

Table des matières

1	Introduction	5
1.1	Introduction du cours	6
1.2	Avant-Propos	6
1.3	Phénomènes aléatoires	6
1.4	Deux idées majeures et incontournables	7
1.4.1	La loi des grands nombres	7
1.4.2	Conditionnement et Indépendance	8
1.5	Les variables aléatoires	8
1.5.1	Loi d'une variable aléatoire	8
1.5.2	Simulation de variables aléatoires	9
1.6	Historique	9
1.7	Bibliographie	11
2	Espace de probabilité	15
2.1	Le langage des probabilités	15
2.1.1	Expériences et événements aléatoires	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	22
2.2.3	Modèles d'urnes	23
2.3	Définition générale des Probabilités	27
2.3.1	Pourquoi la définition précédente ne suffit-elle pas ?	27
2.3.2	Les ensembles dénombrables	27
2.3.3	Tribu	28
2.3.4	Définition d'une probabilité	29
2.3.5	Probabilités sur un espace dénombrable	32

2.3.6	Variable aléatoire et sa loi	32
2.4	Conditionnement et indépendance	36
2.4.1	Probabilités conditionnelles	36
2.4.2	Indépendance	40
2.4.3	Le Lemme de Borel-Cantelli	42
3	Espace fini ou dénombrable	45
3.1	Prérequis : quelques résultats utiles sur les séries	45
3.2	Variables aléatoires discrètes	46
3.3	Espérance des variables aléatoires discrètes	47
3.3.1	Définition	47
3.3.2	Propriétés de l'espérance des variables aléatoires discrètes	49
3.3.3	Variance et écart-type	50
3.3.4	Un résultat fondamental - Moments d'une variable aléatoire	52
3.4	Fonction génératrice d'une variable aléatoire à valeurs entières	53
3.5	Variables aléatoires discrètes usuelles	54
3.5.1	Variable aléatoire de Bernoulli	54
3.5.2	Variable aléatoire binomiale	55
3.5.3	Probabilité de succès et variable aléatoire géométrique	57
3.5.4	Variable aléatoire de Poisson	58
3.6	Lois conditionnelles et indépendance	60
3.6.1	Lois conditionnelles	60
3.6.2	Propriétés de l'espérance conditionnelle	62
3.6.3	Variables aléatoires indépendantes	63
3.6.4	Somme de variables aléatoires indépendantes	66
4	Variables aléatoires réelles et Vecteurs aléatoires	69
4.1	Les variables aléatoires réelles	69
4.2	Les lois de variables aléatoires réelles	71
4.2.1	Fonction de répartition	71
4.2.2	Variables aléatoires de loi à densité	74
4.2.3	Variable aléatoire uniforme sur $[0, 1]$ et générateurs de nombres aléatoires	77
4.2.4	Simulation d'une variable aléatoire par inversion de la fonction de répartition	78
4.3	Espérance des variables aléatoires réelles	78
4.3.1	Définition	78

4.4	Variables aléatoires de carré intégrable	82
4.4.1	Variance et Covariance	82
4.4.2	Approximation linéaire	84
4.5	Calcul de l'espérance pour une variable aléatoire à densité	84
4.5.1	Un résultat général fondamental	84
4.5.2	Calculs d'espérances dans le cas avec densité	85
4.6	Exemples fondamentaux de variables à densité	86
4.6.1	Variable aléatoire uniforme sur $[a, b]$	86
4.6.2	Variable aléatoire exponentielle	87
4.6.3	Variable aléatoire de loi gamma	90
4.6.4	Variables aléatoires normales (ou variables gaussiennes)	90
4.7	Des inégalités fameuses	94
4.7.1	Inégalité de Bienaymé-Chebyshev	94
4.7.2	Inégalité de Cauchy-Schwarz	96
4.7.3	Inégalité de Jensen	96
4.8	Vecteurs aléatoires	97
4.8.1	Vecteurs aléatoires	98
4.8.2	Moments d'un vecteur aléatoire	99
4.8.3	Densités marginales et conditionnelles	99
4.9	Variables aléatoires indépendantes	102
4.9.1	Indépendance de deux variables aléatoires	102
4.9.2	Suite de variables aléatoires indépendantes	104
4.10	Calculs de lois	106
4.10.1	Un théorème d'identification	106
4.10.2	Recherche de densité	107
4.11	Simulation de suites indépendantes de variables aléatoires	110
4.11.1	Inversion de la fonction de répartition	111
4.11.2	Méthode du rejet	112
5	Convergences et loi des grands nombres	115
5.1	Convergences de variables aléatoires	116
5.2	La loi des grands nombres	119
5.3	Méthode de Monte-Carlo	123
6	Fonctions caractéristiques et convergence en loi	125
6.1	La fonction caractéristique	125

6.1.1	Définition et premières propriétés	125
6.1.2	Exemples	127
6.1.3	Propriété fondamentale	128
6.1.4	Somme de variables aléatoires indépendantes	130
6.1.5	Fonction caractéristique et moments	130
6.2	Vecteurs gaussiens	131
6.3	Convergence en loi	135
6.4	Le théorème de la limite centrale	137
6.5	Intervalle de confiance	140
6.5.1	Sondages	140
6.5.2	Précision dans la méthode de Monte-Carlo	143
7	Modèles dynamiques aléatoires	145
7.1	Marche aléatoire simple	145
7.2	Processus de branchement	149
7.2.1	Sommes aléatoires de v.a. indépendantes	149
7.2.2	Processus de branchement	151
7.2.3	Percolation sur un arbre	155
7.3	Files d'attente	156
7.3.1	Un modèle simple en temps discret	156
7.3.2	Stabilité : étude analytique	158
7.3.3	Stabilité : expérience de simulation	161
7.4	Suites récurrentes aléatoires discrètes	163
7.4.1	Probabilités de transition	163
7.4.2	Stabilité	166

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 cours 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 (mouvement de particules, formation de gouttes d'eau), de la météorologie, de la biologie (mutation du génôme), de l'écologie (déplacement des oiseaux migrateurs pendant la grippe aviaire), de la médecine (traitement d'images), de l'économie (marchés boursiers), ou de la sociologie.

1.1 Introduction du cours

Ce cours introduit toutes les notions de base de la théorie des probabilités et permet d’acquérir le raisonnement probabiliste. La 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 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).

Soulignons que les probabilités sont en lien étroit avec la vie quotidienne. À ce titre, elles s’appuient sur un passage du concret à l’abstrait : **la modélisation**, ce qui les rend **difficiles, mais palpitantes**. L’apprentissage de ce raisonnement probabiliste sera développé dans le cours et les PC. Par ailleurs, il est important pour vous d’apprendre à fabriquer, et à utiliser à des fins numériques, des expérimentations fictives sur machine, ce qu’on appelle **des simulations**. C’est le but des projets personnels que vous aurez à développer.

1.2 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 maintenant 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 théorie aléatoire de la finance. On peut 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.

1.3 Phénomènes aléatoires

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 1.3.1 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.

On peut 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 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.

Tous 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. Tous les effets physiques dans la nature fonctionnent ainsi, et chaque phénomène déterministe est inévitablement accompagné d'écarts aléatoires. Néanmoins pour certains phénomènes, on peut négliger les éléments aléatoires et remplacer le phénomène réel par un schéma simplifié, en sélectionnant pour ce faire les paramètres les plus importants. Ce qui justifie par exemple l'approche de la mécanique classique.

1.4 Deux idées majeures et incontournables

Deux idées majeures justifient 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 ce qu'il vous faudra assimiler en profondeur.

1.4.1 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 allons être capables de donner des renseignements sur ce type de phénomènes. L'idée majeure est que ces renseignements seront donnés par **la répétition de l'expérience**.

En effet, quand on observe un grand nombre de phénomènes aléatoires, on y décèle généralement des lois régissant les résultats, tout à fait déterminées, stables. Par exemple,

quelle que soit la pièce non truquée avec laquelle on joue à Pile ou Face, quel que soit l'endroit où l'on joue, si on lance 1000 fois la pièce, on aura environ 50% de piles, 50% de faces. De même, si l'on étudie la répartition des tailles d'un groupe d'individus, quel que soit l'échantillon pris dans ce groupe, on aura toujours une courbe des répartitions de même type. On va ainsi pouvoir **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**, qu'on appellera **Loi des grands nombres**, qui légitime l'utilisation d'une modélisation mathématique.

La **Théorie des Probabilités** va essayer de modéliser tous ces types de situations aléatoires, aussi différents soient-ils, par une approche unifiée, et reposant sur une abstraction mathématique.

1.4.2 Conditionnement et Indépendance

Il faudra bien comprendre dans la suite que la construction d'un modèle probabiliste repose fondamentalement sur l'information que l'on connaît *a priori* sur l'expérience aléatoire. Ce modèle permet de quantifier les probabilités de réalisations de certains résultats de l'expérience. Ce qui est fondamental est que **si l'information change, les probabilités de réalisation changent**. (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 1,75m). La richesse du modèle que nous allons construire réside dans le fait que si l'information change par rapport au modèle initial, on pourra calculer les nouvelles chances de réalisation. Tout 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), on dit que ces phénomènes sont indépendants. Cette notion d'**indépendance entre les probabilités de réalisation** va être une hypothèse fondamentale dans toute la théorie.

1.5 Les variables aléatoires

1.5.1 Loi d'une variable aléatoire

Vous allez dans ce cours découvrir une nouvelle manière de "penser" une fonction.

Les fonctions que l'on va étudier peuvent prendre différentes valeurs, obtenues au hasard. On les appelle des variables aléatoires, car elles varient aléatoirement. Plutôt que de chercher les antécédents de chaque valeur possible de la fonction, nous allons nous intéresser à la chance pour la fonction d'être égale à une de ces valeurs ou à un ensemble de ces valeurs. C'est cela que l'on appellera la loi de la variable. Cette notion de **loi d'une variable aléatoire** est à la base du raisonnement probabiliste moderne.

1.5.2 Simulation de variables aléatoires

La méthode de Monte-Carlo, du nom du quartier où se trouve le casino de MONACO, 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 basé 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 massive 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 :

- la méthode `Math.random` en JAVA,
- les fonctions `rand` et `grand` sous SCILAB.

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 là, on peut simuler de nombreux types de loi.

1.6 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. 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 oeuvres de Kepler (1571-1630) et de Galilée (1564-1642). Le calcul probabiliste se développe au cours du 17^{ème} 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 17^{ème} 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 17^{ème} 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.

Ainsi, 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), Leibnitz (1646-1716), et plus tard Laplace (1749-1827), Poisson (1781-1840) (*Recherches sur la probabilité des jugements en matières criminelles 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 il faut, pour développer plus avant la théorie, faute d'outils mathématiques puissants, attendre Laplace (1749-1827) qui 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 la théorie. Dans les pays anglo-saxons, se développe également l'outil statistique, étude des données et analyse de ce que l'on peut en tirer (ne pas oublier que le mot "statistique" vient du mot "état", et que cela a été, depuis cette époque, un outil puissant pour les organismes de décisions). Cela ne peut se faire qu'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 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.

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 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 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 déjà introduit le mouvement brownien en 1900 pour modéliser la dynamique d'un cours boursier. Ce processus aléatoire, évoluant de manière apparemment erratique, s'est avéré être l'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, Paul Lévy (1886-1971, Professeur à l'Ecole Polytechnique) 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 K. Itô (prix Gauss 2006), qui définit une notion d'intégrale par rapport au mouvement brownien et, grâce à elle, 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 suscité une grande émotion.

L'Ecole française de Probabilités est très active. D'anciens professeurs de l'Ecole Polytechnique, tels Métivier ou Neveu, en ont été les piliers.

Vous avez, j'espère, tous entendu parler de la médaille Fields décernée en août 2006 à Wendelin Werner (Université Paris 11, ENS). "Communiqué de Presse : *Les travaux de Wendelin Werner sont à l'interaction de la physique et des mathématiques. Ils introduisent des idées et des concepts nouveaux combinant la théorie des probabilités et l'analyse complexe, pour comprendre les phénomènes critiques de certaines transitions de phase (par exemple, la transition liquide/gaz)*".

Les probabilités continuent à se développer énormément aujourd'hui, de plus en plus en fait, alimentées en particulier de manière essentielle par la physique théorique, le développement des réseaux de télécommunications, la finance, et plus récemment, par la biologie. Par ailleurs, les modèles probabilistes se développent aussi autour des statistiques, dans de multiples domaines d'applications.

1.7 Bibliographie

Cours de l'Ecole : J.M. Bony : Intégration et Analyse hilbertienne.

F. Comets, N. El Karoui, J. Neveu : Probabilités.

F. Comets : Aléatoire. Introduction aux probabilités et à la simulation aléatoire.

Ouvrages généraux :

P. Billingsley : *Probability and Measure*, Wiley, New York (1979).

L. Breiman : *Probability*, Addison Wesley 1968.

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.

G. Grimmett, D. Stirzaker : *Probability and Random Processes*, Oxford University Press 1992.

J. Jacod, P. Protter : *L'essentiel en théorie des probabilités*, Cassini, 2003.

Y. Lacroix, L. Mazliak : *Probabilités, variables aléatoires, convergences, conditionnement, ellipses*, 2006.

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.

Pour savoir ce qui vous attend (si vous le souhaitez) l'année prochaine :

M. Benaim, N. El Karoui : *Promenade aléatoire*, Cours de l'Ecole, 2001.

Bibliographie des Mathématiciens : <http://turnbull.mcs.st-and.ac.uk/~history/>

Pour les prérequis en Analyse et Algèbre : replongez-vous dans vos livres de classes préparatoires.

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 aléatoires

a) Expérience Aléatoire

Définition 2.1.1 On appelle **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 d'états** (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. On lance deux pièces à Pile ou Face : $\Omega = \{PP, PF, FP, FF\}$
- Exemple 2. On lance un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Exemple 3. On envoie une fléchette sur une cible circulaire de 30 cm de diamètre et l'expérience consiste à décrire l'impact de la flèche dans un repère orthonormé de

centre le centre de la cible :

$$\Omega = \{(x, y), \sqrt{x^2 + y^2} \leq 15\}.$$

- Exemple 4. On étudie la durée de vie d'une ampoule électrique :

$$\Omega = [0, +\infty[.$$

- Exemple 5 : Romeo 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. On observe les 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 limitées à gauche sur $[t_1, t_2]$, à valeurs réelles.

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.

On verra également ultérieurement que le modèle abstrait que l'on va construire permettra de s'affranchir du fait que Ω décrit précisément tous les résultats possibles de l'expérience.

b) Événements aléatoires

Définition 2.1.2 On appelle **événement aléatoire** (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 lancé de deux dés,

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

est un événement aléatoire, 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 Romeo, l'ensemble "Juliette se fait attendre plus de 3/4 d'heure" est l'événement aléatoire $]3/4, 1]$.

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 aléatoire.

Ainsi, les événements aléatoires sont des ensembles. On va 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 . On écrit $A \subset \Omega$ et on dit que A est inclus dans Ω .

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. On le note $\emptyset$.

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

Complémentaire : si $A \subset \Omega$, son complémentaire (dans Ω) est l'ensemble des points de Ω n'appartenant pas à A . On le note 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 , tels que $B \subset A$, on note $A \setminus B$ l'ensemble des points qui sont dans A mais pas dans B . On a donc $A \setminus B = A \cap B^c$.

La réunion et l'intersection sont des opérations commutatives et associatives. On a $A \cup B = B \cup A$ et $A \cap B = B \cap A$, et aussi $A \cup (B \cup C) = (A \cup B) \cup C$ et $A \cap (B \cap C) = (A \cap B) \cap C$, ensembles qu'on note 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 , on note $\cup_{i \in I} A_i$ **la réunion de cette famille**, i.e. l'ensemble des points appartenant **à au moins** l'un des A_i . De même, on note $\cap_{i \in I} A_i$ **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. On appelle **partition** de Ω toute famille $(A_i)_{i \in I}$ telle que les ensembles A_i soient disjoints deux-à-deux ($A_i \cap A_j = \emptyset$, $\forall i, j$), et que $\cup_{i \in I} A_i = \Omega$.

d) Modélisation ensembliste des événements aléatoires

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

CORRESPONDANCE ENTRE OPERATIONS ENSEMBLISTES ET EVENEMENTS ALEATOIRES :

Si A et B sont deux événements,

- NON : la réalisation de l'événement contraire à A est représenté 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 $A \cap B = \emptyset$, on dit que A et B sont incompatibles. Cela décrit le fait qu'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.

On note $\mathcal{A}$ l'ensemble de tous les événements. Il modélise l'**information** que l'on peut obtenir à partir des résultats de l'expérience. On peut avoir (mais pas toujours, on verra pourquoi plus loin), $\mathcal{A} = \mathcal{P}(\Omega)$, ensemble de toutes les parties de Ω .

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 on doit avoir $A \cap B \in \mathcal{A}$, $A \cup B \in \mathcal{A}$, $A^c \in \mathcal{A}$, mais aussi $\Omega \in \mathcal{A}$, $\emptyset \in \mathcal{A}$.

2.1.2 Probabilité - Premières propriétés

On cherche à définir, pour un ensemble possible de réalisations de l'expérience $A \in \mathcal{A}$, la vraisemblance qu'on accorde a priori à A (avant le résultat de l'expérience). On cherche donc à associer à chaque événement A un nombre $P(A)$ compris entre 0 et 1, qui représente la chance que cet événement soit réalisé à la suite de l'expérience.

APPROCHE INTUITIVE

Pensons à un jeu de Pile ou Face. Si nous jouons 3 fois, une suite de 3 résultats n'offre presque aucune information. En revanche, si nous jouons 10000 fois et si nous obtenons 5003 Pile et 4997 Face, il est tentant de dire que la probabilité que la pièce tombe sur Pile est la même que celle qu'elle tombe sur Face, à savoir $1/2$.

Cette idée intuitive est à la base du raisonnement probabiliste.

Considérons un événement A lié à une expérience aléatoire donnée $\mathcal{E}$. Supposons que l'on répète n fois l'expérience $\mathcal{E}$. On note n_A le nombre de fois où A est réalisé ;

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

est la fréquence de réalisation de A sur ces n coups. On a les propriétés suivantes :

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

$$\text{Si } A \text{ et } B \text{ sont disjoints, on a } f_n(A \cup B) = f_n(A) + f_n(B).$$

L'approche intuitive et naturelle consiste à définir $P(A)$ comme étant la limite quand n tend vers l'infini des fréquences d'apparition $f_n(A)$. On donnera 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.

Intuitivement,

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

Des propriétés évidentes des fréquences, on en déduit immédiatement que

Proposition 2.1.3 *Une Probabilité sera définie sur les ensembles aléatoires liés à l'expérience et vérifiera les propriétés essentielles suivantes :*

$$\bullet \quad 0 \leq P(A) \leq 1, \quad (2.1.2)$$

$$\bullet \quad P(\Omega) = 1, \quad (2.1.3)$$

$$\bullet \quad P(A \cup B) = P(A) + P(B) \quad \text{si} \quad A \cap B = \emptyset, \quad (2.1.4)$$

et il en découle que

Corollaire 2.1.4 *Une probabilité vérifiera de plus :*

$$\bullet \quad P(\emptyset) = 0 \quad (2.1.5)$$

$$\bullet \quad P(A) + P(A^c) = 1, \quad (2.1.6)$$

$$\bullet \quad P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) \text{ , si les } A_i \text{ sont deux-à-deux disjoints,} \quad (2.1.7)$$

$$\bullet \quad P(A \cup B) + P(A \cap B) = P(A) + P(B), \quad (2.1.8)$$

$$\bullet \quad P(A) \leq P(B) \quad \text{si} \quad A \subset B. \quad (2.1.9)$$

Preuve. (2.1.5) se montre en appliquant (2.1.4) avec $A = B = \emptyset$, (2.1.6) s'obtient de la même façon avec A et A^c . Pour prouver (2.1.8), on décompose l'ensemble A en l'union des deux ensembles disjoints $A \cap B$ et $A \setminus B$. De même pour B . La figure 2.1 peut vous aider à comprendre. (2.1.9) se déduit de (2.1.4) avec $B = A \cup (B \setminus A)$. $\square$

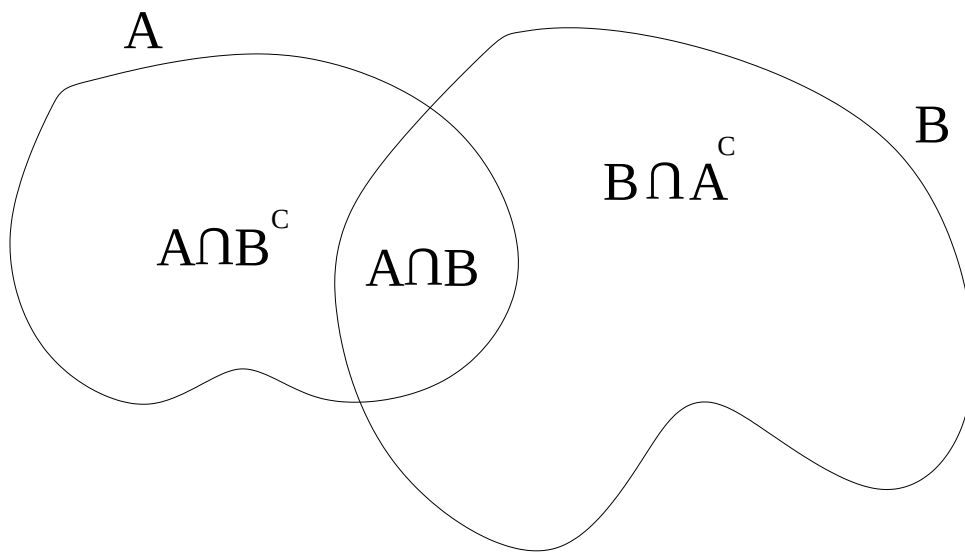


FIG. 2.1 – A , B , $A \cap B^c$, $A \cap B$, $B \cap A^c$ et $A \cup B$

Définition 2.1.5 Un **modèle probabiliste** sera un triplet $(\Omega, \mathcal{A}, P)$ constitué de l'espace Ω , de l'ensemble des événements $\mathcal{A}$, et de la famille des $P(A)$ pour $A \in \mathcal{A}$. On peut ainsi considérer 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. (On verra ultérieurement qu'elle doit satisfaire une propriété supplémentaire nécessaire dès que Ω n'est pas un espace fini).

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, on suppose que l'espace de probabilité Ω est un ensemble fini. On prend alors $\mathcal{A} = \mathcal{P}(\Omega)$.

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

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

On a donc 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 , on aura la proposition fondamentale suivante.

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

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

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

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

On a alors

$$\forall A \in \mathcal{P}(\Omega), \quad P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p_{\omega}. \quad (2.2.11)$$

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

Preuve. Soit d'abord P une probabilité sur Ω , et soit $p_\omega = P(\{\omega\})$. Il est alors évident que $0 \leq p_\omega \leq 1$, et (2.2.11) découle de (2.1.7), puisque n'importe quelle partie A de Ω est réunion disjointe (et finie) des singletons $\{\omega\}$, pour les $\omega \in A$. On a 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 $P(\Omega) = 1$.

Inversement, considérons n nombres $(p_i)_{1 \leq i \leq n}$ vérifiant (2.2.10). On pose $P(\{\omega_i\}) = p_i$ et pour tout $A \subset \Omega$, on définit ainsi $P(A)$ 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 :

$$\Omega = \{w_1, w_2\} \quad \text{et} \quad p_{w_1} = p ; p_{w_2} = 1 - p.$$

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, $\Omega = \{P, F\}$, que l'on assimile plus simplement à $\{0, 1\}$. Si la pièce est équilibrée, on supposera que $p = 1/2$. Mais cette probabilité peut aussi modéliser la probabilité de réalisation 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 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 On dit que la probabilité P sur l'espace fini Ω est uniforme si $p_\omega = P(\{\omega\})$ ne dépend pas de ω . On a 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 P est une probabilité uniforme, on déduit de (2.2.11) que

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

de sorte que le calcul des probabilités se ramène, dans ce cas, à des dénombrements : on est 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, lancé 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, attention à bien préciser l'espace de probabilité sur lequel vous travaillez.

Cette remarque prend toute son ampleur dans ce paragraphe, où l'on va développer différents "modèles d'urnes" qu'on peut également voir comme des modèles de prélèvement d'échantillons dans une population, au cours d'un sondage. Ils interviennent aussi en contrôle de fabrication, ou dans de multiples autres situations. Si les couleurs des boules d'une urne ne vous inspirent pas, transcrivez 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, ..., N_k boules de couleur k . On posera $p_i = \frac{N_i}{N}$ la proportion de boules de couleur i . On tire au hasard n boules de cette urne, $n \leq N$, et on s'intéresse à la répartition des couleurs dans l'échantillon obtenu. On note 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 $n_1 + n_2 + \dots + n_k = n$. On va considérer trois façons de tirer les boules au hasard : tirage avec remise, tirage sans remise, tirage simultané.

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 p et n deux entiers tels que $p \leq n$, nous allons souvent utiliser, dans la suite, la valeur $\binom{n}{p}$ du nombre de parties à p éléments dans un ensemble à n éléments. Nous utiliserons dans ce cours la notation anglo-saxonne, à savoir :

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

Tirage exhaustif ou simultané - La loi hypergéométrique

On tire toutes les boules d'un coup. Ω est alors l'ensemble de toutes les parties possibles de taille n d'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 $\binom{N_1}{n_1} \dots \binom{N_k}{n_k}$. La probabilité cherchée vaut alors

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

On appelle cette distribution la distribution polygénométrique. Dans le cas de deux couleurs,

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

et on parle de loi hypergénométrique.

Ainsi, si dans une fabrication en série, on sait que parmi N pièces usinées, M sont à mettre au rebut, et si on prend au hasard 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. On remplace la boule tirée dans l'urne avant le tirage suivant. On peut donc tirer plusieurs fois la même boule. Ω est alors l'ensemble de tous les n -uplets d'éléments de l'urne ; toutes les répétitions étant possibles, $\text{card}(\Omega) = N^n$. On munit Ω de sa probabilité uniforme. Le nombre de n -uplets de répartition $n_1, n_2, \dots, n_k$ est alors égal à $\frac{n!}{n_1! n_2! \dots n_k!} N_1^{n_1} \dots N_k^{n_k}$: 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, on a N_i possibilités pour chaque boule de couleur i . On a donc finalement

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

On appelle cette probabilité une distribution multinomiale. Dans le cas particulier où $k = 2$, $p_1 = p$ 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 . (On reconnaît les coefficients du binôme, d'où son nom). Attention, les paramètres ne jouent pas le même rôle : $n \in \mathbb{N}$ 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 : on pourra vérifier que

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

Tirage sans remise

On tire maintenant successivement les boules de l'urne, mais sans les replacer dans l'urne après tirage. 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, on peut 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 exhaustif.

Ainsi, on a équivalence du tirage sans remise et du tirage simultané, du point de vue de la composition de l'échantillon, et l'on peut donc 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

On se place 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 $p < 1$. On montre alors très simplement que

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

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 l'équirépartition pour 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 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 pour que la somme des points amenés dépasse dix est égale à la probabilité pour que cette somme ne dépasse pas dix. (Cela vous 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. On munit Ω de la probabilité 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", on remarque 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 de réaliser 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 est

$$1 - \left(\frac{5}{6}\right)^4 \simeq 0.5177 > \frac{1}{2}.$$

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 est :

$$1 - \left(\frac{35}{36}\right)^{24} \simeq 0.4914 < \frac{1}{2}.$$

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 !

EXERCICE 2.2.11 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).

Si un professeur a 40 élèves dans sa classe, quelle chance a-t-il de se tromper s'il affirme qu'au moins deux de ses élèves sont nés le même jour ?

2.3 Définition générale des Probabilités

2.3.1 Pourquoi la définition précédente ne suffit-elle pas ?

Lorsque l'espace d'états Ω n'est pas fini, la Définition 2.2.1 n'est pas suffisante.

On va s'en rendre compte sur l'ensemble suivant. Jouons à Pile ou Face.

Si on joue n fois à Pile ou Face, 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 , et si on suppose que la pièce n'est pas truquée, la probabilité de chaque tirage est uniforme et on se retrouve dans le cadre de la combinatoire. Ainsi, pour tout $A \subset \Omega$,

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

Supposons maintenant que le jeu se poursuive indéfiniment. L'espace d'états 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 $P(A)$ 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), on a $P(A_n) = 2^{-n}$. 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

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

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

On dit qu'un ensemble E est dénombrable s'il est en bijection avec $\mathbb{N}$, c'est-à-dire si l'on peut énumérer ses points en une suite $(x_n)_{n \in \mathbb{N}}$. C'est le cas de l'ensemble $\mathbb{N}$ lui-même,

de $\mathbb{Z}$, de $\mathbb{Q}$, ou 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 $a < b$.

Enonçons quelques propriétés des ensembles dénombrables.

- Tout ensemble dénombrable est infini. (Mais la réciproque est fausse comme 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) $\emptyset \in \mathcal{A}$ et $\Omega \in \mathcal{A}$.
- (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 dénombrable, i.e. si $(A_n)_{n \in \mathbb{N}}$ est une suite d'éléments de $\mathcal{A}$, alors $\bigcap_{n \in \mathbb{N}} A_n$ et $\bigcup_{n \in \mathbb{N}} A_n$ sont dans $\mathcal{A}$.

Noter que $\mathcal{A}$ est alors 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 $n \geq 1$.

Noter 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 l'ensemble des parties de Ω (composées donc 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 l'on va être capable de définir sa probabilité de réalisation $P(A)$.

Comme ces ensembles A sont des sous-ensembles de résultats possibles de l'expérience, la tribu modélise **l'information** que nous donne cette expérience.

Exemple 2.3.2 • $\mathcal{A} = \{\emptyset, \Omega\}$ est la tribu grossière, ou triviale : c'est la plus petite tribu de Ω .

- L'ensemble $\mathcal{P}(\Omega)$ des parties de Ω est une tribu sur Ω . C'est celle que l'on choisira 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.

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 l'ensemble A 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 $\mathbb{R}$.

A 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 :

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

et l'on en déduit que tout intervalle ouvert appartient à $\mathcal{C}$, d'où le résultat. $\square$

2.3.4 Définition d'une probabilité

Passons maintenant à la définition générale d'une probabilité.

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

•

$$P(\Omega) = 1. \quad (2.3.18)$$

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

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

L'axiome (2.3.19), dit "axiome de σ -additivité" est plus fort que (2.1.4). Pour le voir, on commence par appliquer (2.3.19) avec $A_n = \emptyset$ pour tout $n \in \mathbb{N}$: si $a = P(\emptyset)$, on obtient $\sum_n a = a$, ce qui entraîne $a = 0$. Ensuite, si $A, B \in \mathcal{A}$ sont disjoints, on applique (2.3.19) avec $A_0 = A, A_1 = B$ et $A_n = \emptyset$ pour tout $n \geq 2$, ce qui donne $P(A \cup B) = P(A) + P(B) + \sum_{n \geq 2} P(\emptyset) = P(A) + P(B)$, d'où (2.1.4).

Remarquons que (2.3.19) entraîne en particulier que la série de terme général $P(A_n)$ est convergente.

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, on utilise les notations suivantes. Si (A_n) est une suite décroissante de parties de Ω , i.e. $A_{n+1} \subset A_n$ pour tout n et si $A = \cap_n A_n$, on écrit $A_n \downarrow A$. De même, si la suite (A_n) est croissante, i.e. $A_n \subset A_{n+1}$ pour tout n et si $A = \cup_n A_n$, on écrit $A_n \uparrow A$.

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

- (i) *La propriété de σ -additivité (2.3.19)*
- (ii) *Pour toute suite $(A_n)_n$ **croissante**,*

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

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

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

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) 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 P vérifie (2.1.4), elle vérifie (2.1.7) et on a $P(B_n) = \sum_{p \leq n} P(A_p)$ qui croît vers $\sum_n P(A_n)$ et aussi vers $P(B)$ par (ii). On a 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, on a

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

la dernière égalité provenant de (2.1.7). On a donc le résultat. $\square$

La propriété (2.3.19) donne la probabilité de la réunion $\cup_n A_n$ en fonction des $P(A_n)$, lorsque les A_n sont deux-à-deux disjoints. Si ce n'est pas le cas, on a tout de même la majoration suivante, très utile dans la pratique :

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

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

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

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

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), on a $P(C) + P(B \cap A_k) = P(B) + P(A_k)$, donc $P(C) \leq P(B) + P(A_k)$, et on en déduit immédiatement que (2.3.23) est satisfait pour k .

b) Passons maintenant au cas où I est dénombrable. On peut supposer sans restriction que $I = \mathbb{N}^*$. On pose $B_n = \cup_{i=1}^n A_i$, qui croît vers l'ensemble $C = \cup_{n \in I} A_n$. D'après (a), on a

$$P(B_n) \leq \sum_{i=1}^n P(A_i).$$

Mais le membre de gauche ci-dessus croît vers $P(C)$ en vertu de la proposition précédente, tandis que le membre de droite croît vers $\sum_{n \in I} P(A_n)$. En passant à la limite, on obtient donc (2.3.22). $\square$

Définition 2.3.10 On appelle le triplet $(\Omega, \mathcal{A}, P)$ un **espace de probabilité**.

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

Remarquons que l'on peut construire de nombreuses probabilités distinctes sur le même espace $(\Omega, \mathcal{A})$. On verra beaucoup d'exemples ultérieurement, mais on peut s'en convaincre rapidement dans le cadre du jeu de Pile ou Face, suivant que la pièce est truquée ou non truquée. (On peut définir sur $\{0, 1\}$ différentes lois de Bernoulli, suivant le paramètre $p \in [0, 1]$ que l'on choisit).

La définition suivante est fondamentale en Probabilités. Elle introduit une notion de "vrai" ou "faux" qui dépend de la probabilité choisie sur l'espace d'états.

Définition 2.3.11 Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité. Un événement de probabilité nulle est dit négligeable, et une propriété est vraie P -presque-sûrement (en abrégé P -p.s.), si l'ensemble des $\omega \in \Omega$ pour lesquels elle est vraie est de probabilité égale à 1.

2.3.5 Probabilités sur un espace dénombrable

Supposons que Ω soit dénombrable. On peut alors numéroter ses éléments par $\{w_0, w_1, \dots, w_n, \dots\}$, $n \in \mathbb{N}$. 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 sa valeur sur les singletons.

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

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

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

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

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 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.) $\square$

2.3.6 Variable aléatoire et sa loi

En théorie moderne des probabilités, on préfère prendre un point de vue fonctionnel plutôt qu'ensembliste, et travailler sur les variables aléatoires plutôt que sur les événements.

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

- le nombre de 6 obtenus dans un lancé 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 $(\Omega, \mathcal{A}, P)$, et une variable aléatoire X est une application de Ω dans un ensemble F , qui à ω associe la valeur $X(\omega) \in F$.

En pratique, l'ensemble F pourra être soit $\mathbb{N}$ ou tout ensemble au plus dénombrable (on parlera de **variable aléatoire discrète**), soit $\mathbb{R}$ (on parlera de **variable aléatoire réelle**), soit $\mathbb{R}^d$ (on parlera de **vecteur aléatoire**), mais aussi 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, on va préférer s'intéresser aux chances de réalisations des valeurs de X plutôt qu'aux chances de réalisation des résultats de l'expérience. Ainsi donc, grâce à une variable aléatoire X , on peut transporter la structure abstraite du modèle probabiliste $(\Omega, \mathcal{A}, P)$ sur l'espace d'arrivée F (dont la structure est mieux connue), en posant pour $B \subset F$

$$P_X(B) = P(X^{-1}(B)) = P(\{\omega, X(\omega) \in B\}). \quad (2.3.25)$$

L'application P_X , définie sur les parties de F et à valeurs dans $[0, 1]$, appelée **loi de la variable X** , ou distribution de X , est plus facile à caractériser que P , puisque F est connu.

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

Comme on l'a dit, F sera en général, dans ce polycopié, égal à $\mathbb{R}$ ou $\mathbb{R}^d$, ou à un sous-ensemble de ces ensembles, comme $\mathbb{N}$, mais vous rencontrerez ultérieurement des cas où F pourra être un ensemble beaucoup plus gros, comme par exemple un ensemble de fonctions. Dans tous les cas, comme $P(A)$ n'est définie que pour les A appartenant à $\mathcal{A}$, la formule (2.3.25) ne permet de définir $P_X(B)$ que pour les ensembles B tels que $X^{-1}(B) = \{X \in B\} \in \mathcal{A}$, d'où l'intérêt de la proposition suivante.

Proposition 2.3.13 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 P_X définie pour $B \in \mathcal{F}$ par

$$P_X(B) = P(X^{-1}(B)) = P(X \in B) \quad (2.3.26)$$

définit une probabilité sur la tribu $\mathcal{F}$ de F , appelée **la loi de X** .

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

$$X^{-1}(\emptyset) = \emptyset, X^{-1}(F) = \Omega, X^{-1}(B^c) = (X^{-1}(B))^c,$$

$$X^{-1}(\cap_i A_i) = \cap_i X^{-1}(A_i), X^{-1}(\cup_i A_i) = \cup_i X^{-1}(A_i). \quad (2.3.27)$$

□

On remarquera qu'en général, $\mathcal{F} \neq \mathcal{P}(F)$, même si on a $\mathcal{A} = \mathcal{P}(\Omega)$. Comme P_X ne peut être définie que sur $\mathcal{F}$, ceci constitue une première raison, d'ordre mathématique, pour qu'en général, une probabilité soit définie sur une tribu qui peut être strictement plus petite que l'ensemble de toutes les parties.

Les variables que l'on rencontrera dans ce cours seront soit à valeurs dans un ensemble dénombrable, soit à valeurs dans $\mathbb{R}$ ou dans $\mathbb{R}^d$. On les appellera respectivement des variables aléatoires discrètes, réelles ou des vecteurs aléatoires. Les lois seront alors des probabilités respectivement sur un ensemble dénombrable, sur $\mathbb{R}$ ou sur $\mathbb{R}^d$. On a 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 aléatoires au Chapitre 4.

Exemple 2.3.14 Etudions un lancer de deux dés que l'on suppose bien équilibrés. Dans ce cas, l'espace d'états 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 aléatoire $A \subset \Omega$. Puisque les dés sont équilibrés, on définit la probabilité de A par $P(A) = \frac{\text{card}(A)}{36}$ où $\text{card}(A)$ est le cardinal de A et désigne le nombre de points contenus dans A . On vérifie aisément que les propriétés (2.1.2), (2.1.3) et (2.1.4) sont satisfaites, et on a $P(\{\omega\}) = \frac{1}{36}$ pour chaque singleton $\{\omega\}$.

L'application $X : \Omega \rightarrow \{1, 2, \dots, 12\}$ définie par $X(i, j) = i + j$ est la variable aléatoire "somme des résultats des deux dés". Elle a pour loi

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

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

EXERCICE 2.3.15 Montrer la formule de Poincaré :

$$P(\cup_{m=1}^n A_m) = p_1 - p_2 + \dots + (-1)^{n-1} p_n = \sum_{k=1}^n (-1)^{k-1} p_k \quad (2.3.28)$$

où

$$p_k = \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}). \quad (2.3.29)$$

Exemple 2.3.16 Un facteur possède n lettres adressées à n destinataires distincts. Il est totalement ivre et poste complètement 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 ?

Solution :

1) Il y a une seule possibilité de bonne remise de lettres parmi les $n!$ possibilités. La probabilité de bonne répartition est donc $\frac{1}{n!}$.

2) Numérotions de 1 à n les lettres et numérotions par les mêmes numéros les boîtes qui sont censées leur correspondre. Appelons A_i l'événement "La lettre numéro i arrive dans la boîte numéro i ". Ainsi, A_i sera réalisé si la remise de la lettre i est fixée, donc si les $n-1$ autres sont aléatoires. Ainsi, $P(A_i) = \frac{(n-1)!}{n!}$. De même, pour deux numéros quelconques i_1 et i_2 , on aura $P(A_{i_1} \cap A_{i_2}) = \frac{(n-2)!}{n!}$.

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.

$$\begin{aligned} P(E) &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \\ &= \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^n \frac{(-1)^{k-1}}{k!}. \end{aligned}$$

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

$$P(E^c) = 1 - P(E) = \sum_{k=0}^n \frac{(-1)^k}{k!},$$

qui tend donc vers $e^{-1} = 0,368$ quand n tend vers l'infini.

4) $d_n = n!P(E^c)$.

□

EXERCICE 2.3.17

1) Soit $\theta > 0$. Montrer que la suite de nombres définie par $p_n = e^{-\theta} \frac{\theta^n}{n!}$ caractérise une probabilité sur $\mathbb{N}$.

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

2.4 Conditionnement et indépendance**2.4.1 Probabilités conditionnelles**

La notion de **probabilité conditionnelle** est l'une des plus fructueuses de la théorie des probabilités. 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), si on lance 2 dés et que l'on cherche 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, on a, dans chaque cas, calculé le rapport du nombre de résultats favorables sur le nombre de cas possibles. On se rend compte qu'il est indispensable de bien définir l'espace de probabilité lié à l'expérience **munie de l'information a priori**. On remarque également que l'information a priori a changé la valeur de la probabilité de l'événement aléatoire.

L'approche intuitive pour formaliser cette notion est de 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 égal 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 si l'on fait tendre n vers l'infini, on aboutira à la notion suivante.

Définition 2.4.1 Soit A et B deux événements, avec $P(B) > 0$. **La probabilité conditionnelle de A sachant B** est le nombre

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (2.4.30)$$

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

Proposition 2.4.2 1) Soit $(\Omega, \mathcal{A}, P)$ 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 $P(A|B)$ définit une nouvelle probabilité sur Ω , appelée probabilité conditionnelle sachant B .

2) Si $P(A) > 0$ et $P(B) > 0$, on a

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A).$$

Preuve. On a $0 \leq P(A|B) \leq 1$. Par ailleurs, les propriétés (2.3.18) et (2.3.19) pour $P(\cdot|B)$ proviennent des mêmes propriétés pour P et des remarques suivantes : on a $\Omega \cap B = B$, 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 2.4.3 Formule des probabilités composées : Si $A_1, \dots, A_n$ sont des événements de Ω tels que $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$,

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}). \quad (2.4.31)$$

Preuve. On fait une démonstration par récurrence. Si $n = 2$, les formules (2.4.30) et (2.4.31) sont les mêmes. Supposons (2.4.31) vraie pour $n-1$, et soit $B = A_1 \cap A_2 \cap \dots \cap A_{n-1}$. D'après (2.4.30), on a $P(B \cap A_n) = P(B)P(A_n|B)$. En remplaçant $P(B)$ par sa valeur donnée par (2.4.31) avec $n-1$, on obtient (2.4.31) pour n . $\square$

Proposition 2.4.4 Formule des probabilités totales : soit $(B_i)_{i \in \mathbb{N}}$ une partition finie ou dénombrable d'événements de Ω , tels que $P(B_i) > 0$ pour chaque i . Pour tout $A \in \mathcal{A}$, on a alors

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i). \quad (2.4.32)$$

Preuve. On a $A = \cup_{i \in I} (A \cap B_i)$, les $(A \cap B_i)$ sont deux-à-deux disjoints, et $P(A \cap B_i) = P(A|B_i)P(B_i)$. Il suffit alors d'appliquer (2.3.19). $\square$

Théorème 2.4.5 Formule de Bayes : Sous les mêmes hypothèses que dans la proposition 2.4.4, et si $P(A) > 0$,

$$\forall i \in \mathbb{N}, P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}. \quad (2.4.33)$$

Preuve. Le dénominateur de (2.4.33) vaut $P(A)$ d'après (2.4.32), tandis que (2.4.30) implique

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

□

EXERCICE 2.4.6 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.

Exemple 2.4.7 Un individu est tiré au hasard d'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 ($1 - 0,001$ 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 $P(A) = 10^{-4}$ d'où $P(A^c) = 0,9999$, $P(B|A) = 0,99$ et $P(B|A^c) = 0,001$. On trouve alors

$$P(A|B) = \frac{10^{-4} \times 0,99}{10^{-4} \times 0,99 + 0,9999 \times 0,001} \simeq 0,09.$$

Autre exemple :

Exemple 2.4.8 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

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

La formule de Bayes donne alors

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

Exemple 2.4.9 Le jeu des 3 portes : jeu populaire télévisé (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 ?

Solution : 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

$$P(A) = P(B) = P(C) = \frac{1}{3}.$$

Par ailleurs,

$$P(E|A) = \frac{1}{2}, P(E|B) = 0, P(E|C) = 1$$

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,

$$P(E) = P(E|A)P(A) + P(E|B)P(B) + P(E|C)P(C) = \frac{1}{2} \times \frac{1}{3} + \frac{1}{3} = \frac{1}{2}.$$

On en déduit que

$$P(A|E) = \frac{P(A \cap E)}{P(E)} = \frac{P(E|A)P(A)}{P(E)} = \frac{1}{3}.$$

Ainsi $P(C|E) = \frac{2}{3}$. Il faut que le joueur réviser son choix.

EXERCICE 2.4.10 L'inspecteur chargé d'une enquête criminelle est à un certain stade convaincu à 60% de la culpabilité d'un suspect donné. On découvre alors une nouvelle pièce à conviction permettant d'affirmer que le criminel cherché est gaucher. Or 7% des individus dans la population sont gauchers. Comment l'inspecteur doit-il réapprécier la culpabilité du suspect, s'il se trouve que le suspect est gaucher ?

2.4.2 Indépendance

La notion d'indépendance est absolument fondamentale en probabilités et on verra 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, on en déduit les définitions suivantes.

Si B est un événement de probabilité strictement positive, A sera dit **indépendant** de B si

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

On remarque que cette formule se symétrise et la notion d'indépendance se définit finalement comme suit.

Définition 2.4.11 Deux événements aléatoires A et B de probabilité strictement positive sont **indépendants** si et seulement si

$$P(A \cap B) = P(A)P(B). \quad (2.4.34)$$

La probabilité de voir A réalisé ne dépend pas de la réalisation ou non de B .

Remarque 2.4.12 1) Cette notion est une notion liée au choix de la probabilité P et n'est pas une notion ensembliste. ¹

2) Si $P(A) > 0$ et $P(B) > 0$, alors

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

Exemple 2.4.13

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

¹Cela n'a en particulier rien à voir avec le fait que A et B soient disjoints ou non.

2. On tire une carte au hasard dans un jeu de 52 cartes. $A = \{ \text{la carte est un valet} \}$; $B = \{ \text{la carte est un pique} \}$. Il est facile de voir que $P(A) = \frac{4}{52}$, $P(B) = \frac{13}{52}$, et $P(A \cap B) = P(\text{la carte est le valet de pique}) = \frac{1}{52}$. Ainsi, les événements A et B sont indépendants.

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

Proposition 2.4.14 *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 .*

On généralise cette notion à une suite finie ou infinie d'événements de la manière suivante.

Définition 2.4.15 Une suite $(A_n)_n$ d'événements de $(\Omega, \mathcal{A}, P)$ est dite indépendante si

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

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 $P(A \cap B \cap C) = P(A)P(B)P(C)$: par exemple, prenons un lancé de dé avec $A = \{1, 2, 3\}$, $B = \{2, 4, 6\}$ et $C = \{1, 2, 4, 5\}$. On a $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{2}{3}$. On a bien $P(A \cap B \cap C) = P(A)P(B)P(C)$ mais $P(A \cap B) \neq P(A)P(B)$.

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 $A = \{ \text{Face au premier lancé} \}$, $B = \{ \text{Face au deuxième lancé} \}$ 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

Considérons une suite d'espaces de probabilité $(\Omega_n, \mathcal{A}_n, P_n)$. On a vu que ces espaces modélisent des expériences aléatoires. On souhaiterait construire un espace de probabilité rendant toutes ces expériences indépendantes les unes des autres.

Si on a uniquement deux espaces $(\Omega_1, \mathcal{A}_1, P_1)$ et $(\Omega_2, \mathcal{A}_2, P_2)$, on prend $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$, cette notation désignant la tribu produit de $\mathcal{A}_1$ et $\mathcal{A}_2$ 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) et $P(A_1 \times A_2) = P_1(A_1)P_2(A_2)$ caractérise une probabilité sur $\mathcal{A}$ que l'on appelle probabilité produit, notée $P_1 \otimes P_2$.

On peut 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

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

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

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

pour tous $k = 1, 2, \dots$ et $A_i \in \mathcal{A}_i$, et rende ainsi les expériences aléatoires indépendantes. En particulier, en prenant tous les espaces coordonnés égaux, cela nous permettra de modéliser la même expérience répétée une infinité (dénombrable) de fois de manière indépendante.

EXERCICE 2.4.16 Un système comprenant n composants est appelé système en parallèle s'il fonctionne dès qu'au moins l'un de ces composants fonctionne. Si le i ème composant d'un tel système fonctionne indépendamment de tous les autres et avec une probabilité $p_i, i = 1, \dots, n$, quelle est la probabilité que le système fonctionne ?

Même question si le système est un système où les composants sont placés en série.

EXERCICE 2.4.17 *Problème de la ruine du joueur.* Deux joueurs A et B misent sur les résultats successifs du jet répété d'une pièce. A chaque jet, A reçoit une unité de la part de B si Pile est sorti tandis qu'il paie une unité à B dans le cas contraire. Ils poursuivent le jeu tant qu'aucun des deux n'est ruiné. On suppose que les jets sont indépendants et que le côté Pile de la pièce apparaît avec une probabilité p . Soient encore i et $N - i$ les fortunes initiales de A et B respectivement. Quelle est la probabilité que A gagne ? Quelle est la probabilité qu'il y ait un gagnant ?

Nous allons maintenant voir un théorème fameux, dans lequel intervient fondamentalement la notion d'indépendance.

2.4.3 Le Lemme de Borel-Cantelli

Considérons une suite $(A_n)_n$ d'événements aléatoires. On définit l'ensemble $\limsup_n A_n$ comme étant l'ensemble

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

Remarquons que

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

De même,

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

Théorème 2.4.18 Soit $(A_n)_n$ une suite d'événements de $\mathcal{A}$.

- Si la série $\sum_n P(A_n) < +\infty$, alors $P(\limsup_n A_n) = 0$, c'est à dire que P -presque sûrement, un nombre fini au plus de A_n sont réalisés.
- Si de plus la suite $(A_n)_n$ est indépendante, alors

$$\sum_n P(A_n) = +\infty \implies P(\limsup_n A_n) = 1. \quad (2.4.35)$$

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 $P(A) \in]0, 1[$.

La première partie de ce lemme est un outil précieux pour démontrer qu'une propriété est vraie P -presque sûrement (on en verra un exemple dans la preuve donnée de la loi forte des grands nombres au paragraphe 4.2). La deuxième partie caractérise entièrement, dans le cas indépendant, le fait que $P(\limsup_n A_n)$ vaut 0 ou 1 suivant la convergence ou la divergence de la série.

Preuve. La preuve du lemme de Borel-Cantelli n'est pas très compliquée.

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

où $\lim \downarrow$ désigne la limite d'une suite décroissante.

Comme la série $\sum_n P(A_n)$ est convergente, le reste de cette série tend vers 0 et donc $P(\limsup_n A_n) = 0$.

Pour la deuxième partie, écrivons pour m entier,

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

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

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

et l'on conclut finalement que pour tout p , $P(\cup_{i=p}^\infty A_i) = 1$, ce qui implique finalement que $P(\limsup_n A_n)$ vaut 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 $n \geq 1$, on a $P(A_n) = P(A_1) > 0$, d'où $\sum_n P(A_n) = +\infty$. Il résulte du lemme (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 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".

EXERCICE 2.4.19 Il n'existe pas de probabilité P sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ telle que $P(k\mathbb{N}) = \frac{1}{k}$ pour tout entier k , où $k\mathbb{N}$ désigne l'ensemble des entiers multiples de k .

EXERCICE 2.4.20 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 $P(\limsup A_k)$ vaut 0 ou 1 selon que $p < \frac{1}{2}$ ou que $p \geq \frac{1}{2}$. Comment interprétez-vous ce résultat ?

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 pour vous va être de comprendre que ce n'est pas la variable aléatoire, en tant que fonction précise de l'alea, 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 on a $u_n \geq 0$ pour tout n , la suite S_n est croissante, donc elle tend toujours vers une limite S . 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, et on note S et S' leur limites respectives (dans $\mathbb{R}_+$). Pour tout n il existe un entier $m(n)$ tel que $v(i) \leq m(n)$ dès que $i \leq n$. Comme $u_i \geq 0$, on a donc clairement $S'_n \leq S_{m(n)} \leq S$, donc en passant à la limite on obtient $S' \leq S$. On montre de même que $S \leq S'$, et donc $S = S'$.

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

S7 Si $u_n \geq 0$, on peut "sommer par paquets". Cela signifie la chose suivante : soit $(A_i)_{i \in I}$ une partition finie ou dénombrable de $\mathbb{N}^*$. Pour chaque $i \in I$ on pose $v_i = \sum_{n \in A_i} u_n$: si A_i est fini, c'est une somme ordinaire ; sinon, v_i est elle-même la somme d'une série à termes positifs. On a alors la propriété que $\sum_n u_n = \sum_{i \in I} v_i$. Cette dernière somme est de nouveau la somme d'une série à termes positifs si $I = \mathbb{N}^*$. La démonstration de ce résultat est tout-à-fait analogue à celle de (S5) ci-dessus.

S8 Si la série $\sum_n u_n$ est absolument convergente, on a la même propriété (S7) ci-dessus.

S9 Théorème de Fubini pour les séries : soit $(a_{mn})_{m,n \in \mathbb{N}}$ une série double. On suppose 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.

On suppose donnée une probabilité P sur Ω (muni de la tribu de ses parties), caractérisée par les $p_\omega = P(\{\omega\})$. 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, on sait que cette probabilité est caractérisée par les nombres

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

Ainsi,

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

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

Exemple 3.2.2 1) 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 $p(x_i)$ en ordonnées.

EXERCICE 3.2.3 *Représenter le diagramme "en bâtons" de la loi de probabilité de la somme des nombres obtenus lors du jet de deux dés équilibrés.*

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

On suppose 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, on obtient

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

où $f_n(\{\omega\})$ est la fréquence de réalisation du singleton $\{\omega\}$ au cours des n expériences. 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 certainement 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 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

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

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 $E(X)$ de la série ne dépend pas de la manière dont les ω sont ordonnés).

Remarque : $E(X)$ 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 allusion aux problèmes de jeux et d'espérance de gain. Cette terminologie a été introduite par Pascal et est très imagée.

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

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

En particulier, on remarque que l'espérance de X ne dépend que de la loi de X .

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

$$E(X) = \sum_{\omega \in \Omega} p_\omega X(\omega) = \sum_{x_i \in F} \left(\sum_{\omega; X(\omega)=x_i} p_\omega x_i \right) = \sum_{x_i \in F} x_i P(\{\omega; X(\omega) = x_i\}) = \sum_{x_i \in F} x_i p_i^X.$$

□

Analogie avec une notion de mécanique : Le concept d'espérance est à rapprocher de la notion de *centre de gravité* d'un groupe de masses au sens de la mécanique. Considérons en effet une variable X de loi de probabilité $\{(x_i, P(x_i)), i \geq 1\}$. On montre que si les masses $P(x_i), i \geq 1$ 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 $E(X)$. En effet, il suffit d'établir que la somme des moments des forces gravitationnelles par rapport au point d'abscisse $E(X)$ est nulle. En d'autres termes, il suffit de montrer que

$$0 = \sum_i (x_i - E(X))P(x_i),$$

ce qui est immédiat.

Exemple 3.3.3 On choisit au hasard un nombre compris entre 1 et 10. On doit 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 :

- Votre i ème question est du type "Est-ce i ?", i étant égal à 1, 2, ..., 10 :
Soit A_k l'événement : "le nombre $k \in \{1, \dots, 10\}$ a été choisi". Alors

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

et

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

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

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

3.3.2 Propriétés de l'espérance des variables aléatoires discrètes

Définition 3.3.4 On note L^1 l'ensemble de toutes les variables aléatoires X qui admettent une espérance, c'est à dire telles que $\sum_{\omega \in \Omega} p_\omega |X(\omega)| = E(|X|) < +\infty$. On dit dans ce cas que X est intégrable. L'ensemble L^1 dépend de Ω et de la probabilité P .

Les propriétés suivantes sont immédiates :

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

$$E(aX + bY) = aE(X) + bE(Y). \quad (3.3.4)$$

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

$$|E(X)| \leq E(|X|). \quad (3.3.5)$$

- *L'espérance est positive :*

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

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

$$E(X) \leq E(Y). \quad (3.3.7)$$

- 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 :

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

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

EXERCICE 3.3.6 Espérance d'un nombre moyen de rencontres. Chacun des N hommes d'une assemblée jette son chapeau au milieu de la pièce. On mélange les chapeaux et chacun en ramasse un au hasard. Quel est le nombre moyen de rencontres, à savoir le nombre moyen d'hommes ayant récupéré leur chapeau.

EXERCICE 3.3.7 Gestion de stock. Un marchand de journaux reçoit chaque jour la visite d'un nombre aléatoire de clients représenté pour un jour fixé par une variable aléatoire entière positive X . Nous supposons connue la loi de X (grâce, par exemple, à une étude statistique portant sur les ventes des jours précédents). Le marchand se propose d'optimiser le nombre k de journaux qu'il commande à l'éditeur, compte-tenu de ce que

- (i) chaque journal vendu lui rapporte un bénéfice a ,
- (ii) chaque journal invendu lui vaut une perte b ,
- (iii) chaque client insatisfait lui coûte une somme c , dont le montant représente en terme monétaire le risque de perdre ce client définitivement.

Si G_k représente le gain aléatoire du marchand de journaux, déterminer k afin de rendre le profit $E(G_k)$ maximum.

Indication : Pour déterminer la valeur optimale de k , on pourra étudier comment varie le profit quand le stock k augmente d'une unité.

3.3.3 Variance et écart-type

On va maintenant introduire un second ensemble de variables aléatoires, l'ensemble L^2 des variables aléatoires X réelles telles que le carré X^2 soit dans L^1 . On dit dans ce cas que X est de carré intégrable.

Proposition 3.3.8 L^2 est un sous-espace vectoriel de L^1 , et si $X \in L^2$ on a

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

Preuve. Soit X et Y deux variables aléatoires réelles et $a \in \mathbb{R}$. Comme $(aX + Y)^2 \leq 2a^2X^2 + 2Y^2$, si X et Y sont dans L^2 , on déduit 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 (3.3.9) a déjà été vue en (3.3.5). Pour la seconde, on peut se contenter du cas où X est positive. Soit alors $a = E(X)$ et $Y = X - a$. D'après (3.3.4) il vient

$$E(Y^2) = E(X^2) - 2aE(X) + a^2 = E(X^2) - a^2,$$

et $E(Y^2) \geq 0$ par (3.3.6). Donc $a^2 \leq E(X^2)$, ce qui est le résultat cherché. $\square$

Définition 3.3.9 Si $X \in L^2$, sa **variance** est définie par

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

et on la note également σ^2 , ou σ_X^2 . En vertu de (3.3.6) elle est positive, et sa racine carrée positive σ ou σ_X s'appelle **l'écart-type** de X .

En développant le carré $(X - E(X))^2$ comme dans la preuve ci-dessus, on voit aussi que

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

L'écart-type est une grandeur qui mesure la moyenne de l'écart des valeurs de X à sa moyenne, d'où son nom.

Exemple 3.3.10 Le loto.

Le joueur coche 6 numéros sur une grille qui en comporte 49. Les 6 numéros gagnants sont déterminés par tirage au sort. Soit N le nombre de numéros gagnants de la grille. Quelle est la loi de N ? En juin 2000, pour une mise de 2 Francs, on recevait le gain $G=g(N)$ suivant :

n numéros gagnants	gain $g(n)$	probabilité
6	2 132 885 F	$7,2 \cdot 10^{-8}$
5	3 575 F	$7,8 \cdot 10^{-5}$
4	94 F	$9,7 \cdot 10^{-4}$
3	11 F	$7,8 \cdot 10^{-2}$

Calculer le gain moyen. Calculer le bénéfice moyen du joueur. Qu'en conclure ? Calculer la valeur de l'écart-type de G . Quelle en est l'interprétation ?

Solution : Le gain moyen est

$$\begin{aligned} E(G) &= \sum_n g(n)P(N = n) \\ &= 11 \times 7.8 \cdot 10^{-2} + 94 \times 9.7 \cdot 10^{-4} + 3575 \times 7.8 \cdot 10^{-5} + 2132885 \times 7.2 \cdot 10^{-8} \\ &= 1,16 \text{ F.} \end{aligned}$$

Vous pourrez vérifier que l'écart-type vaut 572. Ainsi le bénéfice moyen du joueur, qui vaut $E(G) - 2 = -0.84 \text{ F}$, est négatif, et le jeu est défavorable au joueur. La grande valeur de l'écart-type vient de ce que parfois (mais très rarement), le jeu peut rapporter beaucoup.

3.3.4 Un résultat fondamental - Moments d'une variable aléatoire

On a vu en Proposition 3.3.2 que si $E(|X|) < +\infty$, alors $E(X)$ ne dépend que de la loi P_X de X .

Plus généralement, on considère maintenant une fonction f de F dans $\mathbb{R}$. Ainsi $Y = f(X)$ est une variable aléatoire réelle. On peut aussi considérer f comme une variable aléatoire sur l'espace F muni de la probabilité P_X . On a alors le résultat fondamental suivant, qui montre la cohérence de la notion d'espérance.

Théorème 3.3.11 *Avec les hypothèses précédentes, la variable aléatoire $Y = f(X)$ sur (Ω, P) est dans L^1 si et seulement si la variable aléatoire f sur (F, P_X) est dans L^1 . Dans ce cas, les espérances de ces deux variables aléatoires sont égales, et on a en particulier*

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

Preuve. Comme on peut sommer par paquets par (S7) dans une série à termes positifs, on voit comme pour (3.3.3) que les deux expressions de droite de (3.3.11) sont égales si on remplace f par $|f|$; elles sont donc finies simultanément, donc d'après la définition de L^1 , on a $f(X) \in L^1(\Omega, P) \Leftrightarrow f \in L^1(F, P_X)$.

Si ces propriétés sont réalisées, en utilisant cette fois (S8) on voit 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 3.3.12 *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) :*

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

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 $\mathbb{N}$, et donc, comme nous l'avons vu, caractérisée par une suite de nombres compris entre 0 et 1 et de somme 1. Le but de cette partie 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.

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

Définition 3.4.1 La fonction génératrice de X est la fonction définie sur l'intervalle $[0, 1]$ par la formule suivante (rappelons (3.3.11)) :

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

Comme on le voit 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 de l'énoncé sont alors bien connues. 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 $X \in L^1$ il faut et il suffit que G_X soit dérivable à gauche en $x = 1$, et dans ce cas $E(X) = G'_X(1)$.

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

$$\lim_{x \uparrow 1} \sum_n u_n(x) = \sum_n \lim_{x \uparrow 1} u_n(x). \quad (3.4.13)$$

Si $x < 1$ on a

$$\frac{G_X(x) - G_X(1)}{x - 1} = \sum_n p_n \frac{x^n - 1}{x - 1} = \sum_n p_n (1 + x + \dots + x^{n-1}),$$

et les fonctions $u_n(x) = p_n(1+x+\dots+x^{n-1})$ sont croissantes et positives, avec $\lim_{x \uparrow 1} u_n(x) = np_n$. Le résultat découle alors de (3.4.13). $\square$

Plus généralement, la même démonstration prouve que la variable aléatoire $X(X-1)\dots(X-p)$ est dans L^1 (et donc que X admet un moment d'ordre $p+1$), si et seulement si G_X est $p+1$ fois dérivable à gauche en $x=1$, et on a alors

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

Pour se rappeler cette formule, on peut dériver formellement la série (3.4.12) terme à terme, $p+1$ fois, au point $x=1$, ce qui donne

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

et le membre de droite ci-dessus égale le membre de gauche de (3.4.14) lorsque ce dernier existe, d'après (3.3.11). On peut aussi bien dériver $p+1$ fois les deux membres de l'égalité $G_X(x) = E(x^X)$, en échangeant les dérivées et le signe espérance (là encore, c'est une manipulation "formelle").

Il peut être beaucoup plus rapide et simple, pour calculer les moments d'une variable aléatoire (espérance, variance,...), d'utiliser directement la fonction génératrice, comme nous allons le voir dans les exemples ci-dessous.

3.5 Variables aléatoires discrètes usuelles

3.5.1 Variable aléatoire de Bernoulli

On lance une pièce n fois. On associe 1 à Pile et 0 à Face. L'espace des résultats de l'expérience sera donc

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

On suppose que les lancers sont indépendants les uns des autres. On suppose par ailleurs que 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, on prendra $p = \frac{1}{2}$.

• 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

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

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*

$$E(X) = p ; \quad (3.5.15)$$

$$Var(X) = p(1 - p). \quad (3.5.16)$$

Preuve. On a

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

□

Remarque 3.5.2 Si A est un événement aléatoire, l'espérance de la variable de Bernoulli 1_A se calcule aisément : $E(1_A) = 1 \times P(A) + 0 \times P(A^c)$, soit

$$E(1_A) = P(A),$$

ce qui montre que la notion d'espérance généralise la notion de probabilité.

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 la valeurs entières entre 0 et n . Calculons sa loi : Pour tout $k \in \{0, \dots, n\}$, on cherche

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

Cela veut dire que dans les n lancers on a obtenu 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 on fixe 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, on obtient que

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

Définition 3.5.3 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 $k \in \{0, \dots, n\}$,

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

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

$$G_X(x) = (1 - p + px)^n \quad (3.5.17)$$

$$E(X) = np \quad (3.5.18)$$

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

Preuve.

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

En dérivant G_X et en considérant la dérivée de G_X en $x = 1$, on obtient que

$$E(X) = \sum_{k=0}^n k \binom{n}{k} p^k (1 - p)^{n-k} = G'_X(1) = pn.$$

Si on dérive deux fois en 1 ce polynôme G_X , on obtient

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

et par suite, puisque $E(X) = np$, on en déduit que

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

□

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 $1 \leq k \leq n$,

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

car on a supposé que nos lancers étaient indépendants (voir Section 2.6.2). De même,

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

Si l'on fait (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

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

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 à valeurs dans $\mathbb{N}^*$ telle que $\forall k \in \mathbb{N}^*$,

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

Proposition 3.5.6 Soit X une variable de loi la loi géométrique de paramètre p sur $\mathbb{N}^*$.

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

$$G_X(x) = \frac{px}{1 - (1-p)x}. \quad (3.5.22)$$

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 nécessaire est $1/p$. Il faut donc s'attendre à lancer 6 fois un dé équilibré avant d'obtenir le premier 1, en moyenne.

Preuve.

$$E(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} < +\infty. \quad (3.5.23)$$

Calculons la fonction génératrice

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

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

On en déduit que $E(X) = \frac{1}{p}$. □

3.5.4 Variable aléatoire de Poisson

Définition 3.5.7 On dira que X est une **variable aléatoire de Poisson** de paramètre $\theta > 0$ si X prend ses valeurs dans $\mathbb{N}$ et sa loi est caractérisée pour tout $k \in \mathbb{N}$ par

$$P(X = k) = e^{-\theta} \frac{\theta^k}{k!}. \quad (3.5.24)$$

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 . On a

$$E(X) = \theta, \quad (3.5.25)$$

$$Var(X) = \theta, \quad (3.5.26)$$

$$G_X(x) = e^{\theta(x-1)}. \quad (3.5.27)$$

Preuve. Soit X une variable aléatoire de Poisson de paramètre θ . On a

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

Par ailleurs,

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

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

Il est facile de montrer que $G_X(x) = e^{\theta(x-1)}$. En dérivant, on retrouve $E(X) = \theta$. En dérivant deux fois, (3.4.14) donne $E(X(X-1)) = \theta^2$, donc on retrouve la valeur de $Var(X)$. □

La loi de Poisson comme limite de lois binomiales.

Soit

$$\begin{aligned} p_j(a_n, n) &= \binom{n}{j} (a_n)^j (1 - a_n)^{n-j} \quad \text{si } j \leq n \\ &= 0 \quad \text{si } j \geq n+1, \end{aligned} \quad (3.5.28)$$

avec $a_n \in [0, 1]$. Ainsi, $(p_j(a_n, n))_{j \in \mathbb{N}}$ est l'extension naturelle sur $E = \mathbb{N}$ de la loi binomiale de paramètre a_n et de taille n . Supposons alors que $na_n \rightarrow \theta \in \mathbb{R}_+^*$ lorsque $n \rightarrow +\infty$. En développant les combinaisons $\binom{n}{j}$, il est facile de vérifier que

$$p_j(a_n, n) \rightarrow p_j = e^{-\theta} \frac{\theta^j}{j!}, \quad \forall j \in \mathbb{N}.$$

Ce résultat est très utile pour les calculs numériques : si $n a_n$ est "petit", on peut remplacer la loi binomiale par la loi de Poisson, ce qui conduit à des calculs plus simples. La loi de Poisson modélise la probabilité d'un événement rare ($a_n \sim \frac{\theta}{n}$ petit) dans une suite infinie d'événements (n grand).

On verra plus loin une autre approximation de la loi binomiale $B(a_n, n)$ lorsque n est grand et a_n n'est pas petit.

On peut citer beaucoup d'exemples de variables aléatoires qui obéissent à la loi de Poisson, (parce qu'on approxime ainsi une loi binomiale) :

- Le nombre de centenaires dans une communauté humaine
- Le nombres de fautes de frappe par page de ce polycopié
- Le nombre de paquets de biscuits pour chien vendus dans un magasin donné en l'espace d'un jour
- Le nombre de clients entrant au bureau de poste de l'Ecole en l'espace d'un jour
- Le nombre de bactéries dans un volume de liquide fixé

Exemple numérique : Pour effectuer un contrôle de qualité, on procède au tirage au hasard de 20 objets dans une fabrication industrielle. Supposons savoir que 5% des objets de cette fabrication sont défectueux. On suppose que le nombre total d'objets fabriqués est énorme. On fait l'hypothèse que le nombre X d'objets défectueux prélevés suit une loi binomiale de paramètres $n = 20$ et $p = 0,05$; (Voir Section 2.2.3). L'approximation par la loi de Poisson π_1 de paramètre $np = 1$ est ici satisfaisante comme le montre le tableau suivant :

x	$P(N = x)$	$\pi_1(x)$
0	0,358	0,368
1	0,377	0,368
2	0,189	0,184
3	0,060	0,061
4	0,013	0,015
...	...	...

TAB. 3.1 – Approximation de la loi binomiale par la loi de Poisson $n = 20$, $p = 1/20$, $\theta = 1$

3.6 Lois conditionnelles et variables aléatoires indépendantes

3.6.1 Lois conditionnelles

On rappelle ici que les notions de probabilités conditionnelles et d'événements aléatoires indépendants ont été introduites au Chapitre 2.6.

Dans ce paragraphe, on considère deux variables aléatoires X et Y définies sur le même espace Ω fini ou dénombrable, muni de la probabilité P . On suppose X et Y à valeurs respectivement dans F et G , que l'on peut supposer eux-mêmes finis ou dénombrables. On pose $p_i^X = P(X = x_i)$ pour $x_i \in F$, et $p_j^Y = P(Y = y_j)$ pour $y_j \in G$.

La connaissance des deux lois P^X et P^Y ne donne aucune information sur les liens qui peuvent unir les comportements aléatoires de X et de Y . Il est plus intéressant de considérer le couple $Z = (X, Y)$ comme une variable aléatoire discrète à valeurs dans le produit cartésien $F \times G$, et on note $P_Z = (p_k^Z, z_k \in F \times G)$ sa loi.

Pour $z_k = (x_i, y_j) \in F \times G$

$$p_k^Z = P(Z = z_k) = P(X = x_i \text{ et } Y = y_j).$$

Les lois P^X et P^Y s'appellent **les lois marginales du couple de variables aléatoires X et Y** .

En règle générale, les lois marginales ne suffisent pas à connaître la loi de Z . Considérons par exemple l'expérience suivante. On lance à la fois 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. (Voir l'exemple introductif de la section 2.4.1).

Il est donc naturel de s'intéresser, pour chaque valeur fixée x_i de X , à la loi de Y ayant a priori l'information que $X = x_i$.

Dans toute la suite de ce paragraphe, on pourra échanger les rôles de X et de Y .

Définition 3.6.1 Soit $x_i \in F$ tel que $P(X = x_i) > 0$. On appelle **loi conditionnelle de Y sachant $X = x_i$** la probabilité définie sur G par

$$p_j^{Y|X=x_i} = P(Y = y_j | X = x_i) \quad \forall y_j \in G. \quad (3.6.29)$$

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 .

On a en fait les relations suivantes.

Proposition 3.6.2 *Il est équivalent de connaître les $(p_k^Z : z_k \in F \times G)$ d'une part, $(z_k = (x_i, y_j))$, les $(p_i^X : x_i \in F)$ et les $(p_j^{Y|X=x_i} : y_j \in G)$ pour les $x_i \in F$ tels que $p_i^X > 0$ d'autre part, via les formules :*

$$P(X = x_i) = \sum_{y_j \in G} P(Z = (x_i, y_j)), \quad (3.6.30)$$

$$p_j^{Y|X=x_i} = \frac{P(Z = (x_i, y_j))}{P(X = x_i)} \quad \text{si } P(X = x_i) > 0, \quad (3.6.31)$$

$$P(Z = (x_i, y_j)) = \begin{cases} P(X = x_i) P(Y = y_j | X = x_i) & \text{si } P(X = x_i) > 0 \\ 0 & \text{sinon} \end{cases} \quad (3.6.32)$$

Preuve. Il suffit de montrer les trois formules de l'énoncé. D'abord, 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 (3.6.30) découle de la σ -additivité (2.3.19). (3.6.31) vient de la formule (2.4.30). Enfin (3.6.32) découle de (3.6.31) si $p_i^X > 0$, tandis que si $p_i^X = P(X = x_i) = 0$ on a *a fortiori* $P(X = x_i, Y = y_j) = P(Z = (x_i, y_j)) = 0$ d'après (2.2.11). $\square$

Puisque pour i fixé tel que $p_i^X > 0$, $\{(y_j, p_j^{Y|X=x_i}), y_j \in G\}$ définit une probabilité, on peut de même que pour les lois marginales définir les notions d'espérance, de variance ou de moments pour cette loi conditionnelle.

Par exemple, on peut définir la notion d'espérance conditionnellement à $\{X = x_i\}$, dès que Y est intégrable.

Définition 3.6.3 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

$$E(Y|X = x_i) = \sum_j y_j p_j^{Y|X=x_i} = \sum_j y_j P(Y = y_j | X = x_i). \quad (3.6.33)$$

Remarque 3.6.4 La série définie en (3.6.33) est bien convergente. En effet, on peut utiliser l'observation suivante

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

et le théorème de Fubini (voir (S9)).

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

Définition 3.6.5 On appelle **espérance conditionnelle de Y sachant X** la variable aléatoire

$$E(Y|X) = \psi(X), \quad \text{avec} \quad \psi(x) = E(Y|X = x), \quad (3.6.34)$$

pour x tel que $P(X = x) > 0$, 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 .

3.6.2 Propriétés de l'espérance conditionnelle

L'espérance conditionnelle possède la propriété fondamentale suivante.

Théorème 3.6.6 Si Y est intégrable, alors $\psi(X) = E(Y|X)$ est intégrable, et

$$E(\psi(X)) = E(Y).$$

Preuve.

Supposons que X soit une variable aléatoire discrète à valeurs $(x_i)_i$ et Y une variable aléatoire discrète à valeurs $(y_j)_j$

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

On a utilisé 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 $E(Y)$ en conditionnant par une variable auxiliaire X :

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

Il généralise la formule de Bayes (2.4.5), qui correspond ici à $Y = \mathbf{1}_A$, et $B_i = \{X = x_i\}$. On l'écrit souvent sous forme

$$E(E(Y|X)) = E(Y). \quad (3.6.35)$$

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

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

De plus,

$$E(Y g(X) | X) = g(X) E(Y | X) \quad (3.6.36)$$

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

Solution : On a 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

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

soit $E(K | N) = pN$. D'après le théorème 3.6.6,

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

puisque $\lambda = E(N)$.

3.6.3 Variables aléatoires indépendantes

Remarque : La formule (3.6.30) de la proposition précédente montre que l'on peut calculer les lois marginales P_X et P_Y de X et de Y à partir de la loi P_Z de Z .

En revanche, **attention**, la connaissance des lois marginales ne suffit pas, en général, à retrouver la loi de $Z = (X, Y)$, comme le prouve l'exemple suivant : Z vaut $(1, 1)$ et $(-1, -1)$ avec probabilité $\frac{p}{2}$ et $(1, -1)$ et $(-1, 1)$ avec probabilité $\frac{1-p}{2}$, où $p \neq \frac{1}{2}$.

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

Définition 3.6.8 Les variables aléatoires X et Y sont dites **indépendantes** si pour toutes parties $A \subset F$, $B \subset G$ on a

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B). \quad (3.6.37)$$

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

On a la propriété fondamentale suivante.

Proposition 3.6.9 *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 $P(Z = (x_i, y_j)) = P(X = x_i)P(Y = y_j) = p_i^X p_j^Y$ 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 3.6.10 (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 3.6.8 de l'indépendance est symétrique en X et Y , on peut 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 (3.6.37). Inversement, supposons (ii). En sommant par paquets dans la série à termes positifs, on obtient pour $A \subset F$ et $B \subset G$:

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

donc on a (i). Enfin, l'équivalence (ii) $\Leftrightarrow$ (iii) provient de (3.6.31) et (3.6.32). $\square$

Proposition 3.6.11 *Supposons les variables aléatoires X et Y indépendantes, et soit 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 dans L^1 , et on a*

$$E(f(X)g(Y)) = E(f(X)) E(g(Y)). \quad (3.6.38)$$

Preuve. Exactement comme dans la démonstration précédente, on peut écrire

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

qui est fini par hypothèse : par suite $f(X)g(Y)$ appartient à L^1 . En utilisant alors (S8), la même démonstration montre qu'on a les égalités ci-dessus en enlevant les valeurs absolues : cela donne (3.6.38). $\square$

Exemple 3.6.12 On considère n variables aléatoires de Bernoulli $(X_i)_{i=1}^n$ indépendantes et de paramètre $p \in]0, 1[$.

La probabilité d'obtenir la suite $x_1 \cdots x_n$, où $x_i \in \{0, 1\}$, est égale à

$$P(X_i = x_i, 1 \leq i \leq n) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}, \quad (3.6.39)$$

avec $x_i \in \{0, 1\}$, et l'on retrouve les résultats donnés au chapitre 2 (modèles d'urnes).

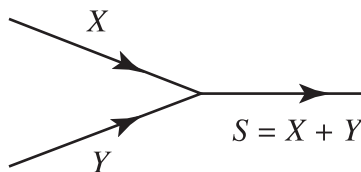
En particulier, on trouve en utilisant (3.6.39) que la somme $S = X_1 + \cdots + X_n$ suit une loi binomiale $\mathcal{B}(n, p)$.

Ainsi donc, la loi binomiale $\mathcal{B}(n, p)$ est la loi de la somme de n variables aléatoires indépendantes de loi de Bernoulli de paramètre p .

Comment simuler n variables aléatoires de Bernoulli indépendantes et de paramètre $p \in]0, 1[$? On va utiliser un générateur de nombres au hasard, algorithme qui fournit une suite de nombres compris entre 0 et 1, qui ont 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 cours de Terminale ou Chapitre suivant pour la définition d'une loi uniforme). On pose alors :

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

Exemple 3.6.13 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'autoroute après le raccordement. X et Y sont modélisées par des variables aléatoires de loi de Poisson de paramètres respectifs $\lambda > 0$ et $\mu > 0$. Ainsi, par exemple, $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k \in \mathbb{N}$. Les variables aléatoires X et Y sont supposées indépendantes. On cherche à déterminer la loi de S et l'espérance conditionnelle de X sachant S .

Solution :

a) Par hypothèse, $P(X = k, Y = \ell) = p_X(k) p_Y(\ell) = \frac{\lambda^k \mu^\ell e^{-(\lambda+\mu)}}{k! \ell!}$ pour $k, \ell \geq 0$. Mais cette probabilité est égale à $P(X = k, S = k + \ell)$, si bien que, avec $n = k + \ell$,

$$p_{X,S}(k, n) = (\lambda/\mu)^k \mu^n e^{-(\lambda+\mu)} \frac{1}{k!(n-k)!}$$

pour $0 \leq k \leq n$, et $p_{X,S}(k, n) = 0$ sinon. D'après la formule des marginales, il vient en sommant sur k

$$\begin{aligned} p_S(n) &= \sum_k p_{X,S}(k, n) = \sum_{k=0}^n (\lambda/\mu)^k \binom{n}{k} \times \frac{1}{n!} \mu^n e^{-(\lambda+\mu)} \\ &= \frac{1}{n!} (\lambda + \mu)^n e^{-(\lambda+\mu)} \end{aligned}$$

pour $n \geq 0$. Donc S suit la loi de Poisson de paramètre $\lambda + \mu$.

b) La loi conditionnelle de X sachant $S = n$ ($n \geq 0$) est par définition

$$p_k^{X|S=n} = \frac{p_{X,S}(k, n)}{p_S(n)} = \binom{n}{k} \left(\frac{\lambda}{\lambda + \mu} \right)^k \left(\frac{\mu}{\lambda + \mu} \right)^{n-k}$$

pour $k \in [0, n]$ entier, et $p_k^{X|S=n} = 0$ sinon. Cette loi est la loi binomiale $B\left(n, \frac{\lambda}{\lambda + \mu}\right)$.

On peut alors calculer $E(X | S = n)$ en utilisant (3.6.33), 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, $E(X | S = n) = n \frac{\lambda}{\lambda + \mu}$, ou encore

$$E(X | S) = S \frac{\lambda}{\lambda + \mu}.$$

EXERCICE 3.6.14 On admet que le nombre de clients du bureau de poste de l'Ecole pendant une journée est une variable aléatoire poissonnienne de paramètre λ . On note par p la probabilité qu'une personne entrant dans le bureau de poste soit une femme. Montrer que dans ce cas, le nombre de femmes et celui des hommes parmi les clients quotidiens sont des variables aléatoires indépendantes, suivant des lois de Poisson de paramètres respectifs λp et $\lambda(1 - p)$.

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 F et G soient égaux à l'ensemble $\mathbb{Z}$ des entiers relatifs. Alors, si X et Y sont deux variables aléatoires et $Z = (X, Y)$,*

$$P(X + Y = i) = \sum_{j \in \mathbb{Z}} P(Z = (j, i - j)) = \sum_{j \in \mathbb{Z}} P(Z = (i - j, j)). \quad (3.6.40)$$

En particulier si X et Y sont indépendantes, on a

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

Preuve. (3.6.41) découle immédiatement de (3.6.40) et de (ii) de la proposition 3.6.9. Pour (3.6.40), 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 $j \in \mathbb{Z}$. $\square$

Proposition 3.6.16 *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 . On a alors pour tout $s \in [0, 1]$*

$$G_U(s) = G_X(s)G_Y(s). \quad (3.6.42)$$

Preuve. Il suffit de remarquer que $G_U(s) = E(s^U) = E(s^{X+Y})$ et $G_X(s) = E(s^X)$ et $G_Y(s) = E(s^Y)$ pour $s \in [0, 1]$, et d'appliquer (3.6.38). $\square$

Exemple 3.6.17 1) Soit X et Y des variables aléatoires indépendantes de lois binomiales respectives $\mathcal{B}(n, p)$ 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, on en déduit que $X + Y$ suit **la loi binomiale** $\mathcal{B}(n + m, p)$, ce que l'on savait au vu de l'exemple 3.6.12.

2) Soit X et Y des variables aléatoires indépendantes de lois de Poisson de paramètres respectifs θ et ζ . D'après (3.5.27), $U = X + Y$ 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 $\theta + \zeta$.

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 l'on considère maintenant peuvent être à valeurs dans tout $\mathbb{R}$. On souhaite que la tribu $\mathcal{E}$ définie dans la Proposition 2.3.13 contienne la tribu borélienne 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 l'espace d'état Ω muni de la tribu $\mathcal{A}$ des événements. Une application X de Ω dans $\mathbb{R}$ est une **variable aléatoire** réelle si $X^{-1}(B) \in \mathcal{A}$ pour tout $B \in \mathcal{B}(\mathbb{R})$.

On a 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, on note $\mathcal{R}$ 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 : vu la proposition 2.3.6, on a 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, $A = \{x \in \mathbb{R}^n : f(x) > a\}$ 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 on a

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

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

Comme application de cet énoncé ou de la partie 1) de sa preuve, on a les propriétés suivantes, où X, Y et les $(X_n)_{n \in \mathbb{N}^*}$ sont des variables aléatoires réelles :

Proposition 4.1.3

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

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

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

(pour le voir, on remarque que $\{\sup_i X_i \leq a\} = \cap_i \{X_i \leq a\}$, qui est dans $\mathcal{A}$ par hypothèse, et on applique la partie (1) de la preuve précédente ; même chose pour l'inf) ;

$$X_n(\omega) \rightarrow Z(\omega), \quad \forall \omega \quad \Rightarrow \quad Z \text{ est une variable aléatoire ;} \quad (4.1.4)$$

(On a en effet $Z = \inf_n Y_n$, où $Y_n = \sup_{i \geq n} X_i$, et on applique deux fois (4.1.3)) ;

$$Z = 1_A \quad (\text{indicatrice de } A) \text{ est une variable aléatoire} \Leftrightarrow A \in \mathcal{A} ; \quad (4.1.5)$$

(il suffit de remarquer que $Z^{-1}(B)$ égale $\emptyset, A, A^c$ ou Ω selon que $B \cap \{0, 1\}$ égale $\emptyset, \{1\}, \{0\}$ ou $\{0, 1\}$).

Dans le cas où $n = 1$ (le cas général sera vu plus tard), on peut généraliser la Proposition 4.1.2 à des fonctions f beaucoup plus générales, appelée 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. Pour nous, 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

Maintenant que la notion de variable aléatoire réelle est bien établie, on peut définir rigoureusement la loi P_X de X comme la probabilité introduite en Section 2.3.3, définie sur $\mathbb{R}$ muni de la tribu borélienne $\mathcal{B}(\mathbb{R})$, et vérifiant, rappelons-le, que pour tout $B \in \mathcal{B}(\mathbb{R})$,

$$P_X(B) = P(X \in B).$$

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 donc X une variable aléatoire réelle et P_X sa loi.

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

Le résultat fondamental sur $\mathbb{R}$ et lié à cette propriété de la tribu borélienne, est que la probabilité 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 P_X sa loi. On appelle **fonction de répartition de X** la fonction suivante :

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

Exemple 4.2.2 Si P_X est la mesure de Dirac en 0 (voir Section 2.3.4), c'est-à-dire si $X \equiv 0$, 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. On verra ci-dessous une caractérisation des points où elle est continue.

Proposition 4.2.3 La fonction de répartition F_X caractérise la probabilité P_X et elle vérifie les trois conditions suivantes :

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

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

1) D'après (4.2.6), on a $P_X([x, y]) = F(y) - F(x)$ pour tous $x < y$. Par suite, si $B = \cup_{i=1}^n]x_i, y_i]$, avec $x_i < y_i < x_{i+1}$, on a

$$P_X(B) = \sum_{i=1}^n P_X([x_i, y_i]) = \sum_{i=1}^n F(y_i) - F(x_i),$$

car les intervalles sont disjoints. Puisque $P_X([x, +\infty[) = 1 - F(x)$, on en déduit finalement que F caractérise la restriction de 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 (qu'on admettra) montre que la connaissance de P_X sur cette algèbre suffit à déterminer entièrement 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), on remarque 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, limite que l'on notera $F(x-)$. En remarquant que $] - \infty, y[= \lim_n] - \infty, y_n]$ si $y_n \uparrow y$, on peut facilement obtenir que pour $x < y$,

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

En particulier,

$$P_X(\{x\}) = F(x) - F(x-)$$

est le saut de la fonction F au point x . On a donc

Proposition 4.2.4

$$P_X(\{x\}) = P(X = x) = 0 \Leftrightarrow F \text{ est continue en } x.$$

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 absolument nécessaire d'introduire les tribus en probabilités, malgré la complexité que cela engendre. Si on ne le faisait pas, ce qui reviendrait à prendre (sans le dire) la tribu $\mathcal{P}(\mathbb{R})$, il n'existerait que très peu de probabilités sur $\mathbb{R}$, à savoir les "probabilités discrètes" que l'on a introduites aux chapitres précédents.

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

Etudions 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 P_X de X est caractérisée par la suite $p_n = P_X(\{n\}) = P(X = n)$. La fonction de répartition F de X vaut alors

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

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 P_X de X est caractérisée pour tout $i \in E$ par $q_i = P_X(\{i\}) = P(X = i)$, 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. On retrouve bien (4.2.8) dans le cas où $E = \mathbb{N}$. On voit que 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 tout-à-fait être partout dense dans $\mathbb{R}$, par exemple il peut être égal à l'ensemble des rationnels $\mathbb{Q}$. Si alors $q_i > 0$ pour tout $i \in \mathbb{Q}$, la fonction F est discontinue en tout rationnel, et continue partout ailleurs....

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

En prenant la fonction F nulle sur $\mathbb{R}_-$, $F(x) = x$ sur $[0, 1]$ et $F(x) = 1$ pour $x \geq 1$ qui satisfait bien les hypothèses du théorème 4.2.5, on obtient 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 vos connaissances préalables, et que l'intégrale généralisée $\int_{-\infty}^{+\infty} |f(x)| dx$ existe (donc est finie), ce qui implique aussi que l'intégrale généralisée $\int_{-\infty}^{+\infty} f(x) dx$ existe.

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

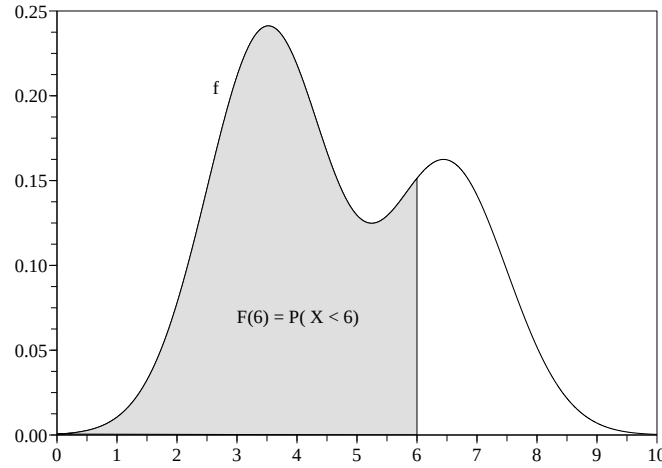
$$\int_{-\infty}^{+\infty} f(x) dx = 1. \quad (4.2.10)$$

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.

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

FIG. 4.1 – Graphe de la densité f .

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 P_X admet la densité f , et donc si pour tout réel x ,

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

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

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

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

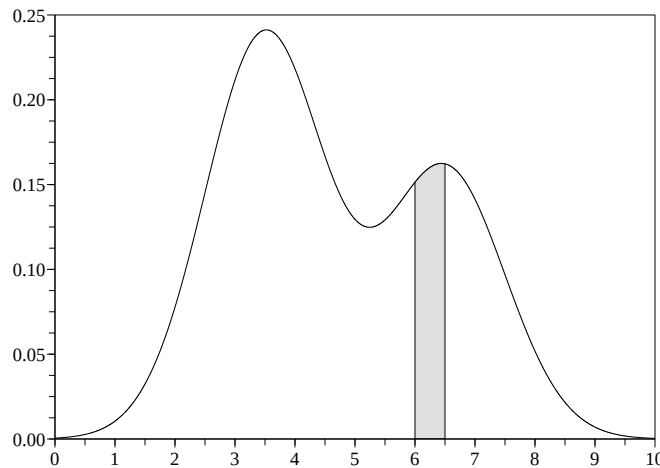
$$F'(x) = f(x).$$

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 FONDAMENTALE : Voici une interprétation “intuitive” de la densité f de P_X . Si Δx est un “petit” accroissement de la quantité x , on a (si du moins f est continue en x) :

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

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

- Remarque 4.2.11** 1) La fonction de répartition est, de manière évidente, entièrement déterminée par la probabilité P_X . Il n'en est pas de même de la densité, lorsqu'elle existe : si en effet on a (4.2.11) et si on pose $g(x) = f(x)$ si $x \notin \mathbb{Q}$ et $g(x) = f(x) + 1$ si $x \in \mathbb{Q}$, alors g est encore une densité de P_X .
- 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" : si on se donne d'une part une fonction f positive intégrable et d'intégrale strictement positive, et une partie finie ou dénombrable F de $\mathbb{R}$ et des $q_i > 0$ indicés par $i \in F$, tels que

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

Alors, la fonction

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

est une fonction de répartition, et la probabilité P_X associée n'admet pas de densité, et n'est pas non plus discrète.

EXERCICE 4.2.12 La durée de fonctionnement d'un ordinateur avant sa première panne est une variable aléatoire positive de densité donnée par

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

Quelle est la probabilité que cette durée de fonctionnement soit comprise entre 50 et 150 heures ?

Quelle est la probabilité que l'ordinateur fonctionne moins de 100 heures ?

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** : dans le passé il y a eu quelques déboires...

Actuellement, tous les générateurs sont basés sur des procédés purement déterministes par une récurrence $x_{n+1} = \phi(x_n)$ 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 donc périodique de période $p \leq m$; mais 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 de ne pas dépasser le dixième de la période). Ces générateurs n'ont donc, en fait, rien d'aléatoires et c'est pourquoi on les qualifie, aussi, de **pseudo-aléatoires**.

Sans être les plus récents, les **générateurs congruentiels linéaires** sont très utilisés. Ils produisent une suite d'entiers

$$x_{n+1} = ax_n + b \quad \text{modulo } m,$$

à partir d'une valeur initiale x_0 . Bien entendu, les entiers m, a et b doivent être soigneusement choisis (on exige au moins que la période p soit maximale, c'est-à-dire égale à m). Exemples de tels générateurs :

- la fonction **rand** de SCILAB utilise les valeurs $m = 2^{31}$, $a = 843314861$ et $b = 453816693$;
- la méthode **Math.random** de JAVA utilise les valeurs $m = 2^{48}$, $a = 25214903917$ et $b = 11$.

Par contre, la fonction **grand** de SCILAB est basée un générateur plus performant et plus moderne : le **Mersenne twister**² de période colossale $2^{19937} - 1$.

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

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

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

A partir d'une variable aléatoire uniforme X sur $[0, 1]$, générée comme ci-dessus, on peut simuler une variable aléatoire réelle Y de loi quelconque, dès lors que l'on connaît sa fonction de répartition F . On définit la fonction "inverse" continue à gauche de F par :

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

La fonction G n'est pas à proprement parler l'inverse (ou 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

$$G(x) \leq y \Leftrightarrow x \geq F(y),$$

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

Proposition 4.2.13 *Si on pose $Y = G(X)$, on obtient une variable aléatoire de fonction de répartition F .*

Preuve. On remarque facilement que si $y \in \mathbb{R}$, on a

$$P(Y \leq y) = P(G(X) \leq y) = P(X \leq F(y)) = F(y)$$

puisque X suit une loi uniforme sur $[0, 1]$. □

Exemple 4.2.14 Simulation d'une variable aléatoire discrète. Soit $p = (p_1, p_2, \dots)$ une probabilité sur un ensemble dénombrable $\{x_1, x_2, \dots\}$ avec $x_1 < x_2 < \dots$. La fonction de répartition associée est $F(x) = \sum_{x_i \leq x} p_i$, qui 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

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

qui définit une variable aléatoire X de loi p .

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 "suffisamment petites" (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).

On appelle **variable aléatoire étagée** toute variable aléatoire X 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

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

Construisons dans une première étape l'espérance des variables aléatoires positives. Si X est une telle variable, on considère une suite X_n de variables aléatoires positives étagées croissant vers X , par exemple

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

Comme $X_n \leq X_{n+1}$, on a $E(X_n) \leq E(X_{n+1})$ par (3.3.7), et on peut poser

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

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 .

Soit enfin X une variable aléatoire de signe quelconque. Elle 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 4.3.1 On dit que la variable aléatoire X est **intégrable** si les valeurs $E(X^+)$ et $E(X^-)$ sont toutes les deux finies. Dans ce cas, l'espérance mathématique de X est le nombre

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

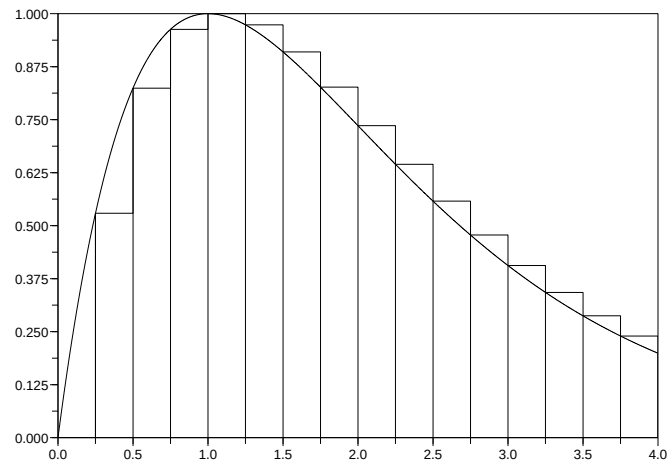


FIG. 4.3 – Découpage “à la Riemann”, en 16 intervalles de même longueur de l’abscisse

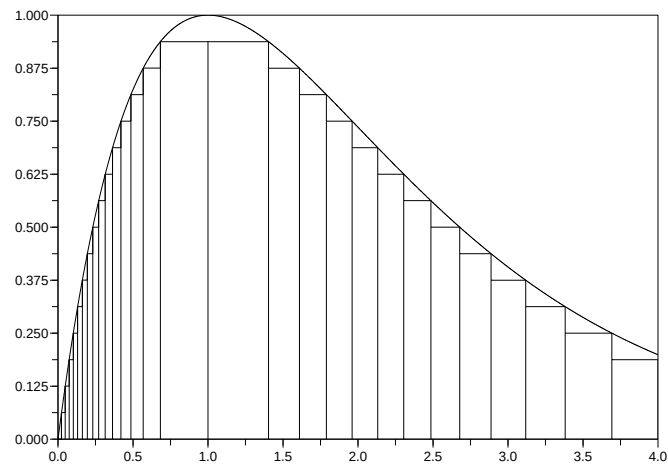


FIG. 4.4 – Découpage “à la Lebesgue”, en 16 intervalles de même longueur de l’ordonnée

On appelle 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.

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

Proposition 4.3.2 *Soit X une variable aléatoire. Alors*

$$X \in L^1 \Leftrightarrow E(|X|) < +\infty.$$

On peut vérifier sans peine que ces définitions coïncident avec celle du paragraphe 3.3 lorsque Ω est fini ou dénombrable. On montre aussi (par passage à la limite) que les propriétés données dans la Proposition 3.3.5 sont vraies dans notre cadre général.

On les rappelle ici :

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

$$E(aX + bY) = aE(X) + bE(Y). \quad (4.3.19)$$

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

$$|E(X)| \leq E(|X|). \quad (4.3.20)$$

- **Positivité :**

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

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

$$E(X) \leq E(Y). \quad (4.3.22)$$

- L'espérance d'une variable constante est égale à cette constante :

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

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

Remarque : Cette notion d'intégrale $\int X(\omega)P(d\omega)$, 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. Gardons à l'esprit que ces notation et 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 , on définit aussi, comme au Chapitre 3, l'espace L^2 des variables aléatoires X telles que le carré X^2 soit dans L^1 .

Définition 4.4.1 On dit qu'une variable aléatoire X est de carré intégrable si la variable aléatoire X^2 est intégrable, c'est-à-dire si son espérance est finie.

Proposition 4.4.2 L^2 est un sous-espace vectoriel de L^1 , et si $X \in L^2$,

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

La preuve est identique à celle de la proposition 3.3.8.

Définition 4.4.3 Si $X \in L^2$, sa **variance** est l'espérance de la variable aléatoire $(X - E(X))^2$, et on la note $Var(X)$ 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 .

On a évidemment encore

$$Var(X) = E(X^2) - E(X)^2. \quad (4.4.24)$$

"La variance est égale à la moyenne du carré 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 $|XY| \leq \frac{1}{2}(X^2 + Y^2)$. On peut alors définir la notion suivante.

Définition 4.4.4 Si X et Y sont dans L^2 , la variable aléatoire $(X - E(X))(Y - E(Y))$ est dans L^1 . On appelle **covariance** de X et Y l'espérance de cette variable aléatoire, et on la note $cov(X, Y)$:

$$cov(X, Y) = E((X - E(X))(Y - E(Y))). \quad (4.4.25)$$

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

$$\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}}. \quad (4.4.26)$$

Notez que, en utilisant la linéarité de l'espérance, on obtient

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

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 $a, b, a', b' \in \mathbb{R}$:

$$E((aX + b)(a'Y + b')) = aa'E(XY) + ab'E(X) + a'bE(Y) + bb',$$

$$E(aX + b) E(a'Y + b') = aa'E(X)E(Y) + ab'E(X) + a'bE(Y) + bb'.$$

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

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

et en particulier

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

On en déduit 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 $E(X)$ et d'écart-type $\sigma_X > 0$, alors la variable aléatoire

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

est d'espérance nulle et d'écart-type 1. On dit 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

On considère deux variables aléatoires de carré intégrable pour lesquelles on suppose que les variances et la covariance sont connues. On souhaite 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 $E((Y - (aX + b))^2)$. Il s'agit de déterminer les constantes a et b telles que $E((Y - (aX + b))^2)$ soit minimum. Or, par linéarité,

$$E((Y - (aX + b))^2) = E(Y^2) - 2aE(Y) - 2bE(XY) + a^2 + 2abE(X) + b^2E(X^2).$$

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

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

On vérifie aisément que ces valeurs donnent bien un minimum pour $E((Y - (aX + b))^2)$, et déterminent ainsi la meilleure approximation linéaire de Y basée sur X au sens du carré moyen de l'erreur.

Cette erreur vaut

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

Le carré moyen de l'erreur vaut alors

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

On constate 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 ci-dessus 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 3.3.11. On considère la variable aléatoire X de Ω dans $\mathbb{R}$ et une fonction mesurable h de $\mathbb{R}$ dans $\mathbb{R}$. On a vu que $Y = h(X)$ est une variable aléatoire sur Ω .

On dira que h est intégrable si h appartient à $L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$.

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

$$E(h(X)) = \int_{\Omega} h(X(\omega))P(d\omega) = \int_{\mathbb{R}} h(x)P_X(dx). \quad (4.5.31)$$

Preuve. (4.5.31) est évident lorsque h ne prend qu'un nombre fini de valeurs, grâce à (4.3.15) et au fait que $P(X^{-1}(B)) = P_X(B)$. Si h est positive, on l'approche par des fonctions étagées h_n comme en (4.3.16), et on en déduit encore (4.5.31), les trois membres étant simultanément finis ou infinis. Le résultat général s'en déduit par différence. $\square$

Commentaire : Si on connaît la loi de X , on pourra donc calculer $E(h(X))$ pour toute fonction h intégrable.

Par exemple, si X est une variable aléatoire discrète, chargeant par exemples les entiers k avec probabilité p_k , on retrouve que

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

4.5.2 Calculs d'espérances dans le cas avec densité

Nous allons dans ce paragraphe voir comment on calcule l'espérance d'une variable aléatoire réelle, ou plus généralement d'une fonction d'une variable aléatoire réelle, lorsque cette dernière admet une densité. **On rappelle que X admet la densité f si la fonction de répartition de P_X est dérivable de dérivée f .**

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

$$\int_{-\infty}^{+\infty} |g(x)|f(x)dx < \infty, \quad (4.5.32)$$

et dans ce cas on a

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x)dx. \quad (4.5.33)$$

En particulier,

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx \quad (4.5.34)$$

et

$$Var(X) = \int_{-\infty}^{+\infty} (x^2 - E(X))^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left(\int_{-\infty}^{+\infty} xf(x)dx \right)^2. \quad (4.5.35)$$

Nous n'avons pas les éléments permettant de montrer 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 $i \in \mathbf{Z}$. Ainsi, $X_n(\omega) \rightarrow X(\omega)$ pour tout ω , et par continuité de g on a $g(X_n) \rightarrow g(X)$. De plus, comme X_n est une variable aléatoire discrète, on a en utilisant (4.3.15) et (4.2.12) :

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

et le dernier terme ci-dessus tend vers le second membre de (4.5.33) lorsque $n \rightarrow \infty$, par approximation de Riemann.

4.6 Exemples fondamentaux de variables à densité

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

On généralise ici la notion de variable uniforme sur $[0, 1]$. On se donne ici deux réels $a < b$, et la probabilité P_X admet la densité

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{sinon.} \end{cases} \quad (4.6.36)$$

En vertu de la Proposition 4.2.10, on pourrait 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 $[a, b]$ correspond au fait que si on choisit un point selon la probabilité uniforme, on a "autant de chances" de tomber au voisinage de chaque point de l'intervalle $[a, b]$, ce qui explique le nom "uniforme". Remarquer aussi que $P_X(\{x\}) = 0$ pour tout x (comme pour toutes les probabilités avec densité) : on a 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*

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2} \quad ; \quad Var(X) = \frac{(b-a)^2}{12}. \quad (4.6.37)$$

(Si on a la même chance de tirer "un quelconque point" de l'intervalle $[a, b]$, il est naturel d'obtenir que l'espérance est 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 7h25 et 7h30. La probabilité d'attendre moins de 5 minutes est donc

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

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

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

4.6.2 Variable aléatoire exponentielle

Définition 4.6.3 On dit que X suit la loi exponentielle de paramètre $\lambda > 0$ si P_X admet la loi de densité

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ \lambda e^{-\lambda x} & \text{sinon.} \end{cases} \quad (4.6.38)$$

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

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-\lambda x} & \text{sinon.} \end{cases} \quad (4.6.39)$$

De plus,

$$E(X) = \frac{1}{\lambda}, \quad Var(X) = \frac{1}{\lambda^2}. \quad (4.6.40)$$

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

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.

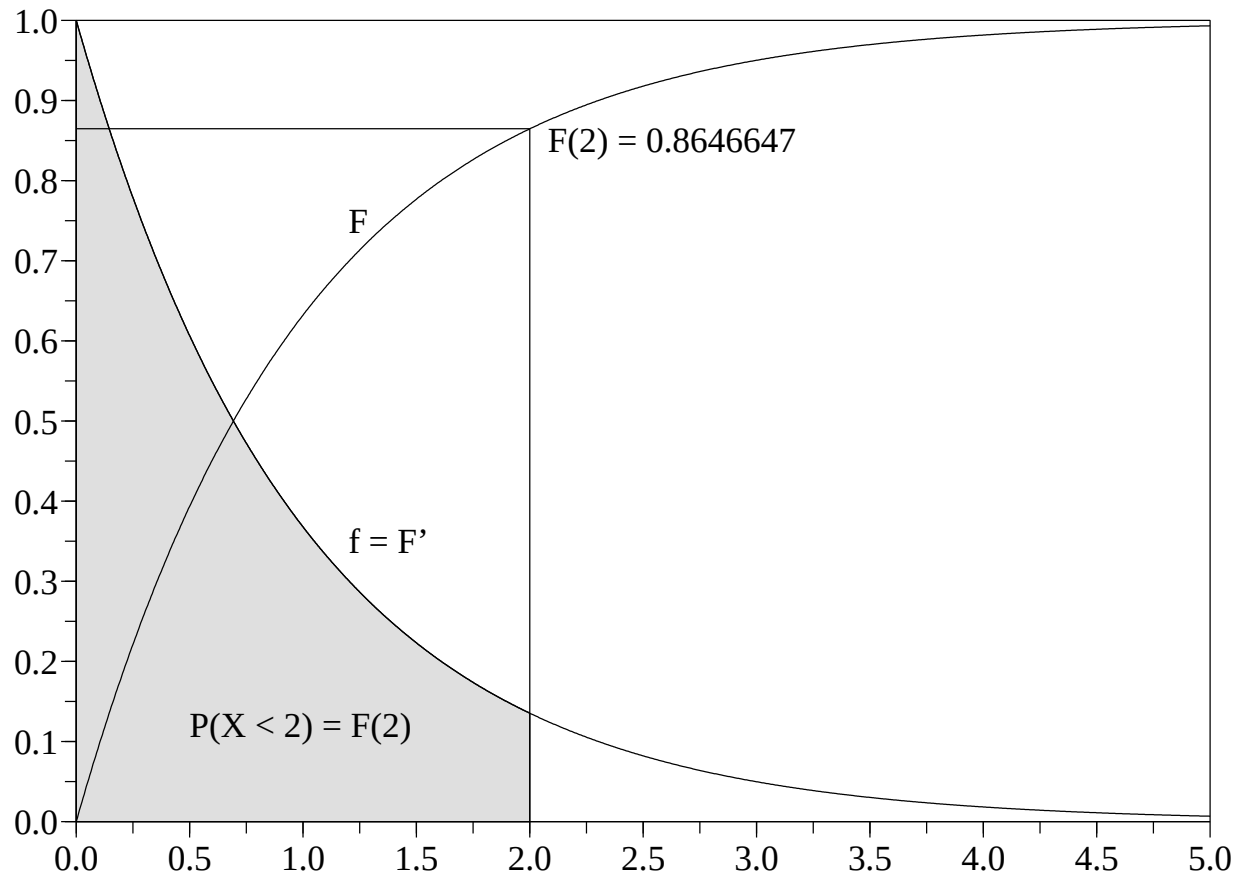


FIG. 4.5 – Graphe de la densité $f(x) = e^{-x}$ et de la fonction de répartition $F(x) = 1 - e^{-x}$

Exemple 4.6.5 On suppose que la durée de vie d'une conversation téléphonique mesurée en minutes, est une variable aléatoire exponentielle de paramètre $\lambda = \frac{1}{10}$. Vous arrivez à une cabine téléphonique et quelqu'un entre juste devant vous. Avec quelle probabilité devez-vous attendre plus de 10 mns ? entre 10 et 20 mns ?

Solution : X désigne la durée de la conversation de la personne précédente.

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

et

$$P(10 < X < 20) \approx 0.233.$$

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 $P(X > s) > 0$ pour tout $s \in \mathbb{R}$. On a $P(X > t + s | X > t) = P(X > s)$ pour tous $s, t > 0$ si et seulement si X suit une loi exponentielle.

Preuve. Soit $G(t) = P(X > t) = 1 - F(t)$, où F est la fonction de répartition de X . D'après (2.4.30), 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 (4.6.39). $\square$

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 (4.6.38). On calcule aisément $F(x) = 1 - e^{-\lambda x}$ pour $x \geq 0$, $F(x) = 0$ pour $x < 0$; $F^{-1}(u) = -(1/\lambda) \ln(1 - u)$. Ainsi, avec U une v.a. uniforme, $X' = -(1/\lambda) \ln(1 - U)$ convient, mais on lui préférera

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

qui suit la même loi (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 d'abord que la fonction gamma est définie pour $\alpha \in]0, \infty[$ par

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx. \quad (4.6.41)$$

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

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{1}{\Gamma(\alpha)} \theta^\alpha x^{\alpha-1} e^{-\theta x} & \text{sinon.} \end{cases} \quad (4.6.42)$$

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

On remarque 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 $E(X)$, sa variance σ_X^2 , et l'espérance $E(X^\beta)$ pour tout $\beta > -\alpha$, sont données par

$$E(X) = \frac{\alpha}{\theta}, \quad \sigma_X^2 = \frac{\alpha}{\theta^2}, \quad E(X^\beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} \frac{1}{\theta^\beta} \quad (4.6.43)$$

On utilisera le fait que $E(X^\beta) = \int_0^\infty x^\beta f(x) dx$ et (4.6.41). En revanche si $\beta \leq -\alpha$, on a $E(X^\beta) = +\infty$.

Lorsque $\alpha = n$, la loi gamma représente la loi du temps d'attente avant la n -ième occurrence d'un certain type d'événements indépendants (voir Exercice 6.1.10).

4.6.4 Variables aléatoires normales (ou variables gaussiennes)

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

Définition 4.6.9 On appelle **variable aléatoire normale centrée réduite** une variable aléatoire X de loi de densité

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (4.6.44)$$

Pour vérifier que cette fonction est d'intégrale 1, on remarque que $I = \int_{-\infty}^{+\infty} f(x)dx$ vérifie

$$I^2 = \int \int_{\mathbb{R}^2} f(x)f(y)dxdy = \int_0^{2\pi} d\theta \int_0^\infty \frac{1}{2\pi} e^{-\rho^2/2} \rho d\rho$$

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

Proposition 4.6.10 *Soit X une variable normale centrée réduite. Alors*

$$E(X) = 0 \quad , \quad Var(X) = 1. \quad (4.6.45)$$

La loi de X est dans ce cas notée $\mathcal{N}(0,1)$, et les qualificatifs “centrée” et “réduite” proviennent précisément de (4.6.45).

Cette proposition est immédiate. $E(X) = 0$ car c'est l'intégrale d'une fonction impaire ; pour calculer $Var(X)$ on peut faire une intégration par parties.

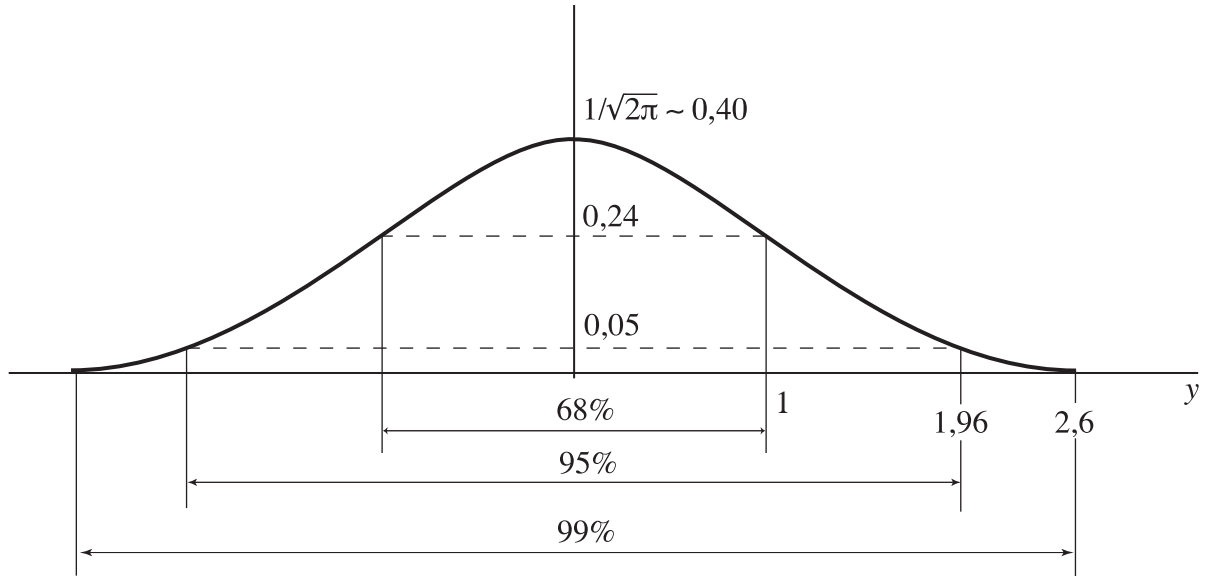


FIG. 4.6 – La courbe de Gauss f et l'aire sous-tendue

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

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x-m)^2}{2\sigma^2}. \quad (4.6.46)$$

On voit que f est une densité en se ramenant par le changement de variable $x \mapsto \frac{x-m}{\sigma}$ à la fonction (4.6.44). 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 autres confrères pour devenir le théorème connu sous le nom de théorème central limite, 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. On peut 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 (4.6.45) 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. Par exemple, pour une variable X de loi normale centrée réduite, en utilisant la symétrie de la densité et la table ci-dessous, on aura $P(|X| < 2) = 2P(0 < X < 2) = 2(P(X < 2) - \frac{1}{2}) = 2(0,9772 - 0,5) = 0,9544$.

De même, on pourra utiliser la table pour l'exercice suivant.

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 avec 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 ci-dessus.

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

Voici le tableau donnant des valeurs de $\Pi(x) = P(X \leq x)$, lorsque X suit la loi gaussienne centrée réduite.

Bien-sûr, les logiciels classiques de statistiques permettent d'obtenir les valeurs données par la table.

Valeurs de la fonction de répartition Π de la loi de Gauss $\mathcal{N}(0, 1)$

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
$\Pi(x)$	0,99865	0,99903	0,99931	0,99952	0,99966	0,99977	0,999841	0,999928	0,999968	0,999997

Remarque 4.6.13 Dans ces exemples, il est très 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 on sait que la loi de X est dans une certaine classe de lois (lois exponentielles, lois normales), on pourra 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 de

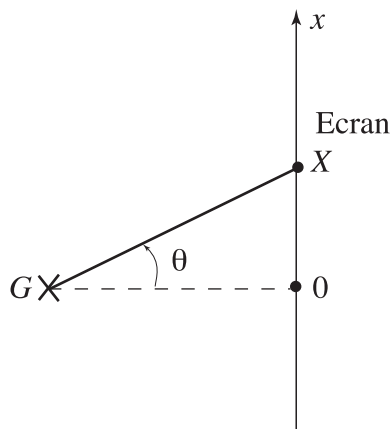


FIG. 4.7 – Gyrophare

l'abscisse X du point d'impact du rayon lumineux sur un écran plan infini situé à distance 1 de G , comme indiqué sur le dessin 4.7.

L'angle θ est une variable aléatoire de densité $q(\theta) = \frac{1}{\pi} \mathbf{1}_{]-\frac{\pi}{2}, \frac{\pi}{2}[}(\theta)$ 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

$$F(x) = P(X \leq x) = P(\theta \leq \arctan x) = \int_{-\infty}^{\arctan x} q(\theta) d\theta = \frac{1}{\pi} \arctan x + \frac{1}{2}.$$

La fonction F est de classe C^1 et a pour dérivée

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}. \quad (4.6.47)$$

Par conséquent, la variable aléatoire X a la loi de densité f , que l'on appelle **loi de Cauchy**.

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

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

ce qui est faux. Ainsi, une variable de Cauchy n'est pas intégrable. (La variable aléatoire X n'a pas d'espérance).

4.7 Des inégalités fameuses

4.7.1 Inégalité de Bienaymé-Chebyshev

Théorème 4.7.1 Soit $X \in L^2$ et $a > 0$. On a alors les inégalités suivantes

- *Inégalité de Markov :*

$$P(|X| \geq a) \leq \frac{E(X^2)}{a^2}. \quad (4.7.48)$$

- *Inégalité de Bienaymé-Chebyshev :*

$$P(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}. \quad (4.7.49)$$

Preuve. On a $X^2 \geq a^2 1_{[a, \infty[}(|X|)$, donc $E(X^2) \geq a^2 E(1_{[a, \infty[}(|X|)) = a^2 P(|X| \geq a)$, ce qui donne la première formule (inégalité de Markov). La seconde découle de la première appliquée à $X - E(X)$. $\square$

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 $E(X)$ 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 (4.7.49), comme on va s'en convaincre ci-dessous.

Exemple 4.7.3 Posons $m = E(X)$ et $\sigma^2 = Var(X)$. Alors, pour tout $a > 0$,

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

soit encore

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

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

$$P(X \in]m - 2\sigma, m + 2\sigma]) \geq 1 - \frac{1}{4} \approx 0.75; \quad P(X \in]m - 3\sigma, m + 3\sigma]) \geq 1 - \frac{1}{9} \approx 0.88.$$

Ces approximations sont universelles mais très grossières. Soit X une variable normale de loi $N(m, \sigma)$. En utilisant la table numérique de la loi normale donnée ci-dessus, on obtient

$$P(X \in]m - 2\sigma, m + 2\sigma]) \approx 0.95 \geq 1 - \frac{1}{4}; \quad P(X \in]m - 3\sigma, m + 3\sigma]) \approx 0.997 \geq 1 - \frac{1}{9}.$$

Les renseignements ainsi obtenus sont beaucoup plus précis.

4.7.2 Inégalité de Cauchy-Schwarz

Proposition 4.7.4 *si X et Y sont dans L^2 , alors $XY \in L^1$, et on a l'inégalité de Cauchy-Schwarz :*

$$|E(XY)| \leq E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}. \quad (4.7.50)$$

On a égalité dans l'inégalité si et seulement si les deux variables aléatoires sont presque-sûrement proportionnelles.

Preuve. Comme $|XY| \leq \frac{1}{2}(X^2 + Y^2)$, on a $XY \in L^1$ dès que $X, Y \in L^2$. Enfin, pour tout $x \in \mathbb{R}$ on a d'après la linéarité et la positivité de l'espérance :

$$x^2 E(X^2) + 2xE(XY) + E(Y^2) = E[(xX + Y)^2] \geq 0.$$

Mais ceci n'est possible que si ce trinôme en x n'a au plus qu'une seule racine réelle ; son discriminant doit donc être négatif ou nul, ce qui donne immédiatement (4.7.50).

Le discriminant est nul si et seulement si il y a une racine double x_0 et dans ce cas, $Y(\omega) = -x_0 X(\omega)$ pour presque tout ω . $\square$

On en déduit en particulier que le coefficient de corrélation défini par 4.4.26 vérifie

$$-1 \leq \rho(X, Y) \leq 1. \quad (4.7.51)$$

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 est une fonction convexe. Alors*

$$E(f(X)) \geq f(E(X)).$$

Preuve. Puisque f est convexe, on sait 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 $a = E(X)$. Il suffit alors de prendre l'espérance pour obtenir le résultat. $\square$

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, $E(u(X)) \leq u(m)$ ce qui rend le placement sûr préférable. Si par contre u est convexe, le placement risqué est meilleur puisque $E(u(X)) \geq u(m)$.

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

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

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*

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy. \quad (4.8.52)$$

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 (4.8.52) est encore vraie et on peut définir $\int_{\mathbb{R} \times \mathbb{R}} f(x, y) dx dy$ comme la valeur commune :*

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

Remarquons l'analogie avec le théorème de Fubini (S9) pour les séries. En fait, la théorie de l'intégration 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 $(f_n)_n$ telle $\sum_n \int_{\mathbb{R}} |f_n(x)| dx < +\infty$, alors*

$$\sum_n \int_{\mathbb{R}} f_n(x) dx = \int_{\mathbb{R}} \left(\sum_n f_n(x) \right) dx.$$

Le Corollaire 4.2.6 se généralise alors de la manière suivante.

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

$$\lambda \left(\prod_{i=1}^d [a_i, b_i] \right) = \prod_{i=1}^d (b_i - a_i),$$

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

On peut 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 $\mathbb{R}^d$ de mesure de Lebesgue $\lambda(V) \in]0, +\infty[$ est définie pour tout borélien $A \subset V$ par

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

La probabilité que le résultat tombe dans une partie A de V est proportionnelle au volume de cette partie. Si $d = 1$ 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 X à valeurs dans $\mathbb{R}^n$, est formé de n variables aléatoires réelles, qui sont les composantes de X : $X = (X_1, \dots, X_n)$.

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

$$F(x_1, \dots, x_n) = P\left(X \in \prod_{i=1}^n]-\infty, x_i]\right). \quad (4.8.54)$$

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

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

$$\int_{\mathbb{R}^n} f(x) dx = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1, \quad (4.8.55)$$

et si

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(y_1, \dots, y_n) dy_1 \dots dy_n. \quad (4.8.56)$$

Exactement comme en dimension 1, on a :

Proposition 4.8.6 Soit 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 $\mathbb{R}$. On a alors $g(X) \in L^1$ si et seulement si

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} |g(x_1, \dots, x_n)| f(x_1, \dots, x_n) dx_1 \dots dx_n < \infty, \quad (4.8.57)$$

et dans ce cas on a

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

Pour simplifier, on écrit aussi $E(g(X)) = \int_{\mathbb{R}^n} g(x) f(x) dx$.

4.8.2 Moments d'un vecteur aléatoire

Si les composantes X_i du vecteur aléatoire $X = (X_1, \dots, X_n)$ sont dans L^1 , on peut définir le vecteur espérance $E(X) = (E(X_1), \dots, E(X_n))$.

Si les composantes X_i du vecteur aléatoire $X = (X_1, \dots, X_n)$ sont dans L^2 , la **matrice des covariances** de X est la matrice $n \times n$ dont les éléments sont les $c_{i,j} = \text{cov}(X_i, X_j)$.

Proposition 4.8.7 *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 4.8.8 *Soit X un vecteur aléatoire n -dimensionnel, de matrice de covariance C . Soit A une matrice $m \times n$ et Y le vecteur aléatoire m -dimensionnel $Y = AX$. La matrice de covariance de Y est alors $AC^t A$, où $^t A$ est la transposée de A .*

Preuve. Calcul immédiat.

□

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. On considère donc une variable aléatoire Z à valeurs dans $\mathbb{R}^2$ de loi à densité, et on note X et Y ses deux composantes : $Z = (X, Y)$. Comme dans le cas discret, on peut obtenir la loi des marginales X et Y à partir de la loi de Z . Ce résultat se généralise sans peine à une dimension supérieure.

Proposition 4.8.9 *Supposons que Z admette une densité f . Alors, X et Y admettent les densités f_X et f_Y suivantes sur $\mathbb{R}$:*

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx. \quad (4.8.59)$$

Les fonctions f_Y et f_Z s'appellent les **densités marginales** de f . Notez que la réciproque de cette proposition est fausse : les variables aléatoires X et Y peuvent avoir des densités sans que le couple $Z = (X, Y)$ en ait une. Supposons par exemple que $X = Y$; si $\Delta = \{(x, x) : x \in \mathbb{R}\}$ est la diagonale de $\mathbb{R}^2$, on a évidemment $P_Z(\Delta) = 1$, tandis que si la formule (4.8.58) était valide pour P_Z , on aurait $P_Z(\Delta) = E(1_\Delta(Z)) = \int_{\mathbb{R}^2} 1_\Delta(z) f(z) dz = 0$. En particulier, faites 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 $x \in \mathbb{R}$, on a par (4.8.56) :

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

Donc si f_X est définie par (4.8.59), on a $P(X \leq x) = \int_{-\infty}^x f_X(u) du$, ce qui montre que f_X est la densité de X . Pour Y on opère de la même manière. $\square$

Exemple 4.8.10 On lance une fléchette sur une cible circulaire de rayon unité. Le joueur est suffisamment loin de la cible pour que l'on suppose le point M d'impact de la fléchette 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

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

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

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

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

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}. \quad (4.8.60)$$

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

L'interprétation de (4.8.60) est la suivante : la fonction $f_{Y|X=x}$ est la densité de la "loi conditionnelle de Y sachant que $X = x$ ". Bien-sûr, on a $P(X = x) = 0$ 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 , on a comme en (4.2.12), et lorsque f est continue :

$$\begin{aligned} f_X(x)\Delta x &\approx P(x \leq X \leq x + \Delta x), \\ f(x, y)\Delta x \Delta y &\approx P(x \leq X \leq x + \Delta x, y \leq Y \leq y + \Delta y). \end{aligned}$$

Par suite

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

Puisque $f_{Y|X=x}$ est une densité, on peut 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 4.8.12 Soit Y une variable aléatoire intégrable.

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

$$E(Y|X = x) = \int_{\mathbb{R}} y f_{Y|X=x}(y) dy. \quad (4.8.61)$$

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

$$E(Y|X) = \psi(X), \quad \text{avec} \quad \psi(x) = E(Y|X = x). \quad (4.8.62)$$

Remarque 4.8.13 1) $\psi(x)$ n'est défini que pour $x \notin B = \{u, f_X(u) = 0\}$. Mais $P(X \in B) = \int_B f_X(u) du = 0$ entraîne que (4.8.62) définit bien l'espérance conditionnelle $\psi(X) = E(Y|X)$ avec probabilité 1.

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

3) On peut bien-sûr intervertir les rôles de X et de Y dans tous les résultats.

En remplaçant les sommes par des intégrales, on peut adapter les preuves de la Section 3.6.1 et obtenir dans le cadre de lois à densités les mêmes propriétés de l'espérance conditionnelle que l'on résume ici.

Proposition 4.8.14 *Soit Y une variable aléatoire intégrable, alors*

1)

$$E(Y) = E(E(Y|X)) = \int_{\mathbb{R}} E(Y|X=x) f_X(x) dx.$$

2) *Supposons de plus que Z est intégrable. Alors*

$$E(aY + bZ | X) = a E(Y | X) + b E(Z | X),$$

3) $E(Y | X) \geq 0$ si $Y \geq 0$,

4) $E(1 | X) = 1$,

5) *Si g est mesurable positive ou bornée, $E(Y g(X) | X) = g(X) E(Y | X)$,*

6) *Pour toute fonction h mesurable positive ou bornée sur $\mathbb{R}^2$,*

$$\begin{aligned} E(h(X, Y)) &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y) f(x, y) dx dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y) f_X(x) f_{Y|X=x}(y) dx dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(x, y) f_Y(y) f_{X|Y=y}(x) dx dy. \end{aligned}$$

Exemple 4.8.15 Soit X et Y de densité jointe $f_{X,Y}(x, y) = \frac{1}{x} \mathbf{1}_T(x, y)$ où T est le triangle $T = \{0 < y < x < 1\}$. On calcule $f_X(x) = \int f_{X,Y}(x, y) dy = \mathbf{1}_{]0,1[}(x)$, 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 $E(Y | X = x)$ est donnée par le milieu $x/2$ de l'intervalle portant cette loi uniforme, et l'on a $E(Y | X) = \frac{X}{2}$.

4.9 Variables aléatoires indépendantes

4.9.1 Indépendance de deux variables aléatoires

Dans la suite de ce paragraphe, on considère un couple (X, Y) de vecteurs aléatoires, X et Y sont respectivement à valeurs dans $\mathbb{R}^m$ et $\mathbb{R}^n$.

Définition 4.9.1 Les vecteurs aléatoires X et Y sont **indépendants** si pour tous boréliens A et B dans les espaces correspondants on a

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B). \quad (4.9.63)$$

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 X et Y sont indépendantes, les variables aléatoires $g(X)$ et $h(Y)$ sont aussi indépendantes. Si de plus $g(X)$ et $h(Y)$ sont dans L^1 , alors le produit $g(X)h(Y)$ est aussi dans L^1 , et on a

$$E(g(X)h(Y)) = E(g(X)) E(h(Y)). \quad (4.9.64)$$

Preuve. La première assertion est évidente par définition même de l'indépendance. Pour le reste, on remarque d'abord que (4.9.64) se réduit à (4.9.63) si g et h sont des fonctions indicatrices. Comme les deux membres de (4.9.64) sont linéaires en g et en h , on a aussi (4.9.64) lorsque g et h sont des fonctions étagées. D'après (4.3.17) on en déduit qu'on a aussi (4.9.64) pour g et h positives quelconques. Si g et h sont de signe quelconque, on a donc (4.9.64) pour $|g|$ et $|h|$, donc si $g(X)$ et $h(Y)$ sont dans L^1 on en déduit que $g(X)h(Y) \in L^1$. Enfin, dans ce cas, par différence (en considérant $g = g^+ - g^-$ et $h = h^+ - h^-$ et en développant le produit) on obtient (4.9.64) 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 , (en fait il suffit qu'elles soient dans L^1), on a $\text{Cov}(X, Y) = 0$ et $\rho(X, Y) = 0$ ($\rho(X, Y)$ est le coefficient de corrélation).

On sait que $|\rho(X, Y)| \leq 1$ par (4.7.51). Si les variables aléatoires X et Y sont indépendantes on a donc $|\rho(X, Y)| = 0$; au contraire si $|\rho(X, Y)|$ est proche de 1, les variables aléatoires sont "fortement dépendantes", d'où le nom de "coefficient de corrélation".

Le résultat suivant est un analogue de l'équivalence (i) $\Leftrightarrow$ (ii) dans la proposition 3.6.9.

Proposition 4.9.4 Soit 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 $Z = (X, Y)$ admette (sur $\mathbb{R}^2$) la densité suivante :

$$f(x, y) = f_X(x)f_Y(y). \quad (4.9.65)$$

Preuve. Si on a indépendance, il vient par (4.9.63) :

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) = \int_{-\infty}^x f_X(u)du \int_{-\infty}^y f_Y(v)dv,$$

ce qui montre que $\mu = P_Z$ vérifie (4.8.56) avec f donnée par (4.9.65).

Inversement, supposons qu'on ait (4.9.65). Le même calcul que ci-dessus montre alors que $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$ pour tous $x, y \in \mathbb{R}$. Mais, exactement comme dans la preuve de la proposition 4.2.3, on peut montrer que ces relations s'étendent en $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ pour tous boréliens A et B , d'où le résultat. $\square$

4.9.2 Suite de variables aléatoires indépendantes

Si on a une famille finie $X_1, \dots, X_n$ 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 variables aléatoires $X_1, \dots, X_n$ sont **indépendantes** (ou, "mutuellement indépendantes") si pour tous boréliens $A_1, \dots, A_n$ on a

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^n P(X_i \in A_i). \quad (4.9.66)$$

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 dans L^2 , les variances vérifient

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i). \quad (4.9.67)$$

Preuve. On utilise ici (4.4.30). Puisque

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i,j=1; i \neq j}^n \text{Cov}(X_i, X_j).$$

Comme ici $\text{Cov}(X_i, X_j) = 0$ si $i \neq j$, on a le résultat. $\square$

Enfin, si on a une suite **infinie** $(X_n)_{n \in \mathbb{N}^*}$,

Définition 4.9.7 La suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires est dite **indépendante** si pour tout n la famille finie $X_1, \dots, X_n$ 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 $(X_n)_n$ entraîne celle de*

- 1) *toute sous-suite $(X_{i_k})_k$,*
- 2) *toute sous-suite de vecteurs issus de $(X_n)_n$,*
- 3) *toute suite de la forme $(f_1(X_1), \dots, f_n(X_n))_n$.*

Exemple 4.9.9 On considère l'ensemble $\Omega = [0, 1[$ muni de la tribu borélienne restreinte à cet ensemble, et de la mesure de Lebesgue. À chaque réel ω , nous associons son développement dyadique (unique si l'on impose que les w_i ne sont pas tous égaux à 1 à partir d'un certain rang) :

$$w = \sum_{i \geq 1} \frac{w_i}{2^i}, \quad w_i \in \{0, 1\}.$$

L'application $X_i : \Omega \rightarrow \{0, 1\}$, qui à ω associe $X_i(\omega) = w_i$ est une variable aléatoire sur Ω . En effet, pour $x_i \in \{0, 1\}$, $1 \leq i \leq n$,

$$\{X_i = x_i\} = \bigcup_{x_1, \dots, x_{i-1}} \left[\sum_{j=1}^i \frac{x_j}{2^j}, \sum_{j=1}^i \frac{x_j}{2^j} + \frac{1}{2^i} \right[,$$

qui est bien un élément de la tribu borélienne de $\Omega = [0, 1[$, et

$$P(\{X_i = x_i\}) = \frac{1}{2^i} \sum_{x_1, \dots, x_{i-1}} 1 = \frac{1}{2}.$$

Montrons l'indépendance des $(X_i)_{1 \leq i \leq n}$. On a

$$\bigcap_{1 \leq i \leq n} \{X_i = x_i\} = \left[\sum_{i=1}^n \frac{x_i}{2^i}, \sum_{i=1}^n \frac{x_i}{2^i} + \frac{1}{2^n} \right[,$$

si bien que

$$P\left(\bigcap_{1 \leq i \leq n} \{X_i = x_i\}\right) = \frac{1}{2^n},$$

qui est la mesure de Lebesgue de l'intervalle ci-dessus. Cela montre que les variables aléatoires X_i sont indépendantes, et de loi de Bernoulli de paramètre $\frac{1}{2}$. Ainsi, sur Ω , nous venons de construire 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 au problème de la construction d'une suite de variables aléatoires indépendantes de lois données.

Théorème 4.9.10 *Soit, pour chaque entier n , une probabilité μ_n sur $\mathbb{R}^{p(n)}$, où $p(n)$ est un entier non nul. Il existe alors un espace Ω muni d'une tribu $\mathcal{A}$ et d'une probabilité P , sur lequel on peut définir une suite (X_n) de variables aléatoires indépendantes, chaque X_n étant de loi μ_n .*

4.10 Calculs de lois

4.10.1 Un théorème d'identification

Dans la sous-section suivante, on va utiliser le théorème suivant, que l'on ne démontrera pas, réciproque de (4.5.33).

Théorème 4.10.1 *1) Si il existe une probabilité μ sur $\mathbb{R}$ telle que pour toute fonction continue bornée h ,*

$$E(h(X)) = \int_{\mathbb{R}} h(x)\mu(dx), \quad (\text{voir notation en (4.3.18)})$$

alors la loi de X est égale à μ .

2) Si il existe une fonction f telle que pour toute fonction continue bornée h ,

$$E(h(X)) = \int_{\mathbb{R}} h(x)f(x)dx,$$

alors la loi de X admet la densité f .

L'idée de la preuve repose sur le fait que si on pouvait prendre $h = 1_{]-\infty, y]}$, on obtiendrait la fonction de répartition de μ , ou (4.2.11).

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 une variable aléatoire X 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 sera égale à

$$P(X \leq a)\delta_a + \mathbf{1}_{[a, +\infty[} f(x)dx.$$

(Une partie à densité et une masse au point a). Pour s'en convaincre, on considère une fonction h continue bornée sur $\mathbb{R}$, et on calcule

$$\begin{aligned} E(h(Y)) &= E(h(\max(a, X))) = E(h(a)\mathbf{1}_{\{a \geq X\}}) + E(h(X)\mathbf{1}_{\{a < X\}}) \\ &= h(a)P(X \leq a) + \int_a^{+\infty} h(x)f(x)dx. \end{aligned}$$

On utilise 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 telle 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 $E(h(Y)) = E(h \circ g(X))$ sous la forme $\int h(y)f_Y(y)dy$ pour une fonction convenable f_Y , et une suffisamment grande classe de fonctions h .

La fonction f_Y sera alors la densité cherchée. Or, (4.5.33) implique

$$E(h(Y)) = E(h \circ g(X)) = \int_{-\infty}^{+\infty} h \circ g(x)f_X(x)dx, \quad (4.10.68)$$

et on fait 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 $a = 0$, on a alors $Y = b$ et la loi de Y (sans densité) est la masse de Dirac en b . Si au contraire $a \neq 0$, on fait le changement de variable $y = ax + b$ dans (4.10.68), ce qui donne

$$E(h(Y)) = \int_{-\infty}^{+\infty} h(ax + b)f_X(x)dx = \int_{-\infty}^{+\infty} h(y)f_X\left(\frac{y-b}{a}\right)\frac{1}{|a|}dy$$

(on peut considérer séparément les cas $a > 0$ et $a < 0$). Donc

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right)\frac{1}{|a|}. \quad (4.10.69)$$

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 g est décroissante sur $\mathbb{R}_-$ et croissante sur $\mathbb{R}_+$. Le changement de variable $y = x^2$ donne alors

$$\begin{aligned} E(h(Y)) &= \int_{-\infty}^0 h(x^2)f_X(x)dx + \int_0^{+\infty} h(x^2)f_X(x)dx \\ &= \int_0^{+\infty} h(y)f_X(-\sqrt{y})\frac{1}{2\sqrt{y}}dy + \int_0^{+\infty} h(y)f_X(\sqrt{y})\frac{1}{2\sqrt{y}}dy, \end{aligned}$$

et on a donc

$$f_Y(y) = (f_X(-\sqrt{y}) + f_X(\sqrt{y})) \frac{1}{2\sqrt{y}}. \quad (4.10.70)$$

On en déduit 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 $X = (X_1, \dots, X_n)$ une variable aléatoire de densité f_X sur $\mathbb{R}^n$. Soit g une fonction de $\mathbb{R}^n$ dans $\mathbb{R}^m$, et $Y = g(X)$. Plusieurs cas sont à considérer :

a) On a $m > n$: Le vecteur Y n'admet pas de densité.

b) On a $m = n$: On fait, comme dans le cas unidimensionnel, le changement de variable $y = g(x)$ dans l'intégrale n -uple qui remplace le terme de droite de (4.10.68).

Rappelons cette formule de changement de variable. D'abord, supposons que g soit une bijection continûment différentiable de A dans B (deux ouverts de $\mathbb{R}^n$). Son **jacobien** est le déterminant $J(g)(x)$ dont les composantes sont les dérivées partielles $\partial g_i(x)/\partial x_j$, où g_i est la $i^{\text{ème}}$ composante de g . On a alors

$$\int_A h \circ g(x) f_X(x) dx = \int_B h(y) f_X \circ g^{-1}(y) \frac{1}{|J(g)(g^{-1}(y))|} dy \quad (4.10.71)$$

(attention à la valeur absolue de $J(g)$). Rappelons d'ailleurs que $1/J(g)(g^{-1}(y))$ est le jacobien de la transformation inverse g^{-1} au point $y \in B$. Si alors $f_X(x) = 0$ en dehors de A , on obtient que Y admet la densité

$$f_Y(y) = 1_B(y) f_X \circ g^{-1}(y) \frac{1}{|J(g)(g^{-1}(y))|}. \quad (4.10.72)$$

Lorsque g est simplement continûment différentiable, il existe souvent une partition finie $(A_i)_{1 \leq i \leq d}$ de l'ensemble $\{x : f_X(x) > 0\}$, telle que g soit injective sur chaque A_i , et on note $B_i = g(A_i)$ l'image de A_i par g . On découpe alors l'intégrale selon les A_i , on applique (4.10.71) à chaque morceau, et on fait la somme. On obtient alors

$$f_Y(y) = \sum_{i=1}^d 1_{B_i}(y) f_X \circ g^{-1}(y) \frac{1}{|J(g)(g^{-1}(y))|}, \quad (4.10.73)$$

où g^{-1} est bien définie sur chaque B_i (comme image réciproque de la restriction de g à A_i).

On a $m < n$: On commence par “compléter” Y , en essayant de construire une application g' de $\mathbb{R}^n$ dans $\mathbb{R}^n$ dont les m premières composantes coïncident avec les composantes de

g , et pour laquelle on puisse appliquer (4.10.72) ou (4.10.73). On obtient ainsi la densité $f_{Y'}$ de $Y' = g'(X)$. Puis on applique l'extension évidente de (4.8.59) :

$$f_Y(y_1, \dots, y_m) = \int \dots \int_{\mathbb{R}^{n-m}} f_{Y'}(y_1, \dots, y_m, y_{m+1}, \dots, y_n) dy_{m+1} \dots dy_n. \quad (4.10.74)$$

Donnons les exemples importants suivants.

- **Coordonnées polaires.**

Soit $X = (U, V)$ un vecteur aléatoire de $\mathbb{R}^2$, et $Y = (R, \Theta)$ ses coordonnées polaires. La transformation g est un difféomorphisme de $A = \mathbb{R}^2 \setminus \{0\}$ dans $B =]0, \infty[\times]0, 2\pi]$, et son inverse g^{-1} s'écrit facilement : $u = r \cos \theta$, $v = r \sin \theta$. Le jacobien de g^{-1} au point (r, θ) est r , donc (4.10.72) entraîne

$$f_Y(r, \theta) = r f_X(r \cos \theta, r \sin \theta) 1_B(r, \theta). \quad (4.10.75)$$

Par exemple si U et V sont indépendantes et de loi $\mathcal{N}(0, 1)$, (4.9.65) entraîne que $f_X(u, v) = \frac{1}{2\pi} \exp -\frac{u^2+v^2}{2}$, donc (4.10.75) implique

$$f_Y(r, \theta) = \frac{1}{2\pi} r e^{-r^2/2} 1_{]0, \infty[}(r) 1_{]0, 2\pi]}(\theta). \quad (4.10.76)$$

En particulier les variables aléatoires R et Θ sont indépendantes, la première suit la loi de densité $re^{-r^2/2} 1_{]0, \infty[}(r)$, et la seconde est uniforme sur $[0, 2\pi]$.

- **Loi beta.**

Soit $X = (U, V)$ 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 X , il faut d'abord "compléter" Y . On prend par exemple $Y' = (Y, Z)$, avec $Z = U + V$, ce qui correspond à $g(u, v) = (\frac{u}{u+v}, u+v)$. Cette application est bijective de $A =]0, \infty[^2$ dans $B =]0, 1[\times]0, \infty[$, et on a $g^{-1}(y, zy) = (yz, z(1-y))$, qui a pour jacobien z . Comme $f_X(u, v) = \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1} v^{\beta-1} e^{-\theta(u+v)} 1_A(u, v)$, (4.10.72) entraîne

$$f_{Y'}(y, z) = \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} z^{\alpha+\beta-1} y^{\alpha-1} (1-y)^{\beta-1} e^{-\theta z} 1_B(y, z). \quad (4.10.77)$$

Il reste à appliquer (4.10.74) :

$$\begin{aligned} f_Y(y) &= \int f_{Y'}(y, z) dz \\ &= \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1} 1_{]0, 1[}(y) \int_0^\infty z^{\alpha+\beta-1} e^{-\theta z} dz \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1} (1-y)^{\beta-1} 1_{]0, 1[}(y) \end{aligned} \quad (4.10.78)$$

(utiliser (4.6.41)). On appelle **loi bêta** de paramètres α et β la loi admettant cette densité.

On obtient aussi facilement la densité de Z : en effet, (4.10.78) montre que $f_{Y'}(y, z)$ est le produit de $f_Y(y)$ par la fonction

$$f_Z(z) = \frac{\theta^{\alpha+\beta}}{\Gamma(\alpha+\beta)} z^{\alpha+\beta-1} e^{-\theta z} 1_{]0, \infty[}(z),$$

qui d'après (4.8.59) est la densité de la variable aléatoire Z . On a 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 $\frac{U}{U+V}$.*

- **Produit de convolution.**

Si U et V sont indépendantes de densités f_U et f_V , on peut de la même manière trouver la densité de la somme $Z = U + V$. Là encore on “complète” en $T = (U, Z)$ (par exemple), correspondant à la bijection $g(u, v) = (u, u + v)$ sur $\mathbb{R}^2$, dont le jacobien est 1. Par suite (4.10.72) conduit à :

Proposition 4.10.5 *Si U et V sont deux variables aléatoires indépendantes de densités f_U et f_V , alors $Z = U + V$ admet la densité*

$$f_Z(z) = \int_{-\infty}^{+\infty} f_U(u) f_V(z - u) du = \int_{-\infty}^{+\infty} f_U(z - v) f_V(v) dv. \quad (4.10.79)$$

Etant donnée l'importance de la formule ci-dessus, on lui donne un nom : on dit que la fonction f_Z est 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.

EXERCICE 4.10.6 *Cet exercice vous donne une méthode (usuelle) pour simuler deux variables aléatoires indépendantes de loi normale.*

Soient U et V deux variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. On pose $R = \sqrt{-2 \ln U}$ et $T = 2\pi V$.

Déterminer la loi du couple de variables aléatoires (R, T) , puis celle de la variable R^2 .

Montrer que les variables aléatoires $X = R \cos T$ et $Y = R \sin T$ sont indépendantes, et identifier les lois respectives de X et de Y .

4.11 Simulation de suites indépendantes de variables aléatoires

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 P_Y .

4.11.1 Inversion de la fonction de répartition

Généralisons la méthode introduite en Section 4.2.4.

Supposons que l'on connaît la fonction de répartition F des variables Y_i . À partir d'une suite de variables aléatoires uniformes sur $[0, 1]$ et indépendantes, on peut simuler une suite de variables aléatoires réelles indépendantes de fonction de répartition F . Comme en Section 4.2.4, on définit la fonction "inverse" continue à gauche de F par :

$$G(x) = \inf\{y : F(y) \geq x\}, \quad \forall x \in]0, 1[. \quad (4.11.80)$$

Proposition 4.11.1 *Si on pose $Y_n = G(X_n)$, on obtient une suite de variables aléatoires indépendantes, de loi 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 en Section 4.2.4. $\square$

Expérience de simulation en SCILAB

Un exemple de simulation vous est présenté sur la page WEB du cours à l'adresse :

www.enseignement.polytechnique.fr/mathematiques-appliquees/Aleatoire
(puis cliquer sur le lien "Histogrammes et simulations de variables aléatoires").

On y simule n réalisations indépendantes $X_1, X_2, \dots, X_n$ de variables aléatoires de même loi exponentielle de paramètre $\lambda = 1$. Une manière commode de représenter ces tirages est d'en tracer un **histogramme** : on partitionne $[0, 5]$ en intervalles I_k et pour chacun d'eux on calcule la proportion de réalisations qui tombent dedans, proportion que l'on normalise 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

$$\hat{f}_n(x) = \frac{1}{\lambda(I_k)} \frac{\text{card}\{i : i \leq n, X_i \in I_k\}}{n} \quad \text{pour } x \in I_k.$$

On a noté $\lambda(I)$ la longueur de l'intervalle I , et comme précédemment $\text{card}(A)$ le cardinal de l'ensemble fini A .

Ici on a choisi la partition de $[0, 5]$ en intervalles I_k avec

$$\begin{aligned} I_1, \dots, I_{20} & \quad \text{intervalles consécutifs de } [0, 2[\text{ de longueur } 1/10 \text{ et} \\ I_{21}, \dots, I_{35} & \quad \text{intervalles consécutifs de } [2, 5[\text{ de longueur } 1/5. \end{aligned}$$

Au fur et à mesure de la simulation, on constate que $\hat{f}_n$ "approche" la vraie densité f (ou plutôt, la fonction en escalier égale aux moyennes de f sur les intervalles) lorsque n

augmente. On justifiera cette observation par un résultat de convergence (loi des grands nombres) au chapitre 5. On peut recommencer l'expérience et constater la convergence vers la même limite. Le script SCILAB est disponible sur cette page : vous pouvez en modifier les divers paramètres comme la loi simulée, le nombre de classes de l'histogramme. . .

4.11.2 Méthode du rejet

Cette méthode s'applique lorsque la probabilité P_Y admet une densité f , et lorsqu'on connaît une autre probabilité ν , telle que

- on peut simuler des variables aléatoires de loi ν (par exemple par la méthode précédente) ;

- ν admet une densité g telle que

$$f(x) \leq ag(x), \quad \forall x \in \mathbb{R}^d, \quad (4.11.81)$$

pour une constante connue a (nécessairement $a \geq 1$, et même $a > 1$ si $P_Y \neq \nu$, 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).

On suppose aussi qu'on dispose, d'une part de la suite $X_1, \dots, X_n, \dots$ ci-dessus (constituée de variables aléatoires indépendantes de loi uniforme), et d'autre part d'une suite $Z_1, \dots, Z_n, \dots$ potentiellement infinie de variables aléatoires indépendantes de loi ν et indépendantes des $(X_n)_n$. On pose alors

$$N = \inf(n \in \mathbb{N}^* : f(Z_n) > aX_n g(Z_n)), \quad Y = Z_n \quad \text{si } N = n. \quad (4.11.82)$$

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 P_Y et N et Y sont indépendantes.*

Preuve. La première assertion entraîne que la variable Y est bien définie dans la seconde formule (4.11.82).

Notons $A_n = \{f(Z_n) > aX_n g(Z_n)\}$. Les événements A_n sont indépendants et ont même probabilité. On pose $P(A_n) = \alpha$ (on calculera α plus tard), de sorte que $P(N > n) = (1 - \alpha)^n$, et pour toute fonction h on a

$$E(h(Y)) = \sum_{n=1}^{\infty} E(h(Z_n) 1_{A_n} 1_{\{N > n-1\}}).$$

Les variables aléatoires $h(Z_n) 1_{A_n}$ d'une part et $1_{\{N > n-1\}}$ d'autre part sont indépendantes, donc

$$E(h(Y)) = \sum_{n=1}^{\infty} E(h(Z_n) 1_{A_n}) (1 - \alpha)^{n-1}. \quad (4.11.83)$$

Enfin $E(h(Z_n)1_{A_n})$ vaut

$$\int h(z)g(z) \left(\int_0^1 1_{\{f(z) > axg(z)\}} dx \right) dz = \int h(z)g(z) \frac{f(z)}{ag(z)} dz = \frac{1}{a} \int h(z)f(z) dz.$$

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 $E(h(Z_n)1_{A_n})$ par sa valeur dans (4.11.83), on obtient que $E(h(Y)) = \int h(z)f(z)dz$, d'où le résultat. L'indépendance est immédiate. $\square$

Nous avons ainsi obtenu une variable aléatoire de loi P_Y . Pour obtenir une suite de telles variables aléatoires indépendantes, il faut répéter la même procédure.

On peut comparer les deux méthodes. La première est très simple à mettre en oeuvre, **si on connaît explicitement la fonction G** , ce qui est assez rare dans la pratique. La seconde nécessite la connaissance de f , g et a , et aussi le fait qu'on sache préalablement simuler selon la loi ν (si on peut par exemple utiliser la première méthode pour cette loi) : ces conditions sont assez souvent remplies. Elle est malheureusement parfois longue à mettre en oeuvre (sa "longueur" est proportionnelle à N).

Un cas particulier : on suppose 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 $a = C(c - b)$. On peut adapter la Proposition 4.11.2 et montrer que si $(U_k)_k$ et $(V_k)_k$ des suite de de variables aléatoires indépendantes et de loi respectivement uniforme sur le rectangle $[b, c]$ et sur le rectangle $[0, C]$, alors

$$Y = U_N \quad \text{avec} \quad N = \min\{k \geq 1 : V_k \leq f(U_k)\}$$

définit une variable aléatoire de densité f .

On va 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 . On rejette 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 $E(N) = C(c - b)$. 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.

EXERCICE 4.11.3 Dans le cadre de la proposition 4.11.2,

- montrer que la loi conditionnelle de U_1 sachant que $V_1 \leq f(U_1)$ a pour densité f ;
- montrer qu'il en est de même pour la loi de U_n sachant $N = n$ ($\forall n \geq 1$).

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 réalisations d'un événement converge vers la probabilité de réalisation de cet événement. Ainsi, notre modèle est bien cohérent avec l'intuition. Par ailleurs, 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 d'états Ω (muni de la tribu $\mathcal{A}$ et de la probabilité P). Nous voulons étudier la répartition des valeurs d'une variable aléatoire X de loi P_X , réalisées au cours d'une succession de n expériences aléatoires indépendantes. Par exemple, vous interviewez n étudiants choisis au hasard parmi vos concitoyens et vous leur demandez s'ils aiment les brocolis. (Ici, la réponse sera 0 ou 1, et la variable aléatoire associée X sera une variable de loi de Bernoulli 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 P_X . Nous sommes intéressés au comportement aléatoire de cette suite de résultats, et en particulier par 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

$$\lim_{n \rightarrow +\infty} \frac{X_1 + \cdots + X_n}{n},$$

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

Nous allons étudier, dans ce paragraphe, des modes de convergence impliquant la proximité des variables aléatoires elle-mêmes, contrairement au cas de la convergence en loi qui sera étudiée ultérieurement.

On considère une suite $(X_n)_{n \geq 1}$ de variables aléatoires, définies sur un espace de probabilité $(\Omega, \mathcal{A}, P)$, et à valeurs dans $\mathbb{R}^d$. On considère également sur le même espace une variable aléatoire “limite” X . On notera $|\cdot|$ la valeur absolue dans $\mathbb{R}$ ou la norme dans $\mathbb{R}^d$.

Définition 5.1.1 a) La suite (X_n) converge **presque sûrement** vers X , ce qui s’écrit $X_n \rightarrow X$ p.s., s’il existe un ensemble $N \in \mathcal{A}$ de probabilité nulle, tel que $X_n(\omega) \rightarrow X(\omega)$ quand $n \rightarrow \infty$, pour tout $\omega \notin N$.

b) La suite (X_n) converge **en probabilité** vers X , ce qui s’écrit $X_n \xrightarrow{P} X$, si pour tout $\varepsilon > 0$ on a

$$P(|X_n - X| \geq \varepsilon) \rightarrow 0 \quad \text{quand } n \rightarrow \infty. \quad (5.1.1)$$

c) La suite (X_n) converge **en moyenne** vers X , ce qui s’écrit $X_n \xrightarrow{L^1} X$, si X_n et X sont dans L^1 et si

$$E(|X_n - X|) \rightarrow 0 \quad \text{quand } n \rightarrow \infty. \quad (5.1.2)$$

Remarque : La convergence presque-sûre est la plus proche de la convergence simple des fonctions. Mais ici, on permet à certains ω de ne pas vérifier $X_n(\omega) \rightarrow X(\omega)$, si toutefois, leur probabilité de réalisation 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 telles que

$$P(X_n = 1) = \frac{1}{n} \quad ; \quad P(X_n = 0) = 1 - \frac{1}{n}.$$

Pour tout $\varepsilon \in]0, 1[$, on a $P(|X_n| > \varepsilon) = \frac{1}{n}$, qui tend vers 0 quand n tend vers l'infini. Ainsi, la suite $(X_n)_n$ tend vers $X = 0$ en probabilité. Comme $E(X_n) = \frac{1}{n}$, elle tend aussi en moyenne vers 0. Mais considérons maintenant une suite $(Y_n)_n$ de variables aléatoires de Bernoulli telles que

$$P(Y_n = n^2) = \frac{1}{n} \quad ; \quad P(Y_n = 0) = 1 - \frac{1}{n}.$$

Par le même argument que ci-dessus, on voit que la suite $(Y_n)_n$ converge en probabilité vers 0, mais en revanche, $E(Y_n) = n$, et la suite ne converge pas en moyenne vers 0.

Rappelons un des résultats les plus importants de la théorie de l'intégration, qui relie deux des modes de convergence que nous venons de définir. (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 5.1.3 (de Lebesgue, ou de convergence dominée) : Si les variables aléatoires X_n convergent presque sûrement sur Ω vers une limite X et si $|X_n| \leq Z$ pour tout n , avec $Z \in L^1$, alors X_n et X sont dans L^1 et on a $E(|X_n - X|) \rightarrow 0$. En particulier, $E(X_n) \rightarrow E(X)$.

D'autres relations existent entre ces convergences, que nous allons explorer maintenant.

Proposition 5.1.4 La convergence presque-sûre ou la convergence en moyenne entraînent la convergence en probabilité.

Preuve. Soit $A_{n,\varepsilon} = \{\omega, |X_n(\omega) - X(\omega)| \geq \varepsilon\}$.

a) Supposons que $X_n \rightarrow X$ p.s. et soit N l'ensemble de probabilité nulle en dehors duquel on a $X_n(\omega) \rightarrow X(\omega)$. Si $\omega \notin N$ on a $\omega \notin A_{n,\varepsilon}$ pour tout $n \geq n_0$, où n_0 dépend de ω et de ε , ce qui implique que les variables aléatoires $Y_{n,\varepsilon} = 1_{N^c \cap A_{n,\varepsilon}}$ tendent simplement vers 0 quand $n \rightarrow \infty$. Comme on a aussi $0 \leq Y_{n,\varepsilon} \leq 1$, le théorème de convergence dominée (Théorème 5.1.3) entraîne que $E(Y_{n,\varepsilon}) \rightarrow 0$. Mais

$$P(A_{n,\varepsilon}) \leq P(N^c \cap A_{n,\varepsilon}) + P(N) = P(N^c \cap A_{n,\varepsilon}) = E(Y_{n,\varepsilon}) \xrightarrow{n \rightarrow \infty} 0,$$

et on en déduit (5.1.1).

b) Supposons que $X_n \xrightarrow{L^1} X$. Pour $\varepsilon > 0$ on a $1_{A_{n,\varepsilon}} \leq \frac{1}{\varepsilon} |X - X_n|$, au sens où pour tout $\omega \in \Omega$, $1_{A_{n,\varepsilon}}(\omega) \leq \frac{1}{\varepsilon} |X(\omega) - X_n(\omega)|$. On en déduit que

$$P(A_{n,\varepsilon}) \leq \frac{1}{\varepsilon} E(|X_n - X|) \xrightarrow{n \rightarrow \infty} 0,$$

d'où encore (5.1.1). $\square$

La convergence en probabilité n'entraîne pas la convergence en moyenne, comme on l'a vu dans l'exemple donné ci-dessus. Si les X_n ne sont pas trop grands, il y a cependant équivalence entre les deux modes de convergence. En voici un exemple.

Proposition 5.1.5 *S'il existe une constante a telle que $|X_n| \leq a$ presque-sûrement, il y a équivalence entre $X_n \xrightarrow{P} X$ et $X_n \xrightarrow{L^1} X$.*

Preuve. Etant donnée la proposition précédente, dont on reprend les notations, il suffit de montrer que la convergence en probabilité implique la convergence en moyenne, lorsque $|X_n| \leq a$.

Comme $|X_n| \leq a$, on a $\{|X| > a + \varepsilon\} \subset A_{n,\varepsilon}$, et donc $P(|X| > a + \varepsilon) \leq P(A_{n,\varepsilon})$. En faisant tendre n vers $+\infty$, on en déduit que $P(|X| > a + \varepsilon) = 0$. Ceci est vrai pour tout ε , donc on a

$$P(|X| > a) = 0. \quad (5.1.3)$$

Comme $|X_n| \leq a$ on a aussi

$$|X_n - X| \leq \varepsilon + (|X_n| + |X|)1_{A_{n,\varepsilon}} \leq \varepsilon + 2a1_{A_{n,\varepsilon}}$$

sur l'ensemble $\{|X| \leq a\}$, qui est de probabilité 1. Donc il vient

$$E(|X_n - X|) \leq \varepsilon + 2aP(A_{n,\varepsilon}).$$

On déduit (par (5.1.1)) que $\limsup_n E(|X_n - X|) \leq \varepsilon$, et comme ε est arbitrairement proche de 0, on a en fait (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 5.1.4, mais "à peine plus", comme le montre le résultat suivant, donné sans démonstration.

Proposition 5.1.6 *Si $X_n \xrightarrow{P} X$, il existe une sous-suite (n_k) telle que $X_{n_k} \rightarrow X$ p.s. quand $k \rightarrow \infty$.*

Exemple 5.1.7 Soit $\Omega = \mathbb{R}$ muni de sa tribu borélienne et P la probabilité uniforme sur $[0, 1]$. Soit $X_n = 1_{A_n}$, où A_n est un intervalle de $[0, 1]$ de longueur $1/n$. On a alors $E(X_n) = 1/n$, donc la suite X_n tend vers $X = 0$ en moyenne, et donc en probabilité. Supposons que les A_n sont placés bout-à-bout, en recommençant en 0 chaque fois qu'on

arrive au point 1. On voit qu'on parcourt indéfiniment l'intervalle $[0, 1]$ (car la série de terme général $1/n$ diverge) : donc la suite numérique $X_n(\omega)$ ne converge pour aucun ω , et on n'a pas $X_n \rightarrow X$ p.s. ; cependant comme la série $\sum_n 1/n^2$ converge, on voit que $X_{n^2} \rightarrow X = 0$ p.s.

Proposition 5.1.8 *Soit f une fonction continue de $\mathbb{R}^d$ dans $\mathbb{R}$.*

a) *Si $X_n \rightarrow X$ p.s., alors $f(X_n) \rightarrow f(X)$ p.s.*

b) *Si $X_n \xrightarrow{P} X$, alors $f(X_n) \xrightarrow{P} f(X)$.*

Preuve. (a) est évident. Pour (b) remarquons d'abord que si $K > 0$ et $\varepsilon > 0$,

$$\{|f(X_n) - f(X)| \geq \varepsilon\} \subset \{|X| > K\} \cup \{|X| \leq K, |f(X_n) - f(X)| \geq \varepsilon\}. \quad (5.1.4)$$

La fonction f est uniformément continue sur $\{x : |x| \leq 2K\}$, donc il existe $\eta > 0$ tel que $|x - y| < \eta$ et $|x| \leq 2K$, $|y| \leq 2K$ impliquent $|f(x) - f(y)| < \varepsilon$. Donc, en choisissant η suffisamment petit, $|x| \leq K$ entraîne $|y| \leq 2K$, et (5.1.4) implique :

$$\{|f(X_n) - f(X)| \geq \varepsilon\} \subset \{|X| > K\} \cup \{|X_n - X| \geq \eta\},$$

$$P(|f(X_n) - f(X)| \geq \varepsilon) \leq P(|X| > K) + P(|X_n - X| \geq \eta).$$

D'après l'hypothèse il vient

$$\limsup_n P(|f(X_n) - f(X)| \geq \varepsilon) \leq P(|X| > K). \quad (5.1.5)$$

Enfin $\lim_{K \rightarrow \infty} P(|X| > K) = 0$, (on le montre grâce au théorème de convergence dominée), et donc dans (5.1.5) la $\limsup$ est nulle. On obtient ainsi le résultat. $\square$

5.2 La loi des grands nombres

Reprenons la modélisation introduite dans l'introduction de ce chapitre. On considère une suite $(X_n)_{n \geq 1}$ de variables aléatoires réelles **indépendantes et de même loi**. On considère la moyenne des n premières variables aléatoires, i.e.

$$M_n = \frac{1}{n}(X_1 + \dots + X_n), \quad (5.2.6)$$

et 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**.

Commençons par un résultat partiel, mais que nous démontrons complètement.

Théorème 5.2.1 *On considère une suite de variables aléatoires X_n indépendantes, de même loi, et de carré intégrable ($X_n \in L^2$). On note $m = E(X_n)$ leur moyenne commune. On a alors*

$$M_n = \frac{X_1 + \dots + X_n}{n} \rightarrow m \quad \text{p.s. et en moyenne,} \quad (5.2.7)$$

quand n tend vers l'infini. On a donc aussi convergence en probabilité. On a même un peu plus que la convergence en moyenne, à savoir que si $n \rightarrow +\infty$,

$$E((M_n - m)^2) \rightarrow 0. \quad (5.2.8)$$

Preuve. Notons σ^2 la variance commune à toutes les variables X_n , qui existe puisqu'on a supposé $X_n \in L^2$. En vertu de la linéarité de l'espérance et de (4.9.67) et (4.4.29), on a

$$E(M_n) = m, \quad E((M_n - m)^2) = \text{Var}(M_n) = \frac{\sigma^2}{n}, \quad (5.2.9)$$

d'où (5.2.8).

Comme $E(|Y|)^2 \leq E(Y^2)$, on en déduit que $E(|M_n - m|) \rightarrow 0$, donc on a aussi la convergence de M_n vers m en moyenne.

Il reste à montrer la convergence presque-sûre. Ce résultat est beaucoup plus délicat.

Quitte à remplacer X_n par $X_n - m$ (donc M_n par $M_n - m$), on peut supposer que $m = 0$.

Montrons tout d'abord que la sous-suite $(M_{n^2})_n$ converge p.s. vers 0.

D'après l'inégalité de Bienaymé-Chebyshev, (5.2.9) implique que pour $q \in \mathbb{N}^*$,

$$P(|M_{n^2}| \geq \frac{1}{q}) \leq \frac{\sigma^2 q^2}{n^2}.$$

Donc si $A_{n,q} = \{|M_{n^2}| \geq \frac{1}{q}\}$, on a $\sum_{n \geq 1} P(A_{n,q}) < \infty$. Posons ensuite $B_{n,q} = \cup_{m \geq n} A_{m,q}$ et $C_q = \cap_{n \geq 1} B_{n,q}$. En appliquant le théorème de Borel-Cantelli (Théorème 2.4.18) on montre alors que $P(C_q) = 0$.

Par suite si on pose $N = \cup_{q \in \mathbb{N}^*} C_q$, on obtient $P(N) \leq \sum_{q=1}^{\infty} P(C_q) = 0$ en vertu de (2.3.22). Maintenant, si $\omega \notin N$, alors $\omega \in \cap_{q \in \mathbb{N}^*} (C_q)^c$. Ainsi, pour tout $q \geq 1$ on a $\omega \notin C_q$, 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 p.s. vers 0.

Pour tout entier n on note $p(n)$ l'entier tel que $p(n)^2 \leq n < (p(n) + 1)^2$. On a

$$M_n - \frac{p(n)^2}{n} M_{p(n)^2} = \frac{1}{n} \sum_{p=p(n)^2+1}^n X_p,$$

et comme pour la seconde égalité de (5.2.9), on a

$$\begin{aligned} E \left(\left(M_n - \frac{p(n)^2}{n} M_{p(n)^2} \right)^2 \right) &= \frac{n - p(n)^2}{n^2} \sigma^2 \\ &\leq \frac{2p(n) + 1}{n^2} \sigma^2 \leq \frac{2\sqrt{n} + 1}{n^2} \sigma^2, \end{aligned}$$

parce que $p(n) \leq \sqrt{n}$. D'après Bienaymé-Chebyshev (encore), on a

$$P \left(\left| M_n - \frac{p(n)^2}{n} M_{p(n)^2} \right| \geq a \right) \leq \frac{2\sqrt{n} + 1}{n^2} \frac{\sigma^2}{a^2}.$$

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 $p(n)^2/n \rightarrow 1$. On en déduit que $M_n \rightarrow 0$ p.s. $\square$

Plus généralement, on a le résultat suivant, admis sans démonstration. Pour cette preuve, on pourra consulter le cours de deuxième année "Promenade aléatoire".

Théorème 5.2.2 *On suppose que les X_n sont des variables aléatoires indépendantes, de même loi, et intégrables, et on note $m = E(X_n)$ leur moyenne commune. On a alors*

$$\frac{X_1 + \dots + X_n}{n} \rightarrow m \quad \text{p.s. et en moyenne,} \quad (5.2.11)$$

quand n tend vers l'infini.

Revenons à "l'approche par les fréquences" du Chapitre 2. Soit un événement A . On répète l'expérience, et on note 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

$$f_n(A) = \frac{1}{n}(X_1 + \dots + X_n) = M_n.$$

Par ailleurs, les X_i sont indépendantes et de même loi et $E(X_i) = P(X_i = 1) = P(A)$. Donc (5.2.7) implique que $f_n(A) \rightarrow P(A)$ p.s. : On obtient 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 .

Exemple 5.2.3 Considérons une suite X_n de variables aléatoires ne prenant que les valeurs 0 et 1. On définit 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 P une probabilité sous laquelle les X_n sont indépendantes et de même loi, avec $P(X_n = 1) = p \in]0, 1[$. (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). 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 P n'est pas une somme de mesures de Dirac.

Cet exemple montre que lorsqu'on étudie la convergence des variables aléatoires, il est **indispensable** d'introduire la convergence p.s., puisqu'on n'a généralement pas la convergence simple (i.e. pour tout ω).

EXERCICE 5.2.4 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 $n \geq 1$,

$$P(X_n = \frac{1}{u_n}) = u_n; P(X_n = 0) = 1 - u_n.$$

1) Calculer $E(X_n)$.

2) On suppose que $\sum_{n \geq 1} u_n < +\infty$. En déduire que $X_n \xrightarrow{P} 0$. Montrer en fait que $X_n \xrightarrow{\text{p.s.}} 0$ quand n tend vers l'infini. (On pourra considérer les ensembles $A_n = \{|X_n| > \varepsilon\}$ et appliquer le théorème de Borel-Cantelli). 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 5.2.5 Montrer que la loi forte des grands nombres reste vraie pour des variables aléatoires indépendantes positives d'espérance commune égale à $+\infty$: $\frac{X_1 + \dots + X_n}{n} \xrightarrow{\text{p.s.}} +\infty$. (On pourra considérer pour un réel $M > 0$ donné les variables aléatoires $X_i \wedge M$).

5.3 Méthode de Monte-Carlo

Montrons comment on peut 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. On applique la loi des grands nombres aux variables aléatoires $f(X_i)$ qui vérifient bien toutes les hypothèses voulues puisque f est bornée. On a alors

$$\lim_n \frac{f(X_1) + \dots + f(X_n)}{n} = E(f(X)) = \int_0^1 f(x)dx.$$

□

En choisissant des variables aléatoires de loi uniforme sur $[a, b]$, on peut de même obtenir une approximation d'une intégrale définie sur l'intervalle $[a, b]$.

Ce résultat se généralise à toutes les dimensions.

On veut calculer l'intégrale $I = \int_A f(x)dx$, où f est une fonction mesurable bornée et A est le cube $\{x = (x_1, \dots, x_d) : |x_i| \leq \alpha \ \forall i\}$ de $\mathbb{R}^d$. Pour calculer I , on peut simuler une suite $X_1, \dots, X_n$ de variables aléatoires indépendantes et de loi uniforme sur A : cela revient à dire que si chaque 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

$$I_n = \frac{(2\alpha)^d}{n} (f(X_1) + \dots + f(X_n)). \quad (5.3.12)$$

En effet la loi uniforme sur A admet la densité $g(x) = \frac{1}{(2\alpha)^d} 1_A(x)$, donc l'espérance des $f(X_i)$ 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, et qui sera l'objet du chapitre suivant, va donner un contrôle de cette erreur.

L'avantage de cette méthode est qu'elle marche même si la fonction f est très irrégulière (alors que les méthodes déterministes de type "méthode du trapèze" ne marchent que si

la fonction f est continue). En outre elle marche indépendamment de la dimension d , le temps de calcul étant proportionnel à d (en effet, tirer une variable X de loi uniforme sur A revient à tirer ses d composantes, chacune selon la loi uniforme sur $[0, 1]$), 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 le temps de calcul est grosso modo proportionnel à une constante à la puissance d . On verra 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.

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 qu'on appelle la **fonction caractéristique** d'une variable aléatoire , et qui dans d'autres branches des mathématiques s'appelle aussi **transformée de Fourier**.

6.1.1 Définition et premières propriétés

On notera $\langle x, y \rangle$ le produit scalaire de deux vecteurs de $\mathbb{R}^n$. Si $u \in \mathbb{R}^n$, la fonction (complexe) $x \mapsto e^{i\langle u, x \rangle}$ est continue, de module 1. Donc si X est une variable aléatoire à valeurs dans $\mathbb{R}^n$, on peut considérer $e^{i\langle u, X \rangle}$ comme une variable aléatoire à valeurs complexes. Ses parties réelle $Y = \cos(\langle u, X \rangle)$ et imaginaire $Z = \sin(\langle u, X \rangle)$ 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 $e^{i\langle u, X \rangle}$ est

$$E(e^{i\langle u, X \rangle}) = E(Y) + iE(Z) = E(\cos(\langle u, X \rangle)) + iE(\sin(\langle u, X \rangle)).$$

Définition 6.1.1 Si X est une variable à valeurs dans $\mathbb{R}^n$, sa **fonction caractéristique** est la fonction ϕ_X de $\mathbb{R}^n$ dans $\mathbb{C}$ définie par

$$\phi_X(u) = E(e^{i\langle u, X \rangle}). \quad (6.1.1)$$

Si X est à valeurs réelles, ϕ_X est définie de $\mathbb{R}$ dans $\mathbb{C}$, et vaut :

$$\forall u \in \mathbb{R}, \quad \phi_X(u) = E(e^{iuX}). \quad (6.1.2)$$

Nous remarquerons que la fonction caractéristique ne dépend en fait **que de la loi** P_X de X . C'est la "transformée de Fourier" de la loi P_X .

Nous verrons que cette fonction porte bien son nom, en ce sens qu'elle va caractériser la loi P_X . C'est une notion, qui, de ce point de vue, généralise la fonction génératrice que nous avons vue au Chapitre 2, et formellement on a

$$\phi_X(u) = G_X(e^{iu}).$$

Proposition 6.1.2 ϕ_X est une fonction de module inférieur ou égal à 1, continue, avec

$$\phi_X(0) = 1 ; \quad \phi_X(-u) = \overline{\phi_X(u)}.$$

Preuve. $|z|$ désigne le module d'un nombre complexe z .

Comme $E(Y)^2 \leq E(Y^2)$ pour toute variable aléatoire réelle Y , on a :

$$|\phi_X(u)|^2 = (E(\cos \langle u, X \rangle))^2 + (E(\sin \langle u, X \rangle))^2 \leq E(\cos^2 \langle u, X \rangle + \sin^2 \langle u, X \rangle),$$

et donc $|\phi_X(u)| \leq 1$.

Pour montrer la continuité, considérons une suite $u_p \rightarrow u$. On a convergence simple des $e^{i\langle u_p, X \rangle}$ vers $e^{i\langle u, X \rangle}$ et la majoration du module par la constante 1, qui est dans L^1 . En appliquant le théorème de convergence dominée (Théorème 5.1.3), on obtient que $\phi_X(u_p) \rightarrow \phi_X(u)$, et la fonction ϕ_X est continue. $\square$

Proposition 6.1.3 Si X est une variable aléatoire à valeurs dans $\mathbb{R}^n$, si $a \in \mathbb{R}^m$ et si A est une matrice $m \times n$, on a :

$$\phi_{a+AX}(u) = e^{i\langle a, u \rangle} \phi_X({}^t A u), \quad \forall u \in \mathbb{R}^m, \quad (6.1.3)$$

où ${}^t A$ désigne la transposée de la matrice A .

Preuve. On a $e^{i\langle u, a+AX \rangle} = e^{i\langle u, a \rangle} e^{i\langle {}^t A u, X \rangle}$. En effet, $\langle u, AX \rangle = \langle {}^t A u, X \rangle$. Il suffit alors de prendre les espérances pour obtenir le résultat. $\square$

6.1.2 Exemples

1) **X suit une loi binomiale $\mathcal{B}(p, n)$**

$$\begin{aligned} E(e^{iuX}) &= \sum_{k=0}^n e^{iuk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n \binom{n}{k} (e^{iu}p)^k (1-p)^{n-k} \\ &= (e^{iu}p + 1 - p)^n. \end{aligned}$$

Ainsi,

$$\phi_X(u) = (pe^{iu} + 1 - p)^n. \quad (6.1.4)$$

2) **X suit une loi de Poisson de paramètre θ**

$$E(e^{iuX}) = \sum_{k \geq 0} e^{iuk} \frac{\theta^k}{k!} e^{-\theta} = e^{\theta e^{iu}} e^{-\theta} = e^{\theta(e^{iu}-1)}.$$

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 $g(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. On constate 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 $g(x)e^{sx} = g(x-s)e^{s^2/2}$. On veut montrer que (6.1.6) reste vraie pour s complexe. En développant en série entière de s les deux membres de cette égalité, on peut facilement justifier que

$$\int_{-\infty}^{+\infty} e^{sx} g(x) dx = \sum_n \frac{s^n}{n!} \int_{-\infty}^{+\infty} x^n g(x) dx \quad ; \quad e^{s^2/2} = \sum_n \frac{s^{2n}}{2^n n!}.$$

En identifiant les coefficients de s^n , on en déduit 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 calcul direct des intégrales. On déduit de ces résultats que si s est complexe, les développements des deux membres de (6.1.6) sont encore égaux, et donc que

$$\phi_X(u) = e^{-u^2/2}. \quad (6.1.7)$$

En effet, on identifie les développements de Taylor de chaque membre. (Le rayon de convergence de la série entière est infini).

6) **X suit une loi normale $\mathcal{N}(m, \sigma^2)$**

On sait qu'une variable aléatoire X admettant cette loi s'écrit $X = m + \sigma Y$, où Y suit la loi $\mathcal{N}(0, 1)$. D'après (6.1.3) et (6.1.7) on a donc

$$\phi_X(u) = e^{i u m - u^2 \sigma^2 / 2}. \quad (6.1.8)$$

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 6.1.4 *La fonction caractéristique ϕ_X caractérise la loi de la variable aléatoire X . Ainsi, si deux variables ont même fonction caractéristique, elles ont même loi.*

La preuve de ce résultat est difficile et on ne la donnera pas ici.

Corollaire 6.1.5 *Soit $X = (X_1, \dots, X_n)$ une variable aléatoire à valeurs dans $\mathbb{R}^n$. Les composantes X_i sont indépendantes si et seulement si pour tous $u_1, \dots, u_n \in \mathbb{R}$ on a*

$$\phi_X(u_1, \dots, u_n) = \prod_{j=1}^n \phi_{X_j}(u_j). \quad (6.1.9)$$

Preuve. On a $\langle u, X \rangle = \sum_{j=1}^n u_j X_j$. Si les X_i sont indépendantes et comme $e^{i\langle u, X \rangle} = \prod_j e^{iu_j X_j}$, on obtient immédiatement (6.1.5), en utilisant (4.9.64).

Supposons inversement qu'on ait (6.1.5). On peut 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 $X' = (X'_1, \dots, X'_n)$ on a donc, en utilisant la condition nécessaire et (6.1.5) que $\phi_{X'} = \phi_X$. Ainsi, X et X' ont même loi, ce qui entraîne que pour tous boréliens A_j on a

$$P\left(\bigcap_j \{X_j \in A_j\}\right) = P\left(\bigcap_j \{X'_j \in A_j\}\right) = \prod_j P(X'_j \in A_j) = \prod_j P(X_j \in A_j),$$

d'où l'indépendance cherchée. $\square$

La transformée de Laplace : Lorsque X est une variable aléatoire à valeurs dans $\mathbb{R}_+$, on définit sa transformée de Laplace par

$$\psi_X(\lambda) = E(e^{-\lambda X}), \quad \lambda \in \mathbb{R}_+. \quad (6.1.10)$$

C'est une fonction définie sur $\mathbb{R}_+$, indéfiniment dérivable sur $]0, \infty[$, et formellement on a $\psi_X(\lambda) = \phi_X(i\lambda)$; 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 P_X .

Si de plus X est une variable aléatoire à valeurs dans $\mathbb{N}$, de fonction génératrice G_X , alors on a $\psi_X(\lambda) = G_X(e^{-\lambda})$.

Exemple 6.1.6 Transformée de Laplace d'une loi gamma. Soit X une variable aléatoire de loi $\Gamma(\alpha, \theta)$. Alors,

$$\psi_X(\lambda) = \frac{1}{\Gamma(\alpha)} \theta^\alpha \int_0^{+\infty} e^{-\lambda x} x^{\alpha-1} e^{-\theta x} dx = \frac{1}{\Gamma(\alpha)} \theta^\alpha \int_0^{+\infty} e^{-(\lambda+\theta)x} x^{\alpha-1} dx = \frac{\theta^\alpha}{(\lambda + \theta)^\alpha}. \quad (6.1.11)$$

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 variables aléatoires indépendantes

Proposition 6.1.7 *Si X et Y sont deux variables aléatoires indépendantes à valeurs dans $\mathbb{R}^n$, la fonction caractéristique de la somme $X + Y$ est donnée par*

$$\phi_{X+Y} = \phi_X \phi_Y. \quad (6.1.12)$$

Preuve. Comme $e^{i\langle u, X+Y \rangle} = e^{i\langle u, X \rangle} e^{i\langle u, Y \rangle}$, il suffit d'appliquer (4.9.64). $\square$

Exemple 6.1.8 Soit X, Y indépendantes, et $Z = X + Y$:

- 1) X suit loi normale $\mathcal{N}(m, \eta^2)$ et Y suit une loi $\mathcal{N}(m', (\eta')^2)$: alors Z suit une loi $\mathcal{N}(m + m', \eta^2 + (\eta')^2)$;

Cela découle de (6.1.8) et (6.1.12).

- 2) 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 (6.1.12).

- 3) 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 (6.1.12).

6.1.5 Fonction caractéristique et moments

Proposition 6.1.9 *Soit X un vecteur aléatoire de $\mathbb{R}^n$. Si la variable aléatoire $|X|^m$ (où $|X|$ désigne la norme euclidienne du vecteur X) est dans L^1 pour un entier m , la fonction ϕ_X est m fois continûment différentiable sur $\mathbb{R}^n$, et on a pour tout choix des indices $i_1, \dots, i_m$:*

$$\frac{\partial^m}{\partial u_{i_1} \partial u_{i_2} \dots \partial u_{i_m}} \phi_X(u) = i^m E(e^{i\langle u, X \rangle} X_{i_1} X_{i_2} \dots X_{i_m}) \quad (6.1.13)$$

(les X_j sont les composantes de X).

En prenant $u = 0$ ci-dessus, cette formule permet de calculer $E(X_{i_1} X_{i_2} \dots X_{i_m})$ en fonction des dérivées à l'origine de ϕ_X . Par exemple, si X est à valeurs réelles, on a

$$E(X) = i \phi'_X(0), \quad (\text{resp. } E(X^2) = -\phi''_X(0)) \quad (6.1.14)$$

dès que $X \in L^1$ (resp. $X \in L^2$).

Preuve. On se contente du cas $m = 1$, le cas général se montrant de la même manière, par récurrence sur m . Soit $v_j = (0, \dots, 0, 1, 0, \dots, 0)$ le $j^{\text{ème}}$ vecteur de base de $\mathbb{R}^n$. On a

$$\frac{\phi_X(u + tv_j) - \phi_X(u)}{t} = E \left(e^{i\langle u, X \rangle} \frac{e^{itX_j} - 1}{t} \right). \quad (6.1.15)$$

Soit $t_p \rightarrow 0$. 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 $2|X_j|$, qui par hypothèse est dans L^1 . Donc par le théorème 5.1.3 (de Lebesgue), on en déduit que (6.1.15) converge vers $iE(e^{i\langle u, X \rangle} X_j)$ quand $t \rightarrow 0$. On en déduit que la première dérivée partielle de ϕ_X par rapport à u_j existe et est donnée par la formule (6.1.13). Enfin, on montre comme dans la proposition précédente que cette dérivée est continue. $\square$

EXERCICE 6.1.10 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. On pourra vérifier que pour tout t , $\phi(t) = \left(\phi\left(\frac{t}{\sqrt{2^n}}\right) \right)^{2^n}$, et en déduire que ϕ a bien la forme voulue.

EXERCICE 6.1.11 Montrer que si X et Y sont deux variables aléatoires positives indépendantes, alors $\psi_{X+Y} = \psi_X \psi_Y$. En déduire qu'une variable aléatoire de loi $\Gamma(n, \theta)$, pour $n \in \mathbb{N}^*$, a même loi que la somme de n variables aléatoires indépendantes de loi exponentielle de paramètre θ . En déduire que la fonction caractéristique d'une variable aléatoire de loi $\Gamma(n, \theta)$ vaut $\phi(u) = \left(\frac{\theta}{\theta - iu} \right)^n$.

6.2 Vecteurs gaussiens

Définition 6.2.1 Un vecteur aléatoire $X = (X_1, \dots, X_n)$ à valeurs dans $\mathbb{R}^n$ est appelé un **vecteur gaussien** si toute combinaison linéaire $\sum_{j=1}^n a_j X_j$ 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 X est gaussien (utiliser l'exemple 6.1.8 (1)).

Contre-exemple : Si les composantes X_i sont normales mais pas indépendantes, il se peut que X ne soit pas 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 $X = (X_1, X_2)$ n'est pas un vecteur gaussien, puisque $0 < P(X_1 + X_2 = 0) < 1$ (donc $X_1 + X_2$ ne suit pas une loi normale).

Théorème 6.2.3 *X est un vecteur gaussien si et seulement si sa fonction caractéristique s'écrit*

$$\phi_X(u) = e^{i\langle u, m \rangle - \frac{1}{2}\langle u, Cu \rangle}, \quad (6.2.16)$$

où $m \in \mathbb{R}^n$ et C est une matrice $n \times n$ symétrique non-négative; dans ce cas on a $m = E(X)$ (i.e. $m_j = E(X_j)$ pour chaque j) et C est la matrice des covariances de X .

Preuve. a) Condition suffisante : supposons (6.2.16). Pour toute combinaison linéaire $Y = \sum_j a_j X_j = \langle a, X \rangle$ on a, pour $v \in \mathbb{R}$:

$$\phi_Y(v) = \phi_X(va) = e^{iv\langle m, a \rangle - \frac{v^2}{2}\langle a, Ca \rangle},$$

donc Y suit la loi $\mathcal{N}(\langle a, m \rangle, \langle a, Ca \rangle)$.

b) Condition nécessaire : Soit C la matrice des covariances de X et m son vecteur moyenne (noter que ces quantités existent, car chaque composante X_j , étant gaussienne, est dans L^2). Si $Y = \langle a, X \rangle = \sum_{j=1}^n a_j X_j$ avec $a \in \mathbb{R}^n$, un calcul simple montre que

$$E(Y) = \langle a, m \rangle, \quad \text{Var}(Y) = \langle a, Ca \rangle.$$

Par hypothèse Y suit une loi normale, donc vu ce qui précède sa fonction caractéristique est

$$\phi_Y(v) = e^{iv\langle a, m \rangle - \frac{v^2}{2}\langle a, Ca \rangle}.$$

Mais $\phi_Y(1) = \phi_{\langle a, X \rangle}(1) = E(e^{i\langle a, X \rangle}) = \phi_X(a)$, d'où (6.2.16). $\square$

Corollaire 6.2.4 *Si X est un vecteur gaussien, ses composantes sont indépendantes si et seulement si la 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 (6.2.16) et le Corollaire 6.1.5. $\square$

Proposition 6.2.5 *Soit X un vecteur gaussien à valeurs dans $\mathbb{R}^n$, de moyenne 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 A telles que $X = m + AY$, où $Y = (Y_1, \dots, Y_n)$.*

Preuve. Comme C est une matrice symétrique non-négative (cf. proposition 4.8.7), il existe une matrice orthogonale A et une matrice diagonale Λ dont les éléments diagonaux vérifient $\lambda_j \geq 0$, et telle que la matrice de covariance de X s'écrive $C = A\Lambda^t A$. Soit $Y = {}^t A(X - m)$. Alors Y est un vecteur gaussien de covariance $C' = {}^t A C A = \Lambda$ et de moyenne nulle. Les composantes Y_j de Y répondent à la question. $\square$

Proposition 6.2.6 *Le vecteur gaussien X admet une densité sur $\mathbb{R}^n$ si et seulement si sa matrice de covariance est non-dégénérée (ou inversible, ou de valeurs propres toutes strictement positives).*

Preuve. Reprenons la preuve de la proposition précédente : les λ_j qui y paraissent sont les valeurs propres de C .

Si $\lambda_j > 0$ pour tout j , le vecteur aléatoire Y admet la densité suivante sur $\mathbb{R}^n$:

$$f_Y(y) = \prod_{j=1}^n \frac{1}{\sqrt{2\pi\lambda_j}} e^{-y_j^2/2\lambda_j}.$$

Comme $X = m + AY$ on en déduit que X admet la densité

$$f_X(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(C)}} e^{-\frac{1}{2} \langle x-m, C^{-1}(x-m) \rangle}. \quad (6.2.17)$$

Si au contraire C n'est pas inversible, il existe $a \in \mathbb{R}^n$ tel que $a \neq 0$ et $Ca = 0$. La variable aléatoire $Z = \langle X, a \rangle$ a pour variance $\langle a, Ca \rangle = 0$ et pour moyenne $z = \langle m, a \rangle$, donc $P(Z = z) = 1$. Ainsi, avec une probabilité 1, le vecteur X est dans un hyperplan H orthogonal à a , i.e. $P(X \in H) = 1$. Or, si X admettait la densité f , on aurait $P(X \in H) = \int_H f(x) dx$, donc $P(X \in H) = 0$ 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 gaussiennes $N(0, 1)$ indépendantes.

L'intérêt de cette loi provient essentiellement de son utilisation massive en statistiques, 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, on obtient immédiatement que la loi du χ^2 à d degrés de liberté est la loi $\Gamma(d/2, 1/2)$ de densité

$$C_d \exp(-y/2) y^{d/2-1} \quad (6.2.18)$$

sur $\mathbb{R}_+$, pour une constante appropriée C_d .

En utilisant la définition, on montre que si Y suit une loi de χ^2 à d degrés de liberté, alors

$$E(Y) = d \quad , \quad \text{Var}(Y) = 2d. \quad (6.2.19)$$

On peut montrer les résultats suivants.

Proposition 6.2.8 *Soit n variables aléatoires $(X_1, \dots, X_n)$ indépendantes, de même loi normale $\mathcal{N}(m, \sigma^2)$. Considérons*

$$\bar{X} = \frac{X_1 + \dots + X_n}{n}; \quad V = \frac{1}{n} ((X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2).$$

Alors

- 1) $\bar{X}$ suit une loi normale d'espérance m et de variance σ^2/n ,
- 2) $\frac{n}{\sigma^2} ((X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2)$ est une variable de χ^2 à $n - 1$ degrés de liberté,
- 3) les variables aléatoires $\bar{X}$ et V sont indépendantes.

Preuve. Comme les variables aléatoires sont indépendantes et de loi normale, le vecteur $X = (X_1, \dots, X_n)$ est un vecteur gaussien, et il est donc immédiat que $\bar{X}$ suit une loi normale d'espérance la moyenne des espérances des X_i et de variance la somme des variances divisée par n^2 .

On va simplifier le problème en se ramenant à des variables aléatoires centrées réduites. Posons pour chaque $1 \leq i \leq n$, $\xi_i = \frac{X_i - m}{\sigma}$. Il est alors facile de montrer que

$$V = \frac{\sigma^2}{n} \left(\sum_{i=1}^n \xi_i^2 - \eta_1^2 \right),$$

où $\eta_1 = \frac{\xi_1 + \dots + \xi_n}{\sqrt{n}}$.

Considérons une matrice $n \times n$ orthogonale U dont la première ligne est formée de coordonnées toutes égales à $1/\sqrt{n}$ (les autres termes étant choisis arbitrairement). Désignons par η_j ($1 \leq j \leq n$) les coordonnées du vecteur aléatoire obtenu en appliquant U au vecteur aléatoire (colonne) ξ soit

$$\begin{pmatrix} \eta_1 \\ \dots \\ \eta_n \end{pmatrix} = U \begin{pmatrix} \xi_1 \\ \dots \\ \xi_n \end{pmatrix}.$$

Alors η_1 est bien compatible avec la définition précédente. Puisque U est orthogonale, et que ξ est un vecteur gaussien formé de variables aléatoires indépendantes, il en est de même pour η , et de plus, on a que $\sum_{i=1}^n \eta_i^2 = \sum_{i=1}^n \xi_i^2$. Ceci entraîne que

$$V = \frac{\sigma^2}{n} \left(\sum_{i=1}^n \eta_i^2 - \eta_1^2 \right) = \frac{\sigma^2}{n} \sum_{i=2}^n \eta_i^2.$$

Ainsi, $\frac{n}{\sigma^2} V$ est bien une variable aléatoire de chi-deux à $n - 1$ degrés de liberté, indépendante de η_1 , et donc de $\bar{X}$. $\square$

6.3 Convergence en loi

Nous allons introduire maintenant une nouvelle notion de convergence de suites de variables aléatoires. (Reportez-vous au Chapitre 5 pour revoir les définitions 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”, mais pas que les variables aléatoires elles-mêmes sont proches. Enfin, l’on verra également que toutes les convergences introduites au Chapitre 5 entraînent la convergence en loi.

On considère des variables aléatoires X_n et X , toutes à valeurs dans le même espace $\mathbb{R}^d$, mais pouvant éventuellement être définies sur des espaces d’états différents.

Définition 6.3.1 On dit que la suite (X_n) **converge en loi** vers X , et on écrit $X_n \xrightarrow{\mathcal{L}} X$, si pour toute fonction f **continue bornée** sur $\mathbb{R}^d$ on a

$$E(f(X_n)) \rightarrow E(f(X)),$$

quand n tend vers l’infini.

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, si

$$\lim_n P(X_n = x_i) = P(X = x_i), \quad \forall 1 \leq i \leq N,$$

la suite $(X_n)_n$ converge en loi vers X .

Il suffit d’écrire que

$$E(f(X_n)) = \sum_{i=1}^N f(x_i) P(X_n = x_i).$$

En Section 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.

Cette notion de convergence est plus faible que la notion de convergence en probabilité.

Proposition 6.3.3 *Si $X_n \xrightarrow{P} X$, alors $X_n \xrightarrow{\mathcal{L}} X$.*

Ainsi, les convergences en moyenne et presque-sûre entraînent également la convergence en loi, par la proposition 5.1.4.

Preuve. Soit $f \in C_b(\mathbb{R}^d)$. D'après la proposition 5.1.8, on a $f(X_n) \xrightarrow{P} f(X)$, et comme f est bornée, la proposition 5.1.5 entraîne que $f(X_n)$ converge aussi en moyenne vers $f(X)$. Comme $|E(Y)| \leq E(|Y|)$ pour toute variable aléatoire réelle Y , on en déduit qu'à fortiori $E(f(X_n)) \rightarrow E(f(X))$. $\square$

Proposition 6.3.4 *Soit 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 $F_n(x) \rightarrow F(x)$ pour tout x en lequel F est continue.*

Notez 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 $f_{p,b} \in C_b(\mathbb{R})$ telle que

$$1_{]-\infty, b]} \leq f_{p,b} \leq 1_{]-\infty, b+1/p]}. \quad (6.3.20)$$

On a alors $E(f_{p,b}(X_n)) \rightarrow E(f_{p,b}(X))$ si $n \uparrow \infty$. De (6.3.20) on déduit d'abord que $F_n(a) = P(X_n \leq a) \leq E(f_{p,a}(X_n))$ et $E(f_{p,a}(X)) \leq F(a + \frac{1}{p})$; donc $\limsup_n F_n(a) \leq F(a + \frac{1}{p})$ pour tout p , donc aussi $\limsup_n F_n(a) \leq F(a)$. On en déduit ensuite que $F_n(a) \geq E(f_{p,a-1/p}(X_n))$ et $E(f_{p,a-1/p}(X)) \geq F(a - \frac{1}{p})$; 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 $F_n(a) \rightarrow F(a)$.

b) Inversement, supposons que $F_n(x) \rightarrow F(x)$ pour tout $x \in T$, où T est une partie dense de $\mathbb{R}$. Soit $f \in C_b(\mathbb{R})$ et $\varepsilon > 0$. Soit $a, b \in T$ avec $F(a) \leq \varepsilon$ et $F(b) \geq 1 - \varepsilon$. Il existe n_0 tel que

$$n \geq n_0 \quad \Rightarrow \quad P(X_n \notin]a, b]) = 1 - F_n(b) + F_n(a) \leq 3\varepsilon. \quad (6.3.21)$$

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

$$g(x) = \sum_{i=1}^k f(a_i) 1_{[a_{i-1}, a_i]}(x)$$

vérifie $|f - g| \leq \varepsilon$ sur $]a, b]$. Si $M = \sup_x |f(x)|$, il vient alors

$$|E(f(X_n)) - E(g(X_n))| \leq M P(X_n \notin]a, b]) + \varepsilon, \quad (6.3.22)$$

et de même pour X . Enfin $E(g(X_n)) = \sum_{i=1}^k f(a_i)(F_n(a_{i-1}) - F_n(a_i))$, et de même pour X , par définition de g . Comme $F_n(a_i) \rightarrow F(a_i)$ pour tout i , on en déduit l'existence de n_1 tel que

$$n \geq n_1 \Rightarrow |E(g(X_n)) - E(g(X))| \leq \varepsilon. \quad (6.3.23)$$

D'après (6.3.21), (6.3.22) et (6.3.23) on a

$$n \geq \sup(n_0, n_1) \Rightarrow |E(f(X_n)) - E(f(X))| \leq 3\varepsilon + 5M\varepsilon.$$

ε étant arbitraire, on en déduit que $E(f(X_n)) \rightarrow E(f(X))$, et on a le résultat. $\square$

Théorème 6.3.5 (*Théorème de Lévy*) Soit X_n des variables aléatoires à valeurs dans $\mathbb{R}^d$.

- a) Si la suite $(X_n)_n$ converge en loi vers X , alors ϕ_{X_n} converge simplement vers ϕ_X .
- b) Si les ϕ_{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 X et $X_n \xrightarrow{\mathcal{L}} X$.

Preuve. (Nous ne démontrons pas (b), qui est assez difficile).

Pour obtenir (a), il suffit de remarquer que $\phi_{X_n}(u) = E(g_u(X_n))$ et $\phi_X(u) = E(g_u(X))$, où g_u est la fonction continue bornée $g_u(x) = e^{i\langle u, x \rangle}$ et d'appliquer la définition 6.3.1 (en séparant parties réelle et imaginaire). $\square$

Exemples : 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 $\lambda_n \rightarrow \lambda$.

6.4 Le théorème de la limite centrale

Ce théorème est aussi connu sous le nom : théorème central limite. Plus simplement, il apparaît souvent sous l'abréviation TCL.

La situation est la même que dans le chapitre précédent concernant la loi des grands nombres (Théorème 5.2.1) : les X_n sont des variables aléatoires **indépendantes, de même loi, et dans L^2** . On note m et σ^2 la moyenne et la variance des X_n , et aussi

$$S_n = X_1 + \dots + X_n \quad (6.4.24)$$

(ainsi $M_n = \frac{S_n}{n}$). Nous avons vu que $\frac{S_n}{n}$ converge vers m p.s. 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 $\frac{S_n}{n} - m$, on est amené 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 . 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 universel 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, dans L^2 , 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. Soit ϕ la fonction caractéristique de $X_n - m$, et $Y_n = (S_n - nm)/\sigma\sqrt{n}$.

Comme les X_n sont indépendantes et d'après (6.1.3), la fonction caractéristique de Y_n est

$$\phi_n(u) = \phi\left(\frac{u}{\sigma\sqrt{n}}\right)^n. \quad (6.4.25)$$

Comme $E(X_n - m) = 0$ et $E((X_n - m)^2) = \sigma^2$, (6.1.13) entraîne

$$\phi(u) = 1 - \frac{u^2\sigma^2}{2} + u^2 o(|u|) \quad \text{quand } u \rightarrow 0.$$

Comme $\phi(0) = 1$ et que ϕ est continue en 0, il est facile de voir que pour u fixé et n assez grand,

$$|\phi\left(\frac{u}{\sigma\sqrt{n}}\right) - 1| \leq 1/2.$$

On peut généraliser la notion de logarithme aux complexes z tels que $|z - 1| \leq 1/2$ (cf. cours de Mathématiques). La fonction $\ln z$ définie sur le disque $\{z \in \mathcal{C} : |z - 1| \leq 1/2\}$ 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 $\varepsilon_n(u) \rightarrow 0$ quand $n \rightarrow \infty$, et on en déduit alors immédiatement que $\phi_n(u) \rightarrow \exp -u^2/2$. Le résultat découle alors du théorème 6.3.5. $\square$

Remarque : On peut déduire de ce résultat que $n^\alpha | \frac{S_n}{n} - m |$ converge vers 0 (resp. $+\infty$) en probabilité lorsque $\alpha < 1/2$ (resp. $\alpha > 1/2$).

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. $P(X_i = 1) = p$ et $P(X_i = 0) = 1 - p$. On a alors $m = p$ et $\sigma^2 = p(1 - p)$.

Nous voulons calculer $P(S_n \leq x)$ 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), on peut 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, on utilise les théorèmes précédents :

$$\frac{S_n}{n} \xrightarrow{P} p, \tag{6.4.26}$$

$$\frac{S_n - np}{\sqrt{np(1 - p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1). \tag{6.4.27}$$

Π désignant la fonction de répartition de la loi $\mathcal{N}(0, 1)$, on a

$$P(S_n \leq x) \simeq \Pi \left(\frac{x - np}{\sqrt{np(1 - p)}} \right). \tag{6.4.28}$$

Rappelons que la fonction de répartition Π ci-dessus est "tabulée", et une table de cette fonction est fournie au chapitre 4.6.

Exemple 6.4.3 On lance 1000 fois une pièce (que l'on suppose non truquée). Quelle est la probabilité d'obtenir plus de 545 fois le côté Face ?

Solution : Bien-sûr, le calcul exact utilisant les lois binomiales serait rébarbatif et atroce. On écrit :

$$P(S_n > 545) = P\left(\frac{S_n - n/2}{\sqrt{1000}/2} > \frac{45}{\sqrt{1000}/2}\right) \simeq \int_{\frac{90}{\sqrt{1000}}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

En utilisant la table numérique, on obtient

$$P(S_n > 545) \simeq 1 - \Pi(2, 84) \simeq 0,0023.$$

Le théorème 6.4.1 admet une version multidimensionnelle, de preuve similaire. On suppose que les X_n sont des vecteurs aléatoires à valeurs dans $\mathbb{R}^d$, indépendants et de même loi. On suppose que les composantes des X_n sont dans L^2 . On a ainsi un vecteur moyenne $m = E(X_n)$, et une matrice de covariance $C = (c_{ij})$ avec c_{ij} = la covariance des composantes i et j de X_n . On a alors le :

Théorème 6.4.4 *Les vecteurs aléatoires $\frac{S_n - nm}{\sqrt{n}}$ convergent en loi vers un vecteur aléatoire gaussien centré (i.e. de moyenne nulle), de matrice de covariance C .*

Remarquons ici que la vitesse de convergence est toujours en $\frac{1}{\sqrt{n}}$, indépendante de la dimension d .

6.5 Intervalles de confiance

Dans les deux exemples suivants, nous allons voir comment le Théorème de loi des grands nombres et le Théorème de la limite centrale permettent d'**estimer** un paramètre inconnu, et de dire à un pourcentage d'erreurs près, fixé a priori, dans quelle fourchette d'estimation se trouve ce paramètre.

6.5.1 Sondages

A la veille d'une élection portant sur le choix entre deux candidats, on effectue un sondage afin de déterminer la proportion p de votes pour un des candidats, appelé AA. 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 AA, $X_i = 0$ sinon. En pratique, on évite d'interroger deux fois un même individu (tirage sans remise), de sorte que *a priori*, les X_i sont dépendants. Mais nous avons vu au chapitre 2.2.3 que lorsque la taille de la population totale est grande, il y a peu de différence entre les

tirages avec remise et sans remise. Nous nous placerons donc dans cette hypothèse, et supposerons ainsi les X_i indépendantes. Il faut bien comprendre qu'ici, l'espace d'états Ω est l'ensemble de tous les échantillons de n individus, et que choisir un échantillon est un choix particulier ω de Ω .

Les variables aléatoires X_i , $i \leq n$, définies sur Ω , sont supposées indépendantes et de même loi de Bernoulli $P(X_i = 1) = p = 1 - P(X_i = 0)$. Nous nous trouvons avec le problème statistique suivant : **comment estimer le paramètre p inconnu au vu des observations $X_1, \dots, X_n$?** Intuitivement, aucune information sur p n'est contenue dans l'ordre des réponses, et il suffit de résumer le sondage $X_1, \dots, X_n$ par le nombre d'intentions de vote $S_n = X_1 + \dots + X_n$. Ce raisonnement heuristique peut d'ailleurs se quantifier ainsi. La loi conditionnelle des observations $(X_1, \dots, X_n)$ sachant $S_n = k$ ne dépend pas du paramètre inconnu p (c'est la probabilité uniforme sur les $\binom{n}{k}$ suites $(x_1, \dots, x_n)$ comportant k "uns" et $n - k$ "zéros"). Par conséquent, la valeur de S_n contient toute l'information contenue dans le sondage sur p . On cherche alors à estimer p en utilisant une fonction $\hat{X}_n$ de S_n , appelée **estimateur** de p , et le choix naturel est la proportion $\hat{X}_n$ de vote pour AA dans l'échantillon,

$$\hat{X}_n = \frac{S_n}{n}.$$

Cet estimateur est " bon " en moyenne, au sens où son espérance est égale au paramètre que l'on cherche à estimer : $E(\hat{X}_n) = p$, on dit alors qu'il est **sans biais**. La loi des grands nombres nous apprend qu'il est " bon " lorsque l'échantillon est de grande taille puisque

$$\hat{X}_n \xrightarrow[n \rightarrow \infty]{} p$$

en probabilité et presque-sûrement ; l'estimateur converge vers le paramètre à estimer.

Le sondage donne 1300 intentions de votes pour AA, et 1200 pour son adversaire. La valeur de l'estimateur $\hat{X}_n$, calculée sur notre échantillon particulier (un " ω " particulier), appelée estimation $\hat{p}_n$ de p , prend la valeur 0,52. Peut-on quantifier la confiance à donner à cette estimation numérique ? Plus précisément, cette valeur 0,52 est-elle significativement différente de 0,5 ?

Il faut tenir compte des possibilités de fluctuations dues à l'échantillonnage : si on change d'échantillon, il est clair que la valeur de l'estimation va changer (on peut imaginer que sur un autre échantillon de 2500 personnes interrogées, l'estimation vaille 0,49). Toutefois, on connaît la loi de la variable aléatoire $\hat{X}_n$ définie sur l'ensemble Ω des échantillons de taille n . La réponse dépend alors de l'amplitude ε des fluctuations de $\hat{X}_n$ autour de p , et comme la taille n de l'échantillon est grande, nous pouvons alors utiliser l'approximation gaussienne (Théorème 6.4.1)

$$P(|\hat{X}_n - p| \leq \varepsilon) = P\left(\frac{\sqrt{n}|\hat{X}_n - p|}{\sqrt{p(1-p)}} \leq \frac{\sqrt{n}\varepsilon}{\sqrt{p(1-p)}}\right) \simeq \int_{-a}^a g(x)dx,$$

où g est la densité de la loi normale centrée réduite et $a = \frac{\sqrt{n\varepsilon}}{\sqrt{p(1-p)}}$.

Cette probabilité sera environ égale à 0,95, si $a = 1.96$ et à 0,99 si $a = 2.6$. (Voir la table numérique de la fonction de répartition de la loi normale, Section 4.6)

Ce résultat nous dit que l'erreur $\hat{p}_n - p$ commise, en prenant $\hat{p}_n$ comme approximation de p ne dépassera pas, avec une probabilité de 95%, resp. de 99% environ, le seuil

$$\frac{1.96\sqrt{p(1-p)}}{\sqrt{n}}, \quad \text{resp.} \quad \frac{2.6\sqrt{p(1-p)}}{\sqrt{n}}.$$

Ce seuil dépend malencontreusement du paramètre p inconnu, mais la fonction $\sqrt{p(1-p)} = \sigma(p)$ est majorée par sa valeur maximale $1/2$, de sorte qu'en remplaçant le facteur $\sqrt{p(1-p)}$ par $1/2$ dans les seuils précédents, on ne fera qu'augmenter notre quasi-certitude. Plus généralement, si on se donne *a priori* la probabilité α que l'erreur $\hat{p}_n - p$ ne dépasse pas un seuil donné, ce seuil sera égal à $\frac{s_\alpha}{2\sqrt{n}}$, où s_α est tel que $2(1 - \Pi(s_\alpha)) = \alpha$.

En conclusion, nous obtenons le résultat suivant (en remplaçant 1,96 par 2).

Proposition 6.5.1 *Intervalle de confiance pour l'estimation de p .*

Soit $\alpha \in]0, 1[$, proche de 1 dans la pratique et appelé le niveau de confiance asymptotique de l'intervalle. Dès que n est assez grand ($n \geq 50$, np et $n(1-p) \geq 5$ en pratique), l'intervalle

$$I_\alpha = [\hat{p}_n - \frac{s_\alpha}{2\sqrt{n}}; \hat{p}_n + \frac{s_\alpha}{2\sqrt{n}}],$$

où s_α est tel que $\int_{-s_\alpha}^{s_\alpha} g(x)dx = \alpha$, est l'intervalle de confiance de p de niveau asymptotique α , au sens où

$$\lim_n P \left(p \in \left[\hat{X}_n - \frac{s_\alpha}{2\sqrt{n}}; \hat{X}_n + \frac{s_\alpha}{2\sqrt{n}} \right] \right) = \alpha.$$

Ainsi

$$\left[\hat{p}_n - \frac{1}{\sqrt{n}}; \hat{p}_n + \frac{1}{\sqrt{n}} \right] \tag{6.5.29}$$

est l'intervalle de confiance de niveau 95% et

$$\left[\hat{p}_n - \frac{1,3}{\sqrt{n}}; \hat{p}_n + \frac{1,3}{\sqrt{n}} \right] \tag{6.5.30}$$

celui de niveau 99%.

Dans notre exemple, **l'intervalle de confiance à 95% pour p** est $[0, 50; 0, 54]$. Les instituts de sondage annoncent d'ailleurs leurs résultats ainsi (sous forme de fourchette d'estimation).

Remarque 6.5.2 On dit souvent que l'intervalle $\left[\hat{X}_n - \frac{s_\alpha}{2\sqrt{n}}; \hat{X}_n + \frac{s_\alpha}{2\sqrt{n}} \right]$ est l'intervalle de confiance de p de niveau asymptotique α . Il faut bien faire attention que cet intervalle est aléatoire.

6.5.2 Précision dans la méthode de Monte-Carlo

Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction intégrable, telle que le calcul analytique de $I = \int_{\mathbb{R}^d} f(x)dx$ soit compliqué ou impossible. Nous avons vu alors, chapitre 5.3, une méthode probabiliste par **simulation**. Comparée aux méthodes d'intégration numériques usuelles, la méthode de Monte-Carlo est surtout intéressante en dimension d grande.

Soit p une densité de probabilité sur $\mathbb{R}^d$, dont le support contient celui de f ($f(x) \neq 0 \Rightarrow p(x) > 0$). Alors $I = \int_{\mathbb{R}^d} f(x)dx = \int_{\mathbb{R}^d} \frac{f(x)}{p(x)} p(x)dx$, et I s'écrit donc comme

$$I = E(\varphi(X)) \quad \text{avec} \quad \varphi = f/p \quad , \quad X \text{ de densité } p .$$

Par hypothèse, f est intégrable, cela entraîne que $\varphi(X) \in L^1$.

Si p est une densité facilement simulable, on sait générer des v.a. $X_1, \dots, X_n$ indépendantes et de densité p . Nous avons vu que la moyenne

$$\hat{I}_n = \frac{1}{n}(\varphi(X_1) + \dots + \varphi(X_n))$$

constitue une approximation (en fait, une estimation au sens statistique du terme) de I . La loi des grands nombres entraîne que $\hat{I}_n$ converge vers I (en probabilité, et presque-sûrement) lorsque $n \rightarrow \infty$, et le théorème de la limite centrale nous renseigne, lorsque $E(\varphi(X_1)^2) < \infty$, 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 avec un niveau de confiance de 95%, de la forme

$$\left[\hat{I}_n - \frac{1,96\sigma}{\sqrt{n}} \quad , \quad \hat{I}_n + \frac{1,96\sigma}{\sqrt{n}} \right] .$$

Cela veut dire que la probabilité que la simulation obtenue ω ne donne pas un résultat appartenant à cet intervalle vaut 0,05.

Ici,

$$\sigma^2 = \text{Var}(\varphi(X_1)) = \int_{\mathbb{R}^d} \frac{f(x)^2}{p(x)} dx - I^2 \tag{6.5.31}$$

est le plus souvent inconnu, de sorte qu'il faut estimer sa valeur elle aussi. On définit alors l'estimateur

$$\hat{\Sigma}_n = \left(\frac{1}{n} \sum_{i=1}^n \varphi(X_i)^2 - \hat{I}_n^2 \right)^{1/2} ,$$

qui vérifie $\hat{\Sigma}_n \rightarrow \sigma$ d'après la loi des grands nombres. On utilise alors le théorème de Slutsky, que l'on donne ci-après sans démonstration (Théorème 6.5.3), pour montrer que l'on peut alors considérer l'intervalle de confiance

$$\left[\hat{I}_n - \frac{1,96\hat{\Sigma}_n}{\sqrt{n}} \quad , \quad \hat{I}_n + \frac{1,96\hat{\Sigma}_n}{\sqrt{n}} \right] , \tag{6.5.32}$$

où l'on a remplacé σ par l'estimateur $\hat{\Sigma}_n$. On peut vérifier que cet intervalle (aléatoire) constitue encore un intervalle de confiance pour I de niveau 95%. Si on connaît les valeurs numériques de $\hat{I}_n$ et $\hat{\Sigma}_n$, calculées à partir des valeurs simulées, on obtiendra la fourchette d'estimation de la méthode.

Théorème 6.5.3 *Soit $(X_n)_n$ et $(Y_n)_n$ deux suites de vecteurs aléatoires. Supposons que (X_n) converge en loi vers X et que $|X_n - Y_n|$ converge vers 0 en probabilité. Alors (Y_n) converge en loi vers X .*

Chapitre 7

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 développons ici quelques exemples de phénomènes aléatoires qui évoluent au cours du temps.

Ce chapitre, fondamental pour les applications, peut être vu comme une introduction au cours "Promenade aléatoire" de deuxième année.

7.1 Marche aléatoire simple

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 $X_n = -1$ avec probabilité $1 - p$. Les variables X_n , $n \geq 1$, sont supposées indépendantes, et la fortune initiale du joueur est une constante, $S_0 = k$. Donc

$$S_n = k + \sum_{i=1}^n X_i \quad (7.1.1)$$

est somme de n variables aléatoires indépendantes et de même loi, encore appelée **marche aléatoire**.

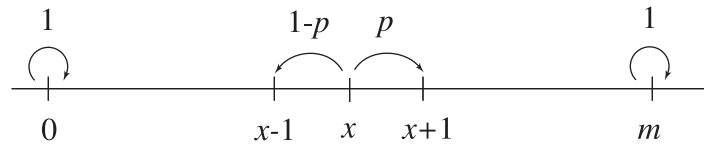
En réalité, dans le problème particulier de la ruine du joueur, le jeu s'arrête dès que le joueur est ruiné, ou que son adversaire l'est. Le modèle (7.1.1) de la marche aléatoire décrit aussi le mouvement d'une particule sur l'ensemble $\mathbb{Z}$ des entiers relatifs, qui se déplace à chaque instant entier $n = 1, 2, \dots$, d'un pas vers la droite avec probabilité p et d'un

pas vers la gauche avec probabilité $1 - p$, indépendamment des déplacements précédents. Cette promenade aléatoire peut être plus compliquée, avec des sauts X_i à valeurs réelles, ou encore à valeurs dans $\mathbb{R}^d$ auquel cas la particule se déplace dans l'espace à d dimensions. Les marches aléatoires peuvent modéliser aussi le nombre de clients dans une file d'attente (les instants entiers n correspondent alors aux instants de départ ou d'arrivée des clients), la position d'un grain de pollen en suspension, ou la valeur de l'indice CAC 40 tous les lundis matins.

Dans l'étude des réactions nucléaires, on s'intéresse à l'histoire d'une particule (comme un neutron), qui se déplace selon une promenade aléatoire dans un conteneur, dont les parois sont constituées d'un matériau absorbant. La promenade aléatoire s'arrête dès que la particule touche le bord du conteneur, où elle est absorbée.

Revenons au problème de la ruine du joueur. En réalité, le jeu s'arrête dès que le joueur est ruiné ($S_n = 0$) ou que son adversaire l'est ($S_n = m$, si on désigne par $m - k$ la fortune initiale de l'adversaire). La formule (7.1.1) n'est vraie que jusqu'à la fin du jeu, après quoi S_n reste constant.

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 $\mu_k = P_k(R)$ lorsque le joueur démarre avec une fortune initiale k (l'indice k de P_k fait référence au point de départ de S_n). D'après les formules de conditionnement (voir Chapitre 2),

$$P_k(R) = P_k(R | X_1 = +1) P(X_1 = +1) + P_k(R | X_1 = -1) P(X_1 = -1) . \quad (7.1.2)$$

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

$$\mu_k = p \mu_{k+1} + (1 - p) \mu_{k-1} \quad (7.1.3)$$

pour $k \in \{1, \dots, m - 1\}$, avec les conditions aux limites $\mu_0 = 1$, $\mu_m = 0$. L'équation (7.1.3) 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 (7.1.3) sont de la forme $\mu_k = \alpha r_1^k + \beta r_2^k$, et on détermine α et β par les conditions aux limites. on obtient

$$P_k(R) = \frac{\left(\frac{1-p}{p}\right)^m - \left(\frac{1-p}{p}\right)^k}{\left(\frac{1-p}{p}\right)^m - 1} \quad , \quad 0 \leq k < m \quad (7.1.4)$$

Procédant de même pour l'événement G : " le joueur finit par gagner ", on constate que $P_k(G)$ satisfait (7.1.3), avec d'autres conditions aux limites ($\mu_0 = 0$, $\mu_m = 1$). On trouve alors que $P_k(G) = 1 - P_k(R)$, si bien que le jeu s'arrête en un temps fini, avec probabilité 1.

- ii) $p = 1/2$ (**jeu équitable**). Alors $r_1 = r_2 = 1$, les solutions de (7.1.3) sont de la forme $\mu_k = \alpha + \beta k$, et tenant compte des conditions limites on obtient

$$P_k(R) = 1 - \frac{k}{m} \quad , \quad P_k(G) = \frac{k}{m} \quad , \quad (7.1.5)$$

et ici encore le jeu finit par s'arrêter.

Durée du jeu : Déterminons la durée moyenne du jeu, $V_k = E_k(\nu)$ avec

$$\nu = \min\{k \geq 1 : S_n = 0 \text{ ou } m\},$$

et E_k désigne l'espérance sous P_k . Comme dans (7.1.2), on conditionne suivant les valeurs de X_1

$$E_k(\nu) = E_k(\nu | X_1 = +1) P(X_1 = +1) + E_k(\nu | X_1 = -1) P(X_1 = -1)$$

et on remarque que cette fois $E_k(\nu | X_1 = \pm 1) = V_{k \pm 1} + 1$ puisqu'une unité de temps s'est écoulée. On obtient donc l'équation de récurrence linéaire avec second membre

$$V_k = p(1 + V_{k+1}) + (1 - p)(1 + V_{k-1}) \quad (7.1.6)$$

pour $1 \leq k \leq m - 1$, avec $V_0 = V_m = 0$. On résout cette équation en trouvant solutions générales et solution particulière :

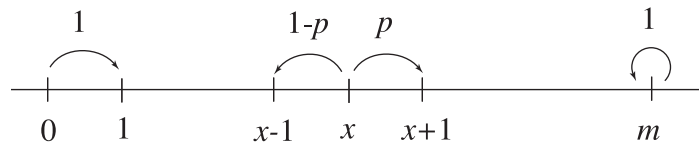
$$\begin{aligned} E_k(\nu) &= \frac{1}{1 - 2p} \left(k - m \frac{1 - \left(\frac{1-p}{p}\right)^k}{1 - \left(\frac{1-p}{p}\right)^m} \right) && \text{si } p \neq 1/2 , \\ &= k(m - k) && \text{si } p = 1/2 . \end{aligned}$$

Barrière réfléchissante : Dans le problème précédent, on suppose 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

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

TAB. 7.1 – Le destin du joueur

joueur, qui décide d'arrêter le jeu. On dit ici que le point 0 est réfléchissant, tandis que le point m reste absorbant. On a 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, 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 (7.1.4) ; ici $p < 1/2$. Son destin est résumé dans le tableau ci-dessous. 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 .

EXERCICE 7.1.1 Montrer que la durée moyenne du jeu (avec barrière réfléchissante) vérifie encore (7.1.6), mais avec d'autres conditions aux limites : $V_m = 0$, $V_0 = 1 + V_1$. Déterminer V_k .

EXERCICE 7.1.2 On considère la marche aléatoire S_n donnée par (7.1.1). On définit les événements

$$A_m = \{\exists n \mid S_n = 0 \text{ , } S_i < m \quad \forall i < n\}$$

où la marche visite 0 avant m ($m \geq k + 1$), et

$$A = \{\exists n \mid S_n = 0\}$$

où la marche finit par visiter 0.

- a) Montrer que la probabilité $P(A_m)$ est donnée par (7.1.4), (7.1.5), selon que $p \neq 1/2$ ou $p = 1/2$.
- b) Montrer que $A_m \subset A_{m+1}$ et calculer la probabilité $P(\bigcup_m A_m)$. (Indication : utiliser a) et la proposition 2.3.8).
- c) Montrer que $A = \bigcup_m A_m$. (Indication : on pourra remarquer que $\{S_n = 0\} \subset A_{n+k+1}$). En déduire que

$$P(A) = 1 \quad \text{ou} \quad P(A) < 1$$

selon que

$$p \leq 1/2 \quad \text{ou} \quad p > 1/2 .$$

7.2 Processus de branchement

Ces modèles sont absolument fondamentaux en Ecologie et 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.

7.2.1 Sommes aléatoires de v.a. indépendantes

Il arrive fréquemment, dans la modélisation probabiliste, que l'on ait à considérer des sommes de variables aléatoires dont le nombre de termes lui-même soit aléatoire.

On peut par exemple considérer

- a) le nombre de filles d'une famille $\sum_{i=1}^{\nu} 1_{F_i}$ 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) ;
- b) 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 7.2.1 *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*

$$g(x) = E(x^{X_n}) \quad (x \in [0, 1], n \geq 1)$$

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

$$G(x) = E(x^{\nu}) ,$$

alors la fonction génératrice de la somme

$$S_\nu = \sum_{m=1}^{\nu} X_m \quad (S = 0 \text{ si } \nu = 0) \quad (7.2.7)$$

des ν premières variables aléatoires X_m , est la fonction composée de g et G , soit

$$E(x^{S_\nu}) = G \circ g(x) \quad (\forall x, \quad x \in [0, 1]) \quad (7.2.8)$$

Preuve. Soient $S_n = \sum_{m=1}^n X_m$, ($n \geq 1$), les sommes partielles des X_m . On suppose par convention que $S_0 = 0$. Alors la variable aléatoire S s'écrit explicitement sur Ω

$$S(\omega) = S_{\nu(\omega)}(\omega) \quad (7.2.9)$$

et dépend donc de deux manières de ω .

Comme les ensembles $(\{\nu = n\}, n \in \mathbb{N})$ forment une partition de Ω , on peut écrire

$$\begin{aligned} E(x^{S_\nu}) &= \sum_n E(x^{S_\nu} \mathbf{1}_{\{\nu=n\}}) = \sum_n E(x^{S_n} \mathbf{1}_{\{\nu=n\}}) \\ &= \sum_n E(x^{S_n}) P(\nu = n) \quad (\text{par indépendance de } S_n \text{ et de } \nu) \\ &= \sum_n g(x)^n P(\nu = n) \quad (\text{par indépendance des } X_i, i \leq n) \\ &= G(g(x)). \end{aligned}$$

□

Exemple 7.2.2 Sous les hypothèses d'indépendance précédentes, si ν suit une loi de Poisson de paramètre θ et les X_m suivent une loi de Bernoulli de paramètre p , on a

$$E(x^\nu) = \exp(\theta(x - 1)) \quad \text{et} \quad E(x^{X_m}) = xp + (1 - p)$$

de sorte que

$$E(x^{S_\nu}) = \exp(\theta(xp + (1 - p) - 1)) = \exp(\theta p(x - 1)),$$

ce qui montre que S_ν suit une loi de Poisson de paramètre θp .

EXERCICE 7.2.3 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 $a > 1$ ($P(N = k) = \frac{1}{a}(1 - \frac{1}{a})^k$), 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 7.2.4 Dédurre de la proposition 7.2.1 que pour des variables aléatoires X_i et N de carré intégrable,

$$E(S_N) = E(N)E(X) \quad , \quad \text{Var}(S_N) = \text{Var}(N)(E(X))^2 + E(N)\text{Var}(X) \quad ,$$

où $E(X)$ et $\text{Var}(X)$ désignent respectivement les espérance et variance communes aux X_i . Comment s'écrivent ces formules lorsque N est déterministe ?

7.2.2 Processus de branchement

La première étude sur les processus de branchement a été motivée par l'étude de la survivance des noms des familles (de nobles anglais), et donc par l'étude de la non-extinction des descendants mâles qui seuls transmettaient le nom. Ces processus sont connus sous le nom de processus de 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 décrit aussi bien la croissance d'une population de cellules, ou du nombre de neutrons dans un réacteur, ou encore d'une épidémie dans une population.

On étudie 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 ξ_i^n d'enfants ($1 \leq i \leq Z_n$) de sorte que

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i^n \quad (n \geq 0). \quad (7.2.10)$$

Les variables aléatoires ξ_i^n sont supposées indépendantes entre elles et de même loi de probabilité caractérisée par la fonction génératrice $g(x) = E(x^\xi)$. Dans toute la suite, nous excluons le cas inintéressant où ξ serait égale à 1 avec probabilité un, c'est-à-dire le cas où $g(x) \equiv x$.

La figure 7.1 représente les individus des générations $0, \dots, 3$, lorsque $Z_0 = 1$, $\xi_1^0 = 2$, $\xi_1^1 = 3$, $\xi_2^1 = 1$, $\xi_1^2 = 2$, $\xi_2^2 = 0$, $\xi_3^2 = 1$, $\xi_4^3 = 3$. Cette figure représente un arbre aléatoire.

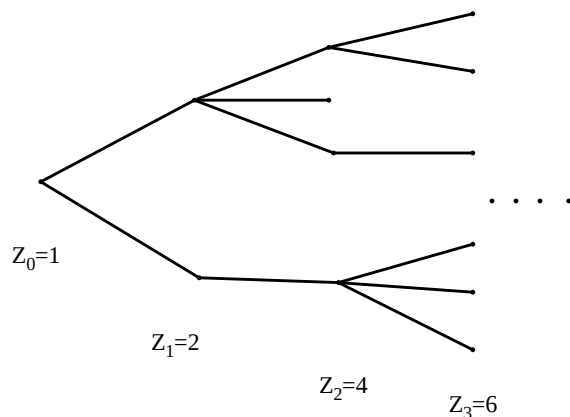


FIG. 7.1 – L'arbre généalogique du processus de branchement

La Proposition 7.2.1 montre alors que la fonction génératrice $G_n(x) = G_{Z_n}(x)$ de Z_n vérifie

$$G_{n+1}(x) = G_{Z_{n+1}}(x) = G_n(g(x)) \quad (n \geq 0) \quad (7.2.11)$$

et si nous supposons que $Z_0 = 1$ de sorte que $G_0(x) \equiv x$, nous trouvons que pour tout $n \geq 1$:

$$G_n(x) = g_n(x) \quad \text{où} \quad g_n = g \circ g \circ \dots \circ g \quad n \text{ fois} . \quad (7.2.12)$$

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

$$A = \cup_n \uparrow \{Z_n = 0\} \quad (7.2.13)$$

et sa probabilité donnée par

$$P(E) = \lim_n \uparrow g_n(0) .$$

(puisque $P(Z = 0) = G_Z(0)$ pour toute v.a. Z entière positive).

Deux cas se présentent, comme nous allons le voir dans le théorème suivant.

Théorème 7.2.5 *Si le nombre moyen d'enfants $E(\xi)$ de chaque individu est inférieur ou égal à 1, la population s'éteint presque sûrement : $P(A) = 1$.*

Si $E(\xi) > 1$, la probabilité d'extinction $P(A)$ est égale au nombre positif v strictement inférieur à 1, solution unique de $g(v) = v$ dans $[0, 1[$.

Remarque 7.2.6 Par indépendance des variables $(\xi_i^n; n \geq 1, i \geq 1)$ on déduit de (7.2.10) et de l'exercice 7.2.4 que $E(Z_n) = E(\xi) \times E(Z_{n-1}) = (E(\xi))^n$ (par récurrence). Cette dernière égalité entraîne que la population s'éteint lorsque $E(\xi) < 1$. En effet, il est facile de voir que

$$P(Z_n \neq 0) \leq E(Z_n) = (E(\xi))^n . \quad (7.2.14)$$

Preuve. La démonstration du théorème repose sur le lemme analytique suivant.

Lemme 7.2.7 *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

$$g'(x) = E(\xi x^{\xi-1}) \quad \text{et} \quad g''(x) = E(\xi(\xi-1)x^{\xi-2})$$

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

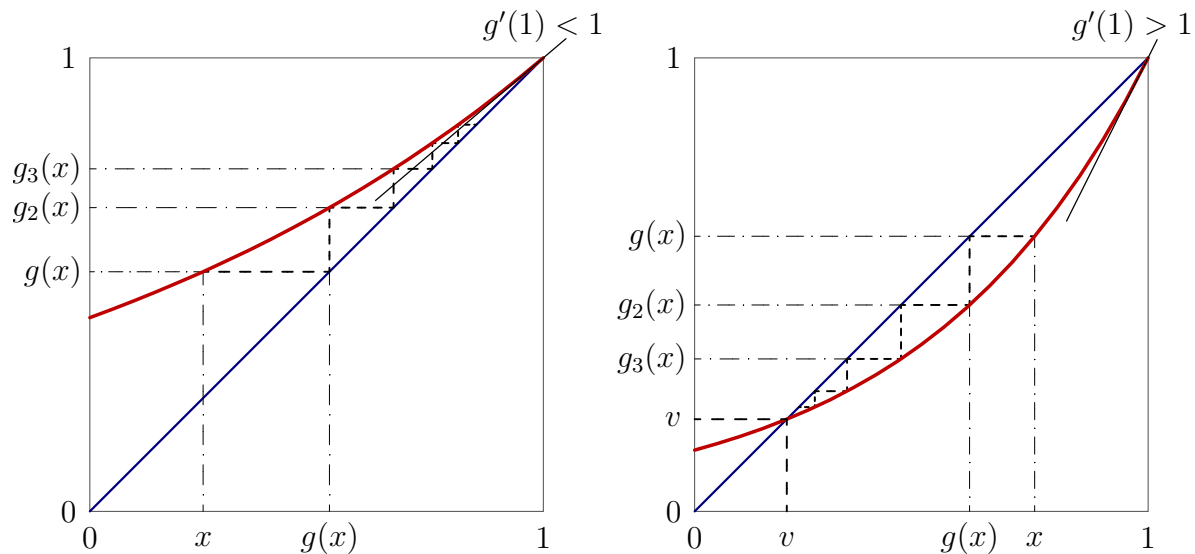


FIG. 7.2 – La fonction g et ses itérées dans les deux cas $E(\xi) \leq 1$ et $E(\xi) > 1$

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 au contraire que $g'(1) > 1$; ceci se voit bien sur la figure 7.2. Ainsi, selon le cas, l'équation de point fixe $g(v) = v$ n'a pas de solution ou au contraire possède une unique solution dans $[0, 1[$.

- 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 $E(\xi) = g'(1) \leq 1$, alors, grâce à (7.2.13),

$$P(A) = \lim_n g_n(0) = 1 .$$

2) si $E(\xi) = g'(1) > 1$, alors

$$P(A) = \lim_n g_n(0) = v .$$

□

Remarque 7.2.8 Lorsque $E(\xi) > 1$, 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 $P(Z_n \rightarrow \infty | A^c) = 1$.

En effet, pour tout entier $a > 0$ et quel que soit x que l'on se sera fixé dans $]0, 1[$, on obtient, en utilisant l'inégalité de Markov (4.7.48),

$$\begin{aligned} P(0 < Z_n \leq a) &\leq x^{-a} \sum_{k \geq 1} x^k P(Z_n = k) \\ &= x^{-a} (g_n(x) - g_n(0)) \rightarrow 0 \quad \text{lorsque } n \rightarrow \infty. \end{aligned}$$

Il en résulte déjà la propriété plus faible que $P(0 < Z_n \leq a | A^c) \xrightarrow[n \rightarrow \infty]{} 0$, $\forall a > 0$.

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 $g'(v) < 1$, on a $\sum_n P(0 < Z_n \leq a) < \infty$, et le lemme de Borel-Cantelli (Théorème 2.4.18) 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 $E(\xi) < 1$, la probabilité de "non extinction à l'instant n " tend vers 0 à une vitesse géométrique puisque, comme on l'a vu,

$$P(Z_n \neq 0) \leq (E(\xi))^n.$$

- b) lorsque $E(\xi) = 1$, cette même probabilité tend beaucoup plus lentement vers zéro. Plus précisément, on peut montrer que si $E(\xi^2) < \infty$,

$$P(Z_n \neq 0) \sim c/n \quad \text{lorsque } n \rightarrow \infty$$

avec $c = 2/E(\xi(\xi - 1))$.

- c) lorsque $E(\xi) > 1$, la taille de la population Z_n explose à une vitesse géométrique lorsqu'il n'y a pas extinction, du moins si $E(\xi^2) < \infty$ car dans ce cas, on peut montrer que

$$Z_n / (E(\xi))^n \rightarrow W$$

lorsque $n \rightarrow \infty$ pour une v.a. W strictement positive sur A^c .

EXERCICE 7.2.9 Branchement avec immigration. Chaque génération du processus de branchement (avec $Z_0 = 1$) est augmentée d'un nombre aléatoire d'immigrants, qui deviennent alors indistinguables des autres individus. On suppose ces nouvelles variables aléatoires indépendantes entre elles et indépendantes des $(\xi_i^n; i \geq 1, n \geq 1)$, de fonction génératrice H . Montrer que la fonction génératrice G_n de Z_n vérifie alors $G_{n+1}(x) = G_n(g(x)) H(x)$.

7.2.3 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 micro-fissures). 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 $u = w_1 w_2 \dots w_n$ de longueur $|u| = n$ ($n \geq 1$, $w_i \in \{a, b\}$). Dans la figure 7.3 on a 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 l'on écrit alors $u \leq v$; on notera $u < v$ si $u \leq v$, $u \neq v$.

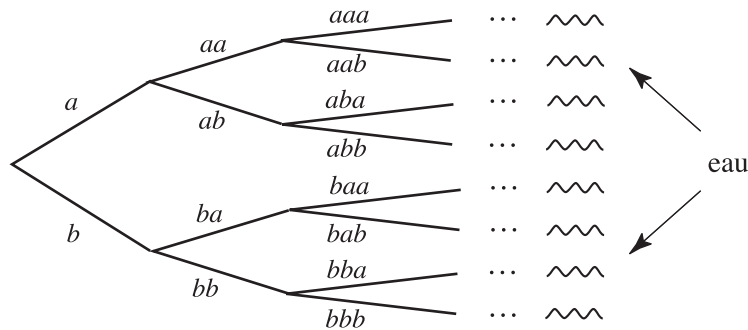


FIG. 7.3 – L'arbre A_3

A chaque arête $u \in A$, on associe une variable aléatoire binaire X_u , et l'on suppose ces variables aléatoires $(X_u)_{u \in A}$ indépendantes et de même loi donnée par $P(X_u = 1) = p = 1 - P(X_u = 0)$, 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. On définit alors

$$S_u = \sum_{v \leq u} X_v \quad , \quad u \in A$$

$$C_n = \{u \in A; |u| = n, S_u = n\} \quad , \quad C = \bigcup_{n \geq 1} C_n$$

L'ensemble C (resp. C_n) est constitué des arêtes de l'arbre (resp. à profondeur n), qui sont ouvertes et reliées à la racine – le centre de la pierre – par des arêtes ouvertes. Notez 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 $(1-p)^2$, $2p(1-p)$, p^2 respectivement. Sa fonction génératrice vaut donc $g(x) = (1-p+px)^2$ comme on l'a 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 on conclut :

(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.

On peut montrer aussi 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 votre tasse.

7.3 Files d'attente

Les modèles de file 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'au cas plus complexes des réseaux de communication (ethernet, internet, . . .). On s'intéresse 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.

7.3.1 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 ξ_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 alors selon la dynamique suivante :

si $X_n = 0$ alors $X_{n+1} = \xi_{n+1}$, et

si $X_n \geq 1$ alors $X_{n+1} = (X_n - 1) + \xi_{n+1}$.

Notant $a^+ = \max\{a; 0\}$ la partie positive de $a \in \mathbb{R}$, X_n satisfait donc la relation de récurrence

$$X_{n+1} = (X_n - 1)^+ + \xi_{n+1} \quad , \quad n \in \mathbb{N} \quad (7.3.15)$$

qui la détermine entièrement en fonction des ξ_n et de l'état initial X_0 que l'on supposera indépendant de la suite $(\xi_n)_n$.

Si q est la loi commune des ξ_n , c'est-à-dire que $P(\xi_n = x) = q(x)$ pour $x \in \mathbb{N}$, la loi conditionnelle de X_{n+1} sachant X_n est donnée par

$$P(X_{n+1} = y | X_n = x) = \begin{cases} q(y) & \text{si } x = 0 , \\ q(y - x - 1) & \text{si } x \geq 1 \end{cases} \quad (7.3.16)$$

(on convient que $q(x) = 0$ si $x \notin \mathbb{N}$).

Pour le bon fonctionnement de la file d'attente, il est nécessaire que la file ne s'engorge pas. On étudie 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 $x \in \mathbb{N}$, la notation P_x désignera la loi de la file d'attente partant de l'état initial constant $X_0 = x$, c'est-à-dire une probabilité sur l'espace de probabilité Ω telle que

$$P_x(X_0 = x) = 1.$$

Remarquons que si la file démarre vide ($X_0 = 0$) ou avec un client en attente ($X_0 = 1$), alors elle sera la même. Là l'instant 1, et donc aussi aux instants ultérieurs, de sorte que l'on a

$$P_0(T_0 < \infty) = P_1(T_0 < \infty).$$

Appelons v cette probabilité que la file se vide : $v := P_0(T_0 < \infty) = P_1(T_0 < \infty)$.

Puisque trivialement $a^+ \geq a$, il résulte de (7.3.15) que $X_{n+1} \geq X_n - 1 + \xi_{n+1}$, pour tout $n \geq 0$, et par sommation,

$$X_n \geq \sum_{k=1}^n \xi_k - n + X_0 . \quad (7.3.17)$$

On en déduit que

$$\frac{X_n}{n} \geq \frac{1}{n} \sum_{k=1}^n \xi_k - 1 + \frac{X_0}{n} . \quad (7.3.18)$$

Or, la loi des grands nombres montre que, avec probabilité 1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \xi_k = E(\xi_1).$$

Ainsi donc, presque-sûrement, $\liminf_n \frac{X_n}{n}$ va être supérieure à $E(\xi_1) - 1$. Le signe de cette différence va donc jouer un grand rôle. Rappelons que le paramètre $\rho = E(\xi_1)$ représente la charge moyenne de travail par unité de temps pour le serveur.

Lorsque $E(\xi_1) > 1$, la **file d'attente est instable, elle s'engorge** :

$$\rho := E(\xi_1) > 1 \Rightarrow \lim_{n \rightarrow \infty} X_n = +\infty \quad \text{avec probabilité 1 .}$$

D'autre part, tant que $n < T_0$, on a $(X_n - 1)^+ = X_n - 1$ et donc on a égalité dans (7.3.17). On en déduit que lorsque $\rho < 1$, le membre de droite de (7.3.17) tend vers $-\infty$ d'après la loi des grands nombres, et l'événement $\{T_0 = \infty\} \subset \{\sum_{k=1}^n \xi_k - n + X_0 \geq 1, \forall n \geq 1\}$ est de probabilité nulle. Ainsi la **file** reste **stable** dans ce cas :

$$\rho := E(\xi_1) < 1 \Rightarrow T_0 < \infty \quad \text{avec probabilité 1 .}$$

Le cas critique $\rho = 1$ est plus difficile. Il est décrit par le résultat plus précis suivant.

7.3.2 Stabilité : étude analytique

Théorème 7.3.1 *Pour la file d'attente (7.3.15) démarrant de $X_0 = 0$ ou 1, la probabilité qu'elle finisse par se vider, qui vaut*

$$v := P_0(T_0 < \infty) = P_1(T_0 < \infty),$$

est la plus petite solution de $g(v) = v$ dans $[0, 1]$, où g est la fonction génératrice de ξ_1

$$g(s) = E(s^{\xi_1}) = \sum_{x \geq 0} s^x q(x) .$$

On en déduit que $v = 1$ si et seulement si $\rho = E(\xi_1) \leq 1$.

Preuve. Chaque période d'activité du serveur peut se décrire à l'aide d'un processus de branchement. La file démarrant de $X_0 = 1$, le client présent à l'instant $n = 0$ est l'ancêtre du processus de branchement. Ses "enfants" sont les ξ_0 clients arrivant dans la file pendant son service ; ils constituent la première génération ($Z_1 = \xi_0$). 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 . On définit récursivement la $(N+1)$ -ème 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 . On peut écrire $Z_0 = 1, Z_1 = \xi_0$, et pour $N \geq 0$

$$Z_{N+1} = \sum_{Z_0 + \dots + Z_{N-1} \leq i \leq Z_0 + \dots + Z_N - 1} \xi_i \quad (7.3.19)$$

(avec convention que la somme sur l'ensemble vide est 0). En particulier, $Z_2 = \sum_{i=1}^{\xi_0} \xi_i$.

L'expression (7.3.19) n'est pas exactement de la forme (7.2.10) trouvée pour le modèle de branchement, mais elle définit bien la même loi : par exemple, la probabilité

$$P(Z_1 = x, Z_2 = y) = P(Z_2 = y | Z_1 = x)P(Z_1 = x) = P(\xi_0 = x) \times P\left(\sum_{i=1}^x \xi_i = y\right)$$

est la même pour (7.3.19) et pour (7.2.10). Plus généralement, la suite $(Z_N, N \geq 1)$ a même loi que le processus de branchement. On décrit 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 le file se vide est égale à $v = P_1(T_0 < \infty) = P(A)$. D'après le théorème 7.2.5, v est donc la plus petite solution dans $[0, 1]$ de l'équation $v = g(v)$. De même, $v = 1 \Leftrightarrow \rho \leq 1$.

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

$$\sum_{N \geq 1} Z_N = T_0 ,$$

et rappelant que $E(Z_N) = (E(\xi_1))^N = \rho^N$, on constate que la durée moyenne de la période d'activité n'est finie que si $\rho < 1$

$$E(T_0) < \infty \Leftrightarrow \rho = E(\xi_1) < 1 .$$

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. Voilà une autre preuve du théorème 7.3.1, qui n'utilise pas le processus de branchement.

(i) En conditionnant suivant les valeurs de ξ_1

$$\begin{aligned} P_1(T_0 < \infty) &= \sum_{x \geq 0} P_1(T_0 < \infty, \xi_1 = x) \\ &= \sum_{x \geq 0} P_1(T_0 < \infty | \xi_1 = x) q(x) \\ &= q(0) + \sum_{x \geq 1} P_x(T_0 < \infty) q(x) . \end{aligned} \tag{7.3.20}$$

En effet, les $(\xi_i, i \geq 1)$ étant indépendants et de même loi, la loi de $(X_2, X_3, \dots)$ conditionnée en $X_1 = x$ (ou $\xi_1 = x$) est la même que celle de $(X_1, X_2, \dots)$ démarrant de $X_0 = x$.

Introduisons, pour tout $x \in \mathbb{N}$, le temps d'atteinte T_x de l'état x , défini par $T_x = \min \{n \geq 1 : X_n = x\}$. Montrons que l'on a :

$$P_x(T_{x-1} < \infty) = \sigma \quad \forall x \geq 1, \quad (7.3.21)$$

$$P_x(T_0 < \infty) = P_x(T_{x-1} < \infty) \times P_{x-1}(T_{x-2} < \infty) \times \dots \times P_1(T_0 < \infty). \quad (7.3.22)$$

Pour montrer (7.3.21), on remarque que pour la file commençant à $x \geq 1$, $T_{x-1} = n$ si et seulement si

$$\begin{aligned} n &= \min \{k \geq 1 : x + (\xi_1 - 1) + \dots + (\xi_k - 1) = x - 1\} \\ &= \min \{k \geq 1 : \xi_1 + \dots + \xi_k = k - 1\}. \end{aligned}$$

Ainsi $P_x(T_{x-1} = n)$ ne dépend pas de x , et donc $P_x(T_{x-1} = n) = P_1(T_0 = n)$, ce qui entraîne que $P_x(T_{x-1} < \infty) = v$, en sommant sur n . On établit (7.3.22) par récurrence, en montrant que $P_x(T_0 < \infty) = P_x(T_{x-1} < \infty) \times P_{x-1}(T_0 < \infty)$, $\forall x \geq 2$. Pour cela on écrit

$$\begin{aligned} P_x(T_0 < \infty) &= \sum_{\ell=1}^{\infty} P_x(T_0 < \infty, T_{x-1} = \ell) \\ &= \sum_{\ell=1}^{\infty} P_x(T_0 < \infty | T_{x-1} = \ell) P_x(T_{x-1} = \ell) \\ &= \sum_{\ell=1}^{\infty} P_{x-1}(T_0 < \infty) P_x(T_{x-1} = \ell) \\ &= P_{x-1}(T_0 < \infty) \sum_{\ell=1}^{\infty} P_x(T_{x-1} = \ell) = P_{x-1}(T_0 < \infty) P_x(T_{x-1} < \infty) \end{aligned}$$

par un argument analogue à celui utilisé dans (7.3.20). En utilisant (7.3.21), (7.3.22) et le fait que $P_1(T_0 < \infty) = v$, (7.3.20) se transforme en

$$v = q(0) + \sum_{x \geq 1} v^x q(x),$$

et la relation $v = g(v)$ est démontrée.

(ii) Nous avons vu dans l'étude du processus de branchement, que l'équation $z = g(z)$, $z \in [0, 1]$, a $z = 1$ 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 (7.3.22), (7.3.21), on a

$$\begin{aligned} P_x(T_0 \leq n) &\leq P_x(T_{x-1} \leq n) \times P_{x-1}(T_{x-2} \leq n) \times \dots \times P_1(T_0 \leq n), \\ P_x(T_{x-1} \leq n) &= P_1(T_0 \leq n), \end{aligned}$$

et, de façon analogue à (7.3.20) on a

$$P_1(T_0 \leq n+1) = q(0) + \sum_{x \geq 1} P_x(T_0 \leq n) q(x) .$$

Il en résulte que $v_n = P_1(T_0 \leq n)$ vérifie

$$v_{n+1} \leq q(0) + \sum_{x \geq 1} v_n^x q(x) = g(v_n) .$$

Mais $v_0 = 0 \leq z$ et par récurrence $v_n \leq g(z) = z$ puisque g est croissante sur $[0, 1]$. On en déduit le résultat désiré

$$v = P_1(T_0 < \infty) = \lim_{n \rightarrow \infty} P_1(T_0 \leq n) = \lim_{n \rightarrow \infty} v_n \leq z .$$

□

Exemple 7.3.2 Pour la file d'attente partant de $X_0 = 1$, montrer que

$$T_0 = 1 + \sum_{i \leq \xi_0} T_0^{(i)}$$

où $T_0^{(i)}$ est la durée totale de service du $i^{\text{ème}}$ individu de la première génération et de ses descendants. Montrer que les $(T_0^{(i)}; i \geq 1)$ sont indépendants. En déduire une équation vérifiée par la fonction génératrice de T_0 .

7.3.3 Stabilité : expérience de simulation

Une simulation vous est présentée sur la page WEB du cours à l'adresse :

www.enseignement.polytechnique.fr/mathematiques-appliquees/Aleatoire

(puis cliquer sur le lien "File d'attente en temps discret").

Dans cet exemple, il y a 0, 1 ou 2 arrivées de clients par unité de temps, et la loi $q = (q(x), x = 0, 1, \dots)$ des arrivées est telle que

$$q(0) + q(1) + q(2) = 1 \tag{7.3.23}$$

de sorte que $\rho = 1 + q(2) - q(0)$. En observant la simulation, on constate expérimentalement que

- (i) Lorsque $\rho > 1$, la file finit par s'engorger.
- (ii) Lorsque $\rho < 1$, la file se vide à intervalles réguliers, et elle reste de taille limitée une bonne partie du temps. La proportion de temps passée dans chaque état $x \in \mathbb{N}$,

$$\pi_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k=x\}} \tag{7.3.24}$$

se stabilise vers une limite $\pi(x)$ lorsque la durée n de la simulation augmente indéfiniment. Par une analyse théorique que nous ne faisons pas ici, on peut montrer que

$$\lim_{n \rightarrow \infty} \pi_n(x) = \pi(x) \quad \text{avec probabilité 1 ,} \quad (7.3.25)$$

(ce résultat est appelé un théorème ergodique), et que dans ce cas précis (7.3.23),

$$\lim_{n \rightarrow \infty} P_y(X_n = x) = \pi(x) \quad (7.3.26)$$

pour tout $x \in \mathbb{N}$ et tout point de départ $X_0 = y$. Lorsque $\rho < 1$, π est une probabilité sur l'espace $\mathbb{N}$ des états, appelée **probabilité invariante**.

Cette probabilité décrit le **comportement en temps long** de la file (cf. (7.3.25), (7.3.26)). Elle décrit aussi le **régime stationnaire** (ou état d'équilibre) de la file, car on peut montrer que si l'état de départ X_0 est aléatoire et distribué selon π , alors X_1 est encore de loi π , et par récurrence X_n aussi ($n \geq 1$).

(iii) Lorsque $\rho = 1$, on constate au fur et à mesure de la simulation, que la **file se vide de plus en plus rarement** (mais sûrement, car $v = 1$) ; entre-temps, elle passe par des valeurs de plus en plus grandes. On peut montrer dans ce cas que pour tout état x

$$\begin{aligned} \lim_{n \rightarrow \infty} \pi_n(x) &= 0 \quad \text{avec probabilité 1 ,} \\ \lim_{n \rightarrow \infty} P_y(X_n = x) &= 0 . \end{aligned}$$

Contrairement à (7.3.25), (7.3.26), ces limites sont nulles, reflétant la faible stabilité de la file.

On peut cependant pousser plus loin les calculs analytiques. Dans l'exercice suivant, on calcule la fonction génératrice de la probabilité invariante π , ce qui détermine $(\pi(x), x \geq 0)$ par développement en série entière. Dans le cas particulier où q vérifie (7.3.23), on constate que π est une généralisation de la loi géométrique.

EXERCICE 7.3.3 Montrer que pour la file d'attente (7.3.15),

- a) ξ_n et X_n sont indépendantes,
- b) pour toute variable aléatoire entière Y , de fonction génératrice G_Y ,

$$E(z^{(Y-1)^+}) = G_Y(0) + \frac{1}{z}(G_Y(z) - G_Y(0)) .$$

- c) En déduire la relation de récurrence

$$G_{X_{n+1}}(z) = G(z) \left(G_{X_n}(0) + \frac{1}{z} [G_{X_n}(z) - G_{X_n}(0)] \right) .$$

Dans le cas $\rho < 1$, nous avons “ observé ” dans (ii) ci-dessus qu’il existe une probabilité invariante π décrivant le régime stationnaire : si X_n est de loi π , alors X_{n+1} aussi. La relation de l’exercice 7.3.3 c) ci-dessus entraîne que la fonction génératrice de π ,

$$G(z) = \sum_{x \geq 0} z^x \pi(x)$$

est solution de

$$G(z) = G(0) g(z) \frac{1-z}{g(z)-z}.$$

(On a écrit : $G_{X_{n+1}} = G_{X_n} = G$).

Ceci détermine G à condition de connaître $G(0)$. Mais $G(1) = 1$ et $\lim_{z \nearrow 1} \frac{g(z)-z}{1-z} = -g'(1) + 1 = -\rho + 1$, d’où $G(0) = 1 - \rho$. Cette quantité représente la probabilité à l’équilibre que le serveur se repose.

Par ailleurs, si X_n est de loi π (on dit qu’on est en régime stationnaire),

$$E(X_n) = G'(1) = \frac{1}{2} \left[\rho + \frac{\text{Var}(\xi_1)}{1-\rho} \right].$$

Notons que ces deux quantités deviennent critiques lorsque ρ tend vers 1.

Files d’attente plus générales : Certains projets de simulation présentent des simulations de files avec un ou plusieurs serveurs, avec temps de service aléatoire, des petits réseaux, et des modèles en temps continu. En temps continu, un modèle courant consiste à supposer que les délais entre les arrivées des clients successifs sont des variables aléatoires indépendantes de même loi exponentielle d’espérance $1/\lambda > 0$. Cela modélise un flot imprévisible de clients avec intensité λ .

7.4 Suites récurrentes aléatoires discrètes

Ce paragraphe unifie les modèles étudiés précédemment.

7.4.1 Probabilités de transition

Définition 7.4.1 Soient E, F deux espaces discrets, et $f : E \times F \rightarrow E$ une fonction. Soit $\xi_n : \Omega \rightarrow F$ une suite de variables aléatoires indépendantes et de même loi. Les variables aléatoires $X_n : \Omega \rightarrow E$ définies par récurrence,

$$X_{n+1} = f(X_n, \xi_{n+1}) \quad , \quad n \in \mathbb{N} \tag{7.4.27}$$

forment une suite appelée **suite récurrente aléatoire**.

La valeur initiale X_0 sera choisie déterministe, ou aléatoire (mais indépendante de $(\xi_n, n \geq 1)$).

L'analogue 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 7.4.2 *Pour tout n , la loi de l'état futur X_{n+1} ne dépend de l'histoire $(X_1, \dots, X_n)$ au temps n que par l'état présent X_n . Cette propriété s'écrit en terme de probabilités conditionnelles*

$$P(X_{n+1} = x \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = P(X_{n+1} = x \mid X_n = x_n), \quad (7.4.28)$$

pour tous $n \in \mathbb{N}$, $x_0, \dots, x_n, x \in E$.

Définition 7.4.3 Une suite $(X_n)_{n \in \mathbb{N}}$ vérifiant la propriété (7.4.28) est appelée une **chaîne de Markov**.

Preuve. La définition (7.4.27) entraîne que

$$P(X_0 = x_0, \dots, X_n = x_n, X_{n+1} = x) = \sum_{y: f(x_n, y) = x} P(X_0 = x_0, \dots, X_n = x_n, \xi_{n+1} = y). \quad (7.4.29)$$

Or les variables $X_0, \dots, X_n$ ne dépendent que de $X_0, \xi_0, \dots, \xi_{n-1}$, elles sont indépendantes de ξ_{n+1} et le terme général du second membre de (7.4.29) est égal à $P(X_0 = x_0, \dots, X_n = x_n) P(\xi_{n+1} = y)$. Par définition des probabilités conditionnelles,

$$P(X_{n+1} = x \mid X_0 = x_0, \dots, X_n = x_n) = \sum_{y: f(x_n, y) = x} P(\xi_{n+1} = y). \quad (7.4.30)$$

En sommant dans (7.4.29) sur $x_0, \dots, x_{n-1}$, on obtient par le même argument que le membre de droite de (7.4.30) est aussi égal à $P(X_{n+1} = x \mid X_n = x_n)$. $\square$

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

$$Q(x, y) = P(X_{n+1} = y \mid X_n = x) \quad (7.4.31)$$

décrivent la loi d'évolution en une étape de la suite; on les appelle **probabilités de transition**.

EXERCICE 7.4.4 Vérifier que si X_0 est de loi $p(\cdot)$,

$$P(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = p(x_0) Q(x_0, x_1) Q(x_1, x_2) \dots Q(x_{n-1}, x_n) .$$

Exemple 7.4.5 1. Promenade aléatoire : une particule se déplace sur le réseau $\mathbb{Z}^d$, en progressant, au hasard, à chaque instant n d'un vecteur ξ_{n+1} , et sa position X_n à l'instant n est donnée par

$$X_{n+1} = X_n + \xi_{n+1}$$

avec les ξ_n indépendants et de même loi $q(\cdot)$. Dans ce cas,

$$\begin{aligned} Q(x, y) &= P(X_{n+1} = y \mid X_n = x) \\ &= q(y - x) \end{aligned}$$

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,

$$X_{n+1} = (X_n + \xi_{n+1}) \mathbf{1}_{X_n \in D} + X_n \mathbf{1}_{X_n \notin D}$$

ce qui est encore de la forme (7.4.27). 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

$$X_{n+1} = |X_n + \xi_{n+1}| .$$

2. Dans le modèle de file d'attente (7.3.15), on a

$$X_{n+1} = (X_n - 1)^+ + \xi_{n+1}$$

et Q est donné par (7.3.16).

3. Pour le processus de branchement, ξ_n est donné par la famille des nombres d'enfants ξ_i^{n-1} à la génération $n-1$. Précisément, on a $\xi_n = \bigcup_{m \geq 0} (\xi_i^{n-1}; i \geq m)$, ce qui décrit bien un ensemble dénombrable, et $Z_{N+1} = \sum_{i \leq Z_N} \xi_i^N$ est bien de la forme (7.4.27).

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 exerce des demandes ξ_{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 $(q(y), y \in \mathbb{N})$. 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

$$\begin{aligned} X_{n+1} &= (X_n - \xi_{n+1})^+ \quad \text{si } X_n > a , \\ &= (b - \xi_{n+1})^+ \quad \text{sinon ,} \end{aligned}$$

étant entendu que toute demande ξ_n n'est satisfaite que dans les limites du stock X_n ou b , d'où l'opérateur $(\)^+$ ci-dessus.

La suite $(X_n, n \in N)$ est une chaîne de Markov de la forme (7.4.27). On s'intéressera ici 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 $\{X_n = 0\}$). La loi q des demandes étant supposée connue (par des statistiques antérieures, par exemple), on pourra chercher à ajuster les constantes a et b le mieux possible, économiquement parlant.

7.4.2 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 7.4.1 est " X a même loi que $f(X, \xi)$ " où l'on suppose que X et ξ sont indépendantes. Ici, ξ est une variable aléatoire de même loi que ξ_1 . Cette équation ne porte pas sur la variable X , mais plutôt sur sa loi, que l'on appelle loi invariante.

Définition 7.4.6 La probabilité π sur E est dite *invariante* pour la suite récurrente aléatoire si

$$\left. \begin{array}{l} X \text{ indépendante de } \xi \\ X \text{ de loi } \pi \end{array} \right\} \Rightarrow f(X, \xi) \text{ suit la loi } \pi .$$

Dans ce cas, si X_0 est de loi π , alors $X_1 = f(X_0, \xi_0)$ est de loi π et par conséquent X_n a pour 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 7.4.7 La probabilité π sur E est invariante si et seulement si :

$$\sum_{x \in E} \pi(x) Q(x, y) = \pi(y) \quad , \quad \forall y \in E .$$

Preuve. En effet, si X est indépendant de ξ et de loi π ,

$$\begin{aligned} P(f(X, \xi) = y) &= \sum_{x \in E} P(f(x, \xi) = y, X = x) \\ &= \sum_{x \in E} P(f(x, \xi) = y) \pi(x) \\ &= \sum_{x \in E} Q(x, y) \pi(x) \end{aligned}$$

d'après (7.4.30) et (7.4.31). Ceci montre la proposition. $\square$

EXERCICE 7.4.8 Réversibilité. Soit π une probabilité sur E telle que

$$\pi(x) Q(x, y) = \pi(y) Q(y, x) \quad \forall x, y \in E.$$

- (i) Montrer que π est invariante
- (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 lorsqu'on renverse le temps.

Exemple 7.4.9 *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* : on représente l'état du système comme un vecteur dans $\{0, 1\}^N$, la $i^{\text{ème}}$ composante valant 0 et 1 selon que la $i^{\text{ème}}$ 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 invariante.
- (ii) *Vision macroscopique* : les particules sont indistinguables. On représente alors le système à l'instant n par le nombre X_n de particules dans l'enceinte de gauche. Le mécanisme d'évolution se traduit, avec des v.a. $(\xi_n, n \geq 1)$ indépendantes de loi uniforme sur $\{1, 2, \dots, N\}$, par

$$X_{n+1} = \begin{cases} X_n + 1 & \text{si } \xi_{n+1} > X_n \\ X_n - 1 & \text{si } \xi_{n+1} \leq X_n \end{cases}$$

qui est de la forme (7.4.27). Ses probabilités de transition sont pour $x \in \{0, \dots, N\}$,

$$Q(x, x+1) = 1 - \frac{x}{n} \quad , \quad Q(x, x-1) = \frac{x}{n}$$

et $Q(x, y) = 0$ sinon. L'espace d'état étant l'espace fini $E = \{0, \dots, N\}$, aucune instabilité ne peut se produire. Plutôt que de chercher les probabilités invariantes à l'aide de la proposition 7.4.7, il est plus rapide d'utiliser l'exercice 7.4.8 (ii), et de chercher les solutions de $\pi(x) Q(x, y) = \pi(y) Q(y, x)$. On trouve aisément que $\pi(x) = \binom{n}{x} (1/2)^N$, c'est-à-dire que π est la loi binomiale $\mathcal{B}(N, 1/2)$.

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

Index

- événement aléatoire, 16
- événements indépendants, 40
- approximation linéaire, 84
- chaîne de Markov, 163
- coefficient de corrélation, 82
- convergence en loi, 135
- convergence en moyenne, 116
- convergence en probabilités, 116
- convergence presque-sûre, 116
- convolution, 110
- corrélation, 82
- covariance, 82
- covariance , 82
- densité, 74
- densité d'un vecteur aléatoire, 98
- densité marginale, 100
- diffusion d'Ehrenfest, 167
- droite des moindres carrés, 84
- ensemble dénombrable, 27
- espérance conditionnelle pour des v.a. à densité, 101
- espérance conditionnelle pour des v.a. discrètes, 61, 62
- espérance des variables aléatoires discrètes, 47, 48
- espace d'états, 15
- espace de probabilités, 31
- expérience aléatoire, 15
- expériences aléatoires indépendantes, 41
- file d'attente, 156
- fonction caractéristique, 125
- fonction de répartition, 71
- fonction génératrice, 53
- fonction génératrice de probabilité invariante, 162
- fonction mesurable, 70
- formule de Bayes, 37
- formule des probabilités composées, 37
- formule des probabilités totales, 37
- histogramme, 111
- indépendance, 63
- intervalle de confiance, 142
- limsup d'ensemble, 42
- loi des grands nombres, 120
- loi beta, 109
- loi binomiale, 56
- loi conditionnelle pour des v.a. discrètes, 60
- loi d'une v.a. discrète, 47
- loi d'une variable aléatoire, 34
- loi de Bernoulli, 54
- loi de Cauchy, 94
- loi forte des grands nombres, 121
- loi gamma, 90
- lois marginales, 60
- méthode de Monte-Carlo, 9, 123, 143
- marche aléatoire, 145
- marche aléatoire, 145
- matrice de covariance, 99
- matrice des covariances, 99
- modèle probabiliste, 21
- modèle d'inventaire, 165
- moment, 52
- partition, 18
- percolation, 155

- probabilité, 29
- probabilité conditionnelle, 36
- probabilité sur Ω fini, 21
- probabilité uniforme, 22
- probabilité invariante, 162, 166
- probabilités de transition, 164
- processus réversible, 167
- promenade aléatoire, 145
- propriété de non-vieillessement, 89
- propriété vraie presque-sûrement, 32
- propriété de Markov, 163
- régime stationnaire, 162
- réversibilité, 167
- sondages, 140
- stabilité, 158, 166
- suite d'événements indépendants, 41
- suite de variables aléatoires indépendantes, 105
- table de la loi de Gauss, 92
- théorème central limite multi-dimensionnel, 140
- théorème de convergence dominée, 117
- théorème de Lévy, 137
- théorème de la limite centrale, 138
- théorème de Slutsky, 144
- transformée de Laplace, 129
- tribu, 28
- tribu borélienne, 29
- tribu engendrée par une partie de Ω , 29
- variable aléatoire, 32
- variable aléatoire de Bernoulli, 54
- variable aléatoire étagée, 79
- variable aléatoire de loi normale, 91
- variable aléatoire de chi-deux, 133
- variable aléatoire exponentielle, 87
- variable aléatoire géométrique, 57
- variable aléatoire normale centrée réduite, 90
- variable aléatoire réelle, 69
- variable aléatoire réelle de carré intégrable, 82
- variable aléatoire uniforme sur $[a, b]$, 86
- variables aléatoires réelles intégrables, 79
- variable binomiale, 56
- variable de Poisson, 58
- variable discrète de carré intégrable, 50
- variable discrète intégrable, 49
- variable gaussienne, 90
- variables aléatoires discrètes, 46
- variance d'une variable aléatoire réelle, 82
- vecteur aléatoire, 98
- vecteur gaussien, 131
- vecteurs aléatoires indépendants, 103