathbb{R}$ , il existe une fonction  $f_{p,b}$  continue bornée sur  $\mathbb{R}$  telle que

$$\mathbf{1}_{]-\infty, b]} \leq f_{p,b} \leq \mathbf{1}_{]-\infty, b+1/p]}. \quad (7.22)$$

Alors  $(\mathbb{E}(f_{p,b}(X_n)))_n$  converge vers  $\mathbb{E}(f_{p,b}(X))$  quand  $n$  tend vers l'infini. De (7.22), nous déduisons d'abord que  $F_n(a) = \mathbb{P}(X_n \leq a) \leq \mathbb{E}(f_{p,a}(X_n))$  et  $\mathbb{E}(f_{p,a}(X)) \leq F(a + \frac{1}{p})$ ; donc  $\limsup_n F_n(a) \leq F(a + \frac{1}{p})$  pour tout  $p$ , donc aussi  $\limsup_n F_n(a) \leq F(a)$ . Nous avons également que  $F_n(a) \geq \mathbb{E}(f_{p,a-1/p}(X_n))$  et  $\mathbb{E}(f_{p,a-1/p}(X)) \geq F(a - \frac{1}{p})$ ; donc  $\liminf_n F_n(a) \geq F(a - \frac{1}{p})$  pour tout  $p$ , donc aussi  $\liminf_n F_n(a) \geq F(a)$  puisque  $F(a-) = F(a)$ . Ces deux résultats impliquent que  $F_n(a) \xrightarrow{n \rightarrow \infty} F(a)$ .

b) Inversement, supposons que  $F_n(x) \xrightarrow{n \rightarrow \infty} F(x)$  pour tout  $x \in T$ , où  $T$  est une partie dense de  $\mathbb{R}$ . Soit  $f$  une fonction continue bornée sur  $\mathbb{R}$  et  $\varepsilon > 0$ . Soient  $a, b \in T$  avec  $F(a) \leq \varepsilon$  et  $F(b) \geq 1 - \varepsilon$ . Il existe  $n_0$  tel que

$$n \geq n_0 \quad \Rightarrow \quad \mathbb{P}(X_n \notin ]a, b]) = 1 - F_n(b) + F_n(a) \leq 3\varepsilon. \quad (7.23)$$

La fonction  $f$  est uniformément continue sur  $[a, b]$ , donc il existe un nombre fini de points  $a_0 = a < a_1 < \dots < a_k = b$  appartenant tous à  $T$  et tels que  $|f(x) - f(a_i)| \leq \varepsilon$  si  $a_{i-1} \leq x \leq a_i$ . Donc

$$g(x) = \sum_{i=1}^k f(a_i) \mathbf{1}_{]a_{i-1}, a_i]}(x)$$

vérifie  $|f - g| \leq \varepsilon$  sur  $]a, b]$ . Si  $M = \sup_x |f(x)|$ , il vient alors

$$|\mathbb{E}(f(X_n)) - \mathbb{E}(g(X_n))| \leq M \mathbb{P}(X_n \notin ]a, b]) + \varepsilon, \quad (7.24)$$

et de même pour  $X$ . Enfin  $\mathbb{E}(g(X_n)) = \sum_{i=1}^k f(a_i)(F_n(a_i) - F_n(a_{i-1}))$ , et de même pour  $X$ , par définition de  $g$ . Comme  $(F_n(a_i))_n$  converge vers  $F(a_i)$  pour tout  $i$ , nous en déduisons l'existence de  $n_1$  tel que

$$n \geq n_1 \quad \Rightarrow \quad |\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))| \leq \varepsilon. \quad (7.25)$$

D'après (7.23), (7.24) et (7.25), nous avons

$$n \geq \sup(n_0, n_1) \quad \Rightarrow \quad |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \leq 3\varepsilon + 5M\varepsilon.$$

$\varepsilon$  étant arbitraire, nous en déduisons que  $(\mathbb{E}(f(X_n)))_n$  converge vers  $\mathbb{E}(f(X))$ , d'où le résultat.  $\square$

**Corollaire 7.23** : Si la suite  $(X_n)_n$  de variables aléatoires réelles converge en loi vers  $X$ , et si la loi de  $X$  a une densité, alors pour tous  $a < b$ ,

$$\mathbb{P}(X_n \in ]a, b]) \xrightarrow{n \rightarrow \infty} \mathbb{P}(X \in ]a, b]).$$

**Preuve.** : La fonction de répartition de  $X$  est alors continue en tout point. (Mais pas nécessairement celles des variables aléatoires  $X_n$ .)  $\square$

**Proposition 7.24** Soient  $(X_n)_n$  et  $X$  des variables aléatoires. Pour que  $X_n \xrightarrow{\mathcal{L}} X$  il faut et il suffit que  $\mathbb{E}(f(X_n)) \xrightarrow{n \rightarrow \infty} \mathbb{E}(f(X))$  pour toute fonction  $f$  lipschitzienne bornée.

**Preuve.** Il suffit de montrer que dans la preuve de la proposition 7.22, nous pouvons remplacer les fonctions continues bornées  $f_{p,b}$  approchant  $\mathbf{1}_{]-\infty, b]}$  par des fonctions lipschitziennes bornées, ce qui est immédiat.  $\square$

**Proposition 7.25** (Théorème de Slutsky) Soient  $(X_n)_n$  et  $(Y_n)_n$  deux suites de variables aléatoires définies sur le même espace de probabilité. Supposons que  $(X_n)_n$  converge en loi vers  $X$  et que  $(X_n - Y_n)_n$  converge vers 0 en probabilité. Alors  $(Y_n)_n$  converge en loi vers  $X$ .

**Preuve.** Grâce à la proposition 7.24, il suffit de montrer que  $\lim_n \mathbb{E}(f(Y_n)) = \mathbb{E}(f(X))$ , pour toute fonction bornée lipschitzienne  $f$ . Nous avons alors  $|f(x)| \leq k$  et  $|f(x) - f(y)| \leq C|x - y|$ , pour des constantes  $k$  et  $C$ . Soit  $\varepsilon > 0$  donné. Nous écrivons

$$\begin{aligned} |\mathbb{E}(f(Y_n)) - \mathbb{E}(f(X_n))| &\leq \mathbb{E}(|f(Y_n) - f(X_n)|) \\ &\leq C\varepsilon + 2k \mathbb{P}(|X_n - Y_n| > \varepsilon). \end{aligned}$$

Le deuxième terme du membre de droite tend vers 0 quand  $n$  tend vers l'infini, car  $(X_n - Y_n)_n$  converge en probabilité vers 0. Comme  $\varepsilon$  est arbitrairement petit, nous en déduisons que  $\lim_{n \rightarrow \infty} |\mathbb{E}(f(Y_n)) - \mathbb{E}(f(X_n))| = 0$ , d'où le résultat.  $\square$

Une généralisation de ce lemme est développée dans l'exercice 7.6.

Le théorème suivant caractérise la convergence en loi à l'aide des fonctions caractéristiques. C'est un critère extrêmement utile dans la pratique.

**Théorème 7.26** (Théorème de Lévy) Soit  $(\mathbf{X}_n)_n$  une suite de vecteurs aléatoires à valeurs dans  $\mathbb{R}^d$ .

a) Si la suite  $(\mathbf{X}_n)_n$  converge en loi vers  $\mathbf{X}$ , alors  $\phi_{\mathbf{X}_n}$  converge simplement vers  $\phi_{\mathbf{X}}$ .

b) Si les  $\phi_{\mathbf{X}_n}$  convergent simplement vers une fonction (complexe)  $\phi$  sur  $\mathbb{R}^d$ , et si cette fonction est **continue** en 0, alors c'est la fonction caractéristique d'une v.a.  $\mathbf{X}$  et  $\mathbf{X}_n \xrightarrow{\mathcal{L}} \mathbf{X}$ .

**Preuve.** (Nous ne démontrons pas (b), qui est assez difficile).

Pour obtenir (a), il suffit de remarquer que  $\phi_{\mathbf{X}_n}(\mathbf{u}) = \mathbb{E}(g_{\mathbf{u}}(\mathbf{X}_n))$  et  $\phi_{\mathbf{X}}(\mathbf{u}) = \mathbb{E}(g_{\mathbf{u}}(\mathbf{X}))$ , où  $g_{\mathbf{u}}$  est la fonction continue bornée  $g_{\mathbf{u}}(\mathbf{x}) = e^{i\langle \mathbf{u}, \mathbf{x} \rangle}$  et d'appliquer la définition 7.18 (en séparant parties réelle et imaginaire).  $\square$

Ce théorème, plus les formules du premier paragraphe donnant les fonctions caractéristiques des lois usuelles, impliquent immédiatement que  $X_n \xrightarrow{\mathcal{L}} X$  dans les cas suivants :

- 1)  $X_n$  suit  $\mathcal{B}(m, p_n)$ ,  $X$  suit  $\mathcal{B}(m, p)$ , et  $p_n \rightarrow p$ .
- 2)  $X_n$  suit  $\mathcal{N}(m_n, \sigma_n^2)$ ,  $X$  suit  $\mathcal{N}(m, \sigma^2)$ , et  $m_n \rightarrow m$ ,  $\sigma_n^2 \rightarrow \sigma^2$ .
- 3)  $X_n$  et  $X$  suivent des lois de Poisson de paramètres  $\theta_n$  et  $\theta$ , et  $\theta_n \rightarrow \theta$ .
- 4)  $X_n$  et  $X$  suivent des lois exponentielles de paramètres  $\lambda_n$  et  $\lambda$ , et  $\lambda_n \rightarrow \lambda$ .

Il permet aussi de prouver le lemme de Slutsky très rapidement, y compris pour des vecteurs aléatoires.

**Théorème 7.27** (Théorème de Slutsky) *Soit  $(\mathbf{X}_n)_n$  et  $(\mathbf{Y}_n)_n$  deux suites de vecteurs aléatoires définies sur le même espace de probabilité à valeurs dans  $\mathbb{R}^d$ . Supposons que  $(\mathbf{X}_n)_n$  converge en loi vers  $\mathbf{X}$  et que  $(\mathbf{X}_n - \mathbf{Y}_n)_n$  converge vers 0 en probabilité. Alors  $(\mathbf{Y}_n)_n$  converge en loi vers  $\mathbf{X}$ .*

**Preuve.** Soit  $\mathbf{u} \in \mathbb{R}^d$ . On a :

$$|\phi_{\mathbf{Y}_n}(\mathbf{u}) - \phi_{\mathbf{X}}(\mathbf{u})| \leq |\phi_{\mathbf{Y}_n}(\mathbf{u}) - \phi_{\mathbf{X}_n}(\mathbf{u})| + |\phi_{\mathbf{X}_n}(\mathbf{u}) - \phi_{\mathbf{X}}(\mathbf{u})|.$$

D'une part, le théorème de Lévy partie a) montre que  $|\phi_{\mathbf{X}_n}(\mathbf{u}) - \phi_{\mathbf{X}}(\mathbf{u})|$  tend vers 0 quand  $n \rightarrow +\infty$ . D'autre part

$$|\phi_{\mathbf{Y}_n}(\mathbf{u}) - \phi_{\mathbf{X}_n}(\mathbf{u})| = |\mathbb{E}(e^{i\langle \mathbf{u}, \mathbf{Y}_n \rangle} - e^{i\langle \mathbf{u}, \mathbf{X}_n \rangle})| \leq \mathbb{E}(|e^{i\langle \mathbf{u}, \mathbf{Y}_n - \mathbf{X}_n \rangle} - 1|)$$

tend vers 0 quand  $n \rightarrow +\infty$  d'après la proposition 6.10.  $\square$

En conclusion de ce paragraphe, nous pouvons résumer dans la figure 6.1 les relations entre les différents modes de convergence.

## 7.4 Le théorème de la limite centrale

Ce théorème est aussi connu sous le nom de théorème central limite. Plus simplement, il apparaît souvent sous l'abréviation TCL.

## Relations entre modes de convergence

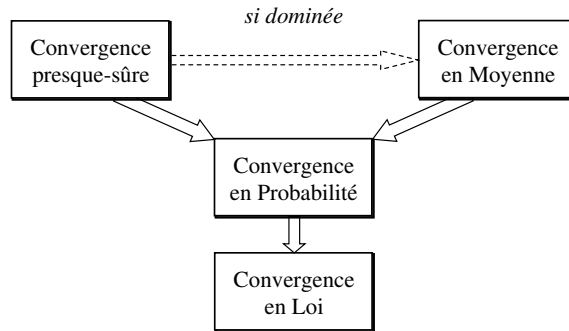


FIGURE 7.1 – Relations entre les différentes notions de convergence

La situation est la même que dans le chapitre précédent concernant la loi des grands nombres (Théorème 6.14). Nous considérons une suite de variables aléatoires  $(X_n)_n$  **indépendantes, de même loi, et de carré intégrable**. Notons  $m$  et  $\sigma^2$  la moyenne et la variance communes aux variables  $X_n$ , et

$$S_n = X_1 + \cdots + X_n \quad (7.26)$$

(ainsi  $M_n = \frac{S_n}{n}$ ). Nous avons vu que  $\frac{S_n}{n}$  converge vers  $m$ , presque-sûrement et en moyenne, et il est naturel de chercher la vitesse à laquelle cette convergence a lieu.

Pour évaluer cette vitesse, c'est-à-dire trouver un équivalent de  $\frac{S_n}{n} - m$ , nous sommes amenés à étudier la limite éventuelle de la suite  $n^\alpha (\frac{S_n}{n} - m)$  pour différentes valeurs de  $\alpha$  : si  $\alpha$  est “petit” cette suite va encore tendre vers 0, et elle va “exploser” si  $\alpha$  est “grand”. On peut espérer que pour une (et alors nécessairement une seule) valeur de  $\alpha$ , cette suite converge vers une limite qui n'est ni infinie ni nulle.

Il se trouve que la réponse à cette question a un aspect “négatif” : la suite  $n^\alpha (\frac{S_n}{n} - m)$  ne converge au sens presque-sûr, ou même en probabilité, pour aucune valeur de  $\alpha$ . (Voir l'exercice 7.8). Elle a aussi un aspect “positif” : cette suite converge, au sens de la convergence en loi, pour la même valeur  $\alpha = 1/2$  quelle que soit la loi des  $X_n$ , et toujours vers une loi normale.

Ce résultat, qui peut sembler miraculeux, a été énoncé par Laplace (1749-1827) et démontré beaucoup plus tard par Lyapounov (1901). Il montre le caractère univer-

sel de la loi normale (d'où son nom!). Il fait l'objet du théorème suivant, appelé **théorème central limite**, ou **de la limite centrale** :

**Théorème 7.28** *Si les  $X_n$  sont des v.a. réelles indépendantes et de même loi, de carré intégrable, de moyenne  $m$  et de variance  $\sigma^2$  avec  $\sigma^2 > 0$ , alors les variables*

$$\frac{S_n - nm}{\sigma\sqrt{n}}$$

*convergent en loi vers une v.a. de loi  $\mathcal{N}(0, 1)$ .*

En d'autres termes,  $\sqrt{n}(\frac{S_n}{n} - m)$  converge en loi vers une variable normale de loi  $\mathcal{N}(0, \sigma^2)$ .

**Preuve.** Nous donnons ici une preuve fondée sur le théorème de Lévy. Une autre preuve sera donnée dans l'exercice 7.7.

Soit  $\phi$  la fonction caractéristique de  $X_n - m$ , et posons  $Y_n = (S_n - nm)/(\sigma\sqrt{n})$ .

Comme les  $X_n$  sont indépendantes et d'après (7.3), la fonction caractéristique de  $Y_n$  est

$$\phi_n(u) = \phi\left(\frac{u}{\sigma\sqrt{n}}\right)^n. \quad (7.27)$$

Comme  $\mathbb{E}(X_n - m) = 0$  et  $\mathbb{E}((X_n - m)^2) = \sigma^2$ , (7.15) entraîne

$$\phi(u) = 1 - \frac{u^2\sigma^2}{2} + u^2 o(|u|) \quad \text{quand } u \rightarrow 0.$$

Comme  $\phi(0) = 1$  et que  $\phi$  est continue en 0, il est facile de voir que pour  $u$  fixé et  $n$  assez grand,

$$\left| \phi\left(\frac{u}{\sigma\sqrt{n}}\right) - 1 \right| \leq 1/2.$$

Il est possible de généraliser la notion de logarithme aux complexes  $z$  tels que  $|z - 1| \leq 1/2$ . (Voir un cours d'analyse complexe.) La fonction  $\ln z$  définie sur le disque  $\{z \in \mathbb{C} : |z - 1| \leq 1/2\}$  admet le même développement limité au voisinage de  $z = 1$  que le logarithme réel. Ainsi,

$$\phi_n(u) = \exp n \ln \left( 1 - \frac{u^2}{2n} + \frac{1}{n} \varepsilon_n(u) \right),$$

où  $\varepsilon_n(u)$  tend vers 0 quand  $n$  tend vers l'infini, et nous en déduisons immédiatement que  $\phi_n(u)$  converge vers  $\exp(-u^2/2)$ . Le résultat découle alors du théorème 7.26.  $\square$

**Remarque** Si les  $X_i$  de l'énoncé du théorème 7.28 suivent des lois  $\mathcal{N}(0,1)$ , alors  $(S_n - nm)/\sqrt{n}$  suit une loi  $\mathcal{N}(0,1)$  pour tout  $n$ . Le théorème de la limite centrale implique que la loi normale centrée réduite est la seule probabilité  $Q$  sur  $\mathbb{R}$  telle que si les  $X_i$  sont de loi  $Q$ , il en est de même de  $(S_n - nm)/\sqrt{n}$ .

**Exemple 7.29 Convergence des lois binomiales.** Supposons que  $S_n$  suive une loi binomiale  $\mathcal{B}(n, p)$ . Cela revient à dire que  $S_n$  a la même loi qu'une somme  $X_1 + \dots + X_n$  de  $n$  variables aléatoires  $X_i$  indépendantes de loi  $\mathcal{B}(1, p)$ , i.e.  $\mathbb{P}(X_i = 1) = p$  et  $\mathbb{P}(X_i = 0) = 1 - p$ . Nous savons alors que  $m = p$  et  $\sigma^2 = p(1 - p)$ .

Nous voulons calculer  $\mathbb{P}(S_n \leq x)$  pour  $x$  fixé et  $n$  grand.

Si  $p$  est très petit, de sorte que  $\theta = np$  ne soit pas trop grand (en pratique,  $\theta \leq 5$  convient), nous pouvons utiliser l'approximation par une loi de Poisson, obtenue au Chapitre 3. Si  $p$  est très proche de 1, de sorte que  $\theta = n(1 - p)$  soit comme ci-dessus, alors  $n - S_n$  suit à son tour une loi proche de la loi de Poisson de paramètre  $\theta$ .

Dans les autres cas, nous utilisons les théorèmes précédents :

$$\frac{S_n}{n} \xrightarrow{P} p, \quad (7.28)$$

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \quad (7.29)$$

Si nous désignons par  $\Phi$  la fonction de répartition de la loi  $\mathcal{N}(0,1)$ , il vient

$$\mathbb{P}(S_n \leq x) \simeq \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right). \quad (7.30)$$

Rappelons que la fonction de répartition  $\Phi$  ci-dessus est "tabulée", et une table de cette fonction est fournie à la section 4.6.

**EXERCICE 7.1** Lançons 1000 fois une pièce (que l'on suppose non truquée). Cherchons la probabilité d'obtenir plus de 545 fois le côté Face.

*Solution* : Bien sûr, le calcul exact utilisant les lois binomiales serait rébarbatif et extrêmement lourd. Nous utilisons le théorème de la limite centrale et écrivons

$$\mathbb{P}(S_{1000} > 545) = \mathbb{P}\left(\frac{S_{1000} - 1000/2}{\sqrt{1000}/2} > \frac{45}{\sqrt{1000}/2}\right) \simeq \int_{\frac{90}{\sqrt{1000}}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

En utilisant la table numérique, nous obtenons

$$\mathbb{P}(S_n > 545) \simeq 1 - \Phi(2,84) \simeq 0,0023.$$

Le théorème 7.28 admet une version multidimensionnelle, de preuve similaire. Considérons des vecteurs aléatoires  $\mathbf{X}_n$  à valeurs dans  $\mathbb{R}^d$ , indépendants et de même loi, dont les composantes sont de carré intégrable. Nous avons ainsi un vecteur moyenne  $\mathbf{m} = \mathbb{E}(\mathbf{X}_n)$ , et une matrice de covariance  $\mathbf{C} = (c_{ij})_{i,j=1}^d$  avec  $c_{ij}$  = la covariance des composantes  $i$  et  $j$  de  $\mathbf{X}_n$ . Nous pouvons alors énoncer le TCL multidimensionnel.

**Théorème 7.30** *Les vecteurs aléatoires  $\frac{S_n - n\mathbf{m}}{\sqrt{n}}$  convergent en loi vers un vecteur aléatoire gaussien centré (i.e. de moyenne nulle) et de matrice de covariance  $\mathbf{C}$ .*

**Remarque 7.31** *La vitesse de convergence est toujours en  $\frac{1}{\sqrt{n}}$ , indépendante de la dimension  $d$ . Cette remarque est fondamentale pour les applications numériques et justifie l'importance des méthodes de Monte-Carlo dans le cas des grandes dimensions.*

## 7.5 Exercices sur le chapitre 7

**EXERCICE 7.2** Soient  $X$  et  $Y$  deux v.a. indépendantes de carré intégrable, de même loi et centrées. Soit  $\phi$  leur fonction caractéristique commune. On suppose que la variable aléatoire  $\frac{X+Y}{\sqrt{2}}$  a même loi que  $X$  et  $Y$ . Montrer que ces variables sont nécessairement de loi normale.

**EXERCICE 7.3** Soit  $X$  une variable aléatoire de Cauchy.

1. Calculer la fonction caractéristique de  $X$ .
2. Montrer que  $\phi_{2X} = (\phi_X)^2$ .

**EXERCICE 7.4** Soient  $X$  et  $Y$  deux variables aléatoires indépendantes et  $a > 0$ .



1. La loi de  $X$  a pour densité  $\frac{a}{2}e^{-|x|^a}$ . Quelle est la fonction caractéristique de  $X$  ?
2. La loi de  $Y$  a pour densité  $\frac{1}{y^2}\mathbf{1}_{y \geq 1}$ .  
Justifier que  $\frac{X}{Y}$  est définie presque-sûrement.  
Calculer la fonction caractéristique de  $\frac{X}{Y}$ .

**EXERCICE 7.5** Soit  $\mathbf{X}$  un vecteur gaussien centré de  $\mathbb{R}^d$ . Soit  $F$  un sous-espace vectoriel de  $\mathbb{R}^d$ .

Le but de l'exercice est de montrer que  $\mathbb{P}(\mathbf{X} \in F) = 0$  ou  $1$ .

1. Soient  $\mathbf{X}_1$  et  $\mathbf{X}_2$  deux vecteurs indépendants de même loi que  $\mathbf{X}$ .  
Montrer que les ensembles  $A(\theta)$ , définis pour  $\theta \in [0, \frac{\pi}{2}[$  par

$$A(\theta) = \{\omega, \mathbf{X}_1(\omega) \cos \theta + \mathbf{X}_2(\omega) \sin \theta \in F ; \mathbf{X}_1(\omega) \sin \theta - \mathbf{X}_2(\omega) \cos \theta \notin F\}$$

sont disjoints pour des  $\theta$  différents.

2. Montrer que  $\mathbb{P}(A(\theta)) = \mathbb{P}(A(0))$ , pour tout  $\theta$ . On montrera que pour tout  $\theta$ ,  

$$\begin{pmatrix} \mathbf{X}_1 \cos \theta + \mathbf{X}_2 \sin \theta \\ \mathbf{X}_1 \sin \theta - \mathbf{X}_2 \cos \theta \end{pmatrix}$$
 a même loi que  $\begin{pmatrix} \mathbf{X}_1 \\ -\mathbf{X}_2 \end{pmatrix}$ .
3. Montrer que  $\mathbb{P}(A(0)) = 0$ . En déduire le résultat.

**EXERCICE 7.6** *Généralisation du théorème de Slutsky.* Montrer que si deux suites de variables aléatoires  $(X_n)_n$  et  $(Y_n)_n$  sont telles que  $X_n$  converge en loi vers  $X$  et  $Y_n$  converge en probabilité vers une constante  $y$ , alors le couple  $(X_n, Y_n)$  converge en loi vers  $(X, y)$ .

En déduire que  $X_n + Y_n$  et  $X_n Y_n$  convergent en loi, respectivement vers  $X + y$  et  $Xy$ .

**EXERCICE 7.7** Une autre preuve du théorème de la limite centrale (due à Lindeberg).

Soit une suite  $(X_n)_n$  de variables aléatoires indépendantes de carré intégrable, centrées et de variance 1. Nous voulons prouver que la suite  $T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$  converge en loi vers une variable aléatoire  $T$  de loi  $\mathcal{N}(0, 1)$ .

Préliminaire : Montrer que  $T_n$  converge en loi vers  $T$  dès que  $\mathbb{E}(f(T_n))$  converge vers  $\mathbb{E}(f(T))$ , pour  $f$  bornée de classe  $C^2$  avec dérivées première et seconde bornées et uniformément continues.

Soit  $(Y_n)_n$  une suite de variables aléatoires indépendantes de loi  $\mathcal{N}(0, 1)$ , indépendantes de la suite  $(X_n)_n$ . Nous savons qu'alors  $T'_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$  est de loi  $\mathcal{N}(0, 1)$ . Le but est de montrer que  $|\mathbb{E}(f(T_n)) - \mathbb{E}(f(T'_n))| \rightarrow 0$ , quand  $n$  tend vers l'infini, pour toute fonction  $f$  bornée de classe  $C^2$  avec dérivées première et seconde bornées et uniformément continues.

1. Notons  $X_{i,n} = \frac{X_i}{\sqrt{n}}$ ,  $Y_{i,n} = \frac{Y_i}{\sqrt{n}}$ , et  $W_{i,n} = \sum_{j=1}^{i-1} Y_{j,n} + \sum_{j=i+1}^n X_{j,n}$ .  
Montrer que

$$f(T_n) - f(T'_n) = \sum_{i=1}^n (f(W_{i,n} + X_{i,n}) - f(W_{i,n} + Y_{i,n})). \quad (7.31)$$

2. Montrer que

$$f(W_{i,n} + X_{i,n}) = f(W_{i,n}) + f'(W_{i,n})X_{i,n} + \frac{1}{2}f''(W_{i,n})(X_{i,n})^2 + R_{X,i,n},$$

et que pour tout  $\varepsilon > 0$ , il existe  $\delta > 0$ , tel que

$$|R_{X,i,n}| \leq (X_{i,n})^2 (\varepsilon \mathbf{1}_{|X_{i,n}| \leq \delta} + \|f''\|_{\infty} \mathbf{1}_{|X_{i,n}| > \delta}).$$

3. En déduire que

$$|\mathbb{E}(f(T_n)) - \mathbb{E}(f(T'_n))| \leq 2\varepsilon + \|f''\|_{\infty} \mathbb{E}(X_1^2 \mathbf{1}_{|X_1| > \delta\sqrt{n}} + Y_1^2 \mathbf{1}_{|Y_1| > \delta\sqrt{n}}).$$

4. Conclure.

**EXERCICE 7.8** Soient  $(X_j)_j$  des variables aléatoires indépendantes et de même loi, de carré intégrable, avec  $E(X_1) = 0$ , et  $\mathbb{E}(X_1^2) = \sigma^2 > 0$ . Soit  $S_n = \sum_{i=1}^n X_i$ .

1. Montrer que  $\frac{S_n}{\sqrt{n}}$  ne converge pas en probabilité.
2. Montrer que  $(n^\alpha |\frac{S_n}{n} - m|)_n$  converge en probabilité vers 0, (respectivement vers  $+\infty$ ), quand  $\alpha < 1/2$ , (respectivement  $\alpha > 1/2$ ).

## Chapitre 8

# Statistique : Estimation

*Tout ce qui existe dans l'Univers est le fruit du hasard et de la nécessité.*

Démocrite

La statistique est la science des données. Un statisticien travaille sur des données (résultats d'un sondage, données météorologiques,...) et essaie de traiter des problèmes de différents types :

- effectuer une prévision, par exemple sur le résultat d'une élection qui aura lieu prochainement à partir d'un sondage,
- quantifier la certitude liée à une prévision, par exemple par une fourchette dans le cas d'un sondage,
- répondre à une question comme "le candidat C sera-t-il élu", "le réchauffement climatique est-il réel?", "ce médicament est-il efficace?", pour aider à la prise de décision, comme l'adoption de protocoles réduisant la production de CO<sub>2</sub> ou la mise sur le marché d'un nouveau médicament.

La première question peut être abordée dans le cadre de l'estimation statistique, la seconde dans le cadre de la théorie des intervalles (ou des régions) de confiance, et la troisième dans le cadre de la théorie des tests de décision. On va dans les chapitres qui viennent donner quelques éléments de réponse pour chacune de ces questions. Dans ce chapitre, on aborde la question de l'estimation ponctuelle.

## 8.1 Introduction à l'estimation

Le but de l'estimation statistique est le suivant. On observe des réalisations d'un phénomène aléatoire (sexe d'un nouveau-né, température ou pluviométrie journalière, ...) qu'on appelle observations. Ces observations sont des copies indépendantes et identiquement distribuées d'une loi inconnue qu'on cherche à retrouver. On se limitera au cas paramétrique dans ce cours, c'est-à-dire qu'on supposera que cette loi appartient à une famille connue de lois qui dépend d'un ou de plusieurs paramètres inconnus. On souhaite déterminer, ou plus exactement estimer, la valeur de ce(s) paramètre(s), à partir des observations.

**Exemple 8.1** *Le sexe, fille ( $F$ ) ou garçon ( $G$ ), d'un nouveau-né est modélisé par une loi  $P_\theta$  sur  $\{F, G\}$  de paramètre inconnu  $\theta \in [0, 1]$ , avec  $P_\theta(\{G\}) = \theta$ , resp.  $P_\theta(\{F\}) = 1 - \theta$ , la probabilité qu'une naissance donne un garçon, resp. une fille. On souhaite connaître la proportion de garçons à la naissance. Ceci revient à chercher à estimer  $\theta$ . Pour estimer  $\theta$ , on observe les sexes des  $n$  nouveaux-nés dans une maternité.*

**Exemple 8.2** *La durée de vie d'une ampoule électrique produite par une usine est modélisée par une loi exponentielle de paramètre  $\lambda \in ]0, +\infty[$ . Pour estimer  $\lambda$ , on laisse allumées  $n$  ampoules jusqu'à ce qu'elles grillent, on observe donc  $n$  durées de vie, à partir desquelles on essaye d'estimer  $\lambda$ .*

D'une manière générale, un modèle statistique est défini de manière similaire à un espace de probabilité, à la différence essentielle près que le dernier élément du triplet n'est pas une probabilité, mais une famille paramétrique de probabilités.

**Définition 8.3** Un modèle statistique est un triplet  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  où  $\mathcal{X}$  est l'espace fondamental (l'ensemble des valeurs possibles des observations),  $\mathcal{A}$  est une tribu sur  $\mathcal{X}$ , et  $\mathcal{P}$  est une famille de probabilités sur  $(\mathcal{X}, \mathcal{A})$ . Un modèle statistique est dit paramétrique s'il existe un entier  $p$  et un ensemble  $\Theta \subset \mathbb{R}^p$  tels que la famille de probabilités  $\mathcal{P}$  s'écrit sous la forme :

$$\mathcal{P} = \{P_\theta, \theta \in \Theta\},$$

où, pour tout  $\theta \in \Theta$ ,  $P_\theta$  est une probabilité sur  $(\mathcal{X}, \mathcal{A})$ .

On notera génériquement  $\theta$  le paramètre d'une famille paramétrique de probabilités.

**Exemple 8.4** *Dans l'exemple 8.1, le modèle statistique associé est le triplet  $(\{F, G\}, \mathcal{P}(\{F, G\}), \mathcal{P})$  où :*

$$\mathcal{P} = \{P_\theta, \theta \in [0, 1]\},$$

et  $P_\theta(\{F\}) = 1 - \theta$  et  $P_\theta(\{G\}) = \theta$ . Ici  $\Theta = [0, 1]$ .

**Exemple 8.5** Dans l'exemple 8.2, le modèle statistique est le triplet  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P})$  où  $\mathcal{P}$  est l'ensemble des lois exponentielles :

$$\mathcal{P} = \{P_\lambda, \lambda \in ]0, +\infty[ \},$$

et  $P_\lambda$  est la probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  de densité  $p_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{]0, +\infty[}(x)$ , avec la notation  $\mathbf{1}_A(x) = 1$  si  $x \in A$  et 0 sinon.

De manière générale, les observations forment un  $n$ -uplet de répliques indépendantes et identiquement distribuées.

**Définition 8.6** Un  $n$ -échantillon est un vecteur aléatoire  $\mathbf{X} = (X_i)_{i=1, \dots, n}$  de  $n$  v.a. indépendantes et identiquement distribuées (i.i.d.) de même loi qui appartient au modèle statistique.

**Notation.** Le vecteur aléatoire  $\mathbf{X} = (X_i)_{i=1, \dots, n}$  est défini sur un espace  $\Omega$  qu'on ne précisera pas (on peut cependant prendre pour  $\Omega$  l'ensemble  $\mathcal{X}^n$  de tous les échantillons de taille  $n$  possibles, et de ce fait, les observations peuvent être vues comme une réalisation particulière d'un  $n$ -échantillon). On notera respectivement  $\mathbb{P}_\theta(\mathbf{X} \in A)$  et  $\mathbb{E}_\theta(f(\mathbf{X}))$  la probabilité de l'événement  $\{\mathbf{X} \in A\}$  et l'espérance de  $f(\mathbf{X})$  lorsque les variables  $X_i$  sont indépendantes et identiquement distribuées de loi  $P_\theta$ .

On observe un  $n$ -échantillon  $\mathbf{X}$ . On souhaite construire à partir de cet échantillon une quantité  $\mathbf{f}(\mathbf{X})$  qui s'approche du paramètre inconnu  $\theta$ . Une telle quantité est appelée estimateur de  $\theta$  et est souvent notée  $\hat{\theta}_n$ . Un estimateur ne dépend que de l'échantillon  $\mathbf{X}$  et ne dépend pas de  $\theta$ , et est donc entièrement caractérisé par la fonction déterministe  $\mathbf{f} : \mathcal{X}^n \rightarrow \Theta$ .

**Exemple 8.7** Dans l'exemple 8.1, on souhaite connaître la proportion de garçons à la naissance. Le modèle statistique sous-jacent est  $(\{F, G\}, \mathcal{P}(\{F, G\}), \mathcal{P})$  avec  $\mathcal{P} = \{P_\theta, \theta \in [0, 1]\}$  l'ensemble des lois  $P_\theta$  telles que  $P_\theta(\{G\}) = \theta$ . Pour estimer  $\theta$ , on observe toutes les naissances dans une maternité, et on note les résultats  $X_1, \dots, X_n$ , avec  $X_i = G$  si la  $i^{\text{ème}}$  naissance donne un garçon, et  $X_i = F$  si c'est une fille. C'est un  $n$ -échantillon de la loi  $P_\theta$ . Un estimateur intuitif de  $\theta$  est la proportion empirique de garçons, c'est-à-dire le nombre de garçons observés divisé par le nombre total de naissances :

$$\hat{\theta}_n = \frac{\text{Card}(i = 1, \dots, n, X_i = G)}{n} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_G(X_i).$$

Cet estimateur est de la forme  $\hat{\theta}_n = f(X_1, \dots, X_n)$ , avec  $f(x_1, \dots, x_n) = (1/n) \sum_{i=1}^n \mathbf{1}_G(x_i)$ .

**Exemple 8.8** Dans l'exemple 8.2, les temps de vie des  $n$  ampoules du lot sont des variables aléatoires  $X_1, \dots, X_n$  indépendantes et identiquement distribuées, de loi exponentielle de paramètre  $\lambda$  inconnu. Notons  $\bar{X}_n$  la moyenne empirique :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

dont on sait qu'elle est proche de l'espérance  $\mathbb{E}_\lambda(X_1) = 1/\lambda$  d'après la loi forte des grands nombres. On peut donc proposer comme estimateur de  $\lambda$  :  $\hat{\lambda}_n = 1/\bar{X}_n$ , c'est-à-dire  $\hat{\lambda}_n = f(\mathbf{X})$  avec  $f(\mathbf{x}) = n/\sum_{i=1}^n x_i$ . Cependant, on pourrait s'y prendre autrement. Par exemple, en notant

$$\check{\lambda}_n = \left( \frac{1}{2n} \sum_{i=1}^n X_i^2 \right)^{-1/2},$$

on sait aussi que  $\check{\lambda}_n$  est proche de  $(\frac{1}{2}\mathbb{E}_\lambda(X_1^2))^{-1/2} = \lambda$  d'après la loi forte des grands nombres. Ceci montre que  $\hat{\lambda}_n$  et  $\check{\lambda}_n$  sont tous les deux susceptibles de nous donner une estimation raisonnable de  $\lambda$  (lorsque  $n$  est grand). La question est de savoir lequel est le meilleur (une fois qu'on aura clarifié ce qu'on entend par bon).

Plus généralement, on peut s'intéresser à l'estimation de  $\mathbf{g}(\boldsymbol{\theta})$  où  $\mathbf{g} : \Theta \rightarrow \mathbb{R}^q$ . En pratique, l'estimation de  $\boldsymbol{\theta}$  correspond au choix  $q = p$  et  $\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$ . Lorsque  $p \geq 2$ , l'estimation des  $q$  premières coordonnées de  $\boldsymbol{\theta}$  avec  $q < p$  correspond au choix  $\mathbf{g}(\theta_1, \dots, \theta_p) = (\theta_1, \dots, \theta_q)$ .

L'exemple 8.1 est particulièrement simple, mais l'exemple 8.2 sur la durée de vie d'une ampoule montre qu'on peut avoir plusieurs estimateurs raisonnables possibles. De plus, il s'agit de savoir si l'estimateur proposé est "bon". Pour répondre à cette question, il faudra d'abord éclaircir ce que nous entendons par "bon estimateur".

## 8.2 Qualités d'un estimateur

Souvent, le nombre d'observations  $n$  est suffisamment grand pour qu'on puisse exploiter les propriétés limites, ou asymptotiques, des estimateurs (asymptotique dans le sens  $n \rightarrow +\infty$ ). C'est ici qu'interviennent les théorèmes limites introduits dans les chapitres précédents. Dans ce paragraphe, on examine les qualités d'une suite d'estimateurs  $(\hat{\boldsymbol{\theta}}_n)_{n \in \mathbb{N}^*}$  définis par une suite de fonctions  $\mathbf{f}_n : \mathcal{X}^n \rightarrow \Theta$ .

**Exemple 8.9** Dans l'exemple 8.1 on peut examiner la suite  $\hat{\theta}_n = f_n(X_1, \dots, X_n)$  définie par la suite de fonctions :

$$f_n : \left\{ \begin{array}{l} \{F, G\}^n \rightarrow \mathbb{R}, \\ (x_1, \dots, x_n) \mapsto \frac{1}{n} \sum_{i=1}^n \mathbf{1}_G(x_i). \end{array} \right.$$

La propriété de convergence est essentielle.

**Définition 8.10** [Convergence] On dit que  $\hat{\theta}_n$  est convergent si, pour tout  $\theta \in \Theta$  :

$$\mathbb{P}_\theta \left( \hat{\theta}_n \xrightarrow{n \rightarrow +\infty} \theta \right) = 1.$$

La propriété de convergence dit que, si la taille de l'échantillon est suffisamment grande, alors la valeur donnée par l'estimateur devient proche de  $\theta$  lorsque l'échantillon est tiré avec la loi  $\mathbb{P}_\theta$ . Il est important de réclamer que cette propriété soit vérifiée pour tous les  $\theta \in \Theta$ , car on ne connaît pas a priori la vraie valeur de  $\theta$  (celle avec laquelle les  $X_i$  ont été tirés).

**Exemple 8.11** Dans l'exemple 8.1, sous  $\mathbb{P}_\theta$ , les variables aléatoires  $\mathbf{1}_G(X_i)$  sont des variables i.i.d. de loi de Bernoulli de paramètre  $\theta$ . L'application de la loi forte des grands nombres prouve la convergence de l'estimateur proposé :

$$\mathbb{P}_\theta \left( \lim_{n \rightarrow +\infty} \hat{\theta}_n = \theta \right) = \mathbb{P}_\theta \left( \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=1}^n \mathbf{1}_G(X_j) = \theta \right) = 1.$$

**Définition 8.12** [Biais] Un estimateur est dit non-biaisé si  $\mathbb{E}_\theta(\hat{\theta}_n) = \theta$  pour tout  $\theta \in \Theta$  et pour tout  $n \in \mathbb{N}^*$ .

Un estimateur est dit asymptotiquement non-biaisé si  $\lim_{n \rightarrow +\infty} \mathbb{E}_\theta(\hat{\theta}_n) = \theta$  pour tout  $\theta \in \Theta$ .

Un estimateur non-biaisé est évidemment asymptotiquement non-biaisé.

**Exemple 8.13** Dans l'exemple 8.1, on utilise successivement la linéarité de l'espérance et le fait que l'espérance d'une v.a. de Bernoulli est égale à son paramètre pour écrire :

$$\mathbb{E}_\theta(\hat{\theta}_n) = \frac{1}{n} \sum_{j=1}^n \mathbb{E}_\theta(\mathbf{1}_G(X_j)) = \mathbb{E}_\theta(\mathbf{1}_G(X_1)) = \theta.$$

L'estimateur  $\hat{\theta}_n$  est donc non-biaisé.

La propriété de non-biais est souvent recherchée, car elle stipule que l'estimateur est bon en moyenne, mais elle est insuffisante. Pour qualifier un estimateur, on utilise le risque quadratique moyen défini comme suit.

**Définition 8.14** [Risque quadratique moyen] Si  $g : \Theta \rightarrow \mathbb{R}$ , alors le risque quadratique moyen d'un estimateur  $\hat{g}_n$  de  $g(\boldsymbol{\theta})$  est la moyenne quadratique des écarts à la valeur à estimer :

$$\text{RQM}_{\boldsymbol{\theta}}(\hat{g}_n) = \mathbb{E}_{\boldsymbol{\theta}} [(\hat{g}_n - g(\boldsymbol{\theta}))^2].$$

On peut donner une expression différente du risque quadratique moyen en développant le carré :

$$\begin{aligned} \text{RQM}_{\boldsymbol{\theta}}(\hat{g}_n) &= \mathbb{E}_{\boldsymbol{\theta}} [(\hat{g}_n - \mathbb{E}_{\boldsymbol{\theta}}(\hat{g}_n) + \mathbb{E}_{\boldsymbol{\theta}}(\hat{g}_n) - g(\boldsymbol{\theta}))^2] \\ &= \mathbb{E}_{\boldsymbol{\theta}} [(\hat{g}_n - \mathbb{E}_{\boldsymbol{\theta}}(\hat{g}_n))^2] + 2\mathbb{E}_{\boldsymbol{\theta}} [\hat{g}_n - \mathbb{E}_{\boldsymbol{\theta}}(\hat{g}_n)] (\mathbb{E}_{\boldsymbol{\theta}}(\hat{g}_n) - g(\boldsymbol{\theta})) \\ &\quad + (\mathbb{E}_{\boldsymbol{\theta}}(\hat{g}_n) - g(\boldsymbol{\theta}))^2. \end{aligned}$$

Le second terme du membre de droite est nul car  $\mathbb{E}_{\boldsymbol{\theta}}[\hat{g}_n - \mathbb{E}_{\boldsymbol{\theta}}(\hat{g}_n)] = \mathbb{E}_{\boldsymbol{\theta}}(\hat{g}_n) - \mathbb{E}_{\boldsymbol{\theta}}(\hat{g}_n) = 0$ . Le risque quadratique moyen se décompose donc en une partie “variance” et une partie “biais” :

$$\text{RQM}_{\boldsymbol{\theta}}(\hat{g}_n) = \text{Var}_{\boldsymbol{\theta}}(\hat{g}_n) + (\mathbb{E}_{\boldsymbol{\theta}}(\hat{g}_n) - g(\boldsymbol{\theta}))^2. \quad (8.1)$$

Si un estimateur est sans biais, son risque quadratique moyen est égal à sa variance. Entre deux estimateurs, on choisira celui au risque le plus faible. Ainsi, de deux estimateurs non-biaisés, on choisira celui de variance plus faible.

**Exemple 8.15** Dans l'exemple 8.1, on utilise le fait que la variance de la somme de v.a. indépendantes est égale à la somme des variances pour obtenir :

$$\text{RQM}_{\theta}(\hat{\theta}_n) = \text{Var}_{\theta}(\hat{\theta}_n) = \frac{1}{n} \text{Var}_{\theta}(\mathbf{1}_G(X_1)) = \frac{\theta(1-\theta)}{n}.$$

On remarque que le risque décroît avec  $n$  comme  $1/n$ , résultat assez général qu'on va retrouver dans la suite, mais pas toujours vrai.

Enfin, lorsque l'estimateur est convergent, on peut se poser la question de la forme de ses fluctuations. On peut penser, d'après le théorème de la limite centrale, que les fluctuations d'un estimateur convergent sont d'ordre  $1/\sqrt{n}$  et à statistique gaussienne quand  $n \rightarrow \infty$ . C'est souvent le cas, mais pas toujours comme on va le voir dans la suite.

**Définition 8.16** [Normalité asymptotique] Soit  $g : \Theta \rightarrow \mathbb{R}$ . Un estimateur  $\hat{g}_n$  de  $g(\boldsymbol{\theta})$  est dit asymptotiquement normal s'il existe deux fonctions déterministes  $m_n(\boldsymbol{\theta})$  et  $\sigma_n(\boldsymbol{\theta})$  telles que, sous  $\mathbb{P}_{\boldsymbol{\theta}}$ , la suite de variables aléatoires  $(\hat{g}_n - m_n(\boldsymbol{\theta}))/\sigma_n(\boldsymbol{\theta})$  converge en loi quand  $n \rightarrow +\infty$  vers une loi gaussienne centrée réduite.



Très souvent, mais pas toujours,  $m_n(\boldsymbol{\theta}) = g(\boldsymbol{\theta})$  et  $\sigma_n(\boldsymbol{\theta}) = \sigma(\boldsymbol{\theta})/\sqrt{n}$  pour une fonction  $\sigma(\boldsymbol{\theta})$  qu'on appelle écart-type asymptotique (et  $\sigma^2(\boldsymbol{\theta})$  est appelée variance asymptotique).

**Exemple 8.17** Dans l'exemple 8.1, on utilise le théorème de la limite centrale pour obtenir que, sous  $\mathbb{P}_\theta$  :

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \theta(1 - \theta)),$$

en loi. La variance asymptotique est ici  $\sigma^2(\theta) = \theta(1 - \theta)$ .

### 8.3 Estimateurs empiriques

On considère ici un modèle statistique de la forme  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P})$  associé à des observations réelles. Il arrive très souvent que le modèle statistique considéré soit (ou puisse être) paramétré par la moyenne et/ou la variance de la loi inconnue. Par exemple, le modèle gaussien qu'on étudiera en détail rentre dans ce cadre :

$$\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 \in ]0, +\infty[ \}. \quad (8.2)$$

On va introduire et discuter ici les estimateurs dits “empiriques” de ces quantités.

**Proposition 8.18 (Estimateur de la moyenne)** Soit  $(X_i)_{i=1, \dots, n}$  un  $n$ -échantillon d'un modèle statistique  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P})$  dont les lois sont intégrables et dont le paramètre  $\boldsymbol{\theta}$  contient la moyenne, i.e., il existe une fonction  $g : \Theta \rightarrow \mathbb{R}$  telle que  $g(\boldsymbol{\theta})$  est la moyenne  $\mu_\theta$  de la loi  $P_\theta$ . La moyenne empirique  $\bar{X}_n$  définie par :

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j, \quad (8.3)$$

est un estimateur de la moyenne qui satisfait les propriétés suivantes :

- (i)  $\bar{X}_n$  est non-biaisé.
- (ii)  $\bar{X}_n$  est convergent.
- (iii) Si de plus les lois  $P_\theta$  sont de carré intégrable, alors  $\text{RQM}_\theta(\bar{X}_n) = \sigma_\theta^2/n$ , avec  $\sigma_\theta^2$  la variance de la loi  $P_\theta$  et  $\bar{X}_n$  est asymptotiquement normal, avec

$$\sqrt{n}(\bar{X}_n - \mu_\theta) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma_\theta^2),$$

en loi.

**Preuve.** Il n'y a rien de nouveau dans cette proposition qui découle de la linéarité de l'espérance, de la loi forte des grands nombres, et du théorème de la limite centrale.

□

D'après la formule de Huygens, la variance d'une v.a. réelle  $X$  de carré intégrable est donnée par :

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

En copiant ce qu'on a fait pour l'estimateur de la moyenne, on peut proposer comme estimateur du premier moment la moyenne empirique  $\bar{X}_n$ , et comme estimateur du second moment la somme  $\frac{1}{n} \sum_{j=1}^n X_j^2$ . On obtient alors un estimateur de la variance, noté  $\bar{V}_n$  :

$$\bar{V}_n = \frac{1}{n} \sum_{j=1}^n X_j^2 - \bar{X}_n^2 = \frac{1}{n} \sum_{j=1}^n X_j^2 - \left( \frac{1}{n} \sum_{j=1}^n X_j \right)^2, \quad (8.4)$$

qu'on appelle variance empirique. Cette somme peut se réécrire différemment.

**Proposition 8.19 (Estimateur de la variance)** *Soit  $(X_i)_{i=1,\dots,n}$  un  $n$ -échantillon d'un modèle statistique  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P})$  dont les lois sont de carré intégrable et dont le paramètre  $\theta$  contient la variance, i.e., il existe une fonction  $g : \Theta \rightarrow \mathbb{R}$  telle que  $g(\theta)$  est la variance  $\sigma_\theta^2$  de la loi  $P_\theta$ . La variance empirique de l'échantillon est définie par :*

$$\bar{V}_n = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2. \quad (8.5)$$

Cet estimateur de la variance vérifie les propriétés suivantes.

(i) L'estimateur est biaisé et asymptotiquement non-biaisé :

$$\mathbb{E}_\theta(\bar{V}_n) = \frac{n-1}{n} \sigma_\theta^2.$$

(ii) L'estimateur est convergent.

(iii) Si les lois  $P_\theta$  sont de quatrième moment fini, alors le risque quadratique moyen de  $\bar{V}_n$  est d'ordre  $1/n$  :

$$\text{RQM}_\theta(\bar{V}_n) = \frac{\mu_\theta^{(4)} - \sigma_\theta^4}{n} + O\left(\frac{1}{n^2}\right),$$

où  $\mu_\theta^{(4)} = \mathbb{E}_\theta(X_1^4) - 4\mathbb{E}_\theta(X_1)\mathbb{E}_\theta(X_1^3) + 6\mathbb{E}_\theta(X_1^2)\mathbb{E}_\theta(X_1)^2 - 3\mathbb{E}_\theta(X_1)^4$ , et  $\bar{V}_n$  est asymptotiquement normal :

$$\sqrt{n}(\bar{V}_n - \sigma_\theta^2) \xrightarrow[n \rightarrow +\infty]{} \mathcal{N}(0, \mu_\theta^{(4)} - \sigma_\theta^4),$$

en loi.

$\mu^{(4)}$  est le moment centré d'ordre 4, dont la définition générale est, pour  $p \in \mathbb{N}^*$  et une v.a. réelle  $X$  de moment d'ordre  $p$  fini :

$$\mu^{(p)} = \mathbb{E}[(X - \mathbb{E}(X))^p].$$

En développant le polynôme  $(X - \mathbb{E}(X))^p$ , on peut aussi écrire que :

$$\mu^{(p)} = \sum_{j=0}^p (-1)^{p-j} \binom{p}{j} \mathbb{E}(X^j) \mathbb{E}(X)^{p-j}.$$

On peut noter que le second moment centré  $\mu^{(2)}$  est égal à la variance  $\sigma^2$  de  $X$ . Donc  $\mu^{(4)} - \sigma^4$  est la variance de la v.a. réelle  $X^{(2)} = (X - \mathbb{E}(X))^2$  :

$$\mu^{(4)} - \sigma^4 = \text{Var}(X^{(2)}) = \mathbb{E} \left[ \left( X^{(2)} - \mathbb{E}(X^{(2)}) \right)^2 \right].$$

**Preuve.** Pour montrer que (8.4) et (8.5) sont équivalents, on part de la seconde expression, et on développe les carrés :

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2 &= \frac{1}{n} \sum_{j=1}^n X_j^2 - \frac{2}{n} \sum_{j=1}^n X_j \bar{X}_n + \bar{X}_n^2 = \frac{1}{n} \sum_{j=1}^n X_j^2 - 2\bar{X}_n^2 + \bar{X}_n^2 \\ &= \frac{1}{n} \sum_{j=1}^n X_j^2 - \bar{X}_n^2. \end{aligned}$$

On va maintenant calculer les deux premiers moments de cet estimateur. On commence par introduire les v.a.  $\tilde{X}_i = X_i - \mu_{\theta}$  (avec  $\mu_{\theta} = \mathbb{E}_{\theta}(X_1)$ ), qui sont indépendantes et identiquement distribuées, de moyenne 0 et de variance  $\sigma_{\theta}^2$ . On a alors :

$$\bar{V}_n = \frac{1}{n} \sum_{j=1}^n (\tilde{X}_j + \mu_{\theta})^2 - \left( \frac{1}{n} \sum_{j=1}^n \tilde{X}_j + \mu_{\theta} \right)^2 = \frac{1}{n} \sum_{j=1}^n \tilde{X}_j^2 - \left( \frac{1}{n} \sum_{j=1}^n \tilde{X}_j \right)^2. \quad (8.6)$$

Pour  $p$  entier, le  $p^{\text{eme}}$  moment de  $\bar{V}_n$  se développe en somme d'espérances de la forme  $\mathbb{E}_{\theta}(\tilde{X}_{i_1} \cdots \tilde{X}_{i_{2p}})$ . Il faut prendre soin de distinguer les indices  $i_j$  qui sont égaux, parce que, si une variable  $\tilde{X}_{i_k}$  apparaît une seule fois dans le produit  $\tilde{X}_{i_1} \cdots \tilde{X}_{i_{2p}}$ , alors, par indépendance des variables  $\tilde{X}_i$ , l'espérance  $\mathbb{E}_{\theta}(\tilde{X}_{i_1} \cdots \tilde{X}_{i_{2p}}) = \mathbb{E}_{\theta}(\tilde{X}_{i_k}) \mathbb{E}_{\theta}(\prod_{j \neq k} \tilde{X}_{i_j})$  est nulle car  $\mathbb{E}_{\theta}(\tilde{X}_{i_k}) = 0$ . Ainsi, pour les deux premiers moments, on obtient :

$$\begin{aligned} \mathbb{E}_{\theta}(\bar{V}_n) &= \frac{1}{n} \sum_{j=1}^n \mathbb{E}_{\theta}(\tilde{X}_j^2) - \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}_{\theta}(\tilde{X}_i \tilde{X}_j) = \frac{1}{n} \sum_{j=1}^n \mathbb{E}_{\theta}(\tilde{X}_1^2) - \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}_{\theta}(\tilde{X}_1^2) \\ &= \mathbb{E}_{\theta}(\tilde{X}_1^2) - \frac{1}{n} \mathbb{E}_{\theta}(\tilde{X}_1^2) = \frac{n-1}{n} \sigma_{\theta}^2, \end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}(\bar{V}_n^2) &= \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_i^2 \tilde{X}_j^2) - \frac{2}{n^3} \sum_{i,j,k=1}^n \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_i \tilde{X}_j \tilde{X}_k^2) + \frac{1}{n^4} \sum_{i,j,k,l=1}^n \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_i \tilde{X}_j \tilde{X}_k \tilde{X}_l) \\
&= \frac{1}{n} \left( \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^4) + (n-1) \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^2)^2 \right) - \frac{2}{n^2} \left( \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^4) + (n-1) \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^2)^2 \right) \\
&\quad + \frac{1}{n^3} \left( \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^4) + 3(n-1) \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^2)^2 \right) \\
&= \frac{(n-1)^2}{n^3} \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^4) + \frac{(n-1)(n^2-2n+3)}{n^3} \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^2)^2 \\
&= \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^2)^2 + \frac{1}{n} \left( \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^4) - 3 \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^2)^2 \right) + O\left(\frac{1}{n^2}\right).
\end{aligned}$$

On peut alors calculer le risque quadratique moyen :

$$\text{RQM}_{\boldsymbol{\theta}}(\bar{V}_n) = \mathbb{E}_{\boldsymbol{\theta}}(\bar{V}_n^2) - 2\sigma_{\boldsymbol{\theta}}^2 \mathbb{E}_{\boldsymbol{\theta}}(\bar{V}_n) + \sigma_{\boldsymbol{\theta}}^4 = \frac{\mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^4) - \mathbb{E}_{\boldsymbol{\theta}}(\tilde{X}_1^2)^2}{n} + O\left(\frac{1}{n^2}\right).$$

En tenant compte du fait que  $\tilde{X}_1 = X_1 - \mu_{\boldsymbol{\theta}}$ , on obtient le résultat annoncé.

Pour montrer la convergence, il faut se servir de l'expression (8.4) et utiliser la loi forte des grands nombres sur chacun des deux termes du membre de droite :

$$\bar{V}_n = \frac{1}{n} \sum_{j=1}^n X_j^2 - \left( \frac{1}{n} \sum_{j=1}^n X_j \right)^2 \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}_{\boldsymbol{\theta}}(X_1^2) - \mathbb{E}_{\boldsymbol{\theta}}(X_1)^2 \text{ avec probabilité } 1.$$

Enfin, pour prouver la normalité asymptotique, on invoque le théorème de la limite centrale pour prouver la convergence en loi suivante :

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \tilde{X}_i^2 - \sigma_{\boldsymbol{\theta}}^2 \right) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathcal{N}(0, \mu_{\boldsymbol{\theta}}^{(4)} - \sigma_{\boldsymbol{\theta}}^4).$$

On a :

$$\sqrt{n}(\bar{V}_n - \sigma_{\boldsymbol{\theta}}^2) = \sqrt{n} \left( \frac{1}{n} \sum_{j=1}^n \tilde{X}_j^2 - \sigma_{\boldsymbol{\theta}}^2 \right) - Y_n,$$

avec

$$Y_n = \sqrt{n} \left( \frac{1}{n} \sum_{j=1}^n \tilde{X}_j \right)^2.$$

Or  $(Y_n)_n$  converge vers 0 en probabilité, car pour tout  $\epsilon > 0$  :

$$\mathbb{P}_{\boldsymbol{\theta}}(|Y_n| > \epsilon) \leq \frac{\mathbb{E}_{\boldsymbol{\theta}}(Y_n)}{\epsilon} = \frac{\sqrt{n} \text{Var}_{\boldsymbol{\theta}}(X_1)}{n\epsilon} \xrightarrow[n \rightarrow +\infty]{} 0.$$

En utilisant le théorème de Slutsky, on trouve alors que

$$\sqrt{n}(\bar{V}_n - \sigma_{\boldsymbol{\theta}}^2) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathcal{N}(0, \mu_{\boldsymbol{\theta}}^{(4)} - \sigma_{\boldsymbol{\theta}}^4),$$

en loi, comme annoncé.  $\square$

Il est facile de proposer un estimateur non-biaisé de la variance, il suffit de prendre la variance empirique  $\bar{V}_n$  et de la multiplier par  $n/(n-1)$  (si  $n \geq 2$ ). On obtient alors l'estimateur appelé variance empirique non-biaisée dont les propriétés sont décrites dans la proposition suivante.

**Proposition 8.20** *Soit  $(X_i)_{i=1,\dots,n}$  un  $n$ -échantillon d'un modèle statistique dont les lois sont de carré intégrable et dont le paramètre  $\theta$  contient la variance, i.e., il existe une fonction  $g : \Theta \rightarrow \mathbb{R}$  telle que  $g(\theta)$  est la variance  $\sigma_\theta^2$  de la loi  $P_\theta$ . La variance empirique non-biaisée  $V_n$  est définie par :*

$$V_n = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2. \quad (8.7)$$

C'est un estimateur de la variance qui satisfait les propriétés suivantes.

(i)  $V_n$  est non-biaisé : pour tout  $n \geq 2$ ,  $\mathbb{E}_\theta(V_n) = \sigma_\theta^2$ .

(ii)  $V_n$  est convergent.

(iii) Si les lois sont de quatrième moment fini, alors le risque quadratique moyen de la variance empirique est d'ordre  $1/n$  :

$$\text{RQM}_\theta(V_n) = \frac{\mu_\theta^{(4)} - \sigma_\theta^4}{n} + O\left(\frac{1}{n^2}\right),$$

et  $V_n$  est asymptotiquement normal :

$$\sqrt{n}(V_n - \sigma_\theta^2) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \mu_\theta^{(4)} - \sigma_\theta^4),$$

en loi.

Noter que les risques quadratiques moyens de  $V_n$  et de  $\bar{V}_n$  sont les mêmes à l'ordre  $1/n$ . C'est normal, car la différence entre ces deux estimateurs se situe au niveau du biais. Or le risque quadratique moyen d'ordre  $1/n$  mesure une erreur d'ordre  $1/\sqrt{n}$ , il est donc complètement insensible au biais qui est plus petit, d'ordre  $1/n$ .

**Preuve.** Le premier point découle de la linéarité de l'espérance et du fait que  $V_n = \frac{n}{n-1} \bar{V}_n$ .

Pour montrer le deuxième point, il faut utiliser l'expression :

$$V_n = \frac{n}{n-1} \times \left[ \frac{1}{n} \sum_{j=1}^n X_j^2 \right] - \frac{n}{n-1} \times \left[ \frac{1}{n} \sum_{j=1}^n X_j \right]^2.$$

On applique la loi forte des grands nombres à chacun des crochets. Avec probabilité 1, le premier crochet converge vers  $\mathbb{E}_\theta(X_1^2)$  et le second crochet converge vers  $\mathbb{E}_\theta(X_1)$ .

Comme  $n/(n-1)$  converge vers 1, on obtient que  $V_n$  converge vers  $\mathbb{E}_\theta(X_1^2) - \mathbb{E}_\theta(X_1)^2$  avec probabilité 1, ce qui donne la convergence de l'estimateur.

Le troisième et dernier point découle du calcul suivant :

$$\text{RQM}_\theta(V_n) = \mathbb{E}_\theta(V_n^2) - 2\sigma_\theta^2 \mathbb{E}_\theta(V_n) + \sigma_\theta^4,$$

et des développements des deux premiers moments de  $\bar{V}_n$  obtenus précédemment. La convergence en loi s'obtient à partir du résultat correspondant pour  $\bar{V}_n$  et du théorème de Slutsky.  $\square$

## 8.4 Méthode de substitution

On décrit ici une méthode générale de construction d'estimateurs convergents, qu'on va appliquer sur des cas particuliers dans la suite. Soit  $(X_i)_{i=1,\dots,n}$  un  $n$ -échantillon d'un modèle statistique  $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ . Soit  $\mathbf{g} : \Theta \rightarrow \Theta' \subset \mathbb{R}^d$  une fonction. On suppose qu'on dispose d'un estimateur convergent  $\hat{\mathbf{g}}_n$  de  $\mathbf{g}(\theta)$ . Si  $\phi : \Theta' \rightarrow \Theta'' \subset \mathbb{R}^q$  est une fonction continue, alors  $\phi(\hat{\mathbf{g}}_n)$  est un estimateur convergent de  $\phi(\mathbf{g}(\theta))$  :

$$\mathbb{P}_\theta \left( \phi(\hat{\mathbf{g}}_n) \xrightarrow{n \rightarrow +\infty} \phi(\mathbf{g}(\theta)) \right) = 1.$$

C'est une conséquence directe de la continuité de  $\phi$ . La principale application de cette méthode générale est la méthode des moments qu'on décrit ci-dessous.

## 8.5 Méthode des moments

Soit  $(X_i)_{i=1,\dots,n}$  un  $n$ -échantillon d'un modèle statistique  $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ . L'objectif est de construire un estimateur convergent de  $\theta$ . La méthode des moments consiste à trouver une fonction  $\mathbf{g} : \Theta \rightarrow \Theta' \subset \mathbb{R}^d$  inversible et de fonction réciproque continue et une fonction  $\boldsymbol{\psi} : \mathcal{X} \rightarrow \mathbb{R}^d$  telle que  $\mathbb{E}_\theta(|\boldsymbol{\psi}(X_1)|) < \infty$  et  $\mathbf{g}(\theta)$  peut s'écrire comme l'espérance de  $\boldsymbol{\psi}(X_1)$  pour tout  $\theta \in \Theta$  :

$$\mathbf{g}(\theta) = \mathbb{E}_\theta(\boldsymbol{\psi}(X_1)).$$

L'estimateur des moments de  $\theta$  est alors

$$\hat{\boldsymbol{\theta}}_n = \mathbf{g}^{-1} \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\psi}(X_i) \right).$$

L'estimateur des moments est convergent car c'est un cas particulier de la méthode de substitution. En effet,  $\frac{1}{n} \sum_{i=1}^n \boldsymbol{\psi}(X_i)$  est un estimateur convergent de  $\mathbf{g}(\theta) = \mathbb{E}_\theta(\boldsymbol{\psi}(X_1))$  d'après la loi forte des grands nombres, et  $\mathbf{g}^{-1}$  est continue.

**Exemple 8.21** On considère le modèle statistique  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P})$  avec  $\mathcal{P} = \{\beta(\alpha, \beta), \alpha, \beta \in ]0, \infty[^2\}$ . On rappelle que la loi  $\beta(\alpha, \beta)$  sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  est à densité (cf (5.17)) :

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \mathbf{1}_{]0,1[}(x).$$

Ici le paramètre  $\theta = (\alpha, \beta)$  est de dimension 2. On désire estimer  $\theta$ . Si  $X_1$  a pour loi  $\beta(\alpha, \beta)$ , alors

$$\mathbb{E}_{\theta}(X_1) = \frac{\alpha}{\alpha + \beta} = g \text{ et } \mathbb{E}_{\theta}(X_1(1 - X_1)) = \frac{\alpha\beta}{(\alpha + \beta)(\alpha + \beta + 1)} = h.$$

D'une part, en inversant le système précédent, on trouve que

$$\alpha = \frac{gh}{g - h - g^2} \text{ et } \beta = \frac{(1 - g)h}{g - h - g^2}.$$

D'autre part, si  $(X_i)_{i=1, \dots, n}$  est un  $n$ -échantillon de loi  $\beta(\alpha, \beta)$ , alors on peut construire à l'aide de la loi forte des grands nombres des estimateurs convergents de  $g$  et  $h$  :

$$\hat{g}_n = \frac{1}{n} \sum_{i=1}^n X_i \text{ et } \hat{h}_n = \frac{1}{n} \sum_{i=1}^n X_i(1 - X_i).$$

On déduit de la méthode des moments que

$$\hat{\alpha}_n = \frac{\hat{g}_n \hat{h}_n}{\hat{g}_n - \hat{h}_n - \hat{g}_n^2} \text{ et } \hat{\beta}_n = \frac{(1 - \hat{g}_n) \hat{h}_n}{\hat{g}_n - \hat{h}_n - \hat{g}_n^2}$$

sont des estimateurs convergents de  $\alpha$  et  $\beta$ .

**Exemple 8.22** On considère le modèle uniforme  $\mathcal{P} = \{P_{\theta}, \theta \in ]0, +\infty[ \}$ , où  $P_{\theta}$  est la loi uniforme sur  $[0, \theta]$ . Si  $X_1$  a pour loi  $P_{\theta}$ , alors on a :

$$\mathbb{E}_{\theta}(X_1) = \frac{\theta}{2}.$$

Par conséquent, si  $(X_i)_{i=1, \dots, n}$  est un  $n$ -échantillon de loi  $P_{\theta}$ , alors on peut construire à l'aide de la loi forte des grands nombres un estimateur convergent de  $\theta$  :

$$\hat{\theta}_n = \frac{2}{n} \sum_{i=1}^n X_i.$$

Cet estimateur est non-biaisé, son RQM est

$$\text{RQM}_{\theta}(\hat{\theta}_n) = \text{Var}_{\theta}(\hat{\theta}_n) = \frac{4}{n} \text{Var}_{\theta}(X_1) = \frac{\theta^2}{3n},$$

et il est asymptotiquement normal :

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \theta^2/3).$$

On va voir ci-dessous qu'on peut en fait construire un bien meilleur estimateur de  $\theta$ .

## 8.6 Maximum de vraisemblance

Dans cette section on considère un modèle statistique  $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$  et on supposera :

- soit que pour tout  $\theta \in \Theta$ , la loi  $P_\theta$  est discrète et à valeurs dans un même espace  $\mathcal{X}$  au plus dénombrable. On pose alors  $p(x, \theta) = P_\theta(\{x\})$  pour tout  $x \in \mathcal{X}$ . On pose pour tout  $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$  :

$$p_n(\mathbf{x}, \theta) = p(x_1, \theta) \cdots p(x_n, \theta),$$

qu'on appelle vraisemblance. À  $\theta$  fixé, en tant que fonction de  $\mathbf{x}$ , c'est la loi d'un  $n$ -échantillon  $(X_i)_{i=1}^n$  sous  $\mathbb{P}_\theta$ .

- soit que pour tout  $\theta \in \Theta$ , la loi  $P_\theta$  est à densité par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ . On note alors  $p(x, \theta)$  la densité de la loi sous  $\mathbb{P}_\theta$  et  $p_n(\mathbf{x}, \theta)$  la densité jointe d'un  $n$ -échantillon, qu'on appelle vraisemblance.

**Définition 8.23** On suppose que pour toute réalisation  $\mathbf{x} = (x_1, \dots, x_n)$  d'un  $n$ -échantillon  $\mathbf{X} = (X_1, \dots, X_n)$ , il existe une unique valeur  $\theta_n(\mathbf{x}) \in \Theta$  qui maximise la vraisemblance (vue comme fonction de  $\theta$ ) de la réalisation  $\mathbf{x}$  :

$$\theta_n(\mathbf{x}) = \operatorname{argmax}_{\theta \in \Theta} p_n(\mathbf{x}, \theta).$$

Alors l'estimateur  $\hat{\theta}_n = \theta_n(\mathbf{X})$  est appelé Estimateur du Maximum de Vraisemblance (EMV) de  $\theta$ .

L'EMV est le paramètre qui maximise la vraisemblance des données étant donné le paramètre. On peut expliquer son principe par une interprétation bayésienne. Pour simplifier la présentation, supposons provisoirement que  $\mathcal{X}$  et  $\Theta$  sont finis. Avant de recueillir des données, on ne sait rien sur le paramètre  $\theta$ , à part qu'il est dans  $\Theta$ , donc on peut considérer que le paramètre est une variable aléatoire discrète de loi uniforme sur  $\Theta$ . Dans ce cadre, on peut donc considérer que la loi jointe des données et du paramètre est :

$$P(\mathbf{x}, \theta) = p_n(\mathbf{x}, \theta) \frac{1}{\operatorname{card}(\Theta)},$$

car  $p_n(\mathbf{x}, \theta)$  est la loi d'un  $n$ -échantillon sachant le paramètre  $\theta$ . Par le théorème de Bayes, la loi du paramètre sachant les données est :

$$P(\theta|\mathbf{x}) = \frac{P(\mathbf{x}, \theta)}{P(\mathbf{x})}, \text{ avec } P(\mathbf{x}) = \sum_{\theta' \in \Theta} P(\mathbf{x}, \theta'),$$

ce qui s'écrit donc :

$$P(\theta|\mathbf{x}) = \frac{p_n(\mathbf{x}, \theta)}{\operatorname{card}(\Theta)P(\mathbf{x})}.$$



Le mode de cette loi, i.e. le paramètre  $\theta$  le plus probable dans  $\Theta$  sachant les données  $\mathbf{x}$ , est l'élément de  $\Theta$  qui maximise  $P(\theta|\mathbf{x})$ , qui est aussi celui qui maximise  $p_n(\mathbf{x}, \theta)$  : c'est donc l'EMV.

**Exemple 8.24** Dans le modèle de Bernoulli,  $\mathcal{X} = \{0, 1\}$ ,

$$\mathcal{P} = \{\mathcal{B}(1, \theta), \theta \in [0, 1]\},$$

la vraisemblance est :

$$p_n(\mathbf{x}, \theta) = \prod_{i=1}^n \theta^{1_{\{x_i=1\}}}(1-\theta)^{1_{\{x_i=0\}}},$$

avec la notation  $\mathbf{1}_a(x) = 1$  si  $x = a$  et 0 sinon. Ceci s'écrit aussi :

$$p_n(\mathbf{x}, \theta) = [\theta^{\bar{x}_n}(1-\theta)^{1-\bar{x}_n}]^n, \quad (8.8)$$

avec  $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ . A  $\mathbf{x}$  fixé, en tant que fonction de  $\theta$ , on remarque que la vraisemblance est maximale lorsque la fonction  $\theta \mapsto \theta^{\bar{x}_n}(1-\theta)^{1-\bar{x}_n}$  est maximale. Le maximum sur  $[0, 1]$  est unique et est atteint au point où la dérivée de la fonction s'annule, en  $\theta = \bar{x}_n$ . Le maximum de la vraisemblance redonne ici l'estimateur empirique.

**Exemple 8.25** Dans le modèle gaussien

$$\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 \in ]0, \infty[ \},$$

la vraisemblance est

$$p_n(\mathbf{x}, (\mu, \sigma^2)) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right). \quad (8.9)$$

On va voir ci-dessous que le maximum de vraisemblance, i.e., la position du maximum de  $(\mu, \sigma^2) \mapsto p_n(\mathbf{x}, (\mu, \sigma^2))$ , est ici aussi l'estimateur empirique.

Comme la fonction logarithme est strictement croissante, il revient au même de maximiser la vraisemblance  $\theta \mapsto p_n(\mathbf{x}, \theta)$  ou de maximiser la log-vraisemblance  $\theta \mapsto l_n(\mathbf{x}, \theta)$  définie par :

$$l_n(\mathbf{x}, \theta) = \ln(p_n(\mathbf{x}, \theta)).$$

On pose également  $l(x, \theta) = \ln(p(x, \theta))$ . Les calculs sont parfois plus aisés avec la log-vraisemblance notamment parce que  $l_n(\mathbf{x}, \theta) = l(x_1, \theta) + \dots + l(x_n, \theta)$ , ce qui simplifie le calcul des dérivées.

**Exemple 8.26** On considère le modèle gaussien  $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$ . Pour tout  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ , la vraisemblance est donnée par (8.9). Donc la log-vraisemblance s'écrit

$$l_n(\mathbf{x}, (\mu, \sigma^2)) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - n \frac{(\bar{x}_n - \mu)^2 + \bar{v}_n}{2\sigma^2},$$

en notant  $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$  et  $\bar{v}_n = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$ . A  $\sigma^2 > 0$  fixé, on voit que la log-vraisemblance est maximale pour  $\mu = \bar{x}_n$  et vaut alors  $-\frac{n}{2}(\ln(2\pi) + f(\sigma^2))$  avec

$$f(s) = \ln(s) + \frac{\bar{v}_n}{s}.$$

On cherche donc maintenant à minimiser  $f(s)$  pour  $s \in ]0, +\infty[$ . Comme la dérivée  $f'(s) = 1/s - \bar{v}_n/s^2$  est négative sur  $]0, \bar{v}_n]$  et positive sur  $[\bar{v}_n, +\infty[$ , la fonction  $f$  atteint son minimum en  $\bar{v}_n$ . On conclut donc que la log-vraisemblance est maximale en  $(\bar{x}_n, \bar{v}_n)$ . Ainsi l'EMV de  $(\mu, \sigma^2)$  est le couple moyenne empirique, variance empirique  $(\bar{X}_n, \bar{V}_n)$ . Notons qu'on obtient également l'EMV en résolvant le système

$$\begin{cases} \frac{\partial}{\partial \mu} l_n(\mathbf{x}, (\mu, \sigma^2)) = 0 \\ \frac{\partial}{\partial (\sigma^2)} l_n(\mathbf{x}, (\mu, \sigma^2)) = 0 \end{cases} \iff \begin{cases} n \frac{\bar{x}_n - \mu}{\sigma^2} = 0 \\ -\frac{n}{2} \left( \frac{1}{\sigma^2} - \frac{(\bar{x}_n - \mu)^2 + \bar{v}_n}{\sigma^4} \right) = 0 \end{cases}$$

Comme  $\mathbb{E}_{(\mu, \sigma^2)}[(\bar{X}_n, \bar{V}_n)] = (\mu, \frac{n-1}{n}\sigma^2)$ , l'EMV est un estimateur biaisé. Il est convergent d'après la loi forte des grands nombres. Pour démontrer qu'il est asymptotiquement normal, on remarque que d'après la proposition 9.4 le vecteur aléatoire  $(\bar{X}_n, \bar{V}_n)$  a la même loi que  $(\frac{1}{n} \sum_{j=1}^n X_j, \frac{\sigma^2}{n} \sum_{j=2}^n Y_j^2)$  où  $(Y_i)_{i=1, \dots, n}$  est une suite de variables indépendantes et identiquement distribuées suivant la loi gaussienne centrée réduite indépendante de  $(X_i)_{i=1, \dots, n}$ . On en déduit que :

$$\sqrt{n} \left( \begin{pmatrix} \bar{X}_n \\ \bar{V}_n \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \right) \stackrel{\text{loi}}{=} \sqrt{n} \left( \frac{1}{n} \sum_{j=1}^n \begin{pmatrix} X_j \\ \sigma^2 Y_j^2 \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \right) - \begin{pmatrix} 0 \\ \frac{\sigma^2 Y_1^2}{\sqrt{n}} \end{pmatrix}.$$

D'après le théorème de la limite centrale vectoriel, le premier terme du membre de droite converge en loi vers la loi gaussienne centrée de matrice de covariance égale à celle du vecteur  $(X_1, \sigma^2 Y_1^2)$  c'est-à-dire

$$\mathbf{C} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}.$$

Le second terme du membre de droite converge presque sûrement vers  $(0, 0)$ . Par le théorème de Slutsky (théorème 7.25), on conclut que l'Estimateur du Maximum de Vraisemblance  $(\bar{X}_n, \bar{V}_n)$  est asymptotiquement normal de matrice de covariance asymptotique  $\mathbf{C}$ .

**Exemple 8.27** On termine par un exemple apparemment simple, mais dans lequel on voit que la normalité asymptotique n'est pas automatique. On considère à nouveau le modèle uniforme  $\mathcal{P} = \{P_\theta, \theta \in ]0, +\infty[ \}$ , où  $P_\theta$  est la loi uniforme sur  $[0, \theta]$ . La vraisemblance d'un  $n$ -échantillon est alors :

$$p_n(\mathbf{x}, \theta) = \theta^{-n} \prod_{i=1}^n \mathbf{1}_{[0, \theta]}(x_i) = \theta^{-n} \mathbf{1}_{[0, +\infty[} \left( \min_{i=1, \dots, n} (x_i) \right) \mathbf{1}_{[0, \theta]} \left( \max_{i=1, \dots, n} (x_i) \right).$$

La vraisemblance est maximale pour  $\theta_n(\mathbf{x}) = \max_{i=1, \dots, n}(x_i)$ . L'EMV est donc  $\hat{\theta}_n = \max_{i=1, \dots, n}(X_i)$ . Il est facile de calculer sa loi. Sa fonction de répartition est :

$$\mathbb{P}_\theta(\hat{\theta}_n \leq x) = \begin{cases} 1 & \text{si } x \geq \theta, \\ \mathbb{P}_\theta(X_1 \leq x)^n = \left(\frac{x}{\theta}\right)^n & \text{si } 0 \leq x < \theta, \\ 0 & \text{si } x < 0. \end{cases}$$

Donc, sous  $\mathbb{P}_\theta$ ,  $\hat{\theta}_n$  est une v.a. à densité :

$$p_{\hat{\theta}_n}(x) = n \frac{x^{n-1}}{\theta^n} \mathbf{1}_{[0, \theta]}(x).$$

Ceci permet de montrer que l'EMV  $\hat{\theta}_n$  est biaisé :

$$\mathbb{E}_\theta(\hat{\theta}_n) = \int_0^\theta x p_{\hat{\theta}_n}(x) dx = \frac{n\theta}{n+1} = \theta - \frac{\theta}{n+1},$$

avec un biais d'ordre  $1/n$ . La variance est

$$\text{Var}_\theta(\hat{\theta}_n) = \mathbb{E}_\theta(\hat{\theta}_n^2) - \mathbb{E}_\theta(\hat{\theta}_n)^2 = \frac{n}{n+2} \theta^2 - \frac{n^2}{(n+1)^2} \theta^2 = \frac{n}{(n+1)^2(n+2)} \theta^2,$$

qui est d'ordre  $1/n^2$ . Le RQM est donc d'ordre  $1/n^2$  lui aussi :

$$\text{RQM}_\theta(\hat{\theta}_n) = \text{Var}_\theta(\hat{\theta}_n) + (\mathbb{E}_\theta(\hat{\theta}_n) - \theta)^2 = \frac{2}{(n+1)(n+2)} \theta^2,$$

ce qui est plus petit que les RQM des EMV observés dans les modèles précédents, gaussiens par exemple, et aussi plus petit que le RQM de l'estimateur empirique pour le modèle uniforme obtenu par la méthode des moments dans l'exemple 8.22. De plus, il apparaît clairement qu'on n'est pas dans le régime du théorème de la limite centrale, où on s'attend à une variance en  $1/n$ . Effectivement,

$$\mathbb{P}_\theta(n(\hat{\theta}_n - \theta) \leq x) = \begin{cases} 1 & \text{si } x \geq 0, \\ \left(1 + \frac{x}{n\theta}\right)^n & \text{si } -n\theta \leq x < 0, \\ 0 & \text{si } x < -n\theta, \end{cases}$$

et donc, pour tout  $x \geq 0$  :

$$\mathbb{P}_\theta(n(\hat{\theta}_n - \theta) \leq -\theta x) \xrightarrow{n \rightarrow +\infty} e^{-x},$$

ce qui montre que

$$n(\hat{\theta}_n - \theta) \xrightarrow{n \rightarrow +\infty} -\theta Z$$

en loi, où  $Z$  est une v.a. exponentielle de paramètre 1. Les fluctuations de l'EMV sont d'ordre  $1/n$  et de loi exponentielle lorsque  $n$  est grand. On est donc loin de la normalité asymptotique.

Finalement, on peut débiaiser l'EMV en considérant l'estimateur :

$$\tilde{\theta}_n = \frac{n+1}{n} \hat{\theta}_n = \frac{n+1}{n} \max_{i=1, \dots, n} (X_i).$$

L'estimateur  $\tilde{\theta}_n$  est non-biaisé, son RQM est :

$$\text{RQM}_\theta(\tilde{\theta}_n) = \text{Var}_\theta(\tilde{\theta}_n) = \frac{(n+1)^2}{n^2} \text{Var}_\theta(\hat{\theta}_n) = \frac{1}{n(n+2)} \theta^2,$$

qui est plus petit que le RQM de  $\hat{\theta}_n$ , et il satisfait :

$$n(\tilde{\theta}_n - \theta) \xrightarrow{n \rightarrow +\infty} -\theta(Z-1),$$

en loi, où  $Z$  est une v.a. exponentielle de paramètre 1. Notez que  $\mathbb{E}((Z-1)^2) = 1 < 2 = \mathbb{E}(Z^2)$ , ce qui montre que la variance asymptotique de  $\tilde{\theta}_n$  est deux fois plus petite que celle de  $\hat{\theta}_n$ . Le "petit" débiaisage s'avère bien utile ici.

En fait, la normalité asymptotique se rencontre avec des modèles dits réguliers, pour lesquels la vraisemblance a de bonnes propriétés de régularité que nous ne détaillerons pas ici, mais qui réclament en particulier que le support (en  $x$ ) de la loi  $p(x, \theta)$  soit indépendant de  $\theta$ . Le modèle uniforme traité ci-dessus viole cette condition, mais c'est en fait une bonne chose du point de vue de l'estimation, puisqu'on tombe sur un estimateur qui converge plus vite que ce que prévoit le théorème de la limite centrale.

## 8.7 Exercices sur le chapitre 8

**EXERCICE 8.1** Montrer que la densité de la loi  $T_n$  est de la forme (9.5).

**EXERCICE 8.2** Approximation des lois  $\chi_n^2$  et  $T_n$ .  
Soit  $(Z_n)_n$  une suite de variables aléatoires réelles avec  $Z_n$  de loi  $\chi_n^2$ .

1. Montrer que  $((Z_n - n)/\sqrt{2n})_n$  converge en loi vers une loi gaussienne centrée réduite.
2. En déduire que  $(\sqrt{2Z_n} - \sqrt{2n-1})_n$  converge en loi vers une loi gaussienne centrée réduite.

La qualité de la seconde approximation est en fait légèrement meilleure que la première.

Soit  $(\zeta_n)_n$  une suite de variables aléatoires réelles avec  $\zeta_n$  de loi  $T_n$ .

3. Montrer que  $(\zeta_n)_n$  converge en loi vers une loi gaussienne centrée réduite.

**EXERCICE 8.3** Réduction de variance dans une méthode de Monte Carlo.

Soit  $g$  une fonction mesurable telle que  $0 \leq g \leq 1$ . On souhaite calculer  $m = \int_0^1 g(x)dx$ . Soient  $X$  et  $Y$  des variables indépendantes et identiquement distribuées, de loi uniforme sur  $[0, 1]$  et

$$U = \mathbf{1}_{Y \leq g(X)}, \quad V = g(X) \quad \text{et} \quad W = \frac{g(X) + g(1-X)}{2}.$$

1. Calculer l'espérance et la variance de  $U$ ,  $V$  et  $W$ .
2. Proposer 3 méthodes de type Monte-Carlo pour calculer  $m$ .

On suppose dans la suite que  $g$  est monotone.

3. Montrer que  $(g(x) - g(y))(g(1-x) - g(1-y)) \leq 0$  pour tous  $x, y$ .  
En déduire que

$$\mathbb{E}(g(X)g(1-X)) = \int_0^1 g(x)g(1-x) dx \leq m^2 \leq \int_0^1 g(x)^2 dx.$$

Comparer les variances de  $U$ ,  $V$ ,  $W$ .

4. Soit  $(X_i)_{i \geq 1}$  une suite de variables aléatoires indépendantes et de même loi uniforme sur  $[0, 1]$ . Des estimateurs

$$A_n = \frac{1}{2n} \sum_{i=1}^{2n} g(X_i), \quad B_n = \frac{1}{2n} \sum_{i=1}^n (g(X_i) + g(1-X_i)),$$

lequel est le meilleur pour calculer  $m$  ?

5. Pour  $g(x) = x^2$ , déterminer pour chaque estimateur  $A_n$  et  $B_n$  combien de simulations sont nécessaires pour obtenir une précision relative de l'ordre de 1% sur le calcul de  $m$  avec probabilité 95%.

**EXERCICE 8.4** Soit  $m$  un entier strictement positif fixé. On considère le modèle binomial à  $m$  fixé,  $\mathcal{X} = \{0, 1, \dots, m\}$ ,

$$\mathcal{P} = \{\mathcal{B}(m, \theta), \theta \in [0, 1]\}.$$

On observe un  $n$ -échantillon  $(X_1, \dots, X_n)$ .

1. Déterminer un estimateur de  $\theta$  par la méthode des moments.
2. Donner l'Estimateur du Maximum de Vraisemblance de  $\theta$ .

**EXERCICE 8.5** On modélise la hauteur maximale annuelle d'un fleuve (exprimée en mètres) par une variable aléatoire dite de Rayleigh de densité  $p(x, a) = \frac{x}{a} \exp(-\frac{x^2}{2a}) \mathbf{1}_{]0, +\infty[}(x)$  où  $a > 0$  est un paramètre inconnu.

1. Calculer l'espérance  $\mathbb{E}_a(X)$  d'une variable aléatoire  $X$  de loi de Rayleigh de paramètre  $a$ . Calculer aussi  $\mathbb{E}_a(X^2)$  et  $\mathbb{E}_a(X^4)$ .
2. On observe un  $n$ -échantillon  $(X_1, \dots, X_n)$  suivant cette loi. Donner l'Estimateur du Maximum de Vraisemblance  $\hat{a}_n$  de  $a$ . Cet estimateur est-il sans biais? convergent? Vérifier qu'il est asymptotiquement normal et identifier la variance asymptotique.
3. Pendant une période de huit ans, on a observé les hauteurs maximales en mètres suivantes pour le fleuve :  $(x_1, \dots, x_8) = (2,5, 1,8, 2,9, 0,9, 2,1, 1,7, 2,2, 2,8)$ . On a  $\sum_{i=1}^8 x_i^2 = 38,69$ . Une compagnie d'assurance estime qu'une crue catastrophique avec une hauteur de 6 mètres au moins n'arrive au plus qu'une fois tous les mille ans. Est-ce justifié?

## Chapitre 9

# Statistique : Intervalle de confiance

*I only believe in statistics that I doctored myself.*

Winston Churchill

### 9.1 Intervalle de confiance et estimation

L'estimation d'un paramètre, même dans le cas d'un estimateur convergent, donnera une valeur différente de la vraie valeur inconnue. Ce qu'on peut dire, c'est que cette valeur inconnue est proche de la valeur estimée, mais tout l'art du statisticien est de quantifier cette erreur par nature aléatoire. Pour répondre rigoureusement au problème de l'estimation d'un paramètre, il est agréable de pouvoir donner un intervalle tel que le paramètre inconnu en fasse partie avec une grande probabilité donnée.

**Définition 9.1** Soit  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  un modèle statistique, avec  $\mathcal{P} = \{P_{\theta}, \theta \in \Theta\}$ . Soit  $g : \Theta \rightarrow \mathbb{R}$ . Soit  $\alpha \in ]0, 1[$ . On dit qu'un intervalle  $I_{\mathbf{X}}$  qui s'exprime en fonction d'un  $n$ -échantillon  $\mathbf{X}$  est un intervalle de confiance pour  $g(\theta)$  de niveau  $1 - \alpha$  si pour tout  $\theta \in \Theta$  :

$$\mathbb{P}_{\theta}(g(\theta) \in I_{\mathbf{X}}) = 1 - \alpha.$$

Lorsque pour tout  $\theta \in \Theta$ , on a  $\mathbb{P}_{\theta}(g(\theta) \in I_{\mathbf{X}}) \geq 1 - \alpha$ , on parle d'intervalle de confiance de niveau  $1 - \alpha$  par excès.

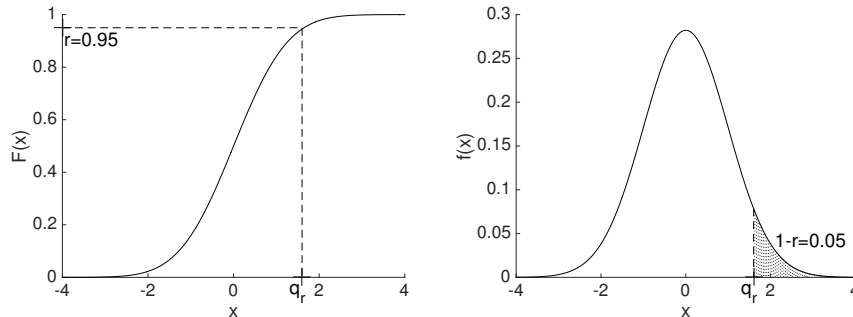


FIGURE 9.1 – Détermination du quantile  $q_r$  d'ordre  $r = 0,95$  d'une loi à partir de la fonction de répartition  $F(x)$  de la loi (gauche) et à partir de la densité  $f(x)$  de la loi (droite). Ici on a pris le cas d'une loi gaussienne centrée réduite.

L'intervalle de confiance  $I_{\mathbf{X}}$  est donc aléatoire dans le sens où ses bornes dépendent de l'échantillon  $\mathbf{X}$ . Lorsqu'on observe un échantillon, on peut affirmer que la vraie valeur  $g(\theta)$  appartient à l'intervalle  $I_{\mathbf{X}}$  construit à partir de l'échantillon observé avec une certitude (ou niveau de confiance) prescrite à l'avance.

Les niveaux usuels sont 90%, 95%, et 99% et correspondent respectivement à  $\alpha = 0,1$ ,  $\alpha = 0,05$  et  $\alpha = 0,01$ .

Pour construire des intervalles de confiance, il est très utile d'introduire la notion de quantile.

**Définition 9.2** On considère la loi d'une variable aléatoire réelle de fonction de répartition  $F$ . Pour  $r \in ]0, 1[$ , on appelle quantile (ou fractile) d'ordre  $r$  de la loi le nombre

$$q_r = \inf \{x \in \mathbb{R}, F(x) \geq r\}.$$

Lorsque la fonction de répartition  $F$  est continue et strictement croissante (par exemple quand la v.a. possède une densité strictement positive, comme sur la figure 9.1), elle est inversible d'inverse  $F^{-1}$  et pour tout  $r \in ]0, 1[$ , on a  $q_r = F^{-1}(r)$ . Par exemple, la médiane est le quantile d'ordre  $1/2$  : Une v.a. réelle a autant de chances d'être plus petite ou plus grande que la médiane. Le premier quartile est le quantile d'ordre  $1/4$  : Une v.a. réelle a une chance sur quatre d'être plus petite et trois chances sur quatre d'être plus grande que le premier quartile.

La fonction de répartition est toujours croissante, ce qui entraîne la croissance de  $r \mapsto q_r$ . Pour construire des intervalles de confiance et des tests, nous utiliserons les propriétés suivantes :



**Proposition 9.3** *On suppose que la loi de la v.a. réelle  $X$  de fonction de répartition  $F$  possède une densité. Les quantiles de la loi satisfont alors les propriétés suivantes.*

1. Pour tout  $r \in ]0, 1[$ ,  $F(q_r) = r$ .
2. Pour tout  $\alpha \in ]0, 1[$ ,  $\mathbb{P}(X \notin [q_{\alpha/2}, q_{1-\alpha/2}]) = \mathbb{P}(X < q_\alpha) = \mathbb{P}(X > q_{1-\alpha}) = \alpha$ .
3. Pour tout  $\alpha \in ]0, 1[$ ,  $\mathbb{P}(X \in [q_{\alpha/2}, q_{1-\alpha/2}]) = \mathbb{P}(X \geq q_\alpha) = \mathbb{P}(X \leq q_{1-\alpha}) = 1 - \alpha$ .
4. Si la loi de  $X$  est symétrique (i.e. la densité est une fonction paire), alors pour tout  $\alpha \in ]0, 1[$ ,  $\mathbb{P}(|X| > q_{1-\alpha/2}) = \alpha$  et  $\mathbb{P}(|X| \leq q_{1-\alpha/2}) = 1 - \alpha$ .

**Preuve.** 1. Pour tout  $y < q_r$ , on a  $F(y) < r$  et par croissance de  $F$ , pour tout  $y > q_r$ ,  $F(y) \geq r$ . Comme  $F$  est continue, on en déduit que  $F(q_r) = r$ .

2. Le résultat se déduit des égalités  $\mathbb{P}(X < q_r) = \mathbb{P}(X \leq q_r) = F(q_r) = r$  et  $\mathbb{P}(X > q_r) = 1 - F(q_r) = 1 - r$ .

3. Ce point s'obtient par passage au complémentaire.

4. Lorsque la densité de  $X$  est une fonction paire, la variable aléatoire  $-X$  a même loi que  $X$ . En outre  $F(0) = 1/2$ , ce qui entraîne que  $q_{1-\alpha/2} > 0$ . Donc :

$$\begin{aligned} \mathbb{P}(|X| > q_{1-\alpha/2}) &= \mathbb{P}(X < -q_{1-\alpha/2}) + \mathbb{P}(X > q_{1-\alpha/2}) \\ &= \mathbb{P}(-X > q_{1-\alpha/2}) + \mathbb{P}(X > q_{1-\alpha/2}) = 2\mathbb{P}(X > q_{1-\alpha/2}) \\ &= \alpha, \end{aligned}$$

et la dernière propriété s'en déduit par passage au complémentaire.  $\square$

Pour obtenir des intervalles de confiance sur la moyenne et la variance d'une loi inconnue dont on a un échantillon, on a besoin de certaines propriétés sur des estimateurs, tels que la moyenne empirique, la variance empirique non-biaisée, la moyenne empirique renormalisée par la variance empirique non-biaisée, etc. Il est établi dans la section précédente que, avec probabilité 1 :

$$\bar{X}_n \xrightarrow{n \rightarrow +\infty} \mu \text{ et } V_n \xrightarrow{n \rightarrow +\infty} \sigma^2,$$

c'est-à-dire que la moyenne empirique (8.3) et la variance empirique non-biaisée (8.7) sont des estimateurs convergents de l'espérance  $\mu$  et de la variance  $\sigma^2$ . Mais on a besoin de plus. Pour résumer, il arrive dans certains cas qu'on puisse caractériser entièrement la loi des estimateurs, ce qui permet de construire des intervalles de confiance exacts (et valables pour tout  $n$ ). Mais le plus souvent la situation est trop compliquée, et on se sert alors des propriétés asymptotiques des estimateurs (en particulier, la normalité asymptotique) pour construire des intervalles de confiance asymptotiques, qui sont valables pour  $n$  suffisamment grand.

## 9.2 Intervalles exacts pour le modèle gaussien

La proposition suivante caractérise la distribution statistique de la moyenne empirique et de la variance empirique non-biaisée dans le cas d'un échantillon à statistique gaussienne.

**Proposition 9.4** Soit  $(X_i)_{i=1,\dots,n}$  un  $n$ -échantillon de loi  $\mathcal{N}(\mu, \sigma^2)$ , avec  $n \geq 2$ . Les v.a. réelles  $\bar{X}_n$  et  $V_n$  définies par (8.3) et (8.7) sont indépendantes pour tout  $n$ . De plus, pour tout  $n$ , on a :

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1), \quad (9.1)$$

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{V_n}} \sim T_{n-1}, \quad (9.2)$$

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi_n^2, \quad (9.3)$$

$$(n-1) \frac{V_n}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \sim \chi_{n-1}^2. \quad (9.4)$$

Les lois  $T_n$  (loi de Student à  $n$  degrés de liberté) et  $\chi_n^2$  (loi de  $\chi^2$  à  $n$  degrés de liberté) sont décrites dans les tables 9.1 et 9.2. La loi de  $\chi^2$  a déjà été introduite dans la définition 7.17, c'est la loi de la somme des carrés de  $n$  variables aléatoires gaussiennes centrées réduites indépendantes, et elle a pour densité (7.20). Pour  $n$  entier strictement positif, la loi  $T_n$  est définie de la manière suivante : Soit  $Z$  une variable aléatoire de loi gaussienne centrée réduite et soit  $U$  une variable indépendante de  $Z$  et distribuée suivant la loi de  $\chi^2$  à  $n$  degrés de liberté. Par définition la variable  $T = Z/\sqrt{U/n}$  suit une loi de Student à  $n$  degrés de liberté. La densité de  $T$  est paire et donnée par :

$$f(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad (9.5)$$

où  $\Gamma$  est la fonction Gamma d'Euler :  $\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} dx$  (voir exercice 8.1). Son espérance ne peut pas être définie pour  $n = 1$  et est nulle pour  $n \geq 2$ . Sa variance est infinie pour  $n \leq 2$  et vaut  $n/(n-2)$  pour  $n \geq 3$ .

La proposition 9.4 donne donc les lois exactes des estimateurs empiriques de la moyenne et de la variance pour tout  $n$ , ce qui va nous permettre de construire des intervalles de confiance exacts pour ces deux paramètres.

**Preuve.** Introduisons les v.a.  $Z_j = (X_j - \mu)/\sigma$ . Ce sont des v.a. gaussiennes indépendantes centrées réduites, et on peut exprimer le  $n$ -échantillon  $(X_i)_{i=1,\dots,n}$

comme  $X_j = \mu + \sigma Z_j$ . La v.a. réelle  $Y_1$  définie par :

$$Y_1 = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} = \frac{1}{\sqrt{n}} \sum_{j=1}^n Z_j \quad (9.6)$$

a une loi gaussienne. Sa moyenne est 0 car les  $Z_j$  sont centrés, et sa variance est 1 car les  $Z_j$  sont indépendants et de variance 1.

On trouve de la même façon :

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = \sum_{j=1}^n Z_j^2,$$

ce qui montre que cette v.a. réelle suit une loi  $\chi_n^2$ .

Considérons maintenant la variance empirique renormalisée (9.4). On peut l'écrire sous la forme :

$$\begin{aligned} (n-1) \frac{V_n}{\sigma^2} &= \frac{1}{\sigma^2} \sum_{i=1}^n \left( (\mu + \sigma Z_i) - \left( \mu + \frac{\sigma}{\sqrt{n}} Y_1 \right) \right)^2 = \sum_{i=1}^n \left( Z_i - \frac{Y_1}{\sqrt{n}} \right)^2 \\ &= \sum_{i=1}^n Z_i^2 - Y_1^2. \end{aligned}$$

Donnons-nous maintenant une base  $(\mathbf{e}_1, \dots, \mathbf{e}_n)$  de  $\mathbb{R}^n$  dont le premier vecteur est le vecteur dont toutes les coordonnées valent  $1/\sqrt{n}$ , les autres vecteurs orthonormaux de la base étant choisis arbitrairement. Appelons  $\mathbf{U}$  la matrice  $n \times n$  orthogonale de vecteurs lignes donnés par les  $\mathbf{e}_j$ , et  $\mathbf{Y}$  le vecteur aléatoire donné par  $\mathbf{Y} = \mathbf{U}\mathbf{Z}$ . On peut noter que la première coordonnée  $Y_1$  est bien donnée par (9.6). Le vecteur  $\mathbf{Y}$  est un vecteur gaussien de moyenne nulle et de matrice de covariance  $\mathbf{C} = \mathbf{U}\mathbf{U}^t = \mathbf{I}$ . Cela veut dire que les  $Y_j$  sont des v.a. gaussiennes indépendantes centrées réduites (voir section 7.2). En utilisant une nouvelle fois le fait que la matrice  $\mathbf{U}$  est orthogonale, on a  $\sum_{i=1}^n Z_i^2 = \sum_{j=1}^n Y_j^2$ , et donc :

$$(n-1) \frac{V_n}{\sigma^2} = \sum_{j=2}^n Y_j^2,$$

où la somme va bien de 2 à  $n$ . La loi de cette v.a. réelle est donc un  $\chi_{n-1}^2$ . Cela montre aussi que les v.a. réelles  $\bar{X}_n$  et  $V_n$  sont indépendantes, car la première ne dépend que de  $Y_1$ , alors que la seconde ne dépend que de  $(Y_j)_{j=2, \dots, n}$ . Enfin, comme les variables aléatoires  $(n-1)\sigma^{-2}V_n \sim \chi_{n-1}^2$  et  $\sqrt{n}\sigma^{-1}(\bar{X}_n - \mu) \sim \mathcal{N}(0, 1)$  sont indépendantes, la v.a. réelle :

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{V_n}} = \sqrt{n-1} \frac{\sqrt{n}\sigma^{-1}(\bar{X}_n - \mu)}{\sqrt{n-1}\sigma^{-1}\sqrt{V_n}}$$

suit une loi  $T_{n-1}$ . □

**Corollaire 9.5** Soit  $(X_i)_{i=1,\dots,n}$  un  $n$ -échantillon de loi  $\mathcal{N}(\mu, \sigma^2)$ . Les risques quadratiques moyens des estimateurs empiriques de la moyenne et de la variance sont :

$$\text{RQM}(\bar{X}_n) = \frac{\sigma^2}{n}, \quad \text{RQM}(V_n) = \frac{2\sigma^4}{n-1}.$$

**Preuve.** Le risque quadratique moyen de la moyenne empirique est déjà connu par la proposition 8.18. Celui de la variance empirique se calcule à partir de la caractérisation (9.4) et du fait qu'une v.a. suivant la loi  $\chi_p^2$  a pour variance  $2p$ .  $\square$

### 9.2.1 Estimation par intervalle de confiance de la moyenne

On observe un  $n$ -échantillon d'une loi gaussienne  $\mathcal{N}(\mu, \sigma^2)$ , et on cherche à estimer la moyenne  $\mu$ . On peut se servir de la moyenne empirique  $\bar{X}_n$ , dont on sait qu'elle va donner une estimation bonne lorsque  $n$  est grand. En pratique, on cherche un intervalle de confiance, c'est-à-dire un intervalle dont on puisse dire qu'il contient la valeur inconnue  $\mu$  avec une probabilité qu'on se fixe.

Supposons dans un premier temps que l'écart-type  $\sigma$  de la loi soit connu. Soit un niveau de confiance  $1 - \alpha$  donné (en général  $\alpha = 0,1, 0,05$ , ou  $0,01$ ). D'après la proposition 9.3, le quantile  $q_{1-\alpha/2}$  d'ordre  $1 - \alpha/2$  de la loi gaussienne centrée réduite est tel que si  $Z \sim \mathcal{N}(0, 1)$ , alors  $\mathbb{P}(-q_{1-\alpha/2} \leq Z \leq q_{1-\alpha/2}) = 1 - \alpha$ . On cherche alors  $q_{1-\alpha/2}$  dans une table de la loi normale ou sur un logiciel. D'après la proposition 9.4 :

$$\mathbb{P}\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \in [-q_{1-\alpha/2}, q_{1-\alpha/2}]\right) = 1 - \alpha.$$

En réécrivant l'événement en question :

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \in [-q_{1-\alpha/2}, q_{1-\alpha/2}] \iff \mu \in \left[\bar{X}_n - \frac{\sigma q_{1-\alpha/2}}{\sqrt{n}}, \bar{X}_n + \frac{\sigma q_{1-\alpha/2}}{\sqrt{n}}\right],$$

on obtient l'intervalle de confiance  $I_{1-\alpha}$  pour  $\mu$  au niveau  $1 - \alpha$  :

$$I_{1-\alpha} = \left[\bar{X}_n - \frac{\sigma q_{1-\alpha/2}}{\sqrt{n}}, \bar{X}_n + \frac{\sigma q_{1-\alpha/2}}{\sqrt{n}}\right].$$

Bien sûr,

- plus on se fixe un niveau de confiance  $1 - \alpha$  élevé, plus l'intervalle est large. Si on demande d'être absolument sûr, i.e.  $\alpha = 0$ , alors  $q_1 = +\infty$  et  $I_1 = \mathbb{R}$ .
- plus la v.a. sous-jacente a une dispersion élevée (i.e. plus  $\sigma$  est grand), plus l'intervalle est grand.
- plus l'échantillon est grand, plus l'intervalle est petit. La largeur de l'intervalle de

confiance est proportionnelle à  $1/\sqrt{n}$ . Donc, à niveau  $1-\alpha$  fixé, pour avoir un intervalle de confiance 2 fois plus petit, il faut 4 fois plus d'observations.

Les calculs précédents sont intéressants, mais en pratique, il est rare qu'on connaisse  $\sigma$ . Dans le cas général où on ne connaît pas  $\sigma$ , on utilise une estimation de celui-ci, donnée par la variance empirique sans biais  $V_n$ .

Soit un niveau de confiance  $1-\alpha$  donné. D'après la proposition 9.3, le quantile  $t_{1-\alpha/2}$  d'ordre  $1-\alpha/2$  de la loi  $T_{n-1}$  est tel que, si  $Z \sim T_{n-1}$ , alors  $\mathbb{P}(-t_{1-\alpha/2} \leq Z \leq t_{1-\alpha/2}) = 1-\alpha$ . D'après la proposition 9.4 :

$$\mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{V_n}} \in [-t_{1-\alpha/2}, t_{1-\alpha/2}]\right) = 1-\alpha.$$

En réécrivant l'événement en question, on obtient l'intervalle de confiance  $I_{1-\alpha}$  pour  $\mu$  au niveau  $1-\alpha$  :

$$I_{1-\alpha} = \left[ \bar{X}_n - \frac{\sqrt{V_n}t_{1-\alpha/2}}{\sqrt{n}}, \bar{X}_n + \frac{\sqrt{V_n}t_{1-\alpha/2}}{\sqrt{n}} \right].$$

L'intervalle de confiance obtenu ainsi a la même comportement qualitatif que celui décrit dans le cas  $\sigma$  connu. Il n'y a rien d'étonnant à cela, car on sait que la variance empirique converge avec probabilité 1 vers la variance théorique. Ceci se traduit aussi par le fait que, lorsque  $n$  est grand, la loi de Student à  $n$  degrés de libertés  $T_n$  devient très proche de la loi gaussienne centrée réduite (voir exercice 8.2). En pratique, on utilise des tables de loi  $T_n$  (voir la table 9.1) pour  $n \leq 30$ , et, lorsque  $n > 30$ , on utilise souvent l'approximation gaussienne. On peut aussi utiliser des logiciels de calcul scientifique.

**Exemple 9.6** *On mesure les durées de vie (en heures) de  $n = 10$  ampoules. On obtient :*

1864, 1934, 2033, 1890, 1997, 1974, 1837, 1903, 2009, 1950.

*On cherche la durée de vie moyenne  $\mu$  d'une ampoule, en supposant que la loi décrivant cette durée de vie est une gaussienne. La moyenne empirique et la variance empirique non-biaisée sont :*

$$\bar{X}_{10} \simeq 1939, \quad V_{10} \simeq 4244.$$

*Un intervalle de confiance pour la moyenne au niveau 95% est donc :*

$$I_{0,95} = \left[ \bar{X}_{10} - \frac{\sqrt{V_{10}}t_{0,975}}{\sqrt{10}}, \bar{X}_{10} + \frac{\sqrt{V_{10}}t_{0,975}}{\sqrt{10}} \right],$$

*où  $t_{0,975}$  est tel que  $\mathbb{P}(Z \leq t_{0,975}) = 0,975$ , avec  $Z \sim T_9$ . En utilisant une table de la loi  $T_9$ , on trouve  $t_{0,975} \simeq 2,26$ , et on obtient ainsi l'intervalle de confiance [1892, 1986] pour la valeur inconnue  $\mu$  au niveau 0,95.*

$n \backslash p$	0,40	0,25	0,10	0,05	0,025	0,01	0,005	0,0005
1	0,324920	1,000000	3,077684	6,313752	12,70620	31,82052	63,65674	636,6192
2	0,288675	0,816497	1,885618	2,919986	4,30265	6,96456	9,92484	31,5991
3	0,276671	0,764892	1,637744	2,353363	3,18245	4,54070	5,84091	12,9240
4	0,270722	0,740697	1,533206	2,131847	2,77645	3,74695	4,60409	8,6103
5	0,267181	0,726687	1,475884	2,015048	2,57058	3,36493	4,03214	6,8688
6	0,264835	0,717558	1,439756	1,943180	2,44691	3,14267	3,70743	5,9588
7	0,263167	0,711142	1,414924	1,894579	2,36462	2,99795	3,49948	5,4079
8	0,261921	0,706387	1,396815	1,859548	2,30600	2,89646	3,35539	5,0413
9	0,260955	0,702722	1,383029	1,833113	2,26216	2,82144	3,24984	4,7809
10	0,260185	0,699812	1,372184	1,812461	2,22814	2,76377	3,16927	4,5869
11	0,259556	0,697445	1,363430	1,795885	2,20099	2,71808	3,10581	4,4370
12	0,259033	0,695483	1,356217	1,782288	2,17881	2,68100	3,05454	4,3178
13	0,258591	0,693829	1,350171	1,770933	2,16037	2,65031	3,01228	4,2208
14	0,258213	0,692417	1,345030	1,761310	2,14479	2,62449	2,97684	4,1405
15	0,257885	0,691197	1,340606	1,753050	2,13145	2,60248	2,94671	4,0728
16	0,257599	0,690132	1,336757	1,745884	2,11991	2,58349	2,92078	4,0150
17	0,257347	0,689195	1,333379	1,739607	2,10982	2,56693	2,89823	3,9651
18	0,257123	0,688364	1,330391	1,734064	2,10092	2,55238	2,87844	3,9216
19	0,256923	0,687621	1,327728	1,729133	2,09302	2,53948	2,86093	3,8834
20	0,256743	0,686954	1,325341	1,724718	2,08596	2,52798	2,84534	3,8495
21	0,256580	0,686352	1,323188	1,720743	2,07961	2,51765	2,83136	3,8193
22	0,256432	0,685805	1,321237	1,717144	2,07387	2,50832	2,81876	3,7921
23	0,256297	0,685306	1,319460	1,713872	2,06866	2,49987	2,80734	3,7676
24	0,256173	0,684850	1,317836	1,710882	2,06390	2,49216	2,79694	3,7454
25	0,256060	0,684430	1,316345	1,708141	2,05954	2,48511	2,78744	3,7251
26	0,255955	0,684043	1,314972	1,705618	2,05553	2,47863	2,77871	3,7066
27	0,255858	0,683685	1,313703	1,703288	2,05183	2,47266	2,77068	3,6896
28	0,255768	0,683353	1,312527	1,701131	2,04841	2,46714	2,76326	3,6739
29	0,255684	0,683044	1,311434	1,699127	2,04523	2,46202	2,75639	3,6594
30	0,255605	0,682756	1,310415	1,697261	2,04227	2,45726	2,75000	3,6460
$+\infty$	0,253347	0,674490	1,281552	1,644854	1,95996	2,32635	2,57583	3,2905

TABLE 9.1 – Table de la loi  $T_n$ . On reporte dans ce tableau les valeurs  $x$  pour lesquelles  $\mathbb{P}(T_n \geq x) = p$ . Pour  $n > 30$ , on adopte souvent l'approximation gaussienne  $T_n \sim \mathcal{N}(0, 1)$ .

### 9.2.2 Estimation par intervalle de confiance de la variance

Les raisonnements qu'on vient d'appliquer pour l'estimation de la moyenne peuvent être repris pour l'estimation de la variance  $\sigma^2$ , lorsque la moyenne  $\mu$  est connue, puis lorsqu'elle ne l'est pas.

Commençons par le cas où la moyenne  $\mu$  de l'échantillon est connu. Donnons-nous le niveau de confiance  $1 - \alpha$ . On commence par choisir  $t_{1-\alpha,l}$  et  $t_{1-\alpha,r}$  tels que, si  $Z \sim \chi_n^2$ , alors  $\mathbb{P}(t_{1-\alpha,l} \leq Z \leq t_{1-\alpha,r}) = 1 - \alpha$ . La loi  $\chi_n^2$  n'est pas symétrique et il y a une infinité de couples possibles  $(t_{1-\alpha,l}, t_{1-\alpha,r})$ . On choisit en général les deux extrémités  $t_{1-\alpha,l}$  et  $t_{1-\alpha,r}$  de sorte que le risque soit également réparti à gauche et à droite, i.e.  $\mathbb{P}(Z < t_{1-\alpha,l}) = \alpha/2$  et  $\mathbb{P}(Z > t_{1-\alpha,r}) = \alpha/2$ . Autrement dit,  $t_{1-\alpha,l}$  est le quantile d'ordre  $\alpha/2$  de la loi  $\chi_n^2$  et  $t_{1-\alpha,r}$  est le quantile d'ordre  $1 - \alpha/2$  de la loi  $\chi_n^2$ . Mais on pourrait parfaitement proposer un intervalle asymétrique. On utilise en pratique des tables de la loi  $\chi_n^2$  (voir la table 9.2) ou des logiciels de calcul scientifique. Lorsque la moyenne  $\mu$  est connue, l'estimateur empirique de la variance est :

$$V_n^* = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

$V_n^*$  est sans biais, convergent, et  $nV_n^*/\sigma^2$  suit la loi  $\chi_n^2$  d'après la proposition 9.4. On peut donc affirmer que :

$$\mathbb{P}\left(\frac{nV_n^*}{\sigma^2} \in [t_{1-\alpha,l}, t_{1-\alpha,r}]\right) = 1 - \alpha.$$

En réécrivant l'événement en question, on trouve un intervalle de confiance au niveau  $1 - \alpha$  de la variance :

$$I_{1-\alpha} = \left[ \frac{nV_n^*}{t_{1-\alpha,r}}, \frac{nV_n^*}{t_{1-\alpha,l}} \right].$$

**Exemple 9.7** Dans l'exemple 9.6, imaginons que le fabricant ait fait des mesures extensives, et qu'il indique sur la boîte la durée de vie moyenne :  $\mu = 1920$ . On recherche la variance  $\sigma^2$  inconnue de la loi de durée de vie. L'estimateur empirique de la variance est :

$$V_{10}^* = \frac{1}{10} \sum_{i=1}^{10} (X_i - 1920)^2 \simeq 4184.$$

On cherche un intervalle de confiance au niveau  $1 - \alpha = 95\%$  de la variance à partir de l'échantillon de 10 ampoules observées. Cet intervalle est :

$$I_{0,95} = \left[ \frac{10V_{10}^*}{t_{0,95,r}}, \frac{10V_{10}^*}{t_{0,95,l}} \right].$$

$n \backslash p$	0,995	0,990	0,975	0,950	0,050	0,025	0,010	0,005
1	0,00004	0,00016	0,00098	0,00393	3,84146	5,02389	6,63490	7,87944
2	0,01003	0,02010	0,05064	0,10259	5,99146	7,37776	9,21034	10,59663
3	0,07172	0,11483	0,21580	0,35185	7,81473	9,34840	11,34487	12,83816
4	0,20699	0,29711	0,48442	0,71072	9,48773	11,14329	13,27670	14,86026
5	0,41174	0,55430	0,83121	1,14548	11,07050	12,83250	15,08627	16,74960
6	0,67573	0,87209	1,23734	1,63538	12,59159	14,44938	16,81189	18,54758
7	0,98926	1,23904	1,68987	2,16735	14,06714	16,01276	18,47531	20,27774
8	1,34441	1,64650	2,17973	2,73264	15,50731	17,53455	20,09024	21,95495
9	1,73493	2,08790	2,70039	3,32511	16,91898	19,02277	21,66599	23,58935
10	2,15586	2,55821	3,24697	3,94030	18,30704	20,48318	23,20925	25,18818
11	2,60322	3,05348	3,81575	4,57481	19,67514	21,92005	24,72497	26,75685
12	3,07382	3,57057	4,40379	5,22603	21,02607	23,33666	26,21697	28,29952
13	3,56503	4,10692	5,00875	5,89186	22,36203	24,73560	27,68825	29,81947
14	4,07467	4,66043	5,62873	6,57063	23,68479	26,11895	29,14124	31,31935
15	4,60092	5,22935	6,26214	7,26094	24,99579	27,48839	30,57791	32,80132
16	5,14221	5,81221	6,90766	7,96165	26,29623	28,84535	31,99993	34,26719
17	5,69722	6,40776	7,56419	8,67176	27,58711	30,19101	33,40866	35,71847
18	6,26480	7,01491	8,23075	9,39046	28,86930	31,52638	34,80531	37,15645
19	6,84397	7,63273	8,90652	10,11701	30,14353	32,85233	36,19087	38,58226
20	7,43384	8,26040	9,59078	10,85081	31,41043	34,16961	37,56623	39,99685
21	8,03365	8,89720	10,28290	11,59131	32,67057	35,47888	38,93217	41,40106
22	8,64272	9,54249	10,98232	12,33801	33,92444	36,78071	40,28936	42,79565
23	9,26042	10,19572	11,68855	13,09051	35,17246	38,07563	41,63840	44,18128
24	9,88623	10,85636	12,40115	13,84843	36,41503	39,36408	42,97982	45,55851
25	10,51965	11,52398	13,11972	14,61141	37,65248	40,64647	44,31410	46,92789
26	11,16024	12,19815	13,84390	15,37916	38,88514	41,92317	45,64168	48,28988
27	11,80759	12,87850	14,57338	16,15140	40,11327	43,19451	46,96294	49,64492
28	12,46134	13,56471	15,30786	16,92788	41,33714	44,46079	48,27824	50,99338
29	13,12115	14,25645	16,04707	17,70837	42,55697	45,72229	49,58788	52,33562
30	13,78672	14,95346	16,79077	18,49266	43,77297	46,97924	50,89218	53,67196

TABLE 9.2 – Table de la loi  $\chi_n^2$ . On reporte dans ce tableau les valeurs  $x$  pour lesquelles  $\mathbb{P}(\chi_n^2 \geq x) = p$ . Pour  $n > 30$ , on adopte souvent l'approximation gaussienne  $\sqrt{2}\chi_n^2 - \sqrt{2n-1} \sim \mathcal{N}(0, 1)$ .



Il faut évaluer les deux nombres  $t_{0,95,l}$  et  $t_{0,95,r}$ . Pour cela, on consulte une table de la fonction de répartition de la loi  $\chi_{10}^2$  et on cherche les niveaux  $t_{0,95,l}$  et  $t_{0,95,r}$  tels que, si  $Z \sim \chi_{10}^2$ , alors  $\mathbb{P}(Z < t_{0,95,l}) = 0,025$  et  $\mathbb{P}(Z < t_{0,95,r}) = 0,975$ . On trouve  $t_{0,95,l} = 3,25$  et  $t_{0,95,r} = 20,48$ . On obtient alors que  $[2043, 12875]$  est un intervalle de confiance pour  $\sigma^2$  au niveau 0,95. Noter que l'intervalle de confiance pour la variance est beaucoup plus large que pour l'espérance. Il est en effet plus difficile d'estimer la variance que la moyenne.

Dans le cas d'un échantillon de taille  $n$  supérieure à 30, on peut admettre que, si  $Z \sim \chi_n^2$ , alors la v.a.  $\sqrt{2}\sqrt{Z} - \sqrt{2n-1}$  suit la loi gaussienne centrée réduite. Pour des échantillons de taille  $n$  supérieure à 30, on utilise cette approximation gaussienne pour déterminer les quantiles de la loi  $\chi_n^2$  (voir l'exercice 8.2). C'est pourquoi les tables de la loi  $\chi_n^2$  s'arrêtent en général à  $n = 30$ . Bien sûr, avec un logiciel de mathématiques ou de statistique, on peut trouver des valeurs approchées des quantiles de la loi  $\chi_n^2$  valables pour des  $n$  plus grands, avec une précision arbitraire.

Supposons maintenant que la moyenne  $\mu$  soit inconnue. On commence par choisir  $t_{1-\alpha,l}$  et  $t_{1-\alpha,r}$  tels que, si  $Z \sim \chi_{n-1}^2$ , alors  $\mathbb{P}(t_{1-\alpha,l} \leq Z \leq t_{1-\alpha,r}) = 1 - \alpha$ . Lorsque la moyenne  $\mu$  est inconnue, l'estimateur empirique non-biaisé de la variance est :

$$V_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

$V_n$  est sans biais, convergent, et  $(n-1)V_n/\sigma^2$  suit la loi  $\chi_{n-1}^2$  d'après la proposition 9.4. Un intervalle de confiance au niveau  $1 - \alpha$  de la variance est donc :

$$I_{1-\alpha} = \left[ \frac{(n-1)V_n}{t_{1-\alpha,r}}, \frac{(n-1)V_n}{t_{1-\alpha,l}} \right].$$

**Exemple 9.8** On reprend l'exemple 9.7, mais sans supposer qu'on connaît l'espérance  $\mu$ . L'estimateur empirique non-biaisé de la variance est alors :

$$V_{10} = \frac{1}{9} \sum_{i=1}^{10} (X_i - \bar{X}_{10})^2 \simeq 4244.$$

Un intervalle de confiance au niveau  $1 - \alpha = 95\%$  de la variance est :

$$I_{0,95} = \left[ \frac{9V_{10}}{t_{0,95,r}}, \frac{9V_{10}}{t_{0,95,l}} \right].$$

On commence par consulter une table de la fonction de répartition de la loi  $\chi_9^2$  et on cherche les niveaux  $t_{0,95,l}$  et  $t_{0,95,r}$  tels que, si  $Z \sim \chi_9^2$ , alors  $\mathbb{P}(Z < t_{0,95,l}) = 0,025$  et  $\mathbb{P}(Z < t_{0,95,r}) = 0,975$ . On trouve  $t_{0,95,l} = 2,70$  et  $t_{0,95,r} = 19,02$ . On obtient alors que  $[2008, 14147]$  est un intervalle de confiance pour  $\sigma^2$  au niveau 0,95. Comme on pouvait s'y attendre, l'intervalle est un peu plus grand que dans le cas où on connaît l'espérance, car il y a plus d'incertitude puisque l'on ne connaît pas la moyenne.

### 9.3 Résultats asymptotiques

Dans le cas non-gaussien, la proposition 9.4 n'est plus vraie, mais en vertu du théorème de la limite centrale, on peut obtenir des résultats similaires sous forme asymptotique. On obtient les distributions de la moyenne et de la variance empirique “pour  $n$  assez grand”. Dans la pratique, on admettra ces résultats lorsque  $n \geq 30$ . Des résultats fins (utilisant le théorème de Berry-Essen ou des inégalités de concentration) permettent de contrôler l'erreur commise, mais ils sortent du cadre de ce cours. Nous nous contenterons de donner une application particulièrement importante de cette méthode pour les intervalles de confiance pour une proportion, c'est-à-dire les sondages.

#### 9.3.1 Intervalles de confiance asymptotiques

On cherche un intervalle de confiance pour un paramètre  $\theta$  réel. On suppose que  $\hat{\theta}_n$  est un estimateur convergent et asymptotiquement normal de  $\theta$ , plus exactement qu'il existe une fonction  $\sigma(\theta)$  de  $\Theta$  dans  $]0, +\infty[$  telle que  $\sqrt{n}(\hat{\theta}_n - \theta)/\sigma(\theta)$  converge en loi vers une loi gaussienne centrée réduite. Comme on l'a vu, c'est un cas assez courant, qui arrive dès qu'on est dans le régime du théorème de la limite centrale. On suppose aussi que la fonction  $\sigma(\theta)$  est continue. La suite de v.a. réelles  $\sigma(\hat{\theta}_n)$  converge alors presque sûrement vers  $\sigma(\theta)$ , et par le théorème de Slutsky,  $\sqrt{n}(\hat{\theta}_n - \theta)/\sigma(\hat{\theta}_n)$  converge en loi vers une loi gaussienne centrée réduite. Par conséquent, pour  $n$  pas trop petit, on peut approcher les probabilités suivantes,

$$\mathbb{P}_\theta\left(\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sigma(\hat{\theta}_n)} \leq u\right), \quad \mathbb{P}_\theta\left(\sqrt{n}\left|\frac{\hat{\theta}_n - \theta}{\sigma(\hat{\theta}_n)}\right| \leq u\right), \quad \mathbb{P}_\theta\left(\sqrt{n}\frac{\hat{\theta}_n - \theta}{\sigma(\hat{\theta}_n)} \geq -u\right),$$

avec  $u > 0$ , par  $\Phi(u)$ ,  $\Phi(u) - \Phi(-u) = 2\Phi(u) - 1$ , et  $1 - \Phi(-u) = \Phi(u)$ , respectivement, où  $\Phi$  est la fonction de répartition de la loi gaussienne centrée réduite. La fonction réciproque  $\Phi^{-1}$  de  $\Phi$  est la fonction quantile de la loi gaussienne (voir la table 5.1). Les quantiles gaussiens suivantes sont particulièrement utiles :

$$\Phi^{-1}(0,975) = 1,96, \quad \Phi^{-1}(0,995) = 2,58. \quad (9.7)$$

Définissons les intervalles aléatoires

$$\begin{aligned} I_{n,1} &= \left[ \hat{\theta}_n - \Phi^{-1}(1 - \alpha) \frac{\sigma(\hat{\theta}_n)}{\sqrt{n}}, +\infty \right], \\ I_{n,2} &= \left[ \hat{\theta}_n - \Phi^{-1}(1 - \alpha/2) \frac{\sigma(\hat{\theta}_n)}{\sqrt{n}}, \hat{\theta}_n + \Phi^{-1}(1 - \alpha/2) \frac{\sigma(\hat{\theta}_n)}{\sqrt{n}} \right], \\ I_{n,3} &= \left] -\infty, \hat{\theta}_n + \Phi^{-1}(1 - \alpha) \frac{\sigma(\hat{\theta}_n)}{\sqrt{n}} \right], \end{aligned}$$

avec une intersection à effectuer éventuellement avec le domaine  $\Theta$  des valeurs de  $\theta$ . On a alors :

$$\mathbb{P}_\theta(\theta \in I_{n,j}) \simeq 1 - \alpha,$$

pour  $j = 1, 2, 3$  et pour  $n$  assez grand.

**Définition 9.9** On appelle intervalle de confiance asymptotique au niveau  $1 - \alpha$  de  $\theta$  un des intervalles aléatoires  $I_{n,j}$ ,  $j = 1, 2, 3$ .

On dit que  $I_{n,2}$  est un intervalle de confiance bilatère, tandis que  $I_{n,1}$  et  $I_{n,3}$  sont des intervalles de confiance unilatères.

Signalons que la méthode de construction des intervalles de confiance asymptotiques peut être étendue sans difficulté au cas où la vitesse de convergence n'est pas  $\sqrt{n}$  et la loi limite n'est pas gaussienne, comme par exemple pour l'EMV du modèle uniforme vu dans l'exemple 8.27.

### 9.3.2 Sondages

A la veille du second tour d'une élection présidentielle, on effectue un sondage afin de déterminer la proportion  $\theta \in [0, 1]$  de votes pour le candidat Monsieur C. On pourrait considérer plus généralement tout problème de sondage avec réponses binaires. Le sondage porte sur  $n = 2500$  individus choisis au hasard dans le corps électoral (excluant les abstentionnistes). On note  $X_i = 1$  si le  $i^{\text{ème}}$  individu interrogé vote pour C,  $X_i = 0$  sinon. En pratique, on évite d'interroger deux fois un même individu (tirage sans remise), de sorte que les  $X_i$  sont dépendants. Mais nous avons vu dans la section 2.2.2 que lorsque la taille  $n$  de l'échantillon est faible comparée à la population totale, il y a peu de différence entre les tirages avec remise et sans remise. Nous nous placerons donc dans cette hypothèse, et supposerons ainsi les  $X_i$  indépendantes.

Les v.a.  $X_i$ ,  $i = 1, \dots, n$ , sont supposées indépendantes et de même loi de Bernoulli  $\mathbb{P}(X_i = 1) = \theta = 1 - \mathbb{P}(X_i = 0)$ . Nous nous trouvons dans le problème statistique suivant : comment estimer le paramètre  $\theta$  inconnu au vu des observations  $X_1, \dots, X_n$  ? Intuitivement, aucune information sur  $\theta$  n'est contenue dans l'ordre des réponses, et il suffit de résumer le sondage  $X_1, \dots, X_n$  par le nombre  $S_n = X_1 + \dots + X_n$  d'intentions de vote pour C. Ce raisonnement heuristique peut se quantifier par le résultat qui suit, qui dit que la loi conditionnelle de  $\mathbf{X}$  sachant  $S_n = k$  est la probabilité uniforme sur les  $\binom{n}{k}$  suites  $(x_1, \dots, x_n)$  comportant  $k$  uns et  $n - k$  zéros.

**Lemme 9.10** Soit  $k \in \{0, \dots, n\}$ . Pour tout  $\mathbf{x} = (x_1, \dots, x_n) \in \{0, 1\}^n$ , on a :

$$\mathbb{P}(\mathbf{X} = \mathbf{x} | S_n = k) = \begin{cases} \frac{1}{\binom{n}{k}} & \text{si } \mathbf{x} \in E_{n,k}, \\ 0 & \text{sinon,} \end{cases}$$

où  $E_{n,k}$  représente l'ensemble des suites  $(x_1, \dots, x_n)$  comportant  $k$  uns et  $n - k$  zéros.

**Preuve.** Soit  $\mathbf{x} = (x_1, \dots, x_n) \in \{0, 1\}^n$ . Par définition d'une probabilité conditionnelle :

$$\mathbb{P}(\mathbf{X} = \mathbf{x} | S_n = k) = \frac{\mathbb{P}(\{\mathbf{X} = \mathbf{x}\} \cap \{S_n = k\})}{\mathbb{P}(S_n = k)}.$$

Comme  $S_n$  suit une loi binomiale de paramètres  $(n, \theta)$  :

$$\mathbb{P}(S_n = k) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}.$$

Si  $\mathbf{x} \notin E_{n,k}$ , alors la somme des  $x_i$  ne vaut pas  $k$ , et donc :

$$\mathbb{P}(\{\mathbf{X} = \mathbf{x}\} \cap \{S_n = k\}) = 0.$$

Si  $\mathbf{x} \in E_{n,k}$ , alors la somme des  $x_i$  vaut  $k$ , et :

$$\mathbb{P}(\{\mathbf{X} = \mathbf{x}\} \cap \{S_n = k\}) = \mathbb{P}(\mathbf{X} = \mathbf{x}) = \theta^k (1 - \theta)^{n-k}.$$

En injectant dans la probabilité conditionnelle, on obtient le résultat cherché.  $\square$

Le lemme précédent montre que la loi conditionnelle des observations  $(X_1, \dots, X_n)$  sachant  $S_n = k$  ne dépend pas du paramètre inconnu  $\theta$ . Par conséquent la valeur de  $S_n$  contient toute l'information sur  $\theta$  contenue dans le sondage. On cherche alors à estimer  $\theta$  par une fonction  $\hat{\theta}_n$  de  $S_n$ , et le choix naturel est la proportion  $\hat{\theta}_n$  de vote pour C dans l'échantillon :

$$\hat{\theta}_n = \frac{S_n}{n}.$$

Cet estimateur est sans biais  $\mathbb{E}_\theta(\hat{\theta}_n) = \theta$  (quel que soit  $\theta$ ), c'est-à-dire qu'il est "bon" en moyenne. Cet estimateur est convergent, c'est-à-dire qu'il est "bon" lorsque l'échantillon est de grande taille. Le sondage donne 1300 intentions de votes pour C, et 1200 pour son adversaire. L'estimateur  $\hat{\theta}_n$  prend la valeur 0,52 mais est-on "sûr" pour autant que C sera élu ? Plus précisément, cette valeur est-elle significativement supérieure à 0,5 ? La réponse dépend de l'amplitude des fluctuations de  $\hat{\theta}_n$  autour de  $\theta$ , et nous utilisons alors l'approximation gaussienne (théorème de la limite centrale) :

$$\mathbb{P}\left(|\hat{\theta}_n - \theta| \leq a\sqrt{\theta(1-\theta)}\right) \simeq \Phi(a) - \Phi(-a) = 2\Phi(a) - 1,$$

où  $\Phi$  est la fonction de répartition de la loi gaussienne centrée réduite. D'après (9.7), l'erreur commise  $|\hat{\theta}_n - \theta|$  ne dépassera pas, quasi certainement (= avec 95%, resp. 99% de probabilité environ), le seuil :

$$\frac{1,96\sqrt{\theta(1-\theta)}}{\sqrt{n}}, \quad \text{resp.} \quad \frac{2,58\sqrt{\theta(1-\theta)}}{\sqrt{n}}.$$

Ce seuil dépend malencontreusement du paramètre  $\theta$  inconnu, mais la fonction  $\sqrt{\theta(1-\theta)}$  est majorée par sa valeur maximale 1/2, de sorte qu'en remplaçant le facteur  $\sqrt{\theta(1-\theta)}$  par 1/2 dans les seuils précédents, on ne fera qu'augmenter notre quasi-certitude. En conclusion, réenonçons le résultat précédent sous la forme sous laquelle il est généralement utilisé.

**Proposition 9.11 (Intervalle de confiance pour l'estimation de  $\theta$ )** *Dès que  $n$  est assez grand ( $n\theta$  et  $n(1-\theta) \geq 10$  en pratique) :*

$$\theta \in \left[ \hat{\theta}_n - \frac{0,98 \text{ (resp. 1,29)}}{\sqrt{n}}, \hat{\theta}_n + \frac{0,98 \text{ (resp. 1,29)}}{\sqrt{n}} \right], \quad (9.8)$$

avec la quasi-certitude de 95% (resp. 99%).

Dans notre exemple, l'intervalle de confiance à 95% pour  $\theta$  est  $[0,50, 0,54]$ . Les instituts de sondage annoncent d'ailleurs leurs résultats ainsi (sous forme de fourchette).

### 9.3.3 Calcul d'intégrale par la méthode de Monte-Carlo

Soit  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction intégrable. Lorsque le calcul analytique de  $I = \int_{\mathbb{R}^d} f(\mathbf{x}) d\mathbf{x}$  n'est pas possible, il existe diverses méthodes d'intégration numérique. Nous décrivons ici une méthode probabiliste par simulation. Soit  $p$  une densité de probabilité sur  $\mathbb{R}^d$ , dont le support contient celui de  $f$  (i.e., si  $\mathbf{x}$  est tel que  $f(\mathbf{x}) \neq 0$  alors  $p(\mathbf{x}) > 0$ ). On a :

$$I = \int_{\mathbb{R}^d} f(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^d} \frac{f(\mathbf{x})}{p(\mathbf{x})} p(\mathbf{x}) d\mathbf{x},$$

et  $I$  s'écrit comme :

$$I = \mathbb{E}(\psi(\mathbf{X})) \text{ avec } \psi = f/p, \mathbf{X} \text{ v.a. de densité } p.$$

Si  $p$  est une densité facilement simulable, on sait générer des v.a.  $\mathbf{X}_1, \dots, \mathbf{X}_n$  indépendantes et de densité  $p$ . La quantité :

$$\hat{I}_n = \frac{1}{n} [\psi(\mathbf{X}_1) + \dots + \psi(\mathbf{X}_n)]$$

constitue une approximation, en fait, une estimation au sens statistique du terme, de  $I$ . Cette méthode de calcul (approximatif) est appelée méthode de Monte-Carlo. Le nom de la méthode fait référence aux jeux de hasard pratiqués à Monte-Carlo.

L'estimateur  $\hat{I}_n$  est en fait la moyenne empirique de la v.a. réelle  $\psi(\mathbf{X})$ , qui a pour espérance  $I$ . La proposition 8.18 donne les propriétés importantes de cet estimateur : il est sans-biais (par linéarité de la moyenne), convergent (par application de la loi des grands nombres), son risque quadratique moyen est  $\text{RQM}(\hat{I}_n) = \sigma^2/n$ , où  $\sigma^2 = \text{Var}(\psi(\mathbf{X}))$ , qui est fini lorsque  $\mathbb{E}(\psi(\mathbf{X})^2) < +\infty$ . Cette variance est donnée par :

$$\begin{aligned} \sigma^2 &= \mathbb{E}(\psi(\mathbf{X})^2) - \mathbb{E}(\psi(\mathbf{X}))^2 = \int_{\mathbb{R}^d} \psi(\mathbf{x})^2 p(\mathbf{x}) d\mathbf{x} - I^2 \\ &= \int_{\mathbb{R}^d} \frac{f(\mathbf{x})^2}{p(\mathbf{x})} d\mathbf{x} - \left( \int_{\mathbb{R}^d} f(\mathbf{x}) d\mathbf{x} \right)^2. \end{aligned} \quad (9.9)$$

Un point remarquable de la méthode est que la variance  $\sigma^2$ , et donc le risque quadratique moyen, dépend explicitement du choix de la densité  $p$ . Comme il y a une infinité de choix possibles pour la densité  $p$ , tout l'art du praticien est de bien choisir la densité pour "réduire la variance". Les techniques de réduction de variance pour les méthodes de Monte Carlo font l'objet d'intenses recherches.

Le théorème de la limite centrale nous renseigne, lorsque  $\mathbb{E}(\psi(\mathbf{X})^2) < +\infty$ , sur la distribution de l'erreur  $\hat{I}_n - I$ . En procédant comme dans le paragraphe précédent, on obtient un intervalle de confiance pour  $I$  à la quasi-certitude 95% de la forme :

$$\left[ \hat{I}_n - \frac{1,96\sigma}{\sqrt{n}}, \hat{I}_n + \frac{1,96\sigma}{\sqrt{n}} \right].$$

Ici, la variance  $\sigma^2$  définie par (9.9) est la plus souvent inconnue, de sorte qu'il faut estimer sa valeur elle aussi. On définit alors :

$$\hat{\sigma}_n = \left( \frac{1}{n} \sum_{i=1}^n \psi(\mathbf{X}_i)^2 - \hat{I}_n^2 \right)^{1/2},$$

qui vérifie  $\hat{\sigma}_n \rightarrow \sigma$  quand  $n \rightarrow +\infty$  d'après la loi des grands nombres, et on utilise alors l'intervalle de confiance asymptotique :

$$\left[ \hat{I}_n - \frac{1,96\hat{\sigma}_n}{\sqrt{n}}, \hat{I}_n + \frac{1,96\hat{\sigma}_n}{\sqrt{n}} \right]. \quad (9.10)$$

L'intervalle (9.10) indique l'erreur de la méthode de Monte-Carlo. Comparée aux méthodes d'intégration numériques usuelles, la méthode de Monte-Carlo est surtout intéressante en dimension  $d$  grande.

## 9.4 Exercices sur le chapitre 9

**EXERCICE 9.1** Soit  $(X_1, \dots, X_n)$  un  $n$ -échantillon de variables aléatoires i.i.d. suivant la loi exponentielle de paramètre  $\theta \in ]0, +\infty[$ .

1. Soit  $S_n = X_1 + \dots + X_n$ . Quelle est la loi de  $S_n$  ?
2. En déduire un intervalle de confiance au niveau  $1 - \alpha \in (0, 1)$  pour  $\theta$ . A l'aide des quantiles des lois de  $\chi^2$  donnés dans la Table 9.2, préciser la mise en œuvre de cet intervalle de confiance pour  $n = 10$  et  $\alpha = 5\%$ .

**EXERCICE 9.2** On veut évaluer la proportion  $p$  de foyers d'un pays disposant d'un poste de télévision et désireux de recevoir les émissions par câble. Ne voulant pas procéder à un recensement complet de la population, on se propose d'estimer cette proportion à partir d'un échantillon de taille  $n$  prélevé au hasard dans la population du pays. On définit une variable aléatoire  $\bar{X}_n$ , dont la réalisation est  $\bar{x}_n$ , fréquence observée dans l'échantillon des ménages concernés par la télévision câblée.

1. Préciser l'espérance et la variance de  $\bar{X}_n$ .
2. Justifier que  $\bar{X}_n$  converge en un sens à préciser vers  $p$ .
3. Soit  $n = 100$  et  $\bar{x}_n = 0,64$ . Déterminer un intervalle de confiance pour  $p$  au niveau 0,9 en utilisant la borne supérieure du produit  $p(1-p)$ . En déduire une fourchette d'estimation pour  $p$  au niveau de confiance 0,9.

**EXERCICE 9.3** La loi de Pareto de paramètre de forme  $\alpha > 0$  et de paramètre d'échelle  $\beta > 0$  est donnée par sa densité  $p(x, (\alpha, \beta)) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}} \mathbf{1}_{[\beta, +\infty[}(x)$ . On observe un  $n$ -échantillon  $(X_1, \dots, X_n)$  suivant cette loi.

1. Déterminer l'Estimateur du Maximum de Vraisemblance du couple  $(\alpha, \beta)$ .
2. Le paramètre d'échelle  $\beta$  est maintenant supposé connu égal à 1. Vérifier que si  $X$  suit la loi de Pareto de paramètres  $\alpha$  et 1,  $\ln(X)$  suit la loi exponentielle de paramètre  $\alpha$ . En déduire un estimateur de  $\alpha$ . Montrer qu'il est convergent et construire un intervalle de confiance asymptotique de niveau  $1 - \eta$  pour  $\alpha$ .

**EXERCICE 9.4** On s'intéresse à la durée de vie de deux composants électroniques se trouvant sur un système solidaire. Si l'un des deux composants tombe en panne, le système tout entier doit être changé. Les durées de vie de ces deux composants sont modélisées par des variables aléatoires exponentielles de paramètres  $\lambda$  et  $\mu$  indépendantes. Formellement, on considère un  $n$ -échantillon  $(X_1, \dots, X_n)$  de loi  $\mathcal{E}(\lambda)$  et un  $n$ -échantillon  $(Y_1, \dots, Y_n)$ , indépendant du précédent, de loi  $\mathcal{E}(\mu)$ , où  $\lambda > 0$  et

$\mu > 0$ . On observe seulement la durée de vie du composant qui est tombé en panne, ce qui veut dire qu'on observe seulement le  $n$ -échantillon  $((Z_1, W_1), \dots, (Z_n, W_n))$  où  $Z_i = \min(X_i, Y_i)$  et  $W_i = 1$  si  $Z_i = X_i$  et 0 si  $Z_i = Y_i$ .

1. Donner les lois de  $Z_i$  et  $W_i$ .
2. Montrer que les variables  $Z_i$  et  $W_i$  sont indépendantes.

Les variables aléatoires  $Z_i$  et  $W_i$  étant indépendantes, la vraisemblance du  $n$ -échantillon  $((Z_1, W_1), \dots, (Z_n, W_n))$  est définie comme étant le produit de la vraisemblance du  $n$ -échantillon  $(Z_1, \dots, Z_n)$  par la vraisemblance du  $n$ -échantillon  $(W_1, \dots, W_n)$ . Dans les questions 3 à 7, on suppose que la loi de la durée de vie du second composant est connue, i.e. que  $\mu$  est connu.

3. Donner l'Estimateur du Maximum de Vraisemblance  $\hat{\lambda}_n$  de  $\lambda$ .
4. Montrer que  $\hat{\lambda}_n$  est convergent.
5. Calculer  $\mathbb{E}_\lambda(\hat{\lambda}_n)$  et en déduire que  $\mathbb{E}_\lambda(\hat{\lambda}_n)$  tend vers  $\lambda$  lorsque  $n$  tend vers l'infini.
6. Vérifier que  $\hat{\lambda}_n$  est asymptotiquement normal et identifier sa variance asymptotique.
7. Donner un intervalle de confiance bilatéral symétrique asymptotique de niveau  $1 - \alpha$  pour  $\lambda$ .
8. Donner l'Estimateur du Maximum de Vraisemblance  $(\hat{\lambda}_n, \hat{\mu}_n)$  de  $(\lambda, \mu)$ .



## Chapitre 10

# Statistique : Test

*The true logic for this world is the calculus of Probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind.*

J. Clerk Maxwell

L'objectif d'un test d'hypothèse est de répondre à une question qu'on peut poser de la manière suivante : au vu de l'observation d'un  $n$ -échantillon, le paramètre  $\theta$  du modèle est-il ou non dans un sous-ensemble de  $\Theta$  appelé hypothèse nulle et noté  $H_0$  ? On retrouve cette situation fréquemment :

- Lors d'un sondage d'intention de vote sur un échantillon de 1000 personnes, on trouve que 520 personnes déclarent vouloir voter pour  $A$  et 480 pour  $B$ . Il est naturel d'annoncer que  $A$  sera élu. Mais est-ce que la marge d'erreur est suffisante pour pouvoir faire une telle annonce sans (trop de) risque de se tromper ? En fait, on est en train d'estimer le paramètre d'une loi de Bernoulli (qui représente l'intention de vote pour le candidat  $A$ ) et de tester l'hypothèse : le paramètre est-il plus grand que  $1/2$  ?

- Si on s'intéresse au changement climatique, on peut par exemple travailler sur les données de température moyenne au mois d'août à Paris. Sur l'ensemble du vingtième siècle, ces températures moyennes en degrés Celsius sont distribuées suivant une loi gaussienne d'espérance 20 et de variance 1,4. Sur les quinze dernières années, on a

observé les températures moyennes suivantes (données fictives) :

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$x_{13}$	$x_{14}$	$x_{15}$
22	19	21	20	18	22	21	18	20	25	21	19	23	20	22

telles que la moyenne empirique est  $\bar{x}_{15} = 20,73$  et la variance empirique non-biaisée est  $v_{15} = 1,91^2$ . On voit bien que la moyenne empirique sur les quinze dernières années dépasse 20, ce qui indiquerait plutôt qu'il y a un réchauffement, mais on ne peut pas en être absolument sûr. A partir des observations, on souhaite construire un test d'hypothèse qui permette de décider s'il y a un réchauffement ou pas (i.e., si la température moyenne a effectivement augmenté ces quinze dernières années par rapport à l'ensemble du vingtième siècle, ou pas). On comprend bien ici que quelle que soit la décision qu'on prenne, on peut se tromper. On peut se tromper en annonçant qu'il y a un réchauffement, alors qu'il n'y en a pas. On peut se tromper en annonçant qu'il n'y a pas de réchauffement, alors qu'il y en a.

## 10.1 Tests et erreurs

On considère  $\mathbf{X} = (X_1, \dots, X_n)$  un  $n$ -échantillon du modèle statistique  $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ . Soit  $(H_0, H_1)$  une partition de l'ensemble  $\Theta$  des paramètres.

On appelle test d'hypothèse une règle de décision qui, au vu de l'observation  $\mathbf{X}$ , permet de décider si  $\theta$  est dans l'ensemble  $H_0$  appelé hypothèse nulle ou si  $\theta$  est dans l'ensemble  $H_1$  appelé hypothèse alternative.

Un test est déterminé par sa région critique  $W$  qui constitue un sous-ensemble (mesurable) de l'ensemble  $\mathcal{X}^n$  des valeurs possibles de  $\mathbf{X}$ . La règle de décision du test associé à  $W$  est la suivante. Lorsqu'on observe  $\mathbf{x} = (x_1, \dots, x_n)$ ,

- si  $\mathbf{x} \in W$ , alors on rejette  $H_0$  et on accepte  $H_1$  i.e. on décide que  $\theta \in H_1$ ,
- si  $\mathbf{x} \notin W$ , alors on accepte  $H_0$  et on rejette  $H_1$  i.e. on décide que  $\theta \in H_0$ .

On appelle erreur de première espèce le rejet de  $H_0$  à tort. C'est la situation où l'échantillon suivait la loi  $\mathbb{P}_\theta$  pour un certain  $\theta \in H_0$ , mais on a annoncé qu'on rejetait  $H_0$ . Cette erreur est mesurée par le risque de première espèce, qui est la fonction :  $\theta \in H_0 \mapsto \mathbb{P}_\theta(\mathbf{X} \in W)$ .

On appelle erreur de seconde espèce le rejet de  $H_1$  à tort. Cette erreur est mesurée par le risque de seconde espèce :  $\theta \in H_1 \mapsto \mathbb{P}_\theta(\mathbf{X} \in W^c) = 1 - \mathbb{P}_\theta(\mathbf{X} \in W)$ . La fonction  $\theta \in H_1 \mapsto \mathbb{P}_\theta(\mathbf{X} \in W)$  s'appelle puissance du test.

**Exemple 10.1** On considère modèle  $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2), \mu \in \{\mu_0, \mu_1\}\}$  avec  $\sigma^2 > 0$  connu et  $\mu_0 > \mu_1$ , on souhaite tester  $H_0 = \{\mu = \mu_0\}$  contre  $H_1 = \{\mu = \mu_1\}$ . On va

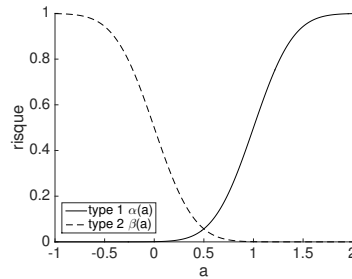


FIGURE 10.1 – Risques de première et deuxième espèce pour l'exemple 10.1 en fonction du seuil  $a$  avec  $\mu_0 = 1$ ,  $\mu_1 = 0$ ,  $\sigma = 0$ , et  $n = 10$ .

bien sûr accepter  $H_0$  (resp.  $H_1$ ) si la moyenne empirique  $\bar{X}_n$  est grande (resp. petite), c'est-à-dire choisir la région critique de la forme  $W = \{\bar{X}_n < a\}$  pour un certain  $a$ . En utilisant le fait que sous  $H_0$ , la moyenne empirique  $\bar{X}_n$  suit la loi  $\mathcal{N}(\mu_0, \sigma^2/n)$ , le risque de première espèce est :

$$\alpha(a) = \mathbb{P}_{(\mu_0, \sigma^2)}(W) = \mathbb{P}_{(\mu_0, \sigma^2)}(\bar{X}_n < a) = \Phi\left(\sqrt{n}\frac{a - \mu_0}{\sigma}\right),$$

avec  $\Phi$  la fonction de répartition de la loi gaussienne centrée réduite. Le risque de seconde espèce est :

$$\beta(a) = \mathbb{P}_{(\mu_1, \sigma^2)}(W^c) = \mathbb{P}_{(\mu_1, \sigma^2)}(\bar{X}_n \geq a) = 1 - \Phi\left(\sqrt{n}\frac{a - \mu_1}{\sigma}\right) = \Phi\left(\sqrt{n}\frac{\mu_1 - a}{\sigma}\right).$$

Il est clair que  $a \rightarrow \alpha(a)$  est croissante alors que  $a \rightarrow \beta(a)$  est décroissante. On dessine sur la figure 10.1 les deux erreurs en fonction du seuil  $a$  choisi.

Idéalement, on voudrait minimiser les risques de première et de deuxième espèce. Mais ceci n'est pas possible en même temps, comme le montre l'exemple précédent. Par convention, on minimise en priorité le risque de première espèce.

**Définition 10.2** Le niveau d'un test défini par sa région critique  $W$  est le nombre

$$\alpha = \sup_{\theta \in H_0} \mathbb{P}_{\theta}(W).$$

Si un test est de niveau  $\alpha$ , alors on sait que, lorsqu'on rejette  $H_0$ , on a au plus une probabilité  $\alpha$  de se tromper. Parmi tous les tests de niveau inférieur à un seuil  $\alpha$  fixé, on souhaite minimiser le risque de seconde espèce ou de manière équivalente maximiser la puissance. En général, on choisit  $\alpha = 10\%$ ,  $\alpha = 5\%$ , ou  $\alpha = 1\%$ .

Comme on décide de minimiser en priorité le risque de première espèce, les rôles de l'hypothèse nulle  $H_0$  et de l'hypothèse alternative  $H_1$  ne sont pas symétriques. Le choix de  $H_0$  parmi deux ensembles constituant une partition de  $\Theta$  dépend donc du problème considéré : on choisit comme hypothèse nulle l'ensemble qu'on ne souhaite surtout pas voir rejeté à tort. Mais c'est une question de point de vue : dans le cadre d'un test sur le réchauffement climatique, un politicien peut vouloir mettre comme hypothèse nulle l'absence d'un réchauffement, car il ne veut surtout pas s'engager dans un protocole contraignant si le réchauffement n'est pas avéré. Un écologiste peut vouloir mettre comme hypothèse nulle l'existence d'un réchauffement, car il estime que les effets d'un réchauffement seraient catastrophiques.

**Définition 10.3** Soit  $(W_n)_n$  une suite de régions critiques où  $n$  désigne la taille de l'échantillon. La suite de tests basée sur  $(W_n)_n$  est dite

- convergente si, pour tout  $\theta \in H_1$ ,

$$\lim_{n \rightarrow +\infty} \mathbb{P}_\theta(W_n) = 1.$$

- de niveau asymptotique  $\alpha$  si

$$\lim_{n \rightarrow +\infty} \sup_{\theta \in H_0} \mathbb{P}_\theta(W_n) = \alpha.$$

Souvent on utilise un estimateur aux bonnes propriétés connues pour construire le test et guider le choix de la région critique.

**Exemple 10.4** On reprend l'exemple 10.1. La région critique est de la forme  $W_n = \{\bar{X}_n < a\}$ . Le choix  $a = (\mu_0 + \mu_1)/2$ , qui peut sembler naturel, ne permet pas de contrôler le risque de première espèce. Pour obtenir ce contrôle de la probabilité de rejeter  $H_0$  à tort, on utilise le fait que sous  $H_0$ , la moyenne empirique  $\bar{X}_n$  suit la loi  $\mathcal{N}(\mu_0, \sigma^2/n)$ . Donc on a

$$\mathbb{P}_{(\mu_0, \sigma^2)}(W_n) = \mathbb{P}_{(\mu_0, \sigma^2)}(\bar{X}_n < a) = \Phi\left(\sqrt{n} \frac{a - \mu_0}{\sigma}\right).$$

En désignant par  $\Phi^{-1}(r)$  le quantile d'ordre  $r$  de la loi gaussienne centrée réduite, le choix  $a = \mu_0 + \sigma \Phi^{-1}(\alpha)/\sqrt{n}$  assure que le niveau du test est  $\alpha$ . Pour ce choix, la probabilité de rejeter  $H_1$  à tort est

$$\mathbb{P}_{(\mu_1, \sigma^2)}(W_n^c) = \mathbb{P}_{(\mu_1, \sigma^2)}(\bar{X}_n \geq a) = \Phi\left(\sqrt{n} \frac{\mu_1 - \mu_0}{\sigma} - \Phi^{-1}(\alpha)\right),$$

qui vérifie

$$\mathbb{P}_{(\mu_1, \sigma^2)}(W_n^c) \leq \Phi(-\Phi^{-1}(\alpha)) = 1 - \Phi(\Phi^{-1}(\alpha)) = 1 - \alpha.$$

De plus  $\mathbb{P}_{(\mu_1, \sigma^2)}(W_n^c) = \mathbb{E}(\mathbf{1}_{[\Phi^{-1}(\alpha) + \sqrt{n}(\mu_0 - \mu_1)/\sigma, +\infty[}(Z))$  (pour  $Z$  une v.a. de loi gaussienne centrée réduite) converge vers zéro lorsque  $n$  tend vers l'infini d'après le théorème de convergence dominée, ce qui montre que le test est convergent.

## 10.2 Modèle gaussien

On considère le modèle gaussien  $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 \in ]0, +\infty[ \}$ . On va proposer des tests pour la moyenne et pour la variance.

### 10.2.1 Tests pour la moyenne $\mu$

Soit  $\mu_0 \in \mathbb{R}$ . On souhaite tester  $H_0 = \{\mu = \mu_0\}$  contre  $H_1 = \{\mu \neq \mu_0\}$  au niveau  $\alpha \in ]0, 1[$ .

D'après la proposition 9.4, si  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  et  $V_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  désignent respectivement la moyenne empirique et l'estimateur empirique non-biaisé de la variance, le rapport

$$\zeta_n = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sqrt{V_n}}$$

suit la loi de Student  $T_{n-1}$  sous  $H_0$  (i.e. si  $\mu = \mu_0$ ).

Sous  $H_1$ , par la loi forte des grands nombres,  $\bar{X}_n - \mu_0$  converge presque sûrement vers  $\mu - \mu_0 \neq 0$  et  $V_n$  converge presque sûrement vers  $\sigma^2 > 0$ . Donc  $\zeta_n$  tend presque sûrement vers  $+\infty$  ou  $-\infty$  lorsque  $n \rightarrow +\infty$ , suivant que  $\mu > \mu_0$  ou  $\mu < \mu_0$ .

On choisit donc  $W_n = \{|\zeta_n| > a\}$  pour la région critique. On note  $t_r(n-1)$  le quantile d'ordre  $r$  de la loi de Student  $T_{n-1}$ . Si  $a \geq t_{1-\alpha/2}(n-1)$ , alors pour tout  $\sigma^2 > 0$  :

$$\mathbb{P}_{(\mu_0, \sigma^2)}(W_n) = \mathbb{P}_{(\mu_0, \sigma^2)}(|\zeta_n| > a) \leq \mathbb{P}_{(\mu_0, \sigma^2)}(|\zeta_n| > t_{1-\alpha/2}(n-1)) = \alpha,$$

ce qui montre que le niveau du test est inférieur à  $\alpha$ . Comme on souhaite ensuite minimiser le risque de seconde espèce  $\mathbb{P}_{(\mu, \sigma^2)}(W_n)$  pour  $\mu \neq \mu_0$ , on choisit au plus juste  $a = t_{1-\alpha/2}(n-1)$ . En conclusion, on choisit la région critique  $W_n = \{|\zeta_n| > t_{1-\alpha/2}(n-1)\}$  et on a alors  $\mathbb{P}_{(\mu_0, \sigma^2)}(W_n) = \alpha$  pour tout  $\sigma^2 > 0$ . Notons que lorsque  $n \rightarrow +\infty$ ,  $t_{1-\alpha/2}(n-1)$  converge vers  $\Phi^{-1}(1 - \alpha/2)$ , le quantile d'ordre  $1 - \alpha/2$  de la loi gaussienne centrée réduite. Donc, par le théorème de convergence dominée,  $\mathbb{P}_{(\mu, \sigma^2)}(W_n) = \mathbb{E}_{(\mu, \sigma^2)}(\mathbf{1}_{t_{1-\alpha/2}(n-1), \infty}(|\zeta_n|))$  tend vers 1 lorsque  $n$  tend vers l'infini. Ainsi le test est convergent.

On peut reprendre le raisonnement précédent pour construire un test de niveau  $\alpha$  pour  $H_0 = \{\mu \leq \mu_0\}$  et  $H_1 = \{\mu > \mu_0\}$ . La région critique est alors  $\{\zeta_n \geq t_{1-\alpha}(n-1)\}$ .

**Exemple 10.5** *On reprend l'exemple des données de température moyenne au mois d'août à Paris.*

1) *On souhaite tester l'hypothèse nulle  $H_0 = \{\mu \leq 20\}$  (absence de réchauffement climatique) contre  $H_1 = \{\mu > 20\}$  (existence d'un réchauffement climatique). C'est le point de vue d'un politicien comme on le décrivait ci-dessus. La région critique est*

$\{\zeta_n \geq t_{1-\alpha}(n-1)\}$  avec  $\zeta_n = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sqrt{V_n}}$ ,  $\mu_0 = 20$  et  $n = 15$ . On observe  $\zeta_{15}^{obs} = \sqrt{15}(20,73 - 20)/1,91 = 1,48$ . Comme  $t_{0,95}(14) = 1,76$ , on accepte  $H_0$  au niveau  $\alpha = 5\%$ . On peut donc conclure qu'il n'y a pas de réchauffement climatique.

2) On souhaite tester l'hypothèse nulle  $H_0 = \{\mu > 20\}$  (existence d'un réchauffement climatique) contre  $H_1 = \{\mu \leq 20\}$  (absence de réchauffement climatique). C'est le point de vue d'un écologiste comme on le décrivait ci-dessus. La région critique est  $\{\zeta_n \leq t_\alpha(n-1)\}$  avec  $\zeta_n = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sqrt{V_n}}$ ,  $\mu_0 = 20$  et  $n = 15$ . On observe  $\zeta_{15}^{obs} = \sqrt{15}(20,73 - 20)/1,91 = 1,48$ . Comme  $t_{0,05}(14) = -1,76$ , on accepte  $H_0$  au niveau  $\alpha = 5\%$ . On peut donc conclure qu'il y a réchauffement climatique.

3) Supposons maintenant que les données de température moyenne au mois d'août à Paris aient la forme :

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$	$x_{13}$	$x_{14}$	$x_{15}$
22	19	21	20	20	22	24	18	20	25	21	19	24	20	23

telles que la moyenne empirique est  $\bar{x}_{15} = 21,20$  et la variance empirique non-biaisée est  $v_{15} = 2,08^2$ . On prend pour hypothèse nulle  $H_0 = \{\mu \leq 20\}$  (absence de réchauffement climatique) contre  $H_1 = \{\mu > 20\}$  (existence d'un réchauffement climatique). On observe alors  $\zeta_{15}^{obs} = \sqrt{15}(21,20 - 20)/2,08 = 2,25$ . Comme  $t_{0,95}(14) = 1,76$ , on rejette  $H_0$  au niveau  $\alpha = 5\%$ . Ainsi on peut conclure à l'augmentation des températures sur les quinze dernières années. Mais si on s'impose le niveau  $\alpha = 1\%$ , alors on accepte  $H_0$ , car  $t_{0,99}(14) = 2,62$ .

Comme le montre l'exemple précédent :

- Quand les données n'apportent que peu d'information, on va toujours accepter l'hypothèse nulle, afin d'éviter de commettre une erreur de première espèce (rejeter à tort l'hypothèse nulle).

- Quand les données sont informatives, mais qu'on impose un niveau  $\alpha$  très proche de 0, on va aussi toujours accepter l'hypothèse nulle, car c'est la seule manière d'être quasi-certain de ne pas commettre une erreur de première espèce.

Tout ceci montre bien que le choix de l'hypothèse nulle est fondamental.

## 10.2.2 Tests pour la variance $\sigma^2$

Soit  $\sigma_0^2 > 0$ . On souhaite tester  $H_0 = \{\sigma^2 \geq \sigma_0^2\}$  contre  $H_1 = \{\sigma^2 < \sigma_0^2\}$ . On introduit

$$\zeta_n = \frac{(n-1)V_n}{\sigma_0^2}.$$

Comme d'après la proposition 9.4,  $(n-1)V_n/\sigma^2$  suit la loi  $\chi_{n-1}^2$  sous  $\mathbb{P}_{(\mu,\sigma^2)}$ ,  $\zeta_n$  prend des valeurs de plus en plus grandes lorsque  $\sigma^2$  croît. En particulier  $\zeta_n$  va avoir tendance à prendre des valeurs plus grandes sous  $H_0$  que sous  $H_1$ . C'est pourquoi on acceptera  $H_0$  si  $\zeta_n$  est grand et on rejettera  $H_0$  si  $\zeta_n$  est petit. On choisit donc une région critique de la forme  $W_n = \{\zeta_n \leq a\}$ . En outre, si  $Z \sim \chi_{n-1}^2$ , alors

$$\begin{aligned} \sup_{(\mu,\sigma^2) \in H_0} \mathbb{P}_{(\mu,\sigma^2)}(\zeta_n \leq a) &= \sup_{\mu \in \mathbb{R}, \sigma^2 \geq \sigma_0^2} \mathbb{P}_{(\mu,\sigma^2)}\left(\frac{(n-1)V_n}{\sigma^2} \leq a \frac{\sigma_0^2}{\sigma^2}\right) \\ &= \sup_{\sigma^2 \geq \sigma_0^2} \mathbb{P}\left(Z \leq a \frac{\sigma_0^2}{\sigma^2}\right) = \mathbb{P}(Z \leq a). \end{aligned}$$

Le choix  $a = x_\alpha(n-1)$  où  $x_r(n-1)$  désigne le quantile d'ordre  $r$  de la loi  $\chi_{n-1}^2$  assure que le niveau du test est  $\alpha$ .

## 10.3 Test du $\chi^2$ (test du chi-deux)

Le test du  $\chi^2$  permet de répondre à des questions telles que “Un dé à six faces est-il pipé ?” Pour cela on observe les fréquences d'apparition des faces lors de  $n$  lancers de ce dé et on les compare au vecteur  $(1/6, \dots, 1/6)$ . Si on constate qu'on s'éloigne significativement de ce vecteur, on peut rejeter l'hypothèse que le dé est équilibré et annoncer que le dé est pipé. La question est de savoir ce qu'on entend par “significativement”.

### 10.3.1 Test d'adéquation à une loi

On observe un  $n$ -échantillon  $(X_1, \dots, X_n)$  de variables aléatoires indépendantes et identiquement distribuées à valeurs dans un espace fini  $\mathcal{X} = \{a_1, \dots, a_k\}$ . La loi est paramétrée par  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  avec  $\mathbb{P}_{\boldsymbol{\theta}}(X_1 = a_j) = \theta_j$  pour  $j \in \{1, \dots, k\}$ . Le paramètre  $\boldsymbol{\theta}$  vit dans l'ensemble  $\Theta = \{\boldsymbol{\theta} \in [0, 1]^k, \sum_{i=1}^k \theta_i = 1\}$ . Pour  $\boldsymbol{\theta}^{(0)} \in \Theta$  fixé, on souhaite tester l'hypothèse nulle  $H_0 = \{\boldsymbol{\theta}^{(0)}\}$  contre l'hypothèse alternative  $H_1 = \Theta \setminus \{\boldsymbol{\theta}^{(0)}\}$ .

**Exemple 10.6** Dans le cas du dé à six faces évoqué plus haut,  $\mathcal{X} = \{1, \dots, 6\}$  et  $\boldsymbol{\theta}^{(0)} = (1/6, \dots, 1/6)$ .

On peut déterminer l'EMV de  $\boldsymbol{\theta}$  et ses propriétés.

**Proposition 10.7** *L'EMV de  $\boldsymbol{\theta}$  est donné par le vecteur des fréquences empiriques d'apparition des différentes valeurs possibles :*

$$\hat{\boldsymbol{\theta}}_n = \left( \frac{N_1(\mathbf{X})}{n}, \dots, \frac{N_k(\mathbf{X})}{n} \right), \quad N_i(\mathbf{x}) = \sum_{j=1}^n \mathbf{1}_{a_i}(x_j), \quad (10.1)$$

avec la notation  $\mathbf{1}_a(x) = 1$  si  $x = a$  et 0 sinon. L'EMV est non-biaisé, convergent, asymptotiquement normal : Sous  $\mathbb{P}_{\boldsymbol{\theta}}$ ,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta})), \quad (10.2)$$

en loi, avec

$$C(\boldsymbol{\theta})_{jj'} = \begin{cases} \theta_j - \theta_j^2 & \text{si } j = j', \\ -\theta_j \theta_{j'} & \text{si } j \neq j'. \end{cases} \quad (10.3)$$

**Preuve.** La vraisemblance et la log-vraisemblance sont égales à :

$$\begin{aligned} p_n(\mathbf{x}, \boldsymbol{\theta}) &= \prod_{j=1}^n \left( \prod_{i=1}^k \theta_i^{\mathbf{1}_{a_i}(x_j)} \right), \\ l_n(\mathbf{x}, \boldsymbol{\theta}) &= \sum_{i=1}^k N_i(\mathbf{x}) \ln(\theta_i). \end{aligned} \quad (10.4)$$

La log-vraisemblance peut aussi écrire :

$$l_n(\mathbf{x}, \boldsymbol{\theta}) = n \sum_{i \in K(\mathbf{x})} \rho_i(\mathbf{x}) \ln(\theta_i), \quad (10.5)$$

où  $K(\mathbf{x}) = \{i = 1, \dots, k : N_i(\mathbf{x}) > 0\}$  et on note

$$\boldsymbol{\rho}(\mathbf{x}) = \left( \frac{N_1(\mathbf{x})}{n}, \dots, \frac{N_k(\mathbf{x})}{n} \right). \quad (10.6)$$

Si il existe  $i \in K(\mathbf{x})$  tel que  $\theta_i = 0$ , alors  $l_n(\mathbf{x}, \boldsymbol{\theta}) = -\infty$ . Si  $\theta_i > 0$  pour tout  $i \in K(\mathbf{x})$ , comme  $\ln(y) = y - 1 + \int_1^y (1/z - 1) dz \leq y - 1$  pour tout  $y > 0$  (avec inégalité stricte dès que  $y \neq 1$ ), on en déduit que :

$$\begin{aligned} l_n(\mathbf{x}, \boldsymbol{\theta}) - l_n(\mathbf{x}, \boldsymbol{\rho}(\mathbf{x})) &= n \sum_{i \in K(\mathbf{x})} \rho_i(\mathbf{x}) \ln \left( \frac{\theta_i}{\rho_i(\mathbf{x})} \right) \\ &\leq n \sum_{i \in K(\mathbf{x})} \rho_i(\mathbf{x}) \left( \frac{\theta_i}{\rho_i(\mathbf{x})} - 1 \right) = n \sum_{i \in K(\mathbf{x})} \theta_i - n \sum_{i \in K(\mathbf{x})} \rho_i(\mathbf{x}) \\ &= n \sum_{i \in K(\mathbf{x})} \theta_i - n \leq 0. \end{aligned}$$



La seconde inégalité est stricte dès que  $\boldsymbol{\theta} \neq \boldsymbol{\rho}(\mathbf{x})$ . Ceci montre que  $\boldsymbol{\rho}(\mathbf{x})$  maximise la vraisemblance et que l'estimateur du maximum de vraisemblance de  $\boldsymbol{\theta}$  est (10.1).

L'EMV  $\hat{\boldsymbol{\theta}}_n$  est non-biaisé car :

$$\mathbb{E}_{\boldsymbol{\theta}}(N_j(\mathbf{X})) = n\mathbb{E}_{\boldsymbol{\theta}}(\mathbf{1}_{a_j}(X_1)) = n\mathbb{P}_{\boldsymbol{\theta}}(X_1 = a_j) = n\theta_j.$$

On trouve aussi que l'EMV est convergent en appliquant la loi forte des grands nombres.

En utilisant le fait que

$$\text{Cov}_{\boldsymbol{\theta}}(\mathbf{1}_{a_j}(X_1), \mathbf{1}_{a_{j'}}(X_1)) = \mathbb{E}_{\boldsymbol{\theta}}(\mathbf{1}_{a_j}(X_1)\mathbf{1}_{a_{j'}}(X_1)) - \mathbb{E}_{\boldsymbol{\theta}}(\mathbf{1}_{a_j}(X_1))\mathbb{E}_{\boldsymbol{\theta}}(\mathbf{1}_{a_{j'}}(X_1))$$

qui est égal à  $C(\boldsymbol{\theta})_{jj'}$  défini par (10.3), on montre par le théorème de la limite centrale vectoriel que l'EMV est asymptotiquement normal avec  $\mathbf{C}(\boldsymbol{\theta})$  pour matrice de covariance asymptotique.  $\square$

Ce résultat va nous permettre de construire un test. On suppose dorénavant que les coefficients de  $\boldsymbol{\theta}^{(0)}$  sont tous non-nuls. L'idée qui est à la base du test est que le vecteur  $\hat{\boldsymbol{\theta}}_n$  est censé être plus proche de  $\boldsymbol{\theta}^{(0)}$  sous  $H_0$  que sous  $H_1$ . Afin de quantifier la "proximité", on utilise la pseudo-distance du  $\chi^2$  :

$$\zeta_n = n \sum_{j=1}^k \frac{(\hat{\theta}_{n,j} - \theta_j^{(0)})^2}{\theta_j^{(0)}}. \quad (10.7)$$

On obtient le comportement asymptotique suivant :

**Proposition 10.8** Soit  $\zeta_n$  défini par (10.7).

1) Sous  $H_0$ ,  $\zeta_n$  converge en loi quand  $n \rightarrow +\infty$  vers une variable aléatoire  $Z$  qui suit une loi de  $\chi^2$  à  $k - 1$  degrés de liberté.

2) Sous  $H_1$ ,  $\zeta_n$  tend presque sûrement vers  $+\infty$ .

**Preuve.** Sous  $H_1$ , les observations sont indépendantes et identiquement distribuées de loi  $\mathbb{P}_{\boldsymbol{\theta}}$  pour un certain  $\boldsymbol{\theta} \neq \boldsymbol{\theta}^{(0)}$ . Donc il existe  $j \in \{1, \dots, k\}$  tel que  $\theta_j \neq \theta_j^{(0)}$ . Par la loi forte des grands nombres,  $\mathbb{P}_{\boldsymbol{\theta}}$ -presque sûrement,  $\hat{\theta}_{n,j} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{a_j}(X_i)$  converge vers  $\theta_j$ , et donc  $n \frac{(\hat{\theta}_{n,j} - \theta_j^{(0)})^2}{\theta_j^{(0)}}$  tend vers  $+\infty$ .

Sous  $H_0$ , les observations sont indépendantes et identiquement distribuées de loi  $\mathbb{P}_{\boldsymbol{\theta}^{(0)}}$ . D'après (10.2), la suite de vecteurs aléatoires

$$\mathbf{Y}_n = \left( \frac{\hat{\theta}_{n,1} - \theta_1^{(0)}}{\sqrt{\theta_1^{(0)}}}, \dots, \frac{\hat{\theta}_{n,k} - \theta_k^{(0)}}{\sqrt{\theta_k^{(0)}}} \right)$$

converge en loi vers un vecteur aléatoire  $\mathbf{Y}$  de loi  $\mathcal{N}(\mathbf{0}, \mathbf{C}^{(0)})$  avec  $C_{jj}^{(0)} = 1 - \theta_j^{(0)}$  et  $C_{jj'}^{(0)} = -\sqrt{\theta_j^{(0)}\theta_{j'}^{(0)}}$  si  $j \neq j'$ . Autrement dit,  $\mathbf{C}^{(0)} = \mathbf{I} - \mathbf{e}_1\mathbf{e}_1^t$  où  $\mathbf{e}_1$  est le vecteur unitaire de coordonnées  $e_{1,j} = \sqrt{\theta_j^{(0)}}$ . Par continuité de l'application  $\mathbf{y} \in \mathbb{R}^k \mapsto \|\mathbf{y}\|^2 \in \mathbb{R}^+$ ,  $\zeta_n = \|\mathbf{Y}_n\|^2$  converge en loi vers  $\|\mathbf{Y}\|^2$ . Tout le travail consiste maintenant à identifier la loi de cette limite. Donnons-nous une base  $(\mathbf{e}_1, \dots, \mathbf{e}_k)$  de  $\mathbb{R}^k$  dont le premier vecteur est  $\mathbf{e}_1$  décrit précédemment. Appelons  $\mathbf{U}$  la matrice  $k \times k$  orthogonale de vecteurs lignes donnés par les  $\mathbf{e}_j$ , et  $\mathbf{Z}$  le vecteur aléatoire donné par  $\mathbf{Z} = \mathbf{U}\mathbf{Y}$ . D'une part, on a  $\|\mathbf{Y}\| = \|\mathbf{Z}\|$ . D'autre part,  $\mathbf{Z}$  est un vecteur gaussien de moyenne nulle et de matrice de covariance  $\mathbf{U}\mathbf{C}^{(0)}\mathbf{U}^t = \mathbf{I} - (\mathbf{U}\mathbf{e}_1)(\mathbf{U}\mathbf{e}_1)^t$ . Or  $\mathbf{U}\mathbf{e}_1 = \mathbf{e}_1$  et donc  $\mathbf{U}\mathbf{C}^{(0)}\mathbf{U}^t$  est la matrice diagonale de coefficients diagonaux  $(0, 1, \dots, 1)$ . Ceci montre que  $Z_1 = 0$  p.s. et que les  $Z_i, i = 2, \dots, k$  sont indépendantes et identiquement distribuées selon la loi gaussienne centrée réduite. Par conséquent la loi limite de  $\zeta_n$  est la loi  $\chi_{k-1}^2$ .  $\square$

En notant  $x_r(k-1)$  le quantile d'ordre  $r$  de la loi  $\chi^2$  à  $k-1$  degrés de liberté, on a  $\mathbb{P}(Z \geq x_{1-\alpha}(k-1)) = \alpha$ . On en déduit le corollaire suivant :

**Corollaire 10.9** *Le test de région critique  $W_n = \{\zeta_n > x_{1-\alpha}(k-1)\}$  est convergent de niveau asymptotique  $\alpha$ .*

**Preuve.** Sous  $H_1$ , on déduit de la proposition 10.8 que  $\mathbf{1}_{[x_{1-\alpha}(k-1), +\infty[}(\zeta_n)$  converge  $\mathbb{P}_{\theta}$ -presque sûrement vers 1. Le théorème de convergence dominée entraîne alors que

$$\mathbb{P}_{\theta}(\zeta_n \geq x_{1-\alpha}(k-1)) = \mathbb{E}_{\theta}(\mathbf{1}_{[x_{1-\alpha}(k-1), +\infty[}(\zeta_n)) \xrightarrow{n \rightarrow +\infty} \mathbb{E}_{\theta}(1) = 1,$$

ce qui assure que le test est convergent. Pour vérifier que son niveau asymptotique est  $\alpha$ , il suffit de vérifier que  $\mathbb{P}_{\theta^{(0)}}(\zeta_n \geq x_{1-\alpha}(k-1))$  converge vers  $\mathbb{P}(Z \geq x_{1-\alpha}(k-1)) = \alpha$  lorsque  $n$  tend vers l'infini, où  $Z$  est de loi  $\chi_{k-1}^2$ . Or, c'est vrai d'après le corollaire 7.23 car  $Z$  est à densité.  $\square$

En pratique, on considère que l'approximation en loi par  $\chi_{k-1}^2$  est valide sous  $H_0$  si  $n \min_{j=1, \dots, k} \theta_j^{(0)} \geq 5$ . Si cette condition n'est pas satisfaite, on peut regrouper les valeurs de  $a_j$  pour lesquelles  $\theta_j^{(0)}$  est trop faible et augmenter ainsi le minimum (on appliquera cette procédure dans l'exemple 10.14).

Le principe du test du  $\chi^2$  se généralise à des lois arbitraires, pas nécessairement à valeurs dans un espace fini. Il suffit de considérer une partition finie  $(A_i)_{i=1, \dots, k}$  de  $\mathcal{X}$  telle que  $\mathbb{P}_{\theta^{(0)}}(X_1 \in A_j) = \tilde{\theta}_j^{(0)}$  avec  $n \min_{j=1, \dots, k} \tilde{\theta}_j^{(0)} \geq 5$ . On peut alors regrouper les observations sous la forme :

$$\hat{\theta}_{n,i} = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{A_i}(X_j), \quad i = 1, \dots, k.$$

$\tilde{\zeta}_n$  défini par :

$$\tilde{\zeta}_n = n \sum_{j=1}^k \frac{(\hat{\theta}_{n,j} - \tilde{\theta}_j^{(0)})^2}{\tilde{\theta}_j^{(0)}}$$

satisfait la proposition 10.8, ce qui permet de construire un test comme dans le corollaire 10.9.

**Exemple 10.10** *On considère un dé à 6 faces. On souhaite tester au niveau 5% si le dé n'est pas pipé. Donc ici  $\mathcal{X} = \{1, \dots, 6\}$  et  $H_0 = \{(1/6, \dots, 1/6)\}$ .*

*Lors de  $n = 100$  lancers du dé on observe les résultats suivants :  $N_1 = 20, N_2 = 13, N_3 = 17, N_4 = 12, N_5 = 23, N_6 = 15$ . On obtient  $\zeta_{100}^{obs} = 5,36$ . Sous l'hypothèse  $H_0$  ("le dé n'est pas pipé"),  $\zeta_{100}$  suit une loi de  $\chi^2$  à 5 degrés de liberté. Si  $Z \sim \chi_5^2$ , on a  $\mathbb{P}(Z \geq 11,07) = 0,05$ . Comme  $\zeta_{100}^{obs} < 11,07$ , on accepte donc, au niveau 5%, l'hypothèse que le dé n'est pas pipé.*

*Lors de  $n = 1000$  lancers du dé on observe les résultats suivants :  $N_1 = 200, N_2 = 130, N_3 = 170, N_4 = 120, N_5 = 230, N_6 = 150$ . On obtient  $\zeta_{1000}^{obs} = 53,6$ . Comme  $\zeta_{1000}^{obs} > 11,07$ , on rejette, au niveau 5%, l'hypothèse que le dé n'est pas pipé, autrement dit, on affirme que le dé est pipé.*

*Dans le premier cas où  $n = 100$ , les proportions observées étaient les mêmes que dans le second cas où  $n = 1000$ , mais on a cependant accepté l'hypothèse que le dé n'était pas pipé car on n'avait pas assez de données pour rejeter avec suffisamment d'assurance cette hypothèse.*

### 10.3.2 Test d'adéquation à une famille de lois

Nous discutons ici une généralisation du test précédent qui permet entre autres de répondre à une question du type : les observations sont-elles géométriques, gaussiennes, etc ? Il ne s'agit plus de tester l'adéquation d'observations à une loi donnée, mais à une famille de lois.

On observe toujours un  $n$ -échantillon  $(X_1, \dots, X_n)$  de variables aléatoires indépendantes et identiquement distribuées à valeurs dans un espace fini  $\mathcal{X} = \{a_1, \dots, a_k\}$ . La loi est paramétrée par  $\theta = (\theta_1, \dots, \theta_k)$  avec  $\mathbb{P}_\theta(X_1 = a_j) = \theta_j$  pour  $j \in \{1, \dots, k\}$ . Le paramètre  $\theta$  vit dans l'ensemble  $\Theta = \{\theta \in [0, 1]^k, \sum_{i=1}^k \theta_i = 1\}$ . On souhaite tester l'hypothèse nulle  $H_0$  contre  $H_1 = \Theta \setminus H_0$  pour un certain sous-ensemble  $H_0 \subset \Theta$ .

**Exemple 10.11** *On considère  $n = 200$  rouleaux d'un jeu de grattage contenant 100 tickets chacun. Pour  $k = 1, \dots, n$ , on observe  $X_k$  le nombre de tickets gagnants dans le  $k$ ème rouleau. On veut tester l'hypothèse  $H_0 =$  "la loi du nombre de tickets gagnants par rouleau est une binomiale" (en d'autres mots, on veut vérifier si les tickets*

*gagnants sont bien uniformément répartis dans les rouleaux, sans présumer de la proportion de tickets gagnants). Ici  $\mathcal{X} = \{0, \dots, 100\}$  et  $H_0 = \{\theta \in \Theta, \exists p \in [0, 1], \theta_j = \binom{100}{j} p^j (1-p)^{100-j} \forall j = 0, \dots, 100\}$ .*

Pour être plus précis, on va paramétrer l'ensemble des lois qui forment l'hypothèse nulle sous la forme  $H_0 = \{\theta^\pi, \pi \in \Pi\}$ , où

- $\Pi$  est une partie d'intérieur non vide de  $\mathbb{R}^h$  avec  $h < k - 1$ ,
- $\pi \mapsto \theta^\pi$  est une application de  $\Pi$  dans  $\Theta$ .

**Exemple 10.12** *On reprend l'exemple 10.11. Ici on peut paramétrer  $H_0$  avec l'ensemble  $\Pi = [0, 1]$  et l'application  $\theta^p : p \in [0, 1] \mapsto (\theta_j = \binom{100}{j} p^j (1-p)^{100-j})_{j=0, \dots, 100} \in \Theta$ .*

L'idée consiste alors à utiliser un estimateur  $\hat{\pi}_n$  de  $\pi$  à valeurs dans  $\Pi$  (très souvent, ce sera l'EMV de  $\pi$ ) et à comparer les vecteurs  $\hat{\theta}_n$  défini par (10.1) et  $\theta^{\hat{\pi}_n}$ . Si ces vecteurs sont suffisamment proches, on pourra accepter l'hypothèse  $H_0$ . Encore une fois, la question essentielle est de savoir ce que veut dire "suffisamment proches".

**Proposition 10.13** *Soit*

$$\zeta_n = n \sum_{j=1}^k \frac{(\hat{\theta}_{n,j} - \theta_j^{\hat{\pi}_n})^2}{\theta_j^{\hat{\pi}_n}}.$$

*Sous des hypothèses de régularité non-précisées (vérifiées en général lorsque  $\hat{\theta}_n$  est l'EMV de  $\theta$  et  $\hat{\pi}_n$  est l'EMV de  $\pi$ ) :*

- Sous  $H_0$ ,  $\zeta_n$  converge en loi vers  $Z \sim \chi_{k-h-1}^2$ .
- Sous  $H_1$ ,  $\zeta_n$  tend presque sûrement vers  $+\infty$ .

Il est essentiel de noter que le nombre de degrés de liberté dans la limite en loi sous  $H_0$  est  $k - h - 1$  et non plus  $k - 1$  comme dans le test d'adéquation à une loi donnée. En effet, il faut estimer le paramètre  $\pi$  ce qui réduit le nombre de degrés de liberté.

Finalement, en procédant comme pour le corollaire 10.9, on conclut que le test de région critique  $W_n = \{\zeta_n \geq x_{1-\alpha}(k - h - 1)\}$  est convergent de niveau asymptotique  $\alpha$ , avec  $x_r(k - h - 1)$  le quantile d'ordre  $r$  de la loi  $\chi^2$  à  $k - h - 1$  degrés de liberté.

**Exemple 10.14** *On reprend l'exemple 10.11-10.12. On note  $N_i = \text{Card}(k = 1, \dots, 200, X_k = i)$  et on observe :*

$i$	0	1	2	3	4	5	6	7	8	9	10	11	12	$\geq 13$
$N_i$	1	7	14	29	36	41	26	17	17	8	1	1	2	0

L'EMV de  $p$  est (voir l'exercice 8.4) :

$$\hat{p}_{200} = \frac{\sum_{k=1}^{200} X_k}{100 \times 200} = \frac{\sum_i i N_i}{100 \times 200} \simeq 0,05.$$

On regroupe par paquets de tailles supérieures à 5 :

$i$	$\leq 1$	2	3	4	5	6	7	8	$\geq 9$
$N_i$	8	14	29	36	41	26	17	17	12
$200\theta_i^{p_{200}}$	7,45	16,29	27,97	35,67	36,00	29,97	21,16	12,93	12,55

et donc, avec  $\hat{\theta}_{200,i} = N_i/200$ ,

$i$	$\leq 1$	2	3	4	5	6	7	8	$\geq 9$
$\hat{\theta}_{200,i}$	0,0400	0,0700	0,1450	0,1800	0,2050	0,1300	0,0850	0,0850	0,0600
$\theta_i^{p_{200}}$	0,0372	0,0814	0,1399	0,1783	0,1800	0,1499	0,1058	0,0647	0,0627

On calcule

$$\zeta_{200}^{obs} = 200 \sum_{j=1}^{200} \frac{(\hat{\theta}_{200,j} - \theta_j^{p_{200}})^2}{\theta_j^{p_{200}}} \simeq 3,74.$$

Or ici on a  $9 - 1 - 1 = 7$  degrés de liberté, le seuil pour le test  $\{\zeta_{200} \geq x_{0,95}(7)\}$  au niveau 0,05 est  $x_{0,95}(7) = 14,07$ . Comme on a  $\zeta_{200}^{obs} < x_{0,95}(7)$ , on accepte l'hypothèse nulle "la loi du nombre de tickets gagnants par rouleau est une binomiale".

**Exemple 10.15** On considère les résultats au concours de l'X de deux lycées :

	Admis	Recalés	Présentés
Henri IV	81	17	98
Le Parc	136	17	153
Total	217	34	251

On désire tester l'hypothèse selon laquelle les élèves des deux lycées ont le même taux de réussite à l'X.

Ici chaque observation est de la forme  $(X_i, Y_i)$  où  $X_i$  est à valeurs dans  $\{H, L\}$  et  $X_i = H$ , resp.  $L$ , signifie que l'étudiant vient du lycée  $H$ , resp. lycée  $L$ , et  $Y_i$  est à valeurs dans  $\{A, R\}$  et  $Y_i = A$ , resp.  $R$ , signifie que l'étudiant a été admis, resp. a été recalé. L'ensemble  $\Theta$  de toutes les lois possibles est de la forme  $\Theta = \{(\theta_{jl})_{j \in \{H,L\}, l \in \{A,R\}} \in [0, 1]^4, \sum_{jl} \theta_{jl} = 1\}$  et l'hypothèse nulle est de la forme  $H_0 = \{(\theta_{jl})_{j \in \{H,L\}, l \in \{A,R\}}, \theta_{HA} = qp, \theta_{LA} = (1-q)p, \theta_{HR} = q(1-p), \theta_{LR} = (1-q)(1-p), p \in [0, 1], q \in [0, 1]\}$ , qui contient toutes les lois pour lesquelles le taux d'admission  $p$  (inconnu) ne dépend pas du lycée. La pseudo-distance du  $\chi^2$  est

$$\zeta_{251}^{obs} = 251 \sum_{j \in \{H,L\}, l \in \{A,R\}} \frac{(\hat{\theta}_{jl} - \hat{q}_j \hat{p}_l)^2}{\hat{q}_j \hat{p}_l},$$

avec

$$\hat{q}_H = \frac{98}{251}, \quad \hat{q}_L = \frac{153}{251}, \quad \hat{p}_A = \frac{217}{251}, \quad \hat{p}_R = \frac{34}{251},$$

$$\hat{\theta}_{HA} = \frac{81}{251}, \quad \hat{\theta}_{HR} = \frac{17}{251}, \quad \hat{\theta}_{LA} = \frac{136}{251}, \quad \hat{\theta}_{LR} = \frac{17}{251}.$$

On trouve :  $\zeta_{251}^{obs} = 1,98$ . Or ici on a  $4 - 2 - 1 = 1$  degré de liberté, le seuil pour le test  $\{\zeta_{251} \geq x_{0,95}(1)\}$  au niveau 0,05 est  $x_{0,95}(1) = 3,84$ . Comme on a  $\zeta_{251}^{obs} < x_{0,95}(1)$ , on accepte l'hypothèse nulle que les deux lycées ont le même taux de réussite.

On peut voir le dernier exemple comme un test d'indépendance (on teste l'indépendance du taux de réussite et du lycée d'origine). Il peut se généraliser comme on l'explique ci-dessous.

### 10.3.3 Test d'indépendance

Ici on suppose que les observations sont des paires  $(X_i, Y_i)$ , par exemple taux de réussite au concours et lycée d'origine, ou encore température et hygrométrie journalières, et on se demande si ces deux quantités sont indépendantes. On peut construire un test d'indépendance qui est en fait un cas particulier du test d'adéquation à une famille de lois qu'on vient de présenter.

On observe un  $n$ -échantillon  $((Y_1, Z_1), \dots, (Y_n, Z_n))$  de vecteurs aléatoires indépendants et identiquement distribués, avec  $Y_i$  à valeurs dans  $\{b_1, \dots, b_d\}$  et  $Z_i$  à valeurs dans  $\{c_1, \dots, c_m\}$ . On pose  $X_i = (Y_i, Z_i)$  à valeurs dans l'espace fini  $\mathcal{X} = \{b_1, \dots, b_d\} \times \{c_1, \dots, c_m\}$  de cardinal  $dm$ . On note  $\theta = (\theta_{jl}, 1 \leq j \leq d, 1 \leq l \leq m)$  où  $\mathbb{P}_\theta(Y_i = b_j, Z_i = c_l) = \theta_{jl}$ . L'ensemble  $H_0$  des  $\theta$  qui donnent des lois  $\mathbb{P}_\theta$  de forme produit correspondant à l'indépendance de  $Y_i$  et  $Z_i$  peut être paramétré par  $H_0 = \{\theta^\pi, \pi \in \Pi\}$ , avec

$$\theta_{jl}^\pi = \begin{cases} \pi_j^{(1)} \pi_l^{(2)} & \text{si } 1 \leq j \leq d-1, 1 \leq l \leq m-1, \\ (1 - \sum_{j'=1}^{d-1} \pi_{j'}^{(1)}) \pi_l^{(2)} & \text{si } j = d, 1 \leq l \leq m-1, \\ \pi_j^{(1)} (1 - \sum_{l'=1}^{m-1} \pi_{l'}^{(2)}) & \text{si } 1 \leq j \leq d-1, l = m, \\ (1 - \sum_{j'=1}^{d-1} \pi_{j'}^{(1)}) (1 - \sum_{l'=1}^{m-1} \pi_{l'}^{(2)}) & \text{si } j = d, l = m, \end{cases}$$

et

$$\Pi = \left\{ \pi = (\pi_j^{(1)} \in [0, 1], 1 \leq j \leq d-1, \pi_l^{(2)} \in [0, 1], 1 \leq l \leq m-1), \right. \\ \left. \text{avec } \sum_{j=1}^{d-1} \pi_j^{(1)} \leq 1, \sum_{l=1}^{m-1} \pi_l^{(2)} \leq 1 \right\},$$

qui est une partie d'intérieur non-vide de  $\mathbb{R}^{d+m-2}$ . On pose

$$\hat{\theta}_{jl} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(b_j, c_l)}(Y_i, Z_i), \quad \hat{\pi}_j^{(1)} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{b_j}(Y_i), \quad \hat{\pi}_l^{(2)} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{c_l}(Z_i).$$

Notons que l'EMV de  $\boldsymbol{\pi} = (\pi_j^{(1)}, 1 \leq j \leq d-1, \pi_l^{(2)}, 1 \leq l \leq m-1)$  est  $\hat{\boldsymbol{\pi}} = (\hat{\pi}_j^{(1)}, 1 \leq j \leq d-1, \hat{\pi}_l^{(2)}, 1 \leq l \leq m-1)$ . On note

$$\zeta_n = n \sum_{j=1}^d \sum_{l=1}^m \frac{(\hat{\theta}_{jl} - \hat{\pi}_j^{(1)} \hat{\pi}_l^{(2)})^2}{\hat{\pi}_j^{(1)} \hat{\pi}_l^{(2)}}$$

(avec la convention que les termes de la somme pour lesquels le dénominateur est nul sont nuls).  $\zeta_n$  mesure la “distance” entre la matrice  $\hat{\boldsymbol{\theta}}$  des fréquences des couples  $(b_j, c_l)$  et la matrice  $\hat{\boldsymbol{\pi}}$  des produits des fréquences marginales. On comprend que, si les deux coordonnées  $Y_i$  et  $Z_i$  sont indépendantes,  $\hat{\boldsymbol{\theta}}$  et  $\hat{\boldsymbol{\pi}}$  doivent être proches. Donc on rejette l'hypothèse  $H_0$  d'indépendance si  $\zeta_n$  est grand. Plus quantitativement, comme la dimension  $h$  de  $\boldsymbol{\Pi}$  est  $d + m - 2$ , on a  $k - h - 1 = dm - d - m + 1 = (d-1)(m-1)$ . On rejette l'hypothèse  $H_0$  d'indépendance au niveau  $\alpha$  si  $\zeta_n$  dépasse  $x_{1-\alpha}((d-1)(m-1))$  et on l'accepte sinon, où  $x_{1-\alpha}((d-1)(m-1))$  le quantile d'ordre  $1 - \alpha$  de la loi  $\chi^2$  à  $(d-1)(m-1)$  degrés de liberté. L'exemple 10.15 est un cas particulier d'application de cette méthode, avec  $d = m = 2$ .

## 10.4 Exercices sur le chapitre 10

**EXERCICE 10.1** On se place dans le modèle exponentiel  $\mathcal{P} = \{\mathcal{E}(\theta), \theta \in \Theta\}$ ,  $\Theta = ]0, +\infty[$ .

- Rappeler  $\mathbb{E}_\theta(X_1)$ .
- Soit  $S_n = X_1 + \dots + X_n$ . En remarquant que  $2\theta S_n \sim \chi_{2n}^2$ , construire un test de niveau  $\alpha$  pour  $H_0 = \{\theta_0\}$  et  $H_1 = H_0^c$ .  
A.N. :  $n = 15$ ,  $\bar{x}_n = 1.47$ ,  $\theta_0 = 1$ ,  $\alpha = 5\%$ .
- Proposer un test pour  $H_0 = [\theta_0, +\infty[$  et  $H_1 = ]0, \theta_0[$ .

**EXERCICE 10.2** On souhaite vérifier la qualité du générateur de nombres aléatoires d'un ordinateur. Pour cela, on procède à 250 tirages dans l'ensemble  $\{0, \dots, 9\}$  et on obtient les résultats suivants :

$x$	0	1	2	3	4	5	6	7	8	9
$N(x)$	28	32	23	26	23	31	18	19	19	31

Tester au niveau  $\alpha = 0,1$  si le générateur produit des entiers uniformément répartis sur  $\{0, \dots, 9\}$ .





# Chapitre 11

## Corrections des exercices

### 11.1 Corrigés des exercices du chapitre 2

*Exercice 2.7* : 1) Soit  $P_n$  cette probabilité. Alors

$$\begin{aligned} P_n &= 1 - \mathbb{P}(\text{il n'y a pas de personnes ayant leur anniversaire le même jour}) \\ &= 1 - \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{(365)^n}. \end{aligned}$$

Le calcul donne :

$$P_4 = 0,016 ; P_{16} = 0,284 ; P_{22} = 0,476 ; P_{40} = 0,891 ; P_{64} = 0,997.$$

2) Nous cherchons le plus petit entier  $n$  tel que  $\frac{365!}{(365-n)!(365)^n} \leq \frac{1}{2}$ . Avec une égalité, la formule de Stirling donne alors qu'approximativement,

$$e^{-n} \left(1 - \frac{n}{365}\right)^{-(365 + \frac{1}{2} - n)} = 0,5.$$

En passant au logarithme et en ne gardant que les termes prépondérants, nous obtenons  $\frac{n^2 - n}{2(365)} = 0,693$ , d'où  $n = 23$ .

*Exercice 2.8* : Cette formule se montre par récurrence sur  $n$ . Elle est triviale pour  $n = 1$ . Remarquons que pour  $n = 2$ ,

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2). \quad (11.1)$$

La formule est donc vérifiée puisque dans ce cas,  $p_1 = \mathbb{P}(A_1) + \mathbb{P}(A_2)$  et  $p_2 = \mathbb{P}(A_1 \cap A_2)$ . Supposons que la formule soit vraie pour toute réunion de  $n - 1$  événements. Montrons-la pour  $n$ . Posons  $B_1 = A_1 \cup \dots \cup A_{n-1}$  et  $B_2 = A_n$ . En appliquant (11.1) à  $B_1$  et  $B_2$  et la formule de récurrence pour calculer  $\mathbb{P}(B_1)$  et  $\mathbb{P}((A_1 \cap A_n) \cup \dots \cup (A_{n-1} \cap A_n))$ , nous obtenons immédiatement le résultat.

*Exercice 2.9 :*

1) Il y a une seule possibilité de bonne remise de lettres parmi les  $n!$  possibilités. La probabilité de bonne répartition est donc  $\frac{1}{n!}$ .

2) Numérotions de 1 à  $n$  les lettres et numérotions par les mêmes numéros les boîtes qui sont censées leur correspondre. Appelons  $A_i$  l'événement "La lettre numéro  $i$  arrive dans la boîte numéro  $i$ ". Ainsi,  $A_i$  sera réalisé si la remise de la lettre  $i$  est fixée à la bonne valeur, indépendamment de la manière aléatoire dont les  $n - 1$  autres lettres sont distribuées. Nous en déduisons que  $\mathbb{P}(A_i) = \frac{(n-1)!}{n!}$ . De même, pour deux numéros quelconques  $i_1$  et  $i_2$ , on aura  $\mathbb{P}(A_{i_1} \cap A_{i_2}) = \frac{(n-2)!}{n!}$ .

L'événement  $E$  "une lettre au moins arrive à la bonne adresse" est égal à  $E = A_1 \cup \dots \cup A_n$ . On peut donc appliquer la formule de Poincaré.

$$\begin{aligned} \mathbb{P}(E) &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) \\ &= \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^n \frac{(-1)^{k-1}}{k!}. \end{aligned}$$

Remarquons que quand  $n$  tend vers l'infini, cette quantité tend vers  $1 - e^{-1} \simeq 0.632$ . On se convaincra que pour des valeurs modérées de  $n$  ( $n=7$ ) on est très proche de cette limite.

3) La probabilité pour qu'aucune lettre n'arrive à destination vaut alors

$$\mathbb{P}(E^c) = 1 - \mathbb{P}(E) = \sum_{k=0}^n \frac{(-1)^k}{k!},$$

qui tend donc vers  $e^{-1} = 0,368$  quand  $n$  tend vers l'infini.

4)  $d_n = n! \mathbb{P}(E^c)$ . □

*Exercice 2.10 :* Il est immédiat de montrer que pour tout  $n \in \mathbb{N}^*$ ,  $0 \leq p_n \leq 1$  et que  $\sum_{n \in \mathbb{N}^*} p_n = 1$ .

*Exercice 2.11* : Notons F et G les événements “fille” ou “garçon”.

1. Une configuration possible (F,F) sur 4 configurations :  $\frac{1}{4}$ .
2. Une configuration possible sur les 2 configurations (F,F), (F,G) :  $\frac{1}{2}$ .
3. Une configuration possible sur les 3 configurations (F,F), (F,G), (G,F) :  $\frac{1}{3}$ .

*Exercice 2.12* : Les probabilités des différents génotypes valent alors, par indépendance

$$\mathbb{P}(AA) = p^2, \quad \mathbb{P}(Aa) = 2pq, \quad \mathbb{P}(aa) = q^2.$$

La proportion d'allèle A dans la deuxième génération sera alors

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A | \text{individu de génotype } AA) \mathbb{P}(AA) + \mathbb{P}(A | \text{individu de génotype } Aa) \mathbb{P}(Aa) \\ &= p^2 + \frac{2pq}{2} = p, \end{aligned}$$

et de même la proportion d'allèle a sera q.

*Exercice 2.13* : Notons H le fait d'être hémophile et NH le contraire. Nous savons que

$$\mathbb{P}(\text{Reine H}) = \frac{1}{2}; \quad \mathbb{P}(3 \text{ fils NH} | \text{Reine H}) = \left(\frac{1}{2}\right)^3,$$

par indépendance des naissances. Nous cherchons à calculer  $\mathbb{P}(\text{Reine H} | 3 \text{ fils NH})$ . Nous allons utiliser la formule de Bayes. Nous avons

$$\begin{aligned} \mathbb{P}(3 \text{ fils NH}) &= \mathbb{P}(3 \text{ fils NH} | \text{Reine H}) \mathbb{P}(\text{Reine H}) \\ &\quad + \mathbb{P}(3 \text{ fils NH} | \text{Reine NH}) \mathbb{P}(\text{Reine NH}) \\ &= \left(\frac{1}{2}\right)^3 \times \frac{1}{2} + \frac{1}{2} = \frac{9}{2^4}, \end{aligned}$$

$$\text{d'où } \mathbb{P}(\text{Reine H} | 3 \text{ fils NH}) = \frac{\frac{1}{2} \times \left(\frac{1}{2}\right)^3}{\frac{9}{2^4}} = \frac{\frac{1}{2^4}}{\frac{9}{2^4}} = \frac{1}{9}.$$

De plus,

$$\mathbb{P}(4^{\text{ème}} \text{ fils H} | 3 \text{ fils NH}) = \mathbb{P}(4^{\text{ème}} \text{ fils H} | 3 \text{ fils H, Reine H}) \mathbb{P}(\text{Reine H} | 3 \text{ fils NH}) = \frac{1}{2} \times \frac{1}{9} = \frac{1}{18}.$$

*Exercice 2.14* : Notons A (resp. B, C) l'événement “la Ferrari est derrière la porte A” (resp. B, C). Soit E l'événement “le présentateur annonce au joueur qu'elle n'est pas derrière la porte B”. On a

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{3}.$$

Par ailleurs,

$$\mathbb{P}(E|A) = \frac{1}{2}, \quad \mathbb{P}(E|B) = 0, \quad \mathbb{P}(E|C) = 1.$$

En effet, si la voiture est derrière A, le présentateur a le choix entre B et C, alors que si elle est derrière C, il n'a pas le choix puisqu'il ne peut pas parler de A. Ainsi,

$$\mathbb{P}(E) = \mathbb{P}(E|A)\mathbb{P}(A) + \mathbb{P}(E|B)\mathbb{P}(B) + \mathbb{P}(E|C)\mathbb{P}(C) = \frac{1}{2} \times \frac{1}{3} + \frac{1}{3} = \frac{1}{2}.$$

On en déduit que

$$\mathbb{P}(A|E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E|A)\mathbb{P}(A)}{\mathbb{P}(E)} = \frac{1}{3}.$$

De même

$$\mathbb{P}(C|E) = \frac{\mathbb{P}(E|C)\mathbb{P}(C)}{\mathbb{P}(E)} = \frac{2}{3}.$$

Il faut donc que le joueur révise son choix.

*Exercice 2.15* : 1)

$$\mathbb{P}(E_2|F_2) = \frac{\mathbb{P}(E_2 \cap F_2)}{\mathbb{P}(F_2)} = \frac{\mathbb{P}(E_2)\mathbb{P}(F_2|E_2)}{\mathbb{P}(F_2)}.$$

$$\mathbb{P}(F_2) = \sum_{n=2}^{+\infty} \mathbb{P}(F_2|E_n)\mathbb{P}(E_n) \quad \text{et} \quad \mathbb{P}(F_2|E_n) = \binom{n}{2} \frac{1}{2^n}.$$

On a  $\mathbb{P}(F_2|E_0) = \mathbb{P}(F_2|E_1) = 0$ .

$$\mathbb{P}(F_2) = \sum_{n=2}^{+\infty} p_n \binom{n}{2} \frac{1}{2^n} = \frac{(1-2a)}{4^2} \sum_{n=2}^{+\infty} n(n-1) \frac{1}{4^{n-2}}.$$

Calcul de la somme : dérivée seconde de la série entière  $\sum_n x^n$ .

On obtient  $\mathbb{P}(F_2) = \frac{8(1-2a)}{27}$ , et

$$\mathbb{P}(E_2|F_2) = \frac{p_2 \frac{1}{4}}{\frac{8(1-2a)}{27}} = \frac{27}{64} \simeq 0,42.$$

2) On calcule  $\mathbb{P}(G_2|F_2) = \frac{\mathbb{P}(G_2 \cap F_2)}{\mathbb{P}(F_2)}$ .

$$\mathbb{P}(G_2 \cap F_2) = \mathbb{P}(E_4)\mathbb{P}(G_2 \cap F_2|E_4) = (1-2a) \frac{1}{2^3} \binom{4}{2} \frac{1}{2^4} = \frac{3(1-2a)}{64}.$$

$$\mathbb{P}(G_2|F_2) = \frac{81}{512} \simeq 0,158.$$

*Exercice 2.16* : Remarquons que les événements aléatoires  $A_k$  sont indépendants. Ainsi, par le lemme de Borel-Cantelli 2.35,  $\mathbb{P}(\limsup_k A_k)$  sera égale à 0 ou 1 suivant que la série de terme général  $\mathbb{P}(A_k)$  converge ou diverge. La réalisation de  $A_k$  entraîne l'existence d'un nombre  $i \in \{2^k, \dots, 2^{k+1} - k\}$  tel que les lancers  $i, i+1, \dots, i+k$  donnent tous Face. En appelant  $A_{k,i}$  cet événement, nous obtenons

$$\mathbb{P}(A_k) \leq \sum_{i=2^k}^{2^{k+1}-k} \mathbb{P}(A_{k,i}) = (2^k - k + 1)p^k \leq 2^k p^k.$$

Donc, si  $p < \frac{1}{2}$ , nous en déduisons que la série  $\sum_k \mathbb{P}(A_k)$  converge, et par le lemme de Borel-Cantelli, que  $\mathbb{P}(\limsup_k A_k) = 0$ .

Nous allons maintenant montrer que si  $p \geq \frac{1}{2}$ , la série diverge. Découpons  $\{2^k, \dots, 2^{k+1} - 1\}$  en à peu près  $\frac{2^k}{k}$  morceaux de longueur  $k$ ,  $\{2^k, \dots, 2^k + k - 1\}, \{2^k + k, \dots, 2^k + 2k - 1\}, \dots$ . Notons  $B_{k,i}$  l'événement "tous les lancers du  $i$ ème morceau donnent Face". Alors

$$\mathbb{P}(A_k) \geq \mathbb{P}(\exists i \in \{1, \dots, [2^k/k]\}; B_{k,i} \text{ est réalisé}).$$

Or les  $B_{k,i}$  ne concernent pas les mêmes lancers et sont donc indépendants. Nous avons alors

$$\mathbb{P}(A_k^c) \leq \mathbb{P}(\forall i B_{k,i}^c \text{ est réalisé}) = (1 - p^k)^{[2^k/k]} \leq \exp(-p^k [2^k/k]).$$

Ainsi, pour  $k$  assez grand,  $\mathbb{P}(A_k) \geq 1 - \exp(-p^k [2^k/k])$ .

Si  $p > 1/2$ ,  $p^k [2^k/k] \rightarrow \infty$  et donc  $\mathbb{P}(A_k) > 1/2$  pour  $k$  assez grand.

Si  $p = 1/2$ ,  $\mathbb{P}(A_k) \geq 1/(2k)$  pour  $k$  assez grand et donc  $\sum_k \mathbb{P}(A_k) = \infty$ .

*Exercice 2.17* : Remarquons que si  $p$  et  $q$  sont des nombres premiers, alors  $p\mathbb{N} \cap q\mathbb{N} = pq\mathbb{N}$ . Par ailleurs, si cette probabilité existe, nous aurons

$$\mathbb{P}(pq\mathbb{N}) = \frac{1}{pq} = \mathbb{P}(p\mathbb{N}) \mathbb{P}(q\mathbb{N}).$$

Les événements  $p\mathbb{N}$  et  $q\mathbb{N}$  sont donc indépendants, et de même pour toute suite finie. Or  $\sum_{p \text{ premier}} \mathbb{P}(p\mathbb{N}) = +\infty$  (série harmonique), et donc par le lemme de Borel-Cantelli, nous avons que  $\mathbb{P}(\limsup_{p \text{ premier}} p\mathbb{N}) = 1$ . Ce qui voudrait dire que  $\mathbb{P}$ -presque sûrement tout nombre entier serait multiple d'une infinité de nombres premiers, et en particulier que l'ensemble des nombres entiers multiples d'une infinité de nombres premiers serait non vide, ce qui est faux.

## 11.2 Corrigés des exercices du chapitre 3

*Exercice 3.2* : Pour  $k \geq 1$ ,  $\mathbb{P}(X = k) = \frac{k}{(k+1)!}$ . Nous en déduisons que  $\mathbb{E}(X) = e - 1$  et  $\text{Var}(X) = 3e - e^2$ .

*Exercice 3.3* : Nous pouvons écrire :  $X = \sum_{k \in \mathbb{N}} \mathbf{1}_{k < X}$ . Ainsi,

$$\mathbb{E}(X) = \mathbb{E} \left( \sum_{k \in \mathbb{N}} \mathbf{1}_{k < X} \right) = \sum_{p=0}^{\infty} p \mathbb{P} \left( \sum_{k \in \mathbb{N}} \mathbf{1}_{k < X} = p \right) = \sum_{p=0}^{\infty} p \mathbb{P}(X = p),$$

où nous avons appliqué le théorème de Fubini pour les séries doubles de termes positifs.

*Exercice 3.4* : Soit  $X_i$  le nombre de blessés dans le  $i$ -ème accident. Nous avons  $Y = \sum_{i=1}^N X_i$ . Ainsi, pour  $s \in [0, 1[$ ,

$$\begin{aligned} G_Y(s) &= \mathbb{E}(s^Y) = \mathbb{E} \left( s^{\sum_{i=1}^N X_i} \right) = \sum_k \mathbb{E} \left( s^{\sum_{i=1}^N X_i} \mid N = k \right) \mathbb{P}(N = k) \\ &= \sum_k \mathbb{E}(s^{X_1}) \dots \mathbb{E}(s^{X_k}) \mathbb{P}(N = k), \end{aligned}$$

par indépendance des variables aléatoires  $N, X_1, \dots, X_k$ . Nous en déduisons que

$$G_Y(s) = \sum_k \mathbb{P}(N = k) (G_X(s))^k = G_N(G_X(s)).$$

Ainsi, en dérivant cette formule en 1 et en utilisant que  $G_X(1) = 1$  et  $G'_Y(1) = \mathbb{E}(Y)$ ,  $G'_X(1) = \mathbb{E}(X)$ ,  $G'_N(1) = \mathbb{E}(N)$ , nous obtenons

$$\mathbb{E}(Y) = \mathbb{E}(N) \mathbb{E}(X) = m \mu.$$

De même, en dérivant de nouveau la fonction génératrice, nous pouvons en déduire la variance de  $Y$ . Tous calculs faits, nous obtenons

$$\text{Var}(Y) = \text{Var}(N) (\mathbb{E}(X))^2 + \mathbb{E}(N) \text{Var}(X) = \sigma^2 \mu^2 + m \tau^2.$$

*Exercice 3.5* : a) Nous savons que  $S$  suit la loi de Poisson de paramètre  $\lambda + \mu$ , comme somme de variables aléatoires indépendantes, de loi de Poisson de paramètres  $\lambda$  et  $\mu$ .

b) La loi conditionnelle de  $X$  sachant  $S = n$  ( $n \geq 0$ ) est par définition

$$p_k^{X|S=n} = \frac{p_{X,S}(k,n)}{p_S(n)} = \binom{n}{k} \left( \frac{\lambda}{\lambda + \mu} \right)^k \left( \frac{\mu}{\lambda + \mu} \right)^{n-k}$$

pour  $k \in [0, n]$  entier, et  $p_k^{X|S=n} = 0$  sinon. Cette loi est la loi binomiale  $\mathcal{B}\left(n, \frac{\lambda}{\lambda + \mu}\right)$ .

On peut alors calculer  $\mathbb{E}(X | S = n)$  en utilisant (3.16), mais il est plus rapide d'utiliser la valeur  $np$  de l'espérance d'une variable de loi  $\mathcal{B}(n, p)$ . Ainsi,  $\mathbb{E}(X | S = n) = n \frac{\lambda}{\lambda + \mu}$ , ou encore

$$\mathbb{E}(X | S) = S \frac{\lambda}{\lambda + \mu}.$$

*Exercice 3.6* : Calculons les lois de  $Z$ ,  $Y$ , et  $X$  :

Loi de  $Z$  :  $\mathbb{P}(Z = k) = \sum_{n=0}^k p_{k,n} = \frac{e^{-2} 2^k}{k!}$ . C'est une loi de Poisson de paramètre 2.

Loi de  $X$  :  $\mathbb{P}(X = n) = \sum_{k=n}^{\infty} p_{k,n} = \frac{e^{-1,04} (1,04)^n}{n!}$ . C'est une loi de Poisson de paramètre 1,04.

Nous remarquons immédiatement que  $Z$  et  $X$  ne sont pas indépendantes puisque

$$\mathbb{P}(Z = k) \mathbb{P}(X = n) \neq \mathbb{P}(Z = k; X = n).$$

Loi de  $Y = Z - X$  :

$$\mathbb{P}(Y = m) = \mathbb{P}(Z = X + m) = \sum_n p_{n+m,n} = \frac{e^{-0,96} (0,96)^m}{m!}.$$

C'est une loi de Poisson de paramètre 0,96.

Nous observons que que

$$\mathbb{P}(X = n, Y = m) = \mathbb{P}(Z = n + m, X = n) = \mathbb{P}(X = n) \mathbb{P}(Y = m).$$

Les v.a.  $X$  et  $Y$  sont indépendantes.

Loi conditionnelle de  $X$  sachant  $Z = k$  : pour tout  $0 \leq n \leq k$ , nous avons

$$\mathbb{P}(X = n | Z = k) = \frac{\mathbb{P}(X = n, Z = k)}{\mathbb{P}(Z = k)} = \binom{k}{n} (0,52)^n (0,48)^{k-n}.$$

La loi conditionnelle de  $X$  sachant  $Z = k$  est donc une loi binomiale  $\mathcal{B}(k; 0,52)$ . Son espérance vaut donc  $\mathbb{E}(X | Z = k) = 0,52 k$ , et  $\mathbb{E}(X | Z) = 0,52 Z$ .

*Exercice 3.7* :

1)

$$\begin{aligned}\mathbb{P}(X_n \neq X_{n-1}) &= \mathbb{P}(X_n = 0, X_{n-1} = 1) + \mathbb{P}(X_n = 1, X_{n-1} = 0) \\ &= (1-p)p + p(1-p) = 2p(1-p).\end{aligned}$$

Par ailleurs,

$$\begin{aligned}\mathbb{P}(A_n \cap A_{n-1}) &= \mathbb{P}(X_n = 0, X_{n-1} = 1, X_{n+1} = 1) + \mathbb{P}(X_n = 1, X_{n-1} = 0, X_{n+1} = 0) \\ &= p^2(1-p) + p(1-p)^2 = p(1-p).\end{aligned}$$

Ainsi

$$\mathbb{P}(A_n \cap A_{n-1}) = \mathbb{P}(A_n)\mathbb{P}(A_{n-1}) \iff p(1-p) = 4p^2(1-p)^2 \iff p = \frac{1}{2}.$$

Cette condition est suffisante pour avoir l'indépendance de la suite  $(A_n)_n$ . En effet, dès que les indices  $k$  et  $n$  sont distants de plus d'une unité,  $A_k$  et  $A_n$  sont indépendants, par définition, car les variables aléatoires  $X_n$  le sont.

2-a) For  $n \geq 2$ ,  $\mathbb{P}(T = n) = \mathbb{P}(X_1 = \dots = X_{n-1}, X_{n-1} \neq X_n) = p^{n-1}(1-p) + p(1-p)^{n-1}$ .

2-b)  $\mathbb{P}(T < \infty) = \mathbb{P}(\cup_{n \geq 2} \{T = n\}) = (1-p) \sum_{n \geq 1} p^n + p \sum_{n \geq 1} (1-p)^n = 1$ .

3)

$$\begin{aligned}\mathbb{P}((X_T, X_{T+1}) = (0, 1)) &= \sum_{n \geq 2} \mathbb{P}((X_n, X_{n+1}) = (0, 1); T = n) \\ &= \sum_{n \geq 2} \mathbb{P}(X_1 = \dots = X_{n-1} = 1, X_n = 0, X_{n+1} = 1) \\ &= \sum_{n \geq 2} (1-p)p^n = p^2.\end{aligned}$$

Puisque nous avons également  $\mathbb{P}((X_T, X_{T+1}) = (1, 0)) = (1-p)^2$ , ces deux quantités sont égales si et seulement si  $p = \frac{1}{2}$ .

*Exercice 3.8 :*

1) Remarquons que  $\{U_p(x) = n\} = \{x = p^n y, y \in \mathbb{N}^*, \text{ premier avec } p\}$ . Ainsi,

$$\mathbb{Q}(U_p = n) = \sum_{y: y \wedge p=1} \mathbb{Q}(p^n y) = \sum_{y: y \wedge p=1} \frac{c}{p^{2n} y^2} = \frac{c}{p^{2n}} \sum_{y: y \wedge p=1} \frac{1}{y^2}.$$

Mais  $\sum_{n=0}^{\infty} \mathbb{Q}(U_p = n) = 1$ , d'où  $\sum_{n=0}^{\infty} \frac{c}{p^{2n}} \sum_{y: y \wedge p=1} \frac{1}{y^2} = 1$ . Ainsi  $K_p = \sum_{y: y \wedge p=1} \frac{1}{y^2}$  vérifie  $cK_p \left( \frac{1}{1-\frac{1}{p^2}} \right) = 1$ , et  $cK_p = 1 - \frac{1}{p^2}$ . Finalement,  $\mathbb{Q}(U_p = n) = \left(1 - \frac{1}{p^2}\right) \frac{1}{p^{2n}}$ .



2)  $\mathbb{Q}(U_p \geq n) = \mathbb{Q}(\{x = p^n y, y \in \mathbb{N}^*\}) = \frac{c}{p^{2n}} \sum_{y \in \mathbb{N}^*} \frac{1}{y^2} = \frac{1}{p^{2n}}$  car  $c \sum_{y \in \mathbb{N}^*} \frac{1}{y^2} = 1$ , puisque  $\mathbb{Q}$  est une probabilité.

3) Comme ci-dessus,

$$\begin{aligned} \mathbb{Q}(U_{p_1} \geq n_1, \dots, U_{p_k} \geq n_k) &= \mathbb{Q}(x, x = p_1^{n_1} \dots p_k^{n_k} y, y \in \mathbb{N}^*) \\ &= \frac{1}{p_1^{2n_1} \dots p_k^{2n_k}}. \end{aligned}$$

Ainsi les événements  $\{U_{p_1} \geq n_1\}, \dots, \{U_{p_k} \geq n_k\}$  sont indépendants. Il en est de même des événements  $\{U_{p_1} = n_1\}, \dots, \{U_{p_k} = n_k\}$ , et donc des variables aléatoires.

4) Le calcul donne  $G(x) = \frac{p^2-1}{p^2-x}$ , pour  $x \in [0, 1]$ . Il en résulte  $\mathbb{E}(U_p) = \frac{1}{p^2-1}$  et  $\text{Var}(U_p) = \frac{p^2}{(p^2-1)^2}$ .

*Exercice 3.9 :*

1-a) C'est évident.

1-b) La condition suffisante est immédiate. Elle est nécessaire car  $\phi \geq 0$ ,  $\phi$  ne s'annule qu'en 0 ou en 1, et  $H(p) = \sum_i \phi(p_i)$  avec  $\sum_i p_i = 1$ .

1-c) On utilise le rappel avec l'inégalité inverse car  $\phi$  est concave.

$$\frac{1}{n} \sum_{i=1}^n \phi(p_i) \leq \phi\left(\frac{1}{n} \sum_{i=1}^n p_i\right) = \phi\left(\frac{1}{n}\right).$$

Donc  $H(p) = \sum_{i=1}^n \phi(p_i) \leq n \phi\left(\frac{1}{n}\right) = \ln n$ . L'égalité a lieu si et seulement si  $p = \left(\frac{1}{n}, \dots, \frac{1}{n}\right)$  car  $\phi$  est strictement concave.

L'interprétation des deux propriétés précédentes est la suivante : l'entropie quantifie l'incertitude sur l'état  $p$ . Dans le cas extrême où  $p$  est une mesure de Dirac, l'entropie est nulle. Elle est maximale quand tous les états ont la même probabilité  $\frac{1}{n}$ .

2-a) Observons que  $\ln Z(0) = \ln n$ .

2-b) On a  $\frac{\mu_\beta(\omega)}{\mu_\beta(\omega')} \xrightarrow{\beta \rightarrow +\infty} 0$  si  $U(\omega) > U(\omega')$ . Nous en déduisons que quand  $\beta$  tend vers l'infini,  $\mu_\beta$  devient une probabilité uniforme sur l'ensemble  $\Omega_{\min} := \{\omega; U(\omega) = \min_{\Omega} U\}$ .

De même que lorsque  $\beta \rightarrow -\infty$ ,  $\mu_\beta$  devient une probabilité uniforme sur  $\Omega_{\max} := \{\omega; U(\omega) = \max_{\Omega} U\}$ . Les limites (3.25) s'en suivent immédiatement.

2-c)  $\beta \rightarrow Z(\beta)$  est non nulle et de classe  $C^\infty$  sur  $\mathbb{R}$ , il en est donc de même pour  $\beta \rightarrow \ln Z(\beta)$ . On calcule

$$\begin{aligned} (\ln Z)'(\beta) &= \frac{Z'(\beta)}{Z(\beta)} = -\frac{1}{Z(\beta)} \sum_{i=1}^n U(\omega_i) e^{-\beta U(\omega_i)} = -\langle U \rangle_{\mu_\beta} \\ (\ln Z)''(\beta) &= \frac{Z''(\beta)}{Z(\beta)} - \left( \frac{Z'(\beta)}{Z(\beta)} \right)^2 = \frac{1}{Z(\beta)} \sum_i U^2(\omega_i) e^{-\beta U(\omega_i)} - (\langle U \rangle_{\mu_\beta})^2 \\ &= \text{Var}_{\mu_\beta}(U) \geq 0. \end{aligned}$$

Ainsi,  $\beta \mapsto \ln Z(\beta)$  est convexe et strictement convexe si et seulement si la variance ne s'annule pas, c'est-à-dire si  $U$  n'est pas constante.

3-a) Nous déduisons de la question précédente que  $\beta \mapsto \langle U \rangle_{\mu_\beta}$  est strictement décroissante. Comme elle est continue, elle est donc bijective. (3.25) nous permet de conclure.

3-b) On a  $H(\mu_\beta) = -\sum_i \mu_\beta(\omega_i) \ln \mu_\beta(\omega_i)$ .

Recherchons les extrémums de  $F : \frac{\partial F}{\partial \eta_i} = -\ln \eta_i - 1 - \beta U(\omega_i) - \lambda = 0$  donne  $\eta(\omega_i) = e^{-\beta U(\omega_i) - \lambda - 1}$ . On détermine  $\lambda$  par la condition que  $\eta$  est une probabilité :

$$\lambda + 1 = \ln \sum_{i=1}^n e^{-\beta U(\omega_i)} = \ln Z(\beta).$$

Ainsi, si  $\mu_\beta$  maximise l'entropie, on aura

$$\ln \mu_\beta(\omega_i) = -\beta U(\omega_i) - \lambda - 1 = -\beta U(\omega_i) - \ln Z(\beta),$$

d'où le résultat.

3-c) Nous allons utiliser l'inégalité de convexité pour la fonction  $\phi$ .

$$\begin{aligned} & -\sum_i p_i \ln p_i + \sum_i (-\beta U(\omega_i) p_i) - \ln Z(\beta) \\ &= -\sum_i p_i \ln p_i + \sum_i \left( p_i \ln e^{-\beta U(\omega_i)} \right) - \sum_i p_i \ln Z(\beta) \\ &= -\sum_i p_i \ln \frac{p_i}{\mu_\beta(\omega_i)} \\ &= -\sum_i \mu_\beta(\omega_i) \frac{p_i}{\mu_\beta(\omega_i)} \ln \frac{p_i}{\mu_\beta(\omega_i)} = \sum_i \mu_\beta(\omega_i) \phi \left( \frac{p_i}{\mu_\beta(\omega_i)} \right) \\ &\leq \phi \left( \sum_i \mu_\beta(\omega_i) \frac{p_i}{\mu_\beta(\omega_i)} \right) = \phi(1) = 0. \end{aligned}$$

Nous avons donc montré que pour toute proba  $p$  :

$$H(p) + \langle (-\beta U) \rangle_p \leq \ln Z(\beta).$$

On sait que  $\mu_\beta$  atteint le maximum de  $p \mapsto H(p) + \langle (-\beta U) \rangle_p$  qui vaut  $\ln Z(\beta)$ . Puisque  $\phi$  est strictement concave, l'inégalité de convexité n'est saturée que si  $p_i/\mu_\beta(\omega_i)$  est une constante, qui ne peut être autre que 1. Donc  $p_i = \mu_\beta(\omega_i)$  et  $\mu_\beta$  est l'unique probabilité qui maximise l'entropie étant donnée l'espérance  $E$ .

Ainsi, nous avons démontré le “principe variationnel” : pour tout  $\beta \in \mathbb{R}$ ,

$$H(\mu_\beta) + \langle (-\beta U) \rangle_{\mu_\beta} = \ln Z(\beta) = \sup_{p \text{ probabilité}} (H(p) + \langle (-\beta U) \rangle_p).$$

De plus,  $\mu_\beta$  est l'unique probabilité qui atteint le maximum.

## 11.3 Corrigés des exercices du chapitre 4

*Exercice 4.1* : Il suffit d'écrire

$$\mathbb{E}((X - a)^2) = \mathbb{E}((X - \mathbb{E}(X))^2) + (\mathbb{E}(X) - a)^2.$$

## 11.4 Corrigés des exercices du chapitre 5

*Exercice 5.5* : Puisque  $g$  est de classe  $C^1$ , nous avons  $g(x) = \int_0^x g'(t) dt$ . Alors

$$\mathbb{E}(g(X)) = \mathbb{E} \left( \int_0^X g'(t) dt \right) = \int_0^\infty \left( \int_0^x g'(t) dt \right) f(x) dx,$$

où  $f$  est la densité de  $X$ . Alors en appliquant le théorème de Fubini, nous obtenons

$$\mathbb{E}(g(X)) = \int_0^\infty g'(t) \left( \int_0^\infty \mathbf{1}_{\{t < x\}} f(x) dx \right) dt = \int_0^\infty g'(t) \mathbb{P}(X > t) dt.$$

Remarquons que pour  $g(t) = t$ , nous obtenons  $\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > t) dt$ .

En fait, cette propriété est vraie pour toute variable aléatoire  $X$  positive. Il faut pour cela appliquer le théorème de Fubini à la mesure abstraite  $\mathbb{P}(d\omega) \otimes dt$ .

*Exercice 5.6* : 1) La position de l'incendie  $X$  suit une loi uniforme sur  $[0, A]$ . Nous cherchons à placer  $a$  tel que  $\mathbb{E}(|X - a|) = \frac{1}{A} \int_0^A |x - a| dx$  soit minimum. Il est immédiat de montrer que ce minimum vaut  $a = \frac{A}{2}$ .

2) Nous devons alors minimiser la quantité  $\int_0^\infty \lambda e^{-\lambda x} |x - a| dx$ . Tous calculs faits, nous obtenons que  $a = \frac{\ln 2}{\lambda}$ .

*Exercice 5.7* : Rappel sur la fonction  $\Gamma$  :  $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$  est convergente si et seulement si  $\alpha > 0$ . De plus  $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ , d'où nous déduisons que pour  $n$  entier,  $\Gamma(n + 1) = n!$ .

Ainsi, si  $X$  est variable aléatoire de densité  $Cx^{\alpha-1} e^{-px}$  sur  $\mathbb{R}_+$ , pour  $\alpha, p > 0$ , alors  $C = \frac{p^\alpha}{\Gamma(\alpha)}$ .

Nous pouvons utiliser ces préliminaires pour montrer que si  $X$  est la durée de fonctionnement de la lampe, alors

1)  $f$  est une densité de probabilité,

2)  $\mathbb{E}(X) = 8$ ,  $\text{Var}(X) = 32$ .

3) Par changement de variable et intégration par parties, nous obtenons également que :  $\mathbb{P}(X \geq 6) = \frac{5}{2} e^{-\frac{3}{2}}$ .

*Exercice 5.8* : Remarquons que  $Y \in [0, c]$ . De plus,  $\mathbb{P}(Y = 0) = \mathbb{P}(X < 0)$ . Donc, si  $\mathbb{P}(X < 0) \neq 0$ ,  $Y$  ne peut pas avoir de densité. Si  $\mathbb{P}(X < 0) = 0$ , introduisons une fonction  $g$  continue et bornée sur  $[0, c]$ . Alors

$$\mathbb{E}(g(Y)) = \int_0^\infty g(ce^{-\alpha x}) f_X(x) dx = \int_0^c g(y) f_X\left(-\frac{1}{\alpha} \ln \frac{y}{c}\right) \frac{1}{\alpha y} dy,$$

grâce au changement de variable :  $y = ce^{-\alpha x} \iff x = -\frac{1}{\alpha} \ln \frac{y}{c}$ , pour  $x \in \mathbb{R}_+$  et  $y \in [0, c]$ . La densité de  $Y$ , dans ce cas, vaut donc

$$f_Y(y) = \frac{1}{\alpha y} f_X\left(-\frac{1}{\alpha} \ln \frac{y}{c}\right) \mathbf{1}_{]0, c[}(y).$$

*Exercice 5.9* : 1)

$$\mathbb{E}(e^{\lambda X}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\lambda x} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\lambda)^2}{2}} e^{\frac{\lambda^2}{2}} dx = e^{\frac{\lambda^2}{2}}.$$

2) Puisque  $e^{\lambda X} \geq e^{\lambda a} \mathbf{1}_{X \geq a}$ , nous avons

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(e^{\lambda X})}{e^{\lambda a}} \leq e^{\frac{\lambda^2}{2} - \lambda a},$$

pour tout  $\lambda > 0$ . En minimisant le terme de droite en  $\lambda$ , (minimum atteint en  $\lambda = a$ ), nous obtenons finalement que

$$\mathbb{P}(X \geq a) \leq e^{-\frac{a^2}{2}}.$$

3)

$$\mathbb{P}(X \geq a) = \frac{1}{\sqrt{2\pi}} \int_a^{+\infty} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_a^{+\infty} \frac{1}{x} x e^{-\frac{x^2}{2}} dx.$$

Une intégration par parties donne

$$\int_a^{+\infty} \frac{1}{x} x e^{-\frac{x^2}{2}} dx = \frac{1}{a} e^{-\frac{a^2}{2}} - \int_a^{+\infty} \frac{1}{x^2} e^{-\frac{x^2}{2}} dx.$$

Or

$$\int_a^{+\infty} \frac{1}{x^2} e^{-\frac{x^2}{2}} dx = \int_a^{+\infty} x \frac{1}{x^3} e^{-\frac{x^2}{2}} dx \leq \frac{1}{a^3} \int_a^{+\infty} x e^{-\frac{x^2}{2}} dx = \frac{1}{a^3} e^{-\frac{a^2}{2}}.$$

Nous en déduisons l'inégalité voulue.

*Exercice 5.10* : 1) Il est immédiat de vérifier que  $f$  est positive et d'intégrale 1.

2)  $\mathbb{P}(X > x) = \left(\frac{a}{x}\right)^\alpha$  si  $x > a$  et  $\mathbb{P}(X > x) = 1$  si  $x \leq a$ .

3) Soit  $X > a$ . Alors

$$\mathbb{P}(X > x + y | X > x) = \left(\frac{a}{x+y}\right)^\alpha \left(\frac{x}{a}\right)^\alpha = \left(\frac{x}{x+y}\right)^\alpha.$$

Cette quantité tend vers 1 quand  $x$  tend vers l'infini.

Cela veut dire que plus  $X$  prend de grandes valeurs, plus elle a de chances d'en prendre de plus grandes. Cela n'est pas vrai pour la loi exponentielle qui n'a pas de mémoire. Cette loi fut introduite par le marquis de Pareto comme modèle de richesse, celui-ci ayant remarqué au début du 20ème siècle que 20% de la population possédait 80% des richesses. D'autres phénomènes ont ce même type de propriété : pour un service, 20% des clients sont responsables de 80% des réclamations ; pour une activité sportive, 20% d'entraînement supplémentaire peut amener 80% de performance en plus.

4)

$$\mathbb{E}(X) = \alpha a^\alpha \int_a^{+\infty} \frac{x}{x^{\alpha+1}} dx < +\infty \iff \alpha > 1.$$

Dans ce cas,  $\mathbb{E}(X) = \frac{\alpha a}{\alpha - 1}$ .

5) De même,

$$\mathbb{E}(X^2) < +\infty \iff \alpha > 2.$$

Dans ce cas,  $\mathbb{E}(X^2) = \frac{\alpha a^2}{\alpha-2}$  et  $\text{Var}(X) = \frac{\alpha a^2}{(\alpha-2)(\alpha-1)^2}$ .

*Exercice 5.11* : 1) Loi de  $X$  : loi de densité

$$f_X(x) = \int f(x, y) dy = \frac{1}{2\pi} |x| e^{-\frac{x^2}{2}} \int e^{-\frac{x^2 y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

$X$  suit une loi normale centrée réduite.

Un calcul analogue montre que la densité de  $Y$  vaut  $f_Y(y) = \frac{1}{\pi} \frac{1}{1+y^2}$ .  $Y$  suit donc une loi de Cauchy. En particulier,  $Y$  n'a pas d'espérance.

Puisque le produit de  $f_X$  et  $f_Y$  n'est clairement pas égal à  $f$ , les deux variables aléatoires  $X$  et  $Y$  ne sont pas indépendantes.

2) Soit  $g$  une fonction continue et bornée sur  $\mathbb{R}^2$ . On a  $\mathbb{E}(g(X, XY)) = \int g(x, xy) f(x, y) dx dy$ . Considérons le changement de variable  $(x, y) \mapsto (x, z = xy)$ . Le jacobien vaut  $x$ . Ainsi,  $\mathbb{E}(g(X, XY)) = \frac{1}{2\pi} \int g(x, z) e^{-\frac{x^2}{2}} e^{-\frac{z^2}{2}} dx dz$ .  $X$  et  $XY$  sont indépendantes, et  $Z = XY$  suit une loi normale centrée réduite.

3)  $X(1+Y) = X + Z$ , avec  $X$  et  $Z$  indépendantes. La somme de deux variables aléatoires indépendantes de loi normale  $\mathcal{N}(0, 1)$  suit une loi normale  $\mathcal{N}(0, 2)$ .

*Exercice 5.12* : 1) Avec les notations du cours, la densité  $f$  du couple  $(X, Y)$  vaut

$$f(x, y) = f_{X|Y=y} f_Y(y) = \mathbf{1}_{\mathbb{R}_+}(x) \frac{1}{y^2} \mathbf{1}_{[1, +\infty[}(y) x e^{-xy}.$$

Soit  $g$  une fonction continue bornée sur  $\mathbb{R}_+^2$ . Le changement de variable  $(x, y) \mapsto (t = xy, y)$ , de jacobien  $y$ , donne alors que  $\mathbb{E}(g(T, Y)) = \mathbb{E}(g(XY, Y)) = \int_1^\infty \int_0^\infty g(t, y) e^{-t} \frac{t}{y^2} dt dy$ . Comme la densité du couple  $(T, Y)$ , qui vaut  $e^{-t} \frac{t}{y^2} \mathbf{1}_{\mathbb{R}_+}(t) \mathbf{1}_{[1, +\infty[}(y)$  s'écrit comme produit d'une fonction de  $t$  et d'une fonction de  $y$ , nous en déduisons que  $T$  et  $Y$  sont indépendantes et que  $T$  a la densité  $te^{-t} \mathbf{1}_{\mathbb{R}_+}(t)$ .

2) La loi de  $X$  a pour densité  $f_X(x) = \int_1^\infty e^{-xy} x dy = e^{-x}$ . Ainsi  $X$  suit une loi exponentielle de paramètre 1 et la loi conditionnelle de  $Y$  sachant  $X = x$  admet la densité  $f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)} = x e^{-x(y-1)} \mathbf{1}_{[1, +\infty[}(y)$ , pour  $x > 0$ .

3) Nous en déduisons que  $\mathbb{E}(Y|X = x) = \int_1^\infty y x e^{-x(y-1)} dy = \left(\frac{x+1}{x}\right) \mathbf{1}_{\mathbb{R}_+}(x)$ . Ainsi,  $E(Y|X) = \frac{X+1}{X}$ .

*Exercice 5.13 :*

Soit  $h$  une fonction continue bornée sur  $\mathbb{R}_+$ . Comme  $X$  et  $Y$  sont indépendantes, nous avons

$$\mathbb{E}(h(XY)) = \sum_n \int_0^1 h(ny) \mathbb{P}(X = n) dy = \sum_n \mathbb{P}(X = n) \int_0^1 h(ny) dy.$$

Pour  $n \neq 0$ , posons  $z = ny$ . Alors

$$\begin{aligned} \mathbb{E}(h(XY)) &= h(0) \mathbb{P}(X = 0) + \sum_n \mathbb{P}(X = n) \frac{1}{n} \int_0^n h(z) dz \\ &= h(0) \mathbb{P}(X = 0) + \int \left( \sum_n \frac{\mathbb{P}(X = n)}{n} \mathbf{1}_{[0, n]}(z) \right) h(z) dz. \end{aligned}$$

Ainsi,  $Z$  admet une densité si et seulement si  $\mathbb{P}(X = 0) = 0$ .

*Exercice 5.14 :*

Loi de  $X$  :

$$\mathbb{P}(X = n) = \beta \int_0^\infty e^{-(\alpha+\beta)y} \frac{(\alpha y)^n}{n!} dy = \frac{\alpha^n}{n!} \beta \frac{\Gamma(n+1)}{(\alpha+\beta)^{n+1}} = \frac{\beta}{\alpha+\beta} \left( \frac{\alpha}{\alpha+\beta} \right)^n.$$

Ainsi,  $X$  suit une loi géométrique de paramètre  $\frac{\beta}{\alpha+\beta}$ .

Loi de  $Y$  :

$$\mathbb{P}(Y \leq y) = \sum_{n \geq 0} \beta \int_0^y e^{-(\alpha+\beta)u} \frac{(\alpha u)^n}{n!} du = \beta \int_0^y e^{-\beta u} du.$$

Ainsi,  $Y$  suit une loi exponentielle de paramètre  $\beta$ .

*Exercice 5.15 :* 1) Soit  $h$  une fonction continue bornée sur  $[0, 1]^n$ . Alors  $\mathbb{E}(h(U_1, \dots, U_n)) = \int_0^1 \dots \int_0^1 h(x_1, x_1 x_2, \dots, x_1 \dots x_n) dx_1 \dots dx_n$ . On pose  $u_1 = x_1, u_2 = x_1 x_2, \dots, u_n = x_1 \dots x_n$ . Le jacobien de cette transformation vaut  $u_1 \dots u_{n-1}$ . Nous en déduisons que  $\mathbb{E}(h(U_1, \dots, U_n)) = \int h(u_1, \dots, u_n) \mathbf{1}_{0 \leq u_1 \leq \dots \leq u_n \leq 1} \frac{1}{u_1 \dots u_{n-1}} du_1 \dots du_n$ . La loi de  $(U_1, \dots, U_n)$  a donc la densité  $\mathbf{1}_{0 \leq u_1 \leq \dots \leq u_n \leq 1} \frac{1}{u_1 \dots u_{n-1}}$ .

2)  $U_n = U_{n-1} X_n$ , et  $U_{n-1}$  et  $X_n$  sont indépendantes. D'où si  $g$  est une fonction continue bornée sur  $[0, 1]^2$ ,

$$\begin{aligned} \mathbb{E}(g(U_n, U_{n-1})) &= \int g(u_{n-1} x_n, u_{n-1}) f_{X_n}(x_n) f_{U_{n-1}}(u_{n-1}) du_{n-1} dx_n \\ &= \int_0^1 \int_0^u g(z, u) f_{U_{n-1}}(u) \frac{1}{u} dz du. \end{aligned}$$

Ainsi, la loi conditionnelle de  $U_n$  sachant  $U_{n-1} = u$  aura la densité  $z \mapsto \frac{1}{u} \mathbf{1}_{z \leq u}$ .

*Exercice 5.16* : 1) Soit  $f$  une fonction continue bornée sur  $\mathbb{R}_+ \times [0, 2\pi[$ . Alors  $\mathbb{E}(f(R, \Theta)) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(r, \theta) e^{-\frac{r^2+\theta^2}{2}} dx dy = \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} f(r, \theta) e^{-\frac{r^2}{2}} r dr d\theta$ . D'où  $R$  admet la loi de densité  $re^{-\frac{r^2}{2}} \mathbf{1}_{r>0}$  et  $\Theta$  la loi uniforme sur  $[0, 2\pi]$ , et  $R$  et  $\Theta$  sont indépendants.

$R^2$  et  $\Theta$  sont indépendants, et si  $h$  est une fonction continue bornée sur  $\mathbb{R}_+$ ,  $\mathbb{E}(h(R^2)) = \int_0^\infty h(r^2) r e^{-r^2/2} dr = \frac{1}{2} \int_0^\infty h(z) e^{-z/2} dz$ .  $R^2$  suit une loi exponentielle de paramètre  $\frac{1}{2}$ .

2) Nous avons vu que  $R^2$  peut se simuler en prenant  $-2 \ln U$ , où  $U$  suit une loi uniforme sur  $[0, 1]$ .  $\Theta$  va être simulée par  $2\pi V$ , où  $V$  suit une loi uniforme sur  $[0, 1]$ , et  $U$  et  $V$  sont indépendants (comme  $R^2$  et  $\Theta$ ). Alors  $X = R \cos \Theta = \sqrt{-2 \ln U} \cos(2\pi V)$  et  $Y = \sqrt{-2 \ln U} \sin(2\pi V)$ .

3) Soit  $g$  une fonction continue bornée sur  $\mathbb{R}_+$ .  $\mathbb{E}(g(\frac{Y}{X})) = \mathbb{E}(g(\tan \Theta)) = \frac{1}{2\pi} \int_0^{2\pi} g(\tan \theta) d\theta = \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} g(\tan \theta) d\theta$ . La fonction  $\theta \mapsto \tan \theta$  est inversible sur  $]-\frac{\pi}{2}, \frac{\pi}{2}[$ . On obtient  $\mathbb{E}(g(\frac{Y}{X})) = \frac{1}{\pi} \int_{\mathbb{R}} f(t) \frac{dt}{1+t^2}$ . Donc  $\frac{Y}{X}$  suit une loi de Cauchy.

*Exercice 5.17* : 1) Soit  $x \leq y$ .

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(Z_k \leq y, \forall k) - \mathbb{P}(x < Z_k \leq y, \forall k) = F(y)^n - (F(y) - F(x))^n,$$

par indépendance des  $Z_k$ . Ainsi,  $F_X(x) = 1 - (1 - F(x))^n$ , et  $F_Y(y) = F(y)^n$ .

2)  $F$  est dérivable de dérivée  $f$ . Ainsi,  $X$  admet une densité qui vaut  $f_X(x) = F'_X(x) = n(1 - F(x))^{n-1} f(x)$ , et de même  $Y$  admet la densité  $f_Y(y) = n(F(y))^{n-1} f(y)$ .

La densité du couple  $(X, Y)$  vaut  $h(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$ , où  $F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ . Ainsi, nous obtenons  $h(x, y) = n(n-1)(F(Y) - F(x))^{n-2} f(x) f(y)$ .

3)  $F(x) = x$ . Nous avons donc  $f_X(x) = \frac{n}{(b-a)^n} (b-x)^{n-1} \mathbf{1}_{]a, b[}(x)$ ,  $f_Y(y) = \frac{n}{(b-a)^n} (y-a)^{n-1} \mathbf{1}_{]a, b[}(y)$ ,  $h(x, y) = \frac{n(n-1)}{(b-a)^n} (y-x)^{n-2} \mathbf{1}_{a \leq x \leq y \leq b}$ .

Les calculs donnent  $\mathbb{E}(X) = a + \frac{b-a}{n+1}$ ,  $\mathbb{E}(Y) = b - \frac{b-a}{n+1}$ ,  $\text{Var}(X) = \text{Var}(Y) = \frac{n(b-a)^2}{(n+1)^2(n+2)}$   
 $\text{Cov}(X, Y) = \frac{(b-a)^2}{(n+1)^2(n+2)}$ , et  $\rho(X, Y) = \frac{1}{n}$ . Ainsi, plus  $n$  est grand, moins les variables  $X$  et  $Y$  sont corrélées, ce qui est raisonnable!



Exercice 5.18 : 1)

$$\mathbb{P}(M > a, D > b, X > Y) = \lambda\mu \int_{y>a} \int_{x>y+b} e^{-\lambda x} e^{-\mu y} dx dy = \frac{\mu}{\lambda + \mu} e^{-\lambda b} e^{-(\lambda+\mu)a}.$$

2)  $\mathbb{P}(X > Y) = \mathbb{P}(M > 0, D > 0, X > Y) = \frac{\mu}{\lambda + \mu}$ ,  $\mathbb{P}(X < Y) = \frac{\lambda}{\lambda + \mu}$ ,  $\mathbb{P}(M > a, X > Y) = \frac{\mu}{\lambda + \mu} e^{-(\lambda+\mu)a}$ ,  $\mathbb{P}(M > a, X < Y) = \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)a}$ . Ainsi,  $M$  suit une loi exponentielle de paramètre  $\lambda + \mu$ . De plus,

$$\mathbb{P}(D > b | X > Y) = \frac{\mathbb{P}(M > 0, D > b, X > Y)}{\mathbb{P}(X > Y)} = e^{-\lambda b}.$$

La loi conditionnelle de  $D$  sachant  $X > Y$  est donc une loi exponentielle de paramètre  $\lambda$ . De même, la loi conditionnelle de  $D$  sachant  $X < Y$  est une loi exponentielle de paramètre  $\mu$ .

3) Nous remarquons que  $\mathbb{P}(M > a, X > Y) = \mathbb{P}(M > a)\mathbb{P}(X > Y)$ .

4) Nous pouvons montrer par récurrence que  $\min_{1 \leq i \leq n} X_i$  suit une loi exponentielle de paramètre  $\sum_{1 \leq i \leq n} \lambda_i$ , que  $\mathbb{P}(X_k = \min_{1 \leq i \leq n} X_i) = \frac{\lambda_k}{\sum_{1 \leq i \leq n} \lambda_i}$ , et que  $\min_{1 \leq i \leq n} X_i$  et  $\{X_k = \min_{1 \leq i \leq n} X_i\}$  sont indépendants.

## 11.5 Corrigés des exercices du chapitre 6

Exercice 6.1 : 1)  $\mathbb{E}(X_n) = 1$ .

2) Puisque  $\sum_{n \geq 1} u_n < +\infty$ , la suite  $(u_n)_n$  tend vers 0 et donc  $\frac{1}{u_n}$  tend vers  $+\infty$ . Ainsi, pour  $\varepsilon$  fixé et  $n$  assez grand,  $\mathbb{P}(X_n > \varepsilon) = \mathbb{P}(X_n = \frac{1}{u_n}) = u_n$  qui tend vers 0. Donc  $X_n \xrightarrow{P} 0$ .

Les ensembles  $A_n = \{|X_n| > \varepsilon\}$  vérifient que  $\sum_n \mathbb{P}(A_n) < \infty$ , par hypothèse. Alors, par la première partie du lemme de de Borel-Cantelli (Théorème 2.35),  $\mathbb{P}(\limsup_n A_n) = 0$ . Ainsi presque-sûrement, au plus un nombre fini de  $A_n$  sont réalisés. Il en résulte que  $X_n \xrightarrow{\text{p.s.}} 0$ .

3) Nous en déduisons par un argument de limite de Césaro que  $\frac{X_1 + \dots + X_n}{n} \xrightarrow{\text{p.s.}} 0$  quand  $n$  tend vers l'infini.

Cé résultat n'est pas en contradiction avec la loi des grands nombres, bien que  $\mathbb{E}(X_n) = 1$ . En effet, les variables aléatoires  $X_n$  n'ont pas les mêmes lois, donc nous ne sommes pas sous les hypothèses de la LGN.

*Exercice 6.2 :* Considérons un réel  $M > 0$  donné. Alors les variables aléatoires bornées  $X_i \wedge M$  satisfont les hypothèses de la loi des grands nombres. Nous en déduisons que  $\frac{X_1 \wedge M + \dots + X_n \wedge M}{n} \xrightarrow{\text{P.S.}} \mathbb{E}(X_1 \wedge M)$ .

Soit maintenant une suite  $M_k$  de réels qui tend vers l'infini. Pour chaque  $k$ , il existe un ensemble négligeable  $N_k$  en dehors duquel pour tout  $\omega$ ,  $\frac{X_1(\omega) \wedge M_k + \dots + X_n(\omega) \wedge M_k}{n} \rightarrow \mathbb{E}(X_1 \wedge M_k)$ . Alors  $N = \cup_k N_k$  est encore négligeable (comme réunion dénombrable d'ensembles négligeables), et pour  $\omega \notin N$ ,  $\forall k$ ,  $\frac{X_1(\omega) \wedge M_k + \dots + X_n(\omega) \wedge M_k}{n} \rightarrow \mathbb{E}(X_1 \wedge M_k)$ . Nous pouvons alors faire tendre  $k$  vers l'infini, et nous en déduisons que pour  $\omega \notin N$ ,  $\frac{X_1(\omega) + \dots + X_n(\omega)}{n} \rightarrow \mathbb{E}(X_1) = +\infty$ , d'où la convergence presque-sure de  $\frac{X_1 + \dots + X_n}{n}$  vers  $+\infty$ .

*Exercice 6.3 :* 1) La loi des grands nombres entraîne que  $M_n = \frac{X_1 + \dots + X_n}{n} \rightarrow_{p.s.} x$  et donc, puisque  $f$  est continue,

$$f\left(\frac{X_1 + \dots + X_n}{n}\right) \rightarrow_{p.s.} f(x).$$

$f$  étant continue sur  $[0, 1]$ , elle est bornée. Ainsi, les variables aléatoires  $f\left(\frac{X_1 + \dots + X_n}{n}\right)$  le sont également. Nous pouvons alors appliquer le théorème de convergence dominée, et

$$\mathbb{E}\left(f\left(\frac{X_1 + \dots + X_n}{n}\right)\right) = \sum_{k=1}^n f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} \rightarrow f(x).$$

On appelle les polynômes  $P_n(x) = \sum_{k=1}^n f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}$  polynômes de Bernstein.

2)  $f$  est uniformément continue sur  $[0, 1]$  :  $\forall \varepsilon > 0$ ,  $\exists \alpha$ , tel que  $\forall x, y \in [0, 1]$ ,  $|x - y| < \alpha \Rightarrow |f(x) - f(y)| < \varepsilon$ . Nous avons alors

$$\begin{aligned} & |\mathbb{E}(f(M_n)) - f(x)| \\ & \leq \mathbb{E}(|f(M_n) - f(x)| \mathbf{1}_{\{|M_n - x| \geq \alpha\}}) + \mathbb{E}(|f(M_n) - f(x)| \mathbf{1}_{\{|M_n - x| < \alpha\}}) \\ & \leq 2\|f\|_\infty \mathbb{P}(|M_n - x| \geq \alpha) + \varepsilon \\ & \leq 2\|f\|_\infty \frac{\text{Var}(X_1)}{n\alpha^2} + \varepsilon \leq \frac{\|f\|_\infty}{2n\alpha^2} + \varepsilon, \end{aligned}$$

par l'inégalité de Bienaymé-Chebyshev, puisque  $\text{Var}(X_1) = x(1-x) \leq \frac{1}{4}$ .

Nous avons donc prouvé : Toute fonction continue sur  $[0, 1]$  est approchée uniformément par une suite de polynômes. (Théorème de Weierstrass)

*Exercice 6.4 :* 1)  $M(u) = e^{u^2 \sigma^2 / 2}$ .

2) Nous avons :  $e^{u(S_n - nm)} \geq e^{ua} \mathbf{1}_{S_n - nm \geq a}$ , d'où

$$\mathbb{P}(S_n - nm \geq a) \leq \frac{\mathbb{E}(e^{u(S_n - nm)})}{e^{ua}} = e^{-ua} (M(u))^n,$$

par indépendance des  $X_i$ .

3)

$$\begin{aligned} \mathbb{P}(|Y_n - m| \geq \varepsilon) &= \mathbb{P}(S_n - nm \geq n\varepsilon) + \mathbb{P}(S_n - nm \leq -n\varepsilon) \\ &\leq e^{-nu\varepsilon} (M(u))^n + e^{-n(-u)(-\varepsilon)} (M(-u))^n = 2e^{-nu\varepsilon} (M(u))^n \\ &\leq 2e^{-nu\varepsilon} e^{nu^2\sigma^2/2}, \quad \forall u \geq 0. \end{aligned}$$

La meilleure majoration va être obtenue en minimisant l'exposant, c'est-à-dire pour  $u = \varepsilon$ . Nous en déduisons l'inégalité de Chernov.

4) Par l'inégalité de Bienaymé-Chebyshev,  $\mathbb{P}(|Y_n - m| \geq \varepsilon) \leq \frac{\text{Var}(Y_n)}{\varepsilon^2}$ . Ainsi,  $\mathbb{P}(|Y_n - m| \leq \varepsilon) \geq \alpha$  dès que  $1 - \frac{\sigma^2}{n\varepsilon^2} \leq 1 - \alpha = 0,05$ . Avec  $\varepsilon = 0,05$ , il vient que  $n \geq 80000$ .

Par l'inégalité de Chernov, nous obtenons  $2e^{-\frac{n\varepsilon^2}{20}} \leq 0,05$ . Il vient que  $n \geq 29512$ . Pour avoir une évaluation du même ordre de la probabilité cherchée, nous pouvons donc prendre un échantillon beaucoup plus petit si nous utilisons l'inégalité de Chernov. A taille d'échantillon fixée, nous aurons une meilleure évaluation avec cette inégalité.

*Exercice 6.5 :*

$$S_n = S_0(1 + R_1) \cdots (1 + R_n) = S_0 \exp\left(n \frac{1}{n} \sum_{k=1}^n \ln(1 + R_k)\right).$$

Or,  $R_1 > -1$ , donc  $-1 < \mathbb{E}(R_1) < \infty$ , et  $\ln(1 + \mathbb{E}(R_1)) < \infty$ . Par l'inégalité de Jensen, nous en déduisons que  $\mathbb{E}(\ln(1 + R_1)) \leq \ln(1 + \mathbb{E}(R_1)) < \infty$ . Nous pouvons alors appliquer la loi des grands nombres :  $\frac{1}{n} \sum_{k=1}^n \ln(1 + R_k) \rightarrow \mathbb{E}(\ln(1 + R_1))$ . Ainsi, pour  $n$  assez grand,  $S_n$  va se comporter presque-sûrement comme  $S_0 \exp(n\mathbb{E}(\ln(1 + R_1)))$ .

Si  $\mathbb{E}(R_1) = 0$ , (et  $R_1$  non identiquement nulle), alors  $\mathbb{E}(\ln(1 + R_1)) < 0$ , et  $S_n$  va décroître exponentiellement vite vers 0.

*Exercice 6.6 :* Les deux premières questions se résolvent comme dans l'exercice 6.3.

3) Supposons que les  $X_i$  suivent des lois uniformes sur  $[0, 1]$ . Alors  $a = \frac{1}{2}$ , et  $\lim_n I_n = g(\frac{1}{2})$ .

4) Le résultat sur la somme de variables aléatoires indépendantes de loi exponentielle se démontre par récurrence. En appliquant ce résultat, nous aurons donc

$$f(a) = \lim_n \frac{1}{a^n(n-1)!} \int_0^\infty f\left(\frac{z}{n}\right) z^{n-1} e^{-\frac{z}{a}} dz = \lim_n \frac{1}{a^n(n-1)!} \int_0^\infty f(t) n^{n-1} t^{n-1} e^{-\frac{nt}{a}} ndt.$$

En posant  $F(t) = \int_0^\infty f(x)e^{-tx}dx$ , nous obtenons  
 $F^{(n-1)}(\frac{n}{a}) = (-1)^{n-1} \int_0^\infty f(t)t^{n-1}e^{-\frac{nt}{a}}dt$ . D'où le résultat.

## 11.6 Corrigés des exercices du chapitre 7

*Exercice 7.2* : Soit  $t$  un réel. Alors  $\phi(t) = \phi_{\frac{X+Y}{\sqrt{2}}}(t) = \phi_{X+Y}(\frac{t}{\sqrt{2}}) = \left(\phi\left(\frac{t}{\sqrt{2}}\right)\right)^2$ .  
 Par récurrence, nous en déduisons que pour tout  $n$ ,  $\phi(t) = \left(\phi\left(\frac{t}{\sqrt{2^n}}\right)\right)^{2^n} = \exp\left(2^n \ln \phi\left(\frac{t}{\sqrt{2^n}}\right)\right)$ . Comme  $X$  est centrée et a un moment d'ordre 2,  $\sigma^2 = \mathbb{E}(X^2)$ ,  $\phi$  est deux fois dérivable en 0 et

$$\phi\left(\frac{t}{\sqrt{2^n}}\right) = 1 - \frac{\sigma^2}{2} \frac{t^2}{2^n} + o\left(\frac{t^2}{2^n}\right).$$

Nous en déduisons que  $\lim_{n \rightarrow \infty} \left(\phi\left(\frac{t}{\sqrt{2^n}}\right)\right)^{2^n} = e^{-\sigma^2 t^2/2}$ . Ainsi,  $\phi(t) = e^{-\sigma^2 t^2/2}$ , et  $X$  et  $Y$  suivent des lois  $\mathcal{N}(0, \sigma^2)$ .

*Exercice 7.3* : 1)  $\phi_X(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{e^{itx}}{1+x^2} dx$ . Nous appliquons le théorème des résidus.  
 Nous considérons la fonction de variable complexe  $f(z) = \frac{e^{itz}}{\pi(1+z^2)} = \frac{e^{itz}}{2i\pi} \left(\frac{1}{z-i} - \frac{1}{z+i}\right)$  qui a les deux pôles  $i$  et  $-i$ . Pour  $t > 0$ , prenons comme contour  $\Gamma$  le demi-cercle supérieur de centre 0 et de rayon  $R$ , qui encercle  $i$ . Alors la formule des résidus donne  $\int_\Gamma f(z)dz = 2i\pi \text{Res}(f, i)$ . Nous avons  $\text{Res}(f, i) = \frac{e^{-t}}{2i\pi}$ , d'où

$$e^{-t} = \int_{-R}^R \frac{e^{itx}}{1+x^2} dx + \int_{\Gamma \setminus [-R, R]} f(z) dz.$$

Mais sur ce contour,  $f(z) = \frac{e^{itR \cos \theta} e^{-tR \sin \theta}}{1+z^2}$  qui tend vers 0 quand  $R$  tend vers l'infini (car  $\sin \theta > 0$ ). Pour  $t < 0$ , nous prenons l'autre contour et nous obtenons  $e^t$ . Ainsi, la fonction caractéristique d'une variable aléatoire de Cauchy vaut  $\phi_X(t) = e^{-|t|}$ .

2)  $\phi_{2X}(t) = e^{-|2t|} = (\phi_X(t))^2$ .

*Exercice 7.4* : 1)  $\phi_X(t) = \int_{-\infty}^{+\infty} e^{itx} \frac{a}{2} e^{-|x|^a} dx = \frac{a^2}{a^2+t^2}$ .

2)  $Y$  a une densité qui ne charge que  $y \geq 1$ . Donc  $Y > 0$  presque-sûrement.

Soit  $f$  la densité de  $X$  et  $g$  celle de  $Y$ . Nous avons (par le théorème de Fubini et la question précédente),

$$\begin{aligned}
\phi_{\frac{X}{Y}}(t) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{it\frac{x}{y}} f(x)g(y)dx dy = \int_{-\infty}^{+\infty} g(y)dy \left( \int_{-\infty}^{+\infty} e^{it\frac{x}{y}} f(x)dx \right) \\
&= \int_{-\infty}^{+\infty} g(y) \left( \frac{y^2 a^2}{y^2 a^2 + t^2} \right) dy = \int_1^{+\infty} \frac{1}{y^2} \frac{y^2 a^2}{y^2 a^2 + t^2} dy \\
&= \frac{a}{|t|} \left( \frac{\pi}{2} - \arctan \left( \frac{a}{|t|} \right) \right).
\end{aligned}$$

*Exercice 7.5 :* 1) Supposons que  $A(\theta) \cap A(\theta') \neq \emptyset$ . Si  $\omega \in A(\theta) \cap A(\theta')$ , alors  $\mathbf{X}_1(\omega) \cos \theta + \mathbf{X}_2(\omega) \sin \theta \in F$  et  $\mathbf{X}_1(\omega) \cos \theta' + \mathbf{X}_2(\omega) \sin \theta' \in F$ . Puisque  $\theta \neq \theta'$ , tous deux dans  $[0, \frac{\pi}{2}]$ ,  $\cos \theta \sin \theta' - \sin \theta \cos \theta' = \sin(\theta - \theta') \neq 0$ . Cela entraîne que  $\mathbf{X}_1(\omega) \in F$  et  $\mathbf{X}_2(\omega) \in F$ . Contradictoire avec le fait que  $\mathbf{X}_1(\omega) \sin \theta - \mathbf{X}_2(\omega) \cos \theta \notin F$ .

2) Comme  $\mathbf{X}_1$  et  $\mathbf{X}_2$  sont indépendantes, le vecteur  $\mathbf{V} = \begin{pmatrix} \mathbf{X}_1 \\ -\mathbf{X}_2 \end{pmatrix}$  est gaussien centré et de matrice de covariance  $\mathbf{C}_V$  diagonale par bloc, chaque bloc étant la matrice de covariance  $\mathbf{C}_X$  de  $\mathbf{X}$  :

$$\mathbf{C}_V = \begin{pmatrix} \mathbf{C}_X & \mathbf{0}_d \\ \mathbf{0}_d & \mathbf{C}_X \end{pmatrix}$$

avec  $\mathbf{0}_d$  la matrice nulle de taille  $d \times d$ . Nous remarquons que  $\mathbf{W} = \begin{pmatrix} \mathbf{X}_1 \cos \theta + \mathbf{X}_2 \sin \theta \\ \mathbf{X}_1 \sin \theta - \mathbf{X}_2 \cos \theta \end{pmatrix} = \mathbf{M} \begin{pmatrix} \mathbf{X}_1 \\ -\mathbf{X}_2 \end{pmatrix}$  avec  $\mathbf{M}$  la matrice de taille  $2d \times 2d$  :

$$\mathbf{M} = \begin{pmatrix} \cos \theta \mathbf{I}_d & \sin \theta \mathbf{I}_d \\ \sin \theta \mathbf{I}_d & -\cos \theta \mathbf{I}_d \end{pmatrix}$$

et  $\mathbf{I}_d$  la matrice identité de taille  $d \times d$ . Nous utilisons alors la formule  $\mathbf{C}_W = \mathbf{M} \mathbf{C}_V \mathbf{M}^t$  et obtenons que  $\mathbf{C}_W = \mathbf{C}_V$ . Les deux vecteurs gaussiens, ayant même espérance et même matrice de covariance, ont donc même loi.

Nous en déduisons en particulier que  $\mathbb{P}(A(\theta)) = \mathbb{P}(A(0))$ , pour tout  $\theta$ .

3) Supposons que  $\mathbb{P}(A(0)) = \eta > 0$ . Alors pour  $n$  angles  $\theta_i$  distincts de  $[0, \frac{\pi}{2}[$ , comme les  $A(\theta_i)$  sont disjoints, nous aurions  $\mathbb{P}(\cup_{i=1}^n A(\theta_i)) = n\eta$ . Pour  $n$  grand, nous obtiendrions que cette probabilité est strictement supérieure à 1 ce qui est absurde. D'où  $\mathbb{P}(A(0)) = 0$ .

*Exercice 7.6 :* Il suffit de prouver que pour  $f_1$  et  $f_2$  des fonctions continues bornées, et  $f_2$  lipschitzienne de constante de lipschitzianité  $C$ ,  $\mathbb{E}(f_1(X_n)f_2(Y_n))$  converge vers  $\mathbb{E}(f_1(X)f_2(y)) = f_2(y) \mathbb{E}(f_1(X))$ , puisque  $y$  est constante.

En nous inspirant de la preuve du théorème de Slutsky 7.27, nous écrivons pour  $\varepsilon > 0$ ,

$$\begin{aligned} & |\mathbb{E}(f_1(X_n)f_2(Y_n)) - \mathbb{E}(f_1(X)f_2(y))| \\ & \leq |f_2(y)| |\mathbb{E}(f_1(X_n)) - \mathbb{E}(f_1(X))| + \mathbb{E}(|f_1(X_n)| |f_2(Y_n) - f_2(y)|) \\ & \leq \|f_2\|_\infty |\mathbb{E}(f_1(X_n)) - \mathbb{E}(f_1(X))| + C\varepsilon \|f_1\|_\infty + 2\|f_1\|_\infty \|f_2\|_\infty \mathbb{P}(|Y_n - y| > \varepsilon). \end{aligned}$$

Les termes de gauche et de droite tendent vers 0 quand  $n$  tend vers l'infini, par hypothèse. Comme l'inégalité est vraie pour tout  $\varepsilon > 0$ , nous en déduisons le résultat.

Puisque  $(x, y) \mapsto x + y$  et  $(x, y) \mapsto xy$  sont des fonctions continues, il est immédiat que  $X_n + Y_n$  et  $X_n Y_n$  convergent en loi respectivement vers  $X + y$  et  $Xy$ .

*Exercice 7.7* : Préliminaire : Il suffit d'adapter la preuve de la proposition 7.22 en approchant les indicatrices d'intervalles  $] -\infty, b]$  par des fonctions bornées de classe  $C^2$  avec dérivées première et seconde bornées et uniformément continues.

1) Nous avons  $W_{i,n} + X_{i,n} = W_{i-1,n} + Y_{i-1,n}$ . Ainsi, par un argument de somme télescopique, nous obtenons

$$\sum_{i=1}^n (f(W_{i,n} + X_{i,n}) - f(W_{i,n} + Y_{i,n})) = f(W_{1,n} + X_{1,n}) - f(W_{n,n} + Y_{n,n}) = f(T_n) - f(T'_n).$$

2) Puisque  $X_{i,n}$  et  $Y_{i,n}$  sont petits pour  $n$  grand, nous allons utiliser un développement de Taylor. Nous pouvons écrire

$$f(W_{i,n} + X_{i,n}) = f(W_{i,n}) + f'(W_{i,n})X_{i,n} + \frac{1}{2}f''(W_{i,n})(X_{i,n})^2 + R_{X,i,n},$$

où  $R_{X,i,n} = \frac{1}{2}(X_{i,n})^2(f''(W_{i,n} + \theta X_{i,n}) - f''(W_{i,n}))$  pour un  $0 \leq \theta \leq 1$ . D'une part nous avons  $|R_{X,i,n}| \leq (X_{i,n})^2 \|f''\|_\infty$ . D'autre part, puisque  $f''$  est uniformément continue, pour  $\varepsilon > 0$  donné, il existe  $\delta > 0$ , tel que  $|R_{X,i,n}| \leq \varepsilon (X_{i,n})^2$ , pour  $|X_{i,n}| \leq \delta$ . Nous en déduisons que

$$|R_{X,i,n}| \leq (X_{i,n})^2 (\varepsilon \mathbf{1}_{|X_{i,n}| \leq \delta} + \|f''\|_\infty \mathbf{1}_{|X_{i,n}| > \delta}).$$

3) Une inégalité similaire à ci-dessus est vrai pour  $Y_{i,n}$ . Nous substituons alors ces approximations de Taylor dans (7.31). En prenant l'espérance, du fait que  $X_{i,n}$  et  $Y_{i,n}$  sont centrées, et que  $\mathbb{E}(X_{i,n}^2) = \frac{1}{n} = \mathbb{E}(Y_{i,n}^2)$ , et que  $X_{i,n}$  et  $Y_{i,n}$  sont indépendantes de  $W_{i,n}$ , nous en déduisons que

$$\begin{aligned} |\mathbb{E}(f(T_n) - \mathbb{E}(f(T'_n)))| & \leq \sum_{i=1}^n \mathbb{E}(|R_{X,i,n}| + |R_{Y,i,n}|) \\ & \leq \sum_{i=1}^n (\varepsilon \mathbb{E}(X_{i,n}^2 + Y_{i,n}^2) + \|f''\|_\infty \mathbb{E}(X_{i,n}^2 \mathbf{1}_{|X_{i,n}| > \delta} + Y_{i,n}^2 \mathbf{1}_{|Y_{i,n}| > \delta})) \\ & \leq 2\varepsilon + \|f''\|_\infty \mathbb{E}(X_1^2 \mathbf{1}_{|X_1| > \delta\sqrt{n}} + Y_1^2 \mathbf{1}_{|Y_1| > \delta\sqrt{n}}). \end{aligned}$$

4) Comme  $\mathbb{E}(X_1^2) = 1$ ,  $\lim_n \mathbb{E}(X_1^2 \mathbf{1}_{|X_1| > \delta\sqrt{n}}) = 0$ , et de même pour  $Y_1$ . Puisque le choix de  $\varepsilon$  est arbitraire, nous en déduisons finalement que  $|\mathbb{E}(f(T_n)) - \mathbb{E}(f(T'_n))| \rightarrow 0$ , quand  $n$  tend vers l'infini, d'où le théorème de la limite centrale.

*Exercice 7.8 :* 1) Si  $\frac{S_n}{\sqrt{n}}$  converge en probabilité, alors  $\frac{S_{2n}}{\sqrt{2n}} - \frac{S_n}{\sqrt{n}}$  converge en probabilité vers 0. Cette différence vaut  $(\frac{1}{\sqrt{2n}} - \frac{1}{\sqrt{n}})(X_1 + \dots + X_n) + \frac{1}{\sqrt{2n}}(X_{2n+1} + \dots + X_{2n})$ . C'est la somme de deux variables aléatoires indépendantes. C'est donc une variable aléatoire centrée de variance  $\frac{1}{n}(n\sigma^2) + \frac{n}{2n}\sigma^2 = \frac{3}{3}\sigma^2$ . Cette quantité ne peut donc pas tendre en probabilité vers 0.

2) Si  $\alpha < 1/2$ , alors pour tout  $\beta > 0$  fixé, on a  $\beta n^{1/2-\alpha} \rightarrow +\infty$  quand  $n \rightarrow \infty$ . Alors pour tout  $A > 0$ ,  $\exists n_0$ , tel que  $n \geq n_0 \iff \beta n^{1/2-\alpha} > A$ . Alors  $\mathbb{P}(n^\alpha |\frac{S_n}{n} - m| > \beta) = \mathbb{P}(\sqrt{n} |\frac{S_n}{n} - m| > \beta n^{1/2-\alpha}) \leq \mathbb{P}(\sqrt{n} |\frac{S_n}{n} - m| > A)$ . Comme  $\sqrt{n}(\frac{S_n}{n} - m)$  converge en loi vers une variable aléatoire de loi  $\mathcal{N}(0, 1)$  ayant une fonction de répartition continue, les fonctions de répartition convergent, et en particulier,  $\limsup_n \mathbb{P}(n^\alpha |\frac{S_n}{n} - m| > \beta) \leq \frac{3}{2} \mathbb{P}(|\mathcal{N}(0, 1)| > A)$ . Si  $A$  tend vers l'infini, nous obtenons que  $\limsup_n \mathbb{P}(n^\alpha |\frac{S_n}{n} - m| > \beta) = 0$ .

Pour  $\alpha > 1/2$ , un raisonnement analogue permet de montrer que  $\limsup_n \mathbb{P}(\frac{1}{n^\alpha |\frac{S_n}{n} - m|} > \frac{1}{\beta}) = 0$ , d'où  $n^\alpha |\frac{S_n}{n} - m|$  tend vers  $+\infty$  en probabilité.

## 11.7 Corrigés des exercices du chapitre 8

*Exercice 8.1* : Soit  $h$  une fonction de  $\mathbb{R}$  dans  $\mathbb{R}$  continue bornée. Soit  $T$  une v.a. de loi  $T_n$ . Avec  $Z$  v.a. de loi  $\mathcal{N}(0, 1)$  et  $U$  de loi  $\chi_n^2$  indépendante de  $Z$ , on a d'après (7.20) :

$$\begin{aligned}
 \mathbb{E}(h(T)) &= \mathbb{E}\left[h\left(\sqrt{n}\frac{Z}{\sqrt{U}}\right)\right] \\
 &= \frac{1}{2^{n/2}\Gamma(n/2)\sqrt{2\pi}} \int_0^{+\infty} du \int_{-\infty}^{+\infty} dz u^{\frac{n}{2}-1} e^{-\frac{u}{2}} e^{-\frac{z^2}{2}} h\left(\sqrt{n}\frac{z}{\sqrt{u}}\right) \\
 &= \frac{2}{2^{n/2}\Gamma(n/2)\sqrt{2\pi}} \int_0^{+\infty} dx \int_{-\infty}^{+\infty} dz x^{n-1} e^{-\frac{x^2+z^2}{2}} h\left(\sqrt{n}\frac{z}{x}\right) \text{ avec } u \mapsto x = \sqrt{u} \\
 &= \frac{2}{2^{n/2}\Gamma(n/2)\sqrt{2\pi n}} \int_0^{+\infty} dx \int_{-\infty}^{+\infty} dt x^n e^{-\frac{x^2+x^2 t^2/n}{2}} h(t) \text{ avec } z \mapsto t = \sqrt{n}\frac{z}{x} \\
 &= \frac{2}{2^{n/2}\Gamma(n/2)\sqrt{2\pi n}} \int_{-\infty}^{+\infty} dth(t) \int_0^{+\infty} dx x^n e^{-\frac{x^2}{2}\left(1+\frac{t^2}{n}\right)} \text{ par Fubini} \\
 &= \frac{1}{2^{n/2}\Gamma(n/2)\sqrt{2\pi n}} \int_{\mathbb{R}} dth(t) \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \int_0^{+\infty} dy y^{\frac{n-1}{2}} e^{-\frac{y}{2}} \text{ avec } x \mapsto y = x^2\left(1+\frac{t^2}{n}\right) \\
 &= \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\Gamma(n/2)} \int_{\mathbb{R}} dth(t) \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}.
 \end{aligned}$$

Ce calcul permet d'identifier la loi de  $T$  et sa densité.

*Exercice 8.2* : 1) Soit  $(X_i)_i$  une suite de variables aléatoires i.i.d. de loi  $\mathcal{N}(0, 1)$ . Par définition de la loi  $\chi_n^2$ ,  $Z_n$  a même loi que  $\sum_{i=1}^n X_i^2$ . On en déduit que  $\mathbb{E}[Z_n] = n\mathbb{E}[X_1^2] = n$  et  $\text{Var}(Z_n) = n\text{Var}(X_1^2) = 2n$ . De plus, par le théorème de la limite centrale, on trouve :

$$\frac{Z_n - n}{\sqrt{2n}} \stackrel{\text{loi}}{=} \frac{\sum_{i=1}^n X_i^2 - n\mathbb{E}[X_1^2]}{\sqrt{\text{Var}(X_1^2)}\sqrt{n}} \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, 1),$$

la convergence ayant lieu en loi.

2) On note  $Y_n := \frac{Z_n - n}{\sqrt{2n}}$  et on écrit

$$\begin{aligned}
 \sqrt{2Z_n} - \sqrt{2n-1} &= \sqrt{2n + 2\sqrt{2n}Y_n} - \sqrt{2n-1} = \sqrt{2n}\sqrt{1 + \frac{\sqrt{2}}{\sqrt{n}}Y_n} - \sqrt{2n-1} \\
 &= \sqrt{2n} - \sqrt{2n-1} + R_n + Y_n,
 \end{aligned}$$

avec

$$R_n = \sqrt{2n}f\left(\frac{\sqrt{2}}{\sqrt{n}}Y_n\right), \quad f(x) = \sqrt{1+x} - 1 - \frac{x}{2}.$$



On montre que pour tout  $x \in ]-1, +\infty[$ ,  $|f(x)| \leq \frac{1}{2}x^2$  et donc

$$|R_n| \leq \frac{1}{2}\sqrt{2n} \left( \frac{\sqrt{2}Y_n}{\sqrt{n}} \right)^2 = \frac{\sqrt{2}Y_n^2}{\sqrt{n}}.$$

Comme  $(Y_n)_n$  converge en loi,  $(Y_n^2/\sqrt{n})_n$  converge en loi vers 0, et donc converge en probabilité vers 0. Donc  $(\sqrt{2n} - \sqrt{2n-1} + R_n)_n$  converge en probabilité vers 0. On peut alors conclure que  $(\sqrt{2}Z_n - \sqrt{2n-1})_n$  converge en loi vers la loi  $\mathcal{N}(0, 1)$  avec le théorème de Slutsky.

3) Soit  $X$  une v.a. gaussienne centrée réduite et soit  $(Y_i)_i$  une suite de variables aléatoires i.i.d. de loi gaussienne centrée réduite indépendante de  $X$ . Pour tout  $n$ ,  $U_n = \sum_{i=1}^n Y_i^2$  est indépendante de  $X$  et suit la loi de  $\chi^2$  à  $n$  degrés de liberté, donc  $X/\sqrt{U_n/n}$  suit la loi de Student à  $n$  degrés de liberté. Donc  $\zeta_n$  et  $X/\sqrt{U_n/n}$  ont même loi pour tout  $n$ . Or  $U_n/n$  converge en probabilité vers 1 d'après la loi des grands nombres, donc  $X/\sqrt{U_n/n}$  converge vers  $\mathcal{N}(0, 1)$  d'après le théorème de Slutsky.

*Exercice 8.3* : 1)  $\mathbb{E}(U) = \mathbb{E}(V) = \mathbb{E}(W) = m$ .  $\text{Var}(U) = m(1-m)$ ,  $\text{Var}(V) = \int_0^1 g^2(x)dx - m^2 \leq \text{Var}(U)$ , car  $g \leq 1$ .  $\text{Var}(W) = \frac{1}{2} \int_0^1 (g^2(x) + g(x)g(1-x))dx - m^2$ .

2) Soient  $X_i$  et  $Y_i$  des variables aléatoires indépendantes de loi uniforme sur  $[0, 1]$ . On peut appliquer la loi des grands nombres aux variables correspondantes  $U_i$ ,  $V_i$  et  $W_i$ . Ainsi, nous aurons :  $\frac{1}{n} \sum_{i=1}^n g(X_i)$ ,  $\frac{1}{2n} \sum_{i=1}^n (g(X_i) + g(1-X_i))$  et  $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{Y_i \leq g(X_i)}$  convergent presque-sûrement vers  $m$ .

3) La monotonie de  $g$  entraîne que  $\mathbb{E}((g(X) - g(Y))(g(1-X) - g(1-Y))) \leq 0$ . Nous en déduisons que  $2\mathbb{E}(g(X)g(1-X)) - 2\mathbb{E}(g(Y)g(1-X)) \leq 0$ . Or,  $\mathbb{E}(g(Y)g(1-X)) = m^2$ . Donc  $\int_0^1 g(x)g(1-x)dx \leq m^2$ . Par ailleurs,  $m^2 \leq \int_0^1 g^2(x)dx$ . Finalement, nous avons

$$\text{Var}(W) \leq \frac{1}{2} \left( \int_0^1 g^2(x)dx - m^2 \right) \leq \frac{1}{2} \text{Var}(V).$$

4)  $\text{Var}(A_n) = \frac{1}{2n} \text{Var}(V)$  et  $\text{Var}(B_n) = \frac{1}{n} \text{Var}(W)$ . La comparaison ci-dessus entraîne que  $B_n$  est le meilleur.

5)  $m = \int_0^1 x^2 dx = \frac{1}{3}$ .  $\text{Var}(g(X)) = \frac{4}{45} = \sigma^2$ . Le TCL nous dit que  $\sqrt{2n} \frac{(A_n - m)}{\sigma}$  converge vers une variable aléatoire de loi  $\mathcal{N}(0, 1)$ . Nous aurons

$$\mathbb{P}(|A_n - m| > \varepsilon) = \mathbb{P} \left( \left| \sqrt{2n} \frac{(A_n - m)}{\sigma} \right| > \frac{\sqrt{2n}\varepsilon}{\sigma} \right) = 0,05 \iff \frac{\sqrt{2n}\varepsilon}{\sigma} = 1,96.$$

En prenant  $\varepsilon = 0,01$ , nous obtenons  $n = 1708$ .

Étudions maintenant l'estimateur  $B_n$ .  $\text{Var}(W) = \frac{1}{2} \int_0^1 (g^2(x) + g(x)g(1-x))dx - m^2 =$

$\frac{1}{180} = \beta^2$ . Le même raisonnement que précédemment nous dit que  $n$  dans ce cas doit vérifier  $\frac{\sqrt{n\varepsilon}}{\beta} = 1,96$  avec  $\varepsilon = 0,01$ . Cela donne  $n = 214$ , ce qui est bien meilleur comme nous l'avions prévu.

*Exercice 8.4* : 1) On sait que  $\mathbb{E}(X_1) = m\theta$ . Donc un estimateur de  $\theta$  est

$$\check{\theta}_n = \frac{1}{mn} \sum_{i=1}^n X_i.$$

2) La vraisemblance est :

$$p_n(\mathbf{x}, \theta) = \prod_{i=1}^n \binom{m}{x_i} \theta^{x_i} (1 - \theta)^{m - x_i},$$

Ceci s'écrit aussi :

$$p_n(\mathbf{x}, \theta) = \left[ \prod_{i=1}^n \binom{m}{x_i} \right] [\theta^{\bar{x}_n} (1 - \theta)^{m - \bar{x}_n}]^n, \quad (11.2)$$

avec  $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$ . A  $\mathbf{x}$  fixé, en tant que fonction de  $\theta$ , on remarque que la vraisemblance est maximale lorsque la fonction  $\theta \mapsto \theta^{\bar{x}_n} (1 - \theta)^{m - \bar{x}_n}$  est maximale. Le maximum sur  $[0, 1]$  est unique et est atteint au point où la dérivée de la fonction s'annule, en  $\theta = \bar{x}_n/m$ . Le maximum de la vraisemblance redonne ici l'estimateur empirique.

*Exercice 8.5* : 1) On calcule

$$\begin{aligned} \mathbb{E}_a(X) &= \int_0^\infty \frac{x^2}{a} \exp\left(-\frac{x^2}{2a}\right) dx = \frac{1}{2} \sqrt{\frac{2\pi}{a}} \left[ \frac{1}{\sqrt{2\pi a}} \int_{-\infty}^\infty x^2 \exp\left(-\frac{x^2}{2a}\right) dx \right] \\ &= \frac{1}{2} \sqrt{\frac{2\pi}{a}} a = \sqrt{\frac{\pi a}{2}}, \end{aligned}$$

où on a reconnu dans l'expression entre crochets la variance d'une loi gaussienne  $\mathcal{N}(0, a)$ . Puis

$$\mathbb{E}_a(X^k) = \int_0^\infty \frac{x^{k+1}}{a} \exp\left(-\frac{x^2}{2a}\right) dx = (2a)^{k/2} \int_0^\infty y^{k/2} e^{-y} dy,$$

et donc  $\mathbb{E}_a(X^2) = 2a$  et  $\mathbb{E}_a(X^4) = 8a^2$ .

2) La log-vraisemblance est égale à :

$$l_n(\mathbf{X}, a) = \sum_{i=1}^n \ln X_i - n \ln a - \frac{1}{2a} \sum_{i=1}^n X_i^2.$$

En tant que fonction de  $a$ , il y a un unique minimum sur  $]0, +\infty[$  atteint au point où la dérivée s'annule, ce qui donne l'EMV :

$$\hat{a}_n = \frac{1}{2n} \sum_{i=1}^n X_i^2.$$

L'estimateur est non-biaisé :

$$\mathbb{E}_a(\hat{a}_n) = \frac{1}{2n} \sum_{i=1}^n \mathbb{E}_a(X_i^2) = a.$$

Le théorème de la limite centrale donne la normalité asymptotique :

$$\sqrt{n}(\hat{a}_n - a) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \sigma^2(a)),$$

en loi, avec  $\sigma^2(a) = \frac{1}{4} \text{Var}_a(X_1^2) = \frac{1}{4} \mathbb{E}_a(X_1^4) - \frac{1}{4} \mathbb{E}_a(X_1^2)^2 = a^2$ .

3) L'EMV de  $a$  est  $\hat{a}_n = 38,69/16 = 2,42$ . La compagnie d'assurance pense que le paramètre  $a$  vérifie  $\mathbb{P}_a(X \geq 6) \leq 10^{-3}$ , ce qui est équivalent à  $\exp(-6^2/(2a)) \leq 10^{-3}$ , soit  $a \leq 18/\ln(1000) \simeq 2,61$ . Comme  $2,42 < 2,61$ , cela semble raisonnable. Mais ici l'incertitude est grande car l'échantillon n'était pas très grand. Prenons l'approximation gaussienne :

$$\mathbb{P}_a\left(\sqrt{n} \frac{\hat{a}_n - a}{\hat{a}_n} \geq -u\right) \simeq \Phi(u),$$

qui est équivalente à :

$$\mathbb{P}_a(a \leq \hat{a}_n(1 + u/\sqrt{n})) \simeq \Phi(u).$$

La probabilité que  $a$  soit plus petit que 2,61 est donc  $\Phi(u)$  avec  $u$  tel que  $\hat{a}_n(1 + u/\sqrt{n}) = 2,61$ , c'est-à-dire  $u = \sqrt{n}(2,61/\hat{a}_n - 1)$ . Avec  $n = 8$  et  $\hat{a}_n = 2,42$ , cela donne  $u = 0,22$  et donc la probabilité que  $a$  soit plus petit que 2,61 est de  $\Phi(u) = 0,59$  seulement.

## 11.8 Corrigés des exercices du chapitre 9

*Exercice 9.1 :* 1) Sous  $\mathbb{P}_\theta$ ,  $X_1, \dots, X_n$  sont indépendantes et identiquement distribuées selon la loi exponentielle de paramètre  $\theta$ , donc  $S_n$  suit la loi  $\Gamma(n, \theta)$  : Pour tout  $z \geq 0$ ,

$$\mathbb{P}_\theta(S_n \leq z) = \frac{1}{(n-1)!} \int_0^{z\theta} x^{n-1} e^{-x} dx.$$

Il est équivalent de dire que  $\theta S_n$  suit la loi  $\Gamma(n, 1)$ , ou que  $2\theta S_n$  suit la loi  $\chi_{2n}^2$ .

2) Si on note  $\Phi_{\chi(2n)}$  la fonction de répartition de la loi  $\chi_{2n}^2$ , alors on peut trouver deux nombres  $a_n, b_n \in [0, +\infty]$  tels que  $\Phi_{\chi(2n)}(a_n) = \alpha/2$  et  $\Phi_{\chi(2n)}(b_n) = 1 - \alpha/2$  (ce sont les quantiles de la loi  $\chi_{2n}^2$  d'ordre  $\alpha/2$  et  $1 - \alpha/2$ , respectivement). On a alors

$$\begin{aligned} \mathbb{P}_\theta\left(\theta \in \left[\frac{a_n}{2S_n}, \frac{b_n}{2S_n}\right]\right) &= \mathbb{P}_\theta(2\theta S_n \in [a_n, b_n]) \\ &= \Phi_{\chi(2n)}(b_n) - \Phi_{\chi(2n)}(a_n) = 1 - \alpha, \end{aligned}$$

ce qui montre que  $\left[\frac{a_n}{2S_n}, \frac{b_n}{2S_n}\right]$  est un intervalle de confiance exact au niveau  $1 - \alpha$  pour  $\theta$ .

Par exemple, pour  $n = 10$  et  $\alpha = 5\%$ , on a  $a_n = 9,59$  et  $b_n = 34,17$  d'après la Table 9.2, et donc on trouve que  $[4,80/S_n, 17,08/S_n]$  est un intervalle de confiance au niveau 95% de  $\theta$ .

*Exercice 9.2 :* 1)  $\mathbb{E}(\bar{X}_n) = p$ ,  $\text{Var}(\bar{X}_n) = \frac{p(1-p)}{n}$ .

2) La loi des grands nombres implique que  $\bar{X}_n$  converge presque-sûrement et dans  $L^1$  vers  $p$ .

3) L'intervalle de confiance aura la forme  $[\bar{X}_n - \frac{c}{2\sqrt{n}}; \bar{X}_n + \frac{c}{2\sqrt{n}}]$ , où  $\int_{-c}^c g(x)dx = 0,9$ , avec  $g$  la densité de la loi normale  $\mathcal{N}(0, 1)$ . La table numérique donne  $c = 1,645$ . Nous en déduisons alors la fourchette d'estimation  $0,64 - \frac{1,645}{20} \leq p \leq 0,64 + \frac{1,645}{20}$ , soit encore  $0,56 \leq p \leq 0,72$ .

*Exercice 9.3 :* 1) La vraisemblance est :

$$p_n(\mathbf{x}, (\alpha, \beta)) = \alpha^n \beta^{n\alpha} \left(\prod_{i=1}^n x_i\right)^{-\alpha-1} \mathbf{1}_{\min(x_i) \geq \beta}.$$

A  $\mathbf{x}$  et  $\alpha$  fixés, cette fonction de  $\beta$  est maximale en  $\beta = \hat{\beta}_n = \min(x_i)$ , et alors

$$p_n(\mathbf{x}, (\alpha, \beta = \min(x_i))) = \alpha^n (\min(x_i))^{n\alpha} \left(\prod_{i=1}^n x_i\right)^{-\alpha-1} = \alpha^n \psi(\mathbf{x})^\alpha \left(\prod_{i=1}^n x_i\right)^{-1},$$

avec  $\psi(\mathbf{x}) = \frac{\min(x_i)^n}{\prod_{i=1}^n x_i} \in [0, 1]$ . Sauf si les  $x_i$  sont tous égaux, on a  $\psi(\mathbf{x}) < 1$  et donc, en tant que fonction de  $\alpha$ , cette fonction est minimale en

$$\hat{\alpha}_n = -\frac{n}{\ln \psi(\mathbf{x})} = \frac{1}{\frac{1}{n} \sum_{i=1}^n \ln(x_i) - \ln \min(x_i)}.$$

Si les  $x_i$  sont tous égaux, alors la fonction est croissante en  $\alpha$  et tend vers  $+\infty$  en  $+\infty$ . Donc l'EMV de  $(\alpha, \beta)$  est

$$(\hat{\alpha}_n, \hat{\beta}_n) = \left( \min(X_i), \frac{1}{\frac{1}{n} \sum_{i=1}^n \ln(X_i) - \ln \min_i(X_i)} \right).$$

Il est bien défini si  $\mathbf{X}$  n'a pas toutes ses coordonnées identiques.

2) On calcule la fonction de répartition de  $Z = \ln X$ . Soit  $z \in \mathbb{R}$ . On a  $\mathbb{P}(Z \leq z) = \mathbb{P}(X \leq e^z)$ . Donc si  $z < 0$ ,  $e^z < 1$  et  $\mathbb{P}(Z \leq z) = 0$ . Si  $z \geq 0$  :

$$\mathbb{P}(Z \leq z) = \int_1^{e^z} \frac{\alpha}{x^{\alpha+1}} dx = 1 - e^{-\alpha z},$$

ce qui montre que  $Z$  suit la loi exponentielle de paramètre  $\alpha$ .

Un estimateur de  $\alpha$  est donc

$$\hat{\alpha}_n = \left[ \frac{1}{n} \sum_{i=1}^n \ln X_i \right]^{-1}.$$

Par la loi forte des grands nombres, on a

$$1/\hat{\alpha}_n \xrightarrow{n \rightarrow +\infty} 1/\alpha$$

$\mathbb{P}_\alpha$ -presque sûrement. Donc  $\hat{\alpha}_n \xrightarrow{n \rightarrow +\infty} \alpha$   $\mathbb{P}_\alpha$ -presque sûrement. Par le théorème de la limite centrale,  $1/\hat{\alpha}_n$  vérifie

$$\sqrt{n}(1/\hat{\alpha}_n - 1/\alpha) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \text{Var}_\alpha(\ln X)),$$

en loi, avec  $\text{Var}_\alpha(\ln X) = \text{Var}_\alpha(Z) = \mathbb{E}_\alpha(Z^2) - \mathbb{E}_\alpha(Z)^2 = 1/\alpha^2$ . En utilisant cette approximation normale et le théorème de Slutsky, on a

$$\mathbb{P}_\alpha \left( \sqrt{n} \hat{\alpha}_n |1/\hat{\alpha}_n - 1/\alpha| \leq \Phi^{-1}(1 - \eta/2) \right) \simeq 1 - \eta,$$

en notant  $\Phi^{-1}(1 - \eta/2)$  le quantile d'ordre  $1 - \alpha/2$  de la loi gaussienne centrée réduite. Ceci s'écrit aussi :

$$\mathbb{P}_\alpha \left( |\hat{\alpha}_n/\alpha - 1| \leq \Phi^{-1}(1 - \eta/2)/\sqrt{n} \right) \simeq 1 - \eta.$$

On en déduit l'intervalle de confiance asymptotique :

$$\mathbb{P}_\alpha \left( \alpha \in \left[ \frac{\hat{\alpha}_n}{1 + \Phi^{-1}(1 - \eta/2)/\sqrt{n}}, \frac{\hat{\alpha}_n}{1 - \Phi^{-1}(1 - \eta/2)/\sqrt{n}} \right] \right) \simeq 1 - \eta.$$

*Exercice 9.4* : 1) D'une part, pour tout  $z > 0$ ,  $\mathbb{P}_{\lambda, \mu}(Z_i \leq z) = 1 - \mathbb{P}_{\lambda, \mu}(\min(X_i, Y_1) > z) = 1 - \exp(-(\lambda + \mu)z)$ , ce qui montre que  $Z_i \sim \mathcal{E}(\lambda + \mu)$ . D'autre part,  $W_i$  ne prend pour valeurs que 0 ou 1, c'est une variable de Bernoulli. On a  $\mathbb{P}_{\lambda, \mu}(W_i =$

1)  $= \mathbb{P}_{\lambda, \mu}(X_i \leq Y_i) = \int_0^\infty \int_x^\infty \mu \exp(-\mu y) dy \lambda \exp(-\lambda x) dx = \frac{\lambda}{\lambda + \mu}$ , ce qui montre que  $W_i \sim \mathcal{B}\left(\frac{\lambda}{\lambda + \mu}\right)$ .

2) Soit  $f$  et  $g$  deux fonctions continues bornées. En décomposant sur les valeurs prises par  $W_i$ , on a

$$\mathbb{E}_{\lambda, \mu}(f(Z_i)g(W_i)) = g(1)\mathbb{E}_{\lambda, \mu}(f(Z_i)\mathbf{1}_{W_i=1}) + g(0)\mathbb{E}_{\lambda, \mu}(f(Z_i)\mathbf{1}_{W_i=0}).$$

Or

$$\begin{aligned} \mathbb{E}_{\lambda, \mu}(f(Z_i)\mathbf{1}_{W_i=1}) &= \mathbb{E}_{\lambda, \mu}(f(Z_i)\mathbf{1}_{X_i \leq Y_i}) = \mathbb{E}_{\lambda, \mu}(f(X_i)\mathbf{1}_{X_i \leq Y_i}) \\ &= \int_0^\infty f(x) \int_x^\infty \mu \exp(-\mu y) dy \lambda \exp(-\lambda x) dx \\ &= \frac{\lambda}{\lambda + \mu} \int_0^\infty f(x) (\lambda + \mu) \exp(-(\lambda + \mu)x) dx = \frac{\lambda}{\lambda + \mu} \mathbb{E}_{\lambda, \mu}(f(Z_i)), \end{aligned}$$

et de même  $\mathbb{E}_{\lambda, \mu}(f(Z_i)\mathbf{1}_{W_i=0}) = \frac{\mu}{\lambda + \mu} \mathbb{E}_{\lambda, \mu}(f(Z_i))$ . On conclut que

$$\mathbb{E}_{\lambda, \mu}(f(Z_i)g(W_i)) = \mathbb{E}_{\lambda, \mu}(f(Z_i))\mathbb{E}_{\lambda, \mu}(g(W_i)),$$

ce qui assure l'indépendance de  $Z_i$  et  $W_i$ .

3) La vraisemblance est :

$$\begin{aligned} p_n((\mathbf{z}, \mathbf{w}), \lambda) &= \prod_{i=1}^n (\lambda + \mu) \exp(-(\lambda + \mu)z_i) \left(\frac{\lambda}{\lambda + \mu}\right)^{\mathbf{1}_{w_i=1}} \left(\frac{\mu}{\lambda + \mu}\right)^{\mathbf{1}_{w_i=0}} \\ &= \exp\left(-(\lambda + \mu) \sum_{i=1}^n z_i\right) \lambda^{\sum_{i=1}^n w_i} \mu^{n - \sum_{i=1}^n w_i} \\ &= (e^{-\lambda})^{\sum_{i=1}^n z_i} \lambda^{\sum_{i=1}^n w_i} \exp\left(-\mu \sum_{i=1}^n z_i\right) \mu^{n - \sum_{i=1}^n w_i}. \end{aligned}$$

La log-vraisemblance est :

$$l_n((\mathbf{z}, \mathbf{w}), \lambda) = -\lambda \sum_{i=1}^n z_i + \ln \lambda \sum_{i=1}^n w_i - \mu \sum_{i=1}^n z_i + \ln \mu \left(n - \sum_{i=1}^n w_i\right).$$

En particulier  $\partial_\lambda l_n((\mathbf{z}, \mathbf{w}), \lambda) = -\sum_{i=1}^n z_i + \sum_{i=1}^n w_i/\lambda$ , qui est une fonction décroissante en  $\lambda$  qui s'annule uniquement en  $\sum_{i=1}^n w_i / \sum_{i=1}^n z_i$ . Donc la log-vraisemblance est maximale en  $\sum_{i=1}^n w_i / \sum_{i=1}^n z_i$ . On en déduit que l'Estimateur du Maximum de Vraisemblance de  $\lambda$  est

$$\hat{\lambda}_n = \frac{\sum_{i=1}^n W_i}{\sum_{i=1}^n Z_i}.$$

4) D'après la loi forte des grands nombres, sous  $\mathbb{P}_\lambda$ , on a  $\frac{1}{n} \sum_{i=1}^n W_i \rightarrow \frac{\lambda}{\lambda+\mu}$  p.s. et  $\frac{1}{n} \sum_{i=1}^n Z_i \rightarrow \frac{1}{\lambda+\mu}$  p.s.. Donc  $\hat{\lambda}_n \rightarrow \lambda$  p.s..

5) Sous  $\mathbb{P}_\lambda$ ,  $\sum_{i=1}^n Z_i \sim \Gamma(n, \lambda + \mu)$  est indépendante de  $\sum_{i=1}^n W_i \sim \mathcal{B}(n, \frac{\lambda}{\lambda+\mu})$ . Donc

$$\begin{aligned} \mathbb{E}_\lambda(\hat{\lambda}_n) &= \mathbb{E}_\lambda\left(\sum_{i=1}^n W_i\right) \mathbb{E}_\lambda\left(\frac{1}{\sum_{i=1}^n Z_i}\right) = \frac{n\lambda}{\lambda+\mu} \int_0^\infty \frac{1}{z} \frac{1}{\Gamma(n)} (\lambda+\mu)^n z^{n-1} e^{-(\lambda+\mu)z} dz \\ &= \frac{n\lambda\Gamma(n-1)}{\Gamma(n)} \int_0^\infty \frac{1}{\Gamma(n-1)} (\lambda+\mu)^{n-1} z^{n-2} e^{-(\lambda+\mu)z} dz = \frac{n\lambda}{n-1}. \end{aligned}$$

Donc  $\hat{\lambda}_n$  est un estimateur biaisé mais asymptotiquement non-biaisé.

6) On a

$$\sqrt{n}(\hat{\lambda}_n - \lambda) = \frac{1}{\bar{Z}_n} \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_i - \lambda Z_i),$$

avec  $\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i$  qui converge  $\mathbb{P}_\lambda$ -p.s. vers  $\mathbb{E}_\lambda(Z_1) = \frac{1}{\lambda+\mu}$  d'après la loi forte des grands nombres. Sous  $\mathbb{P}_\lambda$ , d'après le théorème de la limite centrale,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (W_i - \lambda Z_i) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \text{Var}_\lambda(W_1 - \lambda Z_1)),$$

en loi. On a

$$\text{Var}_\lambda(W_1 - \lambda Z_1) = \text{Var}_\lambda(W_1) + \lambda^2 \text{Var}_\lambda(Z_1) = \frac{\lambda\mu}{(\lambda+\mu)^2} + \frac{\lambda^2}{(\lambda+\mu)^2} = \frac{\lambda}{\lambda+\mu}.$$

D'après le théorème de Slutsky, sous  $\mathbb{P}_\lambda$ ,

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \lambda(\lambda + \mu)),$$

en loi.

7) Sous  $\mathbb{P}_\lambda$ ,  $\hat{\lambda}_n(\hat{\lambda}_n + \mu)$  converge p.s. vers  $\lambda(\lambda + \mu)$ . Donc, toujours d'après le théorème de Slutsky, sous  $\mathbb{P}_\lambda$ ,

$$\frac{\sqrt{n}}{\sqrt{\hat{\lambda}_n(\hat{\lambda}_n + \mu)}} (\hat{\lambda}_n - \lambda) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, 1),$$

en loi. Si  $\Phi^{-1}(r)$  désigne le quantile d'ordre  $r$  de la loi gaussienne centrée réduite, alors

$$\left[ \hat{\lambda}_n - \Phi^{-1}(1 - \alpha/2) \sqrt{\frac{\hat{\lambda}_n(\hat{\lambda}_n + \mu)}{n}}, \hat{\lambda}_n + \Phi^{-1}(1 - \alpha/2) \sqrt{\frac{\hat{\lambda}_n(\hat{\lambda}_n + \mu)}{n}} \right]$$

est un intervalle de confiance bilatéral asymptotique pour  $\lambda$  de niveau  $1 - \alpha$ .

8) A  $\mu$  fixé,  $l_n((\mathbf{z}, \mathbf{w}), (\lambda, \mu))$  est maximum pour  $\lambda = \frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n z_i}$  et vaut alors  $\sum_{i=1}^n w_i \left( \ln \left( \frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n z_i} \right) - 1 \right) + f(\mu)$  avec  $f(\mu) = -\mu \sum_{i=1}^n z_i + (n - \sum_{i=1}^n w_i) \ln \mu$ . La fonction  $f'(\mu) = -\sum_{i=1}^n z_i + \frac{1}{\mu} (n - \sum_{i=1}^n w_i)$  est décroissante et s'annule pour  $\mu = \frac{n - \sum_{i=1}^n w_i}{\sum_{i=1}^n z_i}$ . Donc  $f(\mu)$  est croissante jusqu'à cette valeur, puis décroissante. On conclut donc que l'EMV de  $(\lambda, \mu)$  est

$$(\hat{\lambda}_n, \hat{\mu}_n) = \left( \frac{\sum_{i=1}^n W_i}{\sum_{i=1}^n Z_i}, \frac{n - \sum_{i=1}^n W_i}{\sum_{i=1}^n Z_i} \right).$$

## 11.9 Corrigés des exercices du chapitre 10

*Exercice 10.1 :* 1)  $\mathbb{E}_\theta(X_1) = 1/\theta$ .

2) On a  $S_n \sim \Gamma(n, \theta)$ , donc  $\theta S_n \sim \Gamma(n, 1)$ , soit  $2\theta S_n \sim \chi_{2n}^2$ . On pose  $\zeta_n = 2\theta_0 S_n$ . Sous  $H_0$ ,  $\zeta_n \sim \chi_{2n}^2$ . Sous  $H_1$ ,  $\zeta_n = (\theta_0/\theta)(2\theta S_n)$  va avoir tendance à prendre des valeurs plus petites (si  $\theta > \theta_0$ ) ou plus grandes (si  $\theta < \theta_0$ ) que sous  $H_0$ . On choisit donc  $W_n = \{\zeta_n \notin [x_{\alpha/2}(2n), x_{1-\alpha/2}(2n)]\}$  où  $x_r(2n)$  désigne le quantile d'ordre  $r$  de la loi  $\chi_{2n}^2$ . Ce test a pour niveau  $\alpha$ .  
A.N. :  $\zeta_{15}^{obs} = 44,1$  et d'après la Table 9.2,  $x_{0,025}(30) = 16,79$  et  $x_{0,975}(30) = 46,98$ .

3) La nouvelle région critique est  $W_n = \{\zeta_n \in ]x_{1-\alpha}(2n), +\infty[ \}$ .  
A.N. :  $x_{0,95}(30) = 43,77$ .

*Exercice 10.2 :* Ici  $\Theta = \{(\theta_i)_{i=0}^9, \theta_i \geq 0, \sum_{i=0}^9 \theta_i = 1\}$  et  $H_0 = \{\boldsymbol{\theta}^{(0)}\}$  avec  $\boldsymbol{\theta}^{(0)} = (1/10, \dots, 1/10)$ . On note  $\hat{\theta}_{250,i} = N(i)/250$  et

$$\zeta_{250} = 250 \sum_{i=0}^9 \frac{(\hat{\theta}_{250,i} - \theta_i^{(0)})^2}{\theta_i^{(0)}}.$$

On a  $\zeta_{250}^{obs} \simeq 10,4$ . Sous l'hypothèse  $H_0$ ,  $\zeta_{250}$  suit une loi du  $\chi^2$  à 9 degrés de liberté. Si  $Z \sim \chi_9^2$ , on a  $\mathbb{P}(Z \geq 16,9) = 0,05$ . Comme  $\zeta_{250}^{obs} \leq 16,9$ , on accepte l'hypothèse que le générateur produit des entiers uniformément répartis sur  $\{0, \dots, 9\}$  au niveau 5%.



# Chapitre 12

## Textes et corrigés d'examens

### Examen Ecole Polytechnique 2017

**Exercice** : File de voitures.

Sur une route à une seule voie et un seul sens de circulation, on considère une file infinie de véhicules numérotés  $0, 1, 2, 3, \dots$ . On suppose qu'ils circulent tous à la même vitesse prise égale à 1. Ils sont séparés par des distances  $(L_j)_{j \geq 1}$  comptées de l'arrière du véhicule  $j - 1$  à l'avant du véhicule  $j$  qui suit.

On suppose que pour que le véhicule  $j$  s'arrête, il lui faut, à partir du moment où le véhicule  $j - 1$  qui est devant lui commence à freiner, une distance  $R_j + D_j$  où  $R_j$  est la distance parcourue pendant le délai de réaction du conducteur du véhicule  $j$  et  $D_j$  est la distance d'immobilisation de ce véhicule.

On suppose que les triplets  $(L_j, R_j, D_j)_{j \geq 1}$  sont i.i.d. avec  $L_1, D_1$  et  $R_1$  indépendantes et positives.

On va étudier les risques de collision lorsqu'un objet encombrant tombe du véhicule 0 en tête de file et empêche la circulation. Pour cela on supposera pour simplifier que :

- lorsqu'un automobiliste freine, il tente d'arrêter son véhicule sur la distance la plus courte possible, peu importe s'il laisse de la place devant lui,
- les vitesses de décélération sont telles que les collisions n'ont lieu qu'entre un véhicule en mouvement et un véhicule arrêté.

On pose  $D_0 = 0$  et pour  $n \geq 1$ , on note  $B_n = \{R_n + D_n < L_n + D_{n-1}\}$ .

1. Montrer que  $\mathbb{P}(B_1) \leq \mathbb{P}(B_2)$ . Comparer  $\mathbb{P}(B_2)$  et  $\mathbb{P}(B_3)$ .
2. Dans cette question, on suppose que  $R_1, L_1$  et  $D_1$  possèdent des densités respectivement notées  $p_R, p_L$  et  $p_D$ .

- (a) Exprimer en fonction de  $p_R$ ,  $p_D$  et  $\bar{F}_L(\ell) = \mathbb{P}(\ell < L_1)$  la probabilité pour que le véhicule 1 ne percute pas l'obstacle.
- (b) Calculer cette probabilité lorsque  $p_R$ ,  $p_L$  et  $p_D$  sont les densités exponentielles de paramètres respectifs  $\theta_R$ ,  $\theta_L$  et  $\theta_D$ .
3. Expliquer pourquoi l'événement "les véhicules de 1 à  $n$  s'arrêtent sans collision" s'écrit  $A_n = \bigcap_{k=1}^n B_k$ .
4. On suppose que  $\mathbb{P}(R_1 \geq L_1) > 0$  et on note  $p$  cette probabilité.
- (a) Montrer que  $\mathbb{P}(D_2 \geq D_1) = \frac{1+\mathbb{P}(D_2=D_1)}{2}$  et que  $\mathbb{P}(B_2^c) \geq \frac{p}{2}$ .
- (b) Que peut-on dire des événements  $(B_{2k})_{k \geq 1}$ ? En déduire que  $\lim_{n \rightarrow +\infty} \mathbb{P}(A_{2n}) = 0$ .
- (c) Conclure qu'il y a presque sûrement au moins un accident.
- (d) On suppose que  $D_1$  est déterministe égale à  $d_F$ . Déterminer la loi de l'indice  $N$  du premier véhicule qui n'arrive pas à s'arrêter à temps. Lorsque  $\mathbb{P}(R_1 + d_F < L_1) > 0$ , quelle est la loi conditionnelle de  $N - 1$  sachant  $N \geq 2$ ?

**Problème : Loi de Pareto.**

Pour  $a \in \mathbb{R}$  et  $b > 0$ , on appelle loi de Pareto de paramètre  $(a, b)$  et on note  $\mathcal{P}(a, b)$  la loi de densité  $p_1(x, a, b) = \frac{b}{(x-a)^{b+1}} \mathbf{1}_{[a+1, +\infty[}(x)$ . On note  $X \sim \mathcal{P}(a, b)$  si la variable aléatoire  $X$  est distribuée suivant la loi de Pareto de paramètre  $(a, b)$ . Cette loi est par exemple utilisée pour modéliser la distribution des revenus dans une population.

**I) Analyse probabiliste**

Soit  $X \sim \mathcal{P}(a, b)$ .

- Montrer que  $(X - a)^{-b}$  suit la loi uniforme sur  $[0, 1]$  et en déduire sans calcul que  $\ln(X - a)$  suit la loi exponentielle de paramètre  $b$ .
- A l'aide du changement de variables  $u = \frac{1}{x-a}$  vérifier que

$$\forall c \in ]-1, b[, \mathbb{E}[(X - (a + 1))^c] = b \int_0^1 (1 - u)^c u^{b-c-1} du.$$

Reconnaître l'intégrale d'une densité usuelle à la constante de normalisation près et en déduire que

$$\forall c \in ]-1, b[, \mathbb{E}[(X - (1 + a))^c] = \frac{\Gamma(b - c)\Gamma(c + 1)}{\Gamma(b)}. \quad (12.1)$$

où  $\Gamma$  est la fonction gamma d'Euler définie par  $\forall a > 0, \Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$  qui vérifie  $\Gamma(a + 1) = a\Gamma(a)$  et  $\forall n \in \mathbb{N}^*, \Gamma(n) = (n - 1)!$ .

Les deux questions qui suivent sont indépendantes du reste du problème.

Soient  $b, c > 0$ ,  $Y \sim \mathcal{P}(0, b)$ ,  $Z \sim \mathcal{P}(0, c)$  et  $\varepsilon$  une variable aléatoire de Bernoulli de paramètre  $\frac{c}{b+c}$  indépendantes.

3. Déterminer la densité de  $S = \varepsilon Y + \frac{1-\varepsilon}{Z}$ .
4. Déterminer la loi de  $(R, Z)$  où  $R = Y/Z$ . Les variables aléatoires  $R$  et  $Z$  sont-elles indépendantes? Vérifier que  $R$  et  $S$  ont même loi.

## II) Estimation statistique

On observe une réalisation de  $\mathbf{X} = (X_1, \dots, X_n)$  où les  $X_i$  sont des variables aléatoires indépendantes et identiquement distribuées de loi commune dans  $\{\mathcal{P}(a, b) : a \in \mathbb{R}, b > 0\}$ .

5. Soient  $a \in \mathbb{R}, b > 0$ . Déterminer la vraisemblance  $p_n(\mathbf{x}, a, b)$  pour  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ . Exprimer à l'aide de  $\min_{1 \leq i \leq n} x_i$  la valeur  $a_n(\mathbf{x}) \in \mathbb{R}$  qui maximise  $a \mapsto p_n(\mathbf{x}, a, b)$  pour tout  $(\mathbf{x}, b) \in \mathbb{R}^n \times ]0, +\infty[$ .
6. Calculer la log-vraisemblance  $l_n(\mathbf{x}, a, b)$ . Quelle valeur  $b_n(\mathbf{x}, a) \in ]0, +\infty[$  maximise-t-elle  $b \mapsto l_n(\mathbf{x}, a, b)$  pour  $a \in \mathbb{R}$  et  $\mathbf{x} \in ]a+1, +\infty[^n$  fixés? En déduire l'estimateur du maximum de vraisemblance (EMV)  $(\hat{a}_n, \hat{b}_n)$  du couple  $(a, b)$ .
7. (a) Calculer  $\mathbb{P}_{(a,b)}(X_1 > x)$  pour  $x \geq a+1$ . En déduire que sous  $\mathbb{P}_{(a,b)}$ ,  $M_n = \min_{1 \leq i \leq n} X_i \sim \mathcal{P}(a, nb)$ . Calculer  $\mathbb{E}_{(a,b)}(\hat{a}_n - a)$  pour  $n > 1/b$ . L'EMV est-il sans biais?
  - (b) A l'aide de l'équation (12.1), vérifier que  $\lim_{n \rightarrow \infty} n^3 \mathbb{E}_{(a,b)}[|\hat{a}_n - a|^3] = \frac{6}{b^3}$  et en déduire que  $\mathbb{P}_{(a,b)}(\lim_{n \rightarrow \infty} \sqrt{n}(\hat{a}_n - a) = 0) = 1$ .
8. On pose  $R_n = \frac{1}{n} \sum_{i=1}^n \ln(X_i - a) - \frac{1}{n} \sum_{i=1}^n \ln(X_i - \hat{a}_n)$ .
  - (a) A l'aide de la question 1., donner les comportements asymptotiques lorsque  $n \rightarrow \infty$  sous  $\mathbb{P}_{(a,b)}$  de  $\frac{1}{n} \sum_{i=1}^n \ln(X_i - a)$  et  $\sqrt{n} \left( \frac{1}{b} - \frac{1}{n} \sum_{i=1}^n \ln(X_i - a) \right)$ .
  - (b) Avec l'inégalité  $\forall x > 0, \ln(x) \leq x - 1$ , vérifier que  $R_n \leq \frac{1}{n} \sum_{i=1}^n \frac{\hat{a}_n - a}{X_i - \hat{a}_n}$ . En déduire que  $0 \leq R_n \leq \hat{a}_n - a$ .
  - (c) Montrer que  $(\hat{a}_n, \hat{b}_n)$  converge  $\mathbb{P}_{(a,b)}$  p.s. vers  $(a, b)$  lorsque  $n \rightarrow \infty$ .
  - (d) En écrivant que  $\sqrt{n}(\hat{b}_n - b) = b\hat{b}_n \times \left( \sqrt{n} \left( \frac{1}{b} - \frac{1}{n} \sum_{i=1}^n \ln(X_i - a) \right) + \sqrt{n}R_n \right)$ , vérifier que  $\sqrt{n}(\hat{b}_n - b)$  converge en loi vers  $\mathcal{N}(0, b^2)$  lorsque  $n \rightarrow \infty$ .

## Corrigé de l'examen Ecole Polytechnique 2017

**Exercice** : File de voitures.

1. On a  $B_1 = \{R_1 + D_1 - L_1 < 0\}$ ,  $B_2 = \{R_2 + D_2 - L_2 < D_1\}$  et  $B_3 = \{R_3 + D_3 - L_3 < D_2\}$ . Comme  $(R_1, D_1, L_1) \stackrel{L}{\leq} (R_2, D_2, L_2)$  et, par positivité de  $D_1$ ,  $\{R_2 + D_2 - L_2 < 0\} \subset B_2$ ,  $\mathbb{P}(B_1) = \mathbb{P}(R_2 + D_2 - L_2 < 0) \leq \mathbb{P}(B_2)$ . Comme  $(R_2, D_2, L_2, D_1) \stackrel{L}{\leq} (R_3, D_3, L_3, D_2)$ , on en déduit que  $\mathbb{P}(B_2) = \mathbb{P}(B_3)$ .
2. (a)

$$\begin{aligned} \mathbb{P}(B_1) &= \mathbb{E} [1_{\{R_1 + D_1 < L_1\}}] = \int_0^\infty \int_0^\infty \int_0^\infty 1_{\{r + \delta < \ell\}} p_L(\ell) d\ell p_D(\delta) d\delta p_R(r) dr \\ &= \int_0^\infty \int_0^\infty \bar{F}_L(r + \delta) p_D(\delta) d\delta p_R(r) dr. \end{aligned}$$

- (b) Pour  $\ell \geq 0$ ,  $\bar{F}_L(\ell) = e^{-\theta_L \ell}$  si bien que

$$\mathbb{P}(B_1) = \int_0^\infty \theta_R e^{-(\theta_R + \theta_L)r} dr \int_0^\infty \theta_D e^{-(\theta_D + \theta_L)\delta} d\delta = \frac{\theta_R}{\theta_R + \theta_L} \times \frac{\theta_D}{\theta_D + \theta_L}.$$

3. Montrons le résultat par récurrence sur  $n$ . A partir du moment où l'obstacle tombe du véhicule 0, le véhicule 1, à distance  $L_1$  de cet obstacle, a besoin de la distance  $R_1 + D_1$  pour s'arrêter en freinant. Donc l'événement "le véhicule 1 s'arrête sans collision" s'écrit  $\{R_1 + D_1 < L_1\} = A_1$ . Supposons que pour  $n \geq 2$ , l'événement "les véhicules de 1 à  $n-1$  s'arrêtent sans collision" s'écrive  $A_{n-1}$ . Dans le cas où le véhicule  $n-1$  ne percute pas son prédécesseur, à partir du moment où il commence à freiner
- il parcourt  $D_{n-1}$  avant de s'arrêter,
  - le véhicule  $n$ , à distance  $L_n$  de lui, a besoin de la distance  $R_n + D_n$  pour s'arrêter par freinage.
- Donc l'événement "les véhicules de 1 à  $n$  s'arrêtent sans collision" s'écrit  $A_{n-1} \cap B_n = A_n$ .
4. (a) Par sigma-additivité,  $1 = \mathbb{P}(D_2 = D_1) + \mathbb{P}(D_2 > D_1) + \mathbb{P}(D_1 > D_2)$  où les deux dernières probabilités sont égales puisque  $(D_1, D_2) \stackrel{L}{\leq} (D_2, D_1)$ . Donc  $\frac{1}{2} = \frac{\mathbb{P}(D_2 = D_1)}{2} + \mathbb{P}(D_2 > D_1)$  et  $\mathbb{P}(D_2 \geq D_1) = \mathbb{P}(D_2 = D_1) + \mathbb{P}(D_2 > D_1) = \frac{\mathbb{P}(D_2 = D_1) + 1}{2}$ . Comme  $\{R_2 \geq L_2\} \cap \{D_2 \geq D_1\} \subset \{R_2 + D_2 \geq L_2 + D_1\} = B_2^c$ , par indépendance  $\mathbb{P}(B_2^c) \geq \mathbb{P}(R_2 \geq L_2) \mathbb{P}(D_2 \geq D_1)$  où le premier facteur est égal à  $p$  et le second est minoré par  $1/2$ .
- (b) Les événements  $(B_{2k})_{k \geq 1}$  sont indépendants et équiprobables d'après la question 1.. Comme  $A_{2n} \subset \bigcap_{k=1}^n B_{2k}$ , on a

$$\mathbb{P}(A_{2n}) \leq \mathbb{P}\left(\bigcap_{k=1}^n B_{2k}\right) = (1 - \mathbb{P}(B_2^c))^n \xrightarrow{n \rightarrow +\infty} 0$$

car  $\mathbb{P}(B_2^c) > 0$  d'après la question précédente.

- (c) L'événement "il n'y a pas d'accident" s'écrivant  $\bigcap_{k \geq 1} A_k$ , il est inclus dans  $A_{2n}$  pour tout  $n \geq 1$  et donc de probabilité nulle.

- (d) On a alors  $B_1 = \{R_1 + d_F < L_1\}$  et  $B_k = \{R_k < L_k\}$  pour  $k \geq 2$  si bien que les événements  $(B_k)_{k \geq 1}$  sont indépendants. On a  $\{N = 1\} = B_1^c$  si bien que  $\mathbb{P}(N = 1) = \mathbb{P}(R_1 + d_F \geq L_1)$  et pour  $n \geq 2$ ,  $\{N = n\} = \bigcap_{k=1}^{n-1} B_k \cap B_n^c$  si bien que  $\mathbb{P}(N = n) = \mathbb{P}(R_1 + d_F < L_1)(1-p)^{n-2}p$ . Comme  $\mathbb{P}(N \geq 2) = \mathbb{P}(R_1 + d_F < L_1)$ , lorsque cette probabilité est strictement positive, pour  $n \in \mathbb{N}^*$ ,

$$\mathbb{P}(N - 1 = n | N \geq 2) = \frac{\mathbb{P}(N = n + 1)}{\mathbb{P}(N \geq 2)} = (1 - p)^{n-1}p.$$

La loi conditionnelle est donc la loi géométrique de paramètre  $p$ .

**Problème :** Loi de Pareto.

### I) Analyse probabiliste

1. Pour  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  mesurable bornée,

$$\mathbb{E}[\varphi((X-a)^{-b})] = \int_{a+1}^{\infty} \varphi((x-a)^{-b}) b(x-a)^{-b-1} dx = - \int_1^0 \varphi(u) du = \int_0^1 \varphi(u) du,$$

en effectuant le changement de variables  $u = (x-a)^{-b}$  tel que  $du = -b(x-a)^{-b-1} dx$ . Comme  $\ln(X-a) = -\frac{1}{b} \ln((X-a)^{-b})$ , on conclut avec le paragraphe 4.6.2 du polycopié.

- 2.

$$\begin{aligned} \mathbb{E}[(X - (1+a))^c] &= \int_{a+1}^{\infty} (x - (a+1))^c b(x-a)^{-b-1} dx \\ &= b \int_{a+1}^{\infty} \left(1 - \frac{1}{x-a}\right)^c \frac{1}{(x-a)^{b-c-1}} \frac{dx}{(x-a)^2} \\ &\stackrel{u=1/(x-a)}{=} -b \int_1^0 (1-u)^c u^{b-c-1} du \\ &= \frac{\Gamma(b-c)\Gamma(c+1)}{\Gamma(b)} \int_0^1 \frac{\Gamma(b+1)}{\Gamma(b-c)\Gamma(c+1)} u^{b-c-1} (1-u)^c du \end{aligned}$$

où la dernière intégrale vaut 1 comme intégrale de la densité  $\beta(b-c, c+1)$ .

3. Comme  $\varphi(S) = \varepsilon\varphi(Y) + (1-\varepsilon)\varphi(\frac{1}{2})$ , en utilisant la linéarité de l'espérance et

l'indépendance des variables aléatoires, on a

$$\begin{aligned}\mathbb{E}[\varphi(S)] &= \mathbb{E}[\varepsilon]\mathbb{E}[\varphi(Y)] + \mathbb{E}[1 - \varepsilon]\mathbb{E}[\varphi(1/Z)] \\ &= \frac{c}{b+c} \int_1^\infty \varphi(y)by^{-b-1}dy + \frac{b}{b+c} \int_1^\infty \varphi(1/z)cz^{-c+1}\frac{dz}{z^2} \\ &\stackrel{s=1/z}{=} \frac{c}{b+c} \int_1^\infty \varphi(y)by^{-b-1}dy + \frac{b}{b+c} \int_0^1 \varphi(s)cs^{c-1}ds \\ &= \frac{bc}{b+c} \int_0^\infty \varphi(s)(1_{\{s \leq 1\}}s^{c-1} + 1_{\{s > 1\}}s^{-b-1})ds.\end{aligned}$$

Ainsi  $S$  a pour densité  $\frac{bc}{b+c}(1_{\{0 < s \leq 1\}}s^{c-1} + 1_{\{s > 1\}}s^{-b-1})$ .

4. Pour  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  mesurable bornée, comme par indépendance de  $Y$  et  $Z$  le couple  $(Y, Z)$  a pour densité le produit de celle de  $Y$  par celle de  $Z$ ,

$$\begin{aligned}\mathbb{E}[\varphi(R, Z)] &= \mathbb{E}[\varphi(Y/Z, Z)] = \int_{z=1}^\infty \int_{y=1}^\infty \varphi(y/z, z)by^{-b-1}dycz^{-c-1}dz \\ &\stackrel{r=y/z}{=} \int_{z=1}^\infty \int_{r=1/z}^\infty \varphi(r, z)b(rz)^{-b-1}zdr cz^{-c-1}dz \\ &= \int_{\mathbb{R}^2} \varphi(r, z)bc1_{\{z > 1, rz > 1\}}r^{-b-1}z^{-b-c-1}drdz.\end{aligned}$$

Ainsi  $(R, Z)$  a pour densité  $bc1_{\{z > 1, rz > 1\}}r^{-b-1}z^{-b-c-1}$  qui ne peut pas se mettre sous forme produit à cause du terme  $1_{\{rz > 1\}}$  si bien que  $R$  et  $Z$  ne sont pas indépendantes. La densité marginale de  $R$  est donc

$$\begin{aligned}1_{\{r > 0\}}bcr^{-b-1} \int_{\max(1, 1/r)}^\infty z^{-b-c-1}dz &= 1_{\{r > 0\}}bcr^{-b-1} \left[ \frac{z^{-b-c}}{b+c} \right]_{\max(1, 1/r)}^\infty \\ &= \frac{bcr^{-b-1}}{b+c} (r^{b+c}1_{\{0 < r \leq 1\}} + 1_{\{r > 1\}}).\end{aligned}$$

On conclut que  $R$  et  $S$  ont même densité et donc même loi.

## II) Estimation statistique

5. On a

$$p_n(\mathbf{x}, a, b) = \frac{b^n}{\prod_{i=1}^n (x_i - a)^{b+1}} 1_{\{a \leq \min_{1 \leq i \leq n} x_i - 1\}}.$$

La fonction  $a \mapsto \frac{b^n}{\prod_{i=1}^n (x_i - a)^{b+1}}$  étant positive et strictement croissante sur  $] -\infty, \min_{1 \leq i \leq n} x_i[$ , on en déduit que

$$a_n(\mathbf{x}) = \min_{1 \leq i \leq n} x_i - 1.$$

6. Pour  $\mathbf{x} \in ]a+1, +\infty[^n$ ,  $l_n(\mathbf{x}, a, b) = n \ln(b) - (b+1) \sum_{i=1}^n \ln(x_i - a)$ . La dérivée partielle  $\frac{\partial l_n}{\partial b}(\mathbf{x}, a, b) = \frac{n}{b} - \sum_{i=1}^n \ln(x_i - a)$  s'annule en  $b = \frac{n}{\sum_{i=1}^n \frac{1}{\ln(x_i - a)}}$ . Comme  $\frac{\partial^2 l_n}{\partial b^2}(\mathbf{x}, a, b) = -\frac{n}{b^2} < 0$ ,  $b_n(\mathbf{x}, a) = \frac{n}{\sum_{i=1}^n \frac{1}{\ln(x_i - a)}}$ . On en déduit que l'EMV du couple  $(a, b)$  est

$$(\hat{a}_n, \hat{b}_n) = \left( \min_{1 \leq i \leq n} X_i - 1, \frac{n}{\sum_{i=1}^n \ln(X_i - \hat{a}_n)} \right).$$

7. (a) Pour  $x \geq a+1$ ,

$$\mathbb{P}_{(a,b)}(X_1 > x) = \int_x^\infty b(y-a)^{-b-1} dy = [-(y-a)^{-b}]_x^\infty = (x-a)^{-b}.$$

Comme  $\{M_n > x\} = \bigcap_{i=1}^n \{X_i > x\}$ , par indépendance,  $\mathbb{P}_{(a,b)}(M_n > x) = (\mathbb{P}_{(a,b)}(X_1 > x))^n = (x-a)^{-nb}$ . Ainsi sous  $\mathbb{P}_{(a,b)}$ ,  $M_n$  a même fonction de répartition que  $X_1$  au remplacement près de  $b$  par  $nb$ , si bien que  $M_n \sim \mathcal{P}(a, nb)$ . Pour  $nb > 1$ , en utilisant (12.1), on a

$$\mathbb{E}_{(a,b)}[\hat{a}_n - a] = \mathbb{E}_{(a,b)}[M_n - (a+1)] = \frac{\Gamma(nb-1)\Gamma(2)}{\Gamma(nb)} = \frac{1}{nb-1},$$

si bien que l'EMV est biaisé.

- (b)  $\mathbb{P}_{(a,b)}$  p.s.  $\forall i \geq 1$ ,  $X_i > a+1$ , et  $\forall n \geq 1$ ,  $M_n > a+1$ ,  $|\hat{a}_n - a|^3 = (M_n - (a+1))^3$  et (12.1) assure que pour  $n > 3/b$ ,

$$\mathbb{E}_{(a,b)}[|\hat{a}_n - a|^3] = \frac{\Gamma(nb-3)\Gamma(4)}{\Gamma(nb)} = \frac{6}{(nb-1)(nb-2)(nb-3)}.$$

Ainsi pour  $n \rightarrow \infty$ ,  $\mathbb{E}_{(a,b)}[|\sqrt{n}(\hat{a}_n - a)|^3] \sim \frac{6}{n^{3/2}b^3}$ . Donc

$$\infty > \sum_{n>3/b} \mathbb{E}_{(a,b)}[|\sqrt{n}(\hat{a}_n - a)|^3] = \mathbb{E}_{(a,b)} \left[ \sum_{n>3/b} |\sqrt{n}(\hat{a}_n - a)|^3 \right],$$

si bien que  $\mathbb{P}_{(a,b)} \left( \sum_{n>3/b} |\sqrt{n}(\hat{a}_n - a)|^3 < \infty \right) = 1$ . Comme le terme général d'une série convergente tend vers 0, on en déduit que

$$\mathbb{P}_{(a,b)} \left( \lim_{n \rightarrow \infty} |\sqrt{n}(\hat{a}_n - a)|^3 = 0 \right) = 1.$$

8. (a) D'après la question 1., les variables  $\ln(X_i - a)$  sont i.i.d. suivant la loi exponentielle  $\mathcal{E}(b)$  d'espérance  $1/b$  et variance  $1/b^2$  sous  $\mathbb{P}_{(a,b)}$ . La loi forte des grands nombres assure que  $\mathbb{P}_{(a,b)} \left( \frac{1}{n} \sum_{i=1}^n \ln(X_i - a) \rightarrow_{n \rightarrow \infty} 1/b \right) = 1$ . Le théorème de la limite centrale entraîne que sous  $\mathbb{P}_{(a,b)}$ , on a

$$\sqrt{n} \left( \frac{1}{b} - \frac{1}{n} \sum_{i=1}^n \ln(X_i - a) \right) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, 1/b^2).$$

(b) L'inégalité entraîne que

$$R_n = \frac{1}{n} \sum_{i=1}^n \ln \left( \frac{X_i - a}{X_i - \hat{a}_n} \right) \leq \frac{1}{n} \sum_{i=1}^n \frac{X_i - a - X_i + \hat{a}_n}{X_i - \hat{a}_n}.$$

Comme  $X_i - \hat{a}_n \geq 1$  pour  $i \in \{1, \dots, n\}$  et  $\hat{a}_n > a$ , on en déduit la majoration de  $R_n$ . Sa positivité découle de l'inégalité  $\hat{a}_n > a$ .

(c) La question précédente et la question 7b. entraînent que

$$\mathbb{P}_{(a,b)}((\hat{a}_n, \sqrt{n}R_n) \xrightarrow{n \rightarrow \infty} (a, 0)) = 1.$$

Comme  $\frac{1}{\hat{b}_n} = \frac{1}{n} \sum_{i=1}^n \ln(X_i - a) - R_n$ , avec la question 8a. et la continuité de  $x \mapsto 1/x$  en  $1/b$ , on conclut que  $\mathbb{P}_{(a,b)}((\hat{a}_n, \hat{b}_n) \xrightarrow{n \rightarrow \infty} (a, b)) = 1$ .

(d) Comme sous  $\mathbb{P}_{(a,b)}$ ,  $\sqrt{n}R_n$  converge p.s. vers 0 et

$$\sqrt{n} \left( \frac{1}{\hat{b}_n} - \frac{1}{n} \sum_{i=1}^n \ln(X_i - a) \right) \xrightarrow{\mathcal{L}} Z \sim \mathcal{N}(0, 1/b^2),$$

le théorème de Slutsky assure que

$$\sqrt{n} \left( \frac{1}{\hat{b}_n} - \frac{1}{n} \sum_{i=1}^n \ln(X_i - a) \right) + \sqrt{n}R_n \xrightarrow{\mathcal{L}} Z$$

lorsque  $n \rightarrow \infty$ . Comme  $b\hat{b}_n$  converge p.s. vers  $b^2$ , une nouvelle application du théorème de Slutsky permet de conclure que

$$\sqrt{n}(\hat{b}_n - b) \xrightarrow{\mathcal{L}} b^2 Z \sim \mathcal{N}(0, b^2).$$

## Examen Ecole Polytechnique 2018

**Exercice :** On a mesuré la hauteur moyenne annuelle des eaux du lac Huron en Amérique du Nord chaque année depuis 1875. Si on note  $X_i$  cette hauteur moyenne annuelle pour l'année  $1875 + i$ , avec  $i = 0, 1, \dots$  et si l'on trace  $X_{i+1}$  en fonction de  $X_i$  pour  $i = 0, 1, \dots$ , on remarque un nuage de points très alignés. Aussi pour modéliser les  $X_i$ , on propose la dynamique suivante :

$$X_n = \alpha X_{n-1} + \varepsilon_n \quad \text{pour } n \geq 1, \tag{12.2}$$

où

- $\alpha$  est un réel fixé de  $] -1, 1[$ ;
- $(\varepsilon_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires indépendantes et identiquement distribuées suivant une loi normale (=gaussienne) centrée  $\mathcal{N}(0, \sigma^2)$  avec  $\sigma^2 > 0$ ;



—  $X_0$  est une variable aléatoire indépendante de  $(\varepsilon_i)_{i \geq 1}$ .

1. Montrer que  $X_n = \alpha^n X_0 + \sum_{i=0}^{n-1} \alpha^i \varepsilon_{n-i}$  pour tout  $n \in \mathbb{N}^*$ .
2. Montrer que la suite  $(\alpha^n X_0)_{n \in \mathbb{N}}$  tend presque sûrement vers 0. Montrer que  $(\sum_{i=0}^{n-1} \alpha^i \varepsilon_{n-i})_{n \in \mathbb{N}^*}$  est une suite de variables gaussiennes convergeant en loi vers la loi  $\mathcal{N}(0, \sigma^2(1 - \alpha^2)^{-1})$ . En déduire que  $(X_n)_{n \in \mathbb{N}}$  converge en loi vers  $\mathcal{N}(0, \sigma^2(1 - \alpha^2)^{-1})$ .
3. On suppose maintenant et jusqu'à la fin de ce problème que  $X_0$  suit la loi  $\mathcal{N}(0, \sigma^2(1 - \alpha^2)^{-1})$ . Montrer que pour tout  $n \in \mathbb{N}$ ,  $X_n$  suit cette loi  $\mathcal{N}(0, \sigma^2(1 - \alpha^2)^{-1})$ .
4. Montrer que  $\text{Cov}(X_i, X_j) = \sigma^2(1 - \alpha^2)^{-1} \alpha^{|j-i|}$  pour tout  $(i, j) \in \mathbb{N}^2$ .  
En utilisant la matrice  $\Sigma = (\text{Cov}(X_i, X_j))_{0 \leq i, j \leq n}$  dont on peut montrer qu'elle est inversible (ne pas le faire!), déterminer la densité  $f_{(X_0, \dots, X_n)}$  de la loi de  $(X_0, \dots, X_n)$  (sans calculs...).

**Problème 1 :** Soit  $(m, \lambda) \in \mathbb{R} \times ]0, \infty[$ . On considère la densité par rapport à la mesure de Lebesgue sur  $\mathbb{R}$  définie par :

$$f(x) = \lambda \exp(-\lambda(x - m)) \mathbf{1}_{x \geq m}.$$

**Questions de probabilités :**

1. Soit  $X$  une variable aléatoire de densité  $f$ . Quelle est la loi de  $X - m$ ?
2. Déterminer  $\mathbb{E}(X)$  et  $\text{Var}(X)$ .  
Déterminer la fonction de répartition de  $X$ .
3. On considère  $(X_1, \dots, X_n)$  une famille de variables aléatoires indépendantes distribuées suivant la même loi que  $X$ . On note  $M_n = \min(X_1, \dots, X_n)$ .  
Montrer que  $M_n$  est une variable aléatoire à densité par rapport à la mesure de Lebesgue et déterminer la.  
Montrer que  $(M_n)$  converge en loi et identifier sa limite.
4. Soit  $Z_n = n(M_n - m)$ . Quelle est la loi de  $Z_n$ ?

**Questions de statistique :**

5. On suppose désormais que  $m$  est connu mais que  $\lambda$  est inconnu. Déterminer la vraisemblance  $p_n(x_1, \dots, x_n; \lambda)$ . Montrer que l'estimateur  $\hat{\lambda}_n$  par maximum de vraisemblance de  $\lambda$  est donné par

$$\hat{\lambda}_n = \frac{n}{\sum_{i=1}^n (X_i - m)}.$$

6. En utilisant la Delta méthode, montrer que  $\hat{\lambda}_n$  vérifie :

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \lambda^2).$$

En déduire un intervalle de confiance asymptotique de niveau 95% pour  $\lambda$ .

**Problème 2 :** Le but de cet exercice est d'étudier une modélisation aléatoire de la répartition des nombres premiers [L'idée est de proposer un modèle qui reproduit le fait que le nombre  $\pi(n)$  de nombres premiers entre 1 et  $n$  satisfait  $\pi(n) \sim \frac{n}{\ln n}$  pour  $n \rightarrow \infty$ ].

On considère la suite  $(X_k)_{k \geq 3}$  définie par  $X_k = \mathbf{1}_{\{k \text{ est un nombre premier}\}}$  et on suppose que  $(X_k)_{k \geq 3}$  est une suite de variables aléatoires indépendantes, telles que pour  $k \geq 3$  :

$$\mathbb{P}(X_k = 1) = 1 - \mathbb{P}(X_k = 0) = \frac{1}{\ln k}.$$

On note :

$$S_n = \sum_{k=3}^n X_k \quad \text{et} \quad A_n = \sum_{k=3}^n \frac{1}{\ln k}.$$

Dans la suite, on pourra utiliser le fait que  $A_n \sim \frac{n}{\ln n}$  lorsque  $n \rightarrow \infty$ .

1. Quelle est la probabilité que le nombre de nombres premiers entre 3 et  $n$  soit nul ?
2. Déterminer  $\mathbb{E}(S_n)$ .
3. On note  $\mathcal{E} = \{n \geq 3, X_n = 1\}$ . Montrer que  $\mathbb{P}(\mathcal{E} \text{ est infini}) = 1$ .
4. Calculer  $\text{Var}(S_n)$ . En déduire que  $S_n/A_n$  converge en probabilité vers 1.
5. Pour  $t > 0$ , montrer que  $\mathbb{E}[e^{tS_n}] \leq \exp(A_n(e^t - 1))$ .  
Pour  $\varepsilon \in ]0, 1[$ , en choisissant astucieusement  $t$ , montrer qu'il existe une fonction  $\phi_+ : ]0, 1[ \rightarrow ]0, +\infty[$  vérifiant

$$\mathbb{P}(S_n \geq (1 + \varepsilon)A_n) \leq \exp(-\phi_+(\varepsilon) A_n).$$

De la même manière, on peut montrer (ne pas le faire !) qu'il existe une fonction  $\phi_- : ]0, 1[ \rightarrow ]0, +\infty[$  telle que  $\mathbb{P}(S_n \leq (1 - \varepsilon)A_n) \leq \exp(-\phi_-(\varepsilon) A_n)$ .

6. Déduire de ce qui précède que  $S_n \times \frac{\ln n}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} 1$ .

## Corrigé de l'examen Ecole Polytechnique 2018

### Exercice :

1. Ceci se fait par récurrence. Vrai pour  $n = 1$ , et si vrai à l'ordre  $n$ , alors  $X_{n+1} = \alpha X_n + \varepsilon_{n+1} = \alpha^{n+1} X_0 + \sum_{i=0}^{n-1} \alpha^{i+1} \varepsilon_{n-i} + \varepsilon_{n+1} = \alpha^{n+1} X_0 + \sum_{i=0}^n \alpha^i \varepsilon_{n+1-i}$  donc vrai à l'ordre  $n + 1$ .
2. Comme  $\alpha \in ]-1, 1[$ , on a  $\alpha^n \xrightarrow[n \rightarrow \infty]{} 0$ . Comme on a pour presque tout  $\omega \in \Omega$ ,  $|X_0(\omega)| < \infty$ , on en déduit que  $\alpha^n X_0 \xrightarrow[n \rightarrow +\infty]{p.s.} 0$ .  
Comme les  $(\varepsilon_i)$  sont des variables gaussiennes indépendantes, on sait que toute combinaison linéaire des  $(\varepsilon_i)$  est une variable gaussienne, donc  $\sum_{i=0}^{n-1} \alpha^i \varepsilon_{n-i}$  a

une loi gaussienne. Elle est clairement centrée et sa variance vaut  $\sigma^2 \sum_{i=0}^{n-1} (\alpha^2)^i$  du fait de l'indépendance des  $(\varepsilon_i)$ , donc sa variance vaut  $\sigma^2(1 - \alpha^{2n})(1 - \alpha^2)^{-1} \xrightarrow[n \rightarrow \infty]{} \sigma^2(1 - \alpha^2)^{-1}$ . Comme on sait qu'une suite de variables gaussiennes dont l'espérance et la variance convergent est une suite qui converge en loi vers une variable gaussienne d'espérance et variance limites, on en déduit que  $\sum_{i=0}^{n-1} \alpha^i \varepsilon_{n-i} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2(1 - \alpha^2)^{-1})$ .

En utilisant le Lemme de Slutsky, comme  $\alpha^n X_0 \xrightarrow[n \rightarrow +\infty]{p.s.} 0$  donc  $\alpha^n X_0 \xrightarrow[n \rightarrow +\infty]{P} 0$

et comme on a  $\sum_{i=0}^{n-1} \alpha^i \varepsilon_{n-i} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2(1 - \alpha^2)^{-1})$ , on en déduit que

$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2(1 - \alpha^2)^{-1})$ .

3. Ceci se montre par récurrence, avec l'argument principal que  $X_n$  suit la loi  $\mathcal{N}(0, \sigma^2(1 - \alpha^2)^{-1})$ , donc  $\alpha X_n$  suit la loi  $\mathcal{N}(0, \alpha^2 \sigma^2(1 - \alpha^2)^{-1})$ , avec  $\varepsilon_{n+1}$  une variable gaussienne indépendante de  $X_n$ , alors  $\alpha X_n + \varepsilon_{n+1}$  est une variable gaussienne centrée de variance la somme des variances, soit  $\alpha X_n + \varepsilon_{n+1}$  suit une loi  $\mathcal{N}(0, \alpha^2 \sigma^2(1 - \alpha^2)^{-1} + \sigma^2) = \mathcal{N}(0, \sigma^2(1 - \alpha^2)^{-1})$ .
4. On a pour  $j > i$ , avec la formule de récurrence initiale,  $X_j = \alpha^{j-i} X_i + \sum_{k=0}^{j-i-1} \alpha^k \varepsilon_{j-k}$ , donc  $\text{Cov}(X_i, X_j) = \text{Cov}(X_i, \alpha^{j-i} X_i) + \text{Cov}(\sum_{k=0}^{j-i-1} \alpha^k \varepsilon_{j-k}, X_i)$  car les variables sont indépendantes et ainsi  $\text{Cov}(X_i, X_j) = \alpha^{j-i} \text{Var}(X_i) = \sigma^2 \alpha^{j-i} (1 - \alpha^2)^{-1}$ . Du fait de la symétrie  $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$ , on en déduit le résultat recherché.

Il est clair qu'il existe une matrice réelle  $\mathbf{A}$  telle que le vecteur  $(X_0, X_1, \dots, X_n)^t$  s'écrive comme  $\mathbf{AZ}$ , où  $\mathbf{Z} = (X_0, \varepsilon_1, \dots, \varepsilon_n)^t$ . Comme  $\mathbf{Z}$  est un vecteur gaussien centré car composé de variables gaussiennes centrées indépendantes, on en déduit que  $(X_0, X_1, \dots, X_n)^t$  est également un vecteur gaussien centré. Sa matrice de covariance est bien  $\mathbf{\Sigma}$  d'après la question précédente et comme celle-ci est inversible, la densité de  $(X_0, X_1, \dots, X_n)^t$  est :

$$f_{(X_0, \dots, X_n)}(x_0, \dots, x_n) = \frac{1}{(2\pi)^{(n+1)/2} \sqrt{\det \mathbf{\Sigma}}} \exp\left(-\frac{1}{2} (x_0, \dots, x_n)^t \mathbf{\Sigma}^{-1} (x_0, \dots, x_n)\right).$$

**Problème 1 :**

1. Il est facile de voir que  $X - m$  a pour loi une loi exponentielle de paramètre  $\lambda$ .
2. On déduit de la question précédente que  $\mathbb{E}(X) = m + 1/\lambda$  et  $\text{Var}(X) = 1/\lambda^2$ .
3. Pour  $x \geq m$ , on a  $F_{m_n}(x) := \mathbb{P}(m_n \leq x) = 1 - \prod_{i=1}^n \mathbb{P}(X_i > m) = 1 - (\mathbb{P}(X_1 > m))^n$  du fait de l'indépendance et de l'équidistribution. Or  $\mathbb{P}(X_1 > m) = \int_m^\infty \lambda e^{-\lambda(t-m)} dt = e^{-\lambda(x-m)}$ . On en déduit que  $F_{m_n}(x) = (1 - e^{-\lambda n(x-m)}) \mathbf{1}_{x \geq m}$  qui est une fonction différentiable presque partout, donc  $m_n$  est une variable aléatoire admettant une densité par rapport à la mesure de Lebesgue qui est la dérivée de  $F_{m_n}(x)$  (sauf en  $m$ ), donc sa densité est :

$$f_{m_n}(x) := \lambda n e^{-\lambda n(x-m)} \mathbf{1}_{x \geq m}.$$

Il est clair que pour tout  $x > m$ ,  $F_{m_n}(x) \xrightarrow{n \rightarrow \infty} 1$ , donc  $F_{m_n}$  converge vers la fonction de répartition d'une masse de Dirac en  $m$ . On en déduit que  $m_n$  converge en loi vers  $m$ , et comme converger en loi vers une constante revient à converger en probabilité vers cette constante, on en déduit que  $m_n$  converge en probabilité vers  $m$ .

4. Pour  $x \geq m$ , on a  $F_{Z_n}(x) := \mathbb{P}(Z_n \leq x) = \mathbb{P}(m_n \leq m + x/n) = 1 - e^{-\lambda x}$  : on en déduit que  $Z_n$  suit une loi exponentielle de paramètre  $\lambda$ .
5. On a  $p_n(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n f_\theta(x_i) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n (x_i - m)\right) \prod_{i=1}^n \mathbf{1}_{x_i \geq m}$  du fait de l'indépendance et de l'équidistributivité des  $X_i$ . Par ailleurs, pour les  $x_i \geq m$ , on peut considérer  $\ell(\lambda) = \ln(p_n(x_1, \dots, x_n; \lambda)) = n \ln \lambda - \lambda \sum_{i=1}^n (x_i - \hat{m})$ . En dérivant cette fonction, on obtient  $\ell'(\lambda) = n/\lambda - \sum_{i=1}^n (x_i - \hat{m})$ , un point critique étant atteint en  $\lambda = \hat{\lambda} = n/\sum_{i=1}^n (x_i - \hat{m})$  et en la dérivant encore on obtient  $\ell''(\lambda) = -n/\lambda^2$  toujours négative : le point critique est donc un maximum local et global, d'où l'expression de l'estimateur.
6. D'après le théorème de la limite centrale dont les hypothèses (v.a.i.i.d. de carré intégrable) sont ici respectées, on a

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n (X_i - m) - \frac{1}{\lambda} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{\lambda^2}\right).$$

Il suffit maintenant d'appliquer la méthode Delta avec la fonction  $g(x) = 1/x$ , de dérivée  $g'(x) = -1/x^2$ , pour obtenir le théorème voulu.

Enfin, on a encore  $\sqrt{n}(\hat{\lambda}/\lambda - 1) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$ . Avec  $q_{0.975}$  le quantile à 97.5% de la loi normale centrée réduite, on sait donc que  $\mathbb{P}(\sqrt{n}(\hat{\lambda}/\lambda - 1) \in [-q_{0.975}, q_{0.975}]) = \mathbb{P}(\lambda \in [\hat{\lambda}/(1 + q_{0.975}/\sqrt{n}), \hat{\lambda}/(1 - q_{0.975}/\sqrt{n})]) \xrightarrow[n \rightarrow \infty]{} 0.95$ , intervalle asymptotique à 95% pour  $\lambda$ .

### Problème 2 :

1. On a  $\mathbb{P}(S_n = 0) = \prod_{k=3}^n \frac{1}{\ln k}$ .
2. On a  $\mathbb{E}(S_n) = \sum_{k=3}^n \mathbb{E}(X_k) = \sum_{k=3}^n \frac{1}{\ln k}$ .
3. Soit  $E_k$  l'événement " $X_k = 1$ ". Montrer que  $\mathcal{E}$  est infini avec une probabilité 1 c'est montrer que  $\limsup_{k \rightarrow \infty} E_k = 1$ . D'après le Lemme de Borel-Cantelli, comme les  $X_k$  sont indépendantes, donc les  $E_k$  également, ceci est vrai si  $\sum_{k=3}^{\infty} \mathbb{P}(E_k) = \infty$ . Or  $\sum_{k=3}^{\infty} \mathbb{P}(E_k) = \lim_{n \rightarrow \infty} A_n = \infty$  (d'après son équivalent), donc  $\mathbb{P}(\mathcal{E} \text{ est infini}) = 1$ .
4. On a  $\text{Var}(S_n) = \sum_{k=3}^n \text{Var}(X_k)$  puisque les variables sont indépendantes, et comme les  $X_k$  sont des variables de Bernoulli, on a  $\text{Var}(X_k) = \frac{\ln k - 1}{\ln^2 k}$ . D'où  $\text{Var}(S_n) = \sum_{k=3}^n \frac{\ln k - 1}{\ln^2 k}$ . D'après l'inégalité de Bienaymé-Tchevychev, il est clair que pour tout  $\varepsilon > 0$ ,  $\mathbb{P}(|S_n/A_n - 1| \geq \varepsilon) \leq \text{Var}(S_n)/(\varepsilon A_n)^2$ . Comme on voit facilement que

$\text{Var}(S_n) \sim A_n$  quand  $n \rightarrow \infty$ , on a ainsi  $\mathbb{P}(|S_n/A_n - 1| \geq \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0 : S_n/A_n$  converge bien en probabilité vers 1.

5. On a  $\mathbb{E}[e^{tS_n}] = \prod_{k=3}^n \mathbb{E}[e^{tX_k}]$  car les variables sont indépendantes. Mais on a facilement  $\mathbb{E}[e^{tX_k}] = \frac{1}{\ln k} e^t + (1 - \frac{1}{\ln k}) = 1 + \frac{e^t - 1}{\ln k}$ . D'où  $\ln(\mathbb{E}[e^{tS_n}]) = \sum_{k=3}^n \ln(1 + \frac{e^t - 1}{\ln k})$ . Comme  $\ln(1+u) \leq u$  pour tout  $u > -1$ , on a  $\ln(\mathbb{E}[e^{tS_n}]) \leq \sum_{k=3}^n \frac{e^t - 1}{\ln k} \leq A_n(e^t - 1)$  d'où le résultat.

On a  $\mathbb{P}(S_n \geq (1+\varepsilon)A_n) = \mathbb{P}(e^{tS_n} \geq e^{(1+\varepsilon)tA_n}) \leq \exp(A_n(e^t - 1) - (1+\varepsilon)tA_n)$  grâce à l'inégalité de Markov, soit  $\mathbb{P}(S_n \geq (1+\varepsilon)A_n) \leq \exp(A_n(e^t - 1 - t - \varepsilon t))$ . Soit la fonction  $h_\varepsilon(t) = e^t - 1 - t - \varepsilon t$  pour  $t \in ]0, 1[$ . Alors  $h'_\varepsilon(t) = e^t - 1 - \varepsilon$ , donc  $h'_\varepsilon(t) = 0$  pour  $t = \ln(1 + \varepsilon)$ . De plus  $h'_\varepsilon(t) < 0$  pour  $t < \ln(1 + \varepsilon)$  et  $h'_\varepsilon(t) > 0$  pour  $t > \ln(1 + \varepsilon)$ . On en déduit que  $h_\varepsilon$  atteint son minimum en  $t = \ln(1 + \varepsilon)$  et comme  $h_\varepsilon(0) = 0$  ce minimum est négatif. En choisissant cette valeur pour  $t$  et en posant  $\phi_+(\varepsilon) = -h_\varepsilon(\ln(1 + \varepsilon)) = -\varepsilon + (1 + \varepsilon) \ln(1 + \varepsilon) > 0$ , on a bien  $\mathbb{P}(S_n \geq (1 + \varepsilon)A_n) \leq \exp(-A_n \phi_+(\varepsilon))$ .

6. De ce qui précède, on a pour tout  $\varepsilon \in ]0, 1[$  :

$$\begin{aligned} \mathbb{P}\left(\left|\frac{S_n}{A_n} - 1\right| \geq \varepsilon\right) &= \mathbb{P}(S_n \geq (1 + \varepsilon)A_n) + \mathbb{P}(S_n \leq (1 - \varepsilon)A_n) \\ &\leq \exp(-A_n \phi_+(\varepsilon)) + \exp(-A_n \phi_-(\varepsilon)). \end{aligned}$$

Comme  $A_n \sim n/\ln n$  et comme  $\phi_+(\varepsilon) > 0$  et  $\phi_-(\varepsilon) > 0$ , on a  $\sum_{n=3}^\infty \mathbb{P}(\left|\frac{S_n}{A_n} - 1\right| \geq \varepsilon) < \infty$ . Ce critère caractérisant la convergence presque sûre, on en déduit que  $\frac{S_n}{A_n} \xrightarrow[n \rightarrow +\infty]{p.s.} 1$ . De l'équivalence de  $A_n$ , on en déduit le résultat final.

## Examen Ecole Polytechnique 2019

**Exercice 1 :** Les trois questions de l'exercice sont indépendantes.

1. Soit  $X \sim \mathcal{N}(0, 1)$  et  $Y = RX$  avec  $R$  indépendante de  $X$  et telle que  $\mathbb{P}(R = -1) = \mathbb{P}(R = 1) = 1/2$ . Quelle est la loi de  $Y$ ? Quelle est la covariance entre  $X$  et  $Y$ ? Quelle est la loi de la variable  $S = X + Y$ ? Le vecteur  $(X, Y)$  est-il gaussien?
2. Soit  $X \sim \mathcal{N}(0, 1)$  et  $Y$  une variable telle que pour tout  $x$ , sachant  $X = x$ ,  $Y$  suit une loi normale de moyenne  $x$  et de variance 1. Déterminer la densité  $f(x, y)$  de la loi de  $(X, Y)$ . Calculer  $\mathbb{E}[X|Y]$ .
3. On considère des variables aléatoires  $X_i$  i.i.d. suivant une loi normale centrée réduite. On rappelle que  $\text{Var}(X_1^2) = 2$ . Déduire du théorème central limite et de la méthode delta le résultat de normalité asymptotique vérifié par la suite de variables aléatoires  $(D_n)_{n \geq 1}$ , où  $D_n = \sqrt{X_1^2 + \dots + X_n^2}$ . Déterminer un intervalle de confiance asymptotique pour  $D_n$  à 95 %, et commenter la

distance à l'origine en grande dimension d'un point tiré aléatoirement suivant une loi normale multivariée standard  $\mathcal{N}(\mathbf{0}, \mathbf{I})$ .

**Exercice 2 :** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes telles que, pour tout  $n \geq 1$ , on a  $\mathbb{E}[X_n] = 0$  et  $|X_n| \leq M$  pour un  $M > 0$  déterministe. On rappelle que  $2 \cosh(u) = e^u + e^{-u}$ .

1. Montrer, par exemple grâce à un argument de convexité, que

$$e^{\lambda x} \leq \frac{M-x}{2M} e^{-\lambda M} + \frac{x+M}{2M} e^{\lambda M} \quad \text{pour tout } x \in [-M, M] \text{ et } \lambda > 0.$$

2. En déduire que, pour tout  $i \geq 1$ ,  $\mathbb{E}(e^{\lambda X_i}) \leq \cosh(\lambda M) \leq e^{\frac{1}{2}\lambda^2 M^2}$ . Pour la dernière inégalité, on pourra comparer les séries entières des fonctions  $\cosh(u)$  et  $e^{\frac{u^2}{2}}$ .
3. Soit  $n \geq 1$ . On note  $S_n = X_1 + \dots + X_n$ . Pour tout  $\lambda > 0$  et  $x > 0$ , montrer que l'on a

$$\mathbb{P}\left(\frac{S_n}{n} \geq x\right) \leq e^{n\left(\frac{\lambda^2 M^2}{2} - \lambda x\right)}.$$

4. En déduire que, pour tout  $x > 0$ ,  $\mathbb{P}\left(\frac{S_n}{n} \geq x\right) \leq e^{-\frac{nx^2}{2M^2}}$ , puis que

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq x\right) \leq 2e^{-\frac{nx^2}{2M^2}}.$$

**Exercice 3 :** Sur l'ensemble des densités de probabilité sur  $\mathbb{R}$  (fonctions positives d'intégrale 1), on définit le carré de la distance de Hellinger  $h$  entre  $f$  et  $g$  par

$$h^2(f, g) = \frac{1}{2} \int_{\mathbb{R}} \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx.$$

1. Montrer que

$$h^2(f, g) = 1 - \int_{\mathbb{R}} \sqrt{f(x)g(x)} dx.$$

Vérifier que  $h$  est une distance sur l'ensemble des densités sur  $\mathbb{R}$  (i.e. positivité avec nullité si et seulement si égalité presque partout, symétrie et inégalité triangulaire), avec de plus  $0 \leq h(f, g) \leq 1$ .

2. Soient  $\theta > 0$  et  $f_\theta$  la densité de la loi uniforme sur  $[0, \theta]$  :  $f_\theta(x) = \frac{1}{\theta} \mathbf{1}_{[0, \theta]}(x)$ .

Montrer que  $h^2(f_\theta, f_{\theta'}) = 1 - \sqrt{\frac{\theta}{\theta'}}$  si  $\theta \leq \theta'$ .

3. Soient  $n \geq 1$  et  $X_1, \dots, X_n$  i.i.d. de loi uniforme sur  $[0, \theta]$  pour un  $\theta > 0$  inconnu. Donner l'estimateur du maximum de vraisemblance  $\hat{\theta}_n = \varphi_n(X_1, \dots, X_n)$  de  $\theta$  et identifier la fonction  $\varphi_n$ . Calculer  $\mathbb{E}_\theta \left[ \sqrt{\hat{\theta}_n} \right]$ , et en déduire que  $\mathbb{E}_\theta \left[ h^2(f_\theta, f_{\hat{\theta}_n}) \right] = \frac{1}{2n+1}$  ( $f_{\hat{\theta}_n}$  est la densité de la loi uniforme sur  $[0, \hat{\theta}_n]$ ).

4. Soit la densité

$$f_n^*(x) = 10 \left(1 - \frac{1}{n}\right) \mathbf{1}_{\{0 \leq x \leq \frac{1}{10}\}} + \frac{10}{n} \mathbf{1}_{\{\frac{9}{10} \leq x \leq 1\}}.$$

- (a) Pour tout  $n \geq 1$ , soit  $Y_n$  une variable aléatoire de loi à densité  $f_n^*$ . Montrer que la suite  $(Y_n)_{n \geq 1}$  converge en loi et identifier sa limite.
- (b) Calculer  $h^2(f_n^*, f_{\frac{1}{10}})$ .
5. Soit  $n \geq 1$ . Les variables  $X_1, \dots, X_n$  sont désormais i.i.d. de densité  $f_n^*$ , mais on les croit toujours i.i.d. suivant une densité uniforme  $f_\theta$ , avec  $\theta > 0$  inconnu. En particulier, l'estimateur  $\hat{\theta}_n = \varphi_n(X_1, \dots, X_n)$  est le même que celui défini ci-dessus en question 3.
- (a) Montrer qu'avec probabilité  $1 - (1 - \frac{1}{n})^n$ , on a :  $\frac{9}{10} \leq \hat{\theta}_n \leq 1$ .
- (b) Montrer que, sur l'événement  $\{\frac{9}{10} \leq \hat{\theta}_n \leq 1\}$ , on a :

$$h^2(f_n^*, f_{\hat{\theta}_n}) \geq 1 - \frac{1}{3} \sqrt{1 - \frac{1}{n}} - \frac{1}{3\sqrt{n}}.$$

- (c) En notant  $\mathbb{E}_n^*[\cdot]$  l'espérance par rapport au modèle où les données  $X_1, \dots, X_n$  sont i.i.d. de densité  $f_n^*$ , en déduire que  $\mathbb{E}_n^*[h^2(f_n^*, f_{\hat{\theta}_n})] \geq u_n$ , avec  $\lim_{n \rightarrow \infty} u_n = \frac{2}{3}(1 - \frac{1}{e})$ .
- (d) Conclure quant à la robustesse de l'estimateur du maximum de vraisemblance à une mauvaise spécification du modèle.

## Corrigé de l'examen Ecole Polytechnique 2019

### Exercice 1 :

1. En notant  $\Phi$  la fonction de répartition de la loi normale standard, on a pour tout réel  $y$ , via l'indépendance de  $R$  et  $X$  :

$$\mathbb{P}(Y \leq y) = \mathbb{P}(R = -1, X \geq -y) + \mathbb{P}(R = 1, X \leq y) = \Phi(y),$$

donc  $Y \sim \mathcal{N}(0, 1)$ . Pour la covariance entre  $X$  et  $Y$ , les variables  $R$  et  $X$  étant centrées avec  $R$  indépendante de  $X$ , il vient :

$$\text{Cov}(X, Y) = \mathbb{E}[XY] = \mathbb{E}[RX^2] = 0.$$

Par la même méthode que ci-dessus, la fonction de répartition de  $S$  s'écrit

$$\mathbb{P}(S \leq s) = \frac{1}{2}(\mathbf{1}_{s \geq 0} + \Phi(s/2)).$$

La variable  $S = X + Y$  n'étant pas gaussienne, le vecteur  $(X, Y)$  n'est pas gaussien.

2. En notant  $f(x, y)$  la densité de la loi de  $(X, Y)$ , on a

$$f(x, y) = f_X(x)f_{Y|X=x}(y) = \frac{1}{2\pi} \exp\left(-\frac{x^2}{2}\right) \exp\left(-\frac{(y-x)^2}{2}\right)$$

En notant  $f_{X|Y=y}(x)$  la densité conditionnelle de  $X$  sachant  $Y = y$ , on a (en tant que fonction de  $x$ )

$$f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)} \propto f(x, y) \propto \exp(-(x - y/2)^2),$$

ce qui montre que, sachant  $Y = y$ ,  $X$  suit une loi  $\mathcal{N}(y/2, 1/2)$ . En particulier, on en déduit que  $\mathbb{E}[X|Y] = Y/2$ .

3. Puisque  $\mathbb{E}[X_1^2] = 1$  et  $\text{Var}(X_1^2) = 2$ , le théorème central limite appliqué aux variables  $X_i^2$  donne

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n X_i^2 - 1 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 2).$$

La méthode delta appliquée avec  $g(x) = \sqrt{x}$  permet d'en déduire que

$$\sqrt{n} \left( \frac{D_n}{\sqrt{n}} - 1 \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1/2),$$

ou encore

$$\sqrt{2} (D_n - \sqrt{n}) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

En arrondissant le quantile d'ordre 0.975 d'une loi normale standard à 2, ceci signifie que :

$$\mathbb{P}(\sqrt{n} - \sqrt{2} \leq D_n \leq \sqrt{n} + \sqrt{2}) \xrightarrow[n \rightarrow \infty]{} 0.95.$$

En grande dimension, un point tiré aléatoirement suivant une loi normale standard a environ 95% de chances d'être à distance inférieure à  $\sqrt{2}$  de  $\sqrt{n}$ .

### Exercice 2 :

1. On commence par écrire que

$$\lambda x = \frac{M-x}{2M} \times (-\lambda M) + \frac{x+M}{2M} \times (\lambda M).$$

Par convexité de la fonction exponentielle, on en déduit que, pour tout  $x \in [-M, M]$  et  $\lambda > 0$ ,

$$\exp(\lambda x) \leq \frac{M-x}{2M} \exp(-\lambda M) + \frac{x+M}{2M} \exp(\lambda M).$$



2. Puisque, pour tout  $i \in \mathbb{N}$ ,  $X_i$  est centrée et  $|X_i| \leq M$ , il en découle que

$$\mathbb{E}(\exp(\lambda X_i)) \leq \cosh(\lambda M).$$

Les séries entières des fonctions  $\cosh(u)$  et  $\exp(u^2/2)$  sont

$$\cosh(u) = \sum_{n=0}^{\infty} \frac{u^{2n}}{(2n)!} \quad \text{et} \quad \exp(u^2/2) = \sum_{n=0}^{\infty} \frac{u^{2n}}{2^n \times n!},$$

or  $2^n \times n!$  correspond uniquement au produit des facteurs pairs de  $(2n)!$ . Donc  $2^n \times n! \leq (2n)!$  et on trouve

$$\mathbb{E}(\exp(\lambda X_i)) \leq \cosh(\lambda M) \leq \exp(\lambda^2 M^2/2).$$

3. Par croissance stricte de la fonction exponentielle et via l'inégalité de Markov, on a pour tout  $x > 0$  et tout  $\lambda > 0$ ,

$$\mathbb{P}\left(\frac{S_n}{n} \geq x\right) = \mathbb{P}(\exp(\lambda S_n) \geq \exp(\lambda n x)) \leq \exp(-n \lambda x) \mathbb{E}[\exp(\lambda S_n)].$$

L'indépendance des  $X_i$  et l'inégalité précédente donnent bien

$$\mathbb{P}\left(\frac{S_n}{n} \geq x\right) \leq \exp\left(n\left(\frac{\lambda^2 M^2}{2} - \lambda x\right)\right).$$

4. Ceci étant vrai pour tout  $\lambda > 0$ , un calcul élémentaire de minimum permet de conclure que

$$\mathbb{P}\left(\frac{S_n}{n} \geq x\right) \leq \inf_{\lambda > 0} \exp\left(n\left(\frac{\lambda^2 M^2}{2} - \lambda x\right)\right) = \exp\left(-\frac{nx^2}{2M^2}\right).$$

De plus,

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq x\right) = \mathbb{P}\left(\frac{S_n}{n} \geq x\right) + \mathbb{P}\left(\frac{-S_n}{n} \geq x\right),$$

avec  $-S_n = (-X_1) + \dots + (-X_n)$ , les variables  $-X_i$  centrées et  $|-X_i| \leq M$ , donc la même inégalité s'applique, d'où

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq x\right) \leq 2 \exp\left(-\frac{nx^2}{2M^2}\right).$$

### Exercice 3 :

1. Puisque

$$h(f, g) = \frac{1}{\sqrt{2}} \|\sqrt{f} - \sqrt{g}\|_2,$$

il est clair que  $h$  est une distance. La formule

$$h^2(f, g) = 1 - \int_{\mathbb{R}} \sqrt{f(x)g(x)} \, dx$$

découle du fait que  $f$  et  $g$  intègrent à 1. Une distance étant positive et le dernier terme étant positif, on en déduit aussi que  $0 \leq h(f, g) \leq 1$ .

2. Soit  $\theta > 0$  et  $f_\theta$  la densité de la loi uniforme sur  $[0, \theta]$ . Le fait que  $h^2(f_\theta, f_{\theta'}) = 1 - \sqrt{\frac{\theta}{\theta'}}$  si  $\theta \leq \theta'$  est une conséquence directe de la formule précédente.
3. Soit  $X_1, \dots, X_n$  i.i.d. de loi uniforme sur  $[0, \theta]$  pour un  $\theta > 0$  inconnu. La vraisemblance associée à cet échantillon s'écrit

$$L_n(\theta) = \frac{1}{\theta^n} \mathbf{1}_{\theta \geq X_{(n)}},$$

où  $X_{(n)} = \max_{i=1, \dots, n}(X_i)$ , qui est maximale pour  $\theta = \hat{\theta}_n = X_{(n)}$ . La variable  $\sqrt{\hat{\theta}_n}$  étant positive (et inférieure à  $\sqrt{\theta}$ ), on peut écrire que

$$\mathbb{E}_\theta \left[ \sqrt{\hat{\theta}_n} \right] = \int_0^\infty \mathbb{P} \left( \sqrt{\hat{\theta}_n} > x \right) dx = \int_0^{\sqrt{\theta}} \mathbb{P} \left( \sqrt{\hat{\theta}_n} > x \right) dx,$$

d'où

$$\mathbb{E}_\theta \left[ \sqrt{\hat{\theta}_n} \right] = \int_0^{\sqrt{\theta}} \left( 1 - \mathbb{P} \left( \sqrt{\hat{\theta}_n} \leq x \right) \right) dx = \int_0^{\sqrt{\theta}} \left( 1 - \left( \frac{x^2}{\theta} \right)^n \right) dx,$$

ce qui donne finalement

$$\mathbb{E}_\theta \left[ \sqrt{\hat{\theta}_n} \right] = \sqrt{\theta} \left( 1 - \frac{1}{2n+1} \right).$$

Puisque presque sûrement  $0 < \hat{\theta}_n \leq \theta$ , le calcul d'espérance par conditionnement et les résultats précédents donnent

$$\mathbb{E}_\theta \left[ h^2(f_\theta, f_{\hat{\theta}_n}) \right] = \mathbb{E}_\theta \left[ \mathbb{E}_\theta \left[ h^2(f_\theta, f_{\hat{\theta}_n}) \mid \hat{\theta}_n \right] \right] = \mathbb{E}_\theta \left[ 1 - \sqrt{\frac{\hat{\theta}_n}{\theta}} \right] = \frac{1}{2n+1}.$$

4. Soit la densité

$$f_n^*(x) = 10 \left( 1 - \frac{1}{n} \right) \mathbf{1}_{0 \leq x \leq 1/10} + \frac{10}{n} \mathbf{1}_{9/10 \leq x \leq 1}.$$

- (a) Pour toute fonction continue bornée  $\varphi$ , il est clair que

$$\mathbb{E}[\varphi(Y_n)] = 10 \left( 1 - \frac{1}{n} \right) \int_0^{1/10} \varphi(x) dx + \frac{10}{n} \int_{9/10}^1 \varphi(x) dx \xrightarrow{n \rightarrow \infty} 10 \int_0^{1/10} \varphi(x) dx,$$

ce qui montre que  $(Y_n)$  converge en loi vers une loi uniforme sur  $[0, 1/10]$ .

- (b) Un calcul immédiat donne

$$h^2(f_n^*, f_{1/10}) = 1 - \int_{\mathbb{R}} \sqrt{f_n^*(x) f_{1/10}(x)} dx = 1 - \sqrt{1 - \frac{1}{n}},$$

quantité qui tend bien vers 0 lorsque  $n$  tend vers l'infini.

5. Soit  $n > 1$ .

(a) On rappelle que  $\hat{\theta}_n = \max_{i=1, \dots, n}(X_i)$ . Comme les  $X_i$  sont i.i.d. et que  $\mathbb{P}_n^*(X_1 \geq 9/10) = \int_{9/10}^1 f_n^*(x) dx = 1/n$  :

$$\mathbb{P}_n^*(\hat{\theta}_n \geq 9/10) = 1 - \mathbb{P}_n^*\left(\max_{i=1, \dots, n}(X_i) \leq 9/10\right) = 1 - \mathbb{P}_n^*(X_1 \leq 9/10)^n = 1 - (1 - 1/n)^n$$

(b) Sur l'événement  $\{9/10 \leq \hat{\theta}_n \leq 1\}$ , on a :

$$\begin{aligned} h^2(f_n^*, f_{\hat{\theta}_n}) &= 1 - \int_0^{1/10} \sqrt{10\left(1 - \frac{1}{n}\right) \frac{1}{\hat{\theta}_n}} dx - \int_{9/10}^{\hat{\theta}_n} \sqrt{\frac{10}{n} \frac{1}{\hat{\theta}_n}} dx \\ &= 1 - \sqrt{\left(1 - \frac{1}{n}\right) \frac{1}{10\hat{\theta}_n}} - \left(\hat{\theta}_n - \frac{9}{10}\right) \sqrt{\frac{10}{n} \frac{1}{\hat{\theta}_n}} \\ &\geq 1 - \frac{1}{3} \sqrt{1 - \frac{1}{n}} - \frac{1}{3\sqrt{n}}. \end{aligned}$$

(c) En notant  $\mathbb{E}_n^*[\cdot]$  l'espérance par rapport au modèle où les données  $X_1, \dots, X_n$  sont i.i.d. de densité  $f_n^*$ , on trouve :

$$\begin{aligned} \mathbb{E}_n^*[h^2(f_n^*, f_{\hat{\theta}_n})] &\geq \mathbb{E}_n^*[h^2(f_n^*, f_{\hat{\theta}_n}) \mathbf{1}_{\{9/10 \leq \hat{\theta}_n \leq 1\}}] \\ &= \mathbb{E}_n^*[h^2(f_n^*, f_{\hat{\theta}_n}) | 9/10 \leq \hat{\theta}_n \leq 1] \mathbb{P}_n^*(9/10 \leq \hat{\theta}_n \leq 1) \\ &\geq \left(1 - \frac{1}{3} \sqrt{1 - \frac{1}{n}} - \frac{1}{3\sqrt{n}}\right) \left(1 - (1 - 1/n)^n\right) =: u_n \end{aligned}$$

avec  $\lim_{n \rightarrow \infty} u_n = 2/3 \times (1 - 1/e)$ .

(d) On constate que, alors que  $f_n^*$  est très proche de  $f_{1/10}$  quand  $n$  est grand d'après la question 4, la distance entre la densité  $f_n^*$  et la densité estimée  $f_{\hat{\theta}_n}$  ne tend pas vers 0. L'estimateur du maximum de vraisemblance n'est donc pas robuste à une mauvaise spécification du modèle.



# Bibliographie

## Ouvrages généraux :

M. Benaïm, N. El Karoui : *Promenade aléatoire*. Chaînes de Markov et simulations ; martingales et stratégies, Editions de l'Ecole Polytechnique, 2004.

P. Billingsley : *Probability and Measure*, Wiley, New York (1979).

L. Breiman : *Probability*, Addison Wesley 1968.

J.F. Delmas, B. Jourdain : *Modèles aléatoires : applications aux sciences de l'ingénieur et du vivant*, Springer 2006.

W. Feller : *An introduction to Probability Theory and its Applications*, 2 Vol. Wiley, 1957.

D. Foata, A. Fuchs : *Calcul des probabilités : cours et exercices corrigés*, Dunod 2003.

H.O. Georgii, *Stochastics. Introduction to probability and statistics*. De Gruyter, 2008.

C. Graham : *Chaînes de Markov*, Mathématiques Appliquées pour le Master/SMAI, Dunod, 2008.

G. Grimmett, D. Stirzaker : *Probability and Random Processes*, Oxford University Press, 1992.

J. Jacod, P. Protter : *L'essentiel en théorie des probabilités*, Cassini, 2003.

B. Jourdain : *Probabilités et statistique*, Ellipses, 2009.

Y. Lacroix, L. Mazliak : *Probabilités, variables aléatoires, convergences, conditionnement*, ellipses, 2006.

E. Pardoux : *Processus de Markov et applications. Algorithmes, réseaux, génome et finance*, Mathématiques Appliquées pour le Master/SMAI, Dunod, 2007.

J. Neveu : *Bases mathématiques du calcul des probabilités*, Masson 1964.

**Pour ceux qui veulent tout savoir sur les probabilités du quotidien :**

G. Pagès, C. Bouzitat : *En passant par hasard... Les probabilités de tous les jours*, Vuibert 1999.

**Pour les passionnés d'Histoire des Sciences et de Philosophie :**

P.S. Laplace : *Essai philosophique sur les probabilités*, Christian Bourgeois 1986.

I. Hacking : *L'émergence de la Probabilité*, Seuil 1975.

**Quelques romans probabilistes :**

D. Kehlmann : *Les arpenteurs du monde*, Actes Sud 2006. (Les tribulations de Humboldt et Gauss)

M. Petit : *L'équation de Kolmogoroff*, Folio 2003.

**Pour un choix aléatoire dans une lecture poétique :**

R. Queneau : *100 000 milliards de poèmes*, Gallimard 1961.

**Bibliographie des Mathématiciens :** <http://turnbull.mcs.st-and.ac.uk/~history/>

# Index

- écart-type, 48
- échantillon, 157
- événement, 15
- événements indépendants, 32
  
- approximation linéaire, 79
  
- biais, 159
  
- coefficient de corrélation, 77
- convergence d'estimateur, 159
- convergence de test, 196
- convergence dominée, 122
- convergence en loi, 144
- convergence en moyenne, 120
- convergence en probabilités, 120
- convergence monotone, théorème, 72
- convergence presque-sûre, 120
- convolution, 111
- corrélation, 77
- covariance, 77
  
- densité, 91
- densité marginale, 103
- droite des moindres carrés, 79
  
- ensemble dénombrable, 38
- ensemble négligeable, 22
- espérance conditionnelle de  $Y$  sachant  $X = x_i$ , 58
- espérance conditionnelle pour des v.a. discrètes, 58
- espérance de v.a. discrètes, 45
- espérance de v.a. réelles, 74
- espace de probabilité, 22
- espace de probabilité produit, 34
  
- espace fondamental, 15
- estimateur du maximum de vraisemblance, 168
- expériences aléatoires indépendantes, 34
  
- Fatou, lemme, 122
- fonction caractéristique, 134
- fonction de répartition, 83
- fonction génératrice, 50
- fonction indicatrice, 48
- formule de Bayes, 30
- formule de Huygens, 48
- formule des probabilités composées, 30
- formule des probabilités totales, 30
- fréquence de réalisation, 18
  
- inégalité de Bienaymé-Chebyshev, 79
- inégalité de Cauchy-Schwarz, 80
- inégalité de Jensen, 81
- indépendance, 60
- intervalle de confiance, 175
  
- lemme de Borel-Cantelli, 35
- lemme de Fatou, 122
- log-vraisemblance, 169
- loi des grands nombres, 125
- loi binomiale, 53
- loi conditionnelle pour des v.a. discrètes, 57
- loi d'une v.a. discrète, 44
- loi de Bernoulli, 52
- loi de chi-deux, 143
- loi de Poisson, 55
- loi hypergéométrique, 25
- lois marginales, 57

- méthode de Monte-Carlo, 11, 129
- mesure de Dirac, 24
- mesure de Lebesgue, 68
- mesure de Lebesgue sur  $\mathbb{R}^n$ , 68
- modèle statistique, 156
- moment, 50
  
- niveau d'un test, 195
- normalité asymptotique, 160
  
- partition, 37
- phénomènes aléatoires, 9
- probabilité, 19
- probabilité conditionnelle, 30
- probabilité produit, 34
- probabilité uniforme, 24
- propriété de non-vieillessement, 97
- propriété vraie presque-sûrement, 22
  
- quantile, 176
  
- risque quadratique moyen, 160
  
- sondages, 187
- suite d'événements indépendants, 33
- suite de vecteurs aléatoires indépendantes, 86
  
- table de la loi de Gauss, 99
- test du chi-deux, 199
- théorème central limite multi-dimensionnel, 152
- théorème de convergence dominée, 122
- théorème de convergence monotone, 72
- théorème de Fubini, 102
- théorème de Lévy, 147
- théorème de la limite centrale, 150
- théorème de Slutsky, 148
- transformée de Laplace, 139
- tribu, 16
- tribu borélienne, 18
- tribu engendrée par une partie de  $\Omega$ , 17
- tribu produit, 34
  
- variable aléatoire de Bernoulli, 52
- variable aléatoire étagée, 71
- variable aléatoire centrée réduite, 77
- variable aléatoire de loi normale, 98
- variable aléatoire de chi-deux, 143
- variable aléatoire exponentielle, 96
- variable aléatoire géométrique, 54
- variable aléatoire normale centrée réduite, 98
- variable aléatoire réelle, 70
- variable aléatoire réelle de carré intégrable, 76
- variable aléatoire uniforme, 95
- variable binomiale, 53
- variable de Poisson, 55
- variable discrète de carré intégrable, 48
- variable discrète intégrable, 47
- variable gaussienne, 98
- variables aléatoires discrètes, 43
- variance, 48
- variance d'une variable aléatoire réelle, 76
- variance empirique non-biaisée, 165
- vecteur espérance, 78
- vecteur gaussien, 141
- vecteurs aléatoires indépendants, 85
- vraisemblance, 168