

Aléatoire

Josselin Garnier, Sylvie Méléard

Département de Mathématiques Appliquées
Ecole Polytechnique

Septembre 2017

Table des matières

1	Introduction	7
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
2	Espace de probabilité	15
2.1	Le langage des probabilités	15
2.1.1	Expériences et événements	15
2.1.2	Probabilité - Premières propriétés	19
2.2	Probabilité sur un espace fini - Calcul combinatoire	21
2.2.1	Définition	21
2.2.2	Probabilité uniforme	23
2.2.3	Modèles d'urnes	23
2.3	Définition générale des probabilités	28
2.3.1	Pourquoi la définition précédente ne suffit-elle pas?	28
2.3.2	Les ensembles dénombrables	28
2.3.3	Tribu	29
2.3.4	Définition d'une probabilité	31
2.3.5	Probabilités sur un espace dénombrable	34
2.4	Loi d'une variable aléatoire	35
2.5	Conditionnement et indépendance	38
2.5.1	Probabilités conditionnelles	38
2.5.2	Indépendance	41
2.5.3	Le lemme de Borel-Cantelli	44
2.6	Exercices sur le chapitre 2	46

3	Espace fini ou dénombrable	49
3.1	Prérequis : quelques résultats utiles sur les séries	49
3.2	Variables aléatoires discrètes	51
3.3	Espérance des variables aléatoires discrètes	52
3.3.1	Définition	52
3.3.2	Propriétés de l'espérance des variables aléatoires discrètes	54
3.3.3	Variance et écart-type	55
3.3.4	Un résultat fondamental - Moments d'une variable aléatoire	57
3.4	Fonction génératrice d'une variable aléatoire à valeurs entières	58
3.5	Variables aléatoires discrètes usuelles	60
3.5.1	Variable aléatoire de Bernoulli	60
3.5.2	Variable aléatoire binomiale	61
3.5.3	Probabilité de succès et variable aléatoire géométrique	63
3.5.4	Variable aléatoire de Poisson	64
3.6	Lois conditionnelles et indépendance	66
3.6.1	Lois conditionnelles	66
3.6.2	Espérance conditionnelle	68
3.6.3	Variables aléatoires indépendantes	70
3.6.4	Somme de variables aléatoires indépendantes	72
3.7	Exercices sur le chapitre 3	74
4	Variables aléatoires réelles et vecteurs aléatoires	79
4.1	Les variables aléatoires réelles	79
4.2	Les lois de variables aléatoires réelles	81
4.2.1	Fonction de répartition	81
4.2.2	Variables aléatoires de loi à densité	85
4.2.3	Variable aléatoire uniforme sur $[0, 1]$ et générateurs de nombres aléatoires	88
4.2.4	Simulation d'une variable aléatoire par inversion de la fonction de répartition	89
4.3	Espérance des variables aléatoires réelles	90
4.3.1	Définition	90
4.4	Variables aléatoires de carré intégrable	94
4.4.1	Variance et Covariance	94
4.4.2	Approximation linéaire	96
4.5	Calcul de l'espérance pour une variable aléatoire à densité	97
4.5.1	Un résultat général fondamental	97
4.5.2	Calculs d'espérances dans le cas avec densité	97
4.6	Exemples fondamentaux de variables à densité	98
4.6.1	Variable aléatoire uniforme sur $[a, b]$	98
4.6.2	Variable aléatoire exponentielle	99
4.6.3	Variable aléatoire de loi gamma	102

4.6.4	Variables aléatoires normales (ou variables gaussiennes)	103
4.7	Des inégalités fameuses	108
4.7.1	Inégalité de Bienaymé-Chebyshev	108
4.7.2	Inégalité de Cauchy-Schwarz	109
4.7.3	Inégalité de Jensen	109
4.8	Vecteurs aléatoires	110
4.8.1	Vecteurs aléatoires	112
4.8.2	Moments d'un vecteur aléatoire	113
4.8.3	Densités marginales et conditionnelles	114
4.9	Variables aléatoires indépendantes	118
4.9.1	Indépendance de deux variables aléatoires	118
4.9.2	Suite de variables aléatoires indépendantes	120
4.10	Calculs de lois	122
4.10.1	Un théorème d'identification	122
4.10.2	Recherche de densité	123
4.11	Simulation de suites de variables aléatoires indépendantes	127
4.11.1	Inversion de la fonction de répartition	127
4.11.2	Méthode du rejet	128
4.12	Exercices sur le chapitre 4	131
5	Convergences et loi des grands nombres	137
5.1	Convergences de variables aléatoires	138
5.2	La loi des grands nombres	143
5.3	Méthode de Monte-Carlo	147
5.4	Exercices sur le chapitre 5	149
6	Fonctions caractéristiques et convergence en loi	153
6.1	La fonction caractéristique	153
6.1.1	Définition et premières propriétés	153
6.1.2	Exemples	155
6.1.3	Propriété fondamentale	157
6.1.4	Somme de vecteurs aléatoires indépendants	159
6.1.5	Fonction caractéristique et moments	160
6.2	Vecteurs gaussiens	161
6.3	Convergence en loi	164
6.4	Le théorème de la limite centrale	168
6.5	Exercices sur le chapitre 6	172
7	Statistique	175
7.1	Estimation ponctuelle	176
7.1.1	Qualités d'un estimateur	178
7.1.2	Estimateurs empiriques	181
7.1.3	Méthode de substitution	186
7.1.4	Méthode des moments	186

7.1.5	Maximum de vraisemblance	188
7.2	Intervalle de confiance	192
7.2.1	Intervalle de confiance et estimation	192
7.2.2	Intervalles exacts pour le modèle gaussien	195
7.2.3	Résultats asymptotiques	203
7.3	Tests	208
7.3.1	Tests et erreurs	208
7.3.2	Modèle gaussien	211
7.3.3	Test du χ^2 (test du chi-deux)	213
7.4	Exercices sur le chapitre 7	222
8	Modèles dynamiques aléatoires	227
8.1	Marche aléatoire	228
8.2	Processus de branchement	232
8.2.1	Somme aléatoire de variables aléatoires indépendantes	232
8.2.2	Processus de branchement	234
8.2.3	Percolation sur un arbre	239
8.3	Files d'attente	241
8.3.1	Un modèle simple en temps discret	241
8.3.2	Stabilité : étude analytique	243
8.4	Suites récurrentes aléatoires discrètes	246
8.4.1	Probabilités de transition	246
8.4.2	Stabilité	249
8.5	Exercices sur le chapitre 8	250
9	Corrections des exercices	255
9.1	Corrigés des exercices du chapitre 2	255
9.2	Corrigés des exercices du chapitre 3	260
9.3	Corrigés des exercices du chapitre 4	265
9.4	Corrigés des exercices du chapitre 5	271
9.5	Corrigés des exercices du chapitre 6	274
9.6	Corrigés des exercices du chapitre 7	278
9.7	Corrigés des exercices du chapitre 8	287
10	Textes et corrigés d'examens	293
	Bibliographie	321
	Index	323

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 1.2.1 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 lancé 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.

Des simulations, proposées dans le cours de l'École Polytechnique, peuvent accompagner la lecture de cet ouvrage et en illustrer la compréhension. Elles se trouvent à l'adresse :

http://www.cmapx.polytechnique.fr/~benaych/aleatoire_index.html. Nous remercions en cela la participation de leur auteur Florent Benaych-Georges.

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 Expériences et événements

a) Expérience Aléatoire

Définition 2.1.1 Nous appelons **expérience aléatoire** une expérience \mathcal{E} qui, reproduite dans des conditions identiques, peut conduire à plusieurs résultats possibles, et dont on ne peut prévoir le résultat par avance. L'espace de tous les résultats possibles, appelé **espace fondamental** (associé à l'expérience), sera noté Ω . Un résultat possible de l'expérience est noté classiquement ω . Ainsi, $\omega \in \Omega$.

Les jeux de hasard, tels Pile ou Face, jeux de cartes, loterie, fournissent des exemples d'expériences aléatoires pour lesquels Ω est fini, mais Ω peut être un espace beaucoup plus compliqué.

- Exemple 1. Lancer de deux pièces à Pile ou Face : $\Omega = \{PP, PF, FP, FF\}$.
- Exemple 2. Lancer d'un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Exemple 3. Envoi d'une fléchette sur une cible circulaire de 30 cm de diamètre. L'expérience consiste à décrire l'impact de la flèche dans un repère orthonormé de centre le centre de la cible : $\Omega = \{(x, y) \in \mathbb{R}^2, \sqrt{x^2 + y^2} \leq 15\}$.
- Exemple 4. Durée de vie d'une ampoule électrique : $\Omega = [0, +\infty[$.
- Exemple 5 : Roméo attend Juliette qui lui a promis d'arriver entre minuit et une heure. Quel va être son temps d'attente ? $\Omega = [0, 1]$.
- Exemple 6. Temps de passage des véhicules à une borne de péage : $\Omega = (\mathbb{R}_+)^{\mathbb{N}}$.
- Exemple 7. L'observation d'un prix d'actif financier sur un intervalle de temps $[t_1, t_2]$ conduit à prendre pour Ω l'ensemble $C([t_1, t_2], \mathbb{R}_+)$ des fonctions continues sur $[t_1, t_2]$, à valeurs réelles positives.
- Exemple 8. L'étude de la vitesse d'une molécule dans un gaz raréfié sur un intervalle de temps $[t_1, t_2]$ conduit à prendre pour Ω l'ensemble des fonctions continues à droite et avec limites à gauche sur $[t_1, t_2]$, à valeurs dans \mathbb{R}^3 .

Cette longue liste d'exemples montre que l'espace Ω peut varier énormément dans sa structure, d'une expérience à l'autre. Cela permet de réaliser la richesse de la théorie qu'il faut mettre en place, pour créer un modèle qui englobe tous ces cas.

Nous verrons également ultérieurement que le modèle abstrait que nous allons construire permettra de s'affranchir du fait que Ω décrit précisément tous les résultats possibles de l'expérience.

b) Evénements

Définition 2.1.2 Nous appelons **événement** (associé à l'expérience \mathcal{E}) un sous-ensemble de Ω dont on peut dire au vu de l'expérience s'il est réalisé ou non. Un événement est donc une partie de Ω .

Ainsi, si l'expérience consiste en un lancer de deux dés,

$$A = \{ \text{la somme des deux dés est inférieure à 4} \}$$

est un événement, mais l'ensemble

$$B = \{ \text{le résultat du premier dé lancé est un nombre inférieur à 4} \}$$

n'en est pas un si Ω ne contient que les résultats non ordonnés des tirages.

Pour notre Roméo, l'ensemble "Juliette se fait attendre plus de 3/4 d'heure" est l'événement $]3/4, 1]$.

Si l'on s'intéresse au prix d'un actif financier sur le temps $[t_1, t_2]$, l'ensemble

$$A = \{ \text{le prix est inférieur au seuil } \alpha \} = \left\{ \omega \in C([t_1, t_2], \mathbb{R}), \sup_{t \in [t_1, t_2]} |\omega(t)| \leq \alpha \right\}$$

est un événement.

Ainsi, **les événements sont des ensembles**. Nous allons utiliser le formalisme de la théorie des ensembles, en particulier les opérations élémentaires sur les ensembles, pour décrire diverses possibilités de réalisations d'événements.

c) Rappels sur les ensembles

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

Une partie A de Ω est aussi un ensemble, appelé sous-ensemble de Ω . Dans ce cas, A est dit inclus dans Ω , 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 Ω) est l'ensemble des points

de Ω 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 Ω , $A \setminus B$ 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 $A \cup (B \cap C) = (A \cup B) \cap C$ et $A \cap (B \cup C) = (A \cap B) \cup C$, 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 Ω 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 $\cup_{i \in I} A_i = \Omega$.

d) Modélisation ensembliste des événements

Les événements étant des ensembles, (rappelons-nous qu'une partie de Ω décrit un sous-ensemble de résultats possibles de l'expérience), nous pourrions effectuer les opérations ensemblistes précédemment décrites, avec l'interprétation suivante.

Correspondance entre opérations sur les ensembles et sur les événements :

Si A et B sont deux événements,

- **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 Ω est l'événement certain (tous les résultats de l'expérience prennent leurs valeurs dans Ω).
- IMPOSSIBLE : \emptyset est l'événement impossible.

Nous notons par \mathcal{A} l'ensemble de tous les événements. Nous pourrions prendre $\mathcal{A} = \mathcal{P}(\Omega)$, ensemble de toutes les parties de Ω , mais pas toujours et nous verrons pourquoi dans la suite de ce cours.

Remarque fondamentale : *Pour que la modélisation soit cohérente avec l'intuition, \mathcal{A} doit être stable par les opérations ensemblistes ci-dessus : si $A, B \in \mathcal{A}$, alors $A \cap B \in \mathcal{A}$, $A \cup B \in \mathcal{A}$, $A^c \in \mathcal{A}$, mais aussi $\Omega \in \mathcal{A}$, $\emptyset \in \mathcal{A}$.*

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 ,

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

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). 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 $f_n(A)$:

$$\mathbb{P}(A) = \text{limite de } f_n(A) \text{ quand } n \uparrow +\infty. \quad (2.1.1)$$

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 :

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

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

nous en déduisons immédiatement que

Proposition 2.1.3 *Une probabilité est définie sur les ensembles aléatoires liés à l'expérience et vérifie les propriétés essentielles suivantes :*

- $0 \leq \mathbb{P}(A) \leq 1,$ (2.1.2)
- $\mathbb{P}(\Omega) = 1,$ (2.1.3)
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ si $A \cap B = \emptyset,$ (2.1.4)

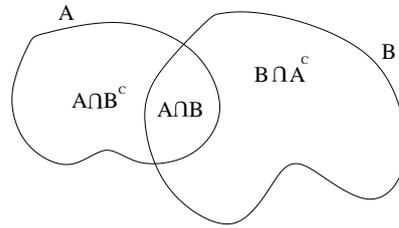
et il en découle que

Corollaire 2.1.4 *Une probabilité vérifie de plus :*

- $\mathbb{P}(\emptyset) = 0$ (2.1.5)
- $\mathbb{P}(A) + \mathbb{P}(A^c) = 1,$ (2.1.6)
- $\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$, si les A_i sont deux-à-deux disjoints, (2.1.7)
- $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B),$ (2.1.8)
- $\mathbb{P}(A) \leq \mathbb{P}(B)$ si $A \subset B.$ (2.1.9)

Preuve. La propriété (2.1.5) se montre en appliquant (2.1.4) avec $A = B = \emptyset$, et (2.1.6) s'obtient de la même façon avec A et A^c . Pour prouver (2.1.8), nous décomposons l'ensemble A en l'union des deux ensembles disjoints $A \cap B$ et son complémentaire $A \setminus B$, et de même pour B , comme cela est représenté dans la figure 2.1. L'inégalité (2.1.9) se déduit de (2.1.4) avec $B = A \cup (B \setminus A)$. \square

Définition 2.1.5 Un **modèle probabiliste** est un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ constitué de l'espace Ω , de l'ensemble des événements \mathcal{A} , et de la famille des $\mathbb{P}(A)$ pour $A \in \mathcal{A}$. Nous pouvons ainsi considérer \mathbb{P} comme une application de \mathcal{A} dans $[0, 1]$, qui vérifie au moins les propriétés (2.1.3) et (2.1.4) données ci-dessus. (Nous verrons ultérieurement qu'elle doit satisfaire une propriété supplémentaire nécessaire dès que Ω n'est pas un espace fini).

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

Nous allons maintenant donner des définitions mathématiques rigoureuses de ces objets.

2.2 Probabilité sur un espace fini - Calcul combinatoire

2.2.1 Définition

Dans ce paragraphe, nous supposons que l'espace de probabilité Ω est un ensemble fini. Dans ce cas, nous choisirons toujours $\mathcal{A} = \mathcal{P}(\Omega)$.

Définition 2.2.1 Une **probabilité sur Ω fini** est une application $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$ qui vérifie (2.1.3) et (2.1.4). Ainsi, elle est caractérisée par :

- $0 \leq \mathbb{P}(A) \leq 1$,
- $\mathbb{P}(\Omega) = 1$,
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ si $A \cap B = \emptyset$.

La probabilité \mathbb{P} satisfait également les propriétés (2.1.5)–(2.1.9).

Comme l'ensemble des singletons $\{\omega\}$, pour $\omega \in \Omega$, est une partition finie de Ω , nous aurons la proposition fondamentale suivante.

Proposition 2.2.2 *Supposons que $\Omega = \{\omega_1, \dots, \omega_n\}$.*

i) Une probabilité \mathbb{P} sur Ω est entièrement caractérisée par ses valeurs sur les singletons, c'est-à-dire par la famille $\{p_{\omega_i} = \mathbb{P}(\{\omega_i\}), \omega_i \in \Omega\}$.

ii) Etant donnée une famille $(p_i)_{1 \leq i \leq n}$ de nombres réels, il lui correspond une unique probabilité \mathbb{P} telle pour tout $\omega_i \in \Omega$, $p_i = \mathbb{P}(\{\omega_i\})$, si et seulement si

$$0 \leq p_i \leq 1, \quad \sum_{i=1}^n p_i = 1. \quad (2.2.10)$$

Nous avons alors

$$\forall A \in \mathcal{P}(\Omega), \quad \mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{i: \omega_i \in A} p_i. \quad (2.2.11)$$

Remarquons que les sommes dans (2.2.11) sont des sommes finies, puisque Ω et donc A , sont de cardinal fini.

Preuve. Soit \mathbb{P} une probabilité sur Ω , et soit $p_\omega = \mathbb{P}(\{\omega\})$. Il est alors évident que $0 \leq p_\omega \leq 1$, et (2.2.11) découle de (2.1.7) puisque toute partie A de Ω est réunion disjointe (et finie) des singletons $\{\omega\}$, pour les $\omega \in A$. Nous en déduisons donc (i) et la condition nécessaire de (ii). En effet, la seconde partie de (2.2.10) découle de (2.2.11) appliqué à $A = \Omega$ et de $\mathbb{P}(\Omega) = 1$.

Inversement, considérons n nombres $(p_i)_{1 \leq i \leq n}$ vérifiant (2.2.10). Nous posons $\mathbb{P}(\{\omega_i\}) = p_i$ et pour tout $A \subset \Omega$, nous définissons $\mathbb{P}(A)$ par (2.2.11). La vérification de (2.1.2), (2.1.3) et (2.1.4) est immédiate. \square

Exemple 2.2.3 Loi de Bernoulli de paramètre $p \in [0, 1]$.

L'espace Ω a deux éléments :

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

Cette probabilité modélise en particulier la chance pour une pièce de tomber sur Pile (ou Face) dans un jeu de Pile ou Face. Dans ce cas, $\Omega = \{P, F\}$ peut être assimilé plus simplement à $\{0, 1\}$. Si la pièce est équilibrée, p sera égal à $1/2$. 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?).

2.2.2 Probabilité uniforme

Un exemple important de probabilité sur un espace fondamental Ω fini est celui de la probabilité uniforme, pour laquelle chaque singleton de Ω a la même chance de réalisation. Ainsi, la définition suivante découle de (2.2.10).

Définition 2.2.4 Nous dirons que la probabilité \mathbb{P} sur l'espace fini Ω est uniforme si $p_\omega = \mathbb{P}(\{\omega\})$ ne dépend pas de ω . Nous avons donc pour tout ω :

$$p_\omega = \frac{1}{\text{card}(\Omega)},$$

où $\text{card}(\Omega)$ désigne le cardinal de Ω , c'est à dire son nombre d'éléments.

Si \mathbb{P} est une probabilité uniforme, nous déduisons de (2.2.11) que

$$\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}, \quad (2.2.12)$$

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

Remarquons que sur un espace fini donné Ω , 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).

2.2.3 Modèles d'urnes

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 . Tirons au hasard n boules de cette urne, $n \leq N$, et intéressons-nous à la répartition des couleurs dans l'échantillon obtenu. Nous notons par $P_{n_1 n_2 \dots n_k}$ la probabilité d'obtenir n_1 boules de couleur 1, n_2 boules de couleur 2, ..., n_k boules de couleur k , avec bien sûr $n_1 + n_2 + \dots + n_k = n$.

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 un résultat différent.

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

Remarque 2.2.5 : 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 :

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

Tirage exhaustif ou simultané - La loi hypergéométrique

Nous tirons toutes les boules en même temps. L'ensemble Ω est alors l'ensemble de toutes les parties possibles de n éléments distincts, et le nombre de résultats possibles est $\binom{N}{n}$. Le nombre de cas favorables donnant la bonne répartition des couleurs est alors égal à $\binom{N_1}{n_1} \dots \binom{N_k}{n_k}$. La probabilité recherchée vaut donc

$$\hat{P}_{n_1 n_2 \dots n_k} = \frac{\binom{N_1}{n_1} \dots \binom{N_k}{n_k}}{\binom{N}{n}}. \quad (2.2.13)$$

Cette distribution s'appelle la distribution polygéométrique. Dans le cas de deux couleurs, elle vaudra

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

qui est appelée distribution (ou loi) hypergéométrique.

Ainsi, si 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 replaçons la boule tirée dans l'urne avant le tirage suivant. Nous pouvons donc tirer plusieurs fois la même boule. L'ensemble Ω 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 Ω de sa probabilité uniforme. 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 $\frac{n!}{n_1!n_2!\dots n_k!}$. Une fois la place des couleurs choisies, 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 $\frac{n!}{n_1!n_2!\dots n_k!} N_1^{n_1} \dots N_k^{n_k}$. Nous avons donc finalement

$$P_{n_1 n_2 \dots n_k} = \frac{n!}{n_1! n_2! \dots n_k!} \frac{N_1^{n_1} \dots N_k^{n_k}}{N^n}. \quad (2.2.14)$$

Cette probabilité est appelée une distribution multinomiale. Dans le cas particulier où $k = 2$, $p_1 = p = N_1/N$ et $p_2 = 1 - p$, la probabilité définie par

$$P_{n_1, n-n_1} = \binom{n}{n_1} p^{n_1} (1-p)^{n-n_1} \quad (2.2.15)$$

sera 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 : le lecteur pourra vérifier que

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

Tirage sans remise

Nous tirons maintenant successivement les boules de l'urne, mais sans les replacer dans l'urne après tirage. L'ensemble Ω est alors l'ensemble des suites de n éléments distincts parmi N et le nombre de cas possibles sera $N(N-1) \dots (N-n+1) = A_N^n$. En raisonnant comme dans le cas avec remise, nous pouvons montrer que le nombre de cas favorables vaut $\frac{n!}{n_1!n_2!\dots n_k!} A_{N_1}^{n_1} \dots A_{N_k}^{n_k}$, ce qui finalement donne pour probabilité la même que celle du cas de tirage simultané.

Ainsi, il y a équivalence du tirage sans remise et du tirage simultané, du point de vue de la composition de l'échantillon, et l'on peut se permettre de ne pas prendre en compte l'ordre des individus dans le tirage.

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 très 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.

Remarque 2.2.6 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 6.3 (Convergence en loi).

Exemple 2.2.7 : 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 : $\frac{2}{7!} = \frac{1}{2520}$.

Exemple 2.2.8 : On tire au hasard quatre cartes d'un jeu de cinquante deux cartes. Quelle est la probabilité pour que, parmi ces quatre cartes, il y ait exactement deux rois ?

Solution : L'hypothèse *au hasard* amène à modéliser cette expérience comme un tirage uniforme dans un certain ensemble Ω qu'il faut préciser. Ici, on prend pour Ω la classe des parties à 4 éléments de l'ensemble de 52 cartes. Le cardinal de Ω est $\binom{52}{4}$ et \mathbb{P} est la probabilité uniforme sur Ω . 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 $\frac{\binom{4}{2} \binom{48}{2}}{\binom{52}{4}}$.

Exemple 2.2.9 : 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. (Cela permettra de construire un jeu parfaitement équitable...)

Solution : L'ensemble Ω 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.

Une difficulté majeure dans ce style de calcul combinatoire est de bien préciser le modèle probabiliste. De célèbres paradoxes sont nés de cette difficulté.

Exemple 2.2.10 Rappelons le 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.*

Cette règle est bonne puisque la probabilité de l'événement qui nous intéresse 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...

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

Cette 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 Définition générale des probabilités

2.3.1 Pourquoi la définition précédente ne suffit-elle pas ?

Lorsque l'espace fondamental Ω n'est pas fini, la définition 2.2.1 n'est pas suffisante.

Nous allons nous en rendre compte sur l'ensemble suivant. Jouons à Pile ou Face.

Si nous jouons n fois, l'espace Ω 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 nous supposons que la pièce n'est pas truquée, la probabilité de chaque tirage est uniforme et nous nous retrouvons dans le cadre de la combinatoire. Ainsi, pour tout $A \subset \Omega$,

$$\mathbb{P}_n(A) = \frac{\text{card}(A)}{2^n}. \quad (2.3.16)$$

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 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 jamais Pile lors des n premiers tirages”. D'après (2.3.16), 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

$$\mathbb{P}(A) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 0. \quad (2.3.17)$$

Pour que ceci soit vrai, les propriétés (2.1.2), (2.1.3), (2.1.4) sont insuffisantes. Il faut ajouter un axiome supplémentaire permettant le passage à la limite dans (2.3.17).

A cet effet, nous devons d'abord caractériser les propriétés que doit satisfaire la classe \mathcal{A} des événements. En effet, si sur un ensemble fini, il est naturel de prendre $\mathcal{A} = \mathcal{P}(\Omega)$, il n'en est plus de même lorsque Ω est infini, ceci pour des raisons mathématiques et des raisons de modélisation qui seront explicitées plus loin. La classe \mathcal{A} doit toutefois satisfaire un certain nombre d'axiomes, et pour les définir, rappelons ce qu'est un ensemble dénombrable.

2.3.2 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 $\{0, 1\}^{\mathbb{N}^*}$, de \mathbb{R} , ni des intervalles $[a, b]$ lorsque $a < b$.

Énonçons quelques propriétés des ensembles dénombrables.

- Tout ensemble dénombrable est infini. (Mais 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.

2.3.3 Tribu

Définition 2.3.1 La classe \mathcal{A} est une **tribu** si elle vérifie les propriétés suivantes :

- (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 et intersection dénombrables, i.e. si $(A_n)_{n \in \mathbb{N}}$ est une suite d'éléments de \mathcal{A} , alors $\bigcup_{n \in \mathbb{N}} A_n$ et $\bigcap_{n \in \mathbb{N}} A_n$ sont dans \mathcal{A} .

Notons qu'une tribu \mathcal{A} contient nécessairement $\emptyset = \Omega^c$ et qu'elle est stable par réunion et intersection finie : si $A, B \in \mathcal{A}$, alors $A \cup B \in \mathcal{A}$ et $A \cap B \in \mathcal{A}$. Il suffit de prendre $A_0 = A$ et $A_n = B$ pour $n \geq 1$.

Notons également que (A3) n'entraîne pas que \mathcal{A} soit stable par réunion ou intersection infinie non dénombrable.

Remarque fondamentale : Dans la modélisation de notre phénomène aléatoire, la tribu représente un ensemble de parties de Ω (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 2.3.2 $\mathcal{A} = \{\emptyset, \Omega\}$ est la tribu grossière, ou triviale : c'est la plus petite tribu de Ω .

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

Définition 2.3.3 Si $C \subset \mathcal{P}(\Omega)$, on appelle tribu engendrée par C la plus petite tribu contenant C . Elle existe toujours, car d'une part $\mathcal{P}(\Omega)$ est une tribu contenant C , et d'autre part l'intersection d'une famille quelconque de tribus est une tribu. Ainsi, la tribu engendrée par C est l'intersection de toutes les tribus contenant C .

Exemple 2.3.4 • La tribu engendrée par un ensemble $A \subset \Omega$ est $\{\emptyset, A, A^c, \Omega\}$.
 • Si $(A_i)_{i \in I}$ est une partition finie ou dénombrable de Ω (i.e. les A_i sont deux-à-deux disjoints et leur réunion est Ω), 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 2.3.5 Si $\Omega = \mathbb{R}$, on appelle **tribu borélienne** la tribu engendrée par la classe des intervalles ouverts de \mathbb{R} .

À titre d'exercice de maniement des tribus, donnons en détail la démonstration du résultat suivant :

Proposition 2.3.6 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 $]a, +\infty[$, il 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 :

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

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

2.3.4 Définition d'une probabilité

Nous sommes maintenant en mesure de donner la définition générale d'une probabilité.

Définition 2.3.7 Une probabilité sur l'espace (Ω, \mathcal{A}) est une application de \mathcal{A} dans $[0, 1]$, notée \mathbb{P} , telle que :

- On a

$$\mathbb{P}(\Omega) = 1. \quad (2.3.18)$$

- Pour toute suite (dénombrable) $(A_n)_n$ d'éléments de \mathcal{A} **deux-à-deux disjoints**, on a

$$\mathbb{P}(\cup_n A_n) = \sum_n \mathbb{P}(A_n). \quad (2.3.19)$$

Remarque : La probabilité \mathbb{P} (dite aussi mesure de probabilité), est une mesure abstraite de masse 1 sur l'espace mesurable (Ω, \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.

L'axiome (2.3.19), dit "axiome de σ -additivité", entraîne en particulier que la série de terme général $\mathbb{P}(A_n)$ est convergente (et de somme $\mathbb{P}(\cup_n A_n)$). Il est plus fort que (2.1.4). Pour le voir, nous commençons par appliquer (2.3.19) 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 (2.3.19) 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 (2.1.4).

Notons que toute probabilité vérifie les propriétés (2.1.5–2.1.9).

Le résultat suivant est très utile dans la pratique, et répond au problème de modélisation que nous nous étions posé en Section 2.3.1. Pour ce résultat, nous utilisons les notations suivantes. Si $(A_n)_n$ est une suite décroissante de parties de Ω , i.e. $A_{n+1} \subset A_n$ pour tout n et si $A = \cap_n A_n$, nous écrivons $A_n \downarrow A$. De même, si la suite $(A_n)_n$ est croissante, i.e. $A_n \subset A_{n+1}$ pour tout n et si $A = \cup_n A_n$, nous écrivons $A_n \uparrow A$.

Proposition 2.3.8 Supposons que $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ vérifie (2.3.18) et (2.1.4). Il y a équivalence entre :

(i) La propriété de σ -additivité (2.3.19).

(ii) Pour toute suite $(A_n)_n$ **croissante**,

$$\mathbb{P}(\cup_n A_n) = \lim_n \uparrow \mathbb{P}(A_n). \quad (2.3.20)$$

(iii) Pour toute suite $(A_n)_n$ **décroissante**,

$$\mathbb{P}(\cap_n A_n) = \lim_n \downarrow \mathbb{P}(A_n). \quad (2.3.21)$$

Ce résultat entraîne 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 (2.1.6), 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 $B_n = \cup_{p \leq n} A_p$ et $B = \cup_n A_n$. Comme \mathbb{P} vérifie (2.1.4), elle vérifie (2.1.7) 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 $\cup_n B_n = A$ 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 (2.1.7). Nous obtenons donc le résultat. \square

La propriété (2.3.19) donne la probabilité de la réunion $\cup_n A_n$ 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 dans la pratique :

Proposition 2.3.9 Soit \mathbb{P} une probabilité, et soit $(A_n)_{n \in I}$ une famille finie ou dénombrable d'événements. On a alors

$$\mathbb{P}(\cup_{n \in I} A_n) \leq \sum_{n \in I} \mathbb{P}(A_n). \quad (2.3.22)$$

Preuve. a) Supposons d'abord l'ensemble I fini. Il s'agit de montrer que pour tout k entier,

$$\mathbb{P}(A_1 \cup \dots \cup A_k) \leq \mathbb{P}(A_1) + \dots + \mathbb{P}(A_k). \quad (2.3.23)$$

Nous montrons cette propriété par récurrence sur k : elle 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 (2.1.8), 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 que (2.3.23) est satisfaite pour k .

b) Considérons maintenant le cas où I est dénombrable. Nous pouvons supposer sans restriction que $I = \mathbb{N}^*$. Posons $B_n = \cup_{i=1}^n A_i$, qui croît vers l'ensemble $C = \cup_{n \in I} A_n$. 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 $\sum_{n \in I} \mathbb{P}(A_n)$. En passant à la limite, nous obtenons donc (2.3.22). \square

Nous avons pu ainsi construire dans toute sa généralité un espace abstrait Ω , 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 généralement représentées par $\omega \in \Omega$.

Définition 2.3.10 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 (Ω, \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. Nous pouvons définir sur $\{0, 1\}$ différentes lois de Bernoulli, en fonction du paramètre $p \in [0, 1]$ choisi.

La définition suivante est fondamentale en théorie des probabilités. Elle introduit une notion de "vrai ou faux" qui dépend de la probabilité choisie sur l'espace fondamental.

Définition 2.3.11 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.

Un événement de probabilité nulle est dit négligeable.

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. Dans ce cas, l'ensemble des ω pour lesquels la propriété est fautive est négligeable.

2.3.5 Probabilités sur un espace dénombrable

Supposons que Ω soit dénombrable. Nous pouvons alors numéroter ses éléments par $\{\omega_0, \omega_1, \dots, \omega_n, \dots\}$, $n \in \mathbb{N}$. La proposition suivante généralise au cas dénombrable la proposition 2.2.2 vue dans le cas fini. Ici, la tribu considérée est l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω .

Proposition 2.3.12 Une probabilité sur un ensemble dénombrable est entièrement caractérisée par ses valeurs sur les singletons.

Etant donnée une suite $(p_n)_n$ de nombres réels tels que

$$0 \leq p_n \leq 1 \quad , \quad \sum_n p_n = 1,$$

il lui correspond une unique probabilité \mathbb{P} telle que pour tout $A \subset E$,

$$\mathbb{P}(A) = \sum_{\omega_n \in A} \mathbb{P}(\{\omega_n\}) = \sum_{n: \omega_n \in A} p_n. \quad (2.3.24)$$

Preuve. Lorsque Ω est fini, ce résultat n'est autre que la proposition 2.2.2. Lorsque Ω est dénombrable, la démonstration est analogue, si ce n'est que pour prouver que \mathbb{P} définie par (2.3.24) vérifie (2.3.19), il faut utiliser la propriété de sommation par paquets pour les séries. (Voir Section 3.1 ci-après.) \square

Exemple 2.3.13 Soit $\theta > 0$ et $p_n = e^{-\theta} \frac{\theta^n}{n!}$. Il est facile de vérifier que $0 \leq p_n \leq 1$ et que $\sum_n p_n = e^{-\theta} \sum_n \frac{\theta^n}{n!} = 1$. La suite $(p_n)_n$ définit une probabilité sur \mathbb{N} , appelée loi de Poisson de paramètre θ . 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 2.3.14 1) Considérons un espace mesurable (Ω, \mathcal{A}) et un point ω_0 fixé dans Ω . Nous pouvons alors définir la probabilité δ_{ω_0} , appelée mesure de Dirac en ω_0 , de la manière suivante : pour $A \in \mathcal{A}$,

$$\delta_{\omega_0}(A) = 1 \text{ si } \omega_0 \in A ; \delta_{\omega_0}(A) = 0 \text{ si } \omega_0 \notin A.$$

Une propriété est alors vraie δ_{ω_0} -presque-sûrement si elle est satisfaite par ω_0 , et l’ensemble $\Omega \setminus \{\omega_0\}$ est un ensemble δ_{ω_0} -négligeable.

2) Soit \mathbb{P} une probabilité définie sur $\Omega = \{\omega_0, \omega_1, \dots, \omega_n, \dots\}$, et $p_n = \mathbb{P}(\{\omega_n\})$. Alors pour tout $A \in \mathcal{A}$, on a

$$\mathbb{P}(A) = \sum_n p_n \delta_{\omega_n}(A).$$

La probabilité \mathbb{P} peut donc s’écrire

$$\mathbb{P} = \sum_n p_n \delta_{\omega_n}.$$

2.4 Loi d’une variable aléatoire

En théorie moderne des probabilités, on préfère prendre un point de vue fonctionnel plutôt qu’ensembliste, et utiliser les variables aléatoires plutôt que les événements. Ce point de vue sera développé dans la suite du cours. Donnons-en ici uniquement les idées de base.

Une variable aléatoire est une grandeur qui dépend du résultat de l’expérience. Par exemple,

- le nombre de 6 obtenus dans un lancer de 3 dés,
- le nombre d’appels dans un central téléphonique pendant une heure,
- la distance du point d’atteinte d’une flèche au centre de la cible,
- la valeur maximale d’un prix d’actif sur un intervalle de temps donné,

sont des variables aléatoires.

En termes mathématiques, on considère un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Une variable aléatoire X est une application de (Ω, \mathcal{A}) dans un ensemble F ,

$$\omega \in \Omega \mapsto X(\omega) \in F.$$

En pratique, l’ensemble F pourra être un ensemble fini ou dénombrable ou \mathbb{R} ou \mathbb{R}^d ou un espace plus sophistiqué tel que l’ensemble $C(\mathbb{R}_+, \mathbb{R}^d)$ des fonctions continues de \mathbb{R}_+ dans \mathbb{R}^d .

Attention : La terminologie, consacrée par l'usage, est très malencontreuse et engendre une difficulté liée au vocabulaire employé. Une variable aléatoire, malgré son nom, n'est pas une variable (au sens de l'analyse), mais une fonction de la variable $\omega \in \Omega$.

Intérêt fondamental : Comme l'espace F est connu dans la pratique, nous allons préférer nous intéresser aux chances de réalisation des valeurs de X plutôt qu'aux chances de réalisation des résultats de l'expérience. Ainsi, grâce à une variable aléatoire X , nous pouvons transporter la structure abstraite du modèle probabiliste $(\Omega, \mathcal{A}, \mathbb{P})$ sur l'espace d'arrivée F (dont la structure est mieux connue), en posant pour $B \subset F$

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega, X(\omega) \in B\}). \quad (2.4.25)$$

NOTATION : Il est usuel de noter l'ensemble $X^{-1}(B) = \{\omega, X(\omega) \in B\}$ par $\{X \in B\}$, ce qui allège les écritures. Rappelons-nous toutefois que cette notation simplifiée désigne un sous-ensemble de Ω .

Comme $\mathbb{P}(A)$ n'est définie que pour les A appartenant à \mathcal{A} , la formule (2.4.25) ne permet de définir $\mathbb{P}_X(B)$ que pour les ensembles B tels que $X^{-1}(B) = \{X \in B\} \in \mathcal{A}$, d'où l'importance de la proposition suivante.

Proposition 2.4.1 a) La famille \mathcal{F} des parties B de F telles que $X^{-1}(B) \in \mathcal{A}$ est une tribu de F .

b) L'application \mathbb{P}_X définie pour $B \in \mathcal{F}$ par

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$$

définit une probabilité sur la tribu \mathcal{F} de F .

Preuve. Les propriétés (A1), (A2) et (A3) pour \mathcal{F} , ainsi que (2.1.3) et (2.3.19) pour \mathbb{P}_X découlent immédiatement des propriétés du même nom pour \mathcal{A} et \mathbb{P} , une fois remarquées les propriétés élémentaires suivantes :

$$\begin{aligned} X^{-1}(\emptyset) &= \emptyset, \quad X^{-1}(F) = \Omega, \quad X^{-1}(B^c) = (X^{-1}(B))^c, \\ X^{-1}(\cap_i A_i) &= \cap_i X^{-1}(A_i), \quad X^{-1}(\cup_i A_i) = \cup_i X^{-1}(A_i). \end{aligned} \quad (2.4.26)$$

□

Définition 2.4.2 La probabilité \mathbb{P}_X , définie sur (F, \mathcal{F}) par (2.4.25) est appelée **loi de la variable** X , ou distribution de X . C'est la mesure image de \mathbb{P} par l'application mesurable X .

Elle sera plus facile à caractériser que \mathbb{P} , puisque F est un ensemble connu (on pourra en particulier utiliser ses propriétés topologiques) alors que Ω est un espace abstrait. Nous remarquerons qu'en général $\mathcal{F} \neq \mathcal{P}(F)$, même si on a $\mathcal{A} = \mathcal{P}(\Omega)$. Comme \mathbb{P}_X ne peut être définie que sur \mathcal{F} , ceci constitue une première raison, d'ordre mathématique, pour qu'une probabilité soit définie sur une tribu qui peut être strictement plus petite que l'ensemble de toutes les parties.

Les variables que nous rencontrerons dans ce cours seront soit à valeurs dans un ensemble dénombrable, soit à valeurs dans \mathbb{R} ou dans \mathbb{R}^d . Nous les appellerons respectivement des variables aléatoires discrètes, réelles ou des vecteurs aléatoires. Leurs lois seront alors des probabilités respectivement sur un ensemble dénombrable, sur \mathbb{R} ou sur \mathbb{R}^d . Nous avons vu ci-dessus des caractérisations simples des probabilités sur un espace fini ou dénombrable. En revanche, décrire les probabilités sur \mathbb{R} ou sur \mathbb{R}^d est beaucoup plus délicat. Nous développerons cela dans le cadre des variables et vecteurs aléatoires au Chapitre 4.

Exemple 2.4.3 Etudions un lancer de deux dés que l'on suppose bien équilibrés. Dans ce cas, l'espace fondamental vaut $\Omega = \{(i, j) : 1 \leq i \leq 6 ; 1 \leq j \leq 6\}$, et il est naturel de prendre ici $\mathcal{A} = \mathcal{P}(\Omega)$. Considérons un événement $A \subset \Omega$. Puisque les dés sont équilibrés, nous définissons la probabilité de A par $\mathbb{P}(A) = \frac{\text{card}(A)}{36}$ où $\text{card}(A)$ est le cardinal de A et désigne le nombre d'éléments de A . C'est la probabilité uniforme sur Ω , et nous avons $\mathbb{P}(\{\omega\}) = \frac{1}{36}$ pour chaque singleton $\{\omega\}$.

L'application $X : \Omega \rightarrow \{2, \dots, 12\}$ définie par $X(i, j) = i + j$ est la variable aléatoire "somme des résultats des deux dés". Elle a pour loi

$$\mathbb{P}_X(B) = \frac{\text{nombre de couples } (i, j) \text{ tels que } i + j \in B}{36}.$$

Par exemple, $\mathbb{P}_X(\{2\}) = \mathbb{P}_X(\{12\}) = \frac{1}{36}$, $\mathbb{P}_X(\{3\}) = \frac{2}{36}$, etc.

EXERCICE 2.4.4 Dessiner le diagramme en bâton de la loi \mathbb{P}_X de X et remarquer qu'elle est très différente de la probabilité uniforme \mathbb{P} définie préalablement sur Ω .

2.5 Conditionnement et indépendance

2.5.1 Probabilités conditionnelles

La notion de conditionnement est l'une des plus fructueuses de la théorie des probabilités, et la différence fondamentale 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 2.5.1 Soit A et B deux événements, avec $\mathbb{P}(B) > 0$. **La probabilité conditionnelle de A sachant B** est le nombre

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (2.5.27)$$

Cela définit bien une probabilité comme l'énonce la proposition suivante.

Proposition 2.5.2 1) Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et B un événement de \mathcal{A} de probabilité strictement positive. Alors l'application de \mathcal{A} dans $[0, 1]$ qui à A associe $\mathbb{P}(A|B)$ définit une nouvelle probabilité sur Ω , appelée probabilité conditionnelle sachant B .

2) 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 (2.3.18) et (2.3.19) 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 (2) est évidente. \square

Proposition 2.5.3 Formule des probabilités composées. Si A_1, \dots, A_n sont des événements de Ω tels que $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$, alors

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}). \quad (2.5.28)$$

Preuve. La démonstration se fait par récurrence. Si $n = 2$, les formules (2.5.27) et (2.5.28) sont les mêmes. Supposons que (2.5.28) soit vraie pour $n - 1$, et soit $B = A_1 \cap A_2 \cap \dots \cap A_{n-1}$. D'après (2.5.27), on a $\mathbb{P}(B \cap A_n) = \mathbb{P}(B)\mathbb{P}(A_n|B)$. En remplaçant $\mathbb{P}(B)$ par sa valeur donnée par (2.5.28) avec $n - 1$, nous obtenons (2.5.28) pour n . \square

Proposition 2.5.4 Formule des probabilités totales. Soit $(B_i)_{i \in I}$ une partition finie ou dénombrable d'événements de Ω , telle que $\mathbb{P}(B_i) > 0$ pour chaque $i \in I$. Pour tout $A \in \mathcal{A}$, on a alors

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap B_i) = \sum_{i \in I} \mathbb{P}(A|B_i) \mathbb{P}(B_i). \quad (2.5.29)$$

Preuve. Nous avons $A = \cup_{i \in I} (A \cap B_i)$. Par hypothèse, les ensembles $(A \cap B_i)$ sont deux-à-deux disjoints, et par ailleurs $\mathbb{P}(A \cap B_i) = \mathbb{P}(A|B_i) \mathbb{P}(B_i)$. Il suffit alors d'appliquer (2.3.19). \square

Théorème 2.5.5 Formule de Bayes. Sous les mêmes hypothèses que dans la proposition 2.5.4, et si $\mathbb{P}(A) > 0$,

$$\forall i \in I, \mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \mathbb{P}(B_i)}{\sum_{j \in I} \mathbb{P}(A|B_j) \mathbb{P}(B_j)}. \quad (2.5.30)$$

Preuve. Le dénominateur de (2.5.30) vaut $\mathbb{P}(A)$ d'après (2.5.29), tandis que (2.5.27) implique

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

□

Exemple 2.5.6 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 les probabilités d'avoir un résultat positif lors de l'application du test si l'individu est séropositif, ou s'il ne l'est pas, sont respectivement égales à 0,99 (c'est la sensibilité du test) et à 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.

Exemple 2.5.7 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

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

2.5.2 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 2.5.8 Deux événements A et B sont **indépendants** si et seulement si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \quad (2.5.31)$$

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

Remarque 2.5.9 1) Cette notion est une notion 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 2.5.10 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 2.5.10 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. $A = \{\text{la carte est une dame}\}$; $B = \{\text{la carte est un cœur}\}$. 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 2.5.11 *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 2.5.12 Une suite $(A_n)_n$ d'événements de (Ω, \mathcal{A}, P) est dite indépendante si

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

pour toute suite finie (i_1, \dots, i_k) d'entiers deux-à-deux distincts.

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 ensembles et l'intersection des 3 ensembles. 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

Considérons une suite d'espaces de probabilité $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$. Nous avons vu que ces espaces modélisent 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 2.3.3). 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 Ω 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 (Ω, \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 égaux, cela nous permettra de modéliser la même expérience répétée une infinité (dénombrable) de fois, de manière indépendante et dans les mêmes conditions.

Exemple 2.5.13 Considérons les lancers successifs et indépendants d'une même pièce de monnaie, telle que la probabilité de tirer Face 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". Soit T la variable aléatoire décrivant le premier lancer pour lequel un Pile est obtenu. Alors, par indépendance des lancers, nous avons pour $k \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}(T = k) &= \mathbb{P}(F_1 \cap F_2 \cap \dots \cap F_{k-1} \cap P_k) \\ &= \mathbb{P}(\text{Face})^{k-1} \mathbb{P}(\text{Pile}) = p^{k-1}(1-p).\end{aligned}$$

Remarquons que

$$\mathbb{P}(T < +\infty) = \sum_{k \geq 1} \mathbb{P}(T = k) = 1,$$

donc $\mathbb{P}(T = +\infty) = 0$. La loi de T est appelée loi géométrique de paramètre p .

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 5).

2.5.3 Le lemme de Borel-Cantelli

Considérons une suite $(A_n)_n$ d'événements. Nous définissons l'ensemble $\limsup_n A_n$ comme étant l'ensemble

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

Remarquons que

$$\begin{aligned}\omega \in \limsup_n A_n &\Leftrightarrow \forall p, \exists n \geq p, \text{ tel que } \omega \in A_n \\ &\Leftrightarrow \omega \text{ appartient à une infinité de } A_n \\ &\Leftrightarrow \sum_n \mathbf{1}_{A_n}(\omega) = +\infty.\end{aligned}$$

De même,

$$\omega \notin \limsup_n A_n \Leftrightarrow \omega \text{ appartient à au plus un nombre fini de } A_n.$$

Théorème 2.5.14 Soit $(A_n)_n$ une suite d'événements de \mathcal{A} .

- 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.

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

$$\sum_n \mathbb{P}(A_n) = +\infty \implies \mathbb{P}(\limsup_n A_n) = 1. \quad (2.5.32)$$

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 5.2. 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

$$\mathbb{P}(\limsup_n A_n) = \lim_p \downarrow \mathbb{P}(\cup_{n \geq p} A_n) \leq \lim_p \downarrow \sum_{n \geq p} \mathbb{P}(A_n), \quad (2.5.33)$$

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 (2.5.33) 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 percutante : 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 termes dont chaque lettre est P ou 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".

2.6 Exercices sur le chapitre 2

EXERCICE 2.6.1 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 2.6.2 Montrer la formule de Poincaré :

$$\mathbb{P}(\cup_{m=1}^n A_m) = p_1 - p_2 + \dots + (-1)^{n-1} p_n = \sum_{k=1}^n (-1)^{k-1} p_k \quad (2.6.34)$$

où

$$p_k = \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}). \quad (2.6.35)$$

EXERCICE 2.6.3 Un facteur possède n lettres adressées à n destinataires distincts. Il est totalement ivre et poste au hasard une lettre par boîte.

Quelle est la probabilité d'obtenir la bonne répartition ?

Quelle est la probabilité qu'une lettre au moins arrive à la bonne adresse ?

Quelle est la probabilité qu'aucune lettre n'arrive à la bonne destination ?

Quel est le nombre d_n de manières différentes de poster les lettres de telle sorte qu'aucune n'arrive à destination ?

EXERCICE 2.6.4 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 2.6.5 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,

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

EXERCICE 2.6.6 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 2.6.7 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 2.6.8 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 2.6.9 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) Quelle est la probabilité pour qu'une famille ayant deux filles aient deux enfants seulement ?

2) Quelle est la probabilité qu'une famille ait deux garçons sachant qu'elle a deux filles ?

EXERCICE 2.6.10 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 2.6.11 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 développer une étude plus systématique des variables aléatoires et de leur loi. 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".

3.1 Prérequis : 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 $S_n = u_1 + \dots + u_n$ la “somme partielle” à l’ordre n .

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.

3.2 Variables aléatoires discrètes

Dans tout ce chapitre, l'espace Ω est fini ou dénombrable.

Nous supposons donnée une probabilité \mathbb{P} sur Ω (muni de la tribu de ses parties $\mathcal{P}(\Omega)$), caractérisée par les $p_\omega = \mathbb{P}(\{\omega\})$. Alors toute application définie sur Ω est une variable aléatoire et l'ensemble F des valeurs de X (i.e. l'ensemble des $X(\omega)$ lorsque ω parcourt Ω) est nécessairement fini ou dénombrable. La loi de X est une probabilité sur F . Par la proposition 2.3.12, nous savons que cette probabilité \mathbb{P}_X est caractérisée par les nombres

$$p_i^X = \mathbb{P}_X(\{x_i\}) = \mathbb{P}(\{\omega, X(\omega) = x_i\}) = \mathbb{P}(X = x_i) = \sum_{\omega: X(\omega)=x_i} p_\omega, \quad \forall x_i \in F. \quad (3.2.1)$$

Ainsi,

Proposition 3.2.1 *La loi d'une variable aléatoire X à valeurs dans un espace dénombrable F est caractérisée par*

$$\{ (x_i, p_i^X), x_i \in F, \text{ avec } p_i^X = \mathbb{P}(X = x_i) \}.$$

Exemple 3.2.2 Une variable aléatoire 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.

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.

3.3 Espérance des variables aléatoires discrètes

Nous supposons ici que $F \subset \mathbb{R}$.

3.3.1 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 ω de l'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. Nous voulons faire tendre n vers l'infini. Si la propriété (2.1.1) intuitée au Chapitre 1 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 Ω est fini), alors la suite $(M_n)_n$ tend vers $\sum_{\omega \in \Omega} p_\omega X(\omega)$: l'espérance mathématique, ou moyenne, est la limite des moyennes arithmétiques lorsque le nombre d'expériences tend vers l'infini. 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 3.3.1 Soit X une variable aléatoire réelle sur l'espace fini ou dénombrable Ω (i.e. une application de Ω dans \mathbb{R}). Son espérance (appelée aussi parfois moyenne) est le nombre

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} p_\omega X(\omega), \quad (3.3.2)$$

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

(Rappelons que, en vertu de (S6), 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 ω sont ordonnés.)

Remarque : $\mathbb{E}(X)$ est un nombre réel qui donne une valeur moyenne résumant la variable aléatoire X .

L'espérance mathématique d'une variable aléatoire 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 3.3.2 *Considérons une variable aléatoire X satisfaisant $\sum_{\omega \in \Omega} p_\omega |X(\omega)| = \mathbb{E}(|X|) < +\infty$. On a alors la formule fondamentale suivante :*

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} p_\omega X(\omega) = \sum_{x_i \in F} x_i \mathbb{P}(X = x_i) = \sum_{x_i \in F} x_i p_i^X. \quad (3.3.3)$$

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

Preuve. La preuve de cette proposition consiste juste à observer que la sommation par paquets est justifiée par (S8) puisque $\sum_{\omega \in \Omega} p_\omega |X(\omega)| = \mathbb{E}(|X|) < +\infty$, puis à écrire

$$\begin{aligned} \mathbb{E}(X) &= \sum_{\omega \in \Omega} p_\omega X(\omega) = \sum_{x_i \in F} \left(\sum_{\omega: X(\omega)=x_i} p_\omega x_i \right) = \sum_{x_i \in F} x_i \mathbb{P}(\{\omega : X(\omega) = x_i\}) \\ &= \sum_{x_i \in F} x_i p_i^X. \end{aligned}$$

□

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. En d'autres termes, il suffit de montrer que

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

ce qui est immédiat.

Exemple 3.3.3 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”. Alors

$$\mathbb{P}(N = k) = \mathbb{P}(N = k | A_k) \mathbb{P}(A_k) = \frac{1}{10},$$

et

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

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

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

3.3.2 Propriétés de l'espérance des variables aléatoires discrètes

Définition 3.3.4 Nous appelons variable aléatoire intégrable une variable aléatoire X qui admet une espérance, c'est à dire telle que $\sum_{\omega \in \Omega} p_{\omega} |X(\omega)| = \mathbb{E}(|X|) < +\infty$. L'ensemble de toutes les variables aléatoires intégrables est noté L^1 . L'ensemble L^1 dépend de Ω et de la probabilité \mathbb{P} .

Les propriétés suivantes sont immédiates :

Proposition 3.3.5 • L^1 est un espace vectoriel, et l'espérance est linéaire sur L^1 : $\forall X, Y \in L^1, \forall a, b \in \mathbb{R}$,

$$\mathbb{E}(aX + bY) = a \mathbb{E}(X) + b \mathbb{E}(Y). \quad (3.3.4)$$

- $X \in L^1 \iff |X| \in L^1$, et dans ce cas,

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|). \quad (3.3.5)$$

- L'espérance est positive :

$$\text{si } X \geq 0 \text{ (i.e. } \forall \omega, X(\omega) \geq 0) \text{ et } X \in L^1, \text{ alors } \mathbb{E}(X) \geq 0. \quad (3.3.6)$$

- Si $X, Y \in L^1$, telles que $X \leq Y$ (i.e. $\forall \omega, X(\omega) \leq Y(\omega)$), alors

$$\mathbb{E}(X) \leq \mathbb{E}(Y). \quad (3.3.7)$$

- L^1 contient toutes les variables aléatoires bornées.
(X est dite bornée s'il existe un réel b tel que $|X(\omega)| \leq b$ pour tout ω).
- L'espérance d'une variable constante est égale à cette constante :

$$\text{Si } X(\omega) = a \text{ pour tout } \omega, \text{ alors } \mathbb{E}(X) = a. \quad (3.3.8)$$

- Si Ω est fini, L^1 contient toutes les variables aléatoires réelles.

Exemple 3.3.6 Supposons que A soit un événement et définissons la variable aléatoire X de la manière suivante : $X(\omega) = 1$ si $\omega \in A$, et $X(\omega) = 0$ si $\omega \notin A$. Cette variable aléatoire est notée $X = \mathbf{1}_A$ et est appelée fonction indicatrice de A . Nous remarquons alors que

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

ce qui donne le lien entre la probabilité d'un événement et l'espérance d'une variable aléatoire.

3.3.3 Variance et écart-type

Nous allons maintenant introduire un second ensemble de variables aléatoires, l'ensemble L^2 des variables aléatoires X réelles telles que le carré X^2 soit intégrable. Nous dirons dans ce cas que X est de carré intégrable.

Proposition 3.3.7 L^2 est un sous-espace vectoriel de L^1 , et si $X \in L^2$ on a

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|) \leq \sqrt{\mathbb{E}(X^2)}. \quad (3.3.9)$$

Preuve. Soient X et Y deux variables aléatoires réelles et $a \in \mathbb{R}$. Si X et Y sont dans L^2 , et comme $(aX + Y)^2 \leq 2a^2X^2 + 2Y^2$, nous déduisons de (3.3.4) 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 (3.3.4).

La première inégalité de (3.3.9) a déjà été vue en (3.3.5). 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 (3.3.4) 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 (3.3.6). Donc $a^2 \leq \mathbb{E}(X^2)$, ce qui est le résultat cherché. \square

Définition 3.3.8 Si $X \in L^2$, sa **variance** est définie par

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

En vertu de (3.3.6), $\text{Var}(X)$ est positive, et sa racine carrée positive σ_X s'appelle **l'écart-type** de X .

En développant le carré $(X - \mathbb{E}(X))^2$ comme dans la preuve ci-dessus, nous obtenons la formule de Huygens :

$$\sigma_X^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (3.3.10)$$

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.

Exemple 3.3.9 (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 σ vient de ce que parfois (mais très rarement) ça rapporte gros.

La forme moderne du loto, mis 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.

3.3.4 Un résultat fondamental - Moments d'une variable aléatoire

Nous avons vu en Proposition 3.3.2 que si $\mathbb{E}(|X|) < +\infty$, alors $\mathbb{E}(X)$ ne dépend que de la loi \mathbb{P}_X de X .

Plus généralement, nous considérons maintenant une fonction f de F dans \mathbb{R} . Ainsi $Y = f(X)$ est une variable aléatoire réelle, qui prend un nombre fini ou dénombrable de valeurs. Nous pouvons aussi considérer f comme une variable aléatoire sur l'espace de probabilité (F, \mathbb{P}_X) . Nous avons alors le résultat fondamental suivant, qui montre la cohérence de la notion d'espérance. Ce théorème s'appelle souvent le théorème de transfert dans la littérature.

Théorème 3.3.10 *Avec les hypothèses précédentes, la variable aléatoire $Y = f(X)$ définie sur (Ω, \mathbb{P}) est intégrable si et seulement si la variable aléatoire f sur (F, \mathbb{P}_X) l'est également. Dans ce cas, les espérances de ces deux variables aléatoires sont égales, et nous avons en particulier que*

$$\mathbb{E}(f(X)) = \sum_{\omega \in \Omega} f(X(\omega)) p_\omega = \sum_{x_i \in F} f(x_i) \mathbb{P}(X = x_i). \quad (3.3.11)$$

Preuve. Comme nous pouvons sommer par paquets par (S7) dans une série à termes positifs, nous voyons comme pour (3.3.3) que les deux expressions de droite de (3.3.11) 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 $f(X) \in L^1(\Omega, \mathbb{P}) \Leftrightarrow f \in L^1(F, \mathbb{P}_X)$.

Si ces propriétés sont réalisées, en utilisant cette fois (S8), nous voyons de la même manière que les deux expressions de droite de (3.3.11) sont aussi égales pour f , ce qui, compte-tenu de (3.3.2), achève la démonstration. \square

Définition 3.3.11 Soit $p \in \mathbb{N}^*$. On dit que la variable aléatoire X admet un moment d'ordre p si la variable aléatoire $X^p \in L^1$, et ce moment vaut alors (en utilisant la proposition précédente) :

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

3.4 Fonction génératrice d'une variable aléatoire à valeurs entières

Considérons une variable aléatoire à valeurs dans \mathbb{N} . Sa loi est une probabilité sur \mathbb{N} . Elle est donc, comme nous l'avons vu, caractérisée par une suite de nombres compris entre 0 et 1 et de somme 1. 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.

Nous considérons une variable aléatoire X à valeurs dans \mathbb{N} , dont la loi est caractérisée par les nombres $p_n = p_n^X = \mathbb{P}(X = n)$.

Définition 3.4.1 La fonction génératrice G_X de X est la fonction définie sur l'intervalle $[0, 1]$ par la formule suivante : $\forall s \in [0, 1]$,

$$G_X(s) = \mathbb{E}(s^X) = \sum_{n=0}^{\infty} p_n s^n. \quad (3.4.12)$$

Comme nous le voyons ci-dessus, la fonction génératrice ne dépend que de la loi de X , c'est à dire de la probabilité $(p_n)_n$.

Proposition 3.4.2 La fonction génératrice 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 3.4.3 *Soit X une variable aléatoire à valeurs entières, de fonction génératrice G_X . Pour que X soit intégrable, il faut et il suffit que G_X soit dérivable à gauche en $s = 1$, et dans ce cas $\mathbb{E}(X) = G_X'(1)$.*

Preuve. Rappelons d'abord un résultat facile sur les séries : si les fonctions $s \mapsto u_n(s)$ sont croissantes et positives sur $[0, 1[$, on peut échanger la limite en s au point 1 et la somme en n :

$$\lim_{s \uparrow 1} \sum_{n \geq 0} u_n(s) = \sum_{n \geq 0} \lim_{s \uparrow 1} u_n(s). \quad (3.4.13)$$

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 (3.4.13). \square

Plus généralement, la même démonstration prouve que

Proposition 3.4.4 *La variable aléatoire $X(X-1)\cdots(X-p)$ 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 en $s = 1$, et nous avons alors*

$$\mathbb{E}(X(X-1)\cdots(X-p)) = G_X^{(p+1)}(1). \quad (3.4.14)$$

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 (3.4.12) 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 (3.4.14) lorsque ce dernier existe, d'après (3.3.11). 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 3.4.5 *Pour calculer les moments d'une variable aléatoire (espérance, variance,...), il peut être beaucoup plus rapide et simple d'utiliser les dérivées de la fonction génératrice plutôt qu'un calcul direct, comme nous le verrons dans le paragraphe ci-dessous.*

3.5 Variables aléatoires discrètes usuelles

3.5.1 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 $1 - p$. Une telle variable est appelée **variable de Bernoulli de paramètre p** . Sa loi est la **loi de Bernoulli de paramètre p** .

Proposition 3.5.1 *L'espérance et la variance d'une variable aléatoire de Bernoulli de paramètre p valent respectivement*

$$\mathbb{E}(X) = p ; \tag{3.5.15}$$

$$\text{Var}(X) = p(1 - p), \tag{3.5.16}$$

et sa fonction génératrice vaut $G_X(s) = 1 - p + ps$.

Preuve. Nous avons

$$\begin{aligned} \mathbb{E}(X) &= p \times 1 + 0 \times (1 - p) = p; \\ \text{Var}(X) &= p - p^2 = p(1 - p). \end{aligned}$$

Le calcul de G_X est immédiat. □

3.5.2 Variable aléatoire binomiale

- Cherchons maintenant le nombre de Piles obtenus sur les n lancers.

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

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

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

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

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

Définition 3.5.2 On dit que la variable aléatoire X est une **variable aléatoire binomiale** de paramètres n et p , que l'on note $\mathcal{B}(n, p)$, 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 .

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 la loi $\mathcal{B}(1, p)$ est la loi de Bernoulli de paramètre p .

Proposition 3.5.3 Soit X une variable binomiale de loi $\mathcal{B}(n, p)$. Alors,

$$G_X(s) = (1 - p + ps)^n, \quad (3.5.17)$$

$$\mathbb{E}(X) = np, \quad (3.5.18)$$

$$\text{Var}(X) = np(1-p). \quad (3.5.19)$$

Preuve.

$$G_X(s) = \sum_{k=0}^n \binom{n}{k} p^k s^k (1-p)^{n-k} = (1-p+ps)^n.$$

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

$$G''_X(1) = \sum_{k=1}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} = E(X(X-1)),$$

et par suite, puisque $\mathbb{E}(X) = np$, nous en déduisons que

$$\text{Var}(X) = E(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X^2) = np(1-p). \quad (3.5.20)$$

□

Exemple 3.5.4 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 l'espérance de X sera égale à $86 \times 0,01 = 0,86$.

Cherchons la probabilité pour que le nombre de médailles soit inférieur à 3. Elle vaut

$$\mathbb{P}(X \leq 3) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3),$$

avec

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

Tous calculs faits, nous trouvons

$$\mathbb{P}(X \leq 3) = 0,9889.$$

Remarquons que la France a remporté 2 médailles d'or.

3.5.3 Probabilité de succès et variable aléatoire géométrique

• Toujours sur ce jeu de n lancers de notre pièce, intéressons-nous au premier temps où nous allons obtenir un Pile. Pour ce faire, définissons la variable aléatoire T par

$$\begin{aligned} T(\omega) &= \inf\{k \in \{1, \dots, n\}, X_k(\omega) = \text{Pile}\}, \text{ si cet ensemble n'est pas vide ;} \\ &= +\infty \quad \text{sinon.} \end{aligned}$$

L'ensemble des valeurs de T est donc $\{1, 2, \dots, n, +\infty\}$. Pour $1 \leq k \leq n$,

$$\mathbb{P}(T = k) = \mathbb{P}(X_1 = \text{Face}, \dots, X_{k-1} = \text{Face}, X_k = \text{Pile}) = p(1-p)^{k-1},$$

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

$$\mathbb{P}(T = +\infty) = (1-p)^n.$$

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 sur \mathbb{N}^* par

$$\mathbb{P}(T = k) = p(1-p)^{k-1},$$

pour $k \in \mathbb{N}^*$. On appelle cette probabilité la loi géométrique de paramètre p .

Définition 3.5.5 Une **variable géométrique** de paramètre $p \in]0, 1[$ est une variable aléatoire à valeurs dans \mathbb{N}^* telle que $\forall k \in \mathbb{N}^*$,

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

Proposition 3.5.6 Soit X une variable aléatoire de loi géométrique de paramètre p sur \mathbb{N}^* . On a :

$$\mathbb{E}(X) = \frac{1}{p}, \quad (3.5.21)$$

$$\text{Var}(X) = \frac{1-p}{p^2}, \quad (3.5.22)$$

$$G_X(s) = \frac{ps}{1-(1-p)s}. \quad (3.5.23)$$

En d'autres termes, en supposant que l'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.

Preuve.

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k p(1-p)^{k-1} < +\infty. \quad (3.5.24)$$

Calculons la fonction génératrice

$$G_X(s) = \sum_{k \geq 1} p(1-p)^{k-1} s^k = \frac{ps}{1-(1-p)s}.$$

Il est immédiat que $G'_X(s) = \frac{p}{(1-s+ps)^2}$ et donc $G'_X(1) = \frac{1}{p}$.

Nous en déduisons que $\mathbb{E}(X) = \frac{1}{p}$. Le calcul de la variance est similaire. \square

3.5.4 Variable aléatoire de Poisson

Définition 3.5.7 Nous dirons que X est une **variable aléatoire 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

$$\mathbb{P}(X = k) = e^{-\theta} \frac{\theta^k}{k!}. \quad (3.5.25)$$

Sa loi est une loi de Poisson de paramètre $\theta > 0$.

Proposition 3.5.8 Soit X une variable aléatoire de Poisson de paramètre θ , alors

$$\mathbb{E}(X) = \theta, \quad (3.5.26)$$

$$\text{Var}(X) = \theta, \quad (3.5.27)$$

$$G_X(s) = e^{\theta(s-1)}. \quad (3.5.28)$$

Preuve. Soit X une variable aléatoire de Poisson de paramètre θ . Son espérance vaut

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k e^{-\theta} \frac{\theta^k}{k!} = \theta.$$

Par ailleurs,

$$E(X(X-1)) = e^{-\theta} \sum_{k=2}^{\infty} k(k-1) \frac{\theta^k}{k!} = \theta^2$$

d'où $\text{Var}(X) = \theta$, vu que $\mathbb{E}(X) = \theta$.

Il est facile de montrer que $G_X(s) = e^{\theta(s-1)}$. En dérivant au point $s = 1$, nous retrouvons $\mathbb{E}(X) = \theta$. En dérivant deux fois, (3.4.14) donne $\mathbb{E}(X(X-1)) = \theta^2$, d'où nous pouvons aussi déduire la valeur de $\text{Var}(X)$. \square

Exemple 3.5.9 Admettons que le nombre d'erreurs Y par page de ce livre suive 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

$$\begin{aligned} p_j(a_n, n) &= \binom{n}{j} (a_n)^j (1 - a_n)^{n-j} & \text{si } j \leq n \\ &= 0 & \text{si } j \geq n + 1. \end{aligned} \quad (3.5.29)$$

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 suite infinie 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 obéissent à 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 3.5.4. 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.

3.6 Lois conditionnelles et variables aléatoires indépendantes

3.6.1 Lois conditionnelles

Nous rappelons ici que les notions de probabilités conditionnelles et d'événements aléatoires indépendants ont été introduites au chapitre 2.5.

Dans ce paragraphe, nous considérons deux variables aléatoires X et Y définies sur le même espace Ω fini ou dénombrable, muni de la probabilité \mathbb{P} . Nous supposons que X et Y sont à valeurs respectivement dans F et G , que nous pouvons supposer eux-mêmes finis ou dénombrables. Nous posons $p_i^X = \mathbb{P}(X = x_i)$ pour $x_i \in F$, et $p_j^Y = \mathbb{P}(Y = y_j)$ pour $y_j \in G$.

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 $F \times G$. Notons $\mathbb{P}_{\mathbf{Z}} = (p_k^{\mathbf{Z}}, \mathbf{z}_k \in F \times G)$ la loi du vecteur aléatoire \mathbf{Z} .

Pour $\mathbf{z}_k = (x_i, y_j) \in F \times G$

$$p_k^{\mathbf{Z}} = \mathbb{P}(\mathbf{Z} = \mathbf{z}_k) = \mathbb{P}(X = x_i \text{ et } Y = y_j).$$

Définition 3.6.1 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 .

Considérons l'expérience suivante. Nous lançons en même temps un dé rouge et un dé bleu. Soit 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 $X = x_i$.

Dans la suite de ce paragraphe, les rôles de X et de Y pourront être échangés.

Définition 3.6.2 Soit $x_i \in F$ tel que $\mathbb{P}(X = x_i) > 0$. On appelle **loi conditionnelle** de Y sachant $X = x_i$ la probabilité sur G définie par

$$p_j^{Y|X=x_i} = \mathbb{P}(Y = y_j | X = x_i) \quad \forall y_j \in G. \quad (3.6.30)$$

La loi conditionnelle de Y sachant $X = x_i$ est donc caractérisée par $\{(y_j, p_j^{Y|X=x_i}), y_j \in G\}$. Ces lois conditionnelles sont a priori différentes pour chaque valeur de x_i et différentes de la loi marginale P^Y .

Nous avons en fait les relations suivantes.

Proposition 3.6.3 Il est équivalent de connaître les $(p_k^Z : z_k \in F \times G)$ avec $z_k = (x_i, y_j)$, et les $(p_i^X : x_i \in F)$ et $(p_j^{Y|X=x_i} : y_j \in G)$ pour $x_i \in F$ tels que $p_i^X > 0$, via les formules :

$$\mathbb{P}(X = x_i) = \sum_{y_j \in F} \mathbb{P}(Z = (x_i, y_j)), \quad (3.6.31)$$

$$p_j^{Y|X=x_i} = \frac{\mathbb{P}(Z = (x_i, y_j))}{\mathbb{P}(X = x_i)} \quad \text{si } \mathbb{P}(X = x_i) > 0, \quad (3.6.32)$$

$$\mathbb{P}(Z = (x_i, y_j)) = \begin{cases} \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j | X = x_i) & \text{si } \mathbb{P}(X = x_i) > 0 \\ 0 & \text{sinon} \end{cases} \quad (3.6.33)$$

Preuve. Il suffit de montrer les trois formules de l'énoncé. Remarquons tout d'abord que l'ensemble $\{X = x_i\}$ est la réunion (finie ou dénombrable) des ensembles deux-à-deux disjoints $\{X = x_i, Y = y_j\} = \{Z = (x_i, y_j)\}$ pour $y_j \in G$, donc (3.6.31) découle immédiatement de la σ -additivité (2.3.19).

(3.6.32) vient de la formule (2.5.27).

Enfin si $p_i^X = \mathbb{P}(X = x_i) > 0$, (3.6.33) découle de (3.6.32) tandis que si $p_i^X = 0$ nous avons *a fortiori* $\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(Z = (x_i, y_j)) = 0$ d'après (2.2.11). \square

3.6.2 Espérance conditionnelle

Pour i fixé tel que $p_i^X > 0$, la loi conditionnelle de Y sachant $X = x_i$ donnée par $\{(y_j, p_j^{Y|X=x_i}), y_j \in G\}$ définit une probabilité. Nous pouvons lui associer son espérance, sa variance ou plus généralement ses moments, dès qu'ils existent. Par exemple, dès que Y est intégrable, son espérance conditionnellement à $\{X = x_i\}$ sera définie comme suit.

Définition 3.6.4 Soit Y une variable aléatoire intégrable. L'espérance conditionnelle de Y sachant $\{X = x_i\}$ est l'espérance de la loi conditionnelle de Y sachant $\{X = x_i\}$, c'est-à-dire

$$\mathbb{E}(Y|X = x_i) = \sum_j y_j p_j^{Y|X=x_i} = \sum_j y_j \mathbb{P}(Y = y_j|X = x_i). \quad (3.6.34)$$

Remarque 3.6.5 La série définie en (3.6.34) est bien convergente. En effet, nous pouvons utiliser l'observation suivante

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

et le théorème de Fubini (voir (S9)).

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 3.6.6 On appelle **espérance conditionnelle de Y sachant X** la variable aléatoire

$$\mathbb{E}(Y|X) = \psi(X), \quad \text{avec} \quad \psi(x) = \mathbb{E}(Y|X = x), \quad (3.6.35)$$

pour x tel que $\mathbb{P}(X = x) > 0$, et $\psi(x) = 0$ sinon.

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 possède la propriété fondamentale suivante.

Théorème 3.6.7 Si Y est intégrable, alors $\psi(X) = \mathbb{E}(Y | X)$ est intégrable, et

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

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 (2.5.29), qui correspond ici à $Y = \mathbf{1}_A$, et $B_i = \{X = x_i\}$. On l'écrit souvent sous forme

$$\mathbb{E}(\mathbb{E}(Y | X)) = \mathbb{E}(Y) . \quad (3.6.36)$$

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,

$$\mathbb{E}(Y g(X) | X) = g(X) \mathbb{E}(Y | X) \quad (3.6.37)$$

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*.)

Exemple 3.6.8 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. Cherchons $\mathbb{E}(K)$.

Solution : Nous avons par hypothèse,

$$p_N(n) = \frac{\lambda^n}{n!} e^{-\lambda} \quad , \quad p^{K|N=n}(k) = \binom{n}{k} p^k (1-p)^{n-k} .$$

D'où

$$\mathbb{E}(K | N = n) = \sum_k k p^{K|N=n}(k) = n p ,$$

soit $\mathbb{E}(K | N) = p N$. D'après le théorème 3.6.7,

$$\mathbb{E}(K) = \mathbb{E}(\mathbb{E}(K | N)) = \mathbb{E}(Np) = p \mathbb{E}(N) = p \lambda$$

puisque $\lambda = \mathbb{E}(N)$.

3.6.3 Variables aléatoires indépendantes

Remarque : La formule (3.6.31) de la proposition précédente montre que l'on peut calculer les lois marginales \mathbb{P}_X et \mathbb{P}_Y de X et de Y à partir de la loi \mathbb{P}_Z de Z .

En revanche, la connaissance des lois marginales ne suffit pas, en général, à retrouver la loi de $Z = (X, Y)$, comme le prouve l'exemple suivant. 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 $p \neq \frac{1}{2}$. Dans ce cas, 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

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

Il est très intéressant d'étudier le cas où l'information que l'on possède sur X ne change rien à la loi de Y , généralisant ainsi la notion d'indépendance introduite pour les événements aléatoires.

Définition 3.6.9 Les variables aléatoires X et Y sont dites **indépendantes** si pour toutes parties $A \subset F$, $B \subset G$ elles vérifient

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B). \quad (3.6.38)$$

Remarque : la notation $\mathbb{P}(X \in A, Y \in B)$ signifie usuellement $\mathbb{P}(\{X \in A\} \cap \{Y \in B\}) = \mathbb{P}(X \in A \text{ et } Y \in B)$.

Nous avons la propriété fondamentale suivante.

Proposition 3.6.10 *Il y a équivalence entre :*

(i) *Les variables aléatoires X et Y sont indépendantes, de lois respectives p^X et 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 $x_i \in F$, $y_j \in G$.*

(iii) *On a $p_j^{Y|X=x_i} = p_j^Y$ pour tout $y_j \in G$ et tout $x_i \in F$ tel que $p_i^X > 0$.*

Remarque 3.6.11 (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 3.6.9 de l'indépendance est symétrique en X et Y , nous pouvons ci-dessus échanger les variables aléatoires X et Y .

Preuve. Pour obtenir (i) \Rightarrow (ii) il suffit de prendre $A = \{x_i\}$ et $B = \{y_j\}$ dans (3.6.38). Inversement, supposons (ii). En sommant par paquets dans la série à termes positifs, nous obtenons pour $A \subset F$ et $B \subset G$:

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

et (i) s'en déduit. Enfin, l'équivalence (ii) \Leftrightarrow (iii) provient de (3.6.32) et (3.6.33). \square

Proposition 3.6.12 *Supposons les variables aléatoires X et Y indépendantes. Soient f et g deux fonctions réelles sur F et G 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*

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X)) \mathbb{E}(g(Y)). \quad (3.6.39)$$

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

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

qui est fini par hypothèse. Par suite, la variable aléatoire $f(X)g(Y)$ est intégrable. En utilisant alors (S8), la même démonstration montre que les égalités ci-dessus sont également vérifiées en enlevant les valeurs absolues : nous obtenons (3.6.39). \square

Exemple 3.6.13 Considérons n variables aléatoires 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

$$\mathbb{P}(X_i = x_i, 1 \leq i \leq n) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}, \quad (3.6.40)$$

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

Nous souhaitons simuler ces n variables aléatoires de Bernoulli indépendantes et de paramètre $p \in]0, 1[$. Nous allons utiliser un générateur de nombres au hasard, i.e. un algorithme qui fournit une suite de nombres compris entre 0 et 1 ayant les mêmes caractéristiques que les tirages $(U_i)_{i=1}^n$ d'une suite de variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. (Voir le chapitre suivant pour la définition d'une loi uniforme.) Nous posons alors

$$X_i = \begin{cases} 1 & \text{si } U_i \leq p \\ 0 & \text{si } U_i > p \end{cases}$$

Il est facile de voir que les variables aléatoires X_i sont indépendantes et de loi de Bernoulli de paramètre p .

Exemple 3.6.14 On admet que le nombre de clients dans un bureau de poste pendant une journée soit une variable aléatoire de Poisson de paramètre λ . 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 variables aléatoires indépendantes, et suivent des lois de Poisson de paramètres respectifs λp et $\lambda(1-p)$. Pour s'en assurer, le lecteur pourra utiliser les calculs de l'exemple 3.6.8.

3.6.4 Somme de variables aléatoires indépendantes

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

Proposition 3.6.15 *Supposons que X et Y soient deux variables aléatoires à valeurs dans \mathbb{Z} et notons $\mathbf{Z} = (X, Y)$. Alors*

$$\mathbb{P}(X + Y = i) = \sum_{j \in \mathbb{Z}} \mathbb{P}(\mathbf{Z} = (j, i - j)) = \sum_{j \in \mathbb{Z}} \mathbb{P}(\mathbf{Z} = (i - j, j)). \quad (3.6.41)$$

En particulier si X et Y sont indépendantes, on a

$$\mathbb{P}(X + Y = i) = \sum_{j \in \mathbb{Z}} \mathbb{P}(X = j)\mathbb{P}(Y = i - j) = \sum_{j \in \mathbb{Z}} \mathbb{P}(X = i - j)\mathbb{P}(Y = j). \quad (3.6.42)$$

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

Preuve. (3.6.42) découle immédiatement de (3.6.41) et de (ii) de la proposition 3.6.10. Pour (3.6.41), il suffit d'appliquer (2.3.19) 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 3.6.16 Ces formules peuvent se généraliser à la somme de n variables aléatoires indépendantes. En particulier, (3.6.40) 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 $\mathcal{B}(n, p)$ doit être interprétée comme la loi de la somme de n variables aléatoires indépendantes, de loi de Bernoulli de paramètre p .

Proposition 3.6.17 *Supposons les variables aléatoires X et Y indépendantes, à valeurs dans $F = G = \mathbb{N}$, 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 $s \in [0, 1]$*

$$G_U(s) = G_X(s)G_Y(s). \quad (3.6.43)$$

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 (3.6.39). \square

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

Exemple 3.6.18

- Soient X et Y des variables aléatoires indépendantes de loi binomiale respectivement $\mathcal{B}(n, p)$ et $\mathcal{B}(m, p)$ (avec le même paramètre p). D'après (3.5.17), $U = X + Y$ vérifie

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

En appliquant encore (3.5.17) et la proposition 3.4.2, nous en déduisons que $X + Y$ suit **la loi binomiale** $\mathcal{B}(n + m, p)$.

- Soient X et Y des variables aléatoires indépendantes de loi de Poisson de paramètres respectifs θ et ζ . D'après (3.5.28), $U = X + Y$ 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 $\theta + \zeta$.

3.7 Exercices sur le chapitre 3

EXERCICE 3.7.1 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 $\mathbb{E}(X)$, $\text{Var}(X)$.

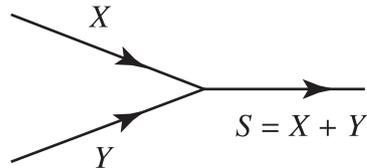
EXERCICE 3.7.2 Si X prend ses valeurs dans \mathbb{N} , montrer que

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k).$$

EXERCICE 3.7.3 Le nombre d'accidents N en une semaine dans une usine est aléatoire d'espérance m et de variance σ^2 . Le nombre d'individus X blessés dans un accident est aléatoire, d'espérance μ et de variance τ^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 , σ^2 , μ et τ^2 .

EXERCICE 3.7.4 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 3.7.5 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 :

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

Montrer que les variables aléatoires Z et X ne sont pas indépendantes mais que Y et X le sont.

Donner la loi conditionnelle de X sachant $Z = k$. En déduire l'espérance conditionnelle de X sachant Z .

EXERCICE 3.7.6 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

$$\mathbb{P}(X_n = 1) = p,$$

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 3.7.7 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 $U_p : \mathbb{N}^* \rightarrow \mathbb{N}$ telle que

$$\forall x \in \mathbb{N}^*, x = \prod_{p \in \mathcal{P}} p^{U_p(x)}.$$

Soit \mathbb{Q} la probabilité sur \mathbb{N}^* définie par

$$\mathbb{Q}(\{x\}) = \frac{c}{x^2}, \quad (0 < c < \infty).$$

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 .

En déduire son espérance et sa variance.

EXERCICE 3.7.8 Probabilités de Gibbs sur un système fini.

Rappel : Considérons ψ 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 Ω .

On définit l'entropie de cette probabilité p en posant

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

1-a) Vérifier que la fonction $\phi : [0, 1] \rightarrow [0, +\infty[$, $\phi(t) = -t \log t$ est strictement concave.

1-b) Montrer que $H(p) = 0$ si et seulement si p est une mesure de Dirac sur Ω .

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 Ω , 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 qui à chaque $\beta \in \mathbb{R}$ associe le nombre

$$Z(\beta) = \sum_{i=1}^n e^{-\beta U(\omega_i)}.$$

La probabilité de Gibbs associée est alors définie pour chaque ω_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 β tend vers $+\infty$, μ_β devient une probabilité uniforme sur l'ensemble $\Omega_{\min} := \{\omega; U(\omega) = \min_{\Omega} U\}$, et que lorsque $\beta \rightarrow -\infty$, μ_β devient une probabilité uniforme sur $\Omega_{\max} := \{\omega; U(\omega) = \max_{\Omega} U\}$. En déduire que

$$\lim_{\beta \rightarrow +\infty} \langle U \rangle_{\mu_\beta} = \min_{\Omega} U \quad ; \quad \lim_{\beta \rightarrow -\infty} \langle U \rangle_{\mu_\beta} = \max_{\Omega} U. \quad (3.7.44)$$

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

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

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

3) Notre but est de rechercher une probabilité μ sur Ω 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 acceptera que, par un théorème de Lagrange, μ doit être un extremum de la fonction

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

où β est un paramètre à déterminer par la contrainte $\langle U \rangle_\mu = E$ et λ un paramètre à déterminer par la contrainte que μ 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 μ_β 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 μ_β est bien un maximum et que c'est en fait l'unique maximum.

Chapitre 4

Variables aléatoires réelles et vecteurs aléatoires

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 Les variables aléatoires réelles

Les variables aléatoires que nous considérons maintenant peuvent être à valeurs dans \mathbb{R} tout entier. Nous souhaitons que la tribu \mathcal{E} définie dans la proposition 2.4.1 contienne la tribu borélienne $\mathcal{B}(\mathbb{R})$ de \mathbb{R} , définie en Définition 2.3.5. La justification de ce choix se trouvera au paragraphe 4.2.1. Cela nous conduit à poser :

Définition 4.1.1 Soit Ω l'espace fondamental muni de la tribu \mathcal{A} des événements. Une application X de Ω dans \mathbb{R} est une **variable aléatoire** réelle si $X^{-1}(B) = \{X \in B\} \in \mathcal{A}$ pour tout $B \in \mathcal{B}(\mathbb{R})$.

Nous avons alors le résultat très utile suivant :

Proposition 4.1.2 Si X_1, \dots, X_n sont des variables aléatoires réelles et si f est une fonction continue de \mathbb{R}^n dans \mathbb{R} , alors $Y = f(X_1, \dots, X_n)$ est une variable aléatoire réelle.

Preuve. 1) Montrons d’abord un résultat auxiliaire, à savoir que si X est une application de Ω dans \mathbb{R} telle que $\{\omega : X(\omega) \leq a\} = X^{-1}(]-\infty, a])$ appartient à la tribu \mathcal{A} pour tout $a \in \mathbb{R}$, alors X est une variable aléatoire réelle.

Pour cela, nous notons \mathcal{R} l’ensemble des $B \in \mathcal{B}(\mathbb{R})$ tels que $X^{-1}(B)$ soit dans \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 $\mathcal{B}(\mathbb{R})$, et elle contient les intervalles $]-\infty, a]$ par construction. Par la proposition 2.3.6, nous avons donc $\mathcal{R} = \mathcal{B}(\mathbb{R})$, ce qui veut dire que $X^{-1}(B) \in \mathcal{A}$ pour tout $B \in \mathcal{B}(\mathbb{R})$, et X est une variable aléatoire.

2) D’après ce qui précède, il suffit de montrer que les ensembles $\{Y \leq a\}$, ou de manière équivalente les ensembles $\{Y > a\}$, sont dans \mathcal{A} pour tout $a \in \mathbb{R}$. Or f étant continue, $A = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) > a\}$ est un ouvert. Il s’écrit donc comme réunion dénombrable $A = \cup_{i \in \mathbb{N}} A_i$ d’ensembles A_i qui sont des “rectangles ouverts” de la forme $A_i = \prod_{j=1}^n]x_{i,j}, y_{i,j}[$, et nous avons

$$\{Y > a\} = \{(X_1, \dots, X_n) \in A\} = \cup_i \{(X_1, \dots, X_n) \in A_i\} = \cup_i \cap_{j=1}^n \{x_{i,j} < X_j < y_{i,j}\}.$$

Comme par hypothèse $\{x_{i,j} < X_j < y_{i,j}\} \in \mathcal{A}$, nous en déduisons le résultat. \square

Comme application de cet énoncé ou de la partie 1) de sa preuve, nous avons les propriétés suivantes.

Proposition 4.1.3 Soient X, Y et $(X_n)_{n \in \mathbb{N}^*}$ des variables aléatoires réelles. Alors,

•

$$X + Y, \quad XY, \quad \frac{X}{Y} \quad \text{si } Y \neq 0, \quad \text{sont des variables aléatoires.} \quad (4.1.1)$$

•

$$\sup_{1 \leq n \leq p} X_n, \quad \inf_{1 \leq n \leq p} X_n, \quad \text{sont des variables aléatoires.} \quad (4.1.2)$$

•

$$\sup_{n \geq 1} X_n, \quad \inf_{n \geq 1} X_n, \quad \text{sont des variables aléatoires.} \quad (4.1.3)$$

(En effet, remarquons que $\{\sup_i X_i \leq a\} = \cap_i \{X_i \leq a\} \in \mathcal{A}$ par hypothèse, et appliquons la partie 1) de la preuve précédente; même chose pour l’inf.)

•

$$\text{Si } X_n(\omega) \xrightarrow{n \rightarrow \infty} Z(\omega), \quad \forall \omega, \quad \text{alors la limite } Z \text{ est une variable aléatoire.} \quad (4.1.4)$$

(Nous avons en effet $Z = \inf_n Y_n$, où $Y_n = \sup_{i \geq n} X_i$, et nous appliquons deux fois (4.1.3)).

•

$Z = \mathbf{1}_A$ (indicatrice de A) est une variable aléatoire $\iff A \in \mathcal{A}$. (4.1.5)

(Il suffit de remarquer que $Z^{-1}(B)$ est égal à \emptyset , A , A^c ou Ω selon que $B \cap \{0, 1\}$ soit égal à \emptyset , $\{1\}$, $\{0\}$ ou $\{0, 1\}$).

Dans le cas où $n = 1$ (le cas général sera vu ultérieurement), nous pouvons généraliser la Proposition 4.1.2 à des fonctions f beaucoup plus générales, appelées fonctions mesurables. Ces fonctions seront définies comme suit.

Définition 4.1.4 Une fonction f de \mathbb{R} dans \mathbb{R} est dite **mesurable** si l'image réciproque par f d'un ensemble de la tribu borélienne de \mathbb{R} est un ensemble de la tribu borélienne.

La partie 1) de la preuve de la Proposition 4.1.2 implique qu'il suffit de montrer que pour tout a réel, $f^{-1}(]-\infty, a])$ est un ensemble de la tribu borélienne. Cette définition est liée à la théorie de la mesure qui dépasse largement le cadre de ce cours. La propriété d'une fonction mesurable que nous retiendrons sera que si X est une variable aléatoire réelle, $f(X)$ l'est aussi.

4.2 Les lois de variables aléatoires réelles

Dès lors que la notion de variable aléatoire réelle est bien établie, nous pouvons définir rigoureusement la loi \mathbb{P}_X de X introduite en Section 2.4. C'est la probabilité définie sur \mathbb{R} muni de la tribu borélienne $\mathcal{B}(\mathbb{R})$, qui vérifie pour tout ensemble borélien $B \in \mathcal{B}(\mathbb{R})$,

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

Nous sommes donc amenés à étudier ce qu'est une probabilité sur \mathbb{R} muni de sa tribu borélienne.

4.2.1 Fonction de répartition

1) Définition

Soit X une variable aléatoire réelle et \mathbb{P}_X sa loi.

Rappelons que la tribu borélienne de \mathbb{R} est engendrée par les intervalles de la forme $] -\infty, a]$, avec $a \in \mathbb{Q}$.

Un résultat fondamental sur \mathbb{R} , et lié à cette propriété de la tribu borélienne, est que la probabilité \mathbb{P}_X est caractérisée par une fonction réelle de variable réelle, objet mathématique que l'on connaît beaucoup mieux.

Définition 4.2.1 Soit X une variable aléatoire et \mathbb{P}_X sa loi. On appelle **fonction de répartition de X** la fonction suivante :

$$F_X(x) = \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X \leq x), \quad \forall x \in \mathbb{R}. \quad (4.2.6)$$

Exemple 4.2.2 Si \mathbb{P}_X est la mesure de Dirac en 0 (voir Exemple 2.3.14), c'est-à-dire si $X = 0$ presque-sûrement, la fonction de répartition est la mesure de Heaviside en 0 : $H(x) = 0$ si $x < 0$ et $H(x) = 1$ si $x \geq 0$.

Remarque : L'exemple ci-dessus montre qu'en général, une fonction de répartition n'est pas une fonction continue. Nous verrons ci-dessous une caractérisation des points où elle est continue.

Proposition 4.2.3 1) La fonction de répartition F_X caractérise la probabilité \mathbb{P}_X .
2) La fonction de répartition vérifie les trois conditions suivantes :

- (i) elle est croissante
- (ii) elle est continue à droite
- (iii) $\lim_{x \downarrow -\infty} F(x) = 0$, $\lim_{x \uparrow +\infty} F(x) = 1$.

Preuve. Puisqu'il n'y a pas d'ambiguïté, notons plus simplement F_X par F .

1) D'après (4.2.6), nous avons $\mathbb{P}_X(]x, y]) = F(y) - F(x)$ 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

$$\mathbb{P}_X(B) = \sum_{i=1}^n \mathbb{P}_X(]x_i, y_i]) = \sum_{i=1}^n (F(y_i) - F(x_i)),$$

car les intervalles sont disjoints. Puisque $\mathbb{P}_X(]x, +\infty]) = 1 - F(x)$, nous en déduisons finalement que F caractérise la restriction de \mathbb{P}_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 (on dit que c'est une algèbre). Un résultat difficile de théorie de la mesure (que nous admettrons, voir le livre de Jacod-Protter pour une preuve), montre que la connaissance de \mathbb{P}_X sur cette

algèbre suffit à déterminer entièrement \mathbb{P}_X sur la tribu engendrée par cette algèbre. Mais nous avons vu (Proposition 2.3.6) que cette tribu est la tribu borélienne.

2) La croissance de F est immédiate. Pour montrer (ii), nous remarquons que si la suite x_n décroît vers x , alors $] - \infty, x_n]$ décroît vers $] - \infty, x]$ et donc $F(x_n)$ décroît vers $F(x)$. (iii) se montre de manière analogue, en remarquant que $] - \infty, x]$ décroît vers \emptyset (resp. croît vers \mathbb{R}) lorsque x décroît vers $-\infty$ (resp. croît vers $+\infty$). \square

Comme F est croissante, elle admet une limite à gauche en chaque point notée $F(x-)$. En remarquant que $] - \infty, y[= \lim_n] - \infty, y_n]$ si $y_n \uparrow y$, nous pouvons facilement obtenir que pour $x < y$,

$$\begin{aligned} \mathbb{P}_X(]x, y]) &= \mathbb{P}(x < X \leq y) = F(y) - F(x), \\ \mathbb{P}_X(]x, y[) &= \mathbb{P}(x < X < y) = F(y-) - F(x), \\ \mathbb{P}_X([x, y]) &= \mathbb{P}(x \leq X \leq y) = F(y) - F(x-), \\ \mathbb{P}_X([x, y[) &= \mathbb{P}(x \leq X < y) = F(y-) - F(x-). \end{aligned} \tag{4.2.7}$$

En particulier,

$$\mathbb{P}_X(\{x\}) = F(x) - F(x-)$$

est le saut de la fonction F au point x . Nous avons donc

Proposition 4.2.4

$$\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x) = 0 \iff F \text{ est continue en } x.$$

La proposition 4.2.3 admet une réciproque, qui est un des résultats les plus difficiles de la théorie de la mesure.

Théorème 4.2.5 (non démontré). *Si F est une fonction réelle sur \mathbb{R} , vérifiant les conditions (i)–(iii) de la Proposition 4.2.3, c’est la fonction de répartition d’une (unique) probabilité μ sur \mathbb{R} muni de sa tribu borélienne. On ne peut pas, en général, définir μ sur la tribu $\mathcal{P}(\mathbb{R})$ de toutes les parties de \mathbb{R} .*

Remarque : Le théorème ci-dessus explique pourquoi, d’un point de vue strictement mathématique, il est nécessaire d’introduire les tribus en probabilités, malgré la complexité que cela engendre. Sinon, cela reviendrait à prendre (sans le dire) la tribu

$\mathcal{P}(\mathbb{R})$ et il n'existerait que très peu de probabilités sur \mathbb{R} , à savoir les “probabilités discrètes” que nous avons introduites aux chapitres précédents.

2) Fonction de répartition d'une variable aléatoire X discrète

Étudions la fonction de répartition d'une variable aléatoire discrète.

- 1) X est identiquement égale au réel $a \in \mathbb{R}$. Alors sa loi est la mesure de Dirac en a et sa fonction de répartition est $F(x) = 1_{[a, \infty[}(x)$.
- 2) X **prend ses valeurs dans** \mathbb{N} . La loi \mathbb{P}_X de X est caractérisée par la suite $p_n = \mathbb{P}_X(\{n\}) = \mathbb{P}(X = n)$. La fonction de répartition F de X vaut alors

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ p_0 + \cdots + p_n & \text{si } n \leq x < n + 1, n \in \mathbb{N}. \end{cases} \quad (4.2.8)$$

La fonction F est une fonction en escalier qui saute de l'amplitude p_n au point n . Puisque $F(x) \in [0, 1]$, F admet au plus k sauts de taille supérieure à $\frac{1}{k}$, pour tout $k \in \mathbb{N}^*$.

- 3) Plus généralement, si X prend ses valeurs dans une partie finie ou dénombrable E de \mathbb{R} , la loi \mathbb{P}_X de X est caractérisée pour tout $i \in E$ par $q_i = \mathbb{P}_X(\{i\}) = \mathbb{P}(X = i)$. La fonction de répartition F de X est alors

$$F(x) = \sum_{i \in E: i \leq x} q_i, \quad (4.2.9)$$

avec la convention qu'une somme “vide” vaut 0. Nous retrouvons bien (4.2.8) dans le cas où $E = \mathbb{N}$. Nous voyons que la fonction F est **purement discontinue**, au sens où elle est complètement caractérisée par ses sauts $\Delta F(x) = F(x) - F(x-)$, via la formule

$$F(x) = \sum_{y \leq x} \Delta F(y).$$

Notons aussi que l'ensemble E , 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 $i \in \mathbb{Q}$, la fonction F nous donne un exemple de fonction discontinue en tout nombre rationnel, et continue partout ailleurs.

3) La mesure de Lebesgue sur $[0, 1]$

La fonction F , nulle sur \mathbb{R}_- , qui vaut $F(x) = x$ sur $[0, 1]$ et $F(x) = 1$ pour $x \geq 1$, satisfait les hypothèses du théorème 4.2.5. Nous obtenons alors le corollaire fondamental suivant.

Corollaire 4.2.6 Si $\mathcal{B}([0, 1])$ désigne la tribu borélienne sur $[0, 1]$, il existe une unique probabilité λ sur $([0, 1], \mathcal{B}([0, 1]))$ telle que la fonction de répartition soit égale à l'identité :

$$\lambda(]x, y]) = y - x, \quad \forall x < y, \quad x, y \in [0, 1].$$

Cette probabilité est appelée **mesure de Lebesgue** sur $[0, 1]$. La fonction F étant continue, cette probabilité ne charge pas les points : $\lambda(\{x\}) = 0, \forall x \in [0, 1]$.

Il existe bien d'autres probabilités, non discrètes, sur \mathbb{R} . Le paragraphe suivant est consacré à un exemple très important, celui des variables aléatoires dont la loi a une densité.

4.2.2 Variables aléatoires de loi à densité

Dans la suite de ce chapitre nous aurons continuellement à considérer des intégrales de fonctions réelles sur \mathbb{R} . L'intervalle d'intégration sera en général \mathbb{R} tout entier, ou parfois un intervalle de la forme $] - \infty, x]$. Il s'agit donc la plupart du temps d'intégrales "généralisées". Dans ce qui suit nous parlerons de **fonctions intégrables**. Cela signifie que la fonction f qu'on intègre est intégrable au sens de Riemann ou de Lebesgue (ce qui est plus général), selon les connaissances préalables du lecteur. Nous supposons plus précisément que l'intégrale généralisée $\int_{-\infty}^{+\infty} |f(x)| dx$ est finie, ce qui implique que l'intégrale généralisée $\int_{-\infty}^{+\infty} f(x) dx$ existe.

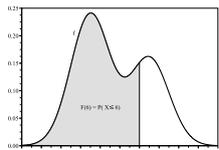
Définition 4.2.7 Une fonction réelle f définie sur \mathbb{R} est une **densité de probabilité**, ou simplement une densité, si elle est positive, intégrable, et vérifie

$$\int_{-\infty}^{+\infty} f(x) dx = 1. \quad (4.2.10)$$

Si f est comme ci-dessus, la fonction

$$F(x) = \int_{-\infty}^x f(y) dy \quad (4.2.11)$$

vérifie les propriétés (i, ii, iii) de la Proposition 4.2.3. C'est donc la fonction de répartition d'une probabilité.

FIGURE 4.1 – Graphe de la densité f et de la fonction de répartition F .

Remarque 4.2.8 $F(x)$ est la surface grise dans la Figure 4.1, délimitée par l'axe $y = 0$, le graphe de f , et la droite verticale d'abscisse x .

Définition 4.2.9 Soit X une variable aléatoire. On dit que X a une loi de densité f (ou par abus de langage "est de densité f "), si \mathbb{P}_X admet la densité f , et donc si pour tout réel x ,

$$\mathbb{P}(X \leq x) = \int_{-\infty}^x f(y)dy.$$

La proposition suivante se démontre immédiatement en utilisant des résultats bien connus d'analyse.

Proposition 4.2.10 Soit X de loi de densité f .

1) Sa fonction de répartition F est continue, de sorte que

$$\mathbb{P}(X = x) = 0, \quad \forall x \in \mathbb{R}.$$

2) La fonction F est dérivable en tout point x où f est continue, et

$$F'(x) = f(x).$$

3) A l'inverse, si la fonction de répartition F de X est dérivable, ou seulement continue partout et dérivable par morceaux, alors X admet la densité

$$F'(x) = f(x).$$

Remarque 4.2.11 Voici une interprétation “intuitive” de la densité f de \mathbb{P}_X . Si Δx est un “petit” accroissement de la quantité x , on a (si du moins f est continue en x) :

$$f(x) \sim \frac{\mathbb{P}_X([x, x + \Delta x])}{\Delta x} = \frac{F(x + \Delta x) - F(x)}{\Delta x}. \quad (4.2.12)$$

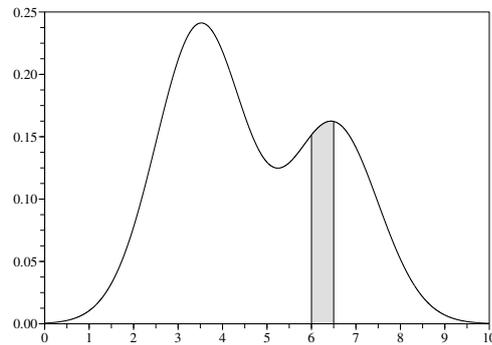


FIGURE 4.2 – $f(6) \sim (F(6,5) - F(6))/0,5$

Remarque 4.2.12 1) La fonction de répartition est, de manière évidente, entièrement déterminée par la probabilité \mathbb{P}_X . Il n’en est pas de même de la densité, lorsqu’elle existe. Si en effet nous avons (4.2.11) et si nous posons $g(x) = f(x)$ si $x \notin \mathbb{Q}$ et $g(x) = f(x) + 1$ si $x \in \mathbb{Q}$, 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.

2) Il existe des variables aléatoires qui n’ont pas de densité : c’est le cas des variables aléatoires discrètes. Il existe des cas “mixtes”. Nous nous donnons d’une part une fonction f positive intégrable et d’intégrale strictement positive, et d’autre part une partie finie ou dénombrable I de \mathbb{R} non vide et des $q_i > 0$ indicés par $i \in I$, tels que

$$\int_{-\infty}^{+\infty} f(x)dx + \sum_{i \in I} q_i = 1.$$

Alors, la fonction

$$F(x) = \int_{-\infty}^x f(y)dy + \sum_{i \in I: i \leq x} q_i \quad (4.2.13)$$

est une fonction de répartition, et la probabilité \mathbb{P}_X associée n'admet pas de densité et n'est pas non plus discrète.

Exemple 4.2.13 La durée de fonctionnement d'un ordinateur avant sa première panne est une variable aléatoire positive de densité donnée par

$$f(x) = \begin{cases} \frac{1}{100}e^{-\frac{x}{100}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Calculons la probabilité que cette durée de fonctionnement X soit comprise entre 50 et 150 heures. Elle vaut

$$\mathbb{P}(X \in [50, 100]) = \int_{50}^{100} \frac{1}{100}e^{-\frac{x}{100}} dx = \frac{\sqrt{e} - 1}{e} \approx 0,24.$$

Calculons la probabilité que l'ordinateur fonctionne moins de 100 heures.

$$\mathbb{P}(X \leq 100) = \int_0^{100} \frac{1}{100}e^{-\frac{x}{100}} dx = \frac{\sqrt{e} - 1}{e} \approx 0,63.$$

4.2.3 Variable aléatoire uniforme sur $[0, 1]$ et générateurs de nombres aléatoires

Les plus simples des variables aléatoires réelles de loi à densité sont les variables uniformes sur $[0, 1]$. Leur densité vaut

$$f(x) = \mathbf{1}_{[0,1]}(x).$$

Une valeur d'une telle variable a autant de chance d'appartenir au voisinage de chaque point de $[0, 1]$.

Il n'est pas facile¹ de trouver un "bon" générateur de nombres au hasard. Actuellement, tous les générateurs 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 ϕ 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 vivement conseillé

1. PARK & MILLER, Random number generators, good ones are hard to find, Commun. ACM (1988), **31**, p. 1192-1201.

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 congruents 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 twister² de période colossale $2^{19937} - 1$. C'est le même générateur que celui utilisé par la bibliothèque Numpy en PYTHON.

4.2.4 Simulation d'une variable aléatoire par inversion de la fonction de répartition

A partir d'une variable aléatoire uniforme U sur $[0, 1]$, générée comme ci-dessus, on peut simuler une variable aléatoire réelle X de loi quelconque, dès lors que l'on connaît sa fonction de répartition F .

Définissons la fonction "inverse" continue à gauche de F par :

$$G(u) = \inf\{x : F(x) \geq u\}, \quad \forall u \in]0, 1[. \quad (4.2.14)$$

La fonction G n'est pas à proprement parler l'inverse (ou la fonction réciproque) de F , puisque celle-ci n'est pas nécessairement bijective de \mathbb{R} dans $]0, 1[$. Elle joue cependant le même rôle, dans la mesure où elle vérifie

$$G(u) \leq x \iff u \leq F(x),$$

et $F(G(u)) = u$ si $0 < u < 1$ et si la fonction F est continue au point $G(u)$.

2. http://fr.wikipedia.org/wiki/Mersenne_Twister

Proposition 4.2.14 Soit U une variable aléatoire de loi uniforme sur $[0, 1]$, F une fonction de répartition et G définie par (4.2.14). La variable aléatoire $X = G(U)$ est alors une variable aléatoire de fonction de répartition F .

Preuve. Nous remarquons facilement que si $x \in \mathbb{R}$, alors

$$\mathbb{P}(X \leq x) = \mathbb{P}(G(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x),$$

puisque U suit une loi uniforme sur $[0, 1]$. \square

Exemple 4.2.15 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

$$X = x_i \quad \text{si} \quad \sum_{j < i} p_j < U \leq \sum_{j < i} p_j + p_i \quad , \quad i \geq 1$$

qui définit une variable aléatoire 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 : Le choix de la valeur de X lorsque $U = 1 - p$ importe peu, car la probabilité que $U = 1 - p$ exactement vaut 0.

4.3 Espérance des variables aléatoires réelles

4.3.1 Définition

Dans cette partie, nous voulons généraliser la notion d'espérance introduite au Chapitre 3 à toutes les variables aléatoires réelles "pas trop grandes" (en un certain sens), par exemple aux variables aléatoires bornées. Nous allons nous contenter de résultats partiels, les résultats complets étant difficiles à démontrer. L'idée naturelle est de se ramener au chapitre précédent en approchant X par une suite de variables aléatoires **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 variable aléatoire (ici c'est \mathbb{R}) mais qu'on connaît très mal son espace de

départ (l'espace abstrait Ω). Aussi, l'idée majeure que nous allons exposer ici, et à la base de la théorie de l'intégration de Lebesgue, est d'approcher nos variables aléatoires par une suite de variables aléatoires obtenues par un découpage de l'espace d'arrivée, et non pas de l'espace de départ (comme dans le cas de l'intégrale de Riemann).

Définition 4.3.1 Nous appellerons **variable aléatoire étagée** toute variable aléatoire Y qui ne prend qu'un nombre fini de valeurs, disons a_1, \dots, a_p . D'après (3.3.2), elle admet donc une espérance donnée par

$$\mathbb{E}(Y) = \sum_{i=1}^p a_i \mathbb{P}(Y = a_i). \quad (4.3.15)$$

Nous allons maintenant définir l'espérance d'une variable aléatoire réelle X . Construisons dans une première étape l'espérance des variables aléatoires positives. Si X est une telle variable, nous considérons une suite $(X_n)_n$ de variables aléatoires positives étagées croissant vers X , par exemple

$$X_n(\omega) = \begin{cases} k2^{-n} & \text{si } k2^{-n} \leq X(\omega) < (k+1)2^{-n} \text{ et } 0 \leq k \leq n2^n - 1 \\ n & \text{sinon.} \end{cases} \quad (4.3.16)$$

Comme $X_n \leq X_{n+1}$, nous avons $\mathbb{E}(X_n) \leq \mathbb{E}(X_{n+1})$ par (3.3.7), et nous pouvons poser

$$\mathbb{E}(X) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n). \quad (4.3.17)$$

Cette limite existe toujours, elle est positive, mais elle peut être infinie. On peut montrer qu'elle ne dépend pas de la suite $(X_n)_n$ approximante choisie, pourvu que celle-ci soit composée de variables aléatoires positives étagées croissant vers X .

Considérons maintenant une variable aléatoire X de signe quelconque. Celle-ci s'écrit $X = X^+ - X^-$, où $X^+ = \sup(X, 0)$ et $X^- = \sup(-X, 0)$ sont les parties positive et négative de X , de sorte que $|X| = X^+ + X^-$, et évidemment, X^+ et X^- sont deux variables aléatoires positives.

Définition 4.3.2 Nous dirons que la variable aléatoire X est **intégrable** si les valeurs $\mathbb{E}(X^+)$ et $\mathbb{E}(X^-)$ sont toutes les deux finies. Dans ce cas, l'espérance mathématique de X est le nombre

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-), \quad \text{noté aussi } \int_{\Omega} X(\omega) \mathbb{P}(d\omega). \quad (4.3.18)$$

Nous appellerons L^1 l'ensemble des variables aléatoires intégrables, que l'on pourra aussi noter $L^1(\Omega, \mathcal{A}, P)$ si l'on souhaite préciser l'espace de probabilité sous-jacent.

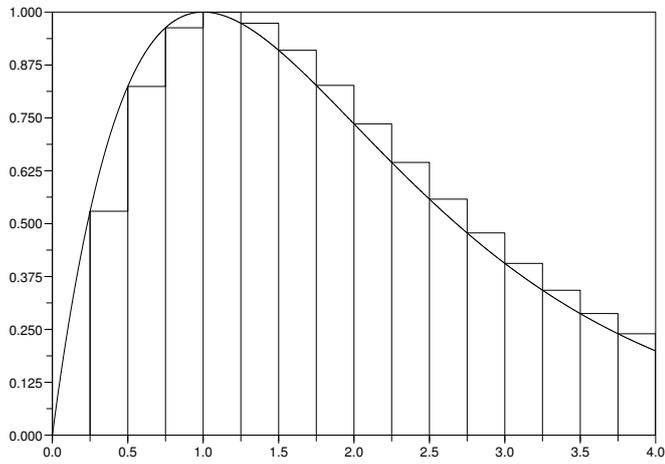


FIGURE 4.3 – Découpage “à la Riemann”, en 16 intervalles de même longueur de l’abscisse

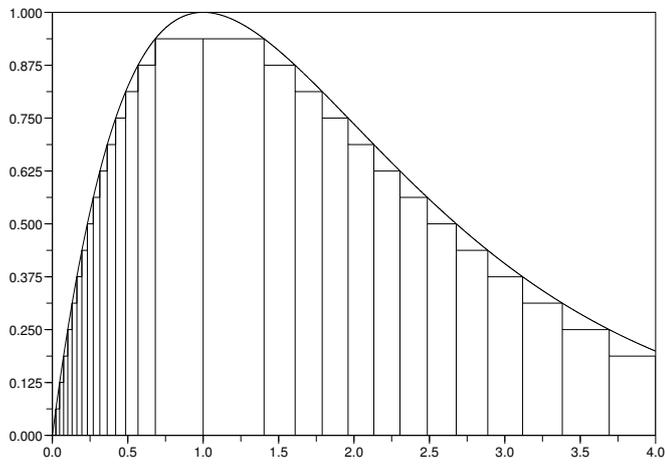


FIGURE 4.4 – Découpage “à la Lebesgue”, en 16 intervalles de même longueur de l’ordonnée

Puisque $|X| = X^+ + X^-$, la proposition suivante est évidente.

Proposition 4.3.3 *Soit X une variable aléatoire. Alors*

$$X \in L^1 \iff \mathbb{E}(|X|) < +\infty.$$

Nous pouvons vérifier sans peine que cette définition coïncide avec celle du paragraphe 3.3 lorsque Ω est fini ou dénombrable. Nous pouvons aussi montrer, (par passage à la limite), que les propriétés données dans la proposition 3.3.5 restent vraies dans notre cadre général.

Nous les rappelons ici :

- **Linéarité** : L^1 est un espace vectoriel, et $\forall X, Y \in L^1, \forall a, b \in \mathbb{R}$,

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y). \quad (4.3.19)$$

- $X \in L^1 \iff |X| \in L^1$, et dans ce cas,

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|). \quad (4.3.20)$$

- **Positivité** :

$$\text{si } X \geq 0 \text{ et } X \in L^1, \text{ alors } \mathbb{E}(X) \geq 0. \quad (4.3.21)$$

Remarquons de plus que si $X \geq 0$ et $\mathbb{E}(X) = 0$, alors $X = 0$ presque-sûrement.

- Si $X, Y \in L^1$ sont telles que $X \leq Y$, alors

$$\mathbb{E}(X) \leq \mathbb{E}(Y). \quad (4.3.22)$$

- L'espérance d'une variable constante (ou en fait presque-sûrement constante) est égale à cette constante :

$$\text{Si } X(\omega) = a \text{ pour tout } \omega, \text{ alors } \mathbb{E}(X) = a. \quad (4.3.23)$$

- S'il existe un réel b tel que $|X(\omega)| \leq b$ pour tout ω , alors $X \in L^1$ et $\mathbb{E}(X) \leq b$.

Remarque : Cette notion d'intégrale $\int X(\omega)\mathbb{P}(d\omega)$, définie sur un espace abstrait muni d'une mesure de probabilité, s'appelle **l'intégrale de Lebesgue** et fait l'objet d'une importante théorie mathématique. Cette notation et cette terminologie d'intégrale sont justifiées par le fait que les propriétés qui étaient vraies pour l'intégrale de Riemann (linéarité, positivité) sont encore vraies dans ce cadre.

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 4.4.1 Nous dirons qu'une variable aléatoire 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 4.4.2 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 3.3.7.

Définition 4.4.3 Si $X \in L^2$, sa **variance** est l'espérance de la variable aléatoire $(X - \mathbb{E}(X))^2$, et on la note $\text{Var}(X)$ ou σ_X^2 , ou σ^2 , s'il n'y a pas d'ambiguïté. Elle est positive, et sa racine carrée positive σ_X s'appelle l'**écart-type** de X .

Nous avons évidemment encore

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (4.4.24)$$

“La variance est égale à la moyenne des carrés moins le carré de la moyenne”.

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).

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 4.4.4 Si X et Y sont dans L^2 , la variable aléatoire $(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))$ est intégrable. On appelle **covariance de X et Y** l'espérance de cette variable aléatoire, et on la note $\text{Cov}(X, Y)$:

$$\text{Cov}(X, Y) = E((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))). \quad (4.4.25)$$

Le **coefficient de corrélation** des variables aléatoires X et Y est le nombre

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \quad (4.4.26)$$

qui est bien défini lorsque $\text{Var}(X) > 0$ et $\text{Var}(Y) > 0$.

Notons que, en utilisant la linéarité de l'espérance, nous obtenons

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y), \quad (4.4.27)$$

et d'ailleurs (4.4.24) est un cas particulier de (4.4.27), car $\text{Var}(X) = \text{Cov}(X, X)$. La linéarité de l'espérance donne immédiatement pour $a, b, a', b' \in \mathbb{R}$:

$$\mathbb{E}((aX + b)(a'Y + b')) = aa' \mathbb{E}(XY) + ab' \mathbb{E}(X) + a'b \mathbb{E}(Y) + bb',$$

$$\mathbb{E}(aX + b) \mathbb{E}(a'Y + b') = aa' \mathbb{E}(X)\mathbb{E}(Y) + ab' \mathbb{E}(X) + a'b \mathbb{E}(Y) + bb'.$$

Donc au vu de (4.4.27), la covariance est une forme bilinéaire sur l'espace vectoriel des variables aléatoires de carré intégrable, et nous avons

$$\text{Cov}(aX + b, a'Y + b') = aa' \text{Cov}(X, Y). \quad (4.4.28)$$

En particulier,

$$\text{Var}(aX + b) = a^2 \text{Var}(X). \quad (4.4.29)$$

Nous en déduisons que les coefficients de corrélation de X et Y et de $aX + b$ et $a'Y + b'$ sont égaux lorsque $aa' > 0$.

Remarque 4.4.5 En vertu des règles énoncées ci-dessus, et si X est une variable aléatoire de carré intégrable, d'espérance $\mathbb{E}(X)$ et d'écart-type $\sigma_X > 0$, alors la variable aléatoire

$$\frac{X - \mathbb{E}(X)}{\sigma_X}$$

est d'espérance nulle et d'écart-type 1. Nous dirons que cette variable aléatoire est **centrée et réduite**.

Proposition 4.4.6 Soient X_1, \dots, X_n des variables aléatoires de carré intégrable, alors il en est de même de la somme, et

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j). \quad (4.4.30)$$

Le calcul est immédiat, grâce aux formules ci-dessus.

4.4.2 Approximation linéaire

Nous considérons deux variables aléatoires de carré intégrable dont nous supposons que les variances et la covariance sont connues. Nous souhaitons trouver la meilleure approximation de Y par une fonction affine de X de la forme $aX + b$, au sens des moindres carrés, c'est à dire qui minimise la quantité $\mathbb{E}((Y - (aX + b))^2)$. Il s'agit de déterminer les constantes a et b telles que $\mathbb{E}((Y - (aX + b))^2)$ soit minimale. Or, par linéarité,

$$\mathbb{E}((Y - (aX + b))^2) = \mathbb{E}(Y^2) - 2a\mathbb{E}(YX) - 2b\mathbb{E}(Y) + a^2\mathbb{E}(X^2) + 2ab\mathbb{E}(X) + b^2.$$

L'annulation des dérivées partielles par rapport à a et b entraîne que les solutions a et b sont

$$\begin{aligned} a &= \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \rho(X, Y) \frac{\sigma_Y}{\sigma_X}; \\ b &= \mathbb{E}(Y) - a\mathbb{E}(X). \end{aligned} \tag{4.4.31}$$

Nous vérifions aisément que ces valeurs donnent bien un minimum pour $\mathbb{E}((Y - (aX + b))^2)$, et déterminent ainsi la meilleure approximation linéaire de Y basée sur X au sens de la distance quadratique moyenne.

Cette approximation linéaire vaut

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

Le carré moyen de l'erreur vaut alors

$$\begin{aligned} E \left(\left(Y - \mathbb{E}(Y) - \rho(X, Y) \frac{\sigma_Y}{\sigma_X} (X - \mathbb{E}(X)) \right)^2 \right) &= \sigma_Y^2 + \rho^2(X, Y) \sigma_Y^2 - 2\rho^2(X, Y) \sigma_Y^2 \\ &= \sigma_Y^2 (1 - \rho^2(X, Y)). \end{aligned}$$

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.

Définition 4.4.7 La droite d'équation $y = ax + b$, où les coefficients a et b sont donnés par les formules (4.4.31) s'appelle **la droite des moindres carrés de Y sur X** .

4.5 Calcul de l'espérance pour une variable aléatoire à densité

4.5.1 Un résultat général fondamental

Le résultat suivant est l'analogue du théorème 3.3.10. Nous considérons une variable aléatoire X de Ω dans \mathbb{R} , de loi \mathbb{P}_X , et une fonction mesurable h de \mathbb{R} dans \mathbb{R} . Nous avons vu que $Y = h(X)$ est une variable aléatoire sur Ω .

Nous dirons que h est intégrable si h appartient à $L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$.

Proposition 4.5.1 *Sous les hypothèses précédentes, h est intégrable si et seulement si $h(X)$ appartient à $L^1(\Omega, \mathcal{A}, \mathbb{P})$, et dans ce cas, les espérances, respectivement de h relativement à \mathbb{P}_X , et de $h(X)$ relativement à \mathbb{P} sont égales. Nous avons*

$$\mathbb{E}(h(X)) = \int_{\Omega} h(X(\omega))\mathbb{P}(d\omega) = \int_{\mathbb{R}} h(x)\mathbb{P}_X(dx). \quad (4.5.32)$$

Preuve. (4.5.32) est évident lorsque h ne prend qu'un nombre fini de valeurs, grâce à (4.3.15) et au fait que $\mathbb{P}(X^{-1}(B)) = \mathbb{P}_X(B)$, par définition de la loi de X . Si h est une fonction positive, nous l'approchons par une suite de fonctions étagées comme en (4.3.16) et nous en déduisons encore (4.5.32), les trois membres étant simultanément finis ou infinis. Le résultat général s'en déduit par différence. \square

Commentaire : Si nous connaissons la loi de X , nous pourrions donc calculer $\mathbb{E}(h(X))$ pour toute fonction h intégrable.

Par exemple, si X est une variable aléatoire discrète, chargeant chaque entier k avec probabilité p_k , nous retrouvons que

$$\int_{\mathbb{R}} h(x)\mathbb{P}_X(dx) = \sum_k h(k)p_k.$$

4.5.2 Calculs d'espérances dans le cas avec densité

Nous allons voir dans ce paragraphe comment calculer l'espérance d'une variable aléatoire réelle, ou plus généralement d'une fonction d'une variable aléatoire réelle,

lorsque cette dernière admet une densité. Nous rappelons que X admet la densité f si la fonction de répartition de \mathbb{P}_X est $x \rightarrow \int_{-\infty}^x f(y)dy$.

Proposition 4.5.2 *Soit X une variable aléatoire réelle admettant la densité f , et soit g une fonction mesurable de \mathbb{R} dans \mathbb{R} . Alors $g(X) \in L^1$ si et seulement si*

$$\int_{-\infty}^{+\infty} |g(x)|f(x)dx < \infty, \quad (4.5.33)$$

et dans ce cas,

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x)dx. \quad (4.5.34)$$

En particulier,

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)dx, \quad (4.5.35)$$

et

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left(\int_{-\infty}^{+\infty} xf(x)dx \right)^2 \quad (4.5.36)$$

Nous n'avons pas les éléments permettant de démontrer ce résultat, mais l'argument "heuristique" suivant permet de comprendre pourquoi il est vrai : supposons f et g continues. Posons $X_n = i/n$ si $i/n \leq X < (i+1)/n$, pour $i \in \mathbb{Z}$. Ainsi, $X_n(\omega) \rightarrow X(\omega)$ pour tout ω , et par continuité de g , nous avons $g(X_n) \rightarrow g(X)$. De plus, comme X_n est une variable aléatoire discrète, nous avons en utilisant (4.3.15) et (4.2.12) :

$$\mathbb{E}(g(X_n)) = \sum_{i \in \mathbb{Z}} g\left(\frac{i}{n}\right) \mathbb{P}\left(\frac{i}{n} \leq X < \frac{i+1}{n}\right) \approx \sum_{i \in \mathbb{Z}} g\left(\frac{i}{n}\right) f\left(\frac{i}{n}\right) \frac{1}{n},$$

et le dernier terme ci-dessus tend vers le second membre de (4.5.34) lorsque $n \rightarrow \infty$, par le théorème d'approximation de Riemann.

4.6 Exemples fondamentaux de variables à densité

4.6.1 Variable aléatoire uniforme sur $[a, b]$

Nous généralisons ici la notion de variable uniforme sur $[0, 1]$. Nous nous donnons deux réels $a < b$, et la probabilité \mathbb{P}_X admet la densité

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{sinon.} \end{cases} \quad (4.6.37)$$

En vertu de la proposition 4.2.10, nous pourrions aussi bien choisir $f(a) = 0$ ou $f(b) = 0$. Au vu de l'interprétation (4.2.12), 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 **nulle** de tomber exactement en un point x fixé à l'avance.

Proposition 4.6.1 Soit X une variable aléatoire de loi uniforme sur $[a, b]$. Alors X est de carré intégrable, et

$$\mathbb{E}(X) = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2} \quad ; \quad \text{Var}(X) = \frac{(b-a)^2}{12}. \quad (4.6.38)$$

(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.)

Les calculs sont immédiats.

Exemple 4.6.2 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. C'est une variable aléatoire 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 minutes est donc

$$\mathbb{P}(10 < X < 15) + \mathbb{P}(25 < X < 30) = \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = \frac{1}{3}.$$

De même, la probabilité d'attendre plus de 10 minutes vaut

$$\mathbb{P}(0 < X < 5) + \mathbb{P}(15 < X < 20) = \frac{1}{3}.$$

4.6.2 Variable aléatoire exponentielle

Définition 4.6.3 Nous dirons que X suit la loi exponentielle de paramètre $\lambda > 0$ si \mathbb{P}_X admet la loi de densité

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ \lambda e^{-\lambda x} & \text{sinon.} \end{cases} \quad (4.6.39)$$

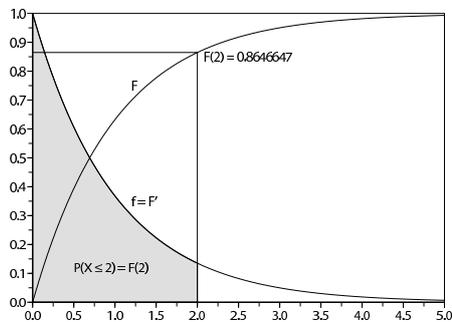


FIGURE 4.5 – Graphe de la densité $f(x) = e^{-x}$ et de la fonction de répartition associée $F(x) = 1 - e^{-x}$.

Proposition 4.6.4 *Si X suit la loi exponentielle de paramètre $\lambda > 0$, alors sa fonction de répartition vaut*

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-\lambda x} & \text{sinon.} \end{cases} \quad (4.6.40)$$

De plus,

$$\mathbb{E}(X) = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}. \quad (4.6.41)$$

Preuve. Les calculs sont très simples. Par exemple, $\mathbb{E}(X) = \int_0^\infty x \lambda e^{-\lambda x} dx$ et $\mathbb{E}(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx$, et $\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. Nous obtenons le résultat par intégrations par parties successives. \square

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 variables aléatoires de loi exponentielle.

Exemple 4.6.5 Supposons que la durée d'une conversation téléphonique mesurée en minutes soit une variable aléatoire exponentielle de paramètre $\lambda = \frac{1}{10}$ (exprimé en minute^{-1}). 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.

$$\mathbb{P}(X > 10) = \int_{10}^{+\infty} \frac{1}{10} e^{-\frac{x}{10}} dx \approx 0.368$$

et

$$\mathbb{P}(10 < X < 20) \approx 0,233.$$

La loi exponentielle possède une propriété importante pour les applications : elle est **sans mémoire**. On dit aussi qu'elle satisfait la propriété de non-vieillessement.

Proposition 4.6.6 (*Propriété de “non-vieillessement”*) Soit X une variable aléatoire positive, telle que $\mathbb{P}(X > s) > 0$ pour tout $s \in \mathbb{R}$. Alors

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s)$$

pour tous $s, t > 0$ si et seulement si X suit une loi exponentielle.

Preuve. Soit $G(t) = \mathbb{P}(X > t) = 1 - F(t)$, où F est la fonction de répartition de X . D'après (2.5.27), 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 $\lambda > 0$. Le résultat s'obtient alors en comparant à (4.6.40) 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.

Simulation d'une variable aléatoire X de loi exponentielle de paramètre $\lambda > 0$:

Sa densité est donnée par (4.6.39). Nous savons que $F(x) = 1 - e^{-\lambda x}$ pour $x \geq 0$, et $F(x) = 0$ pour $x < 0$. Alors,

$$F^{-1}(u) = -(1/\lambda) \ln(1 - u).$$

Nous avons vu au paragraphe 4.2.4 que si U est une variable aléatoire uniforme, la variable aléatoire $X' = -(1/\lambda) \ln(1 - U)$ suit une loi exponentielle de paramètre λ . Mais nous lui préférons

$$X = -\frac{1}{\lambda} \ln U,$$

qui suit la même loi que X' (car U et $1 - U$ ont même loi), et qui est moins coûteuse en temps de calcul.

4.6.3 Variable aléatoire de loi gamma

Rappelons tout d'abord que la fonction gamma est définie pour $\alpha \in]0, +\infty[$ par

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx. \quad (4.6.42)$$

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 $0! = 1$.

Soit $\theta > 0$ et $\alpha > 0$.

Définition 4.6.7 Une variable aléatoire X suit une loi gamma de paramètre d'échelle θ et d'indice α , que l'on note $\Gamma(\alpha, \theta)$, si sa loi admet la densité

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{1}{\Gamma(\alpha)} \theta^\alpha x^{\alpha-1} e^{-\theta x} & \text{sinon.} \end{cases} \quad (4.6.43)$$

Le changement de variable $x \mapsto \theta x$ montre que cette fonction est d'intégrale égale à 1.

Nous remarquons immédiatement que $\Gamma(1, \theta)$ est la loi exponentielle de paramètre θ .

Proposition 4.6.8 Si X est de loi $\Gamma(\alpha, \theta)$, son espérance $\mathbb{E}(X)$, sa variance σ_X^2 , et l'espérance $\mathbb{E}(X^\beta)$ pour tout $\beta > -\alpha$, sont données par

$$\mathbb{E}(X) = \frac{\alpha}{\theta}, \quad \sigma_X^2 = \frac{\alpha}{\theta^2}, \quad \mathbb{E}(X^\beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} \frac{1}{\theta^\beta}. \quad (4.6.44)$$

Cette proposition se montre immédiatement en utilisant le fait que $\mathbb{E}(X^\beta) = \int_0^\infty x^\beta f(x) dx$ et (4.6.42). En revanche si $\beta \leq -\alpha$, nous avons $\mathbb{E}(X^\beta) = +\infty$.

Lorsque $\alpha = n$, la loi gamma représente la loi du temps d'attente avant la n -ième occurrence d'événements indépendants de lois exponentielles de paramètre θ (voir Exercice 6.1.11).

4.6.4 Variables aléatoires normales (ou variables gaussiennes)

Nous introduisons ici les variables aléatoires les plus célèbres en probabilité.

Définition 4.6.9 Nous appelons **variable aléatoire normale centrée réduite** une variable aléatoire X de loi de densité

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (4.6.45)$$

Pour vérifier que la fonction introduite en (4.6.45) est bien d'intégrale 1, nous remarquons que $I = \int_{-\infty}^{+\infty} f(x) dx$ vérifie

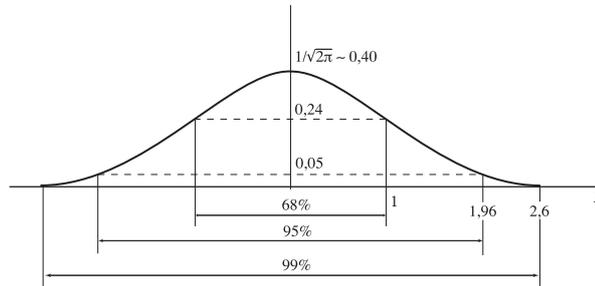
$$I^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x)f(y) dx dy = \int_0^{2\pi} d\theta \int_0^\infty \frac{1}{2\pi} e^{-\rho^2/2} \rho d\rho,$$

(en passant en coordonnées polaires dans l'intégrale double). Un calcul simple montre alors que $I^2 = 1$.

Proposition 4.6.10 Soit X une variable normale centrée réduite. Alors

$$\mathbb{E}(X) = 0, \quad \text{Var}(X) = 1. \quad (4.6.46)$$

La loi de X est dans ce cas notée $\mathcal{N}(0, 1)$, et les qualificatifs “centrée” et “réduite” proviennent précisément de (4.6.46).

FIGURE 4.6 – La courbe de Gauss f et les proportions d’aire sous la courbe

Cette proposition est immédiate. $\mathbb{E}(X) = 0$ car c’est l’intégrale d’une fonction impaire ; pour calculer $\text{Var}(X)$ nous pouvons faire une intégration par parties.

Définition 4.6.11 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

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x - m)^2}{2\sigma^2}. \quad (4.6.47)$$

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 (4.6.45). Le même changement de variable permet de voir que m et σ^2 sont l’espérance et la variance d’une variable aléatoire de loi $\mathcal{N}(m, \sigma^2)$.

La distribution normale fut introduite par De Moivre en 1733. Celui-ci l’utilisa pour approximer une variable aléatoire binomiale quand le paramètre n de celle-ci était 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 au Chapitre 6. 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 définie en (4.6.46) 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 ci-dessous donne les valeurs de la fonction de répartition Φ de la loi normale centrée réduite, pour certaines valeurs x . 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 aurons

$$\mathbb{P}(|X| < 2) = 2\mathbb{P}(0 < X < 2) = 2\left(\mathbb{P}(X < 2) - \frac{1}{2}\right) = 2(0,9772 - 0,5) = 0,9544.$$

Exemple 4.6.12 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 : Remarquons que $\frac{X-m}{\sigma} = \frac{X-270}{10}$ suit une loi $\mathcal{N}(0, 1)$. Nous allons utiliser la table 4.1.

$$\begin{aligned} \mathbb{P}(X > 290 \text{ ou } X < 240) &= \mathbb{P}\left(\frac{X - 270}{10} > 2\right) + \mathbb{P}\left(\frac{X - 270}{10} < -3\right) \\ &= 1 - \Phi(2) + 1 - \Phi(3) = 0,02411. \end{aligned}$$

La table 4.1 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.

Remarque 4.6.13 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 .

Il existe des variables aléatoires qui n'ont pas d'espérance, comme le montre l'exemple suivant.

Exemple 4.6.14 Variable aléatoire de Cauchy. Un gyrophare G envoie un flash lumineux dans une direction aléatoire uniforme d'angle θ . Cherchons la distribution

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 4.1 – 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 $\Phi(x) = 1 - \Phi(-x)$.

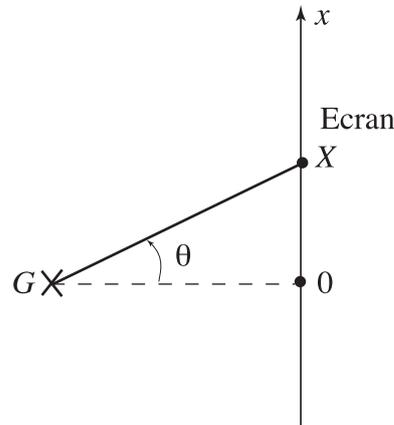


FIGURE 4.7 – Gyrophare

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 4.7.

L'angle θ est une variable aléatoire 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 variable aléatoire 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 a pour dérivée

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}. \quad (4.6.48)$$

Par conséquent, la variable aléatoire X a la loi de densité f .

Définition 4.6.15 Une variable aléatoire réelle X suit une **loi de Cauchy** si elle admet la densité

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Voyons si X est intégrable. Pour cela il faut étudier la convergence de l'intégrale généralisée

$$\int_{-\infty}^{+\infty} \frac{|x|}{\pi(1+x^2)} dx.$$

Cette intégrale diverge (car $f(x) \sim_{x \rightarrow \infty} \frac{1}{|x|}$). Ainsi, la variable aléatoire X n'a pas d'espérance. Une variable de Cauchy n'est donc pas intégrable.

4.7 Des inégalités fameuses

4.7.1 Inégalité de Bienaymé-Chebyshev

Théorème 4.7.1 Soit $p \geq 1$ et $a > 0$. Nous avons les inégalités suivantes

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

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(|X|^p)}{a^p}. \quad (4.7.49)$$

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

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}. \quad (4.7.50)$$

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 4.7.2 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 variable aléatoire 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 (4.7.50), comme nous allons nous en convaincre ci-dessous.

Exemple 4.7.3 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

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

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. Soit X une variable normale de loi $\mathcal{N}(m, \sigma)$. En utilisant la table numérique de la loi normale donnée ci-dessus, nous obtenons

$$\mathbb{P}(X \in]m - 2\sigma, m + 2\sigma]) \approx 0,95 \gg 1 - \frac{1}{4} ; \quad \mathbb{P}(X \in]m - 3\sigma, m + 3\sigma]) \approx 0,997 \gg 1 - \frac{1}{9}.$$

Les renseignements ainsi obtenus sont beaucoup plus précis.

4.7.2 Inégalité de Cauchy-Schwarz

Proposition 4.7.4 *Soient X et Y deux variables aléatoires de carré intégrable, alors $XY \in L^1$, et on a l'inégalité de Cauchy-Schwarz :*

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}. \quad (4.7.51)$$

Il y a égalité dans (4.7.51) si et seulement si les deux variables aléatoires sont presque-sûrement 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 (4.7.51).

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 ω . \square

Nous déduisons en particulier de l'inégalité de Cauchy-Schwarz, que le coefficient de corrélation défini par (4.4.26) vérifie

$$-1 \leq \rho(X, Y) \leq 1. \quad (4.7.52)$$

4.7.3 Inégalité de Jensen

Théorème 4.7.5 *Soit X une variable aléatoire réelle intégrable, et f une fonction mesurable telle que $f(X)$ soit intégrable. Supposons de plus que f soit une fonction convexe. Alors*

$$\mathbb{E}(f(X)) \geq f(\mathbb{E}(X)).$$

Preuve. Puisque f est convexe, nous savons que pour tout $a \in \mathbb{R}$, il existe $\lambda_a \in \mathbb{R}$ tel que pour tout x ,

$$f(x) \geq f(a) + \lambda_a(x - a).$$

Ce résultat est en particulier vrai pour $x = X(\omega)$ et $a = \mathbb{E}(X)$. Il suffit alors de prendre l'espérance pour obtenir le résultat. \square

Applications :

- Si X est une v.a. centrée ($\mathbb{E}(X) = 0$), alors pour tout $\lambda \in \mathbb{R}$, nous avons

$$\mathbb{E}(e^{\lambda X}) \geq 1.$$

- Soit G une variable aléatoire de loi normale centrée réduite. Alors pour tous $x, \sigma, K \in \mathbb{R}$, nous avons

$$\mathbb{E}((xe^{\sigma G - \sigma^2/2} - K)^+) \geq (x - K)^+.$$

car $\mathbb{E}(e^{\sigma G - \sigma^2/2}) = 1$ et la fonction $u \rightarrow (xu - K)^+ = \sup(0, xu - K)$ est convexe.

Exemple 4.7.6 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 risques 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 $\mathbb{E}(u(X)) \geq u(m)$.

4.8 Vecteurs aléatoires

Nous allons généraliser ici la notion de variable aléatoire en considérant maintenant qu'elle peut prendre ses valeurs dans \mathbb{R}^n . De même qu'en dimension 1, la tribu borélienne sur \mathbb{R}^n est la tribu engendrée par les ensembles de la forme $\prod_{i=1}^n]-\infty, x_i]$, pour $x_i \in \mathbb{R}$. Nous munirons dorénavant \mathbb{R}^n de cette tribu. Nous pourrons alors généraliser la notion de fonction mesurable de \mathbb{R}^n dans \mathbb{R} de manière évidente et considérer les intégrales généralisées de telles fonctions. Une fonction mesurable f de \mathbb{R}^n dans \mathbb{R} (par exemple une fonction continue ou continue par morceaux), sera dite intégrable si

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} |f(x_1, \dots, x_n)| dx_1 \cdots dx_n < +\infty.$$

Nous rappelons le théorème de Fubini, essentiel dans les calculs d'intégrales multiples.

Théorème 4.8.1 Soit f une fonction mesurable sur \mathbb{R}^2 .

1) (Théorème de Tonelli) Si f est positive, alors

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy. \quad (4.8.53)$$

2) (Théorème de Fubini) Si f est de signe quelconque mais vérifie que $\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x, y)| dy \right) dx < +\infty$ ou que $\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x, y)| dx \right) dy < +\infty$, alors (4.8.53) est encore vraie et on peut définir $\int_{\mathbb{R} \times \mathbb{R}} f(x, y) dx dy$ comme la valeur commune :

$$\int_{\mathbb{R} \times \mathbb{R}} f(x, y) dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy. \quad (4.8.54)$$

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 4.8.2 Considérons une suite de fonctions $(f_k)_k$ telle que $\sum_k \int_{\mathbb{R}} |f_k(x)| dx < +\infty$, alors

$$\sum_k \int_{\mathbb{R}} f_k(x) dx = \int_{\mathbb{R}} \left(\sum_k f_k(x) \right) dx.$$

Le Corollaire 4.2.6 se généralise alors de la manière suivante.

Théorème 4.8.3 Pour tout entier $n \geq 1$ fixé, il existe une unique fonction λ définie sur la tribu borélienne de \mathbb{R}^n et à valeurs dans \mathbb{R}_+ , qui soit σ -additive, et qui coïncide avec le volume sur les pavés :

$$\lambda(\Pi_{i=1}^n]a_i, b_i]) = \Pi_{i=1}^n (b_i - a_i),$$

où $a_i < b_i$, $i \in \{1, \dots, n\}$. Cette mesure λ s'appelle la **mesure de Lebesgue** sur \mathbb{R}^n .

Nous pouvons alors généraliser la notion d'équiprobabilité vue dans le cas d'un ensemble fini ou sur $[a, b]$. La probabilité uniforme sur un ensemble borélien V de \mathbb{R}^n de mesure de Lebesgue $\lambda(V) \in]0, +\infty[$, est définie pour tout borélien $A \subset V$ par

$$P_V(A) = \frac{\lambda(A)}{\lambda(V)}.$$

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 $V = [a, b]$.

4.8.1 Vecteurs aléatoires

Définition 4.8.4 Un vecteur aléatoire \mathbf{X} à valeurs dans \mathbb{R}^n est formé de n variables aléatoires réelles, qui sont les composantes de \mathbf{X} : $\mathbf{X} = (X_1, \dots, X_n)$.

Comme dans le cas de la dimension 1, sa loi est caractérisée par la fonction de répartition multidimensionnelle $F : \mathbb{R}^n \rightarrow [0, 1]$ définie par

$$F(x_1, \dots, x_n) = \mathbb{P} \left(\mathbf{X} \in \prod_{i=1}^n]-\infty, x_i] \right). \quad (4.8.55)$$

De manière générale, caractériser les fonctions de répartition sur \mathbb{R}^n est assez délicat, de sorte que cette notion est rarement utilisée. Nous allons plus particulièrement nous intéresser aux vecteurs aléatoires à densité.

Définition 4.8.5 On dit que \mathbf{X} admet la densité f si la fonction f est positive sur \mathbb{R}^n , intégrable et vérifie

$$\int_{\mathbb{R}^n} f(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1, \quad (4.8.56)$$

et si

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(y_1, \dots, y_n) dy_1 \dots dy_n. \quad (4.8.57)$$

La proposition suivante généralise à la dimension n la proposition 4.5.2.

Proposition 4.8.6 Soit \mathbf{X} une variable aléatoire à valeurs dans \mathbb{R}^n admettant la densité f , et soit g une fonction mesurable de \mathbb{R}^n dans \mathbb{R} . Alors $g(\mathbf{X})$ est intégrable si et seulement si

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} |g(x_1, \dots, x_n)| f(x_1, \dots, x_n) dx_1 \dots dx_n < +\infty. \quad (4.8.58)$$

Dans ce cas nous avons

$$\mathbb{E}(g(\mathbf{X})) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n. \quad (4.8.59)$$

Pour simplifier, nous écrivons aussi $\mathbb{E}(g(\mathbf{X})) = \int_{\mathbb{R}^n} g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$, avec les notations de l'intégration.

4.8.2 Moments d'un vecteur aléatoire

Définition 4.8.7 Si les composantes X_i du vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$ sont intégrables, nous pouvons définir le vecteur espérance $\mathbb{E}(\mathbf{X}) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))$.

Si les composantes X_i du vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$ sont de carré intégrable, la **matrice de covariance** de \mathbf{X} est la matrice $\mathbf{C}_\mathbf{X} = (c_{i,j})_{1 \leq i,j \leq n}$ de taille $n \times n$ et dont les éléments valent

$$c_{i,j} = \text{Cov}(X_i, X_j).$$

Proposition 4.8.8 La matrice des covariances est symétrique non-négative.

Preuve. La symétrie est évidente. “Non-négative” signifie que $\sum_{i,j=1}^n a_i a_j c_{i,j} \geq 0$ pour tous réels a_i . Un calcul simple montre que

$$\sum_{i,j=1}^n a_i a_j c_{i,j} = \text{Var} \left(\sum_{i=1}^n a_i X_i \right) \geq 0.$$

□

Proposition 4.8.9 Soit \mathbf{X} un vecteur aléatoire n -dimensionnel, de matrice de covariance \mathbf{C} . Soit \mathbf{A} une matrice de taille $m \times n$, et \mathbf{Y} le vecteur aléatoire m -dimensionnel $\mathbf{Y} = \mathbf{A}\mathbf{X}$. La matrice de covariance de \mathbf{Y} est alors $\mathbf{A}\mathbf{C}\mathbf{A}^t$, où \mathbf{A}^t est la transposée de \mathbf{A} .

La preuve provient d'un calcul immédiat.

Exemple 4.8.10 Vecteur gaussien n -dimensionnel.

Un exemple de vecteurs aléatoires est celui des vecteurs gaussiens, que nous étudierons en détail au chapitre 6. Soient $\mathbf{m} \in \mathbb{R}^n$ et \mathbf{C} une matrice symétrique définie positive. Le vecteur $\mathbf{X} \in \mathbb{R}^n$ est un vecteur aléatoire gaussien d'espérance \mathbf{m} et de matrice de covariance \mathbf{C} si sa densité s'écrit

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\mathbf{C})}} e^{-\frac{1}{2} \langle \mathbf{x} - \mathbf{m}, \mathbf{C}^{-1}(\mathbf{x} - \mathbf{m}) \rangle},$$

où $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$.

On a alors $\mathbb{E}(\mathbf{X}) = \mathbf{m}$ et $\mathbf{C}_\mathbf{X} = \mathbf{C}$.

La densité gaussienne f a la forme dessinée sur la figure 4.8.2 pour $n = 2$.

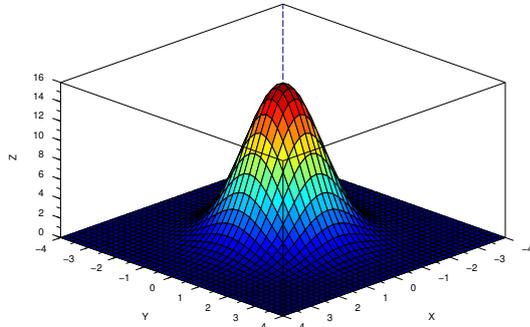


FIGURE 4.8 – Densité gaussienne en dimension deux.

4.8.3 Densités marginales et conditionnelles

Nous allons nous limiter dans la suite de ce paragraphe au cas où $n = 2$, pour simplifier les notations. Nous considérons une variable 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 4.8.11 *Supposons que \mathbf{Z} admette une densité f . Alors, X et Y admettent les densités f_X et f_Y suivantes sur \mathbb{R} :*

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx. \quad (4.8.60)$$

Les fonctions f_X et f_Y s'appellent les **densités marginales** de f . Notons que la réciproque de cette proposition est fautive : les variables aléatoires X et Y peuvent avoir des densités sans que le couple $\mathbf{Z} = (X, Y)$ en ait une.

Supposons par exemple que $X = Y$. Si $\Delta = \{(x, x) : x \in \mathbb{R}\}$ est la diagonale de \mathbb{R}^2 , nous avons évidemment $\mathbb{P}_{\mathbf{Z}}(\Delta) = 1$, tandis que si la formule (4.8.59) était valide pour $\mathbb{P}_{\mathbf{Z}}$, nous aurions $\mathbb{P}_{\mathbf{Z}}(\Delta) = \mathbb{E}(1_{\Delta}(\mathbf{Z})) = \int_{\mathbb{R}^2} 1_{\Delta}(z) f(z) dz = 0$ car Δ a un volume nul dans \mathbb{R}^2 .

En particulier, il faut faire attention au fait que dans le cas général, la densité d'un couple de variables aléatoires à densité n'est pas le produit des densités.

Preuve. Pour tout $x \in \mathbb{R}$, nous avons par (4.8.57) :

$$\mathbb{P}(X \leq x) = \mathbb{P}(\mathbf{Z} \in]-\infty, x] \times \mathbb{R}) = \int_{-\infty}^x du \left(\int_{-\infty}^{+\infty} f(u, v) dv \right).$$

Donc si f_X est définie par (4.8.60), nous obtenons $\mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(u) du$, ce qui montre que f_X est la densité de X . Le raisonnement est analogue pour Y . \square

Exemple 4.8.12 On lance une fléchette sur une cible circulaire de rayon unité. Le joueur est suffisamment loin de la cible pour que le point M d'impact de la fléchette soit supposé uniformément distribué sur la cible. (On décide de n'observer que les lancers qui atteignent 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 variables aléatoires de densité

$$f_{(X,Y)}(x, y) = \frac{1}{\pi} \mathbf{1}_{\{x^2 + y^2 \leq 1\}}$$

uniforme sur le disque, par hypothèse. L'abscisse X est distribuée selon la densité marginale

$$f_X(x) = \int f_{(X,Y)}(x, y) dy = \frac{2}{\pi} (1 - x^2)^{\frac{1}{2}} \mathbf{1}_{[-1,1]}(x).$$

La loi de Y a la même densité.

Donnons maintenant dans ce cadre une définition de la loi conditionnelle de Y sachant $X = x$, comme nous l'avons fait dans le cas discret (Section 3.6.1).

Proposition 4.8.13 *La formule suivante définit une densité sur \mathbb{R} , pour tout x tel que $f_X(x) > 0$:*

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}. \quad (4.8.61)$$

*Cette fonction s'appelle **densité conditionnelle** de Y sachant $X = x$.*

La preuve est immédiate puisque $f_{Y|X=x}$ est une fonction positive d'intégrale 1.

L'interprétation de (4.8.61) 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 : Δx et Δy étant de "petits" accroissements des variables x et y , nous avons comme en (4.2.12), 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

$$\begin{aligned} f_{Y|X=x}(y)\Delta y &\approx \frac{\mathbb{P}(x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y)}{\mathbb{P}(x \leq X \leq x + \Delta x)} \\ &\approx \mathbb{P}(y \leq Y \leq y + \Delta y | x \leq X \leq x + \Delta x). \end{aligned}$$

Puisque $f_{Y|X=x}$ est une densité, nous pouvons en particulier définir l'espérance qui lui est associée, et généraliser ainsi la notion d'espérance conditionnelle vue dans le cas discret, dans le cas où Y est intégrable.

Définition 4.8.14 Soit Y une variable aléatoire intégrable.

1) L'espérance conditionnelle de Y sachant $X = x$ est définie par

$$\mathbb{E}(Y|X = x) = \int_{\mathbb{R}} y f_{Y|X=x}(y) dy. \quad (4.8.62)$$

2) L'espérance conditionnelle de Y sachant X est la variable aléatoire définie par :

$$\mathbb{E}(Y|X) = \psi(X), \quad \text{avec} \quad \psi(x) = \mathbb{E}(Y|X = x). \quad (4.8.63)$$

On peut étendre cette définition à toute variable aléatoire de la forme $h(X, Y)$.

Définition 4.8.15 Soit Y une variable aléatoire et h une fonction mesurable positive ou bornée sur \mathbb{R}^2 .

1) L'espérance conditionnelle de $h(X, Y)$ sachant $X = x$ est définie par

$$\mathbb{E}(h(X, Y)|X = x) = \int_{\mathbb{R}} h(x, y) f_{Y|X=x}(y) dy. \quad (4.8.64)$$

2) L'espérance conditionnelle de $h(X, Y)$ sachant X est la variable aléatoire définie par :

$$\mathbb{E}(h(X, Y)|X) = \psi(X), \quad \text{avec} \quad \psi(x) = \mathbb{E}(h(X, Y)|X = x). \quad (4.8.65)$$

Remarque 4.8.16

1) $\psi(x)$ n'est défini que pour $x \notin B = \{u, f_X(u) = 0\}$. Mais $\mathbb{P}(X \in B) = \int_B f_X(u) du = 0$. Donc (4.8.63) définit bien l'espérance conditionnelle $\psi(X) = \mathbb{E}(Y | X)$ avec probabilité 1.

2) Comme $\mathbb{E}(\mathbb{E}(|Y| | X)) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |y| \frac{f_{X,Y}(x,y)}{f_X(x)} dy \right) f_X(x) dx = \mathbb{E}(|Y|)$ (nous avons utilisé le Théorème 4.8.1 (théorème de Fubini)), l'espérance conditionnelle de Y sachant X est bien définie dès que Y est intégrable.

3) 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 3.6.1 et obtenir dans le cadre des lois à densité les mêmes propriétés de l'espérance conditionnelle résumées ci-dessous.

Proposition 4.8.17 *Soit Y une variable aléatoire .*

1. *Si Y est intégrable, alors*

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = \int_{\mathbb{R}} \mathbb{E}(Y | X = x) f_X(x) dx.$$

2. *Si X et Z sont des variable aléatoire intégrables, alors, pour $a, b \in \mathbb{R}$,*

$$\mathbb{E}(aY + bZ | X) = a \mathbb{E}(Y | X) + b \mathbb{E}(Z | X).$$

3. *Si $Y \geq 0$, alors $\mathbb{E}(Y | X) \geq 0$.*

4. $\mathbb{E}(1 | X) = 1$.

5. *Si h est une fonction mesurable positive ou bornée sur \mathbb{R}^2 , alors*

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

avec

$$\begin{aligned} \mathbb{E}(h(X, Y)) &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y) f(x, y) dx dy, \\ \mathbb{E}(\mathbb{E}(h(X, Y) | X)) &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y) f_X(x) f_{Y|X=x}(y) dx dy. \end{aligned}$$

6. *Si g est mesurable positive ou bornée, alors*

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

Exemple 4.8.18 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 $\mathbb{E}(Y | X) = \frac{X}{2}$.

4.9 Variables aléatoires indépendantes

4.9.1 Indépendance de deux variables aléatoires

Dans la suite de ce paragraphe, nous considérons un couple (\mathbf{X}, \mathbf{Y}) de vecteurs aléatoires. \mathbf{X} et \mathbf{Y} sont respectivement à valeurs dans \mathbb{R}^m et \mathbb{R}^n .

Définition 4.9.1 Les vecteurs aléatoires \mathbf{X} et \mathbf{Y} sont **indépendants** si pour tous boréliens A et B dans les espaces correspondants,

$$\mathbb{P}(\mathbf{X} \in A, \mathbf{Y} \in B) = \mathbb{P}(\mathbf{X} \in A) \mathbb{P}(\mathbf{Y} \in B). \quad (4.9.66)$$

Considérons alors deux fonctions mesurables g et h définies sur \mathbb{R}^m et \mathbb{R}^n .

Proposition 4.9.2 Avec les notations précédentes, et si \mathbf{X} et \mathbf{Y} sont indépendantes, les variables aléatoires $g(\mathbf{X})$ et $h(\mathbf{Y})$ sont aussi indépendantes. Si de plus $g(\mathbf{X})$ et $h(\mathbf{Y})$ sont dans L^1 , alors le produit $g(\mathbf{X})h(\mathbf{Y})$ est aussi intégrable, et nous avons

$$\mathbb{E}(g(\mathbf{X})h(\mathbf{Y})) = \mathbb{E}(g(\mathbf{X})) \mathbb{E}(h(\mathbf{Y})). \quad (4.9.67)$$

Preuve. La première assertion est évidente par définition même de l'indépendance. Pour le reste, nous remarquons d'abord que (4.9.67) se réduit à (4.9.66) si g et h sont des fonctions indicatrices. Comme les deux membres de (4.9.67) sont linéaires en g et en h , nous avons aussi (4.9.67) lorsque g et h sont des fonctions étagées. D'après (4.3.17), nous en déduisons (4.9.67) pour g et h positives quelconques. Si g et h sont de signe quelconque, (4.9.67) est vérifiée pour $|g|$ et $|h|$, donc si $g(\mathbf{X})$ et $h(\mathbf{Y})$ sont

intégrables, nous en déduisons que $g(\mathbf{X})h(\mathbf{Y}) \in L^1$. Enfin, dans ce cas, par différence (en considérant $g = g^+ - g^-$ et $h = h^+ - h^-$ et en développant le produit), nous obtenons (4.9.67) pour g et h elles-mêmes. \square

Si X et Y sont des variables aléatoires réelles, il découle alors de (4.4.27) et du résultat précédent que :

Proposition 4.9.3 *Si les variables aléatoires X et Y sont indépendantes et dans L^2 , alors $\text{Cov}(X, Y) = 0$ et $\rho(X, Y) = 0$ ($\rho(X, Y)$ est le coefficient de corrélation).*

Nous savons que $|\rho(X, Y)| \leq 1$ par (4.7.52). Si les variables aléatoires X et Y sont indépendantes, le coefficient de corrélation $|\rho(X, Y)|$ est donc nul ; 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.

Le résultat suivant est un analogue de l’équivalence (i) \iff (ii) dans la proposition 3.6.10.

Proposition 4.9.4 *Soient X et Y deux variables aléatoires réelles admettant les densités f_X et f_Y . Pour qu’elles soient indépendantes, il faut et il suffit que le couple $\mathbf{Z} = (X, Y)$ admette (sur \mathbb{R}^2) la densité suivante :*

$$f_{\mathbf{Z}}(x, y) = f_X(x) f_Y(y). \quad (4.9.68)$$

Preuve. S’il y a indépendance, il vient par (4.9.66) que

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \mathbb{P}(Y \leq y) = \int_{-\infty}^x f_X(u) du \int_{-\infty}^y f_Y(v) dv,$$

ce qui montre que $\mathbb{P}_{\mathbf{Z}}$ vérifie (4.8.57) avec f donnée par (4.9.68).

Inversement, supposons que (4.9.68) soit satisfaite. Le même calcul que ci-dessus montre alors que $\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \mathbb{P}(Y \leq y)$ pour tous $x, y \in \mathbb{R}$. Mais, comme dans la preuve de la proposition 4.2.3, nous pouvons montrer que ces relations s’étendent en $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B)$ pour tous boréliens A et B , d’où le résultat. \square

4.9.2 Suite de variables aléatoires indépendantes

Etant donnée une famille finie $\mathbf{X}_1, \dots, \mathbf{X}_n$ de vecteurs aléatoires à valeurs respectivement dans $\mathbb{R}^{p(i)}$, $1 \leq i \leq n$, tout ce qui précède s'étend sans difficulté, sauf que les notations deviennent plus compliquées. La seule chose pouvant prêter à confusion est la notion d'indépendance; nous la définissons donc ci-dessous :

Définition 4.9.5 Les vecteurs aléatoires $\mathbf{X}_1, \dots, \mathbf{X}_n$ sont **indépendants** (ou, "mutuellement indépendants") si pour tous boréliens A_1, \dots, A_n on a

$$\mathbb{P}(\mathbf{X}_1 \in A_1, \dots, \mathbf{X}_n \in A_n) = \prod_{i=1}^n \mathbb{P}(\mathbf{X}_i \in A_i). \quad (4.9.69)$$

La proposition suivante est extrêmement utile dans la pratique.

Proposition 4.9.6 Si les X_j , $1 \leq j \leq n$, sont des variables aléatoires indépendantes et de carré intégrable, les variances vérifient

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i). \quad (4.9.70)$$

Preuve. Nous utilisons ici (4.4.30), à savoir

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i,j=1; i \neq j}^n \text{Cov}(X_i, X_j).$$

Comme les X_i sont indépendantes, $\text{Cov}(X_i, X_j) = 0$ si $i \neq j$, d'où le résultat. \square

Considérons maintenant une suite **infinie** $(\mathbf{X}_n)_{n \in \mathbb{N}^*}$.

Définition 4.9.7 La suite $(\mathbf{X}_n)_{n \in \mathbb{N}^*}$ de variables aléatoires est dite **indépendante** si pour tout n , la famille finie $\mathbf{X}_1, \dots, \mathbf{X}_n$ est indépendante.

Il est facile de vérifier (et intuitif aussi) que l'indépendance est préservée par certaines transformations.

Proposition 4.9.8 *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 de la forme $(f_n(\mathbf{X}_n))_n$, où les fonctions f_n sont des fonctions continues, ou même mesurables.*

Exemple 4.9.9 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 ω , nous associons son développement dyadique (unique si l'on impose que les ω_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 ω associe $X_i(\omega) = \omega_i$ est une variable aléatoire sur Ω . En effet, pour $x_i \in \{0, 1\}$, $1 \leq i \leq n$,

$$\{X_i = x_i\} = \bigcup_{x_1, \dots, x_{i-1} \in \{0, 1\}} \left[\sum_{j=1}^i \frac{x_j}{2^j}, \sum_{j=1}^i \frac{x_j}{2^j} + \frac{1}{2^i} \right],$$

qui est bien un élément de la tribu borélienne de $\Omega = [0, 1[$, et

$$\mathbb{P}(\{X_i = x_i\}) = \frac{1}{2^i} \sum_{x_1, \dots, x_{i-1} \in \{0, 1\}} 1 = \frac{1}{2}.$$

Montrons l'indépendance des variables aléatoires $(X_i)_{1 \leq i \leq n}$. Nous avons

$$\bigcap_{1 \leq i \leq n} \{X_i = x_i\} = \left[\sum_{i=1}^n \frac{x_i}{2^i}, \sum_{i=1}^n \frac{x_i}{2^i} + \frac{1}{2^n} \right],$$

si bien que

$$\mathbb{P} \left(\bigcap_{1 \leq i \leq n} \{X_i = x_i\} \right) = \frac{1}{2^n},$$

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 Ω une suite de variables aléatoires 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 $\frac{1}{2}$.

Nous citerons sans démonstration le théorème suivant répondant, dans un cadre général au problème de la construction d'une suite de variables aléatoires indépendantes de lois données.

Théorème 4.9.10 *Etant donnée, pour chaque entier n , une probabilité μ_n sur $\mathbb{R}^{p(n)}$, où $p(n)$ est un entier non nul, il existe un espace Ω 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, chaque \mathbf{X}_n étant de loi μ_n .*

4.10 Calculs de lois

4.10.1 Un théorème d'identification

Dans ce paragraphe, nous allons utiliser le théorème suivant (que nous ne démontrerons pas), réciproque de (4.5.34).

Théorème 4.10.1 *1) Si il existe une probabilité μ sur \mathbb{R} telle que pour toute fonction continue bornée h ,*

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)\mu(dx), \quad (\text{voir notation en (4.3.18)}), \quad (4.10.71)$$

alors la loi de X est égale à μ .

2) Si il existe une fonction f telle que pour toute fonction continue bornée h ,

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)f(x)dx,$$

alors la loi de X admet la densité f .

L'idée de la preuve repose sur le fait que les fonctions continues bornées peuvent approcher une fonction $h = \mathbf{1}_{]-\infty, y]}$, pour laquelle (4.10.71) donne la fonction de répartition de μ .

Utilisons ce résultat dans l'exemple suivant, qui montre qu'il existe des variables aléatoires réelles qui ne sont ni discrètes, ni à densité. Soit X une variable aléatoire de loi de densité f et $a \in \mathbb{R}$ fixé. Considérons la variable aléatoire Y définie par $Y = \max(a, X)$. La loi de Y est égale à

$$\mathbb{P}(X \leq a)\delta_a + \mathbf{1}_{[a, +\infty[} f(x)dx.$$

(Une partie à densité et une masse au point a .) Pour nous en convaincre, nous considérons une fonction h continue bornée sur \mathbb{R} , et nous calculons

$$\begin{aligned}\mathbb{E}(h(Y)) &= \mathbb{E}(h(\max(a, X))) = \mathbb{E}(h(a)\mathbf{1}_{\{a \geq X\}}) + \mathbb{E}(h(X)\mathbf{1}_{\{a < X\}}) \\ &= h(a)\mathbb{P}(X \leq a) + \int_a^{+\infty} h(x)f(x)dx.\end{aligned}$$

Nous utilisons alors le résultat ci-dessus.

4.10.2 Recherche de densité

Un problème important est le suivant. Soit X une variable aléatoire réelle, admettant la densité f_X . Soit g une fonction mesurable, de sorte que $Y = g(X)$ soit aussi une variable aléatoire. Est-ce que Y admet une densité, et si oui, comment la calculer ?

Il convient d'abord de remarquer que cette densité n'existe pas toujours. Si par exemple $g(x) = a$ pour tout x , la loi de Y est la masse de Dirac en a , qui n'a pas de densité.

Pour résoudre ce problème, l'idée consiste à essayer de mettre $\mathbb{E}(h(Y)) = \mathbb{E}(h \circ g(X))$ sous la forme $\int h(y)f_Y(y)dy$ pour une fonction convenable f_Y , et une classe de fonctions h suffisamment grande. La fonction f_Y sera alors la densité cherchée.

Or, (4.5.34) implique

$$\mathbb{E}(h(Y)) = \mathbb{E}(h \circ g(X)) = \int_{-\infty}^{+\infty} h \circ g(x)f_X(x) dx. \quad (4.10.72)$$

Nous souhaitons faire le changement de variable $y = g(x)$ dans cette intégrale. Cela 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 4.10.2 1) Soit $Y = aX + b$, où a et b sont des constantes. Si $a = 0$, nous avons alors $Y = b$ et la loi de Y (sans densité) est la masse de Dirac en b . Si au contraire $a \neq 0$, nous faisons le changement de variable $y = ax + b$ dans (4.10.72), ce qui donne

$$\mathbb{E}(h(Y)) = \int_{-\infty}^{+\infty} h(ax + b)f_X(x) dx = \int_{-\infty}^{+\infty} h(y)f_X\left(\frac{y-b}{a}\right) \frac{1}{|a|} dy$$

(on peut considérer séparément les cas $a > 0$ et $a < 0$). Donc

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \frac{1}{|a|}. \quad (4.10.73)$$

Par exemple :

- 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 $\Gamma(\alpha, \theta/a)$.

2) Soit $Y = X^2$. La fonction $x \rightarrow x^2$ 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 avons donc

$$f_Y(y) = (f_X(-\sqrt{y}) + f_X(\sqrt{y})) \frac{1}{2\sqrt{y}} \mathbf{1}_{]0, +\infty[}(y). \quad (4.10.74)$$

Nous en déduisons le résultat suivant.

Proposition 4.10.3 *Si X suit la loi $\mathcal{N}(0, 1)$, alors X^2 suit la loi $\Gamma(1/2, 1/2)$.*

Dans le cas des vecteurs aléatoires, l'idée est la même. Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire de densité $f_{\mathbf{X}}$ sur \mathbb{R}^n . Soit \mathbf{g} une fonction de \mathbb{R}^n dans \mathbb{R}^m , et $\mathbf{Y} = \mathbf{g}(\mathbf{X})$. Plusieurs cas sont à considérer :

a) $m > n$: Le vecteur \mathbf{Y} n'admet pas de densité.

b) $m = n$: Nous utilisons, comme dans le cas unidimensionnel, le changement de variable $\mathbf{y} = \mathbf{g}(\mathbf{x})$ dans l'intégrale n -uple qui remplace le terme de droite de (4.10.72).

Rappelons cette formule de changement de variable. Supposons tout d'abord que \mathbf{g} soit une bijection continûment différentiable de A dans B (deux ouverts de \mathbb{R}^n). Son **jacobien** est le déterminant $J(\mathbf{g})(\mathbf{x})$ de la matrice jacobienne, dont les composantes sont les dérivées partielles $\partial g_i(\mathbf{x})/\partial x_j$, où g_i est la $i^{\text{ème}}$ composante de \mathbf{g} . Nous avons alors pour toute fonction h continue bornée de B dans \mathbb{R} :

$$\int_A h \circ \mathbf{g}(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_B h(\mathbf{y}) f_{\mathbf{X}} \circ \mathbf{g}^{-1}(\mathbf{y}) \frac{1}{|J(\mathbf{g})(\mathbf{g}^{-1}(\mathbf{y}))|} d\mathbf{y} \quad (4.10.75)$$

(attention à la valeur absolue de $J(\mathbf{g})$). Rappelons d'ailleurs que $1/J(\mathbf{g})(\mathbf{g}^{-1}(\mathbf{y}))$ est le jacobien de la transformation inverse \mathbf{g}^{-1} au point $\mathbf{y} \in B$. Dans le cas où $f_{\mathbf{X}}(\mathbf{x}) = 0$ en dehors de A , nous obtenons que \mathbf{Y} admet la densité

$$f_{\mathbf{Y}}(\mathbf{y}) = 1_B(\mathbf{y})f_{\mathbf{X}} \circ \mathbf{g}^{-1}(\mathbf{y}) \frac{1}{|J(\mathbf{g})(\mathbf{g}^{-1}(\mathbf{y}))|}. \quad (4.10.76)$$

Lorsque \mathbf{g} est simplement continûment différentiable, il existe souvent une partition finie $(A_i)_{1 \leq i \leq d}$ de l'ensemble $\{\mathbf{x} : f_{\mathbf{X}}(\mathbf{x}) > 0\}$, avec \mathbf{g} injective sur chaque A_i . Nous notons $B_i = \mathbf{g}(A_i)$ l'image de A_i par \mathbf{g} . Nous découpons alors l'intégrale selon les A_i , nous appliquons (4.10.75) à chaque morceau, et nous sommes. Nous obtenons alors

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{i=1}^d 1_{B_i}(\mathbf{y})f_{\mathbf{X}} \circ \mathbf{g}^{-1}(\mathbf{y}) \frac{1}{|J(\mathbf{g})(\mathbf{g}^{-1}(\mathbf{y}))|}, \quad (4.10.77)$$

où \mathbf{g}^{-1} est bien définie sur chaque B_i (comme image réciproque de la restriction de \mathbf{g} à A_i).

c) $m < n$: On commence par “compléter” \mathbf{Y} , en essayant de construire une application \mathbf{g}' de \mathbb{R}^n dans \mathbb{R}^n dont les m premières composantes coïncident avec les composantes de \mathbf{g} , et pour laquelle (4.10.76) ou (4.10.77) s'appliquent. Nous obtenons ainsi la densité $f_{\mathbf{Y}'}$ de $\mathbf{Y}' = \mathbf{g}'(\mathbf{X})$. Puis nous appliquons l'extension évidente de (4.8.60) :

$$f_{\mathbf{Y}}(y_1, \dots, y_m) = \int \cdots \int_{\mathbb{R}^{n-m}} f_{\mathbf{Y}'}(y_1, \dots, y_m, y_{m+1}, \dots, y_n) dy_{m+1} \cdots dy_n. \quad (4.10.78)$$

Donnons les exemples importants suivants.

- **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 : $u = r \cos \theta$, $v = r \sin \theta$. Le jacobien de \mathbf{g}^{-1} au point (r, θ) vaut r , donc (4.10.76) entraîne que

$$f_{\mathbf{Y}}(r, \theta) = r f_{\mathbf{X}}(r \cos \theta, r \sin \theta) \mathbf{1}_B(r, \theta). \quad (4.10.79)$$

Par exemple si U et V sont indépendantes et de loi $\mathcal{N}(0, 1)$, (4.9.68) entraîne que $f_{\mathbf{X}}(u, v) = \frac{1}{2\pi} \exp\left(-\frac{u^2+v^2}{2}\right)$, et (4.10.79) implique

$$f_{\mathbf{Y}}(r, \theta) = \frac{1}{2\pi} r e^{-r^2/2} \mathbf{1}_{]0, \infty[}(r) \mathbf{1}_{]0, 2\pi[}(\theta). \quad (4.10.80)$$

En particulier les variables aléatoires R et Θ sont indépendantes, la première suit la loi de densité $re^{-r^2/2} \mathbf{1}_{]0, \infty[}(r)$, et la seconde est uniforme sur $[0, 2\pi]$.

- **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 $\Gamma(\beta, \theta)$. Quelle est la densité de $Y = \frac{U}{U+V}$?

Comme la dimension de Y est plus petite que celle de \mathbf{X} , il faut d'abord "compléter" Y . Nous prenons par exemple $\mathbf{Y}' = (Y, Z)$, avec $Z = U + V$, ce qui correspond à $\mathbf{g}(u, v) = \left(\frac{u}{u+v}, u+v\right)$. Cette application est bijective de $A =]0, \infty[^2$ dans $B =]0, 1[\times]0, \infty[$, et nous avons $\mathbf{g}^{-1}(y, z) = (yz, z(1-y))$, qui a pour jacobien z .

Comme $f_{\mathbf{X}}(u, v) = \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1} v^{\beta-1} e^{-\theta(u+v)} \mathbf{1}_A(u, v)$, (4.10.76) entraîne

$$f_{\mathbf{Y}'}(y, z) = \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha+\beta-1} y^{\alpha-1} (1-y)^{\beta-1} e^{-\theta z} \mathbf{1}_B(y, z). \quad (4.10.81)$$

Il reste à appliquer (4.10.78) :

$$\begin{aligned} f_Y(y) &= \int_0^\infty f_{\mathbf{Y}'}(y, z) dz \\ &= \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1} \mathbf{1}_{]0, 1[}(y) \int_0^\infty z^{\alpha+\beta-1} e^{-\theta z} dz \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1} \mathbf{1}_{]0, 1[}(y) \end{aligned} \quad (4.10.82)$$

(utiliser (4.6.42)). On appelle **loi bêta** de paramètres α et β la loi admettant cette densité.

Nous obtenons aussi facilement la densité de Z . En effet, (4.10.82) montre que $f_{\mathbf{Y}'}(y, z)$ est le produit de $f_Y(y)$ par la fonction

$$f_Z(z) = \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha+\beta)} z^{\alpha+\beta-1} e^{-\theta z} \mathbf{1}_{]0, \infty[}(z),$$

qui d'après (4.8.60) est la densité de la variable aléatoire Z . Nous avons en fait démontré la :

Proposition 4.10.4 *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 α et β .*

- **Produit de convolution.**

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 (4.10.76) conduit à :

Proposition 4.10.5 *Si U et V sont deux variables aléatoires indépendantes de densités respectives f_U et f_V , alors $Z = U + V$ admet la densité*

$$f_Z(z) = \int_{-\infty}^{+\infty} f_U(u)f_V(z-u)du = \int_{-\infty}^{+\infty} f_U(z-v)f_V(v)dv. \quad (4.10.83)$$

La fonction f_Z est appelée le **produit de convolution** des deux fonctions f_U et f_V . Nous verrons intervenir de nouveau le produit de convolution dans le chapitre suivant.

Application : En utilisant ce calcul, nous pouvons montrer que 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.

4.11 Simulation de suites de variables aléatoires indépendantes

Un générateur aléatoire nous permet d’obtenir une suite X_1, \dots, X_n, \dots potentiellement infinie, de variables aléatoires indépendantes et de loi uniforme sur $[0, 1]$. Nous voulons construire une suite Y_1, \dots, Y_n, \dots de variables aléatoires réelles indépendantes, de loi donnée notée \mathbb{P}_Y .

4.11.1 Inversion de la fonction de répartition

Généralisons la méthode introduite au paragraphe 4.2.4.

Supposons que l’on connaisse la fonction de répartition F des variables Y_i . A partir d’une suite de variables aléatoires uniformes sur $[0, 1]$ et indépendantes, nous pouvons simuler une suite de variables aléatoires réelles indépendantes de fonction de répartition F . Comme au paragraphe 4.2.4, nous définissons la fonction “inverse” continue à gauche de F par :

$$G(x) = \inf\{y : F(y) \geq x\}, \quad \forall x \in]0, 1[. \quad (4.11.84)$$

Proposition 4.11.1 *La suite $(Y_n)_n$ définie par $Y_n = G(X_n)$ est une suite de variables aléatoires indépendantes, 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 au paragraphe 4.2.4. \square

Simulation d'une suite de variables aléatoires exponentielles indépendantes

Nous souhaitons simuler n réalisations indépendantes X_1, X_2, \dots, X_n de variables aléatoires de même loi de densité $f(x) = e^{-x} \mathbf{1}_{x>0}$, suivant la méthode préconisée au paragraphe 4.2.4. Remarquons que $\int_5^\infty e^{-x} dx = e^{-5}$ est très petit et nous pouvons nous limiter aux valeurs des variables aléatoires appartenant à l'intervalle $[0, 5]$. Une manière commode de représenter les tirages est d'en tracer un **histogramme**. Nous partitionnons $[0, 5]$ en intervalles I_k . Pour chacun d'eux, nous calculons la proportion de réalisations qui tombent dedans, proportion que nous normalisons par sa longueur. Ainsi, l'histogramme \hat{f}_n associé à cette partition et aux observations X_1, \dots, X_n est la fonction en escalier définie par

$$\hat{f}_n(x) = \frac{1}{\lambda(I_k)} \frac{\text{card}\{i : i \leq n, X_i \in I_k\}}{n} \quad \text{pour } x \in I_k.$$

Nous avons noté $\lambda(I)$ la longueur de l'intervalle I , et comme précédemment $\text{card}(A)$ désigne le cardinal de l'ensemble fini A . Les tailles des intervalles peuvent être adaptées aux valeurs prises par les variables X_i .

Par exemple, pour la loi exponentielle de paramètre 1, nous pouvons choisir la partition de $[0, 5]$ en intervalles I_k avec

$$\begin{array}{ll} I_1, \dots, I_{20} & \text{intervalles consécutifs de } [0, 2[\text{ de longueur } 1/10; \\ I_{21}, \dots, I_{35} & \text{intervalles consécutifs de } [2, 5[\text{ de longueur } 1/5. \end{array}$$

Au fur et à mesure de la simulation, on peut constater que \hat{f}_n "approche" la vraie densité f , lorsque n augmente. Cette observation sera justifiée au chapitre 5 par la loi des grands nombres.

4.11.2 Méthode du rejet

Cette méthode s'applique lorsque la probabilité \mathbb{P}_Y admet une densité f , et lorsqu'on connaît une autre probabilité ν , telle que

- il est possible de simuler des variables aléatoires de loi ν (par exemple par la méthode précédente);
- ν admet une densité g telle que

$$f(x) \leq ag(x), \quad \forall x \in \mathbb{R}^d, \quad (4.11.85)$$

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 variables aléatoires indépendantes de loi uniforme), et d'autre part d'une suite $(Z_n)_n$ potentiellement infinie de variables aléatoires indépendantes de loi ν et indépendantes des $(X_n)_n$. Nous posons alors

$$N = \inf\{n \in \mathbb{N}^*; f(Z_n) > aX_n g(Z_n)\}; \text{ et } Y = Z_n \text{ si } N = n. \quad (4.11.86)$$

Proposition 4.11.2 *La variable aléatoire N est à valeurs dans \mathbb{N}^* et suit une loi géométrique de paramètre $\frac{1}{a}$ (et donc d'espérance a). La variable aléatoire Y suit la loi \mathbb{P}_Y et N et Y sont indépendantes.*

Preuve. La première assertion entraîne que la variable Y est bien définie dans la seconde formule (4.11.86).

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 α plus tard), de sorte que $\mathbb{P}(N > n) = (1 - \alpha)^n$. Pour toute fonction h nous avons

$$\mathbb{E}(h(Y)) = \sum_{n=1}^{\infty} \mathbb{E}(h(Z_n) \mathbf{1}_{A_n} \mathbf{1}_{\{N > n-1\}}).$$

Les variables aléatoires $h(Z_n) \mathbf{1}_{A_n}$ 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

$$\mathbb{E}(h(Y)) = \sum_{n=1}^{\infty} \mathbb{E}(h(Z_n) \mathbf{1}_{A_n}) (1 - \alpha)^{n-1}. \quad (4.11.87)$$

Enfin $\mathbb{E}(h(Z_n) \mathbf{1}_{A_n})$ vaut

$$\int h(z)g(z) \left(\int_0^1 \mathbf{1}_{\{f(z) > axg(z)\}} dx \right) dz = \int h(z)g(z) \frac{f(z)}{ag(z)} dz = \frac{1}{a} \int h(z)f(z) dz.$$

En particulier α égale cette expression lorsque $h = 1$, donc $\alpha = 1/a$, puisque f est une densité. Ainsi, $0 < \alpha < 1$ et la variable aléatoire N est bien à valeurs dans \mathbb{N}^* .

En remplaçant $\mathbb{E}(h(Z_n) \mathbf{1}_{A_n})$ par sa valeur dans (4.11.87), nous obtenons que $\mathbb{E}(h(Y)) = \int h(z)f(z) dz$, d'où le résultat. L'indépendance est immédiate. \square

Nous avons ainsi obtenu une variable aléatoire de loi \mathbb{P}_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 ν . 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 4.11.2 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

$$Y = U_N, \quad \text{avec} \quad N = \min\{k \geq 1 : V_k \leq f(U_k)\},$$

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 4.11.2, la variable aléatoire 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.

4.12 Exercices sur le chapitre 4

EXERCICE 4.12.1 Soit X une v.a. de carré intégrable. Montrer que $\mathbb{E}((X - a)^2)$ atteint son minimum lorsque $a = \mathbb{E}(X)$.

EXERCICE 4.12.2 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 4.12.3 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 λ , où doit-on alors installer le poste à incendie ?

EXERCICE 4.12.4 La durée de fonctionnement d'une lampe suit une loi de densité

$$f(t) = \frac{1}{16}te^{-\frac{t}{4}}\mathbf{1}_{t \geq 0}$$

(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 4.12.5 Soit X une variable aléatoire de densité f_X . Considérons la variable aléatoire

$$Y = ce^{-\alpha X} \mathbf{1}_{\{X > 0\}}, \quad (\alpha > 0, c > 0).$$

Etudier l'existence d'une densité pour Y et donner sa valeur en fonction de f_X .

EXERCICE 4.12.6 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 4.12.7 Considérons deux paramètres $a > 0$ et $\alpha > 0$. On dira que X suit une loi de Pareto de paramètres (a, α) 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 4.12.8 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 4.12.9 Soient X et Y deux v.a. réelles. On suppose que la densité conditionnelle de X sachant $Y = y$ est la densité $\mathbf{1}_{\mathbb{R}_+}(x) y^2 x e^{-xy}$ 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 4.12.10 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 4.12.11 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 4.12.12 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 4.12.13 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, Θ) 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, Θ) et celle de (R^2, Θ) .
- 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 4.12.14 Soit 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 4.12.15 Soient X et Y deux variables aléatoires exponentielles indépendantes de paramètres λ et μ . 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 λ_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 5

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 Ω (muni de la tribu \mathcal{A} et de la probabilité \mathbb{P}). Nous voulons étudier la répartition des valeurs d'une variable aléatoire 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 variable aléatoire 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 variables aléatoires 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 ω ?

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.

5.1 Convergences de variables aléatoires

Dans ce paragraphe, nous allons étudier des modes de convergence impliquant la proximité des variables aléatoires 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 5.1.1 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

$$\mathbb{P}(|\mathbf{X}_n - \mathbf{X}| \geq \varepsilon) \rightarrow 0 \quad \text{quand } n \rightarrow \infty. \quad (5.1.1)$$

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

$$\mathbb{E}(|\mathbf{X}_n - \mathbf{X}|) \rightarrow 0 \quad \text{quand } n \rightarrow \infty. \quad (5.1.2)$$

Remarque : La convergence presque-sûre est la plus proche de la convergence simple des fonctions. Mais ici, nous permettons à certains ω de ne pas vérifier $\mathbf{X}_n(\omega) \rightarrow \mathbf{X}(\omega)$, si toutefois la probabilité de réalisation de l'ensemble de ces ω est nulle.

Ces convergences ne sont pas équivalentes, comme le montre l'exemple suivant.

Exemple 5.1.2 Soit $(X_n)_n$ une suite de variables aléatoires de Bernoulli à valeurs dans $\{0, 1\}$ telles que

$$\mathbb{P}(X_n = 1) = \frac{1}{n} \quad ; \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}.$$

Pour tout $\varepsilon \in]0, 1[$, la probabilité $\mathbb{P}(|X_n| > \varepsilon) = \frac{1}{n}$ tend vers 0 quand n tend vers l'infini. Ainsi, la suite $(X_n)_n$ tend vers $X = 0$ en probabilité. Comme $\mathbb{E}(X_n) = \frac{1}{n}$, elle tend aussi en moyenne vers 0.

Mais considérons maintenant une suite $(Y_n)_n$ de variables aléatoires de Bernoulli à valeurs dans $\{0, n^2\}$ telles que

$$\mathbb{P}(Y_n = n^2) = \frac{1}{n} \quad ; \quad \mathbb{P}(Y_n = 0) = 1 - \frac{1}{n}.$$

Par le même argument que ci-dessus, nous voyons que la suite $(Y_n)_n$ converge en probabilité vers 0, mais en revanche, $\mathbb{E}(Y_n) = n$, et la suite ne converge pas en moyenne vers 0 (ni vers aucune autre limite finie).

Exemple 5.1.3 Soit U une variable aléatoire uniforme sur $[0, 1]$. Posons $Z_n = \mathbf{1}_{\{U \leq \frac{1}{n}\}}$. Alors

$$\mathbb{P}(Z_n = 1) = \frac{1}{n} \quad ; \quad \mathbb{P}(Z_n = 0) = 1 - \frac{1}{n}.$$

Si $\omega \in \{U > 0\}$ est fixé, alors il existe n_0 tel que $U(\omega) > \frac{1}{n_0}$, et donc tel que $Z_n(\omega) = 0$ pour tout $n \geq n_0$. Comme $\mathbb{P}(U > 0) = 1$, ceci montre que la suite $(Z_n)_n$ converge presque-sûrement vers 0.

Nous allons maintenant étudier les liens entre ces différentes convergences.

Le résultat le plus simple est le suivant.

Théorème 5.1.4 *Supposons que les \mathbf{X}_n et \mathbf{X} soient des vecteurs aléatoires intégrables. Alors la convergence en moyenne de $(\mathbf{X}_n)_n$ vers \mathbf{X} entraîne la convergence en probabilité.*

Preuve. Ce résultat est un corollaire de l'inégalité de Markov (4.7.49). En effet, elle implique que pour tout $\varepsilon > 0$,

$$\mathbb{P}(|\mathbf{X}_n - \mathbf{X}| \geq \varepsilon) \leq \frac{\mathbb{E}(|\mathbf{X}_n - \mathbf{X}|)}{\varepsilon}.$$

La convergence vers 0 découle alors de la convergence en moyenne de $(\mathbf{X}_n)_n$ vers \mathbf{X} . \square

Nous avons vu dans l'exemple 5.1.2 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. Ce théorème s'appuie sur un certain nombre de propriétés de l'espérance que nous n'avons pas vues, et nous l'énonçons donc sans démonstration.

Théorème 5.1.5 *(de Lebesgue, ou de convergence dominée) Si la suite de vecteurs aléatoires \mathbf{X}_n converge presque-sûrement sur Ω 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, $\mathbb{E}(\mathbf{X}_n) \xrightarrow{n \rightarrow \infty} \mathbb{E}(\mathbf{X})$.

Attention : la réciproque est fautive. Dans l'exemple 5.1.2, supposons que les variables aléatoires $(X_n)_n$ 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 2.5.14 de Borel-Cantelli, nous en déduisons que pour presque tout ω , une infinité de $X_n(\omega)$ seront supérieurs à ε . 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 ω , un nombre fini au plus de $V_n(\omega)$ seront supérieurs à ε . La suite $(V_n)_n$ converge donc presque-sûrement vers 0.

Nous pouvons donc voir à travers ces exemples que ces notions sont délicates.

Remarque : L'hypothèse de domination est nécessaire. Considérons une suite de variables aléatoires $(T_n)_n$ 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.

D'autres relations existent entre ces convergences, que nous allons explorer maintenant.

Proposition 5.1.6 *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 $\mathbf{X}_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbf{X}$ 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 ω et de ε , ce qui implique que les variables aléatoires $Y_{n,\varepsilon} = \mathbf{1}_{N^c \cap A_{n,\varepsilon}}$ 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 5.1.5) 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 (5.1.1). □

La convergence en probabilité n'entraîne pas la convergence en moyenne, comme nous l'avons vu dans l'exemple 5.1.2. Si les \mathbf{X}_n ne sont "pas trop grands", il y a cependant équivalence entre les deux modes de convergence. En voici un exemple.

Proposition 5.1.7 *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 5.1.4, 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 ε , et donc

$$\mathbb{P}(|\mathbf{X}| > a) = 0. \quad (5.1.3)$$

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 (5.1.1)) que $\limsup_n \mathbb{E}(|\mathbf{X}_n - \mathbf{X}|) \leq \varepsilon$, et comme ε est arbitrairement proche de 0, nous en déduisons (5.1.2). \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 5.1.6, mais "à peine plus", comme le montre le résultat suivant.

Proposition 5.1.8 Si $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$, il existe une sous-suite (n_k) telle que $\mathbf{X}_{n_k} \rightarrow \mathbf{X}$ presque-sûrement quand $k \rightarrow \infty$.

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 2.5.14) appliqué aux ensembles $A_j = \{|\mathbf{X}_{n_{j+1}} - \mathbf{X}_{n_j}| > \frac{1}{2^j}\}$. \square

Exemple 5.1.9 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 ω , 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 5.1.10 *Soit f une fonction continue de \mathbb{R}^d dans \mathbb{R} .*

a) *Si $\mathbf{X}_n \rightarrow \mathbf{X}$ presque-sûrement, alors $f(\mathbf{X}_n) \rightarrow f(\mathbf{X})$ presque-sûrement.*

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 $\varepsilon > 0$,

$$\{|f(\mathbf{X}_n) - f(\mathbf{X})| \geq \varepsilon\} \subset \{|\mathbf{X}| > K\} \cup \{|\mathbf{X}| \leq K, |f(\mathbf{X}_n) - f(\mathbf{X})| \geq \varepsilon\}. \quad (5.1.4)$$

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 η suffisamment petit, $|\mathbf{x}| \leq K$ entraîne $|\mathbf{y}| \leq 2K$, et (5.1.4) 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

$$\limsup_n \mathbb{P}(|f(\mathbf{X}_n) - f(\mathbf{X})| \geq \varepsilon) \leq \mathbb{P}(|\mathbf{X}| > K). \quad (5.1.5)$$

Enfin $\lim_{K \rightarrow \infty} \mathbb{P}(|\mathbf{X}| > K) = 0$ (nous le montrons grâce au théorème de convergence dominée) et donc dans (5.1.5) la \limsup est nulle. Nous obtenons ainsi le résultat. \square

5.2 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**. Soit M_n , la moyenne des n premières variables aléatoires, i.e.

$$M_n = \frac{1}{n}(X_1 + \dots + X_n). \quad (5.2.6)$$

Notre objectif est de montrer que M_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**.

Nous allons démontrer la loi des grands nombres pour des variables aléatoires de carré intégrable. Les hypothèses les plus générales sont données dans l'énoncé du théorème 5.2.2.

Théorème 5.2.1 *Soit $(X_n)_n$ une suite de variables aléatoires indépendantes, de même loi et de carré intégrable ($X_n \in L^2$), et $m = \mathbb{E}(X_n)$ leur moyenne commune. Alors la suite $(M_n)_n$ définie par*

$$M_n = \frac{X_1 + \cdots + X_n}{n} \quad (5.2.7)$$

converge vers m , presque-sûrement et en moyenne, quand n tend vers l'infini. Elle converge donc aussi en probabilité. On a même un peu plus que la convergence en moyenne, à savoir que

$$\mathbb{E}((M_n - m)^2) \xrightarrow{n \rightarrow \infty} 0. \quad (5.2.8)$$

Le résultat sur la convergence en probabilité est appelé **loi faible des grands nombres**. Comme on le voit ci-dessous, sa preuve est presque immédiate. Elle utilise l'inégalité de Bienaymé-Chebyshev. Le résultat est peu informatif et permet d'obtenir certains contrôles d'erreurs. Le résultat prouvant la convergence presque-sure est appelé **loi forte des grands nombres**. Sa preuve est plus délicate et utilise le lemme de Borel-Cantelli. Le résultat est plus fort et donne une information sur la trajectoire $n \rightarrow X_n(\omega)$, pour presque tout ω .

Remarquons également qu'il est nécessaire que les variables aléatoires X_n soient indépendantes. En effet, 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. Notons σ^2 la variance commune à toutes les variables X_n , bien définie puisque X_n est de carré intégrable. En vertu de la linéarité de l'espérance et de (4.9.70) et (4.4.29), nous avons

$$\mathbb{E}(M_n) = m, \quad \mathbb{E}((M_n - m)^2) = \text{Var}(M_n) = \frac{\sigma^2}{n}, \quad (5.2.9)$$

d'où (5.2.8). En appliquant l'inégalité de Bienaymé-Chebyshev on obtient la convergence en probabilité (loi faible des grands nombres).

Comme $\mathbb{E}(|Y|)^2 \leq \mathbb{E}(Y^2)$, nous en déduisons que $\mathbb{E}(|M_n - m|)$ converge vers 0 quand n tend vers l'infini.

La preuve de la convergence presque-sûre est beaucoup plus délicate.

Quitte à remplacer X_n par $X_n - m$ (donc M_n par $M_n - m$), nous pouvons supposer que $m = 0$.

Montrons tout d'abord que la sous-suite $(M_{n^2})_n$ converge presque-sûrement vers 0.

D'après l'inégalité de Bienaymé-Chebyshev (4.7.50), (5.2.9) implique que pour $q \in \mathbb{N}^*$,

$$\mathbb{P}\left(|M_{n^2}| \geq \frac{1}{q}\right) \leq \frac{\sigma^2 q^2}{n^2}.$$

Donc si $A_{n,q} = \{|M_{n^2}| \geq \frac{1}{q}\}$, nous obtenons $\sum_{n \geq 1} \mathbb{P}(A_{n,q}) < \infty$. Posons ensuite $B_{n,q} = \cup_{m \geq n} A_{m,q}$ et $C_q = \cap_{n \geq 1} B_{n,q} = \limsup_n A_{n,q}$. En appliquant le lemme de Borel-Cantelli (Théorème 2.5.14) nous montrons alors que $\mathbb{P}(C_q) = 0$.

Par suite si nous posons $N = \cup_{q \in \mathbb{N}^*} C_q$, nous obtenons $\mathbb{P}(N) \leq \sum_{q=1}^{\infty} \mathbb{P}(C_q) = 0$ en vertu de (2.3.22). Maintenant, si $\omega \notin N$ alors $\omega \in \cap_{q \in \mathbb{N}^*} (C_q)^c$. Ainsi, $\omega \notin C_q$ pour tout $q \geq 1$, donc aussi $\omega \notin B_{n,q}$ pour n assez grand (car $B_{n,q}$ est décroissant en n). Cela veut dire que pour tout $\omega \notin N$, 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, $M_{n^2}(\omega) \rightarrow 0$ si $\omega \notin N$, donc (par définition),

$$M_{n^2} \rightarrow 0 \quad \text{p.s.} \quad (5.2.10)$$

Montrons maintenant que la suite $(M_n)_n$ tend presque-sûrement vers 0.

Pour tout entier n , notons $p(n)$ l'entier tel que $p(n)^2 \leq n < (p(n) + 1)^2$. Alors

$$M_n - \frac{p(n)^2}{n} M_{p(n)^2} = \frac{1}{n} \sum_{p=p(n)^2+1}^n X_p,$$

et puisque les variables aléatoires de la somme sont indépendantes, il vient

$$\begin{aligned} E\left(\left(M_n - \frac{p(n)^2}{n} M_{p(n)^2}\right)^2\right) &= \frac{n - p(n)^2}{n^2} \sigma^2 \\ &\leq \frac{2p(n) + 1}{n^2} \sigma^2 \leq \frac{2\sqrt{n} + 1}{n^2} \sigma^2, \end{aligned}$$

parce que $p(n) \leq \sqrt{n}$. Une nouvelle application de l'inégalité de Bienaymé-Chebyshev donne alors

$$P\left(\left|M_n - \frac{p(n)^2}{n} M_{p(n)^2}\right| \geq a\right) \leq \frac{2\sqrt{n} + 1}{n^2} \frac{\sigma^2}{a^2}.$$

Comme la série $\sum_n \frac{2\sqrt{n}+1}{n^2}$ converge, le même raisonnement que ci-dessus pour (5.2.10) montre que

$$M_n - \frac{p(n)^2}{n} M_{p(n)^2} \rightarrow 0 \quad \text{p.s.}$$

Par ailleurs $M_{p(n)^2} \rightarrow 0$ p.s. d'après (5.2.10), et $p(n)^2/n \rightarrow 1$. Nous en déduisons que $M_n \rightarrow 0$ p.s. \square

Enonçons **la loi forte des grands nombres** dans toute sa généralité. Pour sa preuve, le lecteur pourra consulter le livre “Promenade aléatoire” de Benaïm et El Karoui.

Théorème 5.2.2 *Considérons des variables aléatoires X_n indépendantes, de même loi et intégrables, et notons $m = \mathbb{E}(X_n)$ leur moyenne commune. Alors la moyenne empirique*

$$\frac{X_1 + \cdots + X_n}{n} \tag{5.2.11}$$

converge presque-sûrement et en moyenne vers m , quand n tend vers l'infini.

Revenons à “l'approche par les fréquences” du Chapitre 2. Soit un événement A . Nous répétons l'expérience, et nous notons X_n la variable aléatoire 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

$$f_n(A) = \frac{1}{n}(X_1 + \cdots + X_n) = M_n.$$

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 (5.2.7) 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 (2.1.1), 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 ω . Nous allons voir dans l'exemple suivant qu'il existe un ensemble négligeable sur lequel la convergence n'est pas réalisée.

Exemple 5.2.3 Considérons un jeu de Pile ou Face infini, c'est-à-dire une suite X_n de variables aléatoires ne prenant que les valeurs 0 et 1. Nous définissons cette suite

sur l'espace $\Omega = \{0, 1\}^{\mathbb{N}^*}$, i.e. un point ω 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 ω).

5.3 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 5.3.1 *Considérons une suite $(X_n)_n$ de variables aléatoires indépendantes qui suivent des lois uniformes sur $[0, 1]$, et une fonction f mesurable bornée sur $[0, 1]$, par exemple une fonction continue sur $[0, 1]$.*

Alors

$$\int_0^1 f(x)dx = \lim_n \frac{f(X_1) + \dots + f(X_n)}{n},$$

pour la convergence presque-sûre.

Preuve. Nous appliquons la loi des grands nombres aux variables aléatoires $f(X_i)$ qui vérifient bien toutes les hypothèses voulues puisque f est bornée. Nous avons alors

$$\lim_n \frac{f(X_1) + \dots + f(X_n)}{n} = \mathbb{E}(f(X)) = \int_0^1 f(x)dx.$$

□

En choisissant des variables aléatoires 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 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 variables aléatoires 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 variables aléatoires $(X_{n,j} : n \geq 1, 1 \leq j \leq d)$ sont indépendantes et uniformes sur $[-\alpha, \alpha]$. Une suite de valeurs approchées de I est alors

$$I_n = \frac{(2\alpha)^d}{n} (f(\mathbf{X}_1) + \dots + f(\mathbf{X}_n)). \quad (5.3.12)$$

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 7.2.3.

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 \mathbf{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.

5.4 Exercices sur le chapitre 5

EXERCICE 5.4.1 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 variables aléatoires indépendantes telles que pour tout $n \geq 1$,

$$\mathbb{P}\left(X_n = \frac{1}{u_n}\right) = u_n; \quad \mathbb{P}(X_n = 0) = 1 - u_n.$$

1) Calculer $\mathbb{E}(X_n)$.

2) On suppose que $\sum_{n \geq 1} u_n < +\infty$. En déduire que $X_n \xrightarrow{\mathbb{P}} 0$. 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 5.4.2 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 5.4.3 Soit $(X_n)_n$ une suite de variables aléatoires indépendantes, de même loi de Bernoulli

$$\mathbb{P}(X_n = 1) = x; \quad \mathbb{P}(X_n = 0) = 1 - x,$$

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 5.4.4 *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 que $\forall a \in \mathbb{R}, \forall u \in \mathbb{R}_+$,

$$\mathbb{P}(S_n - nm \geq a) \leq e^{-ua} (M(u))^n,$$

et que $\forall a \in \mathbb{R}, \forall u \in \mathbb{R}_-$,

$$\mathbb{P}(S_n - nm \leq a) \leq e^{-ua} (M(u))^n.$$

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 $\sigma^2 = 10$.

Quelle taille d'échantillon doit-on choisir pour obtenir

$$\mathbb{P}(|Y_n - m| \leq \varepsilon) \geq \alpha,$$

avec $\alpha = 0,95$ et $\varepsilon = 0,05$,

- En utilisant l'inégalité de Bienaymé-Chebyshev ?
- En utilisant l'inégalité de Chernov ?

EXERCICE 5.4.5 (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 5.4.6 Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de même loi, intégrables. On pose $\mathbb{E}(X_1) = a$. Soit f une fonction continue bornée sur \mathbb{R} .

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 exponentielle de paramètre $\frac{1}{a}$, 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 α suit une loi Gamma de paramètres (n, α) , c'est-à-dire de densité $x \mapsto \frac{\alpha^n}{(n-1)!} x^{n-1} e^{-\alpha x}$ pour $x > 0$.

Chapitre 6

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

6.1 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**.

6.1.1 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 va-

riable 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 variables aléatoires réelles. Ces variables aléatoires 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 6.1.1 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

$$\phi_{\mathbf{X}}(\mathbf{u}) = \mathbb{E}\left(e^{i\langle \mathbf{u}, \mathbf{X} \rangle}\right). \quad (6.1.1)$$

Si X est à valeurs réelles, ϕ_X est définie de \mathbb{R} dans \mathbb{C} et vaut

$$\forall u \in \mathbb{R}, \quad \phi_X(u) = \mathbb{E}\left(e^{iuX}\right). \quad (6.1.2)$$

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 6.1.2 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 $\mathbf{u}_p \xrightarrow{p \rightarrow \infty} \mathbf{u}$. 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 5.1.5), 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 6.1.3 *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 :*

$$\phi_{\mathbf{a} + \mathbf{A}\mathbf{X}}(\mathbf{u}) = e^{i\langle \mathbf{u}, \mathbf{a} \rangle} \phi_{\mathbf{X}}(\mathbf{A}^t \mathbf{u}), \quad \forall \mathbf{u} \in \mathbb{R}^m, \quad (6.1.3)$$

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

6.1.2 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,

$$\phi_X(u) = (pe^{iu} + 1 - p)^n. \quad (6.1.4)$$

2) X suit une loi de Poisson de paramètre θ

$$\mathbb{E}(e^{iuX}) = \sum_{k \geq 0} e^{iuk} \frac{\theta^k}{k!} e^{-\theta} = e^{\theta e^{iu}} e^{-\theta} = e^{\theta(e^{iu} - 1)}.$$

Ainsi,

$$\phi_X(u) = e^{\theta(e^{iu} - 1)}. \quad (6.1.5)$$

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 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 ,

$$\int_{-\infty}^{+\infty} e^{sx} g(x) dx = e^{s^2/2}, \quad (6.1.6)$$

puisque $g(x) e^{sx} = g(x-s) e^{s^2/2}$. Nous voulons montrer que (6.1.6) 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 (6.1.6) sont encore égaux, et donc que

$$\phi_X(u) = e^{-u^2/2}. \quad (6.1.7)$$

Remarque 6.1.4 *Au cours de la preuve précédente, nous avons obtenu les moments d'une variable aléatoire X de loi normale centrée réduite,*

$$\mathbb{E}(X^{2n}) = \frac{(2n)!}{2^n n!} ; \quad \mathbb{E}(X^{2n+1}) = 0.$$

Par exemple, $\mathbb{E}(X^4) = 3$.

6) X suit une loi normale $\mathcal{N}(m, \sigma^2)$

Dans ce cas, la variable aléatoire X s'écrit $X = m + \sigma Y$, où Y suit la loi $\mathcal{N}(0, 1)$. D'après (6.1.3) et (6.1.7), sa fonction caractéristique vaut alors

$$\phi_X(u) = e^{i u m - u^2 \sigma^2 / 2}. \quad (6.1.8)$$

6.1.3 Propriété fondamentale

L'intérêt majeur de la fonction caractéristique réside dans le fait qu'elle caractérise la loi de la variable aléatoire.

Théorème 6.1.5 *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

$$\int_{\mathbb{R}^n} g(\mathbf{x}) \mathbb{P}_{\mathbf{X}}(d\mathbf{x}) = \int_{\mathbb{R}^n} g(\mathbf{x}) \mathbb{P}_{\mathbf{Y}}(d\mathbf{x}), \quad (6.1.9)$$

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 (6.1.9) 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 6.1.6 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 $u_1, \dots, u_n \in \mathbb{R}$

$$\phi_{\mathbf{X}}(u_1, \dots, u_n) = \prod_{j=1}^n \phi_{X_j}(u_j), \quad (6.1.10)$$

où $\phi_{\mathbf{X}}$ désigne la fonction caractéristique du vecteur aléatoire \mathbf{X} , et ϕ_{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 (6.1.10), en utilisant (4.9.67).

Supposons inversement qu'on ait (6.1.10). 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 (6.1.10), 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 variable aléatoire à valeurs dans \mathbb{R}_+ , nous pouvons définir sa transformée de Laplace par

$$\psi_X(\lambda) = \mathbb{E}(e^{-\lambda X}), \quad \lambda \in \mathbb{R}_+. \quad (6.1.11)$$

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 variable aléatoire à valeurs dans \mathbb{N} , de fonction génératrice G_X , alors $\psi_X(\lambda) = G_X(e^{-\lambda})$.

Exemple 6.1.7 Transformée de Laplace d'une loi gamma. Soit X une variable aléatoire de loi $\Gamma(\alpha, \theta)$. Alors,

$$\psi_X(\lambda) = \frac{1}{\Gamma(\alpha)} \theta^\alpha \int_0^{+\infty} e^{-\lambda x} x^{\alpha-1} e^{-\theta x} dx = \frac{1}{\Gamma(\alpha)} \theta^\alpha \int_0^{+\infty} e^{-(\lambda+\theta)x} x^{\alpha-1} dx = \frac{\theta^\alpha}{(\lambda+\theta)^\alpha}. \quad (6.1.12)$$

En particulier la transformée de Laplace d'une variable aléatoire de loi exponentielle de paramètre θ , prise en λ , vaut $\frac{\theta}{\lambda+\theta}$.

6.1.4 Somme de vecteurs aléatoires indépendants

Proposition 6.1.8 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

$$\phi_{\mathbf{X}+\mathbf{Y}} = \phi_{\mathbf{X}} \phi_{\mathbf{Y}}. \quad (6.1.13)$$

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 (4.9.67). \square

Exemple 6.1.9 Soient X, Y deux variable aléatoire 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 (6.1.8) et (6.1.13).

- 2) si X et Y suivent des lois de Poisson de paramètres θ et θ' , alors Z suit une loi de Poisson de paramètre $\theta + \theta'$;

Cela découle de (6.1.5) et (6.1.13).

- 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 (6.1.4) et (6.1.13).

6.1.5 Fonction caractéristique et moments

Proposition 6.1.10 Soit \mathbf{X} un vecteur aléatoire de \mathbb{R}^n . Si la variable aléatoire $|\mathbf{X}|^m$ (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 ,

$$\frac{\partial^m}{\partial u_{i_1} \partial u_{i_2} \cdots \partial u_{i_m}} \phi_{\mathbf{X}}(\mathbf{u}) = i^m \mathbb{E} \left(e^{i \langle \mathbf{u}, \mathbf{X} \rangle} X_{i_1} X_{i_2} \cdots X_{i_m} \right) \quad (6.1.14)$$

(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

$$\mathbb{E}(X) = -i \phi'_{\mathbf{X}}(0), \quad (\text{resp. } \mathbb{E}(X^2) = -\phi''_{\mathbf{X}}(0)). \quad (6.1.15)$$

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

$$\frac{\phi_{\mathbf{X}}(\mathbf{u} + t\mathbf{v}_j) - \phi_{\mathbf{X}}(\mathbf{u})}{t} = \mathbb{E} \left(e^{i \langle \mathbf{u}, \mathbf{X} \rangle} \frac{e^{itX_j} - 1}{t} \right). \quad (6.1.16)$$

Soit t_p une suite de réels tendant vers 0. Les variables aléatoires $(e^{it_p X_j} - 1)/t_p$ convergent simplement vers iX_j , en restant bornées en module par la variable aléatoire $|X_j|$ (car $|e^{itx} - 1| = 2|\sin(tx/2)| \leq |tx|$), qui par hypothèse est intégrable. Donc par le théorème 5.1.5 (de Lebesgue), nous en déduisons que (6.1.16) 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 (6.1.14). Enfin, nous pouvons montrer comme dans la Proposition 6.1.2 que cette dérivée est continue. \square

6.2 Vecteurs gaussiens

Définition 6.2.1 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 6.2.2 Si les X_i sont des variables aléatoires normales indépendantes, le vecteur \mathbf{X} est gaussien (cela découle de l'exemple 6.1.9 (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 6.2.3 \mathbf{X} est un vecteur gaussien si et seulement si sa fonction caractéristique s'écrit

$$\phi_{\mathbf{X}}(\mathbf{u}) = e^{i\langle \mathbf{u}, \mathbf{m} \rangle - \frac{1}{2} \langle \mathbf{u}, \mathbf{C} \mathbf{u} \rangle}, \quad (6.2.17)$$

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 (6.2.17). 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

$$\phi_Y(v) = \phi_{\mathbf{X}}(v\mathbf{a}) = e^{i v \langle \mathbf{m}, \mathbf{a} \rangle - \frac{v^2}{2} \langle \mathbf{a}, \mathbf{C}\mathbf{a} \rangle},$$

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

$$\phi_Y(v) = e^{i v \langle \mathbf{a}, \mathbf{m} \rangle - \frac{v^2}{2} \langle \mathbf{a}, \mathbf{C}\mathbf{a} \rangle}.$$

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 (6.2.17). \square

Corollaire 6.2.4 *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 (6.2.17) et le corollaire 6.1.6. \square

Proposition 6.2.5 *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 4.8.8), 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 6.2.6 *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*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(\mathbf{C})}} e^{-\frac{1}{2} \langle \mathbf{x} - \mathbf{m}, \mathbf{C}^{-1}(\mathbf{x} - \mathbf{m}) \rangle}. \quad (6.2.18)$$

Preuve. Reprenons la preuve de la proposition précédente : les λ_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 (6.2.18).

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 variable aléatoire 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 6.2.7 On appelle variable aléatoire de χ^2 (chi-deux) à d degrés de liberté toute variable aléatoire 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 chapitre 7), puisqu'elle permet de construire un test (appelé test du chi-deux), fondamental pour “savoir” (statistiquement parlant) si une variable aléatoire suit une certaine loi donnée a priori ou si deux variables aléatoires sont indépendantes.

Grâce aux propositions 4.10.3 et 4.10.4, nous obtenons immédiatement que la loi du χ^2 à d degrés de liberté est la loi $\Gamma(d/2, 1/2)$ de densité définie sur \mathbb{R}_+ par

$$\frac{1}{2^{d/2} \Gamma(d/2)} \exp(-y/2) y^{d/2-1}. \quad (6.2.19)$$

En utilisant la définition, nous obtenons immédiatement que si Y suit une loi de χ^2 à d degrés de liberté, alors

$$\mathbb{E}(Y) = d \quad , \quad \text{Var}(Y) = 2d. \quad (6.2.20)$$

6.3 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 5 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 6.3.1 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 6.3.2 Un cas très simple est celui où toutes les variables aléatoires 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 3.5.4, nous avons montré la convergence en loi d’une suite de variables aléatoires binomiales vers une variable aléatoire de Poisson, pour un bon choix des paramètres.

Exemple 6.3.3 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 6.3.4 Si $\mathbf{X}_n \xrightarrow{P}_{n \rightarrow \infty} \mathbf{X}$, alors $\mathbf{X}_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} \mathbf{X}$.

Ainsi, les convergences en moyenne et presque-sûre entraînent également la convergence en loi, par la proposition 5.1.6.

Preuve. Supposons que la suite $(\mathbf{X}_n)_n$ converge en probabilité vers \mathbf{X} . Soit f une fonction continue bornée sur \mathbb{R}^d . D'après la proposition 5.1.10, la suite $(f(\mathbf{X}_n))_n$ converge en probabilité vers $f(\mathbf{X})$, et comme f est bornée, la proposition 5.1.7 entraîne que $f(\mathbf{X}_n)$ converge aussi en moyenne vers $f(\mathbf{X})$. En particulier, $(\mathbb{E}(f(\mathbf{X}_n)))_n$ converge vers $\mathbb{E}(f(\mathbf{X}))$. \square

Proposition 6.3.5 Soient X_n et X des variables aléatoires 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

$$1_{]-\infty, b]} \leq f_{p,b} \leq 1_{]-\infty, b+1/p]}. \quad (6.3.21)$$

Alors $(\mathbb{E}(f_{p,b}(X_n)))_n$ converge vers $\mathbb{E}(f_{p,b}(X))$ quand n tend vers l'infini. De (6.3.21), 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

$$n \geq n_0 \quad \Rightarrow \quad \mathbb{P}(X_n \notin]a, b]) = 1 - F_n(b) + F_n(a) \leq 3\varepsilon. \quad (6.3.22)$$

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

$$|\mathbb{E}(f(X_n)) - \mathbb{E}(g(X_n))| \leq M \mathbb{P}(X_n \notin]a, b]) + \varepsilon, \quad (6.3.23)$$

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

$$n \geq n_1 \quad \Rightarrow \quad |\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))| \leq \varepsilon. \quad (6.3.24)$$

D'après (6.3.22), (6.3.23) et (6.3.24), nous avons

$$n \geq \sup(n_0, n_1) \quad \Rightarrow \quad |\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))| \leq 3\varepsilon + 5M\varepsilon.$$

ε étant arbitraire, nous en déduisons que $(\mathbb{E}(f(X_n)))_n$ converge vers $\mathbb{E}(f(X))$, d'où le résultat. \square

Corollaire 6.3.6 : 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 6.3.7 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 6.3.5, 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 6.3.8 (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 6.3.7, 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 ε 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 6.5.5.

Le théorème suivant caractérise la convergence en loi à l'aide des fonctions caractéristiques. C'est un critère extrêmement utile dans la pratique.

Théorème 6.3.9 (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) ϕ sur \mathbb{R}^d , et si cette fonction est **continue** en 0, alors c'est la fonction caractéristique d'une variable aléatoire \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 6.3.1 (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 θ_n et θ , et $\theta_n \rightarrow \theta$.
- 4) X_n et X suivent des lois exponentielles de paramètres λ_n et λ , et $\lambda_n \rightarrow \lambda$.

Il permet aussi de prouver le lemme de Slutsky très rapidement, y compris pour des vecteurs aléatoires.

Théorème 6.3.10 (*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 5.1.7. \square

En conclusion de ce paragraphe, nous pouvons résumer dans la figure 6.1 les relations entre les différents modes de convergence.

6.4 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

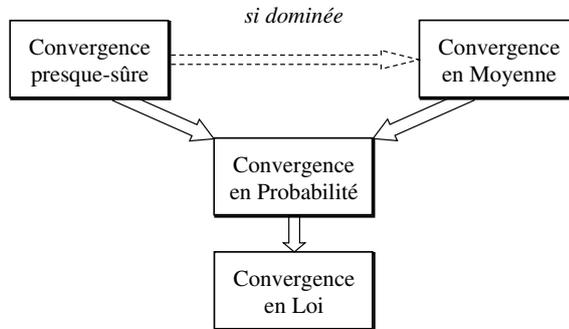


FIGURE 6.1 – Relations entre les différentes notions de convergence

La situation est la même que dans le chapitre précédent concernant la loi des grands nombres (Théorème 5.2.1). Nous considérons une suite de variables aléatoires $(X_n)_n$ **indépendantes, de même loi, et de carré intégrable**. Notons m et σ^2 la moyenne et la variance communes aux variables X_n , et

$$S_n = X_1 + \cdots + X_n \quad (6.4.25)$$

(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 α : si α est “petit” cette suite va encore tendre vers 0, et elle va “exploser” si α est “grand”. On peut espérer que pour une (et alors nécessairement une seule) valeur de α , 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 α . (Voir l'exercice 6.5.7). 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 6.4.1 *Si les X_n sont des variables aléatoires réelles indépendantes et de même loi, de carré intégrable, de moyenne m et de variance σ^2 avec $\sigma^2 > 0$, alors les variables*

$$\frac{S_n - nm}{\sigma\sqrt{n}}$$

convergent en loi vers une variable aléatoire 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 6.5.6.

Soit ϕ 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 (6.1.3), la fonction caractéristique de Y_n est

$$\phi_n(u) = \phi\left(\frac{u}{\sigma\sqrt{n}}\right)^n. \quad (6.4.26)$$

Comme $\mathbb{E}(X_n - m) = 0$ et $\mathbb{E}((X_n - m)^2) = \sigma^2$, (6.1.14) entraîne

$$\phi(u) = 1 - \frac{u^2\sigma^2}{2} + u^2o(|u|) \quad \text{quand } u \rightarrow 0.$$

Comme $\phi(0) = 1$ et que ϕ 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 6.3.9. \square

Remarque : Si les X_i de l'énoncé du théorème 6.4.1 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 6.4.2 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 θ .

Dans les autres cas, nous utilisons les théorèmes précédents :

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} p, \quad (6.4.27)$$

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1). \quad (6.4.28)$$

Si nous désignons par Φ la fonction de répartition de la loi $\mathcal{N}(0, 1)$, il vient

$$\mathbb{P}(S_n \leq x) \simeq \Phi\left(\frac{x - np}{\sqrt{np(1-p)}}\right). \quad (6.4.29)$$

Rappelons que la fonction de répartition Φ ci-dessus est "tabulée", et une table de cette fonction est fournie à la section 4.6.

Exemple 6.4.3 Lançons 1000 fois une pièce (que l'on suppose non truquée). Cherchons la probabilité d'obtenir plus de 545 fois le côté Face.

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 6.4.1 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 6.4.4 *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 6.4.5 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.

6.5 Exercices sur le chapitre 6

EXERCICE 6.5.1 Soient X et Y deux variables aléatoires indépendantes de carré intégrable, de même loi et centrées. Soit ϕ 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 6.5.2 Soit X une variable aléatoire de Cauchy.

1. Calculer la fonction caractéristique de X .
2. Montrer que $\phi_{2X} = (\phi_X)^2$.

EXERCICE 6.5.3 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 6.5.4 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 θ différents.

2. Montrer que $\mathbb{P}(A(\theta)) = \mathbb{P}(A(0))$, pour tout θ . On montrera que pour tout θ ,

$$\begin{pmatrix} \mathbf{X}_1 \cos \theta + \mathbf{X}_2 \sin \theta \\ \mathbf{X}_1 \sin \theta - \mathbf{X}_2 \cos \theta \end{pmatrix}$$
 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 6.5.5 *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 6.5.6 *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

$$f(T_n) - f(T'_n) = \sum_{i=1}^n (f(W_{i,n} + X_{i,n}) - f(W_{i,n} + Y_{i,n})). \quad (6.5.30)$$

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 6.5.7 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 7

Statistique

I only believe in statistics that I doctored myself.

Winston Churchill

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 CO₂ 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 ce chapitre donner quelques éléments de réponse pour chacune de ces questions.

7.1 Estimation ponctuelle

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 7.1.1 Le sexe, fille (F) ou garçon (G), d'un nouveau-né est modélisé par une loi P_θ 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 θ . Pour estimer θ , on observe les sexes des n nouveaux-nés dans une maternité.

Exemple 7.1.2 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 λ , 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 λ .

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 7.1.3 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_θ est une probabilité sur $(\mathcal{X}, \mathcal{A})$.

On notera génériquement θ le paramètre d'une famille paramétrique de probabilités.

Exemple 7.1.4 Dans l'exemple 7.1.1, 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 7.1.5 Dans l'exemple 7.1.2, 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_λ 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 7.1.6 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 Ω qu'on ne précisera pas (on peut cependant prendre pour Ω 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_θ .

On observe un n -échantillon \mathbf{X} . On souhaite construire à partir de cet échantillon une quantité $f(\mathbf{X})$ qui s'approche du paramètre inconnu θ . Une telle quantité est appelée estimateur de θ et est souvent notée $\hat{\theta}_n$. Un estimateur ne dépend que de l'échantillon \mathbf{X} et ne dépend pas de θ , et est donc entièrement caractérisé par la fonction déterministe $f : \mathcal{X}^n \rightarrow \Theta$.

Exemple 7.1.7 Dans l'exemple 7.1.1, 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_θ telles que $P_\theta(\{G\}) = \theta$. Pour estimer θ , 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_θ . Un estimateur intuitif de θ 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 7.1.8 Dans l'exemple 7.1.2, 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 λ 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 λ : $\hat{\lambda}_n = 1/\bar{X}_n$, 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 λ (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 7.1.1 est particulièrement simple, mais l'exemple 7.1.2 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".

7.1.1 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 7.1.9 Dans l'exemple 7.1.1 on peut examiner la suite $\hat{\theta}_n = f_n(X_1, \dots, X_n)$ définie par la suite de fonctions :

$$f_n : \begin{cases} \{F, G\}^n \rightarrow \mathbb{R}, \\ (x_1, \dots, x_n) \mapsto \frac{1}{n} \sum_{i=1}^n \mathbf{1}_G(x_i). \end{cases}$$

La propriété de convergence est essentielle.

Définition 7.1.10 [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 θ lorsque l'échantillon est tiré avec la loi \mathbb{P}_θ . 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 θ (celle avec laquelle les X_i ont été tirés).

Exemple 7.1.11 Dans l'exemple 7.1.1, sous \mathbb{P}_θ , les variables aléatoires $\mathbf{1}_G(X_i)$ sont des variables i.i.d. de loi de Bernoulli de paramètre θ . 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 7.1.12 [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 7.1.13 Dans l'exemple 7.1.1, 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 7.1.14 [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” :

$$\text{RQM}_{\boldsymbol{\theta}}(\hat{g}_n) = \text{Var}_{\boldsymbol{\theta}}(\hat{g}_n) + (\mathbb{E}_{\boldsymbol{\theta}}(\hat{g}_n) - g(\boldsymbol{\theta}))^2. \quad (7.1.1)$$

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 7.1.15 Dans l'exemple 7.1.1, 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 7.1.16 [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 7.1.17 Dans l'exemple 7.1.1, on utilise le théorème de la limite centrale pour obtenir que, sous \mathbb{P}_θ :

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, \theta(1 - \theta)),$$

en loi. La variance asymptotique est ici $\sigma^2(\theta) = \theta(1 - \theta)$.

7.1.2 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 :

$$\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 \in]0, +\infty[\}. \quad (7.1.2)$$

On va introduire et discuter ici les estimateurs dits “empiriques” de ces quantités.

Proposition 7.1.18 (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 $\mu_\boldsymbol{\theta}$ de la loi $P_\boldsymbol{\theta}$. La moyenne empirique \bar{X}_n définie par :

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j, \quad (7.1.3)$$

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 $P_\boldsymbol{\theta}$ sont de carré intégrable, alors $\text{RQM}_\boldsymbol{\theta}(\bar{X}_n) = \sigma_\boldsymbol{\theta}^2/n$, avec $\sigma_\boldsymbol{\theta}^2$ la variance de la loi $P_\boldsymbol{\theta}$ et \bar{X}_n est asymptotiquement normal, avec

$$\sqrt{n}(\bar{X}_n - \mu_\boldsymbol{\theta}) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma_\boldsymbol{\theta}^2),$$

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 :

$$\bar{V}_n = \frac{1}{n} \sum_{j=1}^n X_j^2 - \bar{X}_n^2 = \frac{1}{n} \sum_{j=1}^n X_j^2 - \left(\frac{1}{n} \sum_{j=1}^n X_j \right)^2, \quad (7.1.4)$$

qu'on appelle variance empirique. Cette somme peut se réécrire différemment.

Proposition 7.1.19 (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 θ contient la variance, i.e., il existe une fonction $g : \Theta \rightarrow \mathbb{R}$ telle que $g(\theta)$ est la variance σ_θ^2 de la loi P_θ . La variance empirique de l'échantillon est définie par :*

$$\bar{V}_n = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2. \quad (7.1.5)$$

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_θ 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 σ^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 (7.1.4) et (7.1.5) 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 σ_θ^2 . On a alors :

$$\bar{V}_n = \frac{1}{n} \sum_{j=1}^n (\tilde{X}_j + \mu_\theta)^2 - \left(\frac{1}{n} \sum_{j=1}^n \tilde{X}_j + \mu_\theta \right)^2 = \frac{1}{n} \sum_{j=1}^n \tilde{X}_j^2 - \left(\frac{1}{n} \sum_{j=1}^n \tilde{X}_j \right)^2. \quad (7.1.6)$$

Pour p entier, le p^{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 (7.1.4) 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 :

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \tilde{X}_i^2 - \sigma_{\boldsymbol{\theta}}^2 \right) \xrightarrow[n \rightarrow +\infty]{\text{d.}} \mathcal{N}(0, \mu_{\boldsymbol{\theta}}^{(4)} - \sigma_{\boldsymbol{\theta}}^4).$$

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

$$\sqrt{n}(\bar{V}_n - \sigma_{\boldsymbol{\theta}}^2) \xrightarrow[n \rightarrow +\infty]{\text{d.}} \mathcal{N}(0, \mu_{\boldsymbol{\theta}}^{(4)} - \sigma_{\boldsymbol{\theta}}^4),$$

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 7.1.20 *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 θ contient la variance, i.e., il existe une fonction $g : \Theta \rightarrow \mathbb{R}$ telle que $g(\theta)$ est la variance σ_θ^2 de la loi P_θ . La variance empirique non-biaisée V_n est définie par :*

$$V_n = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2. \quad (7.1.7)$$

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

7.1.3 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 ϕ . La principale application de cette méthode générale est la méthode des moments qu'on décrit ci-dessous.

7.1.4 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 θ . 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 $\psi : \mathcal{X} \rightarrow \mathbb{R}^d$ telle que $\mathbb{E}_\theta(|\psi(X_1)|) < \infty$ et $\mathbf{g}(\theta)$ peut s'écrire comme l'espérance de $\psi(X_1)$ pour tout $\theta \in \Theta$:

$$\mathbf{g}(\theta) = \mathbb{E}_\theta(\psi(X_1)).$$

L'estimateur des moments de θ est alors

$$\hat{\theta}_n = \mathbf{g}^{-1} \left(\frac{1}{n} \sum_{i=1}^n \psi(X_i) \right).$$

L'estimateur des moments est convergent car c'est un cas particulier de la méthode de substitution. En effet, $\frac{1}{n} \sum_{i=1}^n \psi(X_i)$ est un estimateur convergent de $\mathbf{g}(\theta) = \mathbb{E}_\theta(\psi(X_1))$ d'après la loi forte des grands nombres, et \mathbf{g}^{-1} est continue.

Exemple 7.1.21 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 (4.10.82)) :

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbf{1}_{]0,1[}(x).$$

Ici le paramètre $\theta = (\alpha, \beta)$ est de dimension 2. On désire estimer θ . 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 α et β .

Exemple 7.1.22 On considère le modèle uniforme $\mathcal{P} = \{P_{\theta}, \theta \in]0, +\infty[\}$, où P_{θ} est la loi uniforme sur $[0, \theta]$. Si X_1 a pour loi P_{θ} , 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_{θ} , alors on peut construire à l'aide de la loi forte des grands nombres un estimateur convergent de θ :

$$\hat{\theta}_n = \frac{2}{n} \sum_{i=1}^n X_i.$$

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 θ .

7.1.5 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_θ 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 θ fixé, en tant que fonction de \mathbf{x} , c'est la loi d'un n -échantillon $(X_i)_{i=1}^n$ sous \mathbb{P}_θ .

- soit que pour tout $\theta \in \Theta$, la loi P_θ 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}_θ et $p_n(\mathbf{x}, \theta)$ la densité jointe d'un n -échantillon, qu'on appelle vraisemblance.

Définition 7.1.23 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 θ) 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 θ .

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 Θ sont finis. Avant de recueillir des données, on ne sait rien sur le paramètre θ , à part qu'il est dans Θ , donc on peut considérer que le paramètre est une variable aléatoire discrète de loi uniforme sur Θ . 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 θ . 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 θ le plus probable dans Θ sachant les données \mathbf{x} , est l'élément de Θ qui maximise $P(\theta|\mathbf{x})$, qui est aussi celui qui maximise $p_n(\mathbf{x}, \theta)$: c'est donc l'EMV.

Exemple 7.1.24 Dans le modèle de Bernoulli, $\mathcal{X} = \{0, 1\}$,

$$\mathcal{P} = \{\mathcal{B}(1, \theta), \theta \in [0, 1]\},$$

la vraisemblance est :

$$p_n(\mathbf{x}, \theta) = \prod_{i=1}^n \theta^{\mathbf{1}_{1}(x_i)} (1 - \theta)^{\mathbf{1}_0(x_i)},$$

avec la notation $\mathbf{1}_a(x) = 1$ si $x = a$ et 0 sinon. Ceci s'écrit aussi :

$$p_n(\mathbf{x}, \theta) = [\theta^{\bar{x}_n} (1 - \theta)^{1 - \bar{x}_n}]^n, \quad (7.1.8)$$

avec $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$. A \mathbf{x} fixé, en tant que fonction de θ , 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 7.1.25 Dans le modèle gaussien

$$\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 \in]0, \infty[\},$$

la vraisemblance est

$$p_n(\mathbf{x}, (\mu, \sigma^2)) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right). \quad (7.1.9)$$

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 7.1.26 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 (7.1.9). 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 (μ, σ^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 7.2.4 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 :

$$\sqrt{n} \left(\begin{pmatrix} \bar{X}_n \\ \bar{V}_n \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \right) \stackrel{loi}{=} \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^n \begin{pmatrix} X_j \\ \sigma^2 Y_j^2 \end{pmatrix} - \begin{pmatrix} 0 \\ \frac{\sigma^2 Y_1^2}{\sqrt{n}} \end{pmatrix} \right).$$

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 6.3.8), 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 7.1.27 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_θ 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}_θ , $\hat{\theta}_n$ 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 7.1.22. 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 θ . 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.

7.2 Intervalle de confiance

7.2.1 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 7.2.1 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.

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 7.2.2 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 7.1), 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 :

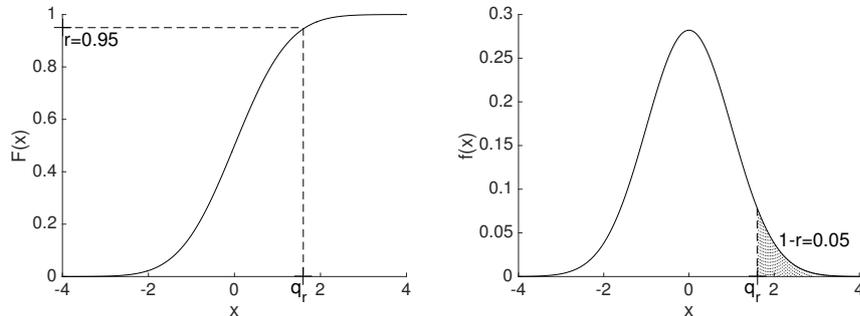


FIGURE 7.1 – Détermination du quantile q_r d'ordre $r = 0,95$ d'une loi à partir de la fonction de répartition $F(x)$ de la loi (gauche) et à partir de la densité $f(x)$ de la loi (droite). Ici on a pris le cas d'une loi gaussienne centrée réduite.

Proposition 7.2.3 *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 $\alpha \in]0, 1[$, $\mathbb{P}(X \notin [q_{\alpha/2}, q_{1-\alpha/2}]) = \mathbb{P}(X < q_{\alpha/2}) = \mathbb{P}(X > q_{1-\alpha/2}) = \alpha$.
3. Pour tout $\alpha \in]0, 1[$, $\mathbb{P}(X \in [q_{\alpha/2}, q_{1-\alpha/2}]) = \mathbb{P}(X \geq q_{\alpha/2}) = \mathbb{P}(X \leq q_{1-\alpha/2}) = 1 - \alpha$.
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 :

$$\bar{X}_n \xrightarrow[n \rightarrow +\infty]{p} \mu \text{ et } V_n \xrightarrow[n \rightarrow +\infty]{p} \sigma^2,$$

c'est-à-dire que la moyenne empirique (7.1.3) et la variance empirique non-biaisée (7.1.7) sont des estimateurs convergents de l'espérance μ et de la variance σ^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.

7.2.2 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 7.2.4 *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 (7.1.3) et (7.1.7) sont indépendantes pour tout n . De plus, pour tout n , on a :*

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1), \quad (7.2.10)$$

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{V_n}} \sim T_{n-1}, \quad (7.2.11)$$

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi_n^2, \quad (7.2.12)$$

$$(n-1) \frac{V_n}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \sim \chi_{n-1}^2. \quad (7.2.13)$$

Les lois T_n (loi de Student à n degrés de liberté) et χ_n^2 (loi de χ^2 à n degrés de liberté) sont décrites dans les tables 7.1 et 7.2. La loi de χ^2 a déjà été introduite dans la définition 6.2.7, c'est la loi de la somme des carrés de n variables aléatoires gaussiennes centrées réduites indépendantes, et elle a pour densité (6.2.19). 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 χ^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 :

$$f(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad (7.2.14)$$

où Γ est la fonction Gamma d'Euler : $\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} dx$ (voir exercice 7.4.1). 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 7.2.4 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 :

$$Y_1 = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} = \frac{1}{\sqrt{n}} \sum_{j=1}^n Z_j \quad (7.2.15)$$

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 χ_n^2 .

Considérons maintenant la variance empirique renormalisée (7.2.13). 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 (e_1, \dots, 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 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 (7.2.15). 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 6.2). 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 χ_{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 7.2.5 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 7.1.18. Celui de la variance empirique se calcule à partir de la caractérisation (7.2.13) et du fait qu'une v.a. suivant la loi χ_p^2 a pour variance $2p$. □

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 μ . 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 μ avec une probabilité qu'on se fixe.

Supposons dans un premier temps que l'écart-type σ 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 7.2.3, 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 7.2.4 :

$$\mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \in [-q_{1-\alpha/2}, q_{1-\alpha/2}]\right) = 1 - \alpha.$$

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 μ 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 σ 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 σ . Dans le cas général où on ne connaît pas σ , 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 7.2.3, 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 7.2.4 :

$$\mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sqrt{V_n}} \in [-t_{1-\alpha/2}, t_{1-\alpha/2}]\right) = 1 - \alpha.$$

En réécrivant l'événement en question, on obtient l'intervalle de confiance $I_{1-\alpha}$ pour μ 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 σ 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

$n \backslash p$	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 7.1 – 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 $T_n \sim \mathcal{N}(0, 1)$.

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 7.4.2). En pratique, on utilise des tables de loi T_n (voir la table 7.1) pour $n \leq 30$, et, lorsque $n > 30$, on utilise souvent l'approximation gaussienne. On peut aussi utiliser des logiciels de calcul scientifique.

Exemple 7.2.6 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 μ 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 μ au niveau 0,95.

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 σ^2 , lorsque la moyenne μ est connue, puis lorsqu'elle ne l'est pas.

Commençons par le cas où la moyenne μ 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 χ_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 χ_n^2 et $t_{1-\alpha,r}$ est le quantile d'ordre $1 - \alpha/2$ de la loi χ_n^2 . Mais on pourrait parfaitement proposer un intervalle asymétrique. On utilise en pratique des tables de la loi χ_n^2 (voir la table 7.2) ou des logiciels de calcul scientifique. Lorsque la moyenne μ 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^*/σ^2 suit la loi χ_n^2 d'après la proposition 7.2.4. 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 7.2.7 Dans l'exemple 7.2.6, imaginons que le fabricant ait fait des mesures extensives, et qu'il indique sur la boîte la durée de vie moyenne : $\mu = 1920$. On

$n \backslash p$	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 7.2 – Table de la loi χ_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 $\sqrt{2\chi_n^2} - \sqrt{2n-1} \sim \mathcal{N}(0, 1)$.

recherche la variance σ^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 :

$$I_{0,95} = \left[\frac{10V_{10}^*}{t_{0,95,r}}, \frac{10V_{10}^*}{t_{0,95,l}} \right].$$

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 χ_{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 σ^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 χ_n^2 (voir l'exercice 7.4.2). C'est pourquoi les tables de la loi χ_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 χ_n^2 valables pour des n plus grands, avec une précision arbitraire.

Supposons maintenant que la moyenne μ 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 μ 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 χ_{n-1}^2 d'après la proposition 7.2.4. 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 7.2.8 On reprend l'exemple 7.2.7, mais sans supposer qu'on connaît l'espérance μ . 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 χ_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 σ^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.

7.2.3 Résultats asymptotiques

Dans le cas non-gaussien, la proposition 7.2.4 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.

Intervalles de confiance asymptotiques

On cherche un intervalle de confiance pour un paramètre θ réel. On suppose que $\hat{\theta}_n$ est un estimateur convergent et asymptotiquement normal de θ , plus exactement qu’il existe une fonction $\sigma(\theta)$ de Θ 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 Φ est la fonction de répartition de la loi gaussienne centrée réduite. La fonction

réciroque Φ^{-1} de Φ est la fonction quantile de la loi gaussienne (voir la table 4.1). Les quantiles gaussiens suivantes sont particulièrement utiles :

$$\Phi^{-1}(0,975) = 1,96, \quad \Phi^{-1}(0,995) = 2,58. \quad (7.2.16)$$

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 Θ des valeurs de θ . 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 7.2.9 On appelle intervalle de confiance asymptotique au niveau $1 - \alpha$ de θ 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 7.1.27.

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 2.2.3 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 supposons 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 θ inconnu au vu des observations X_1, \dots, X_n ? Intuitivement, aucune information sur θ 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 7.2.10 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, θ) :

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

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 θ . Par conséquent la valeur de S_n contient toute l'information sur θ contenue dans le sondage. On cherche alors à estimer θ 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 θ), 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 θ , 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 Φ est la fonction de répartition de la loi gaussienne centrée réduite. D'après (7.2.16), 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 θ 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, réénonçons le résultat précédent sous la forme sous laquelle il est généralement utilisé.

Proposition 7.2.11 (Intervalle de confiance pour l'estimation de θ) *Dès que n est assez grand ($n\theta$ et $n(1-\theta) \geq 10$ en pratique) :*

$$\theta \in \left[\hat{\theta}_n - \frac{0,98 \text{ (resp. 1,29)}}{\sqrt{n}}, \hat{\theta}_n + \frac{0,98 \text{ (resp. 1,29)}}{\sqrt{n}} \right], \quad (7.2.17)$$

avec la quasi-certitude de 95% (resp. 99%).

Dans notre exemple, l'intervalle de confiance à 95% pour θ est [0,50, 0,54]. Les instituts de sondage annoncent d'ailleurs leurs résultats ainsi (sous forme de fourchette).

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 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 7.1.18 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 :

$$\begin{aligned} \sigma^2 &= \mathbb{E}(\psi(\mathbf{X})^2) - \mathbb{E}(\psi(\mathbf{X}))^2 = \int_{\mathbb{R}^d} \psi(\mathbf{x})^2 p(\mathbf{x}) d\mathbf{x} - I^2 \\ &= \int_{\mathbb{R}^d} \frac{f(\mathbf{x})^2}{p(\mathbf{x})} d\mathbf{x} - \left(\int_{\mathbb{R}^d} f(\mathbf{x}) d\mathbf{x} \right)^2. \end{aligned} \quad (7.2.18)$$

Un point remarquable de la méthode est que la variance σ^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 σ^2 définie par (7.2.18) 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 :

$$\left[\hat{I}_n - \frac{1,96\hat{\sigma}_n}{\sqrt{n}}, \hat{I}_n + \frac{1,96\hat{\sigma}_n}{\sqrt{n}} \right]. \quad (7.2.19)$$

L'intervalle (7.2.19) 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.

7.3 Tests

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 θ du modèle est-il ou non dans un sous-ensemble de Θ appelé hypothèse nulle et noté H_0 ? On retrouve cette situation fréquemment :

- Lors d'un sondage d'intension 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 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 réchauffement, alors qu'il n'y en a pas. On peut se tromper en annonçant qu'il y n'a pas de réchauffement, alors qu'il y en a.

7.3.1 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 Θ des paramètres.

On appelle test d'hypothèse une règle de décision qui, au vu de l'observation \mathbf{X} , permet de décider si θ est dans l'ensemble H_0 appelé hypothèse nulle ou si θ 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}_θ 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 7.3.1 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 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 Φ 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 7.2 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 7.3.2 Le niveau d'un test défini par sa région critique W est le nombre

$$\alpha = \sup_{\theta \in H_0} \mathbb{P}_\theta(W).$$

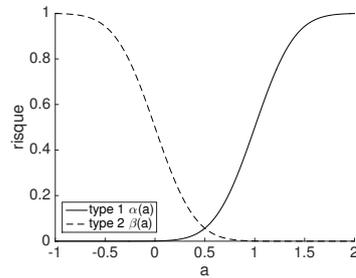


FIGURE 7.2 – Risques de première et deuxième espèce pour l'exemple 7.3.1 en fonction du seuil a avec $\mu_0 = 1$, $\mu_1 = 0$, $\sigma = 0$, et $n = 10$.

Si un test est de niveau α , alors on sait que, lorsqu'on rejette H_0 , on a au plus une probabilité α de se tromper. Parmi tous les tests de niveau inférieur à un seuil α 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 Θ 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 7.3.3 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 α 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 7.3.4 On reprend l'exemple 7.3.1. 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 α . 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.

7.3.2 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.

Tests pour la moyenne μ

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 7.2.4, 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 ζ_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 α . 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 α 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 7.3.5 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 α 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.

Tests pour la variance σ^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 7.2.4, $(n-1)V_n/\sigma^2$ suit la loi χ_{n-1}^2 sous $\mathbb{P}_{(\mu, \sigma^2)}$, ζ_n prend des valeurs de plus en plus grandes lorsque σ^2 croît. En particulier ζ_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 ζ_n est grand et on rejettera H_0 si ζ_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 χ_{n-1}^2 assure que le niveau du test est α .

7.3.3 Test du χ^2 (test du chi-deux)

Le test du χ^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".

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 7.3.6 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 7.3.7 L'EMV de $\boldsymbol{\theta}$ est donné par le vecteur des fréquences empiriques d'apparition des différentes valeurs possibles :

$$\hat{\boldsymbol{\theta}}_n = \left(\frac{N_1(\mathbf{X})}{n}, \dots, \frac{N_k(\mathbf{X})}{n} \right), \quad N_i(\mathbf{x}) = \sum_{j=1}^n \mathbf{1}_{a_i}(x_j), \quad (7.3.20)$$

avec la notation $\mathbf{1}_a(x) = 1$ si $x = a$ et 0 sinon. L'EMV est non-biaisé, convergent, asymptotiquement normal : Sous $\mathbb{P}_{\boldsymbol{\theta}}$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta})), \quad (7.3.21)$$

en loi, avec

$$C(\boldsymbol{\theta})_{jj'} = \begin{cases} \theta_j - \theta_j^2 & \text{si } j = j', \\ -\theta_j \theta_{j'} & \text{si } j \neq j'. \end{cases} \quad (7.3.22)$$

Preuve. La vraisemblance et la log-vraisemblance sont égales à :

$$\begin{aligned} p_n(\mathbf{x}, \boldsymbol{\theta}) &= \prod_{j=1}^n \left(\prod_{i=1}^k \theta_i^{\mathbf{1}_{a_i}(x_j)} \right), \\ l_n(\mathbf{x}, \boldsymbol{\theta}) &= \sum_{i=1}^k N_i(\mathbf{x}) \ln(\theta_i). \end{aligned} \quad (7.3.23)$$

La log-vraisemblance peut aussi écrire :

$$l_n(\mathbf{x}, \boldsymbol{\theta}) = n \sum_{i \in K(\mathbf{x})} \rho_i(\mathbf{x}) \ln(\theta_i), \quad (7.3.24)$$

où $K(\mathbf{x}) = \{i = 1, \dots, k : N_i(\mathbf{x}) > 0\}$ et on note

$$\boldsymbol{\rho}(\mathbf{x}) = \left(\frac{N_1(\mathbf{x})}{n}, \dots, \frac{N_k(\mathbf{x})}{n} \right). \quad (7.3.25)$$

S'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 (7.3.20).

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 (7.3.22), 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 χ^2 :

$$\zeta_n = n \sum_{j=1}^k \frac{(\hat{\theta}_{n,j} - \theta_j^{(0)})^2}{\theta_j^{(0)}}. \quad (7.3.26)$$

On obtient le comportement asymptotique suivant :

Proposition 7.3.8 Soit ζ_n défini par (7.3.26).

1) Sous H_0 , ζ_n converge en loi quand $n \rightarrow +\infty$ vers une variable aléatoire Z qui suit une loi de χ^2 à $k-1$ degrés de liberté.

2) Sous H_1 , ζ_n tend presque sûrement vers $+\infty$.

Preuve. Sous H_1 , les observations sont indépendantes et identiquement distribuées de loi \mathbb{P}_θ pour un certain $\theta \neq \theta^{(0)}$. 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}_θ -presque sûrement, $\hat{\theta}_{n,j} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{a_j}(X_i)$ converge vers θ_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 $\mathbb{P}_{\theta^{(0)}}$. D'après (7.3.21), 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 ζ_n est la loi χ_{k-1}^2 . \square

En notant $x_r(k-1)$ le quantile d'ordre r de la loi χ^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 7.3.9 Le test de région critique $W_n = \{\zeta_n > x_{1-\alpha}(k-1)\}$ est convergent de niveau asymptotique α .

Preuve. Sous H_1 , on déduit de la proposition 7.3.8 que $\mathbf{1}_{[x_{1-\alpha}(k-1), +\infty[}(\zeta_n)$ converge \mathbb{P}_θ -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 α , 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 χ_{k-1}^2 . Or, c'est vrai d'après le corollaire 6.3.6 car Z est à densité. \square

En pratique, on considère que l'approximation en loi par χ_{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 7.3.14).

Le principe du test du χ^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 7.3.8, ce qui permet de construire un test comme dans le corollaire 7.3.9.

Exemple 7.3.10 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"), ζ_{100} suit une loi de χ^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.

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 $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ 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 $\Theta = \{\boldsymbol{\theta} \in [0, 1]^k, \sum_{i=1}^k \theta_i = 1\}$. 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 7.3.11 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 $H_0 = \{\boldsymbol{\theta} \in \Theta, \exists p \in [0, 1], \theta_j = \binom{100}{j} p^j (1-p)^{100-j} \forall j = 0, \dots, 100\}$.

Pour être plus précis, on va paramétrer l'ensemble des lois qui forment l'hypothèse nulle sous la forme $H_0 = \{\boldsymbol{\theta}^{\boldsymbol{\pi}}, \boldsymbol{\pi} \in \boldsymbol{\Pi}\}$, où

- $\boldsymbol{\Pi}$ est une partie d'intérieur non vide de \mathbb{R}^h avec $h < k - 1$,
- $\boldsymbol{\pi} \mapsto \boldsymbol{\theta}^{\boldsymbol{\pi}}$ est une application de $\boldsymbol{\Pi}$ dans Θ .

Exemple 7.3.12 On reprend l'exemple 7.3.11. Ici on peut paramétrer H_0 avec l'ensemble $\boldsymbol{\Pi} = [0, 1]$ 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 $\hat{\boldsymbol{\pi}}_n$ de $\boldsymbol{\pi}$ à valeurs dans $\boldsymbol{\Pi}$ (très souvent, ce sera l'EMV de $\boldsymbol{\pi}$) et à comparer les vecteurs $\hat{\boldsymbol{\theta}}_n$ défini par (7.3.20) et $\boldsymbol{\theta}^{\hat{\boldsymbol{\pi}}_n}$. 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 7.3.13 Soit

$$\zeta_n = n \sum_{j=1}^k \frac{(\hat{\theta}_{n,j} - \theta_j^{\hat{\boldsymbol{\pi}}_n})^2}{\theta_j^{\hat{\boldsymbol{\pi}}_n}}.$$

Sous des hypothèses de régularité non-précisées (vérifiées en général lorsque $\hat{\theta}_n$ est l'EMV de θ et $\hat{\pi}_n$ est l'EMV de π) :

- Sous H_0 , ζ_n converge en loi vers $Z \sim \chi^2_{k-h-1}$.
- Sous H_1 , ζ_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 π ce qui réduit le nombre de degrés de liberté.

Finalement, en procédant comme pour le corollaire 7.3.9, on conclut que le test de région critique $W_n = \{\zeta_n \geq x_{1-\alpha}(k - h - 1)\}$ est convergent de niveau asymptotique α , avec $x_r(k - h - 1)$ le quantile d'ordre r de la loi χ^2 à $k - h - 1$ degrés de liberté.

Exemple 7.3.14 On reprend l'exemple 7.3.11-7.3.12. 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	≥ 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 7.4.4) :

$$\hat{p}_{200} = \frac{\sum_{k=1}^{200} X_k}{100 \times 200} = \frac{\sum_i i N_i}{100 \times 200} \simeq 0,05.$$

On regroupe par paquets de tailles supérieures à 5 :

i	≤ 1	2	3	4	5	6	7	8	≥ 9
N_i	8	14	29	36	41	26	17	17	12
$200\theta_i^{\hat{p}_{200}}$	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	≤ 1	2	3	4	5	6	7	8	≥ 9
$\hat{\theta}_{200,i}$	0,0400	0,0700	0,1450	0,1800	0,2050	0,1300	0,0850	0,0850	0,0600
$\theta_i^{\hat{p}_{200}}$	0,0372	0,0814	0,1399	0,1783	0,1800	0,1499	0,1058	0,0647	0,0627

On calcule

$$\zeta_{200}^{obs} = 200 \sum_{j=1}^{200} \frac{(\hat{\theta}_{200,j} - \theta_j^{\hat{p}_{200}})^2}{\theta_j^{\hat{p}_{200}}} \simeq 3,74.$$

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 7.3.15 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 Θ 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 χ^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.

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 $\boldsymbol{\theta} = (\theta_{jl}, 1 \leq j \leq d, 1 \leq l \leq m)$ où $\mathbb{P}_{\boldsymbol{\theta}}(Y_i = b_j, Z_i = c_l) = \theta_{jl}$. L'ensemble H_0 des $\boldsymbol{\theta}$ qui donnent des lois $\mathbb{P}_{\boldsymbol{\theta}}$ de forme produit correspondant à l'indépendance de Y_i et Z_i peut être paramétré par $H_0 = \{\boldsymbol{\theta}^\pi, \boldsymbol{\pi} \in \boldsymbol{\Pi}\}$, 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

$$\boldsymbol{\Pi} = \left\{ \boldsymbol{\pi} = (\pi_j^{(1)} \in [0, 1], 1 \leq j \leq d-1, \pi_l^{(2)} \in [0, 1], 1 \leq l \leq m-1), \right. \\ \left. \text{avec } \sum_{j=1}^{d-1} \pi_j^{(1)} \leq 1, \sum_{l=1}^{m-1} \pi_l^{(2)} \leq 1 \right\},$$

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). ζ_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 rejettera l'hypothèse H_0 d'indépendance si ζ_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 α si ζ_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 χ^2 à $(d-1)(m-1)$ degrés de liberté. L'exemple 7.3.15 est un cas particulier d'application de cette méthode, avec $d = m = 2$.

7.4 Exercices sur le chapitre 7

EXERCICE 7.4.1 Montrer que la densité de la loi T_n est de la forme (7.2.14).

EXERCICE 7.4.2 Approximation des lois χ_n^2 et T_n .

Soit $(Z_n)_n$ une suite de variables aléatoires réelles avec Z_n de loi χ_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 ζ_n de loi T_n .

3. Montrer que $(\zeta_n)_n$ converge en loi vers une loi gaussienne centrée réduite.

EXERCICE 7.4.3 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 7.4.4 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) .

- Déterminer un estimateur de θ par la méthode des moments.
- Donner l'Estimateur du Maximum de Vraisemblance de θ .

EXERCICE 7.4.5 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.

- 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 $\mathbb{E}_a(X^4)$.
- 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.
- 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 ?

EXERCICE 7.4.6 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 θ . A l'aide des quantiles des lois de χ^2 donnés dans la Table 7.2, préciser la mise en œuvre de cet intervalle de confiance pour $n = 10$ et $\alpha = 5\%$.

EXERCICE 7.4.7 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 7.4.8 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 (α, β) .
2. Le paramètre d'échelle β est maintenant supposé connu égal à 1. Vérifier que si X suit la loi de Pareto de paramètres α et 1, $\ln(X)$ suit la loi exponentielle de paramètre α . En déduire un estimateur de α . Montrer qu'il est convergent et construire un intervalle de confiance asymptotique de niveau $1 - \eta$ pour α .

EXERCICE 7.4.9 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 λ et μ 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 μ est connu.

3. Donner l'Estimateur du Maximum de Vraisemblance $\hat{\lambda}_n$ de λ .
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 λ 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 λ .
8. Donner l'Estimateur du Maximum de Vraisemblance $(\hat{\lambda}_n, \hat{\mu}_n)$ de (λ, μ) .

EXERCICE 7.4.10 On se place dans le modèle exponentiel $\mathcal{P} = \{\mathcal{E}(\theta), \theta \in \Theta\}$, $\Theta =]0, +\infty[$.

1. Rappeler $\mathbb{E}_\theta(X_1)$.
2. Soit $S_n = X_1 + \dots + X_n$. En remarquant que $2\theta S_n \sim \chi_{2n}^2$, construire un test de niveau α pour $H_0 = \{\theta_0\}$ et $H_1 = H_0^c$.
A.N. : $n = 15$, $\bar{x}_n = 1.47$, $\theta_0 = 1$, $\alpha = 5\%$.
3. Proposer un test pour $H_0 = [\theta_0, +\infty[$ et $H_1 =]0, \theta_0[$.

EXERCICE 7.4.11 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 8

Modèles dynamiques aléatoires

Le mystérieux éveille en nous les plus belles émotions. C'est le sentiment fondamental, berceau de l'art et de la science. Celui qui ne l'a pas éprouvé ne sait plus s'émerveiller ou se laisser surprendre, il est pour ainsi dire mort et ses yeux sont éteints.

Albert Einstein, Mein Weltbild - 1934.

Nous allons dans ce chapitre développer quelques exemples de phénomènes aléatoires qui évoluent au cours du temps. Plus précisément, nous nous intéresserons à des suites de variables aléatoires $(Y_n)_n$ qui décrivent à chaque instant n l'état d'un système aléatoire défini par récurrence. Un exemple fondamental est celui des chaînes de Markov, pour lesquelles la loi des états futurs du processus ne dépend du passé que par l'état du processus au présent. Ces processus sont fondamentaux dans de très nombreuses applications, en Physique, en Biologie, en Informatique, en Mathématiques financières...

Nous n'aborderons pas de manière systématique cette notion extrêmement riche, mais nous allons donner quelques exemples de tels processus qui sont très importants en probabilité. Ce chapitre, fondamental pour les applications, peut ainsi être vu comme une introduction à un cours sur les chaînes de Markov. (Voir par exemple les livres de Benaïm-El Karoui, Graham, Pardoux).

Soulignons que jusqu'à présent, nous avons surtout étudié des suites de variables

aléatoires indépendantes. Dans ce chapitre, les variables aléatoires que nous étudierons ne seront pas indépendantes, mais les relations de récurrence qui les définissent vont nous permettre d'en déduire des informations sur leur comportement asymptotique, quand n tend vers l'infini.

8.1 Marche aléatoire

Définition 8.1.1 Nous appelons *marche aléatoire* la suite $(S_n)_n$ définie par

$$S_n = S_0 + \sum_{i=1}^n X_i = S_{n-1} + X_n,$$

où les variables aléatoires réelles $(X_i)_{i \in \mathbb{N}}$ sont indépendantes et de même loi.

Remarquons que les variables aléatoires $S_n, n \in \mathbb{N}$ sont des variables **dépendantes**.

Ces marches aléatoires sont d'une importance fondamentale : elles modélisent le déplacement d'une particule, la promenade d'un individu, un cours de la bourse, la taille d'une population, la taille d'une file d'attente, et d'innombrables autres phénomènes aléatoires.

La marche peut se déplacer sur un réseau régulier ou sur des graphes plus complexes.

Marche aléatoire simple sur \mathbb{Z}^d

Cette marche modélise la dynamique de particules sur le réseau \mathbb{Z}^d . Elle est très utilisée en physique statistique par exemple. Notons (e_1, \dots, e_d) la base canonique de \mathbb{Z}^d . Imaginons une particule qui se déplace sur ce réseau. Sa position à l'instant n , notée par S_n , est définie par la relation de récurrence aléatoire

$$S_0 = 0 \quad ; \quad S_{n+1} = S_n + X_{n+1},$$

où les variables aléatoires X_i sont indépendantes et de même loi sur $\{e_1, \dots, e_d, -e_1, \dots, -e_d\}$. La marche aléatoire est dite symétrique si la loi des X_i est uniforme.

Problème de la ruine du joueur : Un joueur lance une pièce (éventuellement biaisée) de manière répétée. Quand la pièce tombe sur Pile, il reçoit 1 euro de son adversaire, et quand la pièce tombe sur Face, il lui donne 1 euro. Notant S_n la fortune du joueur après n lancers, on a $S_{n+1} = S_n + X_{n+1}$ avec $X_{n+1} = +1$ avec probabilité p (probabilité que la pièce tombe sur Pile) et $X_{n+1} = -1$ avec probabilité $1 - p$. Les

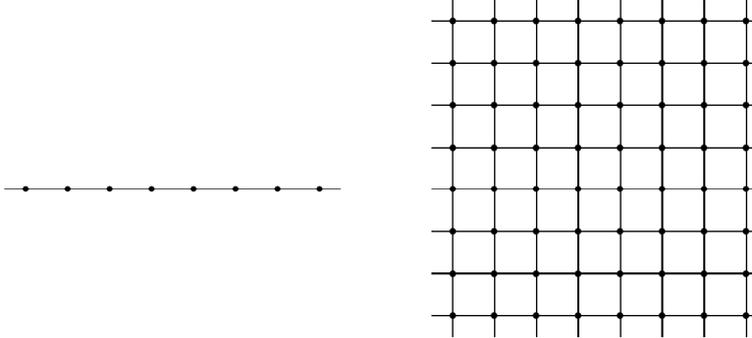


FIGURE 8.1 – Cas à une dimension : la marche peut faire un pas vers la droite avec probabilité p , un pas vers la gauche avec probabilité $1 - p$. Cas à deux dimensions : un pas en avant avec probabilité p_1 , un pas en arrière avec probabilité p_2 , un pas sur le côté avec probabilité p_3 , un pas de l'autre côté avec probabilité p_4 , telles que $p_1 + \dots + p_4 = 1$.

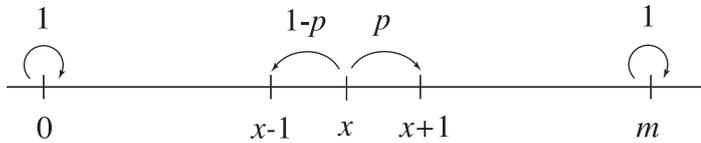
variables X_n , $n \geq 1$, sont supposées indépendantes, et la fortune initiale du joueur est une constante, $S_0 = k$. Donc

$$S_n = k + \sum_{i=1}^n X_i \quad (8.1.1)$$

est une marche aléatoire.

En réalité, dans le problème particulier de la ruine du joueur, la somme totale m d'argent mise en jeu est fixée et le jeu s'arrête dès que le joueur est ruiné, ($S_n = 0$) ou que son adversaire l'est, ($S_n = m$). Remarquons que $m - k$ est la fortune initiale de l'adversaire. La formule (8.1.1) n'est vraie que jusqu'à la fin du jeu, après quoi S_n reste constant. Ainsi cette marche aléatoire est dite absorbée aux points 0 et m . Le modèle (8.1.1) peut aussi décrire le mouvement d'un poisson qui se déplace transversalement dans une rivière. Deux filets de pêche sont placés en travers de la rivière, en position 0 et m . Si le poisson atteint l'un des filets, il est capturé et reste immobile.

L'arrêt du jeu correspond à des **barrières absorbantes** aux points 0 et m . Il est commode de représenter la dynamique de cette marche par le diagramme suivant qui indique la probabilité des différents sauts possibles.



Probabilité de ruine : Notant R l'événement “ le joueur est finalement ruiné ”, nous allons déterminer sa probabilité $\mu_k = \mathbb{P}_k(R)$ lorsque le joueur démarre avec une fortune initiale k . L'indice k de \mathbb{P}_k fait référence au point de départ de S_n . D'après les formules de conditionnement (voir Chapitre 2),

$$\mathbb{P}_k(R) = \mathbb{P}_k(R | X_1 = +1) \mathbb{P}(X_1 = +1) + \mathbb{P}_k(R | X_1 = -1) \mathbb{P}(X_1 = -1). \quad (8.1.2)$$

Si le premier lancer donne Pile, la fortune du joueur devient $k + 1$ et le jeu se déroule alors comme s'il était parti de la fortune $k + 1$. D'où

$$\mu_k = p \mu_{k+1} + (1 - p) \mu_{k-1} \quad (8.1.3)$$

pour $k \in \{1, \dots, m - 1\}$, avec les conditions aux limites $\mu_0 = 1$, $\mu_m = 0$. L'équation (8.1.3) est une récurrence linéaire, pour laquelle on cherche d'abord les solutions de la forme $\mu_k = r^k$: r doit satisfaire l'équation caractéristique

$$p r^2 - r + (1 - p) = 0,$$

dont les solutions sont $r_1 = 1$, $r_2 = (1 - p)/p$. Deux cas se présentent alors.

- i)* $p \neq 1/2$ (jeu biaisé). Les deux racines sont différentes, les solutions de (8.1.3) sont de la forme $\mu_k = \alpha r_1^k + \beta r_2^k$ et nous déterminons α et β par les conditions aux limites. Nous obtenons

$$\mathbb{P}_k(R) = \frac{\left(\frac{1-p}{p}\right)^m - \left(\frac{1-p}{p}\right)^k}{\left(\frac{1-p}{p}\right)^m - 1}, \quad 0 \leq k < m. \quad (8.1.4)$$

Procédant de même pour l'événement G : “ le joueur finit par gagner ”, nous constatons que $\mathbb{P}_k(G)$ satisfait (8.1.3), avec d'autres conditions aux limites ($\mu_0 = 0$, $\mu_m = 1$). Nous trouvons alors que $\mathbb{P}_k(G) = 1 - \mathbb{P}_k(R)$, si bien que le jeu s'arrête en un temps fini, avec probabilité 1.

- ii)* $p = 1/2$ (jeu équitale). Alors $r_1 = r_2 = 1$, les solutions de (8.1.3) sont de la forme $\mu_k = \alpha + \beta k$ et en tenant compte des conditions limites nous obtenons

$$\mathbb{P}_k(R) = 1 - \frac{k}{m}, \quad \mathbb{P}_k(G) = \frac{k}{m}, \quad (8.1.5)$$

et ici encore le jeu finit par s'arrêter.

Durée du jeu : Déterminons la durée moyenne du jeu, $V_k = \mathbb{E}_k(\nu)$ avec

$$\nu = \min\{k \geq 1 : S_n = 0 \text{ ou } m\},$$

et \mathbb{E}_k désigne l'espérance sous \mathbb{P}_k . Comme dans (8.1.2), nous conditionnons suivant les valeurs de X_1

$$\mathbb{E}_k(\nu) = \mathbb{E}_k(\nu | X_1 = 1) \mathbb{P}(X_1 = 1) + \mathbb{E}_k(\nu | X_1 = -1) \mathbb{P}(X_1 = -1)$$

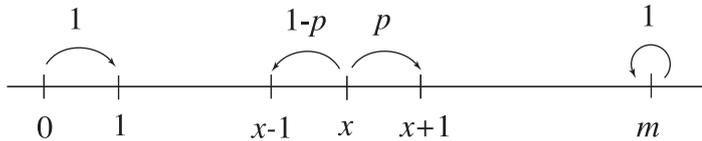
et nous remarquons que cette fois $\mathbb{E}_k(\nu | X_1 = \pm 1) = V_{k\pm 1} + 1$ puisqu'une unité de temps s'est écoulée. Nous obtenons donc l'équation de récurrence linéaire avec second membre

$$V_k = p(1 + V_{k+1}) + (1 - p)(1 + V_{k-1}) \tag{8.1.6}$$

pour $1 \leq k \leq m - 1$, avec $V_0 = V_m = 0$. Nous résolvons cette équation en trouvant solutions générales et solution particulière :

$$\begin{aligned} \mathbb{E}_k(\nu) &= \frac{1}{1 - 2p} \left(k - m \frac{1 - \left(\frac{1-p}{p}\right)^k}{1 - \left(\frac{1-p}{p}\right)^m} \right) && \text{si } p \neq 1/2, \\ &= k(m - k) && \text{si } p = 1/2. \end{aligned}$$

Barrière réfléchissante : Dans le problème précédent, nous supposons que le joueur a un oncle très riche qui lui garantit de ne pas perdre : lorsque $S_n = 0$, l'oncle lui donne 1 euro et $S_{n+1} = 1$. Le jeu s'arrête encore quand la fortune S_n du joueur atteint la valeur m . L'adversaire n'est plus nécessairement ruiné mais le capital de m euros suffit encore au joueur, qui décide d'arrêter le jeu. Nous disons ici que le point 0 est réfléchissant, tandis que le point m reste absorbant. Nous avons le diagramme suivant :



Commentaire : Un joueur dans un casino est dans la situation suivante. Il joue à un jeu défavorable contre un adversaire (presque) infiniment riche, mais qui accepte toujours de jouer tandis que le joueur peut s'arrêter à sa guise. Si le joueur a une fortune initiale de k unités, puis parie une unité à chaque jeu, et joue jusqu'à atteindre la fortune m ou la ruine, alors sa probabilité d'être ruiné est donnée par (8.1.4) ; ici $p < 1/2$. Son destin est résumé dans le tableau 8.1. Notons que, puisque le jeu est défavorable, son gain moyen est négatif même quand il a une probabilité supérieure à $1/2$ d'atteindre la fortune m .

fortune initiale (k)	fortune totale (m)	p	proba. de ruine	gain moyen	durée moyenne
9	10	0,5	0,1	0	9
9	10	0,45	0,21	-1,1	11
90	100	0,5	0,1	0	900
90	100	0,45	0,866	-76,6	756,6
90	100	0,4	0,983	-88,3	441,3
99	100	0,45	0,182	-17,2	171,8

TABLE 8.1 – Le destin du joueur

EXERCICE 8.1.2 Montrer que la durée moyenne du jeu (avec barrière réfléchissante) vérifie encore (8.1.6), mais avec d'autres conditions aux limites : $V_m = 0$, $V_0 = 1 + V_1$. Déterminer V_k .

8.2 Processus de branchement

Ces modèles sont absolument fondamentaux en écologie et pour l'étude de la dynamique des populations, dont ils décrivent l'évolution.

Dans un premier temps, nous nous intéressons à la loi de la somme d'un nombre aléatoire de variables aléatoires indépendantes.

8.2.1 Somme aléatoire de variables aléatoires indépendantes

Il arrive fréquemment, dans la modélisation probabiliste, d'avoir à considérer des sommes de variables aléatoires dont le nombre de termes lui-même est aléatoire.

Nous pouvons par exemple considérer

- le nombre de filles d'une famille $\sum_{i=1}^{\nu} \mathbf{1}_{F_i}$ puisque le sexe de chaque enfant ainsi que le nombre total d'enfants ν sont aléatoires (F_i désigne l'événement : le $i^{\text{ème}}$ enfant est une fille) ;
- le temps de service journalier $\sum_{i=1}^{\nu} X_i$ effectué par un organe de service lorsque les temps de service X_i d'une part et le nombre ν de clients d'autre part sont aléatoires.

La fonction génératrice d'une telle somme s'obtient très simplement lorsque les variables aléatoires en jeu sont toutes indépendantes et à valeurs entières.

Proposition 8.2.1 *Si $(X_n, n \geq 1)$ est une suite infinie de variables aléatoires entières positives, indépendantes et de même loi donnée par la fonction génératrice*

$$g(x) = \mathbb{E}(x^{X_n}) \quad (x \in [0, 1], n \geq 1)$$

et si ν est une variable aléatoire entière positive, indépendante de la suite $(X_n, n \geq 1)$, (c'est-à-dire indépendante du vecteur (X_1, \dots, X_n) quel que soit n), de fonction génératrice

$$G(x) = \mathbb{E}(x^\nu),$$

alors la fonction génératrice de la somme

$$S_\nu = \sum_{m=1}^{\nu} X_m \quad (S = 0 \text{ si } \nu = 0) \quad (8.2.7)$$

des ν premières variables aléatoires X_m , est la fonction composée de g et G , soit

$$\mathbb{E}(x^{S_\nu}) = G \circ g(x) \quad (\forall x, x \in [0, 1]). \quad (8.2.8)$$

Preuve. Soient $S_n = \sum_{m=1}^n X_m$, ($n \geq 1$), les sommes partielles des X_m . Nous supposons par convention que $S_0 = 0$. Alors la variable aléatoire S s'écrit explicitement sur Ω

$$S(\omega) = S_{\nu(\omega)}(\omega) \quad (8.2.9)$$

et dépend donc de deux manières de ω .

Comme les ensembles $(\{\nu = n\}, n \in \mathbb{N})$ forment une partition de Ω , nous pouvons écrire

$$\begin{aligned} \mathbb{E}(x^{S_\nu}) &= \sum_n \mathbb{E}(x^{S_\nu} \mathbf{1}_{\{\nu=n\}}) = \sum_n \mathbb{E}(x^{S_n} \mathbf{1}_{\{\nu=n\}}) \\ &= \sum_n \mathbb{E}(x^{S_n}) \mathbb{P}(\nu = n) \quad (\text{par indépendance de } S_n \text{ et de } \nu) \\ &= \sum_n g(x)^n \mathbb{P}(\nu = n) \quad (\text{par indépendance des } X_i, i \leq n) \\ &= G(g(x)). \end{aligned}$$

□

Exemple 8.2.2 Sous les hypothèses d'indépendance précédentes, si ν suit une loi de Poisson de paramètre θ et si les X_m suivent une loi de Bernoulli de paramètre p , on a

$$\mathbb{E}(x^\nu) = \exp(\theta(x-1)) \quad \text{et} \quad \mathbb{E}(x^{X_m}) = xp + (1-p)$$

de sorte que

$$\mathbb{E}(x^{S_\nu}) = \exp(\theta(xp + (1-p) - 1)) = \exp(\theta p(x-1)),$$

ce qui montre que S_ν suit une loi de Poisson de paramètre θp . Nous avons déjà obtenu ce résultat dans l'exemple 3.6.8

8.2.2 Processus de branchement

Les premières études marquantes sur les processus de branchement, dues à Galton et Watson (1875), ont été en particulier motivées par l'étude de la survivance des noms de familles (de nobles anglais), et donc de la non-extinction des descendants mâles qui seuls transmettaient le nom. Ces processus avaient déjà été introduits par Bienaymé (1845). Ils sont connus sous le nom de processus de Galton-Watson ou processus de Bienaymé-Galton-Watson.

De nombreux phénomènes d'évolution de population peuvent en fait être modélisés par ce type de processus, dont la description est simple, mais qui présentent néanmoins un comportement non trivial. Le modèle peut aussi bien décrire la croissance d'une population de cellules, ou du nombre de neutrons dans un réacteur, ou encore d'une épidémie dans une population.

Nous étudions l'évolution du nombre Z_n d'individus de cette population au cours de ses générations successives $n = 0, 1, 2, \dots$ en supposant que chacun des Z_n individus de la $n^{\text{ième}}$ génération engendre un nombre aléatoire Y_i^n d'enfants ($1 \leq i \leq Z_n$), au temps n , de sorte que

$$Z_{n+1} = \sum_{i=1}^{Z_n} Y_i^n \quad (n \geq 0). \quad (8.2.10)$$

Les variables aléatoires Y_i^n sont supposées indépendantes entre elles et de même loi de probabilité $(p_k, k \in \mathbb{N})$, caractérisée par sa fonction génératrice $g(x) = \mathbb{E}(x^Y)$. Cette loi de Y est appelée la loi de reproduction du processus de Galton-Watson. Dans la suite, nous excluons le cas inintéressant où Y serait égale à 1 avec probabilité un, c'est-à-dire le cas où $g(x) \equiv x$.

La figure 8.2 représente les individus des générations 0 à 3, lorsque $Z_0 = 1$, $Y_1^0 = 2$, $Y_1^1 = 3$, $Y_2^1 = 1$, $Y_1^2 = 2$, $Y_2^2 = 0$, $Y_3^2 = 1$, $Y_4^2 = 3$. Cette figure représente un arbre aléatoire.

Exemple 8.2.3 (i) Division cellulaire. La reproduction est binaire ou la cellule meurt : $\mathbb{P}(Y = 0) = \mathbb{P}(Y = 2) = \frac{1}{2}$. Alors $g(x) = p_0 + p_2x^2$.

(ii) Un chaîne de N nucléotides d'ADN (Voir Delmas-Jourdain). En une unité de temps, la chaîne est répliquée. Chaque nucléotide est copié de façon correcte avec probabilité q . La chaîne est détruite avec probabilité p ou donne naissance à deux molécules avec probabilité $1-p$. Le processus comptant à chaque génération le nombre de chaînes correctes est un processus de Bienaymé-Galton-Watson de loi de reproduction $p_0 = p$ (destruction), $p_1 = (1-p)(1-q^N)$ (non destruction mais réplication incorrecte), $p_2 = (1-p)q^N$ (non destruction mais réplication correcte) et $p_k = 0$ pour $k \geq 3$.

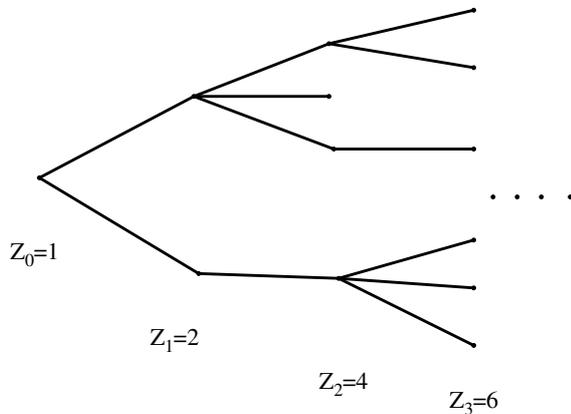


FIGURE 8.2 – L'arbre généalogique du processus de branchement

La Proposition 8.2.1 montre alors que la fonction génératrice $G_n(x) = G_{Z_n}(x)$ de Z_n vérifie pour tout $x \in [0, 1]$,

$$G_{n+1}(x) = G_{Z_{n+1}}(x) = G_n(g(x)) \quad (n \geq 0) \tag{8.2.11}$$

et si nous supposons que $Z_0 = 1$ de sorte que $G_0(x) \equiv x$, nous trouvons que pour tout $n \geq 1$:

$$G_n(x) = g_n(x) \quad \text{où} \quad g_n = g \circ g \circ \dots \circ g \quad n \text{ fois} . \tag{8.2.12}$$

Ce résultat va déjà nous permettre d'étudier le comportement asymptotique de la suite $(Z_n)_{n \geq 0}$ et de calculer notamment la probabilité d'extinction de la population.

Remarquons que $Z_{n+1} = 0$ si $Z_n = 0$ dans le modèle précédent et donc que les événements $\{Z_n = 0\}$ croissent avec n . L'événement $A =$ "Extinction de la population" est donc naturellement défini par

$$A = \cup_n \uparrow \{Z_n = 0\} \tag{8.2.13}$$

et sa probabilité donnée par

$$\mathbb{P}(A) = \lim_n \uparrow g_n(0).$$

(puisque $\mathbb{P}(Z = 0) = G_Z(0)$ pour toute variable aléatoire Z entière positive).

Deux cas se présentent, comme nous allons le voir dans le théorème suivant.

Théorème 8.2.4

(i) Si le nombre moyen d'enfants $\mathbb{E}(Y)$ de chaque individu est inférieur ou égal à 1, la population s'éteint presque sûrement : $\mathbb{P}(A) = 1$.

(ii) Si $\mathbb{E}(Y) > 1$, la probabilité d'extinction $\mathbb{P}(A)$ est égale au nombre positif v strictement inférieur à 1, solution unique de $g(v) = v$ dans $[0, 1[$.

Remarque 8.2.5 Par indépendance des variables $(Y_i^n; n \geq 1, i \geq 1)$ nous déduisons de (8.2.10) et de l'exercice 8.5.2 que $\mathbb{E}(Z_n) = \mathbb{E}(Y) \times \mathbb{E}(Z_{n-1}) = (\mathbb{E}(Y))^n$ (par récurrence). Cette dernière égalité entraîne que la population s'éteint lorsque $\mathbb{E}(Y) < 1$. En effet, il est facile de voir que

$$\mathbb{P}(Z_n \neq 0) \leq \mathbb{E}(Z_n) = (\mathbb{E}(Y))^n. \quad (8.2.14)$$

Dans le théorème 8.2.4, nous obtenons un résultat plus complet.

Preuve. La démonstration du théorème repose sur le lemme analytique suivant.

Lemme 8.2.6 a) Si $g'(1) \leq 1$, alors $g_n(x)$ croît vers 1 lorsque $n \nearrow \infty$ pour tout $x \in [0, 1]$.

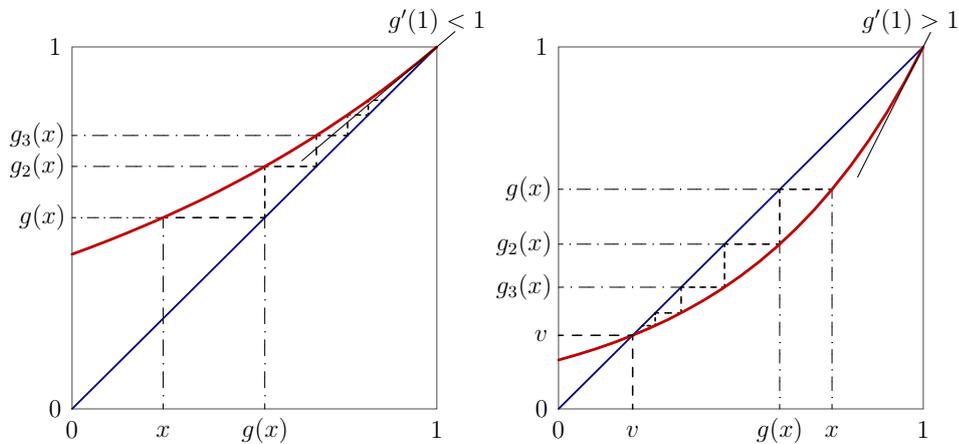
b) Si $g'(1) > 1$, l'équation $g(v) = v$ possède une solution unique dans $[0, 1[$ et $g_n(x)$ croît vers v (resp. décroît) lorsque $n \nearrow \infty$, pour tout $x \in [0, v]$, (resp. tout $x \in [v, 1[$).

Preuve. du lemme.

- a) L'application $x \rightarrow g(x)$ de l'intervalle $[0, 1]$ dans lui-même est croissante et strictement convexe car les dérivées

$$g'(x) = \mathbb{E}(Yx^{Y-1}) \quad \text{et} \quad g''(x) = E(Y(Y-1)x^{Y-2})$$

sont positives en tant qu'espérances de variables aléatoires réelles discrètes positives. De plus $g(1) = 1$. Comme nous avons exclu le cas $g(x) \equiv x$, la courbe g ne coupe pas ou au contraire coupe la diagonale du carré $[0, 1]^2$ en un point distinct de $(1, 1)$, selon que $g'(1) \leq 1$ ou que $g'(1) > 1$; ceci se voit bien sur la figure 8.3. Ainsi, selon le cas, l'équation de point fixe $g(v) = v$ n'a pas de solution ou possède une unique solution dans $[0, 1[$.

FIGURE 8.3 – La fonction g et ses itérées dans les deux cas $\mathbb{E}(Y) \leq 1$ et $\mathbb{E}(Y) > 1$

- b) Lorsque $g'(1) \leq 1$, nous avons $x \leq g(x)$ et donc $g_n(x) \leq g_{n+1}(x)$ (puisque $g_{n+1} = g \circ g_n$) pour tout x . La limite $\lim_n g_n(x)$ est inférieure à 1 et solution de $g(u) = u$, elle ne peut alors valoir que 1.
- c) De même, si $g'(1) > 1$, nous avons $x \leq g(x) \leq v$ ou $x \geq g(x) \geq v$ selon que $x \leq v$ ou que $x \geq v$; il s'en suit que $g_n(x)$ croît (resp. décroît) avec n selon le cas. La limite $\lim_n g_n(x)$, qui est solution de $g(u) = u$ et est strictement inférieure à 1, (du moins si $x \neq 1$), est alors nécessairement égale à v .

□

Ainsi donc, en appliquant les résultats du lemme, nous obtenons immédiatement que

1) si $\mathbb{E}(Y) = g'(1) \leq 1$, alors, grâce à (8.2.13),

$$\mathbb{P}(A) = \lim_n g_n(0) = 1 .$$

2) si $\mathbb{E}(Y) = g'(1) > 1$, alors

$$\mathbb{P}(A) = \lim_n g_n(0) = v .$$

□

Remarque 8.2.7 Lorsque $\mathbb{E}(Y) = 1$, alors d'une part $\mathbb{E}(Z_n) = 1$, et d'autre part Z_n converge presque-sûrement vers 0. Nous en déduisons un contre-exemple pour le

théorème de convergence dominée, où il y a convergence presque-sûre et pas convergence des espérances (ceci est dû au fait que les Z_n ne sont pas bornées). Par ailleurs, si $\mathbb{E}(Y) > 1$, alors $\mathbb{E}(Z_n) = (\mathbb{E}(Y))^n$ tend vers l'infini quand n tend vers l'infini alors que l'ensemble où Z_n tend vers 0 est de probabilité strictement positive. Là-aussi, nous avons un résultat qui peut paraître contraire à l'intuition.

Remarque 8.2.8 Lorsque $\mathbb{E}(Y) > 1$, l'extinction n'est pas certaine. Que se passe-t-il lorsque le processus survit ? En fait, la population "explose" lorsqu'elle ne s'éteint pas, c'est à dire que $\mathbb{P}(Z_n \rightarrow \infty | A^c) = 1$.

En effet, pour tout entier $a > 0$ et quel que soit x que l'on se sera fixé dans $]0, 1[$, nous obtenons, en utilisant l'inégalité de Markov (4.7.49),

$$\begin{aligned} \mathbb{P}(0 < Z_n \leq a) &= \mathbb{P}(Z_n \leq a | Z_n > 0) \mathbb{P}(Z_n > 0) = \mathbb{P}(x^{Z_n} \geq x^a | Z_n > 0) \mathbb{P}(Z_n > 0) \\ &\leq x^{-a} \mathbb{E}(x^{Z_n} | Z_n > 0) \mathbb{P}(Z_n > 0) = x^{-a} \sum_{k \geq 1} x^k \mathbb{P}(Z_n = k) \\ &= x^{-a} (g_n(x) - g_n(0)) \end{aligned}$$

qui tend vers 0 lorsque n tend vers l'infini. Il en résulte déjà la propriété plus faible que $\mathbb{P}(0 < Z_n \leq a | A^c) \xrightarrow[n \rightarrow \infty]{} 0$, $\forall a > 0$.

En outre une analyse plus détaillée de la convergence des g_n montre que cette convergence est exponentielle, i.e. $|g_n(x) - v| \leq c^n |x - v|$ pour $c = g'(v)$ et x inférieur à v . Comme $g'(v) < 1$, cela implique que $\sum_n \mathbb{P}(0 < Z_n \leq a) < \infty$, et le lemme de Borel-Cantelli (Théorème 2.5.14) montre que presque-sûrement, $Z_n = 0$ ou $Z_n > a$ pour tout n à partir d'un certain rang. Comme $Z_n = 0$ implique que $Z_{n'} = 0$ pour tout $n' > n$, deux cas seulement sont possibles : ou bien $Z_n = 0$ à partir d'un certain rang (c'est l'extinction), ou bien $Z_n > a$ à partir d'un certain rang et comme a est arbitraire : $Z_n \rightarrow \infty$ (c'est l'explosion). Ces deux événements complémentaires ont pour probabilités respectives v et $1 - v$.

Signalons pour terminer que :

- a) lorsque $\mathbb{E}(Y) < 1$ (processus de branchement sous-critique), la probabilité de " non extinction à l'instant n " tend vers 0 à une vitesse géométrique puisque, comme nous l'avons vu,

$$\mathbb{P}(Z_n \neq 0) \leq (\mathbb{E}(Y))^n.$$

- b) lorsque $\mathbb{E}(Y) = 1$ (processus de branchement critique), cette même probabilité tend beaucoup plus lentement vers zéro. Plus précisément, on peut montrer que si $\mathbb{E}(Y^2) < \infty$,

$$\mathbb{P}(Z_n \neq 0) \sim c/n \quad \text{lorsque } n \rightarrow \infty$$

avec $c = 2/\mathbb{E}(Y(Y - 1))$. (Voir exercice 8.5.4).

c) lorsque $\mathbb{E}(Y) > 1$ (processus de branchement sur-critique), la taille de la population Z_n explose à une vitesse géométrique lorsqu'il n'y a pas extinction, du moins si $\mathbb{E}(Y^2) < \infty$ car dans ce cas, on peut montrer que

$$Z_n / (\mathbb{E}(Y))^n \rightarrow W$$

lorsque $n \rightarrow \infty$ pour une variable aléatoire W strictement positive sur A^c . (cf. Benaïm-El Karoui).

8.2.3 Percolation sur un arbre

Les modèles de percolation sont utilisés pour décrire la matière molle ou désordonnée : verres en physique, gels et polymères en chimie, ...

Le modèle suivant rend compte d'une pierre poreuse plongée dans l'eau. L'eau ne peut atteindre le centre de la pierre qu'en empruntant une suite de canaux (ou microfissures). Chaque canal est identifié à une arête d'un arbre binaire. Chaque canal peut être "ouvert" ou "fermé", et on note p la proportion de canaux ouverts (i.e., de largeur suffisante pour permettre le passage de l'eau). La question naturelle est de savoir si l'eau va pénétrer jusqu'au coeur de la pierre ou non.

Posons plus précisément le modèle. Considérons l'arbre binaire infini A , dont les arêtes sont numérotées par les suites finies non vides $u = w_1w_2 \dots w_n$ de longueur $|u| = n$ ($n \geq 1$, $w_i \in \{a, b\}$). Dans la figure 8.4, nous avons représenté l'arbre jusqu'à la profondeur 3, c'est-à-dire le sous-ensemble A_3 des arêtes numérotées par les suites de longueur inférieure ou égale à 3. On dit que l'arête u précède l'arête v dans l'arbre si u est un début de v (c'est-à-dire, si v est de la forme $v = uu'$), et nous écrivons alors $u \leq v$; nous noterons $u < v$ si $u \leq v$, $u \neq v$.

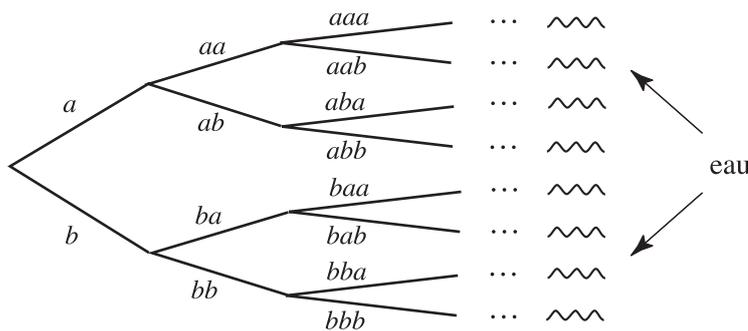


FIGURE 8.4 – L'arbre A_3

A chaque arête $u \in A$ est associée une variable aléatoire binaire X_u , et nous supposons ces variables aléatoires $(X_u)_{u \in A}$ indépendantes et de même loi donnée par $\mathbb{P}(X_u = 1) = p = 1 - \mathbb{P}(X_u = 0)$, où $p \in]0, 1[$. L'arête u est "ouverte" si $X_u = 1$, et permet alors le passage de l'eau ; elle est "fermée" sinon. Nous définissons alors

$$S_u = \sum_{v \leq u} X_v, \quad u \in A,$$

$$C_n = \{u \in A; |u| = n, S_u = n\}, \quad C = \bigcup_{n \geq 1} C_n.$$

L'ensemble C (resp. C_n) est constitué des arêtes de l'arbre (respectivement des arêtes de l'arbre à la profondeur n), qui sont ouvertes et reliées à la racine (le centre de la pierre) par des arêtes ouvertes. Notons que le centre de la pierre sera mouillé si et seulement si C est infini.

Ce modèle est un modèle de branchement. Le cardinal $Z_n = |C_n|$ de C_n est la taille de la population des descendants de la $n^{\text{ième}}$ génération de l'ancêtre (ici, la racine de l'arbre). La loi de reproduction est une loi binomiale $\mathcal{B}(2, p)$, puisque d'un noeud de l'arbre partent 2 canaux en direction de l'eau : ce noeud aura donc 0, 1 ou 2 enfants avec probabilité $(1-p)^2$, $2p(1-p)$, ou p^2 . Sa fonction génératrice vaut donc $g(x) = (1-p+px)^2$ comme nous l'avons vu précédemment. L'ensemble d'extinction correspondant est $\{|C| < \infty\}$.

L'équation $g(x) = x$ a pour racines dans $[0, 1]$:

(i) si $p \leq 1/2$, $x = 1$

(ii) si $p > 1/2$, $x = v$ et $x = 1$ avec $v = \left(\frac{1-p}{p}\right)^2 < 1$.

D'après l'étude précédente, nous concluons que :

(i) si $p \leq 1/2$, l'ensemble aléatoire C des arêtes ouvertes connectées au centre par un chemin d'arêtes ouvertes, est fini avec probabilité un. Le centre de la pierre restera sec.

(ii) si $p > 1/2$, avec probabilité strictement positive, l'ensemble C est infini et l'eau cheminera jusqu'au centre de la pierre.

Nous pouvons aussi montrer que dans ce cas, avec probabilité un, l'eau atteint un nombre infini de noeuds de l'arbre : si elle ne chemine pas nécessairement jusqu'au centre, l'eau atteint cependant des points "voisins" du centre.

Dans le cas $p > 1/2$, on dit qu'il y a **percolation**. Ce terme évoque le percolateur à café, où l'eau se fraie un passage (percole) entre les particules de café moulu jusqu'à la tasse.

8.3 Files d’attente

Les modèles de files d’attente décrivent des unités de service où se présentent des clients selon un flux aléatoire : un serveur à son guichet, un standard téléphonique, un canal de transmission, etc, jusqu’aux cas plus complexes des réseaux de communication (ethernet, internet, ...). Nous nous intéressons à la fluidité du fonctionnement de l’unité (stabilité de la queue, taille, engorgement) en fonction du flux d’arrivée et de la durée de service.

8.3.1 Un modèle simple en temps discret

À chaque instant $n \in \mathbb{N}$, un serveur unique délivre un service de durée 1 au premier client éventuellement en attente. Le flot d’arrivée des clients est homogène, et les nombres Y_n de clients se présentant dans les intervalles de temps $]n - 1, n]$ sont des variables aléatoires entières de même loi, que l’on supposera indépendantes et intégrables. Le nombre X_n de clients présents dans la queue à l’instant $n \in \mathbb{N}$, qui est donc la taille de la file d’attente, évolue selon la dynamique suivante :

si $X_n = 0$ alors $X_{n+1} = Y_{n+1}$, et

si $X_n \geq 1$ alors $X_{n+1} = (X_n - 1) + Y_{n+1}$.

Notant $a^+ = \max(a, 0)$ la partie positive de $a \in \mathbb{R}$, X_n satisfait donc la relation de récurrence

$$X_{n+1} = (X_n - 1)^+ + Y_{n+1} \quad , \quad n \in \mathbb{N}, \quad (8.3.15)$$

qui la détermine entièrement en fonction des Y_n et de l’état initial X_0 que l’on supposera indépendant de la suite $(Y_n)_n$.

Si q est la loi commune des Y_n , c’est-à-dire que $\mathbb{P}(Y_n = i) = q(i)$ pour $i \in \mathbb{N}$, la loi conditionnelle de X_{n+1} sachant X_n est donnée par

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \begin{cases} q(j) & \text{si } i = 0, \\ q(j - i - 1) & \text{si } i \geq 1 \end{cases} \quad (8.3.16)$$

(on convient que $q(i) = 0$ si $i \notin \mathbb{N}$).

Pour le bon fonctionnement de la file d’attente, il est nécessaire que la file ne s’engorge pas. Nous étudions tout d’abord la finitude de la variable aléatoire

$$T_0 = \min\{n \geq 1 : X_n = 0\} \in \mathbb{N} \cup \{+\infty\},$$

qui est le premier instant où la file se vide. Remarquons que $T_0 = +\infty$ si $X_n > 0$ pour tout temps $n \geq 1$.

Ici et dans toute la suite, pour $i \in \mathbb{N}$, la notation \mathbb{P}_i désignera la loi de la file d'attente partant de l'état initial constant $X_0 = i$, c'est-à-dire une probabilité sur l'espace de probabilité Ω telle que

$$\mathbb{P}_i(X_0 = i) = 1.$$

Remarquons que si la file démarre vide ($X_0 = 0$) ou avec un client en attente ($X_0 = 1$), alors elle sera la même à l'instant 1, et donc aussi aux instants ultérieurs, de sorte que nous avons

$$\mathbb{P}_0(T_0 < \infty) = \mathbb{P}_1(T_0 < \infty).$$

Appelons v cette probabilité que la file se vide : $v = \mathbb{P}_0(T_0 < \infty) = \mathbb{P}_1(T_0 < \infty)$.

Puisque trivialement $a^+ \geq a$, il résulte de (8.3.15) que $X_{n+1} \geq X_n - 1 + Y_{n+1}$, pour tout $n \geq 0$, et par sommation,

$$X_n \geq \sum_{k=1}^n Y_k - n + X_0. \quad (8.3.17)$$

Nous en déduisons que

$$\frac{X_n}{n} \geq \frac{1}{n} \sum_{k=1}^n Y_k - 1 + \frac{X_0}{n}. \quad (8.3.18)$$

Or, la loi des grands nombres montre que, avec probabilité 1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Y_k = \mathbb{E}(Y_1).$$

Ainsi, presque-sûrement, $\liminf_n \frac{X_n}{n}$ va être supérieure à $\mathbb{E}(Y_1) - 1$. Le signe de cette différence va donc jouer un grand rôle. Rappelons que le paramètre $\rho = \mathbb{E}(Y_1)$ représente la charge moyenne de travail par unité de temps pour le serveur.

Lorsque $\mathbb{E}(Y_1) > 1$, la file d'attente est instable, elle s'engorge :

$$\rho = \mathbb{E}(Y_1) > 1 \implies \lim_{n \rightarrow \infty} X_n = +\infty \quad \text{avec probabilité 1.}$$

D'autre part, tant que $n < T_0$, nous avons $(X_n - 1)^+ = X_n - 1$ et donc égalité dans (8.3.17). Nous en déduisons que lorsque $\rho < 1$, le membre de droite de (8.3.17) tend vers $-\infty$ d'après la loi des grands nombres, et l'événement $\{T_0 = \infty\} \subset \left\{ \sum_{k=1}^n Y_k - n + X_0 \geq 1, \forall n \geq 1 \right\}$ est de probabilité nulle. Ainsi la file reste stable dans ce cas :

$$\rho = \mathbb{E}(Y_1) < 1 \implies T_0 < \infty \quad \text{avec probabilité 1.}$$

Le cas critique $\rho = 1$ est plus difficile. Il est décrit par le résultat plus précis suivant.

8.3.2 Stabilité : étude analytique

Théorème 8.3.1 *La probabilité que la file d'attente (8.3.15) démarrante de $X_0 = 0$ ou 1 finisse par se vider, qui vaut*

$$v = \mathbb{P}_0(T_0 < \infty) = \mathbb{P}_1(T_0 < \infty),$$

est la plus petite solution de $g(v) = v$ dans $[0, 1]$, où g est la fonction génératrice de Y_1 ,

$$g(x) = \mathbb{E}(x^{Y_1}) = \sum_{i \geq 1} x^i q(i).$$

En particulier, $v = 1$ si et seulement si $\rho = \mathbb{E}(Y_1) \leq 1$.

Preuve. Chaque période d'activité du serveur peut se décrire à l'aide d'un processus de branchement. La file démarrante de $X_0 = 1$, le client présent à l'instant $n = 0$ est l'ancêtre du processus de branchement. Ses "enfants" sont les Y_1 clients arrivant dans la file pendant son service; ils constituent la première génération ($Z_1 = Y_1$). Plus généralement, le client C_2 est un "enfant" du client C_1 si C_2 arrive dans la file pendant le service de C_1 . Nous définissons récursivement la $(N + 1)$ -ième génération comme l'ensemble des enfants d'individus de la N -ième génération. Les tailles $Z_0 = 1, Z_1, Z_2, \dots, Z_N$ des générations successives constituent alors un processus de branchement dont la loi de reproduction est $q = (q(i), i \in \mathbb{N}^*)$. Nous pouvons écrire $Z_0 = 1, Z_1 = Y_1$, et pour $N \geq 0$

$$Z_{N+1} = \sum_{Z_0 + \dots + Z_{N-1} + 1 \leq k \leq Z_0 + \dots + Z_N} Y_k, \quad (8.3.19)$$

(avec la convention que la somme sur l'ensemble vide est 0). En particulier, $Z_2 = \sum_{k=2}^{Z_1+1} Y_k$.

L'expression (8.3.19) n'est pas exactement de la forme (8.2.10) trouvée pour le modèle de branchement, mais elle définit bien la même loi. Par exemple, la probabilité

$$\mathbb{P}(Z_1 = i, Z_2 = j) = \mathbb{P}(Z_2 = j | Z_1 = i) \mathbb{P}(Z_1 = i) = \mathbb{P}(Y_1 = i) \times \mathbb{P}\left(\sum_{k=2}^{i+1} Y_k = j\right)$$

est la même pour (8.3.19) et pour (8.2.10). Plus généralement, la suite $(Z_N, N \geq 1)$ a même loi que le processus de branchement. Nous décrivons ainsi toute la période d'activité du serveur, c'est-à-dire la période de service ininterrompu débutant au temps $n = 0$.

Le processus de branchement Z s'éteint si et seulement si la file finit par se vider. Si A désigne l'événement aléatoire "le processus de branchement Z s'éteint", alors

la probabilité v que la file se vide est égale à $v = \mathbb{P}_1(T_0 < \infty) = \mathbb{P}(A)$. D'après le théorème 8.2.4, v est donc la plus petite solution dans $[0, 1]$ de l'équation $v = g(v)$. De même, $v = 1 \iff \rho \leq 1$.

Ainsi le serveur finit par se reposer si $\rho \leq 1$. Il y a cependant une différence importante entre les cas $\rho < 1$ et $\rho = 1$. Remarquant que la taille totale de la lignée est égale à la durée de la période d'activité, soit

$$\sum_{N \geq 0} Z_N = T_0 ,$$

et rappelant que $\mathbb{E}(Z_N) = ((\mathbb{E}(Y_1))^N = \rho^N$, nous constatons que la durée moyenne de la période d'activité n'est finie que si $\rho < 1$

$$\mathbb{E}(T_0) < \infty \iff \rho = \mathbb{E}(Y_1) < 1 .$$

Le cas $\rho = 1$, stable au sens où $v = 1$, ne l'est que dans un sens faible puisque la durée moyenne de la période d'activité est infinie. Le serveur dans ce cas doit s'attendre à des journées de travail ininterrompu, et au mécontentement de ses clients en attente dans la file! \square

Preuve. Voici une autre preuve du théorème 8.3.1, qui n'utilise pas le processus de branchement.

(i) En conditionnant suivant les valeurs de Y_1

$$\begin{aligned} \mathbb{P}_1(T_0 < \infty) &= \sum_{i \geq 0} \mathbb{P}_1(T_0 < \infty, Y_1 = i) \\ &= \sum_{i \geq 0} \mathbb{P}_1(T_0 < \infty | Y_1 = i) q(i) \\ &= q(0) + \sum_{i \geq 1} P_i(T_0 < \infty) q(i) . \end{aligned} \quad (8.3.20)$$

En effet, les $(Y_k, k \geq 1)$ étant indépendants et de même loi, la loi de (X_2, X_3, \dots) conditionnée en $X_1 = i$ (ou $Y_1 = i$) est la même que celle de (X_1, X_2, \dots) démarrant de $X_0 = i$.

Introduisons, pour tout $i \in \mathbb{N}$, le temps d'atteinte T_i de l'état i , défini par $T_i = \min \{n \geq 1 : X_n = i\}$. Montrons que l'on a :

$$\mathbb{P}_i(T_{i-1} < \infty) = v \quad \forall i \geq 1 , \quad (8.3.21)$$

$$\mathbb{P}_i(T_0 < \infty) = \mathbb{P}_i(T_{i-1} < \infty) \times \mathbb{P}_{i-1}(T_{i-2} < \infty) \times \dots \times \mathbb{P}_1(T_0 < \infty) . \quad (8.3.22)$$

Pour montrer (8.3.21), nous remarquons que pour la file commençant à $i \geq 1$, $T_{i-1} = n$ si et seulement si

$$\begin{aligned} n &= \min \{k \geq 1 : i + (Y_1 - 1) + \cdots + (Y_k - 1) = i - 1\} \\ &= \min \{k \geq 1 : Y_1 + \cdots + Y_k = k - 1\}. \end{aligned}$$

Ainsi $\mathbb{P}_i(T_{i-1} = n)$ ne dépend pas de i , et donc $\mathbb{P}_i(T_{i-1} = n) = \mathbb{P}_1(T_0 = n)$, ce qui entraîne que $\mathbb{P}_i(T_{i-1} < \infty) = v$, en sommant sur n . Nous établissons (8.3.22) par récurrence, en montrant que $\mathbb{P}_i(T_0 < \infty) = \mathbb{P}_i(T_{i-1} < \infty) \times \mathbb{P}_{i-1}(T_0 < \infty)$, $\forall i \geq 2$. Pour cela nous écrivons

$$\begin{aligned} \mathbb{P}_i(T_0 < \infty) &= \sum_{j=1}^{\infty} \mathbb{P}_i(T_0 < \infty, T_{i-1} = j) \\ &= \sum_{j=1}^{\infty} \mathbb{P}_i(T_0 < \infty | T_{i-1} = j) \mathbb{P}_i(T_{i-1} = j) \\ &= \sum_{j=1}^{\infty} \mathbb{P}_{i-1}(T_0 < \infty) \mathbb{P}_i(T_{i-1} = j) \\ &= \mathbb{P}_{i-1}(T_0 < \infty) \sum_{j=1}^{\infty} \mathbb{P}_i(T_{i-1} = j) = \mathbb{P}_{i-1}(T_0 < \infty) \mathbb{P}_i(T_{i-1} < \infty) \end{aligned}$$

par un argument analogue à celui utilisé dans (8.3.20). En utilisant (8.3.21), (8.3.22) et le fait que $\mathbb{P}_1(T_0 < \infty) = v$, (8.3.20) se transforme en

$$v = q(0) + \sum_{i=1}^{\infty} v^i q(i),$$

et la relation $v = g(v)$ est démontrée.

(ii) Nous avons vu dans l'étude du processus de branchement que l'équation $z = g(z)$, $z \in [0, 1]$, admet $z = 1$ pour unique solution si $\rho \leq 1$, et deux solutions $z' < 1$, $z'' = 1$ sinon. Pour compléter la preuve, il suffit donc de montrer que $v \leq z$ si z est un point fixe positif de g . De façon analogue à (8.3.22), (8.3.21), nous avons

$$\begin{aligned} \mathbb{P}_i(T_{i-1} \leq n) &= \mathbb{P}_1(T_0 \leq n), \\ \mathbb{P}_i(T_0 \leq n) &\leq \mathbb{P}_i(T_{i-1} \leq n) \times \mathbb{P}_{i-1}(T_{i-2} \leq n) \times \cdots \times \mathbb{P}_1(T_0 \leq n), \end{aligned}$$

et, de façon analogue à (8.3.20), nous avons

$$\mathbb{P}_1(T_0 \leq n + 1) = q(0) + \sum_{i \geq 1} \mathbb{P}_i(T_0 \leq n) q(i).$$

Il en résulte que $v_n = \mathbb{P}_1(T_0 \leq n)$ vérifie

$$v_{n+1} \leq q(0) + \sum_{i \geq 1} v_n^i q(i) = g(v_n).$$

Mais $v_0 = 0 \leq z$ et par un argument de récurrence, nous en déduisons que $v_n \leq g(z) = z$ puisque g est croissante sur $[0, 1]$. Le résultat désiré en découle :

$$v = \mathbb{P}_1(T_0 < \infty) = \lim_{n \rightarrow \infty} \mathbb{P}_1(T_0 \leq n) = \lim_{n \rightarrow \infty} v_n \leq z.$$

□

8.4 Suites récurrentes aléatoires discrètes

Ce paragraphe unifie les modèles étudiés précédemment.

8.4.1 Probabilités de transition

Définition 8.4.1 Soient E, F deux espaces dénombrables, et $f : E \times F \rightarrow E$ une fonction. Soit $U_n : \Omega \rightarrow F$ une suite de variables aléatoires indépendantes et de même loi. Les variables aléatoires $X_n : \Omega \rightarrow E$ définies par récurrence par

$$X_{n+1} = f(X_n, U_{n+1}) \quad , \quad n \in \mathbb{N} \tag{8.4.23}$$

forment une suite appelée **suite récurrente aléatoire**.

La valeur initiale X_0 sera choisie déterministe ou aléatoire et indépendante de $(U_n, n \geq 1)$.

L'analogie déterministe d'une telle suite aléatoire est l'évolution décrite par une équation de récurrence

$$x_{n+1} = f(x_n)$$

homogène en temps. Cette dépendance au premier ordre se traduit dans le cas aléatoire par la **propriété de Markov**.

Proposition 8.4.2 Pour tout n , la loi de l'état futur X_{n+1} ne dépend du passé (X_1, \dots, X_n) jusqu'au temps n , que par l'état présent X_n . Cette propriété s'écrit en terme de probabilités conditionnelles

$$\mathbb{P}(X_{n+1} = i \mid X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = i \mid X_n = i_n), \tag{8.4.24}$$

pour tous $n \in \mathbb{N}$, $i_0, \dots, i_n, i \in E$.

Définition 8.4.3 Une suite $(X_n)_{n \in \mathbb{N}}$ vérifiant la propriété (8.4.24) est appelée une **chaîne de Markov**.

Preuve. La définition (8.4.23) entraîne que

$$\mathbb{P}(X_0 = i_0, \dots, X_n = i_n, X_{n+1} = i) = \sum_{j: f(i_n, j) = i} \mathbb{P}(X_0 = i_0, \dots, X_n = i_n, U_{n+1} = j). \quad (8.4.25)$$

Or les variables X_0, \dots, X_n ne dépendent que de X_0, U_0, \dots, U_n , elles sont indépendantes de U_{n+1} et le terme général du second membre de (8.4.25) est égal à $\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) \mathbb{P}(U_{n+1} = j)$. Par définition des probabilités conditionnelles, nous en déduisons

$$\mathbb{P}(X_{n+1} = i | X_0 = i_0, \dots, X_n = i_n) = \sum_{j: f(i_n, j) = i} \mathbb{P}(U_{n+1} = j). \quad (8.4.26)$$

En sommant dans (8.4.25) sur i_0, \dots, i_{n-1} , nous obtenons par le même argument que le membre de droite de (8.4.26) est aussi égal à $\mathbb{P}(X_{n+1} = i | X_n = i_n)$. \square

Plus généralement, tout le futur $(X_{n+k}, k \geq 0)$ d'une chaîne de Markov ne dépend de l'histoire (X_0, \dots, X_n) que par l'état présent X_n . Les probabilités

$$Q(i, j) = \mathbb{P}(X_{n+1} = j | X_n = i) \quad (8.4.27)$$

décrivent la loi d'évolution de la suite entre deux temps successifs. On les appelle probabilités de transition.

Proposition 8.4.4 La connaissance de la loi de X_0 et des probabilités de transition $Q(i, j)$ suffit à caractériser la loi de la chaîne. Plus précisément, si X_0 est de loi $p(\cdot)$, alors

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = p(i_0) Q(i_0, i_1) Q(i_1, i_2) \dots Q(i_{n-1}, i_n).$$

Preuve. Nous avons

$$\begin{aligned} & \mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) \\ &= \mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \\ & \quad \mathbb{P}(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \\ &= Q(i_{n-1}, i_n) \mathbb{P}(X_0 = i_0, \dots, X_{n-1} = i_{n-1} | X_0 = i_0, \dots, X_{n-2} = i_{n-2}) \\ & \quad \times \mathbb{P}(X_0 = i_0, \dots, X_{n-2} = i_{n-2}). \end{aligned}$$

Nous obtenons le résultat en itérant le raisonnement. \square

Exemples :

1. Promenade aléatoire : une particule se déplace sur le réseau \mathbb{Z}^d en progressant au hasard à chaque instant n , d'un vecteur \mathbf{U}_{n+1} et sa position \mathbf{X}_n à l'instant n est donnée par

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \mathbf{U}_{n+1}$$

avec les \mathbf{U}_n indépendants et de même loi $q(\cdot)$. Dans ce cas,

$$\begin{aligned} Q(\mathbf{i}, \mathbf{j}) &= \mathbb{P}(\mathbf{X}_{n+1} = \mathbf{j} \mid \mathbf{X}_n = \mathbf{i}) \\ &= q(\mathbf{j} - \mathbf{i}) \end{aligned}$$

Une variante intéressante est de supposer que la particule est absorbée lorsqu'elle sort d'un domaine $D \subset \mathbb{Z}^d$. Dans ce cas,

$$\mathbf{X}_{n+1} = (\mathbf{X}_n + \mathbf{U}_{n+1}) \mathbf{1}_{\mathbf{X}_n \in D} + \mathbf{X}_n \mathbf{1}_{\mathbf{X}_n \notin D}$$

ce qui est encore de la forme (8.4.23). Les questions importantes sont alors : la particule finit-elle par être absorbée ? Quand et où est-elle absorbée ?

Une autre variante est de supposer que la particule est confinée dans D , avec réflexion au bord. Par exemple en dimension $d = 1$, avec $D = \mathbb{N} \subset \mathbb{Z}$, on a

$$X_{n+1} = |X_n + U_{n+1}|.$$

2. Dans le modèle de file d'attente (8.3.15), on a

$$X_{n+1} = (X_n - 1)^+ + Y_{n+1}$$

et Q est donnée par (8.3.16).

3. Pour le processus de branchement, $Z_{N+1} = \sum_{i=1}^{Z_N} Y_i^N$, où Y_i^N est le nombre d'enfants issu du i -ème individu de la N -ième génération, est bien de la forme (8.4.23).
4. Un modèle d'inventaire : Le stock d'un entrepôt est représenté à chaque instant $n \in \mathbb{N}$ par une v.a. entière positive X_n . La clientèle effectue des demandes U_{n+1} entre les instants n et $n + 1$, ($n \in \mathbb{N}$) que l'on suppose être des variables aléatoires entières positives indépendantes et de même loi $(q(i), i \in \mathbb{N})$. La politique de réapprovisionnement consiste, à chaque instant n pour lequel $X_n \leq a$, à réapprovisionner au niveau b ($0 < a < b$), ce qui se fait instantanément. Sous ces hypothèses, nous pouvons écrire que

$$\begin{aligned} X_{n+1} &= (X_n - U_{n+1})^+ \quad \text{si } X_n > a, \\ &= (b - U_{n+1})^+ \quad \text{sinon,} \end{aligned}$$

étant entendu que toute demande U_n n'est satisfaite que dans les limites du stock.

La suite $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov de la forme (8.4.23). Nous pouvons nous intéresser au comportement asymptotique ($n \rightarrow \infty$) de la loi de X_n , à la fréquence des réapprovisionnements nécessaires (événements $\{X_n \leq a\}$) et à celle des ruptures de stock (événements $\{X_n = 0\}$).

8.4.2 Stabilité

L'étude de la stabilité des suites récurrentes déterministes $x_{n+1} = f(x_n)$ passe par l'étude des points fixes de f (les points x tels que $x = f(x)$), et des cycles limites. L'équation analogue pour les suites aléatoires de la définition 8.4.1 est " X a même loi que $f(X, U)$ " où l'on suppose que X et U sont indépendantes. Ici, U est une variable aléatoire de même loi que U_1 . Cette équation ne porte pas sur la variable X , mais sur sa loi, que l'on appelle loi invariante.

Définition 8.4.5 La probabilité π sur E est dite **probabilité invariante** pour la suite récurrente aléatoire si

$$\left. \begin{array}{l} X \text{ indépendante de } U \\ X \text{ de loi } \pi \end{array} \right\} \Rightarrow f(X, U) \text{ suit la loi } \pi .$$

Dans ce cas, si X_0 est de loi π , alors $X_1 = f(X_0, U_0)$ est de loi π et par conséquent X_n a pour loi la loi π . La suite se trouve alors dans un état stationnaire. En particulier elle décrit un phénomène stable.

Pour étudier la stabilité d'une suite, on cherchera d'éventuelles probabilités invariantes. Voilà un premier critère en terme de probabilités de transition.

Proposition 8.4.6 La probabilité π sur E est invariante si et seulement si :

$$\sum_{i \in E} \pi(i) Q(i, j) = \pi(j) \quad , \quad \forall j \in E .$$

Preuve. En effet, si X est indépendant de U et de loi π ,

$$\begin{aligned} \mathbb{P}(f(X, U) = j) &= \sum_{i \in E} \mathbb{P}(f(i, U) = j, X = i) \\ &= \sum_{i \in E} \mathbb{P}(f(i, U) = j) \pi(i) \\ &= \sum_{i \in E} Q(i, j) \pi(i) \end{aligned}$$

d'après (8.4.26) et (8.4.27). Ceci montre la proposition. \square

Exemple 8.4.7 *Modèle de diffusion d'Ehrenfest.* Dans deux enceintes séparées par une paroi poreuse sont réparties N molécules de gaz. À chaque unité de temps une molécule choisie au hasard change d'enceinte.

(i) *Vision microscopique* : l'état du système est représenté comme un vecteur dans $\{0, 1\}^N$, la $i^{\text{ième}}$ composante valant 0 et 1 selon que la $i^{\text{ième}}$ particule est dans l'enceinte de droite ou dans celle de gauche. Le système effectue alors une promenade aléatoire sur le "cube" $\{0, 1\}^N$ dite "au plus proche voisin" : à chaque instant, le système saute d'un sommet du cube à l'un des N sommets voisins. La probabilité uniforme sur $\{0, 1\}^N$ est une probabilité invariante.

(ii) *Vision macroscopique* : les particules sont indistinguables. Le système est alors représenté à l'instant n par le nombre X_n de particules dans l'enceinte de gauche. Le mécanisme d'évolution se traduit ainsi : introduisons une suite de variables aléatoires $(U_n, n \geq 1)$ indépendantes de loi uniforme sur $\{1, 2, \dots, N\}$. Nous posons

$$X_{n+1} = \begin{cases} X_n + 1 & \text{si } U_{n+1} > X_n \\ X_n - 1 & \text{si } U_{n+1} \leq X_n \end{cases}$$

qui est de la forme (8.4.23). Ses probabilités de transition sont pour $i \in \{0, \dots, N\}$,

$$Q(i, i+1) = 1 - \frac{i}{n}, \quad Q(i, i-1) = \frac{i}{n},$$

et $Q(i, j) = 0$ sinon. L'espace d'état est l'espace fini $E = \{0, \dots, N\}$, et aucune instabilité ne peut se produire. Plutôt que de chercher les probabilités invariantes à l'aide de la proposition 8.4.6, il est plus rapide d'utiliser l'exercice 8.5.8 (ii), et de chercher les solutions de $\pi(i) Q(i, j) = \pi(j) Q(j, i)$. Nous trouvons aisément que $\pi(j) = \binom{N}{j} (1/2)^N$, c'est-à-dire que π est la loi binomiale $\mathcal{B}(N, 1/2)$.

8.5 Exercices sur le chapitre 8

EXERCICE 8.5.1 On considère la marche aléatoire S_n donnée par (8.1.1). On définit les événements

$$A_m = \{\exists n \mid S_n = 0, \quad S_i < m \quad \forall i < n\}$$

où la marche visite 0 avant m , ($m \geq k+1$), et

$$A = \{\exists n \mid S_n = 0\}$$

où la marche visite 0.

- 1) Montrer que la probabilité $\mathbb{P}(A_m)$ est donnée par (8.1.4) ou (8.1.5), selon que $p \neq 1/2$ ou $p = 1/2$.
- 2) Montrer que $A_m \subset A_{m+1}$ et calculer la probabilité $\mathbb{P}(\cup_m A_m)$.
- 3) Montrer que $A = \cup_m A_m$. Qu'en déduire ?

EXERCICE 8.5.2 Dédurre de la proposition 8.2.1 que pour des variables aléatoires X_i et N indépendantes et de carré intégrable et $S_N = \sum_{i=1}^N X_i$,

$$\mathbb{E}(S_N) = \mathbb{E}(N)\mathbb{E}(X) \quad , \quad \text{Var}(S_N) = \text{Var}(N)\mathbb{E}(X)^2 + \mathbb{E}(N)\text{Var}(X) \quad ,$$

où $\mathbb{E}(X)$ et $\text{Var}(X)$ désignent respectivement les espérance et variance communes aux X_i . Comment s'écrivent ces formules lorsque N est déterministe ?

EXERCICE 8.5.3 Si N, X_1, X_2, \dots sont des variables aléatoires strictement positives indépendantes, si N suit une loi géométrique sur \mathbb{N} d'espérance a avec $a > 1$ ($\mathbb{P}(N = k) = \frac{1}{a}(1 - \frac{1}{a})^k$ pour tout $k \in \mathbb{N}$), et si les X_i suivent la même loi géométrique sur \mathbb{N} d'espérance b ($b > 1$), montrer que $S_N = \sum_{i=1}^N X_i$ suit une loi géométrique d'espérance ab .

EXERCICE 8.5.4 Nous considérons un processus de Galton-Watson $(Z_n)_n$ critique : sa loi de reproduction est d'espérance $\mathbb{E}(Y) = 1$. Nous supposons que Y est de carré intégrable et nous notons g sa fonction génératrice. Nous voulons montrer que

$$\mathbb{P}(Z_n \neq 0) \stackrel{n \rightarrow \infty}{\sim} \frac{2}{n\mathbb{E}(Y(Y-1))}.$$

1) Notons $m_2 = \mathbb{E}(Y^2)$ et $a = \frac{m_2 - 1}{2}$. Montrer, en utilisant un développement limité de g au voisinage de 1, que pour $x \in [0, 1[$, $\frac{1}{1-g(x)} = \frac{1}{1-x} + a + \Gamma(x)$, avec $\lim_{x \rightarrow 1} \Gamma(x) = 0$.

2) Soit $g_n(x) = g \circ \dots \circ g(x)$, n fois. Montrer que $\frac{1}{1-g_n(x)} = \frac{1}{1-x} + na + \Delta_n(x)$, où $\Delta_n(x) = \sum_{j=0}^{n-1} \Gamma(g_j(x))$.

3) Montrer que $1 - g_n(x) \stackrel{n \rightarrow \infty}{\sim} \frac{2}{n(m_2 - 1)}$.

En déduire le résultat.

EXERCICE 8.5.5 Soient $(X_n)_n, (Y_n)_n, (Z_n)_n$, des suites de variables indépendantes équadistribuées, prenant les valeurs $+1$ et -1 avec probabilité $\frac{1}{2}$.

1) On note $\xi_n = (X_n, Y_n, Z_n)$, et $S_n = \sum_{j=1}^n \xi_j$, avec par convention $S_0 = (0, 0, 0)$.

Soit $\mathbf{a} = (x, y, z) \in \mathbb{Z}^3$ et $N_{\mathbf{a}} = \sum_{n \geq 0} \mathbf{1}_{S_n = \mathbf{a}}$.

1-a) Calculer $\mathbb{P}(S_n = (0, 0, 0))$.

1-b) Montrer que

$$\mathbb{E}(N_{(0,0,0)}) < +\infty.$$

(Et donc en conséquence, $\mathbb{P}(N_{(0,0,0)} < +\infty) = 1$.)

1-c) Reprendre le calcul pour \mathbf{a} quelconque.

2) Soit $r > 0$ et $B_r = \{\mathbf{x} \in \mathbb{Z}^3, |\mathbf{x}| \leq r\}$. Montrer que le nombre de visites de la marche aléatoire S_n dans B_r est presque-sûrement fini.

En déduire que $|S_n|$ converge vers $+\infty$ presque-sûrement, pour toute norme $|\cdot|$ sur \mathbb{R}^3 .

Comparer avec la marche aléatoire simple sur \mathbb{Z} .

EXERCICE 8.5.6 Soit un processus de branchement (Z_n) tel que $Z_0 = 1$ et soit $\phi(x)$ la fonction génératrice de Z_1 . On suppose que Z_1 est de carré intégrable, et on définit $\rho = \mathbb{E}(Z_1)$, $\sigma^2 = \text{Var}(Z_1)$.

1) Rappeler la formule donnant l'expression de la fonction génératrice G_n de Z_n en fonction de ϕ .

2) Calculer $\mathbb{E}(Z_n)$ et $\text{Var}(Z_n)$ en fonction de ρ et de σ^2 .

3) Exprimer la fonction génératrice du couple (Z_n, Z_m) , $n > m$, en fonction de G_m et G_{n-m} .

4) Si $n > m$, montrer que $\mathbb{E}(Z_n Z_m) = \rho^{n-m} \mathbb{E}(Z_m^2)$.

5) On suppose que $\rho > 1$. On considère la suite $(W_n)_n$ définie par $W_n = \frac{Z_n}{\rho^n}$.

Montrer que pour $p > 0$, la suite

$$\mathbb{E}((W_{m+p} - W_m)^2) \text{ converge vers } 0 \text{ quand } m \rightarrow \infty. \quad (*)$$

On acceptera que toute suite de Cauchy au sens (*) converge en moyenne quadratique : il existe donc une variable $W \in L^2$ telle que W_n converge en moyenne quadratique vers W . (i.e. telle que $\mathbb{E}(W_n - W)^2 \rightarrow 0$).

Montrer que

$$\sum_n \mathbb{E}((W_n - W)^2) < \infty.$$

En déduire que $(W_n)_n$ converge presque-sûrement vers W .

Donner la valeur de l'espérance et de la variance de W .

EXERCICE 8.5.7 1) On considère une variable aléatoire Y à valeurs dans \mathbb{N} , de loi géométrique définie pour $p \in]0, 1[$ par

$$\mathbb{P}(Y = k) = p^k(1 - p), \quad \forall k \in \mathbb{N}.$$

Calculer la fonction génératrice de Y . Que vaut l'espérance $\mathbb{E}(Y)$?

2) Soit $(Z_n)_n$ un processus de branchement tel que $Z_0 = 1$. On suppose que

$$\mathbb{E}(x^{Z_1}) = \frac{1 - p}{1 - px}, \quad \text{pour tout } x \in [0, 1[,$$

avec $p \neq \frac{1}{2}$.

Montrer par récurrence sur n que

$$\mathbb{E}(x^{Z_n}) = \frac{\rho^n - 1 - \rho x(\rho^{n-1} - 1)}{\rho^{n+1} - 1 - \rho x(\rho^n - 1)},$$

où l'on a posé $\rho = \frac{p}{1-p}$.

3) On suppose à partir de maintenant que $p < \frac{1}{2}$. Que vaut dans ce cas la probabilité d'extinction de Z_n ?

4) Soit à présent $(Z_n^*)_n$ un nouveau processus de branchement tel que $Z_0^* = 1$. Chaque individu vivant donne naissance à un nombre aléatoire de descendants dont la loi est celle de Z_1 . De plus, à chaque instant de naissance, on augmente la population d'un nouvel individu (modèle de processus de branchement avec immigration).

4-a) Soit $(\hat{Z}_n)_n$ une famille de variables aléatoires indépendantes telles que, pour tout n , \hat{Z}_n a même loi que Z_n . Montrer que Z_n^* a même loi que $1 + \hat{Z}_1 + \dots + \hat{Z}_n$.

4-b) Montrer que

$$\lim_{n \rightarrow \infty} \mathbb{E}(x^{Z_n^*}) = \frac{x(1-\rho)}{1-\rho x} = L(x).$$

4 - c) Montrer que

$$\lim_{n \rightarrow \infty} \mathbb{E}(x^{Z_n} | Z_n > 0) = L(x).$$

Qu'en conclure ?

5) On suppose que la loi de Z_1 est donnée par

$$\mathbb{P}(Z_1 = k) = 2^{-(k+1)}, \quad k \geq 0.$$

Montrer que Z_1 suit une loi géométrique dont on donnera le paramètre.

Montrer par récurrence que $G_n(x) = \mathbb{E}(x^{Z_n})$ vérifie dans ce cas que

$$G_n(x) = \frac{n - (n-1)x}{n+1-nx}.$$

En déduire que

$$\lim_{n \rightarrow \infty} \mathbb{E}\left(\frac{Z_n}{n} | Z_n > 0\right) = 1.$$

Qu'en conclure ?

EXERCICE 8.5.8 Réversibilité.

Soit E un espace dénombrable. Soit $Q(i, j)$ les noyaux de transition d'une chaîne de Markov sur E , et π une probabilité sur E telle que

$$\pi(i) Q(i, j) = \pi(j) Q(j, i) \quad \forall i, j \in E.$$

(i) Montrer que π est une probabilité invariante pour la chaîne.

(ii) Montrer que si l'état initial X_0 est de loi π , alors (X_0, X_1) a même loi que (X_1, X_0) . Le processus est alors appelé réversible, car sa loi ne change pas lorsque l'on inverse le sens du temps.

Chapitre 9

Corrections des exercices

9.1 Corrigés des exercices du chapitre 2

Exercice 2.6.1 : 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 2.6.2 : Cette formule se montre par récurrence sur n . Elle est triviale pour $n = 1$. Remarquons que pour $n = 2$,

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2). \quad (9.1.1)$$

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 (9.1.1) 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 2.6.3 :

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) Numérotons de 1 à n les lettres et numérotons 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 2.6.4 : 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 2.6.5 : 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 2.6.6 : 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 2.6.7 : 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 2.6.8 : 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 2.6.9 : 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 2.6.10 : Remarquons que les événements aléatoires A_k sont indépendants. Ainsi, par le lemme de Borel-Cantelli 2.5.14, $\mathbb{P}(\limsup_k A_k)$ 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 2.6.11 : 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.

9.2 Corrigés des exercices du chapitre 3

Exercice 3.7.1 : 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 3.7.2 : 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 3.7.3 : 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 3.7.4 : 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 λ et μ .

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 $B\left(n, \frac{\lambda}{\lambda + \mu}\right)$.

On peut alors calculer $\mathbb{E}(X | S = n)$ en utilisant (3.6.34), mais il est plus rapide d'utiliser la valeur np de l'espérance d'une variable de loi $B(n, p)$ (voir (3.5.18)).

Ainsi, $\mathbb{E}(X | S = n) = n \frac{\lambda}{\lambda + \mu}$, ou encore

$$\mathbb{E}(X | S) = S \frac{\lambda}{\lambda + \mu}.$$

Exercice 3.7.5 : 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 3.7.6 :

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 3.7.7 :

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 3.7.8 :

1-a) C'est évident.

1-b) La condition suffisante est immédiate. Elle est nécessaire car $\phi \geq 0$, ϕ 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 ϕ 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 ϕ 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 β tend vers l'infini, μ_β devient une probabilité uniforme sur l'ensemble $\Omega_{\min} := \{\omega; U(\omega) = \min_{\Omega} U\}$.

De même que lorsque $\beta \rightarrow -\infty$, μ_β devient une probabilité uniforme sur $\Omega_{\max} := \{\omega : U(\omega) = \max_{\Omega} U\}$. Les limites (3.7.44) s'en suivent immédiatement.

2-c) $\beta \rightarrow Z(\beta)$ est non nulle et de classe C^∞ 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. (3.7.44) 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 λ par la condition que η est une probabilité :

$$\lambda + 1 = \ln \sum_{i=1}^n e^{-\beta U(\omega_i)} = \ln Z(\beta).$$

Ainsi, si μ_β 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 ϕ .

$$\begin{aligned}
 & -\sum_i p_i \ln p_i + \sum_i (-\beta U(\omega_i) p_i) - \ln Z(\beta) \\
 &= -\sum_i p_i \ln p_i + \sum_i (p_i \ln e^{-\beta U(\omega_i)}) - \sum_i p_i \ln Z(\beta) \\
 &= -\sum_i p_i \ln \frac{p_i}{\mu_\beta(\omega_i)} \\
 &= -\sum_i \mu_\beta(\omega_i) \frac{p_i}{\mu_\beta(\omega_i)} \ln \frac{p_i}{\mu_\beta(\omega_i)} = \sum_i \mu_\beta(\omega_i) \phi\left(\frac{p_i}{\mu_\beta(\omega_i)}\right) \\
 &\leq \phi\left(\sum_i \mu_\beta(\omega_i) \frac{p_i}{\mu_\beta(\omega_i)}\right) = \phi(1) = 0.
 \end{aligned}$$

Nous avons donc montré que pour toute proba p :

$$H(p) + \langle (-\beta U) \rangle_p \leq \ln Z(\beta).$$

On sait que μ_β atteint le maximum de $p \mapsto H(p) + \langle (-\beta U) \rangle_p$ qui vaut $\ln Z(\beta)$. Puisque ϕ 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 μ_β 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, μ_β est l'unique probabilité qui atteint le maximum.

9.3 Corrigés des exercices du chapitre 4

Exercice 4.12.1 : Il suffit d'écrire

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

Exercice 4.12.2 : 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 4.12.3 : 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 4.12.4 : Rappel sur la fonction Γ : $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ 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 4.12.5 : 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 4.12.6 : 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 λ , (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 4.12.7 : 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 20^{ème} 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 4.12.8 : 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 4.12.9 : 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 $f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)} = xe^{-x(y-1)} \mathbf{1}_{[1, +\infty[}(y)$, 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 4.12.10 :

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 4.12.11 :

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 β .

Exercice 4.12.12 : 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 4.12.13 : 1) Soit f une fonction continue bornée sur $\mathbb{R}_+ \times [0, 2\pi[$. Alors $\mathbb{E}(f(R, \Theta)) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(r, \theta) e^{-\frac{x^2+y^2}{2}} dx dy = \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} f(r, \theta) e^{-\frac{r^2}{2}} r dr d\theta$. D'où R admet la loi de densité $r e^{-\frac{r^2}{2}} \mathbf{1}_{r>0}$ et Θ la loi uniforme sur $[0, 2\pi]$, et R et Θ sont indépendants.

R^2 et Θ 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 $[0, 1]$. Θ 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 Θ). 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 4.12.14 : 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 4.12.15 : 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 λ . De même, la loi conditionnelle de D sachant $X < Y$ est une loi exponentielle de paramètre μ .

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.

9.4 Corrigés des exercices du chapitre 5

Exercice 5.4.1 : 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 ε 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 2.5.14), $\mathbb{P}(\limsup_n A_n) = 0$. Ainsi presque-sûrement, au plus un nombre fini de A_n sont réalisés. Il en résulte que $X_n \xrightarrow{P.S.} 0$.

3) Nous en déduisons par un argument de limite de Césaro que $\frac{X_1 + \dots + X_n}{n} \xrightarrow{P.S.} 0$ 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 5.4.2 : 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 $\frac{X_1 \wedge M + \dots + X_n \wedge M}{n} \xrightarrow{P.S.} \mathbb{E}(X_1 \wedge M)$.

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 ω , $\frac{X_1(\omega) \wedge M_k + \dots + X_n(\omega) \wedge M_k}{n} \rightarrow \mathbb{E}(X_1 \wedge M_k)$. 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 5.4.3 : 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 5.4.4 : 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 5.4.5 :

$$S_n = S_0(1 + R_1) \cdots (1 + R_n) = S_0 \exp\left(n \frac{1}{n} \sum_{k=1}^n \ln(1 + R_k)\right).$$

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 5.4.6 : Les deux premières questions se résolvent comme dans l'exercice 5.4.3.

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.

9.5 Corrigés des exercices du chapitre 6

Exercice 6.5.1 : 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, $\sigma^2 = \mathbb{E}(X^2)$, ϕ 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 6.5.2 : 1) $\phi_X(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{e^{itz}}{1+x^2} dx$. 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 Γ 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 6.5.3 : 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 6.5.4 : 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 $\mathbf{C}_{\mathbf{V}}$ diagonale par bloc, chaque bloc étant la matrice de covariance $\mathbf{C}_{\mathbf{X}}$ de \mathbf{X} :

$$\mathbf{C}_{\mathbf{V}} = \begin{pmatrix} \mathbf{C}_{\mathbf{X}} & \mathbf{0}_d \\ \mathbf{0}_d & \mathbf{C}_{\mathbf{X}} \end{pmatrix}$$

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 θ .

3) Supposons que $\mathbb{P}(A(0)) = \eta > 0$. Alors pour n angles θ_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 6.5.5 : 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 6.3.10, 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 6.5.6 : Préliminaire : Il suffit d'adapter la preuve de la proposition 6.3.5 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 (6.5.30). 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 ε 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 6.5.7 : 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 $\frac{1}{n}(n\sigma^2) + \frac{n}{2n}\sigma^2 = \frac{3}{2}\sigma^2$. 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 $\limsup_n \mathbb{P}\left(\frac{1}{n^\alpha} \left| \frac{S_n}{n} - m \right| > \frac{1}{\beta}\right) = 0$, d'où $n^\alpha \left| \frac{S_n}{n} - m \right|$ tend vers $+\infty$ en probabilité.

9.6 Corrigés des exercices du chapitre 7

Exercice 7.4.1 : 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 χ_n^2 indépendante de Z , on a d'après (6.2.19) :

$$\begin{aligned} \mathbb{E}(h(T)) &= \mathbb{E}\left[h\left(\sqrt{n}\frac{Z}{\sqrt{U}}\right)\right] \\ &= \frac{1}{2^{n/2}\Gamma(n/2)\sqrt{2\pi}} \int_0^{+\infty} du \int_{-\infty}^{+\infty} dz u^{\frac{n}{2}-1} e^{-\frac{u}{2}} e^{-\frac{z^2}{2}} h\left(\sqrt{n}\frac{z}{\sqrt{u}}\right) \\ &= \frac{2}{2^{n/2}\Gamma(n/2)\sqrt{2\pi}} \int_0^{+\infty} dx \int_{-\infty}^{+\infty} dz x^{n-1} e^{-\frac{x^2+z^2}{2}} h\left(\sqrt{n}\frac{z}{x}\right) \text{ avec } u \mapsto x = \sqrt{u} \\ &= \frac{2}{2^{n/2}\Gamma(n/2)\sqrt{2\pi n}} \int_0^{+\infty} dx \int_{-\infty}^{+\infty} dt x^n e^{-\frac{x^2+x^2 t^2/n}{2}} h(t) \text{ avec } z \mapsto t = \sqrt{n}\frac{z}{x} \\ &= \frac{2}{2^{n/2}\Gamma(n/2)\sqrt{2\pi n}} \int_{-\infty}^{+\infty} dt h(t) \int_0^{+\infty} dx x^n e^{-\frac{x^2}{2}\left(1+\frac{t^2}{n}\right)} \text{ par Fubini} \\ &= \frac{1}{2^{n/2}\Gamma(n/2)\sqrt{2\pi n}} \int_{\mathbb{R}} dt h(t) \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \int_0^{+\infty} dy y^{\frac{n-1}{2}} e^{-\frac{y}{2}} \text{ avec } x \mapsto y = x^2\left(1+\frac{t^2}{n}\right) \\ &= \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\Gamma(n/2)} \int_{\mathbb{R}} dt h(t) \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}. \end{aligned}$$

Ce calcul permet d'identifier la loi de T et sa densité.

Exercice 7.4.2 : 1) Soit $(X_i)_i$ une suite de variables aléatoires i.i.d. de loi $\mathcal{N}(0, 1)$. Par définition de la loi χ_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 $(\sqrt{2Z_n} - \sqrt{2n-1})_n$ 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 χ^2 à n degrés de liberté, donc $X/\sqrt{U_n/n}$ suit la loi de Student à n degrés de liberté. Donc ζ_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 7.4.3 : 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 $n = 1708$.

Etudions 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 7.4.4 : 1) On sait que $\mathbb{E}(X_1) = m\theta$. Donc un estimateur de θ 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 :

$$p_n(\mathbf{x}, \theta) = \left[\prod_{i=1}^n \binom{m}{x_i} \right] [\theta^{\bar{x}_n} (1 - \theta)^{m - \bar{x}_n}]^n, \quad (9.6.2)$$

avec $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$. A \mathbf{x} fixé, en tant que fonction de θ , 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 7.4.5 : 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.

Exercice 7.4.6 : 1) Sous \mathbb{P}_θ , X_1, \dots, X_n sont indépendantes et identiquement distribuées selon la loi exponentielle de paramètre θ , 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 θS_n suit la loi $\Gamma(n, 1)$, ou que $2\theta S_n$ suit la loi χ_{2n}^2 .

2) Si on note $\Phi_{\chi(2n)}$ la fonction de répartition de la loi χ_{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 χ_{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 θ .

Par exemple, pour $n = 10$ et $\alpha = 5\%$, on a $a_n = 9,59$ et $b_n = 34,17$ d'après la Table 7.2, et donc on trouve que $[4,80/S_n, 17,08/S_n]$ est un intervalle de confiance au niveau 95% de θ .

Exercice 7.4.7 : 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 7.4.8 : 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 α fixés, cette fonction de β 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 α , 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 α et tend vers $+\infty$ en $+\infty$. Donc l'EMV de (α, β) 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 α .

Un estimateur de α 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}_α -presque sûrement. Donc $\hat{\alpha}_n \xrightarrow{n \rightarrow +\infty} \alpha$ \mathbb{P}_α -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 $1 - \eta/2$ de la loi gaussienne centrée réduite. Ceci s'écrit aussi :

$$\mathbb{P}_\alpha \left(\left| \hat{\alpha}_n/\alpha - 1 \right| \leq \Phi^{-1}(1 - \eta/2)/\sqrt{n} \right) \simeq 1 - \eta.$$

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 7.4.9 : 1) D'une part, pour tout $z > 0$, $\mathbb{P}_{\lambda,\mu}(Z_i \leq z) = 1 - \mathbb{P}_{\lambda,\mu}(\min(X_i, Y_i) > z) = 1 - \exp(-(\lambda + \mu)z)$, 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 λ 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 λ 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}_λ , 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}_λ , $\sum_{i=1}^n Z_i \sim \Gamma(n, \lambda + \mu)$ 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}_λ -p.s. vers $\mathbb{E}_\lambda(Z_1) = \frac{1}{\lambda+\mu}$ d'après la loi forte des grands nombres. Sous \mathbb{P}_λ , d'après le théorème de la limite centrale,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (W_i - \lambda Z_i) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathcal{N}(0, \text{Var}_\lambda(W_1 - \lambda Z_1)),$$

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}_λ ,

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathcal{N}(0, \lambda(\lambda+\mu)),$$

en loi.

7) Sous \mathbb{P}_λ , $\hat{\lambda}_n(\hat{\lambda}_n + \mu)$ converge p.s. vers $\lambda(\lambda + \mu)$. Donc, toujours d'après le théorème de Slutsky, sous \mathbb{P}_λ ,

$$\frac{\sqrt{n}}{\sqrt{\hat{\lambda}_n(\hat{\lambda}_n + \mu)}} (\hat{\lambda}_n - \lambda) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathcal{N}(0, 1),$$

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 λ de niveau $1 - \alpha$.

8) A μ 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 (λ, μ) est

$$(\hat{\lambda}_n, \hat{\mu}_n) = \left(\frac{\sum_{i=1}^n W_i}{\sum_{i=1}^n Z_i}, \frac{n - \sum_{i=1}^n W_i}{\sum_{i=1}^n Z_i} \right).$$

Exercice 7.4.10 : 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 χ_{2n}^2 . Ce test a pour niveau α .

A.N. : $\zeta_{15}^{obs} = 44,1$ et d'après la Table 7.2, $x_{0,025}(30) = 16,79$ 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 7.4.11 : 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 , ζ_{250} suit une loi du χ^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%.

9.7 Corrigés des exercices du chapitre 8

Exercice 8.5.1 : 1) $\mathbb{P}(A_m)$ est la probabilité que partant de k , la marche revienne à 0 avant d'atteindre m , ce qui est calculé dans (8.1.4) et (8.1.5).

2) L'inclusion est immédiate. Comme les A_m sont croissants, $\mathbb{P}(\cup_m A_m) = \lim_m \mathbb{P}(A_m)$.

Ainsi, pour $p \leq 1/2$, $\mathbb{P}(\cup_m A_m) = 1$, et pour $p > 1/2$, $\mathbb{P}(\cup_m A_m) = \left(\frac{1-p}{p}\right)^k$.

3) Remarquons que $\{S_n = 0\} \subset A_{n+k+1}$ pour chaque n , puisque la marche aléatoire partant de k ne peut pas monter en $n+k+1$ en moins de n étapes. Ainsi, $A = \cup_n \{S_n = 0\} \subset \cup_m A_m$. L'autre inclusion est immédiate. Nous en déduisons que si $p \leq 1/2$, la marche aléatoire atteindra 0 en partant de n'importe quelle valeur k avec probabilité 1 alors que si $p > 1/2$, elle n'atteindra pas 0 avec probabilité strictement positive.

Exercice 8.5.2 : Soit g la fonction génératrice de X et G celle de N . Alors celle de S_N vaut $\Gamma = G \circ g$, comme nous l'avons vu à la proposition 8.2.1. Nous avons donc

$$\Gamma'(x) = G'(g(x)) g'(x).$$

Ainsi, $\Gamma'(1) = G'(1) g'(1)$, d'où $\mathbb{E}(S_N) = \mathbb{E}(N)\mathbb{E}(X)$. Nous avons également $\Gamma''(x) = G''(g(x)) (g'(x))^2 + G'(g(x)) g''(x)$. Or $\Gamma''(1) = \mathbb{E}(S_N^2 - S_N)$. En remplaçant également $g''(1)$ et $G''(1)$ par leurs valeurs, nous trouvons la formule demandée.

Quand N est déterministe, nous retrouvons les formules classiques pour les variables aléatoires indépendantes : $\mathbb{E}(S_N) = N\mathbb{E}(X)$ et $\text{Var}(S_N) = N\text{Var}(X)$.

Exercice 8.5.3 : Nous avons vu que la fonction génératrice d'une loi géométrique de paramètre $\frac{1}{a}$ est égale à $G_a(x) = \frac{x}{a-(a-1)x}$. Ainsi la fonction génératrice de S_N vaudra $G_a \circ G_b(x) = \frac{x}{ab-(ab-1)x}$. Nous reconnaissons la fonction génératrice d'une loi géométrique de paramètre ab .

Exercice 8.5.4 : 1) Comme $g(1) = 1$, $g'(1) = 1$, et $g''(1) = m_2 - 1$, un développement limité de g au voisinage de 1 donne que pour $x \in [0, 1[$,

$$1 - g(x) = (1 - x) - (1 - x)^2 a + (x - 1)^2 \gamma(x),$$

où $\gamma(x)$ tend vers 0 quand x tend vers 1. En prenant les inverses, nous obtenons $\frac{1}{1-g(x)} = \frac{1}{1-x} + a + \Gamma(x)$, avec $\lim_{x \rightarrow 1} \Gamma(x) = 0$.

2) En réitérant cette égalité en changeant x en $g(x)$ nous obtenons la formule pour $n = 2$ et de proche en proche pour n .

3) A x fixé, nous savons que $g_j(x)$ tend vers $1 = m$ quand j tend vers l'infini (car 1 est le seul point fixe de $g(s) = s$). Ainsi, $\Gamma(g_j(x))$ tend vers 0 quand j tend vers l'infini, et il en est de même pour la somme de Césaro $\frac{\Delta_n(x)}{n} = \frac{1}{n} \sum_{j=0}^{n-1} \Gamma(g_j(x))$ quand n tend vers l'infini. Ainsi,

$$1 - g_n(x) = \frac{1}{na} \frac{1}{1 + \frac{1}{na(1-x)} + \frac{\Delta_n(x)}{na}} \sim_{n \rightarrow \infty} \frac{1}{na} = \frac{2}{n(m_2 - 1)}.$$

Exercice 8.5.5 : 1-a) $\mathbb{P}(\mathbf{S}_n = (0, 0, 0)) = (\mathbb{P}(\sum_{j=1}^n X_j = 0))^3$. Or $\mathbb{P}(\sum_{j=1}^n X_j = 0) = 0$ si n est impair, et si $n = 2k$, $\mathbb{P}(\sum_{j=1}^{2k} X_j = 0) = \binom{2k}{k} \left(\frac{1}{2}\right)^{2k}$.

1-b) $\mathbb{E}(N_{(0,0,0)}) = \sum_n \mathbb{P}(\mathbf{S}_n = (0, 0, 0)) = \sum_k \binom{2k}{k} \left(\frac{1}{2}\right)^{2k}$. Or par la formule de Stirling (i.e. $k! \sim_{k \rightarrow \infty} \sqrt{2\pi k} \left(\frac{k}{e}\right)^k$), nous en déduisons que le terme général de la série est équivalent, quand k tend vers l'infini, à $\left(\frac{1}{\sqrt{\pi k}}\right)^3$. La série converge donc, et $\mathbb{E}(N_{(0,0,0)}) < +\infty$.

1-c) Si $\mathbf{a} = (a_1, a_2, a_3)$, alors, par un raisonnement similaire, $\mathbb{P}(\mathbf{S}_n = \mathbf{a}) = \prod_{i=1}^3 \binom{n}{\frac{a_i+n}{2}} \left(\frac{1}{2}\right)^n$. Nous pouvons montrer que $\mathbb{P}(\mathbf{S}_n = \mathbf{a}) \sim \left(\frac{C^{te}}{\sqrt{n}}\right)^3$, et donc la série de terme général $\mathbb{P}(\mathbf{S}_n = \mathbf{a})$ converge. Nous en déduisons que $\mathbb{E}(N_{\mathbf{a}}) = \sum_n \mathbb{P}(\mathbf{S}_n = \mathbf{a}) < +\infty$.

2) Par la question précédente, nous savons que $N_{\mathbf{a}} < +\infty$ presque-sûrement. Comme B_r est fini, il en est de même de $\sum_{\mathbf{a} \in B_r} N_{\mathbf{a}}$. Alors, avec probabilité 1, il existe N tel que $\mathbf{S}_n \notin B_r$ pour tout $n \geq N$. Nous en déduisons que $|\mathbf{S}_n|$ converge vers $+\infty$ presque-sûrement.

Ce comportement, dit transient, est très différent de celui de la marche aléatoire simple sur \mathbb{Z} , qui est récurrente : il y a presque-sûrement une infinité de retours à 0 ou dans chaque état.

Exercice 8.5.6 : 1) $G_n = \phi \circ \dots \circ \phi$, n fois.

2) Par l'exercice 8.5.2, nous obtenons $\mathbb{E}(Z_n) = \rho^n$, et $\text{Var}(Z_n)$ vérifie

$$\text{Var}(Z_{n+1}) = \rho^2 \text{Var}(Z_n) + \sigma^2 \rho^n.$$

Nous pouvons alors montrer par récurrence que $\text{Var}(Z_n) = \sigma^2(\rho^{2n-2} + \dots + \rho^{n-1})$.

3)

$$G(x, y) = \sum_{k, l} \mathbb{P}(Z_m = k, Z_n = l) x^k y^l = \sum_{k, l} \mathbb{P}(Z_m = k) \mathbb{P}(Z_n = l | Z_m = k) x^k y^l.$$

Si $Z_m = k$, alors Z_n est la somme de k variables aléatoires indépendantes de même loi que Z_{n-m} , issu de chacun des m individus de la génération m . Ainsi,

$$\begin{aligned} G(x, y) &= \sum_k \mathbb{P}(Z_m = k) \left(\sum_l \mathbb{P}(Z_{n-m} = l) y^l \right)^k x^k = \sum_k \mathbb{P}(Z_m = k) x^k (G_{n-m}(y))^k \\ &= G_m(x G_{n-m}(y)). \end{aligned}$$

4) On a $\mathbb{E}(Z_m Z_n) = \frac{\partial^2}{\partial x \partial y} G(x, y)(1, 1)$. On calcule

$$\frac{\partial^2}{\partial x \partial y} G(x, y)(x, y) = G'_m(x G_{n-m}(y)) G'_{n-m}(y) + G_{n-m}(y) x G'_{n-m}(y) (G_m)''(x G_{n-m}(y)).$$

Nous en déduisons que $\mathbb{E}(Z_m Z_n) = \rho^n + \rho^{n-m} (G_m)''(1) = \rho^n + \rho^{n-m} (\mathbb{E}(X_m^2) - \rho^m) = \rho^{n-m} \mathbb{E}(Z_m^2)$.

5) On a $\mathbb{E}(W_n^2) = \frac{\mathbb{E}(Z_n^2)}{\rho^{2n}} = \frac{\text{Var}(Z_n) + \rho^{2n}}{\rho^{2n}} = 1 + \sigma^2 \frac{\rho^{n-1}}{\rho^{2n}} (1 + \rho + \dots + \rho^{n-1}) = 1 + \frac{\sigma^2}{\rho(\rho-1)} \left(1 - \frac{1}{\rho^n}\right)$.

Nous avons $\mathbb{E}((W_{n+p} - W_n)^2) = \mathbb{E}((W_{n+p})^2) - \mathbb{E}((W_n)^2)$, car $\mathbb{E}(W_{n+p} W_n) = \rho^p \mathbb{E}((W_n)^2)$ par la question précédente. Ainsi,

$$\mathbb{E}((W_{n+p} - W_n)^2) = \frac{\sigma^2}{\rho(\rho-1)} \left(\frac{1}{\rho^n} - \frac{1}{\rho^{n+p}} \right) = \frac{\sigma^2}{\rho(\rho-1)} (\rho^p - 1) \frac{1}{\rho^{n+p}},$$

qui tend vers 0 quand n tend vers l'infini.

En faisant tendre p vers l'infini dans l'expression ci-dessus, nous obtenons que $\mathbb{E}((W - W_n)^2) = \frac{\sigma^2}{\rho(\rho-1)} \frac{1}{\rho^n}$, et $\rho > 1$. Ainsi la série de terme général $\mathbb{E}((W - W_n)^2)$ converge. Par suite, la variable aléatoire $\sum_n (W_n - W)^2$ est finie presque-sûrement, et donc en particulier son terme général tend vers 0. Nous en concluons donc que $(W_n - W)^2$ converge presque-sûrement vers 0.

Nous avons $\mathbb{E}(W) = \lim_n \mathbb{E}(W_n) = 1$ et $\text{Var}(W) = \lim_n \text{Var}(W_n) = \frac{\sigma^2}{\rho(\rho-1)}$.

Exercice 8.5.7 : 1) Soit $x \in [0, 1[$. Alors $g(x) = \mathbb{E}(x^Y) = \sum_k x^k p^k (1-p) = \frac{1-p}{1-px}$. $\mathbb{E}(Y) = g'(1) = \frac{p}{1-p} = \rho$.

2) Rappelons que la fonction génératrice de Z_n vaut $g \circ \dots \circ g$, n fois. La formule se vérifie aisément. Remarquons que $\rho - 1$ se met en facteur dans chaque terme.

3) Si $p < \frac{1}{2}$, alors $\frac{p}{1-p} < \frac{1}{2}$. Ainsi, l'espérance de la loi de reproduction est inférieure à 1, nous sommes dans le cas d'un branchement sous-critique et la probabilité d'extinction vaut 1.

4-a) A chaque temps n , Z_n^* est composé de un migrant, plus la première génération \hat{Z}_1 issue du migrant précédent, plus la deuxième génération issue du migrant arrivé au temps d'avant plus ... plus la population issue de l'ancêtre. Comme les racines de ces processus de branchement sont indépendantes, les arbres qui en sont issus sont indépendants, et de même loi. Ainsi, Z_n^* a même loi que $1 + \hat{Z}_1 + \dots + \hat{Z}_n$, où les \hat{Z}_n sont indépendants et \hat{Z}_n de loi Z_n .

4-b) Nous en déduisons que $\mathbb{E}(x^{Z_n^*}) = xG_1(x) \cdots G_n(x)$, où G_n est la fonction génératrice de Z_n . Ainsi, par la question 2),

$$\mathbb{E}(x^{Z_n^*}) = x \prod_{j=1}^n \frac{\rho^j - 1 - \rho x(\rho^j - 1)}{\rho^{j+1} - 1 - \rho x(\rho^j - 1)} = x \frac{\rho - 1}{\rho^{n+1} - 1 - \rho x(\rho^n - 1)}.$$

Comme $\rho < 1$, ρ^n tend vers 0. Ainsi, $\lim_{n \rightarrow \infty} \mathbb{E}(x^{Z_n^*}) = x \frac{1-\rho}{1-\rho x} = x \frac{(1-2p)}{1-p(1+x)} = L(x)$.

4 - c) Par la question précédente,

$$\begin{aligned} \mathbb{E}(x^{Z_n} | Z_n > 0) &= \frac{\mathbb{E}(x^{Z_n} \mathbf{1}_{Z_n > 0})}{\mathbb{P}(Z_n > 0)} = \frac{\mathbb{E}(x^{Z_n}) - \mathbb{P}(Z_n = 0)}{1 - \mathbb{P}(Z_n = 0)} \\ &= \frac{\frac{\rho^n - 1 - \rho x(\rho^n - 1)}{\rho^{n+1} - 1 - \rho x(\rho^n - 1)} - \frac{\rho^n - 1}{\rho^{n+1} - 1}}{1 - \frac{\rho^n - 1}{\rho^{n+1} - 1}}. \end{aligned}$$

Le calcul montre que cette probabilité conditionnelle converge vers $\frac{x(1-\rho)}{1-\rho x}$ quand n tend vers l'infini. Cette fonction génératrice est non triviale, ce qui veut dire que le processus conditionné à ne jamais s'éteindre converge vers une variable aléatoire strictement positive, et finie presque-sûrement, de même loi que la limite de Z_n^* lorsque n tend vers l'infini.

5) La loi de Z_1 est alors une loi géométrique de paramètre $\frac{1}{2}$. Ainsi l'espérance vaut 1 et le branchement est critique.

La récurrence donnant la valeur de G_n s'établit sans problème.

On a $\mathbb{P}(Z_n = 0) = \frac{n}{n+1}$. Nous en déduisons que

$$\mathbb{E}\left(\frac{Z_n}{n} \mid Z_n > 0\right) = \frac{\mathbb{E}\left(\frac{Z_n}{n} \mathbf{1}_{Z_n > 0}\right)}{\mathbb{P}(Z_n > 0)} = \frac{\frac{1}{n}}{1 - \frac{n}{n+1}} = \frac{n+1}{n},$$

qui tend vers 1 quand n tend vers l'infini. Ainsi, dans ce cas critique, l'espérance du processus conditionné à la non-extinction tend vers l'infini quand n tend vers l'infini.

Exercice 8.5.8 : (i) Nous devons montrer que $\sum_i \pi(i) Q(i, j) = \pi(j)$. Par la propriété

de réversibilité, nous avons

$$\sum_i \pi(i) Q(i, j) = \sum_i \pi(j) Q(j, i) = \pi(j) \sum_i Q(j, i) = \pi(j).$$

(ii) Soit $(i, j) \in E^2$. Nous avons

$$\mathbb{P}((X_0, X_1) = (i, j)) = \pi(i) Q(i, j) = \pi(j) Q(j, i) = \mathbb{P}((X_1, X_0) = (i, j)).$$

Nous en déduisons le résultat.

Chapitre 10

Textes et corrigés d'examens

Examen Ecole Polytechnique 2014

Problème : Impression recto-verso.

Pour décider ou non de remplacer une imprimante standard qui imprime d'un seul côté par une imprimante recto-verso plus chère, on souhaite quantifier l'économie de papier ainsi réalisée.

Les deux parties peuvent être traitées de façon indépendante. On pourra y utiliser la version suivante du théorème de Slutsky : si la suite de vecteurs aléatoires $((X_n, Y_n))_n$ à valeurs dans $\mathbb{R} \times \mathbb{R}$ est telle que $(X_n)_n$ converge en loi vers une limite déterministe x et $(Y_n)_n$ converge en loi vers Y , alors la suite $(X_n \times Y_n)_n$ converge en loi vers $x \times Y$.

I) Modélisation :

On modélise les nombres de pages $(X_i)_{i \geq 1}$ des documents successivement envoyés à l'impression par une suite de variables aléatoires à valeurs dans \mathbb{N}^* indépendantes et identiquement distribuées suivant la loi $\pi : \forall i, k \in \mathbb{N}^*, \mathbb{P}(X_i = k) = \pi(k)$ avec $\sum_{k \in \mathbb{N}^*} k\pi(k) < +\infty$.

1. On note Y_i le nombre de feuilles nécessaire pour imprimer le i -ème document en recto-verso. Exprimer la loi de cette variable aléatoire en fonction de π .
2. Exprimer $\frac{X_i - Y_i}{X_i}$ en fonction de la parité de X_i . En déduire que l'espérance de la proportion de feuilles économisées sur l'impression d'un document en

recto-verso par rapport à son impression en recto est

$$P = \frac{1}{2} - \frac{1}{2} \sum_{k \in \mathbb{N}} \frac{\pi(2k+1)}{2k+1}.$$

3. Dans cette question, on se place dans le cas $\pi(k) = \rho(1-\rho)^{k-1}$ où X_i suit la loi géométrique de paramètre $\rho \in]0, 1[$.

- (a) Vérifier que Y_i suit une loi géométrique de paramètre à préciser.
 (b) En admettant que

$$\sum_{k \in \mathbb{N}} \frac{(1-\rho)^{2k+1}}{2k+1} = \int_0^{1-\rho} \frac{1}{1-x^2} dx,$$

en déduire P .

4. (a) Que représente $P_n = 1 - \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i}$?
 (b) Vérifier que P_n converge presque sûrement vers une limite P_∞ à préciser lorsque n tend vers l'infini.
 (c) Pour quantifier l'économie à long terme, faut-il utiliser P ou bien P_∞ ?
 (d) Lorsque π est la loi géométrique de paramètre $\rho \in]0, 1[$, vérifier que $P_\infty = \frac{1-\rho}{2-\rho}$. Donner et interpréter les limites de P_∞ pour $\rho \rightarrow 0$ et $\rho \rightarrow 1$.
5. On suppose $\sum_{k \in \mathbb{N}^*} k^2 \pi(k) < +\infty$. Pour $n \in \mathbb{N}^*$, on pose $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ainsi que $\bar{W}_n = \frac{1}{n} \sum_{i=1}^n W_i$ où $W_i = Y_i - \frac{\mathbb{E}(Y_1)}{\mathbb{E}(X_1)} X_i$ pour $i \in \mathbb{N}^*$.
- (a) Montrer que la variable aléatoire W_1 est de carré intégrable et centrée. En déduire que la suite $(\sqrt{n} \bar{W}_n)_n$ converge en loi vers une limite dont on précisera les paramètres en fonction de $\text{Var}(W_1)$.
- (b) En remarquant que $\sqrt{n}(P_\infty - P_n) = \frac{1}{\bar{X}_n} \times \sqrt{n} \bar{W}_n$, conclure que la suite $(\sqrt{n}(P_\infty - P_n))_{n \in \mathbb{N}^*}$ converge en loi vers une limite que l'on précisera.

II) Analyse statistique :

On observe un échantillon (X_1, \dots, X_n) de n nombres de pages de documents envoyés à l'impression. On suppose que les X_i sont indépendantes et identiquement distribuées de loi géométrique de paramètre $\rho \in]0, 1[$.

- Pour $(x_1, \dots, x_n) \in \mathbb{N}^{*n}$, on note $p(x_1, \dots, x_n, \rho) = \mathbb{P}((X_1, \dots, X_n) = (x_1, \dots, x_n))$. Trouver la valeur $\rho(x_1, \dots, x_n)$ qui maximise $\rho \mapsto \ln(p(x_1, \dots, x_n, \rho))$ sur l'intervalle $]0, 1[$. En déduire l'estimateur du maximum de vraisemblance $\hat{\rho}_n = \rho(X_1, \dots, X_n)$ de ρ .
- On pose $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Calculer $\mathbb{E}(\bar{X}_n)$. En déduire que $\mathbb{E}(1/\bar{X}_n) \geq \rho$. Que pensez-vous du biais de l'estimateur $\hat{\rho}_n$?

3. Vérifier que $\hat{\rho}_n$ converge presque sûrement vers ρ lorsque $n \rightarrow \infty$.
4. En remarquant que $\sqrt{n}(\hat{\rho}_n - \rho) = \frac{\rho}{\bar{X}_n} \times \sqrt{n}\left(\frac{1}{\rho} - \bar{X}_n\right)$ montrer que $\sqrt{n}(\hat{\rho}_n - \rho)$ converge en loi vers une limite dont on précisera les paramètres en fonction de ρ .

Exercice : Courte paille.

On considère le jeu de la courte paille entre deux joueurs. Une troisième personne choisit un bâton puis elle le coupe en deux. Elle présente alors les deux morceaux aux joueurs en cachant la différence de longueur dans sa main. Le premier joueur choisit un morceau et le second prend l'autre. Celui qui a le morceau le plus long gagne. La longueur U du bâton est supposée suivre la loi uniforme sur $[0, 1]$ et les deux morceaux finaux ont pour longueur $X = UV$ et $Y = U(1 - V)$ avec V de loi uniforme sur $[0, 1]$ indépendante de U .

1. Calculer $\mathbb{E}[X]$ et $\text{Var}[X]$.
2. Donner la loi de $-\ln(U)$. En déduire sans calcul que $\xi = -\ln(X)$ suit la loi $\Gamma(2, 1)$. Déterminer la loi de X .
3. Quelle est la loi de (X, Y) (faire attention au domaine image du changement de variables)? Les longueurs des morceaux sont-elles indépendantes? Comparer les lois de (Y, X) et de (X, Y) . Le jeu est-il équitable? Retrouver la densité de X .
4. On vous propose de choisir le premier un morceau et de décider si vous en changez avant de regarder le second morceau. Dans cette question, on cherche à déterminer une stratégie qui maximise la probabilité de gagner. Soit w un seuil dans $]0, 1[$. On note $Z = \mathbf{1}_{\{w < X\}}X + \mathbf{1}_{\{w \geq X\}}Y$.

(a) Montrer que

$$\{Z > \min(X, Y)\} = \{\{X > Y\} \cap \{w < X\}\} \cup \{\{X < Y\} \cap \{w \geq X\}\}.$$

- (b) Pour $w \in [1/2, 1[$, en déduire que $\mathbb{P}(Z > \min(X, Y)) = \mathbb{P}(w < X) + \frac{1}{2}$ puis calculer $q(w) = \mathbb{P}(w < X)$.
- (c) Remarquer que $\mathbb{P}(X < Y, w \geq X) = \mathbb{P}(X > Y, w \geq X) + \mathbb{P}(X > w \geq Y)$ et en déduire que si $w \in]0, 1/2]$,

$$\mathbb{P}(Z > \min(X, Y)) = \frac{1}{2} + q(w) \text{ avec } q(w) = w(1 - \ln(4w)).$$

(d) Trouver w^* qui maximise q et donner $q(w^*)$. Commenter.

Corrigé de l'examen Ecole Polytechnique 2014

Problème : Impression recto-verso.

I) Modélisation :

1. Pour $k \in \mathbb{N}^*$, $\{Y_i = k\} = \{X_i = 2k - 1\} \cup \{X_i = 2k\}$, d'où $\mathbb{P}(Y_i = k) = \pi(2k - 1) + \pi(2k)$.
2. $\frac{X_i - Y_i}{X_i} = \begin{cases} \frac{X_i - X_i/2}{X_i} = \frac{1}{2} & \text{si } X_i \text{ pair} \\ \frac{X_i - (X_i+1)/2}{X_i} = \frac{1}{2} - \frac{1}{2X_i} & \text{si } X_i \text{ impair} \end{cases}$ d'où $P = \mathbb{E}\left(\frac{X_i - Y_i}{X_i}\right) = \frac{1}{2} - \frac{1}{2} \sum_{k \in \mathbb{N}} \frac{\pi(2k+1)}{2k+1}$.
3. (a) D'après la question 1, pour $k \in \mathbb{N}^*$, $\mathbb{P}(Y_i = k) = \rho(1-\rho)^{2k-2} + \rho(1-\rho)^{2k-1} = (2\rho - \rho^2)(1-\rho)^{2k-2} = (2\rho - \rho^2)(1 - (2\rho - \rho^2))^{k-1}$ si bien que Y_i suit la loi géométrique de paramètre $2\rho - \rho^2$.

(b) On a

$$\begin{aligned} \sum_{k=0}^{+\infty} \frac{(1-\rho)^{2k+1}}{2k+1} &= \sum_{k=0}^{+\infty} \int_0^{1-\rho} x^{2k} dx = \int_0^{1-\rho} \sum_{k=0}^{+\infty} (x^2)^k dx = \int_0^{1-\rho} \frac{dx}{1-x^2} \\ &= \int_0^{1-\rho} \frac{1}{2(1+x)} + \frac{1}{2(1-x)} dx = \frac{1}{2} \left[\ln\left(\frac{1+x}{1-x}\right) \right]_0^{1-\rho} \\ &= \frac{1}{2} \ln\left(\frac{2-\rho}{\rho}\right). \end{aligned}$$

Donc $P = \frac{1}{2} - \frac{1}{2} \sum_{k=0}^{+\infty} \frac{\rho(1-\rho)^{2k}}{2k+1} = \frac{1}{2} - \frac{\rho}{2(1-\rho)} \sum_{k=0}^{+\infty} \frac{(1-\rho)^{2k+1}}{2k+1} = \frac{1}{2} - \frac{\rho}{4(1-\rho)} \ln\left(\frac{2-\rho}{\rho}\right)$.

4. (a) P_n représente la proportion de feuilles réellement économisées au bout de n documents envoyés à l'impression.
- (b) Les variables $(X_i)_{i \geq 1}$ étant i.i.d. et intégrables, la loi forte des grands nombres assure $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p.s.} n \rightarrow \infty \mathbb{E}(X_1)$. La caractéristique i.i.d. des X_i se transmet aux Y_i qui sont intégrables puisque $1 \leq Y_i \leq X_i$. Donc $\frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{p.s.} n \rightarrow \infty \mathbb{E}(Y_1)$. Comme l'application $(x, y) \mapsto 1 - \frac{y}{x}$ est continue sur $\mathbb{R}^* \times \mathbb{R}$ et que $\mathbb{E}(X_1) \geq 1$, on conclut que $P_n = 1 - \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i} \xrightarrow{p.s.} n \rightarrow \infty 1 - \frac{\mathbb{E}(Y_1)}{\mathbb{E}(X_1)} = P_\infty$.
- (c) Il faut utiliser la limite P_∞ de la proportion de feuilles économisées lorsque le nombre de documents imprimés devient très grand.
- (d) D'après la proposition 3.5.6 et la question 3a, $\mathbb{E}(X_1) = \frac{1}{\rho}$, $\mathbb{E}(Y_1) = \frac{1}{\rho(2-\rho)}$ si bien que $P_\infty = \frac{1-\rho}{2-\rho}$. Pour $\rho \rightarrow 0$, $P_\infty \rightarrow 1/2$: les nombres de pages des documents imprimés sont grands ($\mathbb{E}(X_1) = 1/\rho \rightarrow +\infty$) si bien que l'influence de leur parité devient négligeable. L'économie est donc pratiquement de la moitié. À l'inverse, lorsque $\rho \rightarrow 1$, on n'envoie pratiquement à l'impression que des documents ne comportant qu'une seule page. On ne gagne donc pratiquement rien à faire du recto-verso. On retrouve bien $P_\infty \rightarrow 0$.
5. (a) Comme $|W_i| \leq (1 + \frac{\mathbb{E}(Y_1)}{\mathbb{E}(X_1)})X_i$, l'intégrabilité de X_i^2 entraîne celle de W_i^2 . Par linéarité de l'espérance, $\mathbb{E}(W_i) = \mathbb{E}(Y_i) - \frac{\mathbb{E}(Y_1)}{\mathbb{E}(X_1)} \mathbb{E}(X_i) = 0$. Le caractère i.i.d.

des couples (X_i, Y_i) se transmet aux variables aléatoires W_i . Le théorème de la limite centrale assure que $\sqrt{n}\bar{W}_n = \sqrt{n}(\bar{W}_n - \mathbb{E}(W_1)) \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} V \sim \mathcal{N}_1(0, \text{Var}(W_1))$.

- (b) La question précédente, la convergence en loi de $\frac{1}{\bar{X}_n}$ vers la constante $\frac{1}{\mathbb{E}(X_1)}$ déduite de la convergence presque sûre établie à la question 4b et la version du théorème de Slutsky énoncée au début du problème assurent que

$$\sqrt{n}(P_\infty - P_n) = \frac{1}{\bar{X}_n} \times \sqrt{n}\bar{W}_n \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} \frac{V}{\mathbb{E}(X_1)} \sim \mathcal{N}_1\left(0, \text{Var}\left(\frac{Y_1}{\mathbb{E}(X_1)} - \frac{X_1\mathbb{E}(Y_1)}{(\mathbb{E}(X_1))^2}\right)\right).$$

II) Analyse statistique :

1. Par indépendance des $X_i \sim \text{Geo}(\rho)$, $p(x_1, \dots, x_n; \rho) = \rho^n(1 - \rho)^{\sum_{i=1}^n x_i - n}$.

Donc

$$\frac{\partial \ln p}{\partial \rho}(x_1, \dots, x_n, \rho) = \frac{n}{\rho} - \frac{\sum_{i=1}^n x_i - n}{1 - \rho} = \frac{n - \rho \sum_{i=1}^n x_i}{\rho(1 - \rho)}.$$

est positive sur $]0, \frac{n}{\sum_{i=1}^n x_i}]$ et négative sur $[\frac{n}{\sum_{i=1}^n x_i}, 1[$. Ainsi $\rho(x_1, \dots, x_n) = \frac{n}{\sum_{i=1}^n x_i}$ et

$$\hat{\rho}_n = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}_n}.$$

2. $\mathbb{E}(\bar{X}_n) = \mathbb{E}(X_1) = 1/\rho$. Comme la fonction $x \mapsto 1/x$ est convexe sur $]0, +\infty[$, l'inégalité de Jensen entraîne que $\mathbb{E}(1/\bar{X}_n) \geq 1/(1/\rho) = \rho$. L'inégalité est même stricte car la fonction $x \mapsto 1/x$ est strictement convexe et la variable aléatoire \bar{X}_n non constante. L'estimateur $\hat{\rho}_n$ est donc biaisé.
3. La loi forte des grands nombres entraîne que $\bar{X}_n \xrightarrow{p.s.}_{n \rightarrow \infty} \mathbb{E}(X_1) = 1/\rho$. On conclut par continuité de $x \mapsto \frac{1}{x}$ en ce point.
4. Le théorème de la limite centrale assure que $\sqrt{n}(\frac{1}{\rho} - \bar{X}_n) \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} Z \sim \mathcal{N}_1(0, \text{Var}(X_1))$ où $\text{Var}(X_1) = \frac{1-\rho}{\rho^2}$. Avec l'égalité de l'énoncé, la convergence en loi de $\left(\frac{\rho}{\bar{X}_n}\right)_n$ vers ρ^2 déduite de la question précédente et la généralisation du théorème de Slutsky donnée au début du problème, on conclut que $\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{\mathcal{L}}_{n \rightarrow \infty} \rho^2 Z \sim \mathcal{N}_1(0, \rho^2(1 - \rho))$.

Exercice : Courte paille.

1. Puisque U et V sont indépendantes, $\mathbb{E}[X] = \mathbb{E}[UV] = \mathbb{E}[U]\mathbb{E}[V] = 1/4$. De même, $\mathbb{E}[X^2] = \mathbb{E}[U^2]\mathbb{E}[V^2] = 1/9$ et $\text{Var}(X) = 1/9 - 1/16 = 7/144$.

2. D'après la technique de simulation du paragraphe 4.6.2, la v.a. $-\ln(U)$ suit la loi exponentielle de paramètre 1 qui est également la loi $\Gamma(1, 1)$. Par la proposition 4.10.4, $\xi = -\ln(U) - \ln(V)$ suit la loi $\Gamma(2, 1)$. Pour $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ bornée, on a donc

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(e^{-\xi})] = \int_0^\infty \varphi(e^{-z}) z e^{-z} dz = - \int_0^1 \varphi(x) \ln(x) dx$$

en effectuant le changement de variables $x = e^{-z}$. Donc X possède la densité $-\mathbf{1}_{\{0 < x < 1\}} \ln(x)$.

3. Soient $u, v \in]0, 1[$. On a

$$\begin{cases} x = uv \\ y = u(1-v) \end{cases} \iff \begin{cases} u = x+y \\ v = \frac{x}{x+y} \end{cases}$$

et

$$\begin{cases} 0 < u < 1 \\ 0 < v < 1 \end{cases} \iff \begin{cases} 0 < x+y < 1 \\ 0 < \frac{x}{x+y} < 1 \end{cases} \iff \begin{cases} 0 < x+y < 1 \\ 0 < x, 0 < y \end{cases}.$$

Enfin, la matrice jacobienne $\frac{D(x,y)}{D(u,v)} = \begin{pmatrix} v & u \\ 1-v & -u \end{pmatrix}$ a pour déterminant $-u$ si bien que $\frac{dxdy}{dudv} = |u| = |x+y|$. Pour une fonction $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ bornée, on a donc :

$$\begin{aligned} \mathbb{E}[\varphi(X, Y)] &= \int_0^1 \int_0^1 \varphi(uv, u(1-v)) dudv \\ &= \int_{\mathbb{R}^2} \varphi(x, y) \mathbf{1}_{\{x>0\}} \mathbf{1}_{\{y>0\}} \mathbf{1}_{\{x+y<1\}} \frac{dxdy}{|x+y|}. \end{aligned}$$

La densité de (X, Y) vaut donc

$$p(x, y) = \mathbf{1}_{\{x>0\}} \mathbf{1}_{\{y>0\}} \mathbf{1}_{\{x+y<1\}} \frac{1}{x+y}$$

Du fait du facteur $\frac{\mathbf{1}_{\{x+y<1\}}}{x+y}$, cette densité ne peut pas se mettre sous forme du produit d'une fonction de x par une fonction de y . Les variables aléatoires X et Y ne sont donc pas indépendantes. Comme p est une fonction symétrique, le couple (X, Y) a même loi que le couple (Y, X) , ce qui entraîne que $\mathbb{P}(X < Y) = \mathbb{P}(Y < X) = \frac{1 - \mathbb{P}(X=Y)}{2} = \frac{1-0}{2}$ si bien que le jeu est équitable. Enfin, la densité marginale de X est

$$\begin{aligned} \int_{\mathbb{R}} p(x, y) dy &= \mathbf{1}_{\{0 < x\}} \int_{\mathbb{R}} \mathbf{1}_{\{0 < y < 1-x\}} \frac{dy}{x+y} = \mathbf{1}_{\{0 < x < 1\}} \int_0^{1-x} \frac{dy}{x+y} \\ &= \mathbf{1}_{\{0 < x < 1\}} [\ln(x+y)]_0^{1-x}. \end{aligned}$$

4. (a) On a $\{X > Y\} \cap \{Z > \min(X, Y)\} = \{X > Y\} \cap \{w < X\}$ et $\{X < Y\} \cap \{Z > \min(X, Y)\} = \{X < Y\} \cap \{w \geq X\}$ d'où on déduit

$$\{Z > \min(X, Y)\} = \{\{X > Y\} \cap \{w < X\}\} \cup \{\{X < Y\} \cap \{w \geq X\}\}.$$

- (b) Comme l'union est disjointe, $\mathbb{P}(Z > \min(X, Y)) = \mathbb{P}(\{X > Y\} \cap \{w < X\}) + \mathbb{P}(\{X < Y\} \cap \{w \geq X\})$. Lorsque $w \in [1/2, 1[$, comme $X + Y = U$ prend ses valeurs dans $[0, 1]$, $\{\{X > Y\} \cap \{w < X\}\} = \{w < X\}$ et $\{X < Y\} \cap \{w \geq X\} = \{X < Y\}$. Donc $\mathbb{P}(Z > \min(X, Y)) = \mathbb{P}(w < X) + \frac{1}{2}$ et $q(w) = \mathbb{P}(w \leq X) = -\int_w^1 \ln(x) dx = 1 - w + w \ln w$.
- (c) Comme (X, Y) a même loi que (Y, X) ,

$$\begin{aligned} \mathbb{P}(X < Y, w \geq X) &= \mathbb{P}(Y < X, w \geq Y) \\ &= \mathbb{P}(\{Y < X, w \geq X\} \cup \{Y < X, X > w \geq Y\}) \\ &= \mathbb{P}(Y < X, w \geq X) + \mathbb{P}(X > w \geq Y). \end{aligned}$$

Donc

$$\begin{aligned} \mathbb{P}(Z > \min(X, Y)) &= \mathbb{P}(X > Y, X > w) + \mathbb{P}(X > Y, w \geq X) + \mathbb{P}(X > w \geq Y) \\ &= \mathbb{P}(X > Y) + \mathbb{P}(X > w \geq Y) = \frac{1}{2} + \mathbb{P}(X > w \geq Y). \end{aligned}$$

Par ailleurs, pour $w \in]0, 1/2]$,

$$\begin{aligned} \mathbb{P}(X > w \geq Y) &= \int \mathbf{1}_{\{x > w \geq y\}} \mathbf{1}_{\{x > 0\}} \mathbf{1}_{\{y > 0\}} \mathbf{1}_{\{x+y < 1\}} \frac{1}{x+y} dx dy \\ &= \int_0^w \left(\int_w^{1-y} \frac{dx}{x+y} \right) dy = - \int_0^w \ln(w+y) dy \\ &= -[(w+y)(\ln(w+y) - 1)]_0^w = w(1 - \ln(4w)). \end{aligned}$$

- (d) La dérivée $q'(w) = \begin{cases} -\ln(4w) & \text{si } w \leq \frac{1}{2} \\ \ln(w) & \text{si } w \geq \frac{1}{2} \end{cases}$ est strictement positive sur $]0, 1/4[$ et strictement négative sur $]1/4, 1[$. Donc $q(w)$ atteint son maximum en $w^* = 1/4$ et $q(w^*) = 1/4$. En changeant le morceau tiré si sa longueur est inférieure à $1/4$, vous avez trois chances sur quatre de gagner.

Examen Ecole Polytechnique 2015

Problème : Estimation de densité

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

On s'intéresse ici au problème suivant : on observe des variables aléatoires $(X_i)_{i \geq 1}$ i.i.d. réelles de loi de densité f continûment différentiable et strictement positive, et de fonction de répartition F . On a donc que pour tout i

$$\mathbb{P}(X_i \in [a, b]) = F(b) - F(a) = \int_a^b f(x) dx.$$

Soit $x \in \mathbb{R}$. On pose

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}$$

et pour $h_n > 0$ que l'on fixera plus tard

$$\widehat{f}_n(x) = \frac{F_n(x + h_n) - F_n(x - h_n)}{2h_n}.$$

Le but de ce problème est d'étudier la convergence de \widehat{f}_n vers f .

1. Montrer que pour tout x , $(F_n(x))_{n \geq 1}$ tend presque sûrement vers $F(x)$.
2. Montrer que pour tout $\omega \in \Omega$, la fonction aléatoire $x \rightarrow \widehat{f}_n(x)$ est une densité de probabilité.
3. On va maintenant donner des conditions sur h_n pour que la convergence de $\widehat{f}_n(x)$ vers $f(x)$ soit vérifiée. On fixe donc $x \in \mathbb{R}$.
 - (a) Montrer que

$$\mathbb{E} \left[\left(\widehat{f}_n(x) - f(x) \right)^2 \right] = \left(\mathbb{E} \left[\widehat{f}_n(x) \right] - f(x) \right)^2 + \text{Var} \left(\widehat{f}_n(x) \right).$$

On appelle cette décomposition de l'erreur quadratique une décomposition biais-variance.

- (b) Montrer que si $h_n \rightarrow 0$ alors $\mathbb{E} \left[\widehat{f}_n(x) \right] \rightarrow f(x)$.
- (c) Montrer que $2nh_n \widehat{f}_n(x)$ suit une loi Binomiale $\mathcal{B}(n, p_{n,x})$ et préciser $p_{n,x}$.
- (d) En déduire que

$$\text{Var} \left(\widehat{f}_n(x) \right) \leq \frac{1}{4nh_n} \frac{p_{n,x}}{h_n}.$$

- (e) En déduire que si $h_n \rightarrow 0$ mais $nh_n \rightarrow \infty$ alors

$$\mathbb{E} \left[\left(\widehat{f}_n(x) - f(x) \right)^2 \right] \rightarrow 0$$

puis que $\widehat{f}_n(x) \rightarrow f(x)$ en probabilité.

4. Nous allons maintenant montrer que la convergence de $\widehat{f}_n(x)$ vers $f(x)$ est en un certain sens uniforme. On note $x_+ = \max(x, 0)$. On suppose ici $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$.

(a) Montrer que

$$\mathbb{E} \left[\int |\hat{f}_n(x) - f(x)| dx \right] = 2\mathbb{E} \left[\int (f(x) - \hat{f}_n(x))_+ dx \right].$$

(b) On note $\phi_n(x) := \mathbb{E} \left[(f(x) - \hat{f}_n(x))_+ \right]$. Montrer que pour tout x , $\phi_n(x) \rightarrow 0$.

(c) Montrer que $\phi_n \leq f$ et en déduire que

$$\mathbb{E} \left[\int |\hat{f}_n(x) - f(x)| dx \right] \rightarrow 0.$$

5. On s'intéresse maintenant au théorème central limite associé à cet estimateur.

(a) Soit $S_n \sim \mathcal{B}(n, p_n)$ avec $np_n \rightarrow \infty$, $p_n \rightarrow 0$ et $0 < p_n < 1$. On note

$$S_n^* = \frac{S_n - np_n}{\sqrt{np_n(1-p_n)}}.$$

Montrer que pour tout $t \in \mathbb{R}$

$$\mathbb{E} \left[e^{itS_n^*} \right] \rightarrow e^{-t^2/2}.$$

Indication : On pourra poser $t_n = \frac{t}{\sqrt{np_n(1-p_n)}}$ et remarquer que $p_n t_n^2 = O(n^{-1})$ et donc $p_n^2 t_n^2 = O(n^{-1})$.

(b) Montrer que

$$\begin{aligned} \sqrt{2nh_n} \frac{\hat{f}_n(x) - f(x)}{\sqrt{f(x)}} &= \sqrt{\frac{p_{n,x}(1-p_{n,x})}{2h_n f(x)}} \times \frac{2nh_n \hat{f}_n(x) - np_{n,x}}{\sqrt{np_{n,x}(1-p_{n,x})}} \\ &\quad + \sqrt{\frac{2nh_n}{f(x)}} \times \left(\frac{p_{n,x}}{2h_n} - f(x) \right). \end{aligned}$$

En déduire que si $nh_n^3 \rightarrow 0$ et $nh_n \rightarrow \infty$ alors

$$\sqrt{2nh_n} \frac{\hat{f}_n(x) - f(x)}{\sqrt{f(x)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Problème : Inégalité de transport et concentration.

Définition : on dit qu'une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est lipschitzienne si

$$\exists L > 0, \forall x, y, |f(x) - f(y)| \leq L|x - y|.$$

On s'intéresse ici à une inégalité dite de transport-information (forme duale) : la variable aléatoire réelle X intégrable satisfait l'inégalité $T_1(C)$ si il existe une constante $C > 0$ telle que pour toute fonction lipschitzienne $f : \mathbb{R} \rightarrow \mathbb{R}$ et pour tout λ

$$(*) \quad \mathbb{E} \left[e^{\lambda f(X)} \right] \leq e^{\lambda \mathbb{E}[f(X)] + \frac{\lambda^2 C \|f\|_{Lip}^2}{2}},$$

où l'on note $\|f\|_{Lip} := \sup_{x,y} \frac{|\phi(x) - \phi(y)|}{|x-y|}$. (On remarquera que f lipschitzienne et X intégrable assure $\mathbb{E}(|f(X)|) < \infty$.)

Notre but ici va être double : donner une application de cette inégalité, puis donner une condition nécessaire et suffisante pour la vérifier.

On supposera dans tout le problème que X a une densité notée g_X .

1. Soit (X_i) une suite de variable réelles indépendantes et identiquement distribuées vérifiant $(*)$ (avec la même constante C), on va d'abord s'intéresser à une version quantitative et non asymptotique de la Loi des Grands Nombres.

(a) Montrer que pour tout λ ,

$$\mathbb{E} \left[e^{\lambda \sum_{i=1}^n (X_i - \mathbb{E}[X])} \right] \leq e^{\frac{n\lambda^2 C}{2}}.$$

(b) Montrer que pour toute variable aléatoire réelle Y intégrable et pour tout $\lambda, r > 0$,

$$\mathbb{P}(Y - \mathbb{E}(Y) \geq r) \leq e^{-\lambda r} \mathbb{E} \left(e^{\lambda(Y - \mathbb{E}(Y))} \right).$$

(c) En déduire que pour tout $r > 0$

$$\mathbb{P} \left(\frac{1}{n} \sum_{k=1}^n X_k - \mathbb{E}(X) \geq r \right) \leq e^{-n \frac{r^2}{2C}}.$$

2. On va maintenant montrer que si il existe $c > 0$ telle que

$$(**) \quad \mathbb{E} \left[e^{c|X|^2} \right] < \infty$$

alors $(*)$ est vérifiée. Nous supposons donc $(**)$ vérifiée dans toute cette question. (Cela assure notamment la finitude de $\mathbb{E} \left[e^{\lambda f(X)} \right]$ pour toute fonction f lipschitzienne et tout λ .)

Soit f une fonction lipschitzienne telle que $\mathbb{E}[f(X)] = 0$ et $\|f\|_{Lip} = 1$.

On note X' une variable aléatoire indépendante de X et de même loi que X .

(a) Montrer que pour tout λ , $\mathbb{E} \left[e^{-\lambda f(X')} \right] \geq 1$, et en déduire

$$\mathbb{E} \left[e^{\lambda f(X)} \right] \leq \mathbb{E} \left[e^{\lambda(f(X) - f(X'))} \right]$$

- (b) Montrer que $f(X) - f(X')$ et $f(X') - f(X)$ ont même loi et en déduire que pour tout $k \in \mathbb{N}$, $\mathbb{E}[(f(X) - f(X'))^{2k+1}] = 0$.
- (c) En déduire que

$$\mathbb{E} \left[e^{\lambda f(X)} \right] \leq \mathbb{E} \left[e^{\lambda(f(X) - f(X'))} \right] \leq 1 + \sum_{k \geq 1} \frac{\lambda^{2k}}{(2k)!} \mathbb{E} [|X - X'|^{2k}].$$

- (d) En notant $C := 2 \sup_{k \geq 1} \left(\frac{k! \mathbb{E} [|X - X'|^{2k}]}{(2k)!} \right)^{1/k}$, montrer que

$$C \leq \frac{4}{c} \sup_{k \geq 1} \left(\frac{(k!)^2}{(2k)!} \mathbb{E} \left[e^{\frac{c}{2} |X - X'|^2} \right] \right)^{1/k} < \infty.$$

Rappel : équivalent de Stirling $n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$.

- (e) Montrer que $(**)$ implique $(*)$ soit avec C définie dans la question précédente

$$\mathbb{E} \left[e^{\lambda f(X)} \right] \leq e^{C \frac{\lambda^2}{2}}.$$

3. Nous allons maintenant montrer la réciproque : soit $N \sim \mathcal{N}(0, 1)$ indépendante de X .

- (a) Calculer $\mathbb{E}(e^{aN})$ pour a fixé.
- (b) Montrer que pour $a^2 \leq 1$, $\mathbb{E} \left[e^{\frac{a^2}{2} N^2} \right] = \frac{1}{\sqrt{1-a^2}}$.
- (c) Montrer que pour a donné

$$\mathbb{E} \left[e^{\frac{a^2}{2} (|X| - \mathbb{E}(|X|))^2} \right] = \mathbb{E} \left[e^{aN(|X| - \mathbb{E}(|X|))} \right].$$

- (d) En utilisant $(*)$, montrer que si $a \leq \sqrt{C}$ alors

$$\mathbb{E} \left[e^{\frac{a^2}{2} (|X| - \mathbb{E}(|X|))^2} \right] \leq \frac{1}{\sqrt{1-a^2 C}}$$

et conclure.

Corrigé de l'examen Ecole Polytechnique 2015

1. Comme les (X_i) sont i.i.d., $(\mathbf{1}_{\{X_i \leq x\}})$ sont i.i.d. et bornées et on peut donc appliquer la Loi forte des Grands Nombres, et donc

$$F_n(x) \xrightarrow{p.s.} \mathbb{E}(\mathbf{1}_{\{X_i \leq x\}}) = F(x).$$

2. La fonction $\hat{f}_n(x)$ est positive, il reste à vérifier que son intégrale vaut 1

$$\int \hat{f}_n(x) dx = \frac{1}{n} \sum_{i=1}^n \frac{1}{2h_n} \int_{X_i-h_n}^{X_i+h_n} dx = 1.$$

3. (a)

$$\begin{aligned} \mathbb{E} \left[\left(\hat{f}_n(x) - f(x) \right)^2 \right] &= \mathbb{E} \left[\left(\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)] + \mathbb{E}[\hat{f}_n(x)] - f(x) \right)^2 \right] \\ &= \mathbb{E} \left[\left(\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)] \right)^2 \right] + \left(\mathbb{E} \left[\hat{f}_n(x) \right] - f(x) \right)^2 \\ &\quad + 2 \left(\mathbb{E} \left[\hat{f}_n(x) \right] - f(x) \right) \mathbb{E} \left[\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)] \right] \\ &= \text{Var} \left(\hat{f}_n(x) \right) + \left(\mathbb{E} \left[\hat{f}_n(x) \right] - f(x) \right)^2. \end{aligned}$$

(b) Par définition de la dérivée :

$$\mathbb{E}[\hat{f}_n(x)] = \frac{1}{2h_n} \int_{x-h_n}^{x+h_n} f(u) du \rightarrow f(x).$$

(c) Par définition $(\mathbf{1}_{\{X_i \leq x+h_n\}} - \mathbf{1}_{\{X_i \leq x-h_n\}})$ est une famille de variable i.i.d. prenant les valeurs 0 ou 1 et donc de Bernoulli de paramètre $\int_{x-h_n}^{x+h_n} f(u) du = p_{n,x}$. Une somme i.i.d. de Bernoulli suit alors une loi binomiale $\mathcal{B}(n, p_{n,x})$.

(d) On a $\text{Var}(\mathcal{B}(n, p)) = np(1-p)$ donc

$$\text{Var} \left(\hat{f}_n(x) \right) = \frac{np_{n,x}(1-p_{n,x})}{4n^2h_n^2} \leq \frac{1}{4nh_n} \frac{p_{n,x}}{h_n}.$$

(e) La décomposition biais-variance entraîne qu'il faut montrer que le biais tend vers 0 ce qui fait en question (c), puis que la variance tend vers 0. on s'en assure par la question (e) car $p_{n,x}/h_n$ tend vers $2f(x)$ et $nh_n \rightarrow \infty$. La convergence L^2 implique la convergence en probabilité.

4. (a) On a par 3.(a),

$$\begin{aligned} 0 &= \int (\hat{f}_n(x) - f(x)) dx \\ &= \int (\hat{f}_n(x) - f(x)) \mathbf{1}_{\hat{f}_n(x) \leq f(x)} dx + \int (\hat{f}_n(x) - f(x)) \mathbf{1}_{\hat{f}_n(x) \geq f(x)} dx. \end{aligned}$$

De plus,

$$\int |\hat{f}_n(x) - f(x)| dx = \int (f(x) - \hat{f}_n(x)) \mathbf{1}_{\hat{f}_n(x) \leq f(x)} dx + \int (\hat{f}_n(x) - f(x)) \mathbf{1}_{\hat{f}_n(x) \geq f(x)} dx$$

et la conclusion suit.

- (b) La conclusion provient de Cauchy-Schwarz et de la convergence L^2 de la question 3(f).
- (c) Par définition $x \rightarrow x_+$ est croissante et donc $(f - \widehat{f}_n(x))_+ \leq f$ qui n'est pas aléatoire. On applique ensuite le théorème de convergence dominée.
5. (a) On va ici utiliser le théorème de Lévy et donc s'intéresser à la convergence des fonctions caractéristiques. Notamment on connaît la fonction caractéristique d'une loi binomiale.

$$\begin{aligned} \mathbb{E} \left[e^{itS_n^*} \right] &= e^{-int_n p_n} \mathbb{E} \left[e^{it_n S_n} \right] \\ &= \left(e^{-it_n p_n} (p_n e^{it_n} + (1 - p_n)) \right)^n \\ &= \left(1 - ip_n t_n - \frac{1}{2} p_n^2 t_n^2 + o(1/n) \right)^n \left(1 + ip_n t_n - \frac{1}{2} p_n t_n^2 + o(1/n) \right)^n \\ &= \left(1 - \frac{1}{2} p_n (1 - p_n) t_n^2 + o(1/n) \right)^n \\ &\rightarrow e^{-\frac{t^2}{2}} \end{aligned}$$

- (b) On remarque que $p_{n,x} \sim 2h_n f(x)$ et donc $np_n(x) \rightarrow \infty$ vu que $nh_n \rightarrow \infty$. Par ailleurs, $h_n \rightarrow 0$ car $nh_n^3 \rightarrow 0$. Pour pouvoir utiliser le résultat de 5.(a), il suffit de recentrer par la moyenne plutôt que par $f(x)$ comme suggéré. On montre sans peine qu'il existe D_x borné, car f est continue, telle que

$$\sqrt{2nh_n} |\mathbb{E}[\widehat{f}_n(x)] - f(x)| \leq D_x \sqrt{nh_n^3}$$

qui tend vers 0 par hypothèse. La conclusion suit par Slutsky.

Problème : Inégalité de transport et concentration.

1. (a) C'est une conséquence directe de l'indépendance des (X_i) et de la propriété (*) appliquée à chaque espérance.
- (b) C'est une conséquence directe de Markov et du fait que $x \rightarrow e^{\lambda x}$ est strictement croissante.
- (c) Par l'inégalité de Markov et la question 1.(a), on a donc pour tout λ ,

$$\mathbb{P} \left(\frac{1}{n} \sum_{k=1}^n X_k - \mathbb{E}(X) \geq r \right) \leq e^{-\lambda nr} e^{\frac{n\lambda^2 C}{2}}.$$

On choisit ensuite λ qui minimise $-\lambda nr + \frac{n\lambda^2 C}{2}$ soit $\lambda = \frac{r}{C}$ ce qui donne la borne voulue.

2. (a) Par Jensen, comme $x \rightarrow e^{-x}$ est convexe, $\mathbb{E} \left[e^{-\lambda f(X')} \right] \geq e^{-\mathbb{E}[f(X)]} = 1$.
Puis, vu que X et X' sont indépendantes

$$\mathbb{E} \left[e^{\lambda f(X)} \right] \leq \mathbb{E} \left[e^{\lambda f(X)} \right] \mathbb{E} \left[e^{-\lambda f(X')} \right] \mathbb{E} \left[e^{\lambda(f(X)-f(X'))} \right].$$

- (b) X et X' sont indépendantes et ont même loi donc $f(X) - f(X') \sim f(X') - f(X) \sim -(f(X) - f(X'))$ et donc $f(X) - f(X')$ est une variable symétrique. Ses moments impairs sont donc nuls !
(c) Nous allons simplement utiliser le développement en série entière de la fonction exponentielle et le fait que les moments impairs de $f(X) - f(X')$ sont nuls.

$$\begin{aligned} \mathbb{E} \left[e^{\lambda(f(X)-f(X'))} \right] &= 1 + \sum_{k \geq 1} \frac{\lambda^k}{k!} \mathbb{E} \left[(f(X) - f(X'))^k \right] \\ &= 1 + \sum_{k \geq 1} \frac{\lambda^{2k}}{(2k)!} \mathbb{E} \left[(f(X) - f(X'))^{2k} \right] \\ &= 1 + \sum_{k \geq 1} \frac{\lambda^{2k}}{(2k)!} \mathbb{E} \left[|X - X'|^{2k} \right] \end{aligned}$$

où nous utilisons finalement que $\|f\|_{Lip} \leq 1$.

- (d) Par le développement en série de l'exponentielle on a pour $c > 0$,

$$\mathbb{E} \left[|X - X'|^{2k} \right] \leq 2^k \frac{k!}{c^k} \mathbb{E} \left[e^{\frac{c}{2}|X-X'|^2} \right]$$

et la borne s'en déduit directement. Le fait que cette borne soit finie se déduit de la formule de Stirling et du fait que $\mathbb{E} \left[e^{\frac{c}{2}|X-X'|^2} \right] \leq \mathbb{E} \left[e^{c|X|^2} \right]^2$.

- (e) La définition de C implique

$$\frac{\lambda^{2k}}{(2k)!} \mathbb{E} \left[|X - X'|^{2k} \right] \leq \frac{1}{k!} \left(\frac{C\lambda^2}{2} \right)^k$$

et il ne reste plus qu'à utiliser le développement en série entière de l'exponentielle.

3. (a)

$$\mathbb{E} \left(e^{aN} \right) = \int e^{ax - \frac{x^2}{2}} (2\pi)^{-1/2} dx = e^{\frac{a^2}{2}} \int e^{-\frac{(x-a)^2}{2}} (2\pi)^{-1/2} dx = e^{\frac{a^2}{2}}.$$

- (b) C'est une conséquence directe de la définition de la densité d'une gaussienne.

(c) C'est une application directe de l'indépendance de X et de N et de 3(a).

En effet

$$\mathbb{E} \left[e^{aN(|X| - \mathbb{E}(|X|))} \right] = \mathbb{E} \left[\mathbb{E} \left[e^{aN(|X| - \mathbb{E}(|X|))} \mid X \right] \right] = \mathbb{E} \left[e^{\frac{a^2}{2}(|X| - \mathbb{E}(|X|))^2} \right].$$

(d) Conditionnellement à N indépendante de X , la fonction $x \rightarrow aN(|x| - \mathbb{E}(|X|))$ est $|aN|$ -lipschitzienne et donc on peut appliquer (*) conditionnellement à N pour obtenir

$$\mathbb{E} \left[e^{aN(|X| - \mathbb{E}(|X|))} \right] \leq \mathbb{E} \left[e^{\frac{Ca^2}{2}N^2} \right].$$

Ensuite

$$\mathbb{E} \left[e^{\frac{Ca^2}{2}N^2} \right] = \int e^{-\frac{x^2}{2(1-Ca^2)}} (2\pi)^{-1} dx = \frac{1}{\sqrt{1-a^2C}}$$

ce qui permet de conclure que (**) est vraie.

Sans conditionnement, on écrit tout avec des intégrales usuelles car X a une densité et on fait du Fubini.

Examen Ecole Polytechnique 2016

Problème : Étude d'un estimateur associé à une loi uniforme.

Soit $r > 0$ fixé et $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de loi uniforme sur $[-r\sqrt{3}, r\sqrt{3}]$.

Pour tout $n \geq 1$, on pose $Y_n = X_n^2$.

1. Rappeler l'expression de la densité de X_1 . Calculer l'espérance et la variance de X_1 puis $\mathbb{E}[X_1^4]$. En déduire la variance de Y_1 .
2. Pour tout $n \geq 1$, on pose

$$M_n = \frac{Y_1 + \dots + Y_n}{n} \quad \text{et} \quad T_n = \sqrt{M_n}.$$

Montrer que T_n converge presque-sûrement vers une limite que l'on précisera.

3. On va à présent étudier la vitesse de convergence de T_n vers sa limite. Pour cela, on considère, pour tout $n \geq 1$,

$$U_n = \sqrt{n}(T_n - r) \quad \text{et} \quad V_n = \sqrt{n} \frac{T_n^2 - r^2}{2r} = \sqrt{n} \frac{M_n - r^2}{2r}.$$

Montrer que la suite $(V_n)_{n \geq 1}$ converge en loi vers une loi gaussienne dont on précisera la moyenne et la variance.

4. Montrer que U_n converge en loi vers une limite que l'on précisera.

Indication : on pourra appliquer le théorème de la “méthode delta” vu en cours.

Application. On suppose que les variables aléatoires $(Y_k)_{k \geq 1}$ ci-dessus représentent la quantité d'eau consommée en une journée par chacun des animaux d'un élevage. Plus précisément, Y_k est la quantité d'eau consommée par le k -ème animal un jour donné. On notera n le nombre total d'animaux, supposé très grand, et S_n la quantité d'eau totale consommée en une journée par les n animaux :

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

5. Le but de cette question est d'estimer le paramètre r supposé inconnu. Pour cela, on mesure la quantité totale d'eau consommée par les animaux un jour donné, que l'on note s_n .

En utilisant la variable aléatoire $U_n = \sqrt{S_n} - r\sqrt{n}$, estimer la valeur de r à l'aide d'un intervalle de confiance, de niveau de confiance 0,95, en fonction de n et s_n .

Indication : on supposera que n est suffisamment grand pour que la loi de U_n soit très proche de la loi limite trouvée dans la question précédente et on rappelle que si Z est une variable aléatoire de loi $\mathcal{N}(0, 1)$ alors $\mathbb{P}(|Z| \leq 1,96) = 0,95$.

Problème : Étude microscopique de systèmes de particules, valeurs extrêmes et lois de Poisson.

Certains événements rares, comme les crues d'un fleuve ou la saturation de certains systèmes, se modélisent via ce qu'on appelle la *statistique des valeurs extrêmes*, dont l'objet est l'étude des records dans les séries statistiques. Le problème suivant, qui fait le lien avec un objet appelé *processus ponctuel de Poisson*, porte sur le comptage des valeurs extrêmes. On y met en évidence un principe que l'on applique aussi à l'étude d'un système de particules à l'échelle microscopique.

I) Théorème de la limite de Poisson multivarié.

On rappelle que la *loi de Poisson* de paramètre $\lambda \geq 0$ est la loi d'une variable aléatoire X à valeurs dans \mathbb{N} telle que pour tout $m \in \mathbb{N}$,

$$\mathbb{P}(X = m) = e^{-\lambda} \frac{\lambda^m}{m!}.$$

1. Prouver que dans ce cas, pour tout $t \in \mathbb{R}$, on a $\mathbb{E}[e^{itX}] = \exp\{\lambda(e^{it} - 1)\}$.
Fixons $k \geq 1$ et considérons, pour tout $n \geq 1$, une famille $(T_1(n), \dots, T_n(n))$ de n variables aléatoires indépendantes identiquement distribuées à valeurs dans

$\{0, \dots, k\}$. On pose, pour tout $n \geq 1$ et $\ell \in \{0, \dots, k\}$,

$$p_\ell(n) = \mathbb{P}(T_1(n) = \ell)$$

et on suppose que pour tout $\ell \in \{1, \dots, k\}$, il existe $\lambda_\ell \geq 0$ tel que

$$np_\ell(n) \xrightarrow[n \rightarrow \infty]{} \lambda_\ell.$$

2. Donner la limite de $p_0(n)$ lorsque $n \rightarrow \infty$.

Ainsi, les variables aléatoires $T_1(n), \dots, T_n(n)$ valent très souvent zéro lorsque n devient grand. On va s'intéresser aux autres valeurs qu'elles prennent (rarement, donc), celles de $\{1, \dots, k\}$. Pour $n \geq 1$ et $\ell \in \{1, \dots, k\}$, on pose

$$N_\ell(n) = \text{Card} \{j \in \{1, \dots, n\}, T_j(n) = \ell\}$$

et on va montrer le théorème suivant :

Théorème [Limite de Poisson multivariée] Le vecteur aléatoire

$$(N_1(n), \dots, N_k(n))$$

converge en loi, lorsque $n \rightarrow \infty$, vers un vecteur aléatoire (Y_1, \dots, Y_k) de variables aléatoires indépendantes telles que pour tout $\ell \in \{1, \dots, k\}$, Y_ℓ est distribué selon la loi de Poisson de paramètre λ_ℓ .

3. Soient $t_1, \dots, t_k \in \mathbb{R}$. Montrer que, en posant $t_0 := 0$, on a

$$\mathbb{E} \left[e^{i(t_1 N_1(n) + \dots + t_k N_k(n))} \right] = \left(\sum_{\ell=0}^k p_\ell(n) e^{it_\ell} \right)^n.$$

Indication : On pourra commencer par exprimer chaque variable aléatoire $N_\ell(n)$ à partir des variables aléatoires $\mathbf{1}_{T_j(n)=\ell}$, où $j \in \{1, \dots, n\}$.)

4. Conclure la preuve du théorème “Limite de Poisson multivariée”, en utilisant le fait que

$$p_0(n) = 1 - (p_1(n) + \dots + p_k(n)).$$

Indication : On admettra le fait suivant : si z_n est une suite de nombres complexes qui tend vers $z \in \mathbb{C}$, alors $\left(1 + \frac{z_n}{n}\right)^n \xrightarrow[n \rightarrow \infty]{} e^z$.

II) Limite de Poisson pour les processus ponctuels.

Soient $(X_j)_{j \geq 1}$ des variables aléatoires indépendantes identiquement distribuées à valeurs dans \mathbb{R} . Fixons des suites (non aléatoires) $a_n > 0$, $b_n \in \mathbb{R}$ et posons, pour $n \geq 1$ et $x < y$,

$$N_{x,y}(n) := \text{Card} \left\{ j \in \{1, \dots, n\}, x < \frac{X_j - b_n}{a_n} \leq y \right\}.$$

1. Supposons qu'il existe $f : \mathbb{R} \rightarrow [0, +\infty[$ continue telle que pour tous $x < y$,

$$n\mathbb{P}\left(x < \frac{X_1 - b_n}{a_n} \leq y\right) \xrightarrow{n \rightarrow \infty} \int_x^y f(u) du.$$

En appliquant le théorème "Limite de Poisson multivariée", prouver que pour tous $k \geq 1$ et $x_0 < \dots < x_k$, le vecteur aléatoire

$$(N_{x_0, x_1}(n), \dots, N_{x_{k-1}, x_k}(n)) \quad (10.0.1)$$

converge en loi, lorsque $n \rightarrow \infty$, vers un vecteur (Y_1, \dots, Y_k) de variables aléatoires indépendantes telles que pour tout $\ell \in \{1, \dots, k\}$, Y_ℓ est distribué selon la loi de Poisson de paramètre $\lambda_\ell := \int_{x_{\ell-1}}^{x_\ell} f(u) du$.

2. Application 1 : Étude microscopique d'un système de particules. Soient $(X_j)_{j \geq 1}$ variables aléatoires indépendantes identiquement distribuées à valeurs dans \mathbb{R} ayant une densité g . Pour $n \geq 1$, X_1, \dots, X_n modélisent les positions de n particules. On se fixe $c \in \mathbb{R}$ et on souhaite ici étudier la répartition, pour n grand, de celles de ces particules qui se trouvent au voisinage de c .

- (a) On suppose que g est continue en c . Donner une suite a_n telle que pour tous $x < y$,

$$n\mathbb{P}\left(x < \frac{X_1 - c}{a_n} \leq y\right) \xrightarrow{n \rightarrow \infty} (y - x)g(c).$$

Au regard de la question 1, quelle est alors la limite, pour la convergence en loi, de

$$\text{Card} \left\{ j \in \{1, \dots, n\}, x < \frac{X_j - c}{a_n} \leq y \right\} ?$$

- (b) On suppose que, pour un certain $m \geq 2$, g est de classe \mathcal{C}^{m-1} au voisinage de c , que

$$g(c) = g'(c) = g''(c) = \dots = g^{(m-2)}(c) = 0$$

et que $g^{(m-1)}(c) \neq 0$. Prouver que m est impair et que $g^{(m-1)}(c) > 0$. Donner une suite a_n et une fonction continue $f : \mathbb{R} \rightarrow \mathbb{R}^+$ non identiquement nulle telle que pour tous $x < y$,

$$n\mathbb{P}\left(x < \frac{X_1 - c}{a_n} \leq y\right) \xrightarrow{n \rightarrow \infty} \int_x^y f(u) du.$$

3. Application 2 : Valeurs extrêmes. Soient $(X_j)_{j \geq 1}$ variables aléatoires indépendantes identiquement distribuées à valeurs dans \mathbb{R} ayant une densité g . Les X_j modélisent maintenant les temps d'attente, en secondes, entre les requêtes successives reçues par un serveur web. On cherche à estimer les durées maximales sans requête.

- (a) Prouver que presque sûrement, les valeurs de la suite $(X_j)_{j \geq 1}$ sont deux à deux distinctes.

Indication : On pourra écrire l'événement contraire comme une union, infinie mais dénombrable, d'événements dont on montrera qu'ils sont de probabilité nulle.

On suppose dorénavant les X_j de loi exponentielle de paramètre 1, c'est-à-dire que la densité g des X_j est $g(x) := \mathbf{1}_{x \geq 0} e^{-x}$.

- (b) Montrer que pour tous $x < y$,

$$n\mathbb{P}(x < X_1 - \ln(n) \leq y) \xrightarrow{n \rightarrow \infty} \int_x^y f(u) du,$$

où $f : \mathbb{R} \rightarrow \mathbb{R}^+$ est une fonction que l'on explicitera.

Il résulte de la question précédente que les hypothèses de la question 1 sont satisfaites pour cette fonction f si $a_n = 1$ et $b_n = \ln(n)$. **On admet** que la conclusion de la question 1 reste valable si $x_k = +\infty$, c'est à dire lorsque la dernière coordonnée du vecteur de (10.0.1) est la variable aléatoire

$$N_{x_{k-1}, +\infty}(n) := \text{Card} \left\{ j \in \{1, \dots, n\}, x_{k-1} < \frac{X_j - b_n}{a_n} \right\}.$$

On définit alors, pour tout $1 \leq j \leq n$, $Z_{n,j}$ comme la j -ème plus grande valeur parmi $\{X_1, \dots, X_n\}$.

- (c) Soit $x \in \mathbb{R}$ et $r \geq 1$ entier. Exprimer l'événement $\{Z_{n,r} - \ln(n) \leq x\}$ à partir de la variable aléatoire $N_{x, +\infty}(n)$.
- (d) Montrer que, pour tout entier $r \geq 1$, lorsque $n \rightarrow \infty$, $Z_{n,r} - \ln(n)$ converge en loi. On donnera la fonction de répartition de la loi limite et on montrera qu'elle a une densité que l'on explicitera.
- (e) Donner, pour $x < y$, la limite, lorsque $n \rightarrow \infty$, de

$$\mathbb{P}(\{Z_{n,2} - \ln(n) \leq x\} \cap \{Z_{n,1} - \ln(n) \leq y\}).$$

- (f) On cherche à déduire des questions précédentes l'ordre de grandeur du temps d'attente maximal entre deux requêtes sur une année.
- i. Donner la définition précise et une valeur numérique approchée du nombre t_0 tel que

$$\mathbb{P}(Z_{n,1} - \ln(n) \leq t_0) \xrightarrow{n \rightarrow \infty} 0.99.$$

Indication : $\ln(10) \approx 2.3$.

- ii. À l'aide de la loi des grands nombres, donner la valeur approximative de la durée, en secondes, couverte par n requêtes pour n grand, puis du nombre de requêtes sur une année.

- iii. En déduire le nombre t_1 tel que le temps d'attente maximal entre deux requêtes sur une année est $\leq t_1$ avec probabilité ≈ 0.99 .
Indication : $\ln(3600 \times 24 \times 365) \approx 17.3$.

Corrigé de l'examen Ecole Polytechnique 2016

Problème : Étude d'un estimateur associé à une loi uniforme.

1. La densité de X_1 est $x \rightarrow \frac{1}{2\sqrt{3}r} \mathbf{1}_{[-\sqrt{3}r, \sqrt{3}r]}(x)$.
 X_1 étant bornée, elle admet des moments de tous ordres. Par symétrie, $\mathbb{E}[X_1] = 0$ et pour tout n pair, on a

$$\mathbb{E}[X_1^n] = \frac{1}{2\sqrt{3}r} \int_{-\sqrt{3}r}^{+\sqrt{3}r} x^n dx = \frac{(r\sqrt{3})^n}{n+1}.$$

Ce qui donne, $\mathbb{E}[X_1^2] = \text{Var}[X_1] = r^2$, $\mathbb{E}[X_1^4] = \frac{9r^4}{5}$ puis $\text{Var}[Y_1] = \mathbb{E}[X_1^4] - \mathbb{E}[X_1^2]^2 = \frac{4r^4}{5}$.

2. La loi forte des grands nombres et le fait que la fonction $x \rightarrow \sqrt{x}$ est continue implique immédiatement que T_n converge p.s. vers $\sqrt{\mathbb{E}[Y_1]} = r$.
 3. On a

$$V_n = \frac{\sum_{i=1}^n Y_i - nr^2}{2r\sqrt{n}}.$$

Par application directe du TCL, on voit que $2rV_n/\sqrt{\text{Var}[Y_1]}$ converge en loi vers la loi $\mathcal{N}(0, 1)$. Comme $\text{Var}[X_1^2] = \frac{4r^4}{5}$, on en déduit que V_n converge en loi vers la loi $\mathcal{N}(0, \frac{r^2}{5})$.

4. Le théorème avec la fonction $g(x) = \sqrt{x}$ donne que $\sqrt{n}(g(M_n) - g(\mathbb{E}[Y_1])) = \sqrt{n}(T_n - r)$ converge en loi vers $\mathcal{N}(0, (g'(r^2))^2 \text{Var}[Y_1]) = \mathcal{N}(0, \frac{r^2}{5})$.
 5. Comme n est grand, $\frac{\sqrt{5}}{r}U_n$ est très proche en loi d'une $\mathcal{N}(0, 1)$ et donc

$$\mathbb{P}\left(\frac{\sqrt{5}}{r}U_n \in [-1,96; 1,96]\right) = \mathbb{P}\left(\frac{\sqrt{5}S_n}{r} - \sqrt{5n} \in [-1,96; 1,96]\right) \approx 0,95.$$

On en déduit donc qu'avec une probabilité de ne pas se tromper égale à 95%, on peut affirmer que $(\frac{\sqrt{5}S_n}{r} - \sqrt{5n}) \in [-1,96; 1,96]$ ce qui donne l'intervalle de confiance

$$\left[\frac{\sqrt{5}S_n}{\sqrt{5n} + 1,96}, \frac{\sqrt{5}S_n}{\sqrt{5n} - 1,96}\right].$$

Problème : Étude microscopique de systèmes de particules, valeurs extrêmes et lois de Poisson.

I) Théorème de la limite de Poisson multivarié.

1. On a

$$\mathbb{E} [e^{itX}] = e^{-\lambda} \sum_{m \geq 0} \frac{\lambda^m}{m!} e^{itm} = e^{-\lambda} \sum_{m \geq 0} \frac{(\lambda e^{it})^m}{m!} = e^{-\lambda} e^{\lambda e^{it}} = \exp \{ \lambda (e^{it} - 1) \}.$$

2. Pour tout $\ell \in \{1, \dots, k\}$, $p_\ell(n) = \frac{np_\ell(n)}{n} \xrightarrow[n \rightarrow \infty]{} 0$, donc

$$p_0(n) = p_0(n) = 1 - (p_1(n) + \dots + p_k(n)) \xrightarrow[n \rightarrow \infty]{} 1.$$

3. On a

$$N_\ell(n) = \sum_{j=1}^n \mathbf{1}_{T_j(n)=\ell}.$$

On a donc

$$\begin{aligned} \mathbb{E} \left[e^{i(t_1 N_1(n) + \dots + t_k N_k(n))} \right] &= \mathbb{E} \exp i \sum_{\ell=1}^k t_\ell \sum_{j=1}^n \mathbf{1}_{T_j(n)=\ell} \\ &= \mathbb{E} \exp i \sum_{j=1}^n \sum_{\ell=1}^k t_\ell \mathbf{1}_{T_j(n)=\ell} \\ &= \prod_{j=1}^n \mathbb{E} \exp i \sum_{\ell=1}^k t_\ell \mathbf{1}_{T_j(n)=\ell} \end{aligned}$$

Or pour $t_0 := 0$, pour tout j ,

$$\mathbb{E} \exp i \sum_{\ell=1}^k t_\ell \mathbf{1}_{T_j(n)=\ell} = \mathbb{E} \exp i \sum_{\ell=0}^k t_\ell \mathbf{1}_{T_j(n)=\ell} = \sum_{\ell=0}^k p_\ell(n) e^{it_\ell}.$$

Ainsi,

$$\mathbb{E} \left[e^{i(t_1 N_1(n) + \dots + t_k N_k(n))} \right] = \left(\sum_{\ell=0}^k p_\ell(n) e^{it_\ell} \right)^n$$

4. On a

$$\begin{aligned}
 \mathbb{E} \left[e^{i(t_1 N_1(n) + \dots + t_k N_k(n))} \right] &= \left(1 - \sum_{\ell=1}^k p_\ell + \sum_{\ell=1}^k p_\ell(n) e^{it_\ell} \right)^n \\
 &= \left(1 + \frac{1}{n} \sum_{\ell=1}^k n p_\ell(n) (e^{it_\ell} - 1) \right)^n \\
 &\xrightarrow{n \rightarrow \infty} \exp \sum_{\ell=1}^k \lambda_\ell (e^{it_\ell} - 1) \\
 &= \prod_{\ell=1}^k \mathbb{E}[e^{it_\ell Y_\ell}] \\
 &= \mathbb{E}[e^{i(t_1 Y_1 + \dots + t_k Y_k)}]
 \end{aligned}$$

On conclut alors avec le théorème de Lévy.

II) Limite de Poisson pour les processus ponctuels.

1. Pour tout $n \geq 1$ et $j \in \{1, \dots, n\}$, posons $T_j(n) = \ell$ si il existe $\ell \in \{1, \dots, k\}$ tel que $x_{\ell-1} < \frac{X_j - b_n}{a_n} \leq x_\ell$ et $T_j(n) = 0$ sinon. On est alors dans un cas d'application direct du théorème "Limite de Poisson multivariée".
2. (a) Soit F la fonction de répartition des X_j . On va utiliser le DL d'ordre 1 de F en c : $F(c+u) = F(c) + u g(c) + o(u)$. On a, pour $a_n = n^{-1}$,

$$\begin{aligned}
 n \mathbb{P} \left(x < \frac{X_1 - c}{a_n} \leq y \right) &= n \mathbb{P}(a_n x + c < X_1 \leq a_n y + c) \\
 &= n(F(a_n y + c) - F(a_n x + c)) \\
 &= n(g(c) a_n (y - x) + o(a_n)) \\
 &\xrightarrow{n \rightarrow \infty} (y - x) g(c)
 \end{aligned}$$

Donc les hypothèses de la question 1 sont satisfaites pour $a_n = n^{-1}$, $b_n = c$ et f définie par $f(x) = g(c)$ pour tout $x \in \mathbb{R}$. Donc

$$\text{Card} \left\{ j \in \{1, \dots, n\}, x < \frac{X_j - b_n}{a_n} \leq y \right\}$$

converge en loi vers une loi de Poisson de paramètre $\lambda = g(c)(y - x)$.

(b) On va utiliser le DL d'ordre $m - 1$ de g en c :

$$g(c+u) = \sum_{i=0}^{m-1} \frac{g^{(i)}(c)}{i!} u^i + o(u^{m-1}) = u^{m-1} \frac{g^{(m-1)}(c)}{(m-1)!} (1 + o(1)).$$

La seule possibilité pour que pour tout $\varepsilon > 0$, on n'ait ni

$$\int_{c-\varepsilon}^c g(u)du < 0,$$

ni

$$\int_c^{c+\varepsilon} g(u)du < 0$$

(ce qui est nécessaire car g est une densité de probabilité) est alors que m soit impair et $g^{(m-1)}(c) > 0$. On va alors utiliser le DL d'ordre m de F en c :

$$\begin{aligned} F(c+u) &= \sum_{i=0}^m \frac{F^{(i)}(c)}{i!} u^i + o(u^m) \\ &= F(c) + \sum_{i=1}^m \frac{g^{(i-1)}(c)}{i!} u^i + o(u^m) \\ &= F(c) + \frac{g^{(m-1)}(c)}{m!} u^m + o(u^m). \end{aligned}$$

On a, pour $a_n = n^{-1/m}$ et f définie par $f(u) = \frac{g^{(m-1)}(c)}{(m-1)!} u^{m-1}$ pour tout $u \in \mathbb{R}$,

$$\begin{aligned} n\mathbb{P}(x < \frac{X_1 - c}{a_n} \leq y) &= n\mathbb{P}(a_n x + c < X_1 \leq a_n y + c) \\ &= n(F(a_n y + c) - F(a_n x + c)) \\ &= n(a_n^m \frac{g^{(m-1)}(c)}{m!} (y^m - x^m) + o(a_n^m)) \\ &\xrightarrow[n \rightarrow \infty]{} \int_x^y f(u)du \end{aligned}$$

Donc les hypothèses de la question 1 sont satisfaites pour $a_n = n^{-1/m}$, $b_n = c$ et f définie comme ci-dessus.

3. (a) On introduit l'événement $E := \{\exists i, j \geq 1, i \neq j, X_i = X_j\}$. Cet événement est l'union de la famille dénombrable d'événements $(\{X_i = X_j\})_{i,j \geq 1, i \neq j}$, donc il suffit de montrer que pour tout $i < j$,

$$\mathbb{P}(X_i = X_j) = 0.$$

Or pour tous $i \neq j$, par indépendance de X_i et X_j , la loi de (X_i, X_j) a pour densité $g(x)g(y)$ par rapport à la mesure de Lebesgue dans \mathbb{R}^2 . Donc, en utilisant Fubini pour les fonctions positives puis le fait que tout singleton

est de mesure nulle pour la mesure de Lebesgue sur \mathbb{R} , on a

$$\begin{aligned}\mathbb{P}(X_i = X_j) &= \int_{(x,y) \in \mathbb{R}^2} \mathbf{1}_{y=x} g(x)g(y) dx dy \\ &= \int_{x \in \mathbb{R}} \left(\int_{y \in \mathbb{R}} \mathbf{1}_{y=x} g(y) dy \right) g(x) dx \\ &= \int_{x \in \mathbb{R}} 0 \times g(x) dx \\ &= 0,\end{aligned}$$

ce qui permet de conclure.

- (b) Posons $a_n = 1$ et $b_n = \ln(n)$. On a, pour $x < y$ fixés et n assez grand pour que $a_n x + b_n \geq 0$,

$$\begin{aligned}n\mathbb{P}\left(x < \frac{X_1 - b_n}{a_n} \leq y\right) &= n\mathbb{P}(a_n x + b_n < X_1 \leq a_n y + b_n) \\ &= n(e^{-(a_n x + b_n)} - e^{-(a_n y + b_n)}) \\ &= n(e^{-x - \ln(n)} - e^{-y - \ln(n)}) \\ &= e^{-x} - e^{-y}\end{aligned}$$

donc les hypothèses de la question 1 sont satisfaites pour $f(x) = e^{-x}$.

- (c) Pour tout $x \in \mathbb{R}$, on a l'équivalence

$$Z_{n,r} - \ln(n) \leq x \iff N_{x,+\infty}(n) < r.$$

- (d) Comme $Z_{n,r} - \ln(n) \leq x \iff N_{x,+\infty}(n) < r$, on a

$$\begin{aligned}\mathbb{P}(Z_{n,r} - \ln(n) \leq x) &= \mathbb{P}(N_{x,+\infty}(n) < r) \\ &= \mathbb{P}(N_{x,+\infty}(n) \leq r - 1/2) \\ &\xrightarrow[n \rightarrow \infty]{} \mathbb{P}(Y \leq r - 1/2)\end{aligned}$$

pour Y variable aléatoire de loi $\mathcal{P}(\mu_x)$ pour $\mu_x := \int_x^{+\infty} f(u) du = e^{-x}$ (on a utilisé ici la question 1, ainsi que le fait que la fonction de répartition de toute loi de Poisson est continue en $r - 1/2$). Ainsi, on a, pour tout $x \in \mathbb{R}$,

$$\mathbb{P}(Z_{n,r} - \ln(n) \leq x) \xrightarrow[n \rightarrow \infty]{} F_r(x) := e^{-e^{-x}} \sum_{s=0}^{r-1} \frac{e^{-sx}}{s!}$$

La fonction

$$F_r(x) = \sum_{s=0}^{r-1} \frac{e^{-sx - e^{-x}}}{s!}$$

est strictement croissante (comme limite de la suite de fonctions croissantes $\mathbb{P}(Z_{n,r} - \ln(n) \leq x)$), \mathcal{C}^∞ , de limites en $-\infty$ et $+\infty$ égales à 0 et 1, c'est donc la fonction de répartition d'une loi de densité

$$\begin{aligned} F_r'(x) &= \sum_{s=0}^{r-1} (e^{-x} - s) \frac{e^{-sx - e^{-x}}}{s!} \\ &= e^{-e^{-x}} \sum_{s=0}^{r-1} (e^{-x} - s) \frac{e^{-sx}}{s!} \\ &= e^{-e^{-x}} \left(\sum_{s=1}^r \frac{e^{-sx}}{(s-1)!} - \sum_{s=1}^{r-1} \frac{e^{-sx}}{(s-1)!} \right) \\ &= \frac{e^{-rx - e^{-x}}}{(r-1)!}. \end{aligned}$$

On en déduit que $Z_{n,r} - \ln(n)$ converge en loi vers cette loi.

(e) Comme $Z_{n,r} - \ln(n) \leq x \iff N_{x,+\infty}(n) < r$, on a

$$\begin{aligned} &\mathbb{P}(\{Z_{n,2} - \ln(n) \leq x\} \cap \{Z_{n,1} - \ln(n) \leq y\}) \\ &= \mathbb{P}(\{N_{x,+\infty}(n) < 2\} \cap \{N_{y,+\infty}(n) < 1\}) \\ &= \mathbb{P}(\{N_{x,y}(n) \in \{0, 1\}\} \cap \{N_{y,+\infty}(n) = 0\}) \end{aligned}$$

Soient g, h des fonctions continues sur \mathbb{R} de supports compacts respectivement contenus dans $[-1/2, 3/2]$ et $[-1/2, 1/2]$ telles que $g(0) = g(1) = h(0) = 1$ (de telles fonctions peuvent être choisies affines par morceaux, par exemple $h(t) = (1 - 2|t|)_+$ et $g(t) = h(t) + h(t - 1)$). Alors

$$\mathbb{P}(\{N_{x,y}(n) \in \{0, 1\}\} \cap \{N_{y,+\infty}(n) = 0\}) = \mathbb{E}[g(N_{x,y}(n))h(N_{y,+\infty}(n))].$$

Donc, par la question 1 ainsi que le fait que la fonction $(s, t) \in \mathbb{R}^2 \mapsto g(s)h(t) \in \mathbb{R}$ est continue et bornée,

$$\begin{aligned} \mathbb{P}(\{Z_{n,2} - \ln(n) \leq x\} \cap \{Z_{n,1} - \ln(n) \leq y\}) &= \mathbb{E}[g(N_{x,y}(n))h(N_{y,+\infty}(n))] \\ &\xrightarrow{n \rightarrow \infty} \mathbb{E}[g(Y_1)h(Y_2)] \end{aligned}$$

pour Y_1, Y_2 des variables aléatoires indépendantes de lois $\mathcal{P}(\mu_{x,y})$, $\mathcal{P}(\mu_y)$ avec $\mu_{x,y} := \int_x^y f(u)du = e^{-x} - e^{-y}$ et $\mu_y := \int_y^{+\infty} f(u)du = e^{-y}$. Or, par indépendance,

$$\begin{aligned} \mathbb{E}[g(Y_1)h(Y_2)] &= \mathbb{E}[g(Y_1)]\mathbb{E}[h(Y_2)] \\ &= \mathbb{P}(Y_1 \in \{0, 1\})\mathbb{P}(Y_2 = 0) \\ &= e^{-\mu_{x,y}}(1 + \mu_{x,y})e^{-\mu_y} \\ &= e^{-\mu_{x,y} - \mu_y}(1 + \mu_{x,y}) \\ &= e^{-e^{-x}}(1 + e^{-x} - e^{-y}) \end{aligned}$$

Ainsi,

$$\mathbb{P}(\{Z_{n,2} - \ln(n) \leq x\} \cap \{Z_{n,1} - \ln(n) \leq y\}) \xrightarrow[n \rightarrow \infty]{} e^{-e^{-x}}(1 + e^{-x} - e^{-y}).$$

- (f) i. Par la question 3d, $Z_{n,1} - \ln(n)$ converge en loi vers la loi de fonction de répartition $e^{-e^{-x}}$. Cette fonction étant continue, pour tout $x \in \mathbb{R}$,

$$\mathbb{P}(Z_{n,1} - \ln(n) \leq x) \xrightarrow[n \rightarrow \infty]{} e^{-e^{-x}}.$$

Le nombre t_0 recherché est donc la solution de l'équation

$$e^{-e^{-t_0}} = 0,99,$$

c'est à dire

$$t_0 = -\ln(-\ln(0,99)).$$

Modulo l'approximation $\ln(1 + \varepsilon) \approx \varepsilon$, cette équation s'écrit

$$t_0 = -\ln(-\ln(0,99)) \approx -\ln(10^{-2}) = 2 \ln(10) \approx 4,6.$$

- ii. Soit $S_n = X_1 + \dots + X_n$. On a $\mathbb{E}(X_1) = \int_0^{+\infty} x e^{-x} dx = 1$, donc, par la LGN, p.s., $S_n = n(1+o(1))$ quand $n \rightarrow \infty$. Donc la valeur approximative de la durée couverte par n requêtes pour n grand est n secondes. Le nombre n de requêtes sur une année est donc approximativement $3600 \times 24 \times 365$.
- iii. Par ce qui précède, la durée maximale $Z_{n,1}$ s'écrit $\ln(n) + (Z_{n,1} - \ln(n))$. On a donc, pour

$$t_1 := \ln(n) + t_0 \approx 21,9,$$

et donc la probabilité

$$\mathbb{P}(\text{ temps d'attente maximal entre deux requêtes sur une année } \leq t_1)$$

vaut approximativement 0,99.

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

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, 56
- échantillon, 177
- événement, 16
- événements indépendants, 41

- approximation linéaire, 96

- biais, 179

- chaîne de Markov, 246
- coefficient de corrélation, 95
- convergence d'estimateur, 179
- convergence de test, 210
- convergence en loi, 164
- convergence en moyenne, 138
- convergence en probabilités, 138
- convergence presque-sûre, 138
- convolution, 127
- corrélation, 95
- covariance, 94

- densité, 85
- densité conditionnelle, 115
- densité d'un vecteur aléatoire, 112
- densité de probabilité, 85
- densité marginale, 114
- diffusion d'Ehrenfest, 249
- droite des moindres carrés, 96

- ensemble dénombrable, 28
- ensemble négligeable, 34
- espérance conditionnelle de Y sachant $X = x_i$, 68
- espérance conditionnelle pour des v.a. à densité, 116
- espérance conditionnelle pour des v.a. discrètes, 68
- espérance des variables aléatoires discrètes, 52
- espace de probabilité, 33
- espace de probabilité produit, 43
- espace fondamental, 15
- estimateur du maximum de vraisemblance, 188
- expérience aléatoire, 15
- expériences aléatoires indépendantes, 43

- file d'attente, 241
- fonction caractéristique, 154
- fonction de répartition, 82
- fonction génératrice, 58
- fonction mesurable, 81
- formule de Bayes, 39
- formule de Huygens, 56
- formule des probabilités composées, 39
- formule des probabilités totales, 39
- fréquence de réalisation, 19

- histogramme, 128

- inégalité de Bienaymé-Chebyshev, 107
- inégalité de Cauchy-Schwarz, 109
- inégalité de Jensen, 109
- indépendance, 70
- intervalle de confiance, 193

- lemme de Borel-Cantelli, 44
- limsup d'ensemble, 44
- log-vraisemblance, 189
- loi des grands nombres, 144
- loi à densité, 86

- loi bêta, 126
- loi binomiale, 61
- loi conditionnelle pour des v.a. discrètes, 67
- loi d'une v.a. discrète, 51
- loi d'une variable aléatoire, 37
- loi de Bernoulli, 60
- loi de Cauchy, 107
- loi de chi-deux, 163
- loi de Poisson, 64
- loi de reproduction, 234
- loi forte des grands nombres, 146
- loi géométrique de paramètre p , 63
- loi gamma, 102
- loi hypergéométrique, 24
- lois marginales, 66

- méthode de Monte-Carlo, 11, 147
- marche aléatoire, 228
- marche aléatoire absorbée, 229
- matrice de covariance, 113
- mesure de Dirac, 35
- mesure de Lebesgue, 85
- mesure de Lebesgue sur \mathbb{R}^n , 111
- modèle probabiliste, 20
- modèle statistique, 176
- modèle d'inventaire, 248
- moment, 58

- niveau d'un test, 209
- normalité asymptotique, 180

- partition, 18
- percolation, 240
- phénomènes aléatoires, 9
- probabilité, 31
- probabilité conditionnelle, 38
- probabilité produit, 43
- probabilité sur Ω fini, 21
- probabilité uniforme, 23
- probabilité invariante, 249
- probabilités de transition, 247
- processus réversible, 254
- promenade aléatoire, 228

- propriété de non-vieillessement, 101
- propriété vraie presque-sûrement, 34
- propriété de Markov, 246

- quantile, 193

- risque quadratique moyen, 180
- réversibilité, 254

- sondages, 204
- stabilité, 243, 249
- suite d'événements indépendants, 42
- suite de vecteurs aléatoires indépendantes, 120

- table de la loi de Gauss, 105
- test du chi-deux, 213
- théorème central limite multi-dimensionnel, 172
- théorème de convergence dominée, 140
- théorème de Fubini, 110
- théorème de Lévy, 167
- théorème de la limite centrale, 170
- théorème de Slutsky, 168
- transformée de Laplace, 159
- tribu, 29
- tribu borélienne, 30
- tribu engendrée par une partie de Ω , 30
- tribu produit, 43

- variable aléatoire, 35
- variable aléatoire de Bernoulli, 60
- variable aléatoire étagée, 91
- variable aléatoire centrée réduite, 95
- variable aléatoire de loi normale, 103
- variable aléatoire de chi-deux, 163
- variable aléatoire exponentielle, 99
- variable aléatoire géométrique, 63
- variable aléatoire normale centrée réduite, 103
- variable aléatoire réelle, 79
- variable aléatoire réelle de carré intégrable, 94
- variable aléatoire réelle intégrable, 91
- variable aléatoire uniforme sur $[a, b]$, 98

variable binomiale, [61](#)
variable de Poisson, [64](#)
variable discrète de carré intégrable, [55](#)
variable discrète intégrable, [54](#)
variable gaussienne, [103](#)
variables aléatoires discrètes, [51](#)
variance, [56](#)
variance d'une variable aléatoire réelle, [94](#)
variance empirique non-biaisée, [185](#)
vecteur aléatoire, [111](#)
vecteur espérance, [113](#)
vecteur gaussien, [161](#)
vecteurs aléatoires indépendants, [118](#)
vraisemblance, [188](#)