



M2 - PROBABILITÉS

Cours de pré-rentrée

Laurent Mazliak

24 septembre 2018

Avant-Propos

Que serait-ce que les Jésuites sans la probabilité et que la probabilité sans les Jésuites ?

Otez la probabilité on ne peut plus plaire au monde ; mettez la probabilité on ne peut plus lui déplaire.

B. Pascal, *Pensées* ([22], p.412)

Le présent polycopié est en très grande partie composé à partir de l'ouvrage [17] (mais j'en ai profité pour corriger de nombreuses coquilles qui le parsemaient... sans douter un instant qu'il en reste encore plein !). Il considère connus les principaux résultats de la théorie de la mesure et de l'intégration, et contient évidemment beaucoup plus que ce qui sera fait pendant les deux semaines de pré-rentree.

Un chapitre initial rassemble les principaux résultats de la théorie de la mesure qui seront utilisés en permanence par la suite. Ce chapitre ne contient cependant aucune démonstration. Après avoir défini et étudié des propriétés générales des espaces de probabilités et des variables aléatoires, on passe à l'étude des convergences de suites de variables puis à l'étude du conditionnement. L'étude des vecteurs gaussiens est présentée ensuite ainsi que quelques éléments sur le processus de Poisson, la simulation et les martingales à temps discret (dont le chapitre est repris de [2]). Par contre, les chaînes de Markov, qui font l'objet d'un autre cours n'y sont pas abordées. Pour ce qui concerne les corrections des exercices sur les martingales, ainsi qu'un cours sur les chaînes de Markov on pourra d'ailleurs se reporter à l'ouvrage indispensable (!) [2].

Le cours est émaillé de nombreux exercices, dont beaucoup, repérés par le signe \diamond sont intégralement corrigés dans [17]

Table des matières

| | |
|--|-----------|
| Les probabilités hier et aujourd'hui | 1 |
| 1 Rappels d'intégration | 5 |
| 2 Espaces et mesures de probabilités | 11 |
| 2.1 Espaces de probabilités | 11 |
| 2.2 Théorème d'extension | 12 |
| Application : Construction de la mesure de Lebesgue λ sur $[0, 1]$ | 15 |
| 2.3 Lemme de Borel-Cantelli | 16 |
| 2.4 Probabilité conditionnelle et indépendance | 17 |
| 3 Variables aléatoires | 19 |
| 3.1 Définitions et premières propriétés | 19 |
| 3.2 Lois des variables aléatoires | 20 |
| 3.3 Notions sur les processus à temps discret | 33 |
| 4 Différents types de convergences | 37 |
| 4.1 Convergences | 37 |
| 4.2 Liens entre les différentes convergences | 37 |
| 5 Convergences spatiales | 41 |
| 5.1 Loi des grands nombres | 41 |
| 5.2 Convergence des séries de variables indépendantes | 43 |
| 6 Convergence en loi unidimensionnelle | 49 |
| 6.1 Fonctions de répartition | 49 |
| 6.2 Fonctions caractéristiques | 50 |
| 6.3 Etude de la convergence en loi unidimensionnelle | 53 |
| 6.4 Le théorème limite central | 59 |
| 7 Théorèmes limites dans \mathbb{R}^k | 63 |
| 7.1 Convergence faible et suites tendues | 63 |
| 7.2 Fonctions caractéristiques | 66 |
| 7.3 Quelques prolongements | 67 |
| 8 Espérance conditionnelle | 71 |
| 8.1 Construction de l'espérance conditionnelle | 71 |
| 8.2 Propriétés de l'espérance conditionnelle | 72 |

| | |
|---|------------|
| 9 Lois conditionnelles | 75 |
| 9.1 Lois conditionnelles | 75 |
| 9.2 Cas particuliers | 77 |
| 10 Vecteurs gaussiens | 79 |
| 10.1 Distributions normales dans \mathbb{R}^k | 79 |
| 10.2 Le Théorème Limite Central dans \mathbb{R}^k | 84 |
| 10.3 Vecteurs gaussiens et conditionnement | 84 |
| 11 Simulation | 87 |
| 11.1 Quelques méthodes classiques | 87 |
| 11.1.1 Variables aléatoires discrètes | 87 |
| 11.1.2 Méthode de l'inversion | 88 |
| 11.1.3 Méthode du rejet | 88 |
| 11.2 Simulation de lois particulières | 89 |
| 11.2.1 Loi binomiale $\mathcal{B}(n, p)$ | 89 |
| 11.2.2 Loi de Poisson de paramètre $\lambda > 0$ | 89 |
| 11.2.3 Variable normale réduite | 90 |
| 11.3 Calculs d'espérances | 91 |
| 11.3.1 Méthode de Monte-Carlo | 91 |
| 11.3.2 Suites à discrédance faible | 91 |
| 12 Martingales à temps discret | 93 |
| 12.1 Généralités | 93 |
| 12.2 Premières propriétés | 94 |
| 12.3 Le théorème d'arrêt | 94 |
| 12.4 Inégalités maximales | 95 |
| 12.5 Martingales de carré intégrable | 96 |
| 12.6 Théorèmes de convergence | 96 |
| 12.7 Martingales régulières | 98 |
| Exercices | 99 |
| 13 Processus de Poisson | 111 |
| 13.1 Processus ponctuel | 111 |
| 13.2 Le processus de Poisson | 111 |
| 13.3 Quelques propriétés du processus de Poisson | 113 |
| 13.4 Intégrale stochastique | 116 |
| Bibliographie | 121 |

La théorie des probabilités hier et aujourd'hui

Inclure un bref aperçu historique est légitime au début d'un cours de probabilités, car il semble important que l'étude de cette discipline s'accompagne d'une prise de conscience de sa spécificité. Concevoir le calcul des probabilités simplement comme un sous-chapitre de la théorie de l'intégration, masque en effet complètement les immenses discussions qui eurent lieu pendant des siècles au sujet de sa nature et de ses rapports avec les différentes conceptions qu'on peut se faire du hasard. Si la théorie de la mesure joue incontestablement aujourd'hui un rôle capital dans la théorie des probabilités, un point essentiel que l'histoire nous a légué, c'est que faire des probabilités c'est *autre chose* que faire de l'intégration, même si les problèmes philosophiques que pose la question du hasard n'enflamment plus les esprits. Il y a une démarche probabiliste spécifique. Or, cette différence apparaît en pleine lumière quand on suit la construction élaborée depuis quelques trois cent cinquante ans. La théorie mathématique du hasard s'est construite sur des aller-et-retours permanents entre l'observation des phénomènes et l'élaboration mathématique qui en rend compte. La pratique aujourd'hui du calcul des probabilités a gardé de cette démarche une forme de réflexion très particulière cherchant autant que possible à allier les informations obtenues à partir de l'étude de l'expérience aléatoire et des propriétés des objets mathématiques manipulés.

Mentionnons juste ici une trame et quelques noms. Il est classique de faire démarrer l'histoire du calcul des probabilités à la correspondance échangée entre Blaise Pascal (1623-1662) et Pierre de Fermat (1601-1665) au sujet de problèmes de jeu et notamment du célèbre problème des partis qui amena Pascal à former une notion d'espérance mathématique. Ces débuts officiels furent cependant précédés par un certain nombre de tentatives pour forger un concept de hasard mesurable dont on trouve des traces dès l'Antiquité à travers des "mesures" de risques. L'importance du modèle du jeu comme prototype de l'événement fortuit va s'imposer progressivement au seizième siècle accompagnant l'emprise croissante du rôle de l'argent dans les échanges sociaux, comme en témoignent de merveilleux exemples picturaux comme le tableau *Les Tricheurs* de Georges de la Tour. On peut noter d'ailleurs que le mot *aléatoire* utilisé à partir du 19^{ème} siècle provient lui-même de *alea*, les *dés* en latin. Avant Pascal et Fermat, on peut mentionner des écrits de Jérôme Cardan (1501-1576) et de Galileo Galilei (1564-1642) fournissant des calculs sur les jeux. Il est cependant vrai qu'avec Pascal, on entre dans une autre dimension de la question par la volonté d'élaborer une véritable *géométrie du hasard* selon sa propre formulation. Le premier traité, entièrement consacré à une modélisation mathématique du jeu de dé, est dû à Christiaan Huygens (1629-1695) et date de 1657.

Vers la fin du 17^{ème} siècle cependant, le domaine des sciences de la société (ce qu'on appelait alors l'arithmétique politique) va prendre une place de plus en plus importante dans l'émergence d'une théorie mathématique du calcul des probabilités. L'évaluation des populations et des estimations de mortalité deviennent des disciplines essentielles à la gouvernance moderne et économique des états. On peut dans cette direction citer les écrits de William Petty (1623-1687), Jan de Witt (1625-1672) et Gottfried W. Leibniz (1646-1716). Ce mouvement ira en s'amplifiant et débouchera tout au long

du 18^{ème} siècle sur la naissance de la théorie des populations (ancêtre de notre démographie), à laquelle participent la plupart des grands mathématiciens du temps comme Daniel Bernoulli (1700-1782), Leonhard Euler (1707-1783), Jean-le-Rond d'Alembert (1717-1783), Jean-Antoine de Condorcet (1743-1794). Les questions d'astronomie et le traitement des erreurs de mesure prennent également une grande importance et Pierre-Simon de Laplace (1749-1827) fait des probabilités un outil indispensable pour l'analyse cosmologique.

C'est donc portée par ces problèmes au cours du 18^{ème} siècle que la mathématisation des notions de probabilités s'accroît et les modèles d'urne, les théorèmes limites tels que la loi des grands nombres (théorème de Bernoulli) et des versions initiales du théorème limite central (théorème de Moivre) quantifiant l'écart entre la moyenne des observations et la valeur théorique attendue et la vitesse de convergence, les probabilités géométriques estimant les chances de réalisation de certaines configurations dans un plan, vont amener les mathématiciens à faire intervenir dans la modélisation des propriétés issues de la théorie des fonctions, du calcul intégral et différentiel. Citons les noms de Jakob Bernoulli (1654-1705) qui introduisit la notion de série dans les calculs de probabilités et formula une version élémentaire de ce qu'on nomme aujourd'hui la loi des grands nombres prescrivant une interprétation fréquentielle de la probabilité d'un événement, d'Abraham de Moivre (1667-1754) qui obtint une première version du résultat appelé aujourd'hui théorème limite central pour le schéma de Bernoulli (jeu de pile ou face, voir [18]) et de Georges Louis Leclerc de Buffon (1707-1788) qui développa une première théorie des probabilités géométriques (problème de l'*aiguille de Buffon*, voir Exercice 3.3.19). Ajoutons à cette liste les études consacrées au problème de l'inversion de la probabilité ou de la recherche de la probabilité des causes; on doit y citer Thomas Bayes (1701-1761) qui se posa le problème de l'inférence statistique à partir de probabilités *a posteriori* mais surtout Pierre-Simon de Laplace qui fut le véritable créateur de ce qu'on appelle improprement la *théorie bayésienne*.

Au début du 19^{ème} siècle, à part des essais hasardeux d'application du calcul des probabilités aux sciences morales notamment de la part de Laplace et de son élève Siméon-Denis Poisson (1781-1840), essais qui soulevèrent immédiatement des tempêtes de protestation, l'attention se concentre entraînée par les besoins de la physique expérimentale sur la théorie des erreurs. Adrien-Marie Legendre (1752-1833) formule une première méthode des moindres carrés dont les fondements se trouvent dans la *Théorie analytique* de Pierre-Simon de Laplace. Ce dernier met en avant le rôle de la loi normale et démontre une version assez générale du théorème limite central. Enfin, c'est surtout Carl-Friedrich Gauss (1777-1855) qui développe toute une théorie des erreurs. Pendant tout le 19^{ème} siècle, ces résultats vont être précisés et augmentés. Les mathématiciens russes de l'école de Saint-Petersbourg, à commencer par son fondateur Pafnuty Lvovitch Chebyshev (1821-1894), puis Andreï Andreïevitch Markov (1856-1922) et Aleksandr Mikhaïlovitch Lyapunov (1857-1918), jouent un rôle prépondérant en fournissant des démonstrations de la loi (faible) des grands nombres pour laquelle Chebyshev énonça sa célèbre inégalité qui avait été aussi découverte et démontrée antérieurement par le français Irénée Jules Bienaymé (1796-1878), du théorème limite central (Markov le fit en 1898, concluant une démonstration presque achevée par Chebyshev utilisant la méthode des moments, Lyapunov en 1901 par les fonctions caractéristiques).

On dit souvent que la théorie moderne des probabilités naît avec Andreï Nikolaïevitch Kolmogorov (1903-1987) qui en proposa une axiomatisation en 1933 dans un cadre de théorie de la mesure. En fait, il est sans doute plus juste de voir le tournant

se produire entre 1905 et 1910 quand Emile Borel (1871-1956) entrevoit les applications possibles de la mesure des ensembles et de l'intégrale de Lebesgue au calcul des probabilités et donne un premier énoncé de la loi forte des grands nombres avec une convergence presque sûre. Parmi les précurseurs de Kolmogorov qui utilisèrent les propriétés de la mesure abstraite dans un cadre probabiliste, on peut citer Maurice Fréchet (1878-1973) qui introduisit l'intégration sur un espace abstrait, permettant de définir en toute généralité l'espérance d'une variable aléatoire, et Paul Lévy (1886-1971) qui mit en lumière le rôle fondamental de la convergence en loi. L'axiomatisation de Kolmogorov permit surtout de donner un socle solide à la théorie des processus aléatoires (ou *processus stochastiques*) qui fut le grand sujet d'étude en probabilités tout au long du 20ème siècle. Initiée par les travaux de Andreï Andreevitch Markov et de Henri Poincaré (1854-1912), la théorie des événements en chaîne (*chaînes de Markov*) allait connaître un développement fulgurant avec les recherches de Bohuslav Hostinsky (1884-1951), Maurice Fréchet, Paul Lévy, Harald Cramer (1893-1985), Wolfgang Döblin (1915-1940) et surtout de l'époustouflante école probabiliste soviétique à laquelle appartenaient (liste loin d'être exhaustive !) Andreï Nikolaïevitch Kolmogorov, Alexandr Yakovlevitch Khinchin (1894-1959), Sergueï Natanevitch Bernstein (1880-1968) et Vsevolod Ivanovitch Romanovsky (1879-1954). Dans une direction conjointe, les processus gaussiens, au premier rang desquels se trouve le mouvement brownien, allaient eux aussi se trouver au premier plan, notamment avec l'œuvre immense et très personnelle de Paul Lévy. Une théorie générale des processus stochastiques se mettait alors en place progressivement qui allait aboutir à son exposé systématique par Joseph Lee Doob (1910-2004). En parallèle, un calcul différentiel et intégral des processus aléatoires était construit, le *calcul stochastique*, permettant de définir une notion suffisamment souple et exploitable d'équation différentielle où intervient un bruit aléatoire. Il faut ici mentionner le rôle important joué par l'école japonaise, et notamment par Kyoshi Itô (1915-2008) qui obtint une construction pratique d'une intégrale par rapport au mouvement brownien et grâce à elle, la mise en place d'une version commode d'un calcul différentiel et intégral pour certaines familles de processus stochastiques. Les probabilités se sont ainsi imposées comme un des domaines de recherches et d'applications les plus actifs en mathématiques, et on l'a vue intervenir dans de multiples champs du savoir : en physique naturellement, mais aussi en biologie, en économie, en recherche opérationnelle et dans les sciences humaines. Pour des compléments, le lecteur intéressé peut consulter quelques livres qui abordent l'histoire de la discipline à diverses périodes comme [16], [24], [11], [4], [10] ou [19].

Chapitre premier

Rappels d'intégration

Nous allons dans ce chapitre introductif rappeler les notions les plus importantes de la théorie de la mesure et de l'intégration qui nous serviront en permanence par la suite. Le lecteur est néanmoins invité à se reporter à des textes spécifiques pour de plus amples détails (voir par exemple [9]).

Dans ce qui suit, E désigne un ensemble non vide.

1.1 Définition. On appelle *tribu sur E* un ensemble \mathcal{E} de parties de E tel que

- (i) $\emptyset \in \mathcal{E}$.
- (ii) $A \in \mathcal{E} \Rightarrow A^c \in \mathcal{E}$.
- (iii) $(A_n)_{n \in \mathbb{N}} \in \mathcal{E} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{E}$.

Le couple (E, \mathcal{E}) est dit **ensemble mesurable**.

Quand la condition (iii) est remplacée par la condition beaucoup plus faible

$$(iii') A, B \in \mathcal{E} \Rightarrow A \cup B \in \mathcal{E}$$

on dit que \mathcal{E} est une **algèbre de Boole**. Quand il n'y aura pas d'ambiguïté, on emploiera simplement le terme algèbre pour algèbre de Boole.

On vérifie immédiatement qu'une intersection de tribus est une tribu. De ce fait, quand \mathcal{B} est un ensemble de parties de E , l'intersection de toutes les tribus sur E qui contiennent \mathcal{B} est la plus petite tribu sur E qui contient \mathcal{B} : on l'appelle la **tribu engendrée par \mathcal{B}** et on la note $\sigma(\mathcal{B})$.

Un exemple fondamental est donné quand $E = \mathbb{R}$ et \mathcal{B} est l'ensemble des intervalles ouverts de \mathbb{R} . La tribu engendrée par \mathcal{B} est dite **tribu borélienne** de \mathbb{R} et notée $\mathcal{B}(\mathbb{R})$. Cette définition se prolonge au cas de \mathbb{R}^k et plus généralement d'un espace métrique E quand \mathcal{B} est la famille des ouverts de E .

1.2 Définition. Soient (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurables et $f : E \rightarrow F$ une application. On dit que f est **mesurable** si $f^{-1}(\mathcal{F}) \subset \mathcal{E}$ c'est-à-dire

$$\forall B \in \mathcal{F}, \{x \in E, f(x) \in B\} \in \mathcal{E}.$$

Un cas particulier simple de fonctions mesurables est celui des fonctions continues de \mathbb{R}^k dans \mathbb{R}^m quand ces deux espaces sont munis de leurs tribus boréliennes. Parmi les fonctions mesurables de (E, \mathcal{E}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on a avant tout les **fonctions étagées** de la forme

$$f = \sum_{k=0}^n \alpha_k \mathbb{I}_{A_k}$$

où les A_k sont des ensembles de \mathcal{E} . Rappelons aussi qu'une limite simple d'une suite de fonctions mesurables est une fonction mesurable. A ce sujet, on a l'utile résultat suivant :

1.3 Proposition. *Toute fonction numérique mesurable positive est limite simple d'une suite croissante de fonctions étagées positives.*

Quand $(f_i)_{i \in I}$ est une famille de fonctions de E dans l'espace mesurable (F, \mathcal{F}) , la tribu engendrée par les ensembles $f_i^{-1}(A), A \in \mathcal{F}$ est la plus petite tribu sur \mathcal{E} qui rende mesurables toutes les f_i : c'est la tribu engendrée par les f_i , notée $\sigma(f_i, i \in I)$.

1.4 Définition. Soit (E, \mathcal{E}) un espace mesurable. Une **mesure** sur (E, \mathcal{E}) est une application $\mu : \mathcal{E} \rightarrow \mathbb{R}^+ \cup \{+\infty\}$ telle que pour toute suite (A_n) d'éléments de E disjoints deux à deux on ait

$$\mu\left(\bigcup_{n \geq 0} A_n\right) = \sum_{n \geq 0} \mu(A_n).$$

L'exemple le plus important de mesure est donné par la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, notée traditionnellement λ , qui à un intervalle $]a, b[$ associe $b - a$. Rappelons au passage que son existence n'est pas absolument immédiate et résulte d'un théorème. Une mesure μ sur (E, \mathcal{E}) permet de définir l'**intégrale** d'une fonction mesurable $f : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On considère d'abord des fonctions étagées positives sous la forme

$$f = \sum_{k=0}^n \alpha_k \mathbb{I}_{A_k}$$

pour lesquelles on a

$$\int_E f(x) d\mu(x) = \sum_{k=0}^n \alpha_k \mu(A_k).$$

On prolonge ensuite cette définition pour une fonction mesurable positive f : par la proposition 1.3, celle-ci est limite croissante de fonctions étagées f_k et on définit

$$\int_E f(x) d\mu(x) = \lim \uparrow \int_E f_k(x) d\mu(x).$$

Noter que ce nombre peut être infini. On montre qu'il ne dépend pas de la suite (f_k) choisie.

Enfin, quand f est une fonction mesurable, $f = f^+ - f^-$ (où $a^+ = \max(a, 0)$ et $a^- = \max(-a, 0)$) ; si $\int_E f^+(x) d\mu(x) < +\infty$ et $\int_E f^-(x) d\mu(x) < +\infty$, on dit que f est intégrable (sur E et par rapport à μ) et on pose

$$\int_E f(x) d\mu(x) = \int_E f^+(x) d\mu(x) - \int_E f^-(x) d\mu(x).$$

Noter que si f est intégrable, $|f|$ l'est aussi (puisque $|f| = f^+ + f^-$).

Exercice 1.1

Soit (X, \mathcal{X}, μ) un espace mesuré tel que $\mu(X) < \infty$. Soit (f_n) une suite de fonctions complexes mesurables qui converge simplement sur X . Alors quelque soit $\varepsilon > 0$, il existe un sous ensemble mesurable E de X tel que $\mu(X \setminus E) < \varepsilon$ et que (f_n) converge uniformément sur E . (Théorème d'Egoroff)

L'ensemble des fonctions intégrables sur (E, \mathcal{E}, μ) (ou plus exactement comme usuellement l'ensemble des classes d'équivalence de fonctions intégrables pour la relation d'égalité μ - presque partout) est noté $L^1(E, \mathcal{E}, \mu)$. Plus généralement, si p est un réel ≥ 1 , on définit $L^p(E, \mathcal{E}, \mu)$ comme l'ensemble des fonctions mesurables de (E, \mathcal{E}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telles que $|f|^p \in L^1(E, \mathcal{E}, \mu)$. On note alors

$$\|f\|_p = \left(\int_E |f|^p(x) d\mu(x) \right)^{1/p}.$$

On définit aussi $L^\infty(E, \mathcal{E}, \mu)$ comme l'ensemble des fonctions mesurables de (E, \mathcal{E}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telles qu'il existe $M > 0$ tel que $\mu(\{x \in E, |f(x)| > M\}) = 0$ (on dit que $|f| \leq M$, presque partout (p.p.)) et $\|f\|_\infty$ est la borne inférieure de ces M . L'inégalité suivante permet d'établir des liens entre les espaces L^p .

1.5 Théorème. (inégalité de Hölder) Soient p et q deux réels tels que $\frac{1}{p} + \frac{1}{q} = 1$. Soient $f \in L^p(E, \mathcal{E}, \mu)$ et $g \in L^q(E, \mathcal{E}, \mu)$. Alors $fg \in L^1(E, \mathcal{E}, \mu)$ et

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

Dans le cas d'une mesure de masse totale 1, on a aussi l'inégalité suivante.

1.6 Théorème. (inégalité de Jensen) Soit (E, \mathcal{E}, μ) un espace mesuré tel que $\mu(E) = 1$. Soit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe. Pour toute f fonction intégrable telle que $\phi \circ f$ soit intégrable, on a

$$\phi\left(\int_E f(x) \mu(dx)\right) \leq \int_E \phi \circ f(x) \mu(dx).$$

Exercice 1.2

Soit μ une mesure positive sur (X, \mathcal{B}) , $f : X \rightarrow \mathbb{C}$ mesurable, et $p > 0$. On pose

$$\phi(p) = \int_X |f|^p d\mu = \|f\|_p^p.$$

On pose ensuite $E = \{p : \phi(p) < \infty\}$, et on suppose que $\|f\|_\infty > 0$.

- a) Pour $r < p < s$, avec $r, s \in E$, montrer que $p \in E$.
- b) Montrer que $\log \phi$ est convexe à l'intérieur de E et que ϕ est continue sur E .
- c) D'après a), E est connexe. Est-il nécessairement ouvert, fermé? Peut-il être réduit à un point? Peut-il être n'importe quel sous ensemble connexe de $]0, +\infty[$?
- d) Pour $r < p < s$, montrer que $\|f\|_p \leq \max\{\|f\|_r, \|f\|_s\}$. En conclure $L^p(\mu) \subset L^r(\mu) \cap L^s(\mu)$.
- e) Supposons qu'il existe $r > 0$ tel que $\|f\|_r < \infty$. Montrer qu'alors $\|f\|_p \rightarrow \|f\|_\infty$.

◇ Exercice 1.3

On suppose toujours $\mu(X) = 1$. Soient f et g deux fonctions intégrables et positives telles que $fg \geq 1$. Montrer qu'alors

$$\int_X f d\mu \int_X g d\mu \geq 1.$$

Exercice 1.4

Toujours si $\mu(X) = 1$, soit $h > 0$ mesurable et $A = \int_X h d\mu$. Montrer que

$$\sqrt{1+A^2} \leq \mathbb{E}_\mu(\sqrt{1+h^2}) \leq 1+A.$$

Si μ est la mesure de Lebesgue sur $[0, 1]$ et si h est continue, avec $h = f'$, quelle est l'interprétation géométrique des inégalités ci-dessus ?

Exercice 1.5

Soit $1 < p < \infty$, $f \in L^p(]0, +\infty[)$, relativement à la mesure de Lebesgue. Posons pour $x > 0$,

$$F(x) = \frac{1}{x} \int_0^x f(t) dt.$$

a) Démontrer l'inégalité de Hardy

$$\|F\|_p \leq \frac{p}{p-1} \|f\|_p,$$

qui assure que F envoie L^p dans lui-même.

b) Démontrer que l'on a l'égalité si et seulement si $f = 0$ p.s..

c) Démontrer que la constante $\frac{p}{p-1}$ est la plus petite possible.

d) Si $f > 0$ et $f \in L^1$, montrer que $F \notin L^1$.

Indications : pour le a), considérer d'abord $f \geq 0$ continue à support compact, et intégrer par parties pour obtenir $\|F\|_p^p = -p \int_0^\infty F^{p-1}(x) x F'(x) dx$. Remarquer que $x F' = f - F$, et appliquer l'inégalité de Hölder à $\int F^{p-1} f dx$.

Pour c), faire $f(x) = x^{-1/p}$ sur $[1, A]$, 0 ailleurs, pour A assez grand.

L'importance de la construction de l'intégrale vient de la très grande souplesse apportée aux passages à la limite (contrairement par exemple au cas de l'intégrale de Riemann pour laquelle on a en général besoin de convergences très fortes (type convergence uniforme) pour pouvoir passer à la limite).

1.7 Théorème. (convergence monotone, Fatou-Beppo Levi) Soit (f_n) une suite de fonctions mesurables telle que $0 \leq f_n \leq f_{n+1}$. On note pour tout $x \in E$, $\lim_n f_n(x) = f(x)$. Alors $\int_E f(x) d\mu(x) = \lim_n \int_E f_n(x) d\mu(x)$.

Rappelons le Lemme de Fatou, qui peut permettre d'obtenir des inégalités intéressantes sous des hypothèses très faibles :

1.8 Théorème. (Lemme de Fatou) Soit (f_n) une suite de fonctions mesurables telle que $\forall n, 0 \leq f_n$. On note $f = \liminf_n f_n$. Alors $\int_E f(x) d\mu(x) \leq \liminf_n \int_E f_n(x) d\mu(x)$.

On en déduit le résultat le plus important, le théorème de convergence dominée de Lebesgue

1.9 Théorème. Soit (f_n) une suite de fonctions mesurables telle que $\forall n, |f_n| \leq g$ où g est une fonction intégrable sur (E, \mathcal{E}, μ) . On suppose que p.p. $\lim_n f_n(x) = f(x)$. Alors $\int_E f(x) d\mu(x) = \lim_n \int_E f_n(x) d\mu(x)$.

Rappelons aussi qu'on dit que la mesure μ admet la **densité f par rapport à la mesure ν** si f est une fonction numérique positive mesurable sur \mathbb{R} telle que pour tout $A \in \mathcal{B}(\mathbb{R})$,

$$\mu(A) = \int_A f(x)\nu(dx).$$

Exercice 1.6

Calculer, en les justifiant, les limites suivantes :

$$\lim_{n \rightarrow \infty} \int_0^n \left(1 - \frac{x}{n}\right)^n e^{x/2} dx, \quad \lim_{n \rightarrow \infty} \int_0^n \left(1 + \frac{x}{n}\right)^n e^{-2x} dx.$$

◇ Exercice 1.7

On reprend les hypothèses de l'exercice 1.2 mais en outre on suppose que $\mu(X) = 1$.

- Démontrer que $\|f\|_r \leq \|f\|_s$ si $0 < r < s \leq \infty$.
- A quelles conditions peut-on avoir $0 < r < s \leq \infty$ et $\|f\|_r = \|f\|_s < \infty$?
- En supposant que $\|f\|_r < \infty$ pour un certain $r > 0$, montrer que

$$\lim_{p \rightarrow 0^+} \|f\|_p = \exp\left(\int_X \log |f| d\mu\right),$$

en posant $\exp(-\infty) = 0$.

Une autre notion fondamentale concerne les espaces-produits.

1.10 Définition. Soient (E, \mathcal{E}) et (F, \mathcal{F}) deux espaces mesurables. La tribu sur $E \times F$ engendrée par les pavés $A \times B$ où $A \in \mathcal{E}$ et $B \in \mathcal{F}$ est dite **tribu produit**. On la note $\mathcal{E} \otimes \mathcal{F}$.

Sur l'espace mesurable $(E \times F, \mathcal{E} \otimes \mathcal{F})$, on définit le produit d'une mesure sur (E, \mathcal{E}) par une mesure sur (F, \mathcal{F}) .

1.11 Théorème et Définition. Soit μ (resp. ν) une mesure sur l'espace mesurable (E, \mathcal{E}) (resp. (F, \mathcal{F})). Il existe une unique mesure sur $(E \times F, \mathcal{E} \otimes \mathcal{F})$, notée $\mu \otimes \nu$ et appelée **produit de μ et ν** telle que

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B), A \in \mathcal{E}, B \in \mathcal{F}.$$

Les résultats fondamentaux suivants permettent d'invertir l'ordre d'intégration dans le calcul d'une intégrale par rapport à une mesure produit. Le second résultat sert notamment à démontrer l'intégrabilité d'une fonction de deux variables. Le premier permet quant à lui de calculer une intégrale double par le biais du calcul de deux intégrales simples.

1.12 Théorème. (Fubini-Tonelli) Soit μ (resp. ν) une mesure sur l'espace mesurable (E, \mathcal{E}) (resp. (F, \mathcal{F})). On suppose que μ et ν sont σ -finies (μ est σ -finie s'il existe une suite

$(A_n) \in \mathcal{E}$ telle que $\bigcup_n A_n = E$ et $\mu(A_n) < +\infty$. Soit $f : E \times F \rightarrow \mathbb{R}^+$, $\mathcal{E} \otimes \mathcal{F}$ -mesurable, alors

$$\int_{E \times F} f(x, y) d(\mu \otimes \nu)(x, y) = \int_E \left[\int_F f(x, y) d\nu(y) \right] d\mu(x) = \int_F \left[\int_E f(x, y) d\mu(x) \right] d\nu(y).$$

En outre, la fonction f est intégrable sur $E \times F$ si et seulement si les quantités ci-dessus sont finies.

1.13 Théorème. (Fubini) Soit μ (resp. ν) une mesure sur l'espace mesurable (E, \mathcal{E}) (resp. (F, \mathcal{F})). On suppose que μ et ν sont σ -finies. Soit $f : E \times F \rightarrow \mathbb{R}$, $\mathcal{E} \otimes \mathcal{F}$ -mesurable telle qu'il existe $G : F \rightarrow \mathbb{R}$, intégrable sur (F, \mathcal{F}, ν) pour laquelle ν -p.p.,

$$\int_E |f(x, y)| d\mu(x) \leq G(y).$$

Alors f est intégrable sur $E \times F$ et

$$\int_{E \times F} f(x, y) d(\mu \otimes \nu)(x, y) = \int_E \left[\int_F f(x, y) d\nu(y) \right] d\mu(x) = \int_F \left[\int_E f(x, y) d\mu(x) \right] d\nu(y).$$

Une autre notion très importante est la suivante :

1.14 Définition. Soit μ une mesure finie (i.e. telle que $\mu(\mathbb{R}^d) < +\infty$) sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On définit la **transformée de Fourier** de μ par

$$\hat{\mu}(t) = \int_E e^{-2i\pi \langle t, x \rangle} \mu(dx).$$

$\hat{\mu}$ ainsi définie est une application continue sur \mathbb{R}^d , bornée par $\mu(\mathbb{R}^d)$.

La propriété fondamentale de la transformation de Fourier est qu'il existe une **formule d'inversion** permettant de retrouver μ à partir de $\hat{\mu}$.

1.15 Théorème. Soit μ une mesure finie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Soit h continue et bornée de \mathbb{R}^d dans \mathbb{R} . On a alors

$$\int_{\mathbb{R}^d} h(x) \mu(dx) = \lim_{\sigma \downarrow 0} \int_{\mathbb{R}^d} h(x) \left(\int_{\mathbb{R}^d} \hat{\mu}(u) e^{2i\pi \langle u, x \rangle - 2\pi^2 \sigma^2 |u|^2} du \right) dx.$$

Quand f est une fonction Lebesgue intégrable sur \mathbb{R}^d , les résultats précédents s'adaptent de façon agréable. La transformée de f , notée \hat{f} , est définie comme celle de la mesure $f(u) \lambda_{\mathbb{R}^d}(du)$, soit

$$\hat{f}(u) = \int_{\mathbb{R}^d} e^{-2i\pi \langle u, x \rangle} f(x) dx.$$

Quand $\hat{f}(u)$ est elle-même intégrable, la formule d'inversion s'écrit

$$f(u) = \int_{\mathbb{R}^d} e^{2i\pi \langle u, x \rangle} \hat{f}(u) du.$$

Chapitre 2

Espaces et mesures de probabilités

Depuis Kolmogorov (1933), le modèle habituel pour la présentation d'un problème de probabilités rentre dans le cadre de la théorie de la mesure.

2.1 Espaces de probabilités

2.1.1 Définition. On appelle **espace de probabilités** tout espace mesuré $(\Omega, \mathcal{F}, \mathbf{P})$, où Ω est un ensemble non vide, \mathcal{F} une tribu sur Ω et \mathbf{P} une mesure sur l'espace mesurable (Ω, \mathcal{F}) telle que $\mathbf{P}(\Omega) = 1$.

Rappelons que les éléments de \mathcal{F} s'appellent des **événements**, Ω est l'événement certain, et \emptyset l'événement impossible. Par ailleurs, un événement de probabilité 1 est dit presque sûr. Plus généralement, on dira qu'une propriété est vraie **P**-presque sûrement (en abrégé **P**-p.s.) pour signifier qu'elle est vraie avec probabilité 1.

Rappelons une des multiples versions d'un résultat crucial sur les tribus qui sera souvent utilisé. Un ensemble de parties de Ω , \mathcal{M} , est dit une **classe monotone** si elle est stable par réunion dénombrable croissante et par intersection dénombrable décroissante.

Alors on a

2.1.2 Théorème. (de classe monotone). Soient \mathcal{F}_0 une algèbre de Boole et \mathcal{M} une classe monotone. Si $\mathcal{F}_0 \subset \mathcal{M}$, alors $\sigma(\mathcal{F}_0) \subset \mathcal{M}$.

PREUVE : Tout d'abord, de manière manifeste, on constate qu'une intersection de classes monotones est une classe monotone. On peut donc définir une notion de classe monotone engendrée par un ensemble de parties \mathcal{B} , c'est-à-dire une plus petite classe monotone contenant \mathcal{B} . Par ailleurs, il est manifeste qu'une algèbre de Boole qui est une classe monotone est une tribu. De ce fait, pour montrer le résultat, il suffit de montrer que la classe monotone engendrée par \mathcal{F}_0 , notée \mathcal{C} , est une algèbre de Boole.

Soit $\mathcal{G} = \{A, A^c \in \mathcal{C}\}$. Puisque \mathcal{C} est une classe monotone, \mathcal{G} l'est aussi. Par ailleurs, $\mathcal{F}_0 \subset \mathcal{G}$ et donc $\mathcal{C} \subset \mathcal{G}$. Donc \mathcal{C} est stable par complémentaire.

Soit maintenant $A \in \mathcal{F}_0$. Considérons $\mathcal{G}_1 = \{B, B \cup A \in \mathcal{C}\}$. Clairement, $\mathcal{F}_0 \subset \mathcal{G}_1$ et de plus \mathcal{G}_1 est une classe monotone. De ce fait, $\mathcal{C} \subset \mathcal{G}_1$ et l'union d'un élément quelconque de \mathcal{C} avec un élément de \mathcal{F}_0 est donc dans \mathcal{C} .

Posons $\mathcal{G}_2 = \{B, B \cup C \in \mathcal{C}, \forall C \in \mathcal{C}\}$. D'après l'étape précédente, $\mathcal{F}_0 \subset \mathcal{G}_2$. Par ailleurs, \mathcal{G}_2 est une classe monotone contenant \mathcal{F}_0 donc \mathcal{C} qui est stable par union finie. \square

L'application essentielle du résultat précédent est l'identification d'une mesure μ .

2.1.3 Corollaire. Soient deux mesures μ et ν de masse totale finie sur le même espace mesurable (Ω, \mathcal{F}) . Soit \mathcal{F}_0 une algèbre de Boole incluse dans \mathcal{F} . On suppose que $\mu(A) =$

$\nu(A)$ pour tout $A \in \mathcal{F}_0$. Alors μ et ν coïncident sur $\sigma(\mathcal{F}_0)$. En particulier, si $\sigma(\mathcal{F}_0) = \mathcal{F}$, les deux mesures sont égales.

Notons au passage, à toute fin utile, qu'il n'est absolument pas suffisant que les deux mesures coïncident sur un système générateur de la tribu pour avoir le résultat (penser aux mesures δ_0 et $2\delta_0$ sur l'ensemble des intervalles ouverts $]a, b[$ de \mathbb{R} qui ne contiennent pas 0, où δ_0 désigne la mesure de Dirac en 0).

Ci-dessous, un autre résultat du même genre est proposé.

Exercice 2.1

Soit Ω un ensemble. On appelle π -système un ensemble de parties stable par intersection finie. On appelle λ -système un ensemble de parties qui contient l'ensemble vide, est stable par complémentation et par réunion dénombrable disjointe (cette dernière condition est donc moins exigeante que pour une tribu).

- Montrer que toute intersection de λ -systèmes est un λ -système.
- Montrer qu'un π -système qui est aussi un λ -système est une tribu.
- Soient \mathcal{P} un π -système et \mathcal{L} un λ -système tels que $\mathcal{P} \subset \mathcal{L}$. Montrer que $\sigma(\mathcal{P}) \subset \mathcal{L}$ (ce résultat est connu sous le nom de " **$\pi\lambda$ théorème**").

Enfin, mentionnons sans démonstration une forme fonctionnelle des résultats précédents. Elle ne sera en fait utilisée qu'une seule fois dans ce cours, en 13.4.

2.1.4 Théorème. Soit \mathcal{H} un espace vectoriel de fonctions numériques bornées définies sur un ensemble E . On suppose que \mathcal{H} contient les fonctions constantes, et que pour toute suite $(f_n)_{n \geq 0}$ de fonctions de \mathcal{H} , croissante et uniformément bornée, $\lim_n \uparrow f_n \in \mathcal{H}$.

Soit \mathcal{P} un π -système tel que pour tout $A \in \mathcal{P}$ on a $\mathbb{1}_A \in \mathcal{H}$. Alors toute fonction bornée $\sigma(\mathcal{P})$ -mesurable est dans \mathcal{H} .

2.2 Théorème d'extension

Une **probabilité sur une algèbre (de Boole) \mathcal{A}** se définit à l'instar de ce qu'il en est d'une tribu, mais en ne supposant satisfaite que l'additivité finie. On vérifie la caractérisation suivante.

2.2.1 Lemme. (Continuité en \emptyset) Soit \mathbf{P} une probabilité sur l'algèbre \mathcal{A} . Alors \mathbf{P} est σ -additive sur \mathcal{A} si et seulement si \mathbf{P} est continue en \emptyset , à savoir : si $(A_n)_{n \geq 0}$ est une suite dans \mathcal{A} telle que $A_{n+1} \subset A_n$ et $\bigcap_n A_n = \emptyset$, alors $\mathbf{P}(A_n) \rightarrow 0$.

PREUVE : Supposons que \mathbf{P} soit σ -additive. Alors si la suite (A_n) décroît dans \mathcal{A} vers \emptyset , posons $B_n = A_n \setminus A_{n+1}$. On a $B_n \in \mathcal{A}$ et les B_n sont deux à deux disjoints, d'union A_0 . Donc $\sum_n \mathbf{P}(B_n) = \mathbf{P}(A_0)$. Et $\mathbf{P}(A_N) = \sum_{n \geq N} \mathbf{P}(B_n) \rightarrow 0$.

Réciproquement, soit (B_n) une suite dans \mathcal{A} , de termes deux à deux disjoints, et telle que $\bigcup_n B_n = A_0 \in \mathcal{A}$ (cette condition n'est pas nécessairement remplie puisque \mathcal{A} n'est qu'une algèbre de Boole). Alors si nous posons $A_n = \bigcup_{k \geq n} B_k$, la suite (A_n) est dans \mathcal{A} , et décroît, et $\bigcap_n A_n = \emptyset$. Donc $\mathbf{P}(A_n) \rightarrow 0$. Mais $\mathbf{P}(A_0) = \mathbf{P}(\bigcup_{k < n} B_k \cup A_n) = \sum_{k < n} \mathbf{P}(B_k) + \mathbf{P}(A_n)$, et donc $\mathbf{P}(\bigcup_n B_n) = \sum_n \mathbf{P}(B_n)$. \square

On a de plus la définition suivante

2.2.2 Définition. Soit P une probabilité sur l'algèbre de Boole \mathcal{A} . On dit que P est **sous- σ -additive** si pour tout $A \in \mathcal{A}$ et toute suite $(A_n)_{n \geq 0}$ d'éléments de \mathcal{A} telle que $A \subset \bigcup_{n \geq 0} A_n$, on a

$$\mathbf{P}(A) \leq \sum_{n \geq 0} \mathbf{P}(A_n).$$

On a alors le théorème d'extension suivant.

2.2.3 Théorème. (Caratheodory) Soit \mathbf{P} une probabilité sous- σ -additive sur une algèbre de Boole \mathcal{A} . Alors il existe une unique probabilité $\tilde{\mathbf{P}}$ sur $\sigma(\mathcal{A})$, qui coïncide avec \mathbf{P} sur \mathcal{A} .

PREUVE :

L'unicité de l'extension résulte directement du théorème 2.1.3.

La preuve de l'existence repose sur plusieurs lemmes précisant des propriétés de la **mesure extérieure \mathbf{P}^*** , définie pour $E \subset \Omega$ par

$$\mathbf{P}^*(E) = \inf_{\substack{A_n \in \mathcal{A}, \\ E \subset \bigcup_n A_n}} \sum_n \mathbf{P}(A_n).$$

Considérons

$$\mathcal{M} = \{A \subset \Omega : \forall E \subset \Omega, \mathbf{P}^*(A \cap E) + \mathbf{P}^*(A^c \cap E) = \mathbf{P}^*(E)\}.$$

Nous allons montrer que sur \mathcal{M} , \mathbf{P}^* est une probabilité, que \mathcal{M} est une tribu, et que $\mathcal{A} \subset \mathcal{M}$, et enfin que \mathbf{P}^* et \mathbf{P} coïncident sur \mathcal{A} .

Les propriétés suivantes de \mathbf{P}^* sont élémentaires à démontrer (propriétés de l'inf) :

- (i) : $\mathbf{P}^*(\emptyset) = 0$;
- (ii) : $\mathbf{P}^* \geq 0$ (positivité);
- (iii) : $A \subset B \Rightarrow \mathbf{P}^*(A) \leq \mathbf{P}^*(B)$ (monotonie);

On a en outre

2.2.4 Lemme. \mathbf{P}^* est sous- σ -additive.

PREUVE : Soit $(A_n)_{n \geq 1}$ une suite de parties de Ω . Soit $\varepsilon > 0$. Pour tout $n \geq 0$, soit $(B_k^n)_{k \geq 1}$ une suite dans \mathcal{A} telle que

$$\sum_k P(B_k^n) \leq \mathbf{P}^*(A_n) + \frac{\varepsilon}{2^n}.$$

On a $\bigcup_n A_n \subset \bigcup_{k,n} B_k^n$ et donc

$$\mathbf{P}^*\left(\bigcup_n A_n\right) \leq \sum_{k,n} P(B_k^n) = \sum_n \sum_k P(B_k^n) \leq \sum_n \mathbf{P}^*(A_n) + \sum_n \frac{\varepsilon}{2^n} \leq \sum_n \mathbf{P}^*(A_n) + \varepsilon$$

d'où $\mathbf{P}^*\left(\bigcup_n A_n\right) \leq \sum_n \mathbf{P}^*(A_n)$. □

On en déduit immédiatement la conséquence suivante

2.2.5 Corollaire. Les deux propriétés suivantes sont équivalentes

- (i) $A \in \mathcal{M}$
- (ii) $\mathbf{P}^*(A \cap E) + \mathbf{P}^*(A^c \cap E) \leq \mathbf{P}^*(E)$ pour tout $E \subset \Omega$.

On a

2.2.6 Lemme. \mathcal{M} est une algèbre.

PREUVE : on sait déjà que \mathcal{M} contient \emptyset . Par ailleurs, par définition de \mathcal{M} , elle est clairement stable par complémentation. Soient $A, B \in \mathcal{M}$ et soit $E \subset \Omega$: alors

$$\begin{aligned} \mathbf{P}^*(E) &= \mathbf{P}^*(B \cap E) + \mathbf{P}^*(B^c \cap E) \text{ car } B \in \mathcal{M} \\ &= \mathbf{P}^*(A \cap B \cap E) + \mathbf{P}^*(A^c \cap B \cap E) + \mathbf{P}^*(A \cap B^c \cap E) \\ &\quad + \mathbf{P}^*(A^c \cap B^c \cap E) \text{ car } A \in \mathcal{M}, \\ &\geq \mathbf{P}^*(A \cap B \cap E) + \mathbf{P}^*(A^c \cap E) + \mathbf{P}^*(A \cap B^c \cap E) \text{ grâce à la sous-}\sigma\text{-additivité de } \mathbf{P}^*. \end{aligned}$$

Comme $(A^c \cap E) \cup (A \cap B^c \cap E) = (A^c \cap E) \cup (B^c \cap E) = (A \cap B)^c \cap E$, on a, toujours par la sous- σ -additivité de \mathbf{P}^* , $\mathbf{P}^*(A^c \cap E) + \mathbf{P}^*(A \cap B^c \cap E) \geq \mathbf{P}^*((A \cap B)^c \cap E)$ et donc

$$\mathbf{P}^*(E) \geq \mathbf{P}^*(A \cap B \cap E) + \mathbf{P}^*((A \cap B)^c \cap E).$$

De ce fait, par le Corollaire 2.2.5, $A \cap B \in \mathcal{M}$. □

2.2.7 Lemme. Si les A_n constituent une suite finie ou infinie d'éléments deux à deux disjoints de \mathcal{M} , alors quel que soit $E \subset \Omega$,

$$\mathbf{P}^*(E \cap (\cup_n A_n)) = \sum_n \mathbf{P}^*(E \cap A_n).$$

PREUVE : Soient A_1 et A_2 disjoints dans \mathcal{M} . Alors

$$\begin{aligned} &\mathbf{P}^*(E \cap (A_1 \cup A_2)) \\ &= \mathbf{P}^*(E \cap (A_1 \cup A_2) \cap A_1) + \mathbf{P}^*(E \cap (A_1 \cup A_2) \cap A_1^c) \text{ puisque } A_1 \in \mathcal{M} \\ &= \mathbf{P}^*(E \cap A_1) + \mathbf{P}^*(E \cap A_2) \text{ puisque } A_1 \cap A_2 = \emptyset. \end{aligned}$$

Par récurrence, si $A_1, \dots, A_n \in \mathcal{M}$, on obtient facilement que

$$\mathbf{P}^*(E \cap (\cup_{k=1}^n A_k)) = \sum_{k=1}^n \mathbf{P}^*(E \cap A_k).$$

Soit maintenant une suite $(A_n)_{n \geq 0}$ d'ensembles disjoints dans \mathcal{M} . Par la monotonie de \mathbf{P}^* et par la propriété qu'on vient de démontrer, on a

$$\mathbf{P}^*(E \cap (\bigcup_k A_k)) \geq \mathbf{P}^*(E \cap (\bigcup_{k \leq n} A_k)) = \sum_{k \leq n} \mathbf{P}^*(E \cap A_k).$$

De ce fait, en passant à la limite, $\sum_k \mathbf{P}^*(E \cap A_k) \leq \mathbf{P}^*(E \cap (\bigcup_k A_k))$.

L'inégalité contraire résulte immédiatement de la sous- σ -additivité de \mathbf{P}^* . □

2.2.8 Lemme. \mathcal{M} est une tribu, et \mathbf{P}^* restreinte à \mathcal{M} est σ -additive.

PREUVE : Puisque \mathcal{M} est une algèbre, il suffit, pour montrer qu'elle est une tribu, de montrer qu'elle est stable par unions dénombrables disjointes.

Soit (A_n) une suite disjointe dans \mathcal{M} et posons $U_n = \cup_{k \leq n} A_k$, $n \geq 1$. \mathcal{M} étant une algèbre, on a $U_n \in \mathcal{M}$.

Posons aussi $U = \cup_n A_n$. On a $\mathbf{P}^*(E \cap U_n) + \mathbf{P}^*(E \cap U_n^c) = \mathbf{P}^*(E)$, pour tout n et $E \subset \Omega$.

Par monotonie, $\mathbf{P}^*(E \cap U_n^c) \geq \mathbf{P}^*(E \cap U^c)$. De plus, par le Lemme 2.2.7,

$$\mathbf{P}^*(E \cap U_n) = \sum_{k \leq n} \mathbf{P}^*(E \cap A_k).$$

Pour chaque n , on a donc

$$\mathbf{P}^*(E) \geq \mathbf{P}^*(E \cap U_n) + \mathbf{P}^*(E \cap U^c) \geq \sum_{k \leq n} \mathbf{P}^*(E \cap A_k) + \mathbf{P}^*(E \cap U^c).$$

Par passage à la limite, $\mathbf{P}^*(E) \geq \sum_n \mathbf{P}^*(E \cap A_n) + \mathbf{P}^*(E \cap U^c)$, ce qui par le Lemme 2.2.7 s'écrit $\mathbf{P}^*(E) \geq \mathbf{P}^*(E \cap U) + \mathbf{P}^*(E \cap U^c)$. Donc $U \in \mathcal{M}$ par le Corollaire 2.2.5.

La σ -additivité de \mathbf{P}^* sur \mathcal{M} résulte immédiatement du Lemme 2.2.7 avec $E = \Omega$. \square

2.2.9 Lemme. *On a l'inclusion $\mathcal{A} \subset \mathcal{M}$.*

PREUVE : Soient $A \in \mathcal{A}$, $\varepsilon > 0$, et $E \subset \Omega$. Soit $(A_n)_{n \geq 0}$ une suite d'ensembles dans \mathcal{A} telle que $E \subset \cup_n A_n$ et $\sum_n \mathbf{P}(A_n) \leq \mathbf{P}^*(E) + \varepsilon$. Soient $B_n = A \cap A_n$ et $C_n = A^c \cap A_n$. Notons que B_n et $C_n \in \mathcal{A}$ puisque \mathcal{A} est une algèbre.

On a $E \cap A \subset \cup_n B_n$, et $E \cap A^c \subset \cup_n C_n$. Par définition de \mathbf{P}^* et additivité finie de \mathbf{P} , on a

$$\mathbf{P}^*(E \cap A) + \mathbf{P}^*(E \cap A^c) \leq \sum_n \mathbf{P}(B_n) + \sum_n \mathbf{P}(C_n) = \sum_n \mathbf{P}(A_n) \leq \mathbf{P}^*(E) + \varepsilon,$$

ce qui démontre le lemme en faisant tendre ε vers 0. \square

Fin de la preuve du théorème 2.2.3 : il ne reste plus qu'à vérifier que $\mathbf{P}^* = \mathbf{P}$ sur \mathcal{A} . Par définition de \mathbf{P}^* , $\mathbf{P}^*(A) \leq \mathbf{P}(A)$ si $A \in \mathcal{A}$.

Si $(A_n)_{n \geq 0}$ est une suite dans \mathcal{A} qui recouvre A , par sous- σ -additivité de \mathbf{P} sur \mathcal{A} , il vient $\mathbf{P}(A) \leq \sum_n \mathbf{P}(A_n)$, et en passant à l'inf sur ces recouvrements, on obtient $\mathbf{P}(A) \leq \mathbf{P}^*(A)$. \square

Remarque. La classe \mathcal{M} à laquelle s'étend \mathbf{P} est souvent bien plus large que $\sigma(\mathcal{A})$. Elle contient notamment la complétion de la tribu $\sigma(\mathcal{A})$, c'est-à-dire les sous-ensembles de Ω contenus dans les ensembles de $\sigma(\mathcal{A})$ qui sont de \mathbf{P}^* mesure nulle.

Application : Construction de la mesure de Lebesgue λ sur $[0, 1]$.

On considère l'algèbre \mathcal{A} des unions finies disjointes de sous-intervalles de $[0, 1]$ (vérifier que c'est une algèbre).

Pour $\cup_i I_i \in \mathcal{A}$, on pose $\lambda(\cup_i I_i) = \sum_i \text{longueur}(I_i)$. Il faut vérifier que c'est une bonne définition, puis, que si on considère une union dénombrable disjointe de telles unions, alors la série des sommes de longueurs redonne bien la somme des longueurs initiales. On peut plutôt utiliser le Lemme 2.2.1 et vérifier que si $(A_n) \in \mathcal{A}$ est telle que $A_{n+1} \subset A_n$ et que $\cap_n A_n = \emptyset$, alors $\lambda(A_n) \rightarrow 0$.

On peut alors poursuivre la construction précédente. Donnons nous (A_n) une suite décroissante d'unions finies d'intervalles de $[0, 1]$ telle que $\cap_n A_n = \emptyset$. Nous voulons vérifier que $\lambda(A_n) \rightarrow 0$. Soit $\varepsilon > 0$. Pour chaque n , nous pouvons trouver une union finie disjointe d'intervalles fermés $K_n^0 \subset A_n$ telle que

$$\lambda(A_n \setminus K_n^0) \leq \varepsilon / 2^{n+1}.$$

Posons $K_n = K_1^0 \cap \dots \cap K_n^0$. Alors la suite (K_n) est une suite décroissante de compacts de \mathbb{R} , et d'intersection vide. Il s'ensuit qu'il existe n_0 tel que $n \geq n_0 \Rightarrow K_n = \emptyset$. Alors pour $n \geq n_0$,

$$\lambda(A_n) = \lambda(A_n \setminus K_n) = \lambda(A_n \cap (\cap_{s \leq n} K_s^0)^c) \leq \sum_{s \leq n} \lambda(A_s \cap (K_s^0)^c) \leq \sum_s \varepsilon / 2^{s+1} = \varepsilon,$$

et donc on a montré que $\lambda(A_n) \rightarrow 0$.

Par application du Lemme de continuité en \emptyset , il s'ensuit que λ est une probabilité sur \mathcal{A} . Le théorème d'existence nous donne alors l'existence d'une unique mesure de probabilité sur $([0, 1], \sigma(\mathcal{A}))$ telle que $\lambda(I) = \text{longueur}(I)$ si $I \subset [0, 1]$ est un intervalle. C'est la mesure de Lebesgue sur $[0, 1]$.

Remarquons que $\sigma(\mathcal{A}) = \{A \cap [0, 1] : A \in \mathcal{B}(\mathbb{R})\}$, où $\mathcal{B}(\mathbb{R})$ est la tribu borélienne de \mathbb{R} .

2.3 Lemme de Borel-Cantelli

Le résultat élémentaire principal du modèle probabiliste est le Lemme de Borel-Cantelli qui permet d'obtenir une probabilité faisant intervenir une infinité d'événements sous une hypothèse technique assurant que les événements restent individuellement de probabilité assez petite.

2.3.1 Théorème. (lemme de Borel-Cantelli) Soit (A_n) une suite d'événements de l'espace de probabilités $(\Omega, \mathcal{F}, \mathbf{P})$. On suppose que $\sum_{n \geq 0} \mathbf{P}(A_n) < +\infty$. Soit A l'événement "une infinité de A_i se produisent simultanément". Alors, $\mathbf{P}(A) = 0$.

PREUVE : Il suffit de remarquer que $A = \bigcap_{N \geq 0} \bigcup_{n \geq N} A_n$. On a alors

$$\mathbf{P}(A) = \lim_{N \rightarrow +\infty} \downarrow \mathbf{P}\left(\bigcup_{n \geq N} A_n\right) \leq \lim_{N \rightarrow +\infty} \sum_{n \geq N} \mathbf{P}(A_n) = 0. \quad \square$$

Il peut sembler étonnant qu'une majoration aussi grossière que celle dont résulte le théorème précédent puisse être utile. En fait, il faut plutôt sentir son importance comme provenant du fait qu'on ne peut souvent rien écrire d'autre.

◇ Exercice 2.2

Soit l'espace de probabilités $([0, 1], \mathcal{B}([0, 1]), \mathbf{P})$ où \mathbf{P} est la mesure de Lebesgue.

On définit pour $\omega \in \Omega$, $d_n(\omega) = \omega_n$ où $\omega = \sum_{n \geq 1} \frac{\omega_n}{2^n}$ est la décomposition dyadique

de ω . Soit $\ell_n(\omega)$ la longueur de la chaîne de zéros partant de $d_n(\omega)$: $\ell_n(\omega) = k$ signifie que $d_n(\omega) = \dots = d_{n+k-1}(\omega) = 0$ et $d_{n+k}(\omega) = 1$.

a) Montrer que $\mathbf{P}(\{\omega, \ell_n(\omega) \geq r\}) = 2^{-r}$.

b) Soit $(r_n)_{n \geq 0}$ une suite de réels telle que $\sum_{n \geq 0} 2^{-r_n} < +\infty$.

Montrer que $\mathbf{P}(\{\omega, \ell_n(\omega) \geq r_n \text{ infiniment souvent}\}) = 0$.

c) Soit $e_n(\omega) = \frac{\omega - \sum_{k=1}^{n-1} d_k(\omega)2^{-k}}{2^{-n+1}}$ l'erreur relative commise en estimant ω par son développement dyadique limité à l'ordre $n-1$. Alors, pour $\varepsilon > 0$,

$$\mathbf{P}(\{\omega, e_n(\omega) \leq \frac{1}{n^{1+\varepsilon}} \text{ infiniment souvent}\}) = 0.$$

2.4 Probabilité conditionnelle et indépendance

Commençons par quelques rappels de notions élémentaires.

2.4.1 Définition. Soient A et B deux événements tels que $\mathbf{P}(B) \neq 0$. On appelle **probabilité conditionnelle** de A sachant B le réel $\mathbf{P}(A/B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$.

Il faut bien sûr comprendre intuitivement la définition précédente comme mesurant la probabilité normalisée de ceux des $\omega \in B$ qui réalisent A . En quelque sorte, puisqu'on connaît le fait que B se réalise, l'espace de probabilités peut être restreint à B . Le cas particulier majeur de cette situation est celui où B n'a pas d'influence sur A , ce qui s'exprime aussi en disant que conditionner ou non par B ne change pas la probabilité de A . Ce qui revient à dire que $\mathbf{P}(A/B) = \mathbf{P}(A)$ ou encore $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. Prolongeant cette définition au cas où $\mathbf{P}(B) = 0$, on a la définition classique suivante

2.4.2 Définition. a) Deux événements A et B sont **indépendants** si on a $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. On utilisera souvent la notation $A \perp\!\!\!\perp B$.

b) Deux sous-tribus \mathcal{G} et \mathcal{H} de \mathcal{F} sont indépendantes si et seulement si tout élément de \mathcal{G} et tout élément de \mathcal{H} sont indépendants.

L'exercice suivant montre, si besoin était, qu'il n'y a aucun rapport entre l'indépendance et la disjonction.

Exercice 2.3

Montrer que A et A^c sont indépendants si et seulement si $\mathbf{P}(A) = 0$ ou 1 .

On prolonge comme d'habitude cette définition.

2.4.3 Définition. Une famille d'évènements $(A_i)_{i \in I}$ est indépendante si quel que soit $J \subset I$ fini,

$$\mathbf{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbf{P}(A_j).$$

La famille $(A_i)_{i \in I}$ est **deux à deux indépendante** si quels que soient $i \neq j \in I$, $A_i \perp\!\!\!\perp A_j$.

Soit $(\mathcal{B}_i)_{i \in I}$ une famille de parties de \mathcal{F} . La famille (\mathcal{B}_i) est indépendante si quelle que soit la famille $(A_i) \in \prod \mathcal{B}_i$, elle est indépendante.

Nous dirons que la famille (\mathcal{B}_i) est **deux à deux indépendante** si quels que soient $i \neq j \in I$, $A_i \in \mathcal{B}_i$, et $A_j \in \mathcal{B}_j$, $A_i \perp\!\!\!\perp A_j$.

2.4.4 Proposition. (Borel-Cantelli) Si la suite d'évènements (A_n) est indépendante, alors, si $\sum_n \mathbf{P}(A_n) = \infty$ et si A est l'évènement "une infinité de A_i se produisent simultanément", on a $\mathbf{P}(A) = 1$.

PREUVE : Soit $C_n = \cap_{p \geq n} A_p^c$, alors $\mathbf{P}(\cup_n \uparrow C_n) = \lim_n \uparrow \mathbf{P}(C_n)$. Or si $x \geq 0$, $1 - x \leq e^{-x}$, et donc

$$\mathbf{P}(C_n) = \lim_k \prod_{i=n}^{n+k} \mathbf{P}(A_i^c) = \lim_k \prod_{i=n}^{n+k} (1 - \mathbf{P}(A_i)) \leq \lim_k e^{-\sum_{i=n}^{n+k} \mathbf{P}(A_i)} = 0$$

Donc pour chaque n , $\mathbf{P}(C_n) = 0$ et $\mathbf{P}(C_n^c) = 1$. Or $A = \cap_n C_n^c$ et donc $\mathbf{P}(A) = 1$. \square

Exercice 2.4

Soit $E = \{0, 1\}^{\mathbb{N}}$. On note \mathcal{A} l'ensemble des parties A de E qui dépendent d'un nombre fini de coordonnées, au sens où il existe $n \geq 1$ et $A_n \subset \{0, 1\}^n$ tels que

$$A = A_n \times \{0, 1\} \times \{0, 1\} \times \dots$$

- a) Montrer que \mathcal{A} est une algèbre de Boole.
 b) Montrer qu'il existe une unique probabilité \mathbf{P} sur \mathcal{A} telle que pour tout $(\omega_0, \omega_1, \dots, \omega_n) \in \{0, 1\}^{n+1}$ on ait

$$\mathbf{P}(\{\omega_0\} \times \{\omega_1\} \times \dots \times \{\omega_n\} \times \{0, 1\} \times \{0, 1\} \times \dots) = \frac{1}{2^{n+1}}.$$

- c) Montrer que \mathbf{P} est sous- σ -additive.
 d) Soit $\mathcal{T} = \sigma(\mathcal{A})$. On note \mathbf{P}_∞ la probabilité prolongeant \mathbf{P} sur \mathcal{T} . Sur $(\{0, 1\}^{\mathbb{N}}, \mathcal{T}, \mathbf{P}_\infty)$, posons $A_k = \{\omega_k = 1\}$. Vérifier que la suite (A_k) est indépendante. Evaluer $\sum_k \mathbf{P}_\infty(A_k)$, et en déduire que \mathbf{P}_∞ -presque sûrement, ω possède une infinité de 1 et de 0. Utiliser deux méthodes (par le premier lemme de Borel-Cantelli puis par le second).

Chapitre 3

Variables aléatoires

Dans la suite sauf mention contraire, $(\Omega, \mathcal{F}, \mathbf{P})$ désignera un espace de probabilités.

3.1 Définitions et premières propriétés

Une variable aléatoire va servir à concentrer le hasard sur le résultat d'une expérience aléatoire. Quand on considère une variable aléatoire, on s'intéresse à l'effet produit par le hasard sur une telle expérience. Avec cette idée, la définition suivante est naturelle.

3.1.1 Définition. On appelle **variable aléatoire** à valeurs dans l'espace mesurable (E, \mathcal{E}) toute application mesurable X de (Ω, \mathcal{F}) dans (E, \mathcal{E}) .

Si $E = \mathbb{R}$ muni de sa tribu borélienne, on parle de variables aléatoire réelle (en abrégé, on emploie couramment la notation **v.a.r.**).

Un **vecteur aléatoire** est une application $(X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ dont les composantes sont des variables aléatoires réelles.

Quand $E = \mathbb{R}$, on définit (quand l'intégrale existe) l'**espérance** (ou **moment d'ordre 1**) de X par $\mathbf{E}_{\mathbf{P}}(X) = \int_{\Omega} X(\omega) d\mathbf{P}(\omega)$. Quand il n'y a pas d'ambiguïté sur la probabilité sous-jacente, on écrira \mathbf{E} au lieu de $\mathbf{E}_{\mathbf{P}}$.

On fera naturellement l'abus standard en supposant que cette quantité est toujours définie (et éventuellement infinie) dans le cas où X est à valeurs positives. On dit que la loi est **centrée** si $\mathbf{E}(X) = 0$.

Plus généralement, le **moment d'ordre** $p \geq 1$ de la variable aléatoire réelle X , quand cette quantité est définie, est donné par $\mathbf{E}(X^p)$. L'inégalité de Hölder montre alors que si X admet un moment d'ordre $p \geq 1$, elle admet des moments de tous ordres $1 \leq q \leq p$.

Par ailleurs, la **variance** de X admettant un moment d'ordre 2 est donnée par

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

Quand la variance égale 1, on dit que la loi est **réduite**.

3.1.2 Propriété. La variance est invariante par translation i.e.

$$\text{Var}(X + c) = \text{Var}(X),$$

et une variable sans dispersion est constante \mathbf{P} -p.s., à savoir que

$$\text{Var}(X) = 0 \iff X = \mathbf{E}(X) \mathbf{P} - p.s..$$

3.1.3 Définition. L'**écart-type** de la variable aléatoire X admettant un moment d'ordre 2 est la racine de la variance, $\sigma(X) = \sqrt{\text{Var}(X)}$.

L'espérance d'un vecteur aléatoire est définie par

$$\mathbf{E}((X_1, \dots, X_d)) = (\mathbf{E}(X_1), \dots, \mathbf{E}(X_d)).$$

Pour les vecteurs aléatoires, nous bénéficions des mêmes théorèmes généraux d'inter-version de limite et d'intégrale que pour les variables aléatoires, à ceci près qu'il faut en adapter l'énoncé de sorte à ce qu'ils intègrent simultanément les conditions requises composante à composante, à savoir pour chaque X_i .

3.1.4 Définition. Soit un vecteur aléatoire $X = (X_1, \dots, X_d)$, avec $X_i \in L^2(\mathbf{P})$, la **matrice de dispersion** ou **matrice des variances-covariances** est donnée par

$$\text{Var}(X) = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq d},$$

où $\text{Cov}(X_i, X_j)$ désigne la **covariance des variables** X_i et X_j , définie, pour deux variables aléatoires $Y, Z \in L^2(\mathbf{P})$, par

$$\text{Cov}(Y, Z) = \mathbf{E}((Y - \mathbf{E}(Y))(Z - \mathbf{E}(Z))).$$

◇ Exercice 3.1

Soit un vecteur aléatoire $X = (X_1, \dots, X_n)$ centré, avec $X_i \in L^2(\mathbf{P})$, de matrice de dispersion $D_X = I_n$. Soit A une matrice $n \times n$.

- Quelle est la dispersion du vecteur aléatoire AX ?
- Soit X^* le vecteur transposé du vecteur X . Montrer que la variable aléatoire réelle X^*AX admet une espérance égale à $\text{tr}(A)$.

Rappelons immédiatement deux importantes inégalités de base liées aux moments.

3.1.5 Proposition. Soit X une variable aléatoire réelle.

- (i) (Inégalité de Markov) Si $X \geq 0$, alors pour tout $a > 0$,

$$\mathbf{P}(X > a) \leq \frac{\mathbf{E}(X)}{a}.$$

- (ii) (Inégalité de Bienaymé-Chebyshev) Quand X admet une variance, pour tout $\varepsilon > 0$,

$$\mathbf{P}(|X - \mathbf{E}(X)| > \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}.$$

PREUVE : (i) sera obtenu un peu plus tard comme conséquence d'une propriété plus générale (voir Proposition 3.2.4). (ii) s'obtient en prenant comme variable aléatoire positive $|X - \mathbf{E}(X)|^2$ et $a = \varepsilon^2$ dans (i). □

3.2 Lois des variables aléatoires

Le transfert du hasard se fait à travers le concept central de loi de la variable aléatoire X .

3.2.1 Définition. On appelle **loi** de la variable aléatoire X à valeurs dans l'espace mesurable (E, \mathcal{E}) , la mesure de probabilités sur (E, \mathcal{E}) image de \mathbf{P} par X . On la notera classiquement μ_X . En d'autres termes, μ_X est la mesure définie sur (E, \mathcal{E}) par $\mu_X(A) = \mathbf{P}(X \in A)$ où la notation $(X \in A)$ désigne l'événement $\{\omega, X(\omega) \in A\}$.

Insistons sur l'interprétation capitale de cette notion. Quand on considère la loi de X , on oublie complètement ce qui se passe sur l'espace Ω . Tout se passe comme si le hasard se concentrait désormais sur l'espace E avec la loi μ_X qui devient de ce fait le nouvel espace de probabilités. Ceci peut être particulièrement intéressant si l'ensemble E (contrairement à Ω) s'avère être un ensemble ayant de bonnes propriétés comme \mathbb{R}^d ou des espaces de fonctions. De multiples exemples montreront par la suite l'intérêt d'un tel transfert.

Pour commencer, donnons quelques exemples de lois importantes de variables aléatoires réelles. Il s'agit donc de probabilités sur \mathbb{R} , et même plus précisément sur l'ensemble où la variable prend ses valeurs.

Regardons tout d'abord le cas où la variable serait à valeurs dans \mathbb{N} . La loi pourra alors s'écrire sous la forme $\sum_{n \geq 0} p_n \delta_n(dx)$ où δ_n est la **masse de Dirac** au point n , c'est-à-dire la mesure telle que $\delta_n(A) = \mathbb{I}_A(n)$, et $(p_n)_{n \geq 0}$ est une suite de réels positifs tels que $\sum_n p_n = 1$. Voyons quelques exemples classiques de choix pour la suite (p_n) .

La **Loi de Bernoulli** de paramètre $0 \leq p \leq 1$ est définie par $p_1 = p = 1 - p_0$ (et donc *a fortiori* $p_k = 0, k \geq 2$).

La **loi binomiale** de paramètres $n \in \mathbb{N}^*$ et $0 \leq p \leq 1$ est, elle, définie par

$$p_k = C_n^k p^k (1-p)^{n-k}, 0 \leq k \leq n$$

(et $p_j = 0, j \geq n+1$).

La **loi géométrique** de paramètre $1 > a > 0$ est donnée par

$$p_k = (1-a)a^k, k \geq 0.$$

Enfin, la **loi de Poisson** de paramètre $\lambda > 0$ est donnée par

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!}, k \geq 0.$$

Un autre cas spécialement intéressant est celui où la loi est définie à travers la donnée d'une fonction mesurable positive f , d'intégrale 1 par rapport à la mesure de Lebesgue c'est à dire telle que $\int_{\mathbb{R}} f(x) dx = 1$. On dit alors que f est la **densité** de la loi.

La **loi uniforme** sur $[a, b]$, $a < b$ est définie par le choix $f(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x)$.

La **loi normale** de paramètres $m \in \mathbb{R}$ et $\sigma > 0$ (pour des raisons qui apparaîtront très rapidement, on parle plutôt des paramètres m et σ^2) est définie par le choix

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

Dans le cas particulier où $m = 0$ et $\sigma = 1$, on dit qu'il s'agit d'une loi normale **centrée, réduite**.

La **loi exponentielle** de paramètre λ est définie par le choix

$$f(x) = \lambda e^{-\lambda x} \mathbb{I}_{x > 0}.$$

L'utilisation standard de la notion de loi se fait à travers le résultat suivant.

3.2.2 Proposition. (théorème de transfert) Soit (Ω', \mathcal{B}') un autre espace probabilisable et soit $f : \Omega \rightarrow \Omega'$ une application $(\mathcal{B}, \mathcal{B}')$ -mesurable. Alors pour toute application borélienne $g : \Omega' \rightarrow \mathbb{R}^d$, on a

$$g \in L^1(\mu_f) \iff g \circ f \in \mathcal{L}^1(\mathbf{P}).$$

Si tel est le cas, alors

$$\mathbf{E}_{\mathbf{P}}(g \circ f) = \mathbf{E}_{\mu_f}(g).$$

PREUVE : par définition, si $g : \Omega' \rightarrow \mathbb{R}^d$, son intégrabilité équivaut à celle de toutes ses composantes, et son intégrale est le vecteur des intégrales de ses composantes. Donc restreignons nous à ne traiter que le cas de $g : \Omega' \rightarrow \mathbb{R}$.

Par définition de μ_f , si $A \in \mathcal{B}'$, et $g = \mathbb{1}_A$, $g \circ f = \mathbb{1}_{f^{-1}A}$ sont intégrables, et

$$\mathbf{E}_{\mathbf{P}}(g \circ f) = \mathbf{P}(f^{-1}A) = \mu_f(A) = \mathbf{E}_{\mu_f}(g).$$

Donc l'énoncé est vrai pour g une variable aléatoire simple, i.e.

$$g = \sum_{\text{finie}} \alpha_i \mathbb{1}_{g^{-1}\{\alpha_i\}},$$

où $(g^{-1}\{\alpha_i\})$ est une partition \mathcal{B}' -mesurable de Ω' .

Si g est positive alors elle est limite d'une suite croissante (g_n) de variables aléatoires simples et positives, et donc par Beppo-Levi, et la validité de notre énoncé pour les fonctions simples,

$$\mathbf{E}_{\mu_f}(g) = \lim_n \uparrow \mathbf{E}_{\mu_f}(g_n) = \lim_n \uparrow \mathbf{E}_{\mathbf{P}}(g_n \circ f) = \mathbf{E}_{\mathbf{P}}(g \circ f),$$

car $g_n \circ f \rightarrow g \circ f$ simplement, en étant minorée par zéro, et en croissant.

Il en résulte notamment que si g est positive,

$$g \in L^1(\mu_f) \Leftrightarrow \mathbf{E}_{\mu_f}(g) < \infty \Leftrightarrow \mathbf{E}_{\mathbf{P}}(g \circ f) < \infty \Leftrightarrow g \circ f \in L^1(\mathbf{P}).$$

Donc l'énoncé est vrai pour les variables aléatoires positives. Dans le cas général, on passe par les parties positives $g^+ = \max(0, g)$ et $g^- = \max(0, -g)$. On remarque que $(g \circ f)^+ = g^+ \circ f$ et $(g \circ f)^- = g^- \circ f$, et on écrit

$$\begin{aligned} & g \in L^1(\mu_f) \\ \Leftrightarrow & g^+ \text{ et } g^- \in L^1(\mu_f) \\ \Leftrightarrow & \max(\mathbf{E}_{\mu_f}(g^+), \mathbf{E}_{\mu_f}(g^-)) < \infty \\ \Leftrightarrow & \max(\mathbf{E}_{\mathbf{P}}(g^+ \circ f), \mathbf{E}_{\mathbf{P}}(g^- \circ f)) < \infty \Leftrightarrow (g \circ f)^+ \text{ et } (g \circ f)^- \in L^1(\mathbf{P}) \\ \Leftrightarrow & g \circ f \in L^1(\mathbf{P}). \end{aligned}$$

Puis on conclut par

$$\mathbf{E}_{\mu_f}(g) = \mathbf{E}_{\mu_f}(g^+) - \mathbf{E}_{\mu_f}(g^-) = \mathbf{E}_{\mathbf{P}}((g \circ f)^+) - \mathbf{E}_{\mathbf{P}}((g \circ f)^-) = \mathbf{E}_{\mathbf{P}}(g \circ f). \quad \square$$

3.2.3 Remarque. Supposons que nous ne connaissions de la variable aléatoire $X : \Omega \rightarrow \mathbb{R}$ que sa loi μ_X sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors, bien que nous ne connaissions pas l'application X elle-même, nous pouvons décider de l'intégrabilité de $g(X)$ où $g : \mathbb{R} \rightarrow \mathbb{R}$ est borélienne, et en plus calculer $\mathbf{E}_{\mathbf{P}}(g(X))$. Il suffit en effet pour cela de vérifier que $g \in L^1(\mu_X)$, et si c'est le cas, de calculer

$$\mathbf{E}_{\mathbf{P}}(g(X)) = \mathbf{E}_{\mu_X}(g) = \int_{\mathbb{R}} g(x) d\mu_X(x).$$

3.2.4 Proposition. Soit $\psi : \mathbb{R} \rightarrow \mathbb{R}^+$ borélienne et pour $A \in \mathcal{B}(\mathbb{R})$, posons

$$i_A = \inf\{\psi(x) : x \in A\}.$$

Alors si X est une v.a. sur (Ω, \mathcal{B}, P) ,

$$\mathbf{P}(X \in A) \leq \frac{1}{i_A} \mathbf{E}(\psi(X)).$$

PREUVE : On a

$$\begin{aligned} i_A \mathbf{P}(X \in A) &= \mathbf{E}(i_A \mathbb{I}_A(X)) = \mathbf{E}_{\mu_X}(i_A \mathbb{I}_A(x)) \\ &\leq \mathbf{E}_{\mu_X}(\psi(x) \mathbb{I}_A(x)) \leq \mathbf{E}_{\mu_X}(\psi(x)) = \mathbf{E}(\psi(X)), \end{aligned}$$

où aucune condition d'intégrabilité n'est à vérifier puisque ψ est borélienne positive. \square

La proposition 3.1.5 est une conséquence immédiate de ce résultat avec $a > 0$ et $A =]a, +\infty[$ et $\psi(x) = |x|$. Une autre inégalité est obtenue avec un autre choix de ψ .

3.2.5 Corollaire. (Cantelli) On suppose $X \in L^2(\mathbf{P})$, avec $\mathbf{E}(X) = 0$. On pose $\sigma^2 = \text{Var}(X)$, et on prend $a > 0$, et $\psi(x) = (x + \frac{\sigma^2}{a})^2$. Enfin on prend $A = [a, +\infty[$. Alors on a

$$\mathbf{P}(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}.$$

L'exercice suivant donne la valeur des espérances de lois usuelles.

Exercice 3.2

- a) Si la loi de X est une loi binomiale de paramètres n et p , $\mathbf{E}(X) = np$ et $\text{Var}(X) = np(1-p)$.
- b) Si la loi de X est une loi de Poisson de paramètre $\lambda > 0$, $\mathbf{E}(X) = \text{Var}(X) = \lambda$.
- c) Si la loi de X est une loi uniforme sur $[a, b]$, $\mathbf{E}(X) = \frac{a+b}{2}$ et la variance $\text{Var}(X)$ vaut $\frac{(b-a)^2}{12}$.
- d) Si la loi de X est normale de paramètres m et σ^2 , $\mathbf{E}(X) = m$ et $\text{Var}(X) = \sigma^2$.
- e) Si la loi de X est exponentielle de paramètre $\lambda > 0$, $\mathbf{E}(X) = \frac{1}{\lambda}$ et $\text{Var}(X) = \frac{1}{\lambda^2}$.

Il existe aussi bien entendu des lois ne possédant pas d'espérance :

◇ Exercice 3.3

Soit X une v.a.r. de loi de Cauchy de paramètre 1 i.e. de densité $\frac{1}{\pi} \frac{1}{1+x^2}$. Montrer que X n'a pas d'espérance.

On peut cependant rendre intégrable une variable aléatoire (et même toutes les variables d'une suite) en changeant convenablement la probabilité.

3.2.6 Définition. Soit (E, \mathcal{E}) un espace mesurable.

- a) On dit qu'une mesure μ sur (E, \mathcal{E}) est absolument continue par rapport à une mesure ν sur (E, \mathcal{E}) si pour tout $A \in \mathcal{E}$ tel que $\nu(A) = 0$, on a $\mu(A) = 0$. On écrit alors $\mu \ll \nu$.
- b) On dit qu'une mesure finie μ sur (E, \mathcal{E}) est équivalente à la mesure ν sur (E, \mathcal{E}) si $\mu \ll \nu$ et $\nu \ll \mu$. On note $\mu \equiv \nu$.

Le théorème suivant fait le lien entre l'absolue continuité et l'existence d'une densité.

3.2.7 Théorème. (Radon-Nikodym)

Soient μ et ν deux mesures sur (E, \mathcal{E}) . On suppose $\mu(E) < +\infty$ et ν, σ -finie, et $\mu \ll \nu$. Alors, il existe un unique $h \in L^1(\nu)$ tel que pour $A \in \mathcal{E}$, $\mu(A) = \int_A h d\nu$. On appelle h la dérivée de Radon-Nikodym de μ relativement à ν , et on note

$$d\mu = h d\nu \text{ ou encore } h = \frac{d\mu}{d\nu}.$$

Pour la démonstration, voir par exemple [23].

◇ **Exercice 3.4**

Soit X une variable aléatoire à valeurs dans \mathbb{R} .

a) Montrer que X est presque bornée au sens où $\forall \varepsilon > 0$, il existe $M > 0$ telle que $\mathbf{P}(|X| \geq M) < \varepsilon$.

b) Montrer qu'il existe une probabilité Q équivalente à \mathbf{P} avec $\mathbf{E}_Q(X) < +\infty$.

c) Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires réelles. Montrer qu'il existe une suite de réels strictement positifs $(\lambda_n)_{n \geq 0}$ telle que la série $\sum_{n \geq 0} \lambda_n X_n$ converge p.s.

Déduire l'existence d'une probabilité Q équivalente à \mathbf{P} sous laquelle toutes les X_n admettent un moment d'ordre 1.

Nous avons défini au chapitre précédent l'indépendance d'événements. Regardons maintenant comment se définit l'indépendance de variables aléatoires.

3.2.8 Définition. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles. On dit qu'elles sont indépendantes si, pour tout $n \geq 1$, pour tout choix d'entiers $1 \leq i_1 < i_2 < \dots < i_k \leq n$, et de fonctions mesurables bornées $f : \mathbb{R}^k \rightarrow \mathbb{R}$ et $g : \mathbb{R}^{n-k} \rightarrow \mathbb{R}$, on a

$$\mathbf{E}(f(X_{i_1}, \dots, X_{i_k})g(X_{j_1}, \dots, X_{j_{n-k}})) = \mathbf{E}(f(X_{i_1}, \dots, X_{i_k})) \cdot \mathbf{E}(g(X_{j_1}, \dots, X_{j_{n-k}}))$$

où j_1, \dots, j_{n-k} désignent les entiers de $1, \dots, n$ autres que les i_r .

Quand X et Y sont deux variables aléatoires on notera parfois $X \perp\!\!\!\perp Y$ pour stipuler qu'elles sont indépendantes.

On emploiera parfois l'abréviation classique (mais anglo-saxonne) **i.i.d.** pour signifier *indépendantes et de même loi*.

3.2.9 Théorème. La suite $(X_i)_{i \in I}$ est indépendante si et seulement si quel que soit $J \subset I$ fini, si $\mathcal{X}_J = (X_j)_{j \in J}$, alors

$$\mu_{\mathcal{X}_J} = \bigotimes_{j \in J} \mu_{X_j}.$$

PREUVE : Soit $(A_j)_{j \in J} \in \mathcal{B}(\mathbb{R})^J$. La loi $\mu_{\mathcal{X}_J}$ est entièrement déterminée par ses valeurs du type $\mu_{\mathcal{X}_J}(\prod_{j \in J} A_j)$ (π -système générateur des pavés). Il en est de même de l'autre loi en question sur \mathbb{R}^J , à savoir $\bigotimes_{j \in J} \mu_{X_j}$. Mais par indépendance,

$$\begin{aligned} \mu_{\mathcal{X}_J}(\prod_{j \in J} A_j) &= \mathbf{P}(\cap_{j \in J} X_j^{-1}(A_j)) \\ &= \prod_{j \in J} \mathbf{P}(X_j^{-1}(A_j)) \\ &= \prod_{j \in J} \mu_{X_j}(A_j) \\ &= (\bigotimes_{j \in J} \mu_{X_j})(\prod_{j \in J} A_j). \end{aligned}$$

Donc l'indépendance entraîne que la loi est la loi produit.

Réciproquement, si $\mu_{\mathcal{X}_J} = \bigotimes_{j \in J} \mu_{X_j}$, l'indépendance est évidente. □

3.2.10 Corollaire. Si $X \perp\!\!\!\perp Y$, alors quels que soient $g, h : \mathbb{R} \rightarrow \mathbb{R}$ boréliennes telles que $g(X)$ et $h(Y)$ soient dans $\mathcal{L}^2(\mathbf{P})$, on a

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

En particulier si $X, Y \in L^2(\mathbf{P})$,

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

Et alors $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

PREUVE : il suffit de le vérifier d'abord pour les fonctions simples g, h , puis pour leurs limites croissantes boréliennes, et enfin de considérer les parties positives et négatives.

On a donc

$$\text{Cov}(X, Y) = \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))) = \mathbf{E}(X - \mathbf{E}(X))\mathbf{E}(Y - \mathbf{E}(Y)) = 0^2 = 0.$$

Enfin $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$. □

Un résultat fondamental lié aux suites de variables indépendantes est dû à Kolmogorov :

3.2.11 Théorème. (loi du 0-1) Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires indépendantes. On pose $\mathcal{G} = \bigcap_{n \geq 0} \sigma(X_k, k \geq n)$ ("tribu de queue").

Alors, $\forall A \in \mathcal{G}, \mathbf{P}(A) = 0$ ou 1 .

PREUVE : posons $\mathcal{F}_\infty = \sigma(X_0, X_1, \dots, X_n, \dots)$. Noter qu'évidemment $\mathcal{G} \subset \mathcal{F}_\infty$. Soient $A \in \mathcal{G}$ et $\mathcal{C} = \{B \in \mathcal{F}_\infty, \mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)\}$. On vérifie immédiatement que \mathcal{C} est une classe monotone. De plus, \mathcal{C} contient l'algèbre de Boole $\mathcal{B} = \bigcup_{n \geq 0} \sigma(X_0, \dots, X_n)$. Par le théorème

de classe monotone (Théorème 2.1.2), il contient donc la tribu engendrée par \mathcal{B} , c'est à dire \mathcal{F}_∞ . En particulier, $A \in \mathcal{C}$ et donc A est indépendant de lui même, ce qui équivaut à $\mathbf{P}(A) = 0$ ou 1 . □

Exercice 3.5

Soient A_1, \dots, A_n des événements.

a) Montrer que si A_1, \dots, A_n sont indépendants, il en est de même de $A_1, \dots, A_{n-1}, A_n^c$.

b) Dédurre que $\forall J \subset \{1, \dots, n\}, (A_j^c, j \in J; A_i, i \in \{1, \dots, n\} \setminus J)$ forme un ensemble d'événements indépendants.

c) Montrer que A_1, \dots, A_n sont indépendants si et seulement si les variables aléatoires $\mathbb{I}_{A_1}, \dots, \mathbb{I}_{A_n}$ sont indépendantes.

◇ Exercice 3.6

On se place sur l'espace de probabilités $([0, 1], \mathcal{B}, \lambda)$ où λ est la mesure de Lebesgue sur $[0, 1]$. Soit $x \in [0, 1]$. On écrit x sous la forme de sa décomposition

dyadique $x = \sum_{k=1}^{\infty} \frac{x_k}{2^k}$ et on pose $X_k(x) = x_k$.

a) Montrer que les variables aléatoires $(X_k)_{k \geq 1}$ forment une suite de variables indépendantes et de même loi $\frac{1}{2}(\delta_0 + \delta_1)$.

b) Montrer qu'il existe une suite (Y_n) de variables aléatoires indépendantes sur $([0, 1], \mathcal{B}, \lambda)$ telle que Y_n suit une loi uniforme sur $[0, 1]$.

c) Soit $(\mu_n)_{n \geq 1}$ une suite de mesures de probabilités sur \mathbb{R} . Montrer qu'il existe une suite $(Z_n)_{n \geq 0}$ de variables indépendantes telle que Z_n a pour loi μ_n .

◇ Exercice 3.7

Un joueur va au casino avec une fortune $a \in \mathbb{N}$. A chaque partie, il peut gagner 1 euro avec une probabilité p et perdre 1 euro avec une probabilité $q = 1 - p$.

Son but est de jouer jusqu'à l'obtention de la fortune $c \geq a, c \in \mathbb{N}$ mais il doit s'arrêter s'il est ruiné. On note $s_c(a)$ sa probabilité de succès (atteindre c avant la ruine).

a) Calculer $s_c(0)$ et $s_c(c)$

b) Montrer, pour $a > 0$, en raisonnant sur ce qui s'est passé au premier coup, la relation

$$s_c(a) = ps_c(a+1) + qs_c(a-1)$$

c) Dédurre la valeur de $s_c(a)$

d) Application numérique :

Calculer la valeur précédente avec $a = 900, c = 1000$; $a = 100, c = 20000$ dans les cas $p = 0,5$ et $p = \frac{18}{38}$.

◇ Exercice 3.8

Une princesse a $n \geq 3$ prétendants numérotés par ordre de mérite décroissant $1, 2, \dots, n$. Elle doit en choisir un. Le problème est qu'ils défilent un par un au hasard devant elle et qu'elle ne peut revenir sur son choix si elle en a laissé partir un. Elle doit donc adopter une stratégie pour avoir le plus de chance de choisir le meilleur. . .

Soit Ω l'ensemble des permutations de $\{1, 2, \dots, n\}$. Ω est muni de la probabilité uniforme.

$\sigma \in \Omega$ représente un tirage du hasard (les prétendants défilent dans l'ordre $\sigma(1), \dots, \sigma(n)$).

Pour $1 \leq k \leq n$, on introduit la variable Y_k qui est le rang de $\sigma(k)$ dans l'ensemble $\{\sigma(1), \dots, \sigma(k)\}$ rangé en ordre décroissant : $Y_k = 1$ signifie que $\sigma(k)$ est le plus grand parmi l'ensemble $\{\sigma(1), \dots, \sigma(k)\}$, $Y_k = 2$ signifie qu'il y a exactement un élément de $\{\sigma(1), \dots, \sigma(k)\}$ plus grand que $\sigma(k)$ etc. . .

Par convention, on pose $Y_{n+1} = 0$.

a) Montrer que

$$F : \sigma \rightarrow (Y_1(\sigma), \dots, Y_n(\sigma))$$

définit une bijection de Ω sur $\Pi = \{1\} \times \dots \times \{1, 2, \dots, n\}$

b) En déduire que les variables Y_j sont indépendantes et que Y_k suit la loi uniforme sur $\{1, 2, \dots, k\}$.

c) Soit $\tau_r = \inf\{k > r, Y_k = 1\}$ ($= n + 1$ si cet ensemble est vide).

Calculer la probabilité pour qu'au temps τ_r , le prétendant qui se présente soit le meilleur (i.e. le numero 1).

Comment choisir r^* pour maximiser la probabilité précédente ? Trouver un équivalent de r^* quand n tend vers l'infini.

Exercice 3.9

Soit X une variable aléatoire à valeurs dans \mathbb{N} . La **fonction génératrice** de (la loi de) X est définie pour $0 \leq s \leq 1$ par

$$g_X(s) = \mathbf{E}(s^X) = \sum_{k=0}^{\infty} \mathbf{P}(X = k)s^k.$$

- Calculer les fonctions génératrices des lois de Bernoulli, binomiale, de Poisson.
- Montrer que la fonction génératrice caractérise la loi de X .
- Montrer que si X et Y sont des variables aléatoires indépendantes à valeurs dans \mathbb{N} , on a

$$g_{X+Y}(s) = g_X(s)g_Y(s), 0 \leq s \leq 1.$$

- Montrer que $g'_X(1) = \mathbf{E}(X)$ (cette quantité pouvant éventuellement être infinie).

Exercice 3.10

On reprend le modèle de l'exercice 3.7 mais le jeu change et le joueur est maintenant autorisé à s'endetter (il ne doit plus s'arrêter quand il est ruiné). On s'intéresse au temps d'attente du premier gain par le joueur (c'est à dire au premier instant où sa fortune s'est accrue d'une unité par rapport à sa fortune initiale).

- Posons

$$\phi_n = \mathbf{P}(\text{"au } n \text{ ième coup, pour la première fois, le joueur réalise un gain"}).$$

Par convention, $\phi_0 = 0$.

Calculer ϕ_1 .

- On pose $\Phi(s) = \sum_{n \geq 0} \phi_n s^n$ pour $0 \leq s \leq 1$. Montrer que pour $n > 1$,

$$\phi_n = q(\phi_1 \phi_{n-2} + \cdots + \phi_{n-2} \phi_1)$$

- Déduire que $\Phi(s) - ps = qs\Phi^2(s)$.
- Résoudre l'équation et en déduire Φ .
- Calculer $\sum_{n \geq 0} \phi_n$
- Soit N le numéro du coup où le joueur réalise un gain pour la première fois. Calculer $\mathbf{E}(N)$.

◇ **Exercice 3.11**

Soit $(\xi_{i,j})_{i \geq 0, j \geq 0}$ une famille de variables aléatoires à valeurs dans \mathbb{N} , indépendantes et de même loi de fonction génératrice (voir Exercice 3.9) donnée par $f(s) = \mathbf{E}(s^{\xi_{i,j}}), \forall i, j \geq 0$. On suppose que $\mathbf{P}(\xi_{i,j} = 0) + \mathbf{P}(\xi_{i,j} = 1) < 1$ et que $m = \mathbf{E}(\xi_{i,j}) < +\infty$.

On construit par récurrence une suite de variables aléatoires $(Z_n)_{n \geq 0}$ en posant $Z_0 = 1$ et $Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{n,i}$. Par convention, cette somme vaut 0 si $Z_n = 0$.

- Montrer que pour tout n , les variables $Z_n, \xi_{n,1}, \xi_{n,2}, \dots$ sont indépendantes.
- Soit f_n la fonction génératrice de Z_n . Montrer que $f_n(s) = f^{(n)}(s)$ où on pose $f^{(n)} = f \circ f \circ \dots \circ f$ (n fois).
Calculer $\mathbf{E}(Z_n)$ en fonction de m .
- Montrer que f est strictement croissante et strictement convexe sur $[0, 1]$. Etudier l'équation $f(x) = x$ en fonction des valeurs de m .
- On note q la plus petite solution de l'équation $f(x) = x$ dans $[0, 1]$. Montrer que si $0 \leq t < q, f_n(t)$ tend en croissant vers q .
- Soit l'événement $A = \{\omega \in \Omega, \lim_{n \rightarrow +\infty} Z_n(\omega) = 0\}$. Calculer $\mathbf{P}(A)$.

Exercice 3.12

Soit (X, Y) un couple de variables aléatoires réelles tel que pour $A \in \mathcal{B}(\mathbb{R}^2)$,

$$\mathbf{P}((X, Y) \in A) = \sum_{n \geq 1} \frac{1}{2^{n+1}} \mathbb{I}_A(n, n) + C \int_A (1 - x^2 y^2) \mathbb{I}_{[0,1]^2}(x, y) dm_2(x, y),$$

où C est une constante positive.

- Calculer C .
- Déterminer μ_X , puis μ_Y sans calcul supplémentaire.
- X et Y sont-elles indépendantes ?
- Montrer que X est intégrable et calculer $\mathbf{E}(X)$.

Exercice 3.13

Soient X et Y deux var indépendantes et équidistribuées sur $\{0, \dots, n\}$ et $\{0, \dots, m\}$ respectivement.

Déterminer la loi de $Z = X + Y$: on appelle cette loi la loi *trapèzoïdale*.

◇ **Exercice 3.14**

Soient X et Y deux var indépendantes de lois de Poisson de paramètres respectifs λ et β . Montrer que la loi de $X + Y$ est une loi de Poisson de paramètre $\lambda + \beta$.

Exercice 3.15

a) Si X est une var de loi géométrique, à valeurs dans \mathbb{N} , montrer que quel que soit $k, n \geq 1$,

$$\mathbf{P}(X = n + k | X > n) = \mathbf{P}(X = k).$$

Pourquoi appelle-t-on cette propriété la propriété d'*absence de mémoire*? Existe-t-il d'autres v.a. à valeurs dans \mathbb{N} ayant cette propriété?

b) Soit Y une variable aléatoire réelle positive sans mémoire, c'est-à-dire telle que pour tous $s, t > 0$ on ait

$$P(Y > t + s | Y > t) = P(Y > s).$$

Déterminer la loi de Y .

Exercice 3.16

On note $\{\mu_1 < \mu_2 < \dots\} = \mathcal{P}$ l'ensemble des nombres premiers, et $(N_i)_{i \geq 1}$ une suite de var indépendantes telles que

$$\mathbf{P}(N_i = k) = (1 - \gamma_i) \gamma_i^k, \quad k \geq 0,$$

où $\gamma_i = \mu_i^{-\beta}$, pour chaque i , et un certain $\beta > 1$. Montrer qu'alors

$$M = \prod_{i \geq 1} \mu_i^{N_i}$$

est un entier aléatoire, de loi $\mathbf{P}(M = m) = Cm^{-\beta}$ pour $m \geq 1$ (c'est la distribution dite de "*Dirichlet*"), où C est une constante telle que

$$C = \prod_{i \geq 1} \left(1 - \frac{1}{\mu_i^\beta}\right) = \left(\sum_{m \geq 1} \frac{1}{m^\beta}\right)^{-1}.$$

◇ **Exercice 3.17**

On rappelle la Formule de Poincaré : si (A_1, \dots, A_n) sont des événements,

$$\mathbf{P}\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n (-1)^{k-1} S_k$$

où $S_k = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mathbf{P}(A_{i_1} \cap \dots \cap A_{i_k})$.

A) Sur \mathbb{N}^* , on considère la probabilité μ_s ($s > 1$ réel fixé), définie par

$$\mu_s(\{n\}) = \frac{1}{\zeta(s)n^s}$$

où $\zeta(s) = \sum_{n \geq 1} \frac{1}{n^s}$ (fonction de Riemann).

Pour p premier, on définit les variables aléatoires ρ_p et α_p par $\rho_p(n) = 1$ si p divise n , 0 sinon, et $\alpha_p(n) =$ exposant de p dans la décomposition en facteurs premiers de n .

a) Trouver la loi des variables ρ_p et montrer qu'elles sont indépendantes.

b) Montrer que $\zeta(s) = \prod_{k=1}^{\infty} (1 - \frac{1}{\mu_k^s})^{-1}$ où $\{\mu_k, k \geq 1\}$ désigne la suite des nombres premiers dans \mathbb{N} .

B) Soit $N \geq 2$ un entier fixé. On pose $\Omega = \{1, \dots, N\}^2$ et on note μ_N la probabilité uniforme sur Ω .

On introduit les événements $B = \{\text{les entiers choisis sont premiers entre eux}\}$ et $A_p = \{\text{les entiers choisis sont divisibles par } p\}$ où p est premier.

On note $\mu_1, \mu_2, \dots, \mu_n$ les nombres premiers inférieurs à N et on fixe un entier $m \geq 1$.

a) Montrer que

$$\mu_N\left(\bigcup_{k=1}^m A_{\mu_k}\right) = \sum_{k=1}^m (-1)^{k-1} S_k$$

où $S_k = \frac{1}{N^2} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq m} [N/\mu_{i_1} \dots \mu_{i_k}]^2$.

b) Dédurre, toujours pour m fixé, que

$$\lim_N \mu_N\left(\bigcup_{k=1}^m A_{\mu_k}\right) = 1 - \prod_{k=1}^m \left(1 - \frac{1}{\mu_k^2}\right).$$

c) Conclure que $\lim_{N \rightarrow +\infty} \mu_N(B) = \frac{6}{\pi^2}$ (on rappelle que $\sum_{n \geq 1} \frac{1}{n^2} = \frac{\pi^2}{6}$).

◇ Exercice 3.18

1) Soient X_1, X_2, \dots, X_n des variables réelles indépendantes et de même loi μ . Soit $\sigma \in \mathcal{S}_n$, ensemble des bijections de $\{1, 2, \dots, n\}$ sur lui-même. Montrer que $\forall H \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbf{P}[(X_1, \dots, X_n) \in H] = \mathbf{P}[(X_{\sigma(1)}, \dots, X_{\sigma(n)}) \in H].$$

2) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles indépendantes et de même loi possédant une densité sur \mathbb{R} .

a) On définit $\tilde{\Omega}$ comme l'ensemble des $\omega \in \Omega$ tels que

$$X_k(\omega) \neq X_j(\omega), \forall (j, k) \in \mathbb{N}^2, k \neq j$$

Montrer que $\mathbf{P}(\tilde{\Omega}) = 1$.

b) Sur $\tilde{\Omega}$, on définit la variable aléatoire $T^n(\omega)$ à valeurs dans \mathcal{S}_n comme la permutation ordonnant $(X_1(\omega), \dots, X_n(\omega))$. Autrement dit, $T^n(\omega) = \sigma$ signifie

$$X_{\sigma(1)}(\omega) < X_{\sigma(2)}(\omega) < \dots < X_{\sigma(n)}(\omega)$$

On pose de plus $Y_n(\omega)$ égal au rang de $X_n(\omega)$ parmi $X_1(\omega), \dots, X_n(\omega)$ i.e. $Y_n(\omega) = k$ signifie que parmi $X_1(\omega), \dots, X_n(\omega)$, $k - 1$ valeurs exactement sont inférieures à $X_n(\omega)$.

Montrer que T^n suit une loi uniforme sur \mathcal{S}_n .

c) Montrer que Y_n suit une loi uniforme sur $\{1, 2, \dots, n\}$ et que les variables $Y_1, Y_2, \dots, Y_n, \dots$ sont indépendantes.

d) Soit A_n l'événement $(\max_{k < n} X_k < X_n)$ (on pose $A_1 = \tilde{\Omega}$).

Montrer que A_1, A_2, \dots sont indépendants et que $\mathbf{P}(A_n) = \frac{1}{n}$.

e) Soit $N_n(\omega) = \inf\{n' > n, \omega \in A_{n'}\}$

Montrer que

$$\mathbf{P}(N_n = n + k) = \frac{n}{(n + k - 1)(n + k)}$$

Calculer $\mathbf{E}(N_n)$

◇ Exercice 3.19

Le plan est strié de droites parallèles équidistantes de $2a$. Une aiguille de longueur $2l, l < a$ est jetée au hasard sur le plan au sens où la distance du milieu de l'aiguille à la droite la plus proche est une variable aléatoire X uniforme sur $[0, a]$ et où l'angle que fait l'aiguille avec cette droite est une variable aléatoire φ , indépendante de X , uniforme sur $[0, \pi]$.

Quelle est la probabilité que l'aiguille coupe l'une des parallèles ?

(Aiguille de Buffon).

◇ **Exercice 3.20**

Soit (X, Y) un couple de variables aléatoires réelles. On note m_2 la mesure de Lebesgue sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$, et m la mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On suppose que $\mu_{(X,Y)} \ll m_2$, et on note

$$\frac{d\mu_{(X,Y)}}{dm_2}(x, y) = f(x, y)$$

sa densité (voir Théorème 3.2.7).

- 1) Montrer que $\mu_X \ll m$ et que $\mu_Y \ll m$.
- 2) Exprimer $\frac{d\mu_X}{dm}(x)$ sous la forme d'une intégrale faisant intervenir f . Faire de même en ce qui concerne $\frac{d\mu_Y}{dm}(y)$.
- 3) Soit $A \in \mathcal{B}(\mathbb{R}^2)$. On note $A_y = \{x \in \mathbb{R} : (x, y) \in A\}$. On rappelle que m presque sûrement en y , $A_y \in \mathcal{B}(\mathbb{R})$, et que $y \mapsto m(A_y)$ est mesurable, et que $m_2(A) = \int m(A_y) dy$.

Soient μ et ν deux mesures de probabilités sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, absolument continues relativement à la mesure de Lebesgue m . Montrer qu'alors $\mu \otimes \nu \ll m_2$, et déterminer $\frac{d(\mu \otimes \nu)}{dm_2}$ en fonction de $\frac{d\mu}{dm}$ et de $\frac{d\nu}{dm}$.

- 4) En déduire une CNS pour que $X \perp\!\!\!\perp Y$, concernant la forme algébrique de f .
- 5) On suppose que la loi d'un couple de v.a.r. (X, Y) , $\mu_{(X,Y)}$, a une densité $f(x, y) = \frac{1}{\pi(1+x^2)} \frac{1}{\pi(1+y^2)}$ relativement à m_2 . Les variables X et Y sont-elles indépendantes ?

◇ **Exercice 3.21**

Soit $(A_n)_{n \geq 1}$ une suite d'évènements mesurables d'un espace probabilisé (Ω, \mathcal{B}, P) , telle que $P(A_1) > 0$. On note

$$N_n = \mathbb{I}_{A_1} + \dots + \mathbb{I}_{A_n}, \text{ et } \theta_n = \frac{\sum_{j,k \leq n} P(A_j \cap A_k)}{(\sum_{k \leq n} P(A_k))^2}, \quad n \geq 1.$$

- 1) Montrer que $\limsup_n A_n = \{\lim_n \uparrow N_n = +\infty\}$.
- 2) On note $\mu_k = P(A_k)$ et $m_n = \mu_1 + \dots + \mu_n$. Calculer $\text{Var}(N_n)$ en fonction de θ_n et de m_n .
- 3) Soit $x \in \mathbb{R}$ tel que $m_n > x$. Montrer que $P(N_n \leq x) \leq P(|N_n - m_n| \geq m_n - x)$, et en déduire que $P(N_n \leq x) \leq \frac{(\theta_n - 1)m_n^2}{(m_n - x)^2}$.
- 4) On suppose désormais que $\sum_n \mu_n = +\infty$, et que $\liminf_n \theta_n \leq 1$. Montrer qu'alors quel que soit $x \in \mathbb{R}$, $\liminf_n P(N_n \leq x) = 0$.
- 5) Montrer que $P(\sup_k N_k \leq x) \leq P(N_n \leq x)$. En déduire que $P(\limsup_n A_n) = 1$.
- 6) Déduire de ce qui précède que le second Lemme de Borel-Cantelli est encore valable si l'on suppose les A_i deux à deux indépendants.

Exercice 3.22

Soient X et Y deux var à valeurs entières dont la loi conjointe (i.e. la loi du couple ou vecteur (X, Y)) est donnée par

$$\mathbf{P}((X, Y) = (x, y)) = \frac{C}{(x+y-1)(x+y)(x+y+1)}, \quad x, y \geq 1.$$

Déterminer les lois de $U = X + Y$ et $V = X - Y$.

3.3 Notions sur les processus à temps discret

Nous allons jusqu'à la fin de ce chapitre dire quelques mots concernant les collections de variables aléatoires indéxées par un paramètre temporel discret, autrement dit ce qu'on appelle des processus stochastiques.

3.3.1 Définition. On appelle **filtration** sur l'espace mesurable (Ω, \mathcal{F}) une suite croissante $(\mathcal{F}_n)_{n \in \mathbb{N}}$ de sous-tribus de \mathcal{F} .

L'idée de cette notion est de modéliser un accroissement d'information au cours du temps. On note $\mathcal{F}_\infty = \sigma(\bigcup_{n \geq 0} \mathcal{F}_n)$. Naturellement, il n'y a aucune raison pour que $\bigcup_{n \geq 0} \mathcal{F}_n = \mathcal{F}_\infty$ (en général $\bigcup_{n \geq 0} \mathcal{F}_n$ n'est même pas une tribu), ni que $\mathcal{F} = \mathcal{F}_\infty$. Quand l'espace (Ω, \mathcal{F}) est muni d'une filtration on dit qu'il s'agit d'un espace filtré.

3.3.2 Exemple. Soit $\Omega = [0, 1[$. On considère \mathcal{F}_n l'ensemble des réunions disjointes d'intervalles du type $[k/2^n, k+1/2^n[$, $0 \leq k \leq 2^n - 1$.

\mathcal{F}_n est une filtration sur $(\Omega, \mathcal{B}([0, 1[))$ et $\bigcup_{n \geq 0} \mathcal{F}_n \neq \mathcal{B}([0, 1[) = \mathcal{F}_\infty$.

Etroitement liée à la notion de filtration est celle de temps d'arrêt dont le nom prend son sens grâce à la Proposition 3.3.8 *infra*.

3.3.3 Définition. Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbf{P})$ un espace mesurable filtré. On appelle **temps d'arrêt** sur $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbf{P})$ une variable aléatoire τ à valeurs dans $\mathbb{N} \cup \{\infty\}$ telle que, pour tout $n \in \mathbb{N}$ l'événement $(\tau \leq n)$ appartient à \mathcal{F}_n .

Les principales propriétés élémentaires des temps d'arrêt sont les suivantes.

3.3.4 Proposition. a) Soient τ et σ deux temps d'arrêt. $\tau \wedge \sigma, \tau \vee \sigma$ sont des temps d'arrêt
 b) Si $\tau = \lim_{n \rightarrow +\infty} \uparrow \tau_n$ où les τ_n forment une suite croissante de temps d'arrêt, τ est un temps d'arrêt.

c) τ est un temps d'arrêt si et seulement si $\forall n, (\tau = n) \in \mathcal{F}_n$.

PREUVE :

a) On a $(\tau \wedge \sigma > n) = (\tau > n) \cap (\sigma > n)$ et $(\tau \vee \sigma \leq n) = (\tau \leq n) \cap (\sigma \leq n)$.

b) On a $(\tau \leq n) = \bigcap_{k \geq 0} (\tau_k \leq n)$.

c) Tout d'abord, $(\tau = n) = (\tau \leq n) \cap (\tau \leq n-1)^c$ et donc si τ est un temps d'arrêt, ceci se trouve dans \mathcal{F}_n . Réciproquement, $(\tau \leq n) = \bigcup_{k=0}^n (\tau = k)$ et par hypothèse $(\tau = k) \in \mathcal{F}_k \subset \mathcal{F}_n$. \square

Il convient de remarquer que la dernière propriété énoncée (c) est spécifique au cas discret et ne sera pas valable en temps continu. Il est aussi important de remarquer qu'un temps déterministe fixé considéré comme variable aléatoire constante est un temps d'arrêt.

3.3.5 Proposition et Définition. Soit τ un temps d'arrêt de la filtration (\mathcal{F}_n) . On considère

$$\mathcal{F}_\tau = \{A \in \mathcal{F}, A \cap (\tau \leq n) \in \mathcal{F}_n, \forall n \in \mathbb{N}\}.$$

\mathcal{F}_τ est une tribu dite **tribu du passé de τ** .

La démonstration, immédiate, est laissée en exercice. Deux propriétés importantes des temps d'arrêt sont proposées en exercice.

◇ **Exercice 3.23**

- a) Montrer que si $\tau \leq \sigma$ sont deux temps d'arrêt, $\mathcal{F}_\tau \subset \mathcal{F}_\sigma$.
 b) Montrer que si $\tau = t, t \in \mathbb{N}$, on a $\mathcal{F}_\tau = \mathcal{F}_t$.

On va introduire maintenant les bases de la théorie des processus stochastiques à temps discret. Le cas discret est spécialement simple structurellement, en particulier en ce qui concerne les questions liées à la mesurabilité. L'extension des notions au cas continu représente une étape ultérieure éventuellement délicate (voir le Chapitre 13).

3.3.6 Définition. a) On appelle **processus stochastique indexé par \mathbb{N} à valeurs dans (E, \mathcal{E})** toute suite de variables aléatoires $(X_n)_{n \geq 0}$ à valeurs dans (E, \mathcal{E}) .
 b) Si l'espace de probabilité est filtré par une filtration $(\mathcal{F}_n)_{n \geq 0}$ on dit que le processus $(X_n)_{n \geq 0}$ est **adapté** si pour tout $n \in \mathbb{N}$, X_n est une variable aléatoire \mathcal{F}_n mesurable.
 c) Par ailleurs, à tout processus $(X_n)_{n \geq 0}$ est associée une filtration (dite **naturelle**) définie par $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$. Naturellement, un processus est toujours adapté à sa filtration naturelle.

Un processus stochastique $X = (X_n)_{n \geq 0}$ peut être conçu comme une variable aléatoire à valeurs dans $E^{\mathbb{N}}$. De ce fait, sa loi μ_X est une probabilité sur l'espace mesurable $(E^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ où la tribu $\mathcal{E}^{\otimes \mathbb{N}}$ est engendrée par les ensembles du type $\pi_n^{-1}(B)$ où $B \in E^{\otimes n+1}$ et

$$\pi_n((\omega_i)_{i \geq 0}) = (\omega_0, \omega_1, \dots, \omega_n).$$

L'espace mesurable $(E^{\otimes \mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ est dit **espace canonique** des processus à valeurs dans E . Le processus canonique \tilde{X} est défini sur $E^{\otimes \mathbb{N}}$ par $\tilde{X}_n((\omega_k)_{k \geq 0}) = \omega_n$.

On voit immédiatement que la donnée d'un processus stochastique entraîne l'existence d'une famille de probabilités μ_n sur $(E^{n+1}, \mathcal{E}^{\otimes n+1})$ définies par

$$\mu_n(A_0 \times \dots \times A_n) = \mathbf{P}(X_0 \in A_0, \dots, X_n \in A_n).$$

Les probabilités μ_n s'appellent les répartitions finies du processus. De manière manifeste,

$$\mu_n(dx_0, \dots, dx_n) = \int_E \mu_{n+1}(dx_0, \dots, dx_n, dx_{n+1}). \quad (3.3.1)$$

La question qui se pose naturellement est la réciproque : si une famille de probabilités $(\mu_n)_{n \geq 0}$ sur $(E^{n+1}, \mathcal{E}^{\otimes n+1})_{n \geq 0}$ satisfait la condition de compatibilité précédente (3.3.1), existe-t-il un processus à valeurs dans E admettant les μ_n comme répartitions finies ? La réponse est donnée par le théorème de Kolmogorov suivant dont une démonstration dans deux cas particuliers importants sera faite dans un chapitre ultérieur (Corollaire 9.1.3), à charge pour le lecteur de vérifier l'absence de cercle vicieux. On peut aussi voir une démonstration directe par exemple dans [6], comme conséquence du théorème d'extension de Caratheodory (Théorème 2.2.3).

3.3.7 Théorème. (Kolmogorov) Soit E un espace métrique séparable et soit $(\mu_n)_{n \geq 0}$ une famille de probabilités sur les espaces produit $(E^{n+1}, \mathcal{E}^{\otimes n+1})_{n \geq 0}$ qui satisfait la condition de compatibilité (3.3.1).

Alors il existe une unique probabilité \mathbf{P} sur l'espace canonique telle que sous \mathbf{P} le processus canonique \tilde{X} admette les lois μ_n comme répartitions finies.

Pour terminer cette section, nous allons présenter l'exemple fondamental de temps d'arrêt. Considérons un processus stochastique $(X_n)_{n \geq 0}$ à valeurs dans (E, \mathcal{E}) . Soit $A \in \mathcal{E}$. On appelle temps d'entrée dans A la variable aléatoire à valeurs dans $\mathbb{N} \cup \{+\infty\}$ définie par $\tau_A = \inf\{n \geq 0, X_n \in A\}$ et temps de retour dans A la variable aléatoire à valeurs dans $\mathbb{N} \cup \{+\infty\}$ définie par $\sigma_A = \inf\{n \geq 1, X_n \in A\}$. Dans ces deux définitions, la valeur prise est infinie si l'ensemble considéré est vide. On a alors

3.3.8 Proposition. Pour tout $A \in \mathcal{E}$, τ_A et σ_A sont des temps d'arrêt de la filtration naturelle (\mathcal{F}_n) du processus X .

PREUVE : Soit $n \geq 0$. On a $(\tau_A = n) = (X_0 \notin A, \dots, X_{n-1} \notin A, X_n \in A)$ et ce dernier ensemble est manifestement dans \mathcal{F}_n . D'après la Proposition 3.3.4 c), τ_A est un temps d'arrêt. Pour $n \geq 1$, $(\sigma_A = n) = (X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A)$ qui est aussi dans \mathcal{F}_n . \square

◇ Exercice 3.24

Soit E un ensemble dénombrable. Ω est l'espace des applications de $E^{\mathbb{N}}$ vers E . On note X_n la n -ième projection canonique définie par $X_n(\omega) = \omega_n$ si $\omega = (\omega_k)_{k \geq 0}$. $(\mathcal{F}_n)_{n \geq 0}$ désigne la filtration naturelle du processus $(X_n)_{n \geq 0}$.

Soit $\theta: \Omega \rightarrow \Omega$ l'opérateur de translation, défini par $\theta((\omega_n)_{n \geq 0}) = (\omega_{n+1})_{n \geq 0}$, et $(\theta_n)_{n \geq 0}$ la suite de ses itérées (i.e. $\theta_0 =$ identité et $\theta_n = \theta \circ \dots \circ \theta$, n fois).

a) Montrer que $X_n \circ \theta_m = X_{n+m}$ et $\theta_n^{-1}(\mathcal{F}_m) = \sigma(X_n, \dots, X_{n+m})$.

b) Soient τ et ν deux temps d'arrêt de la filtration $(\mathcal{F}_n)_{n \geq 0}$ (éventuellement $\tau = \nu$).

b1) Montrer que, pour tout entier positif k , $k + \tau \circ \theta_k$ est un temps d'arrêt.

b2) Montrer que $\rho = \nu + \tau \circ \theta_\nu$ est un temps d'arrêt. Montrer que, si on suppose de plus que ν et τ sont finis, $X_\tau \circ \theta_\nu = X_\rho$.

b3) Pour $A \subset E$, notons

$$\tau_A(\omega) = \inf\{n \geq 0; X_n(\omega) \in A\} \quad \sigma_A(\omega) = \inf\{n \geq 1; X_n(\omega) \in A\}$$

les temps d'atteinte et de retour en A . Montrer que, si $A \subset B$, $\tau_A = \tau_B + \tau_A \circ \theta_{\tau_B}$.

Chapitre 4

Différents types de convergences

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires à valeurs dans \mathbb{R} , et soit X une variable aléatoire à valeurs dans \mathbb{R} .

4.1 Convergences

Introduisons d'abord les définitions des différentes convergences.

4.1.1 Définition. a) On dit que la suite de variables (X_n) converge presque sûrement (p.s.) vers X si

$$\mathbf{P}(\{\omega, X_n(\omega) \rightarrow X(\omega)\}) = 1.$$

b) On dit que la suite de variables (X_n) converge dans L^1 vers X si $\mathbf{E}(|X - X_n|) \rightarrow 0$.

c) On dit que la suite de variables (X_n) converge en probabilités vers X si $\forall \varepsilon > 0$,

$$\mathbf{P}(|X_n - X| > \varepsilon) \rightarrow 0.$$

d) On dit que la suite de variables (X_n) converge en loi vers X si pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue et bornée, $\mathbf{E}(f(X_n)) \rightarrow \mathbf{E}(f(X))$.

Les convergences a,b,c) ci-dessus sont dites convergences spatiales car elles font intervenir de manière essentielle l'espace de probabilités Ω . Par contre, on remarque que (comme son nom l'indique), la convergence en loi ne fait intervenir que les lois des variables X_n et X (cf. Proposition 3.2.2). De ce fait, on peut légitimement considérer que la convergence d) est plutôt à voir comme une convergence de mesures et on écrira classiquement $\mu_{X_n} \Rightarrow \mu_X$ pour signifier la convergence en loi en question.

4.2 Liens entre les différentes convergences

Le lien entre ces différentes convergences est donné par la proposition suivante.

4.2.1 Proposition. a) Les convergences dans L^1 et p.s. entraînent chacune la convergence en probabilités.

b) La convergence en probabilités entraîne la convergence en loi.

PREUVE : supposons d'abord que $(X_n)_{n \geq 0}$ converge dans L^1 vers X_∞ . Soit $\varepsilon > 0$. On a alors, par l'inégalité de Markov, $\mathbf{P}(|X_n - X_\infty| > \varepsilon) < \frac{1}{\varepsilon} \mathbf{E}(|X_n - X_\infty|)$ et ce dernier terme tend vers 0, donc la suite converge en probabilités vers X_∞ .

Si $(X_n)_{n \geq 0}$ converge p.s. vers X_∞ , $(\mathbb{I}_{|X_n - X_\infty| > \varepsilon})_{n \geq 0}$ converge p.s. vers 0, et est dominée par 1. Par application du théorème de convergence dominée, $\mathbf{P}(|X_n - X_\infty| > \varepsilon) \rightarrow 0$.

Enfin, supposons que $(X_n)_{n \geq 0}$ converge en probabilités vers X_∞ . Soit f une fonction continue à support compact sur \mathbb{R} . f est alors uniformément continue sur \mathbb{R} : pour $\varepsilon > 0$ donné, il existe $\alpha > 0$ tel que $|x - y| < \alpha \Rightarrow |f(x) - f(y)| < \varepsilon$. On a alors

$$\begin{aligned} & | \mathbf{E}(f(X_n) - f(X_\infty)) | \\ & \leq \mathbf{E}(|f(X_n) - f(X_\infty)|) \\ & = \mathbf{E}(|f(X_n) - f(X_\infty)| \mathbb{I}_{|X_n - X_\infty| < \alpha}) + \mathbf{E}(|f(X_n) - f(X_\infty)| \mathbb{I}_{|X_n - X_\infty| \geq \alpha}) \\ & \leq \varepsilon + 2\|f\|_\infty \mathbf{P}(|X_n - X_\infty| \geq \alpha) \leq 2\varepsilon \end{aligned}$$

pour n assez grand. On a donc $\mathbf{E}(f(X_n)) \rightarrow \mathbf{E}(f(X_\infty))$. On conclut alors en appliquant le Lemme 6.3.3 *infra*. \square

4.2.2 Proposition. Si (X_n) converge en probabilité vers X , il existe une sous-suite (X_{n_k}) qui converge p.s. vers X .

PREUVE : Soit $\varepsilon > 0$. Pour $k \geq 1$, on peut trouver n_k tel que

$$\mathbf{P}(|X_{n_k} - X| > \frac{1}{k}) \leq \frac{1}{k^2}.$$

On a alors

$$\sum_{k=1}^{\infty} \mathbf{P}(|X_{n_k} - X| > \frac{1}{k}) < +\infty$$

et donc, par le lemme de Borel-Cantelli (2.3.1), p.s. à partir d'un rang K , $|X_{n_k} - X| < \frac{1}{k}$. De ce fait, $X_{n_k} \rightarrow X$ p.s. \square

4.2.3 Remarque. Les implications de la Proposition 4.2.1 sont en général les seules valables.

Construisons en effet des contre-exemples aux autres directions.

a) Soit $\Omega = \{0, 1\}$ muni de la probabilité \mathbf{P} telle que $\mathbf{P}(\{0\}) = \mathbf{P}(\{1\}) = \frac{1}{2}$. On définit $X_{2n}(0) = 0, X_{2n}(1) = 1$ et $X_{2n+1}(0) = 1, X_{2n+1}(1) = 0$. Il est alors immédiat que tous les X_n ont même loi, et donc que la suite (X_n) converge en loi. Si on avait une v.a.r. X telle que $X_n \rightarrow_{\mathbf{P}} X$, il existerait $N, \forall n \geq N, \mathbf{P}(|X_n - X| > \frac{1}{4}) < \frac{1}{4}$. Si $N < 2n < 2n+1$, on a

$$\mathbf{P}(|X_{2n} - X_{2n+1}| > \frac{1}{2}) \leq \mathbf{P}(|X - X_{2n+1}|) + \mathbf{P}(|X - X_{2n}|) < \frac{1}{2}.$$

Or $|X_{2n} - X_{2n+1}| = 1$ d'où une contradiction, et la suite ne converge pas en probabilité.

b) Soit $\Omega = [0, 1]$ muni de la mesure de Lebesgue. On considère la variable aléatoire X_n définie par

$$X_n = \begin{cases} 0, & \text{si } 0 \leq x \leq 1 - n^{-1} \\ n^2, & \text{si } 1 - n^{-1} \leq x \leq 1. \end{cases}$$

Clairement, $\mathbf{P}(|X_n| > \varepsilon) = \frac{1}{n}$ et donc $(X_n)_{n \in \mathbb{N}}$ converge en probabilités vers 0. Par contre,

$\mathbf{E}(X_n) = n^2 \times \frac{1}{n} = n$ tend vers $+\infty$ et la suite ne saurait converger dans L^1 .

c) Soit de nouveau $\Omega = [0, 1]$ muni de la mesure de Lebesgue. On définit la suite de variables aléatoires comme suit. X_0 est constante égale à 1. X_1 vaut 1 sur $[\frac{1}{2}, 1]$ et 0 ailleurs. X_2 vaut 1 sur $[0, \frac{1}{2}]$ et 0 ailleurs. X_3 vaut 1 sur $[\frac{2}{3}, 1]$ et 0 ailleurs, X_4 vaut 1 sur $[\frac{1}{3}, \frac{2}{3}]$ et 0 ailleurs, X_5 vaut 1 sur $[0, \frac{1}{3}]$ et 0 ailleurs. Et ainsi de suite, l'intervalle où X_n vaut 1 "se promenant" de droite à gauche en parcourant tout $[0, 1]$ et en se réduisant en longueur. Il est alors clair que la suite converge en probabilité vers 0. Par ailleurs, pour tout $x \in [0, 1]$, il existera toujours des n aussi grands qu'on veut pour lesquels $X_n(x) = 1$ et donc la suite ne converge pas p.s.

◇ Exercice 4.1

Soit (X_n) une suite de variables aléatoires à valeurs 0 ou 1.

On pose $p_n = \mathbf{P}(X_n = 1)$.

- Trouver une condition nécessaire et suffisante pour que la suite (X_n) converge vers 0 dans L^1 .
- Trouver une condition nécessaire et suffisante pour que la suite (X_n) converge vers 0 en probabilités.
- Trouver une condition nécessaire et suffisante pour que la suite (X_n) converge vers 0 p.s.
- Trouver une condition nécessaire et suffisante pour que la suite (X_n) converge vers 0 en loi.

◇ Exercice 4.2

On dit qu'une famille $(X_i)_{i \in I}$ de v.a. réelles sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ est **équi-intégrable** (on dit aussi **uniformément intégrable**) si

$$\lim_{a \rightarrow +\infty} \sup_{i \in I} \int_{\{|X_i| > a\}} |X_i| d\mathbf{P} = 0.$$

On se propose de montrer différentes propriétés d'une telle famille.

- Montrer qu'une famille finie de v.a. intégrables est équi-intégrable.
- Montrer que la famille $(X_i)_{i \in I}$ est équi-intégrable si et seulement si l'on a $\sup_{i \in I} \mathbf{E}|X_i| < +\infty$ et la propriété :
(P) pour tout $\varepsilon > 0$ il existe $\alpha > 0$ tel que $\mathbf{P}(A) < \alpha$ implique

$$\sup_{i \in I} \int_A |X_i| d\mathbf{P} < \varepsilon.$$

- Soit $(X_n)_{n \geq 0}$ une suite de v.a. réelles convergeant dans L^1 vers une v.a. X . Montrer que la suite est équi-intégrable.
- Soit $(X_n)_{n \geq 0}$ une suite équi-intégrable de v.a. réelles convergeant p.s. vers une v.a. X . On note, pour $a > 0$,

$$f_a(x) = -a\mathbf{1}_{\{x < -a\}} + x\mathbf{1}_{\{-a \leq x \leq a\}} + a\mathbf{1}_{\{x > a\}}.$$

- Montrer que $\sup_n \mathbf{E}|X_n| < +\infty$. En déduire que $\mathbf{E}|X| < +\infty$.
- Montrer que, pour tout $a > 0$, $\|f_a(X_n) - f_a(X)\|_1 \rightarrow_n 0$.
- Montrer que

$$\begin{aligned} \sup_n \|X_n - f_a(X_n)\|_1 &\xrightarrow{a \rightarrow +\infty} 0 \\ \|X - f_a(X)\|_1 &\xrightarrow{a \rightarrow +\infty} 0 \end{aligned}$$

- Montrer finalement que X_n converge vers X dans L^1 .

Chapitre 5

Convergences spatiales

5.1 Loi des grands nombres

Commençons par un petit Lemme, qui a son propre intérêt au demeurant :

5.1.1 Lemme. *Si $p \geq 1$ et $X \in L^p(\mathbf{P})$, $X : \Omega \rightarrow \mathbb{R}^+$, alors*

$$\mathbf{E}(X^p) = \int_0^\infty px^{p-1}(1 - F_X(t))dt.$$

PREUVE : On a $1 - F_X(t) = \mathbf{P}(X > t) = \int_0^\infty \mathbb{I}_{]t, +\infty[}(u) d\mathbf{P}_X(u)$, et donc par Fubini-Tonelli, et le théorème de transfert,

$$\begin{aligned} & \int_0^\infty px^{p-1}(1 - F_X(t))dt \\ &= \int_{(\mathbb{R}^+)^2} pt^{p-1} \mathbb{I}_{]t, +\infty[}(u) d\mathbf{P}_X(u) dt \\ &= \int_0^\infty \left(\int_0^u pt^{p-1} dt \right) d\mathbf{P}_X(u) \\ &= \int_0^\infty u^p d\mathbf{P}_X(u) \\ &= \mathbf{E}(X^p). \end{aligned}$$

□

La loi forte “classique” suppose l’indépendance de la suite. Le résultat qui suit est dû au mathématicien indien Etemadi, et date de 1983. Il généralise le cas classique en supposant simplement l’indépendance deux à deux.

5.1.2 Théorème. (Loi forte d’Etemadi) *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendamment distribuées, intégrables, et deux à deux indépendantes. Alors*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \leq n} X_k = \mathbf{E}(X_1) \mathbf{P} - p.s.$$

PREUVE : D’abord remarquons que sous nos hypothèses, les mêmes hypothèses ont lieu pour les deux suites $(X_k^+)_{k \geq 1}$ et $(X_k^-)_{k \geq 1}$. Et donc il est clair que si le résultat est vrai sous l’hypothèse supplémentaire de positivité des X_k , il en découlera dans le cas général par décomposition en parties positives et négatives. Ainsi nous pouvons sans perte de généralité nous restreindre au cas $\mathbf{P}(X_1 \geq 0) = 1$. Nous noterons $S_n = \sum_{k \leq n} X_k$.

Posons $Y_k = X_k \mathbb{I}_{X_k \leq k}$, et $S_n^* = \sum_{k \leq n} Y_k$. Remarquons que la suite (Y_k) est une suite deux à deux indépendante, et que $Y_k \in L^\infty(\mathbf{P})$.

Soit $\alpha > 1$, fixé, et posons $u_n = [\alpha^n]$. Nous allons d’abord démontrer que, si $\varepsilon > 0$ est donné,

$$\Sigma_\varepsilon = \sum_n \mathbf{P} \left[\left| \frac{S_{u_n}^* - \mathbf{E}(S_{u_n}^*)}{u_n} \right| > \varepsilon \right] < \infty.$$

Par indépendance et distribution identique des X_k , on a

$$\text{Var}(S_n^*) = \sum_{k \leq n} \text{Var}(Y_k) \leq \sum_{k \leq n} \mathbf{E}(Y_k^2) = \sum_{k \leq n} \mathbf{E}(X_1^2 \mathbb{I}_{X_1 \leq k}) \leq n \mathbf{E}(X_1^2 \mathbb{I}_{X_1 \leq n}).$$

Par application de l'inégalité de Bienaymé-Chebyshev, il s'ensuit que

$$\Sigma_\varepsilon \leq \frac{1}{\varepsilon^2} \sum_n \frac{1}{u_n} \mathbf{E}(X_1^2 \mathbb{I}_{X_1 \leq u_n}).$$

Posons $K = 2\alpha / (\alpha - 1)$, et soit $x > 0$. Soit $N = N(x) = \min\{n : u_n \geq x\}$: alors $\alpha^N \geq x$, et comme pour $y \geq 1$, on a toujours $y \leq 2[y]$, il s'ensuit que

$$\sum_{u_n \geq x} \frac{1}{u_n} \leq 2 \sum_{n \geq N} \alpha^{-n} = K \alpha^{-N} \leq \frac{K}{x}.$$

Et donc en substituant X_1 à x , il vient, compte tenu de la majoration précédente de Σ_ε , et en appliquant Fubini-Tonelli,

$$\Sigma_\varepsilon \leq \frac{1}{\varepsilon^2} \mathbf{E}(X_1^2 \mathbb{I}_{X_1 > 0} \left(\sum_{u_n \geq X_1} \frac{1}{u_n} \right)) \leq \frac{1}{\varepsilon^2} \mathbf{E}(X_1^2 \mathbb{I}_{X_1 > 0} \frac{K}{X_1}) = \frac{K}{\varepsilon^2} \mathbf{E}(X_1) < \infty.$$

La convergence de Σ_ε entraîne, par le premier lemme de Borel-Cantelli, en prenant une suite $\varepsilon_p \downarrow 0$, que

$$\mathbf{P}(\limsup_n \left| \frac{S_{u_n}^* - \mathbf{E}(S_{u_n}^*)}{u_n} \right| > 0) = 0,$$

autrement dit que **P**-p.s., la suite $(S_{u_n}^* - \mathbf{E}(S_{u_n}^*)) / u_n$ tend vers 0.

D'autre part, $\mathbf{E}(Y_k) \rightarrow \mathbf{E}(X_1)$, par convergence monotone. Et donc puisque $u_n \rightarrow \infty$, il s'ensuit par Césaro, que $\mathbf{E}(S_{u_n}^*) / u_n \rightarrow \mathbf{E}(X_1)$, et donc nous concluons que

$$\frac{S_{u_n}^*}{u_n} \rightarrow \mathbf{E}(X_1) \quad \mathbf{P} - p.s.$$

Ensuite on a, en utilisant le lemme préliminaire,

$$\sum_k \mathbf{P}(X_k \neq Y_k) = \sum_k \mathbf{P}(X_k > k) = \sum_k \mathbf{P}(X_1 > k) \leq \int_0^\infty \mathbf{P}(X_1 > t) dt = \mathbf{E}(X_1) < \infty,$$

et donc à nouveau avec le premier lemme de Borel-Cantelli, il s'ensuit que

$$\mathbf{P}(\limsup_k \{X_k \neq Y_k\}) = 0.$$

Et donc **P**-p.s., il existe $n(\omega)$ tel que si $n \geq n(\omega)$, $(S_n - S_n^*)(\omega) = S_{n(\omega)}(\omega) - S_{n(\omega)}^*(\omega)$, et donc

$$\lim_n \frac{S_n - S_n^*}{n} = 0 \quad \mathbf{P} - p.s.$$

Il s'ensuit que

$$\lim_n \frac{S_{u_n}}{u_n} = \mathbf{E}(X_1) \quad \mathbf{P} - p.s. \quad (5.1.1)$$

Soit $n_0 = \min\{n : \frac{1}{u_n} < \alpha - 1\}$: alors la suite $(u_n)_{n \geq n_0}$ est strictement croissante. Soit $k \geq u_{n_0}$ et soit n l'unique entier tel que $u_n \leq k < u_{n+1}$. Par positivité des X_k , nous avons

$$\frac{u_n}{u_{n+1}} \frac{S_{u_n}}{u_n} \leq \frac{S_k}{k} \leq \frac{u_{n+1}}{u_n} \frac{S_{u_{n+1}}}{u_{n+1}},$$

dont il découle, puisque $u_{n+1} / u_n \rightarrow \alpha$, et compte tenu de (5.1.1) que

$$\text{quel que soit } \alpha > 1, \mathbf{P} \left(\frac{1}{\alpha} \mathbf{E}(X_1) \leq \liminf_k \frac{S_k}{k} \leq \limsup_k \frac{S_k}{k} \leq \alpha \mathbf{E}(X_1) \right) = 1.$$

Il suffit alors d'intersecter les ensembles correspondants de mesures pleines le long d'une suite $\alpha_n \downarrow 1$ pour conclure. \square

5.1.3 Corollaire. *Supposons $X_1^- \in L^1(\mathbf{P})$ et $\mathbf{E}(X_1^+) = +\infty$. Alors, sous les mêmes hypothèses que celles de la loi forte, on a*

$$\frac{S_n}{n} \rightarrow +\infty \quad \mathbf{P} - p.s.$$

PREUVE : Par la loi forte, si nous posons $S_{n,-} = \sum_{k \leq n} X_k^-$, alors $S_{n,-} / n \rightarrow \mathbf{E}(X_1^-)$ P-p.s.. Donc, il suffit de montrer que le corollaire est vrai si $X_1 \geq 0$ et $\mathbf{E}(X_1) = +\infty$. Soit alors $p \in \mathbb{N}$ et posons $X_{k,p} = \min(X_k, p)$. Alors par la loi forte, si $S_{n,p} = \sum_{k \leq n} X_{k,p}$, on a $\lim_n S_{n,p}/n = \mathbf{E}(X_{1,p})$, et puisque $S_{n,p} \leq S_n$, on a

$$\mathbf{E}(X_{1,p}) \leq \liminf_n \frac{S_n}{n}.$$

Comme $X_{1,p} \uparrow X_1$ si $p \rightarrow \infty$, il s'ensuit que $+\infty = \lim_p \uparrow \mathbf{E}(X_{1,p}) \leq \liminf_n \frac{S_n}{n}$. \square

5.2 Convergence des séries de variables indépendantes

Nous allons nous intéresser dans ce paragraphe à un cas particulièrement important d'étude de convergence de variables aléatoires, dont les résultats essentiels marquèrent l'entrée en scène fracassante de Kolmogorov en 1925.

Dans ce paragraphe, et sauf mention contraire, $(X_n)_{n \geq 1}$ désignera une suite de variables aléatoires indépendantes, **centrées et admettant un moment d'ordre 2**. On pose, pour tout n , $S_n = \sum_{k=1}^n X_k$.

5.2.1 Théorème. *On suppose que $\sum_{k \geq 1} \mathbf{E}(X_k^2) < \infty$. Alors $(S_n)_{n \geq 1}$ converge dans L^2 .*

PREUVE : il s'agit de montrer que la suite $(S_n)_{n \geq 0}$ est de Cauchy dans L^2 . Or, par l'indépendance et le centrage des X_k

$$\mathbf{E}(|S_{n+p} - S_n|^2) = \mathbf{E} \left(\left| \sum_{k=n+1}^{n+p} X_k \right|^2 \right) = \sum_{k=n+1}^{n+p} \mathbf{E}(X_k^2),$$

d'où l'on déduit le résultat cherché. \square

Pour regarder la convergence p.s. de ces mêmes séries, on va devoir employer une technique qui s'avèrera très importante dans la théorie des martingales (voir Chapitre 12).

5.2.2 Proposition. (Inégalité de Kolmogorov)

$$\mathbf{P} \left(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon \right) \leq \frac{\mathbf{E}(S_n^2)}{\varepsilon^2}.$$

PREUVE : On a

$$\left[\max_{1 \leq k \leq n} |S_k| \geq \varepsilon \right] = \bigcup_{p=1}^n A_p$$

où

$$A_p = (|S_1| < \varepsilon, |S_2| < \varepsilon, \dots, |S_{p-1}| < \varepsilon, |S_p| \geq \varepsilon).$$

Noter que les A_p sont disjoints et que $S_n - S_p$ est indépendante de $S_p \mathbb{I}_{A_p}$ et d'espérance nulle. De ce fait, pour tout $1 \leq p \leq n$,

$$\mathbf{E}(S_n^2 \mathbb{I}_{A_p}) = \mathbf{E}((S_n - S_p)^2 \mathbb{I}_{A_p}) + \mathbf{E}(S_p^2 \mathbb{I}_{A_p}) \geq \varepsilon^2 \mathbf{P}(A_p)$$

et sommant sur p , $\mathbf{E}(S_n^2) \geq \varepsilon^2 \mathbf{P}(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon)$. □

Une conséquence importante de la Proposition 5.2.2 est le résultat suivant.

5.2.3 Corollaire. *On suppose que S_n converge dans L^2 . Alors, il y a aussi convergence p.s. C'est en particulier le cas sous les hypothèses du Théorème 5.2.1.*

PREUVE : Soit $\varepsilon > 0$. Fixons $n_0 \in \mathbb{N}$. D'après l'inégalité de Kolmogorov, on a

$$\mathbf{P}(\max_{n_0 \leq k \leq n} |S_k - S_{n_0}| \geq \varepsilon) \leq \frac{\mathbf{E}((S_n - S_{n_0})^2)}{\varepsilon^2}.$$

Passant à la limite croissante en n dans le membre de gauche, et utilisant la convergence de $(S_n)_{n \geq 0}$ dans L^2 vers S dans le membre de droite, on obtient

$$\mathbf{P}(\sup_{n_0 \leq k} |S_k - S_{n_0}| \geq \varepsilon) \leq \frac{\mathbf{E}((S - S_{n_0})^2)}{\varepsilon^2}. \quad (5.2.1)$$

On peut construire une suite strictement croissante d'entiers $(n_k)_{k \geq 0}$ telle que

$\mathbf{E}((S - S_{n_k})^2) < \frac{1}{2^k}$. Par le Lemme de Borel-Cantelli, utilisant (5.2.1), on a alors pour presque chaque ω , l'existence d'un rang m_ω , tel que $\sup_{n \geq m_\omega} |S_n - S_{m_\omega}|(\omega) < \varepsilon$, et donc

tel que, pour $n \geq m_\omega$ et $p \geq 0$, $|S_n - S_{n+k}|(\omega) < 2\varepsilon$. C'est dire que la suite (S_n) est p.s. de Cauchy. □

Le premier exercice ci-dessous est un très beau résultat, dû à Lévy, qui montre un autre type d'amélioration de convergence.

◇ Exercice 5.1

Pour X variable aléatoire réelle, on appelle médiane de X tout réel m tel que $\mathbf{P}(X \leq m) \geq \frac{1}{2}$ et $\mathbf{P}(X \geq m) \geq \frac{1}{2}$. Noter que ce réel n'est pas unique en général. On notera systématiquement $\mu(X)$ une médiane de X . Noter que $-\mu(-X)$ est une médiane de X .

On désigne par $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes. On notera en particulier qu'on ne fait aucune hypothèse sur les moments des X_n . On pose

enfin, comme auparavant, $S_n = \sum_{k=1}^n X_k$.

a) Montrer que, pour $\varepsilon > 0$, $\mathbf{P}(|X| > \varepsilon) < \frac{1}{2} \Rightarrow |\mu(X)| < \varepsilon$. Dédire que si $X_n \rightarrow 0$ en probabilités, $\mu(X_n) \rightarrow 0$.

b) Soit $\varepsilon > 0$. Montrer l'inégalité

$$\mathbf{P}\left(\max_{1 \leq k \leq n} [S_k - \mu(S_k - S_n)] \geq \varepsilon\right) \leq 2\mathbf{P}(S_n \geq \varepsilon).$$

Indication : Ecrire que $S_n = [S_k - \mu(S_k - S_n)] + [S_n - S_k + \mu(S_k - S_n)]$.

c) Dédire l'inégalité de Lévy

$$\mathbf{P}\left(\max_{1 \leq k \leq n} |S_k - \mu(S_k - S_n)| \geq \varepsilon\right) \leq 2\mathbf{P}(|S_n| \geq \varepsilon).$$

d) On suppose que $(S_n)_{n \geq 1}$ converge en probabilités vers une variable aléatoire S .

d1) Montrer qu'il existe une sous-suite $(S_{n_j})_{j \geq 0}$ convergeant p.s. vers S telle que

$$\mathbf{P}(|S_{n_{j+1}} - S_{n_j}| \geq \frac{1}{2^j}) < \frac{1}{2^j}.$$

d2) Pour tout $j \geq 1$, on pose

$$T_j = \max_{n_j \leq n \leq n_{j+1}} |S_n - S_{n_j} - \mu(S_n - S_{n_{j+1}})|.$$

Montrer, en utilisant l'inégalité de Lévy, que la suite $(T_j)_{j \geq 1}$ converge p.s. vers 0.

d3) Montrer que $(S_n)_{n \geq 1}$ converge p.s. vers S .

◇ Exercice 5.2

Soit $c > 0$ et supposons que (Z_n) soit une suite indépendante de v.a., centrées, et uniformément bornées par c ($|Z_n| \leq c$). Soit $\varepsilon > 0$ et notons $S_k = Z_1 + \dots + Z_k$. Alors

$$1 - \frac{(\varepsilon + c)^2}{\sum_{k \leq n} \text{Var}(Z_k)} \leq \mathbf{P}\left(\max_{k \leq n} |S_k| \geq \varepsilon\right).$$

(Inégalité de Kolmogorov à gauche)

Exercice 5.3

Soit $(\Omega, \mathcal{B}, P) = ([0, 1], \mathcal{B}, m)$ (m =mesure de Lebesgue), et $f \in L^2$. On pose $I = \int_0^1 f(t)dt$. Soit $(U_i)_{i \geq 1}$ une suite i.i.d., de loi commune la loi uniforme sur $[0, 1]$. On pose

$$I_n = \frac{1}{n}(f(U_1) + \dots + f(U_n)).$$

- a) Montrer que $I_n \xrightarrow[n \rightarrow \infty]{\mathbf{P}\text{-p.s.}} I$;
 b) Utiliser l'inégalité de Chebyshev pour estimer $\mathbf{P}(|I_n - I| > \frac{a}{\sqrt{n}})$.
 (Intégration de Monte-Carlo).

◇ Exercice 5.4

Soit $\{X_n\}_{n \geq 1}$ une suite de v.a. indépendantes.

- a) On suppose que pour un $c > 0$ les trois séries

$$\sum_{n=1}^{\infty} P\{|X_n| \geq c\}, \sum_{n=1}^{\infty} \mathbf{E}(X_n^c), \sum_{n=1}^{\infty} \text{Var}(X_n^c)$$

convergent, où on note $X_n^c = X_n 1_{\{|X_n| \leq c\}}$. Montrer que la série $\sum_{n=1}^{\infty} X_n$ converge p.s.

- b) Montrer que la convergence des séries $\sum_{n=1}^{\infty} \mathbf{E}(X_n)$ et $\sum_{n=1}^{\infty} \mathbf{E}(|X_n|^p)$ pour un nombre réel $1 < p \leq 2$ entraîne la convergence p.s. de la série $\sum_{n=1}^{\infty} X_n$.
 (Théorème des 3 séries)

Exercice 5.5

Soit $f \in \mathcal{C}([0, 1])$, et soit $f_n(x) = \sum_{m=0}^n C_n^m x^m (1-x)^{n-m} f(\frac{m}{n})$ le $n^{\text{ième}}$ polynôme de Bernstein associé à f .

- a) Soit S_n la somme de n variables indépendantes de loi de Bernoulli de paramètre p , $0 < p < 1$. Montrer que $\mathbf{E}f(S_n/n) = f_n(p)$.
 b) Soit $M = \|f\|_{\infty}$. Soit $\varepsilon > 0$. Montrer qu'il existe $\delta > 0$ tel que

$$|\mathbf{E}(f(\frac{S_n}{n})) - f(p)| \leq \varepsilon + 2M\mathbf{P}(|\frac{S_n}{n} - p| > \delta).$$

- c) Montrer que $\mathbf{P}(|\frac{S_n}{n} - p| > \delta) \leq \frac{1}{4n\delta^2}$.
 d) Montrer que $\|f_n(x) - f(x)\|_{\infty} \rightarrow 0$ si $n \rightarrow \infty$.
 (Théorème de Stone-Weierstrass, preuve de Bernstein)

Exercice 5.6

Soit $(X_i)_{i \geq 1}$ une suite i.i.d. dans L^4 telle que $\mu = \mathbf{E}(X_1)$. On veut montrer que $S_n/n \rightarrow \mu$ **P**-p.s..

- Montrer que l'on peut supposer sans perte de généralité que $\mu = 0$.
- Développer $\mathbf{E}(S_n^4)$ pour en déduire que $\mathbf{E}(S_n^4) \leq Cn^2$.
- En déduire avec l'inégalité de Markov que $\mathbf{P}(|S_n| > n\varepsilon) \leq C/(n^2\varepsilon^4)$.
- Montrer à l'aide du premier Lemme de Borel-Cantelli que $\mathbf{P}(\limsup\{|S_n| > \varepsilon\}) = 0$. Conclure.

(Loi forte dans L^4).

Exercice 5.7

Pour $x \in [0, 1]$, on note $(\varepsilon_n(x))_{n \geq 1} \in \{0, 1\}^{\mathbb{N}^*}$ une suite telle que

$$x = \sum_{n \geq 1} \frac{\varepsilon_n(x)}{2^n}.$$

- Montrer que si x n'est pas rationnel de la forme $\frac{p}{2^m}$, alors la suite $(\varepsilon_n(x))$ est unique. Montrer que sinon il en existe au plus deux.
- On note $\pi : \{0, 1\}^{\mathbb{N}^*} \rightarrow [0, 1]$ l'application définie par $\pi((\varepsilon_n)_{n \geq 1}) = \sum_{n \geq 1} \varepsilon_n 2^{-n}$. Sur $\Omega := \{0, 1\}^{\mathbb{N}^*}$, on choisit la tribu $\mathcal{B} = \otimes_{\mathbb{N}^*} \mathbf{P}(\{0, 1\})$. Montrer qu'alors π est mesurable de (Ω, \mathcal{B}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- Soit $p \in]0, 1[$, soit \mathbf{P}_p la probabilité sur $(\{0, 1\}, \mathcal{P}(\{0, 1\}))$ définie par $\mathbf{P}_p(\{1\}) = p$. Soit enfin $\tilde{\mathbf{P}}_p = \otimes_{\mathbb{N}^*} \mathbf{P}_p$ la probabilité produit correspondante sur (Ω, \mathcal{B}) . Montrer que $(\tilde{\mathbf{P}}_{\frac{1}{2}})_\pi = m$, la mesure de Lebesgue sur $[0, 1]$.
- En utilisant la loi forte des grands nombres, une fois, cependant, son utilisation justifiée, montrer que $p \neq q \Rightarrow \tilde{\mathbf{P}}_p \perp \tilde{\mathbf{P}}_q$. Montrer qu'il en est de même de $(\tilde{\mathbf{P}}_p)_\pi$ et de $(\tilde{\mathbf{P}}_q)_\pi$. Ces mesures sont-elles discrètes ? Absolument continues par rapport à la mesure de Lebesgue ?

Indication : Considérer $X_k((\varepsilon_n)_{n \geq 1}) = \varepsilon_k$.

Chapitre 6

Convergence en loi unidimensionnelle

Commençons par étudier quelques outils de représentation des lois.

6.1 Fonctions de répartition

6.1.1 Définition. La fonction de répartition d'une v.a.r. X est la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par

$$F_X(t) = \mathbf{P}(X \leq t) = \mathbf{P}_X(\cdot - \infty, t], \quad t \in \mathbb{R}.$$

C'est un exercice instructif de démontrer que F_X a les propriétés suivantes :

$$\left\{ \begin{array}{ll} (i) & F_X \text{ est croissante} \\ (ii) & F_X \text{ est continue à droite} \\ (iii) & \lim_{t \rightarrow -\infty} F_X(t) = 0 \\ (iv) & \lim_{t \rightarrow +\infty} F_X(t) = 1. \end{array} \right.$$

6.1.2 Lemme. Si $F : \mathbb{R} \rightarrow [0, 1]$ satisfait aux propriétés (i — iv) énoncées précédemment, alors il existe (Ω, \mathcal{B}, P) et une variable aléatoire X sur (Ω, \mathcal{B}) tels que $F = F_X$. De plus si X et Y sont deux variables aléatoires telles que $F_X = F_Y$, alors elles ont même loi, i.e. $\mathbf{P}_X = \mathbf{P}_Y$.

PREUVE : Voir l'exercice 6.1 ci-dessous. □

Le résultat suivant proposé en exercice est un moyen commode de construire une variable aléatoire de loi donnée.

◇ Exercice 6.1

Soit X une variable aléatoire réelle de loi uniforme sur $[0, 1]$. On considère μ une probabilité sur \mathbb{R} et on note $F_\mu(t) = \mu(\cdot - \infty, t]$ (la fonction de répartition de μ). On pose, pour $0 \leq x \leq 1$, $G(x) = \inf\{t \in \mathbb{R}, F_\mu(t) \geq x\}$. Déterminer la loi de la variable aléatoire $G(X)$.

6.1.3 Théorème. La suite $(X_i)_{i \in I}$ est indépendante si et seulement si quel que soit $J \subset I$ fini, si $\mathcal{X}_J = (X_j)_{j \in J}$, alors

$$F_{\mathcal{X}_J}((t_j)_{j \in J}) = \prod_{j \in J} F_{X_j}(t_j) \quad (\text{i.e. } \mu_{\mathcal{X}_J} = \bigotimes_{j \in J} \mu_{X_j}).$$

PREUVE : Soit $(t_j)_{j \in J} \in \mathbb{R}^J$. Si on a l'indépendance, alors

$$F_{\mathcal{X}_J}((t_j)_{j \in J}) = \mathbf{P}(\cap_{j \in J} X_j^{-1}(] - \infty, t_j]) = \prod_{j \in J} \mathbf{P}(X_j^{-1}(] - \infty, t_j]) = \prod_{j \in J} F_{X_j}(t_j).$$

Réciproquement, l'ensemble des pavés $\prod_{j \in J}] - \infty, t_j]$ forme un π -système générateur de $\mathcal{B}(\mathbb{R}^J)$. Si donc l'on suppose que $\mu_{\mathcal{X}_J} = \otimes_{j \in J} \mu_{X_j}$, c'est donc que $\mu_{\mathcal{X}_J}$ et $\otimes_{j \in J} \mu_{X_j}$ coïncident sur ce π -système générateur, et donc sont égales.

Par le Théorème 3.2.9, on obtient l'indépendance. \square

6.2 Fonctions caractéristiques

Une représentation de la loi d'une variable aléatoire réelle se fait aussi par la *fonction caractéristique*. Elle correspond à la transformée de Fourier de la loi de X , μ_X , que les probabilistes utilisent sans la constante 2π .

6.2.1 Définition. Soit X une variable aléatoire à valeurs dans \mathbb{R} . La fonction caractéristique de X est $\varphi_X(t) = \mathbf{E}(e^{itX})$.

Nous allons démontrer une version du théorème d'inversion (Théorème 1.15).

6.2.2 Théorème. Soit φ la fonction caractéristique de la variable aléatoire X de loi μ . Soient $a < b$ deux réels tels que $\mu(\{a\}) = \mu(\{b\}) = 0$. Alors

$$\mu(]a, b]) = \lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt.$$

Avant de démontrer ce théorème, notons que ce résultat implique bien que deux variables aléatoires X et Y ayant même fonction caractéristique ont même loi. En effet, en utilisant le théorème on a $\mu_X(]a, b]) = \mu_Y(]a, b])$ pour tout choix de réels a et b tels que $\mu_X(\{a\}) = \mu_X(\{b\}) = \mu_Y(\{a\}) = \mu_Y(\{b\}) = 0$. Or, les points de masse de μ_X et μ_Y étant au plus dénombrables, ces intervalles forment clairement un π -système engendrant la tribu $\mathcal{B}(\mathbb{R})$.

PREUVE :

Notons $S(t) = \int_0^t \frac{\sin u}{u} du$. S est une fonction continue sur \mathbb{R}_+ et bornée puisque $\lim_{t \rightarrow +\infty} S(t) = \frac{\pi}{2}$. Notons que si on note $\text{sgn}(\theta)$ le signe du réel θ (avec $\text{sgn}(0) = 0$), pour $T \geq 0$, on a

$$\int_0^T \frac{\sin(t\theta)}{t} dt = \text{sgn}(\theta) S(T|\theta|).$$

On a

$$\begin{aligned} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt &= \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \left(\int_{\mathbb{R}} e^{itx} \mu(dx) \right) dt \\ &= \int_{\mathbb{R}} \left(\int_{-T}^T \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt \right) \mu(dx) \end{aligned}$$

par le théorème de Fubini dont l'utilisation est légitime puisque $(t, x) \mapsto \frac{e^{it(x-a)} - e^{it(x-b)}}{it}$ est bornée, et donc intégrable par rapport à la mesure $\lambda \otimes \mu$ qui est finie sur $[-T, T] \times \mathbb{R}$. Notons que, en utilisant la parité des fonctions cos et sin, et puisque $T \geq 0$,

$$\int_{-T}^T \frac{e^{it(x-a)} - e^{it(x-b)}}{it} dt = \int_{-T}^T \frac{\sin(t(x-a)) - \sin(t(x-b))}{t} dt$$

$$= 2 \int_0^T \frac{\sin(t(x-a)) - \sin(t(x-b))}{t} dt = 2 [\operatorname{sgn}(x-a)S(T|x-a|) - \operatorname{sgn}(x-b)S(T|x-b|)].$$

Donc,

$$\frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = \int_{\mathbb{R}} \left[\frac{\operatorname{sgn}(x-a)}{\pi} S(T|x-a|) - \frac{\operatorname{sgn}(x-b)}{\pi} S(T|x-b|) \right] \mu(dx).$$

Remarquons que

$$\lim_{T \rightarrow +\infty} \frac{\operatorname{sgn}(x-a)}{\pi} S(T|x-a|) - \frac{\operatorname{sgn}(x-b)}{\pi} S(T|x-b|) = \psi(x) \text{ avec}$$

$$\psi(x) = \begin{cases} 0 & \text{si } x < a \\ 1/2 & x = a \\ 1 & a < x < b \\ 1/2 & x = b \\ 0 & x > b \end{cases}$$

Par ailleurs, la fonction $\frac{\operatorname{sgn}(x-a)}{\pi} S(T|x-a|) - \frac{\operatorname{sgn}(x-b)}{\pi} S(T|x-b|)$ est bornée donc le théorème de convergence dominée s'applique et on a

$$\lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = \int_{-\infty}^{\infty} \psi(x) \mu(dx) = \frac{1}{2} \mu(\{a\}) + \frac{1}{2} \mu(\{b\}) + \mu(]a, b[) = \mu(]a, b[)$$

puisque $\mu(\{a\}) = \mu(\{b\}) = 0$. □

Ci-dessous sont indiquées les fonctions caractéristiques de quelques lois usuelles.

Exercice 6.2

- a) Montrer que la fonction caractéristique d'une loi uniforme sur $[a, b]$ est $\varphi(t) = \frac{e^{itb} - e^{ita}}{it}$.
- b) Montrer que la fonction caractéristique d'une loi normale $\mathcal{N}(m, \sigma^2)$ est $\varphi(t) = e^{itm - \frac{t^2}{2}}$.
- c) Montrer que la fonction caractéristique d'une loi exponentielle de paramètre $\lambda > 0$ est $\varphi(t) = \frac{\lambda}{\lambda + it}$.
- d) Montrer que la fonction caractéristique d'une loi de Cauchy, de densité $\frac{1}{\pi(1+x^2)}$, est $\varphi(t) = e^{-|t|}$.
- e) Montrer que la fonction caractéristique d'une loi de Poisson de paramètre $\lambda > 0$ est $\varphi(t) = e^{\lambda(e^{it} - 1)}$.

6.2.3 Proposition. Une fonction caractéristique est uniformément continue.

PREUVE : En effet,

$$|\phi_X(t+h) - \phi_X(t)| \leq \mathbf{E}(|e^{itX}(e^{ihX} - 1)|) = \mathbf{E}(|e^{ihX} - 1|).$$

Or, $0 = \lim_{h \rightarrow 0} |e^{ihX} - 1|$. Comme $|e^{ihX} - 1| \leq 2$, on a $\mathbf{E}(|e^{ihX} - 1|) \rightarrow_{h \rightarrow 0} 0$ par convergence dominée. \square

Poursuivons par quelques développements de Taylor : par intégration par parties,

$$\int_0^x (x-s)^n e^{is} ds = \frac{x^{n+1}}{n+1} + \frac{i}{n+1} \int_0^x (x-s)^{n+1} e^{is} ds, \quad (6.2.1)$$

dont il découle par induction que

$$e^{ix} = \sum_{k \leq n} \frac{(ix)^k}{k!} + \frac{i^{n+1}}{n!} \int_0^x (x-s)^n e^{is} ds. \quad (6.2.2)$$

En remplaçant n par $n-1$ dans 6.2.1, et en déduisant la valeur de l'intégrale de droite dans 6.2.1, puis en injectant dans 6.2.2, il vient

$$e^{ix} = \sum_{k \leq n} \frac{(ix)^k}{k!} + \frac{i^n}{(n-1)!} \int_0^x (x-s)^{n-1} (e^{is} - 1) ds. \quad (6.2.3)$$

En estimant les intégrales dans 6.2.2 et 6.2.3 et en séparant les cas $x \geq 0$ et $x < 0$, il vient

$$\left| e^{ix} - \sum_{k \leq n} \frac{(ix)^k}{k!} \right| \leq \min \left\{ \frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!} \right\}. \quad (6.2.4)$$

Remplaçons x par une v.a. X dans 6.2.4. Il s'ensuit que si $\mathbf{E}(|X|^n) < \infty$,

$$\left| \phi_X(t) - \sum_{k \leq n} \frac{(it)^k}{k!} \mathbf{E}(X^k) \right| \leq \mathbf{E} \left[\min \left\{ \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right\} \right]. \quad (6.2.5)$$

Et donc si t est tel que $\lim_n \frac{|t|^n \mathbf{E}(|X|^n)}{n!} = 0$, alors

$$\phi_X(t) = \sum_{k \geq 0} \frac{(it)^k}{k!} \mathbf{E}(X^k). \quad (6.2.6)$$

Si la série 6.2.6 converge sur un voisinage de 0, alors ϕ_X permet de retrouver les moments de X :

$$\phi_X^{(k)}(0) = i^k \mathbf{E}(X^k).$$

Par contre X n'a pas forcément de moments de tous ordres (considérer par exemple la loi de densité proportionnelle à $\frac{1}{x^3+1} \mathbb{I}_{[0,+\infty[}$). Toutefois, on peut énoncer

6.2.4 Proposition. $\mathbf{E}(|X|^k) < \infty$, alors $\phi_X^{(k)}$ est uniformément continue, et

$$\phi_X^{(k)}(t) = \mathbf{E}((iX)^k e^{itX}).$$

PREUVE : Observons que

$$\frac{\phi_X(t+h) - \phi_X(t)}{h} - \mathbf{E}(iXe^{itX}) = \mathbf{E} \left[e^{itX} \frac{e^{ihX} - 1 - ihX}{h} \right].$$

D'après 6.2.4 pour $n = 1$, on a $|e^{ihx} - (1 + ihx)| \leq \min\{\frac{1}{2}x^2, 2|x|\}$, et donc par convergence dominée (par $2|X|$), on en déduit $\phi_X'(t) = \mathbf{E}(iXe^{itX})$, dont l'uniforme continuité se démontre sur le modèle de la preuve de celle de ϕ_X .

Si le moment d'ordre k existe, de proche en proche, on obtient les formules recherchées jusqu'à l'ordre k . □

On en déduit immédiatement la forme du développement de Taylor de la fonction ϕ_X .

6.2.5 Corollaire. Si $\mathbf{E}(|X|^n) < \infty$, alors au voisinage de 0,

$$\phi_X(t) = \sum_{k \leq n} \frac{(it)^k}{k!} \mathbf{E}(X^k) + o(t^n)$$

et en particulier si $n = 2$,

$$\phi_X(t) = 1 + it\mathbf{E}(X) - \frac{t^2}{2}\mathbf{E}(X^2) + o(t^2). \tag{6.2.7}$$

6.3 Etude de la convergence en loi unidimensionnelle

Le premier résultat que nous allons établir est qu'une suite de variables aléatoires qui converge a de bonnes propriétés de bornitude. Introduisons la définition suivante.

6.3.1 Définition. On dit que la suite de v.a.r. (X_n) est tendue si pour tout $\varepsilon > 0$, il existe $K > 0$ tel que $\forall n \geq 0, \mathbf{P}(|X_n| > K) < \varepsilon$.

Exercice 6.3

Montrer que toute variable aléatoire réelle X est tendue.

6.3.2 Proposition. Soit (X_n) une suite de variables aléatoires réelles qui converge en loi vers X . Alors la suite (X_n) est tendue.

PREUVE : On note $\mu_{X_n} = \mu_n$. Soit $\varepsilon > 0$. D'après l'exercice 6.3, on peut trouver $K > 0$ tel que $\mu_X([-K, K]) \geq 1 - \varepsilon$. Soit alors φ_K la fonction qui vaut 1 sur $[-K, K]$, 0 si $|x| > K + 1$ et est raccordée continument de façon linéaire. φ_K est continue et bornée donc, par définition, $\mathbf{E}(\varphi_K(X_n)) \rightarrow \mathbf{E}(\varphi_K(X))$.

De ce fait, pour $n \geq N$, $\mathbf{E}(\varphi_K(X_n)) \geq \mathbf{E}(\varphi_K(X)) - \varepsilon \geq 1 - 2\varepsilon$.

Mais, $\mu_n([- (K + 1), K + 1]) \geq \mathbf{E}(\varphi_K(X_n))$. Finalement, $\forall n \geq N, \mathbf{P}(|X_n| > K + 1) < 2\varepsilon$. Comme les autres n sont en nombre fini, quitte à grandir K , on peut supposer que la même relation est vraie pour tout $n \geq 0$. □

D'une façon générale, on définit classiquement pour une suite de mesures trois types de convergence (étroite, faible, et vague) suivant que les fonctions-test choisies sont continues bornées (cas de la convergence en loi telle que nous l'avons introduite), continues tendant vers 0 à l'infini, ou continues à support compact. Dans le cas qui nous intéresse ici (où μ_{X_n} et μ_X sont des mesures de probabilités), ces trois notions sont identiques, comme le montre le lemme suivant.

6.3.3 Lemme. Soit (X_n) une suite de v.a.r. et X une v.a.r. telles que pour toute f continue à support compact de \mathbb{R} dans \mathbb{R} , $\mathbf{E}(f(X_n))$ converge vers $\mathbf{E}(f(X))$. Alors X_n converge en loi vers X .

PREUVE : La démonstration de la Proposition 6.3.2 n'utilisait que des fonctions à support compact. De ce fait, elle est encore valable ici et on en déduit que la suite (X_n) est tendue. Considérons alors f une fonction continue bornée. On peut trouver une fonction ψ continue à support compact telle que ψ et f coïncident sur $[-K, K]$ et $\|\psi\|_\infty \leq \|f\|_\infty$. Alors

$$|\mathbf{E}(f(X_n) - f(X))| \leq \mathbf{E}(|f(X_n) - \psi(X_n)|) + |\mathbf{E}(\psi(X_n) - \psi(X))| + \mathbf{E}(|\psi(X) - f(X)|).$$

Le terme $|\mathbf{E}(\psi(X_n) - \psi(X))|$ tend vers 0 par hypothèse, et comme $|f - \psi| \leq (\|f\|_\infty + \|\psi\|_\infty)\mathbb{I}_{|x| > K} \leq 2\|f\|_\infty\mathbb{I}_{|x| > K}$, les deux autres termes sont majorés par $2\varepsilon\|f\|_\infty$. \square

La tension d'une suite de variables aléatoires correspond en fait à la relative compacité comme le montre la proposition suivante. Rappelons l'énoncé du théorème de représentation suivant (voir [23]) :

6.3.4 Théorème. (Riesz-Markov) À toute forme linéaire Φ continue sur $C_K(\mathbb{R})$, ensemble des fonctions réelles continues à support compact dans \mathbb{R} , est associée une unique mesure μ sur \mathbb{R} telle que

$$\forall f \in C_K(\mathbb{R}), \Phi(f) = \int_{\mathbb{R}} f(x) d\mu(x).$$

6.3.5 Proposition. Soit (X_n) une suite de variables aléatoires réelles tendue. Alors, il existe une sous suite $(X_{n_k})_{k \geq 0}$ qui converge en loi.

PREUVE : On note de nouveau $\mu_{X_n} = \mu_n$. La démonstration repose sur le procédé diagonal de Cantor. Choisissons $(f_k)_{k \geq 0}$ une suite de fonctions continues à support compact dense pour la convergence uniforme dans l'ensemble des fonctions réelles à support compact. La suite de terme général $\mu_n(f_0) = \int_{\mathbb{R}} f_0(x) \mu_n(dx)$ étant bornée dans \mathbb{R} , elle admet une sous-suite convergente $\mu_{n_k^0}(f_0)$ vers un réel noté $\mu(f_0)$. Considérons la suite $\mu_{n_k^0}(f_1)$: elle admet une sous-suite notée $\mu_{n_k^1}(f_1)$ convergente vers un réel $\mu(f_1)$. Ainsi de suite, on arrive à construire par récurrence une sous suite (μ_{p_n}) de (μ_n) telle que pour tout $k \geq 0$, $(\mu_{p_n}(f_k))$ soit une suite convergente vers $\mu(f_k)$.

Soit h une fonction continue à support compact. On va montrer que $(\mu_{p_n}(h))$ converge : sa limite sera notée $\mu(h)$. Soit $\varepsilon > 0$. Il existe f_k telle que $\|f_k - h\|_\infty < \varepsilon$. De ce fait,

$$|\mu_{p_n}(h) - \mu_{p_m}(h)| \leq |\mu_{p_n}(h) - \mu_{p_n}(f_k)| + |\mu_{p_n}(f_k) - \mu_{p_m}(f_k)| + |\mu_{p_m}(f_k) - \mu_{p_m}(h)|,$$

et les trois termes précédents peuvent être rendus inférieurs à ε pour m et n assez grands. De ce fait, $(\mu_{p_n}(h))$ est une suite de Cauchy. μ définit une forme linéaire sur $C_K(\mathbb{R})$. De plus, comme $|\mu(h)| = \lim_{n \rightarrow +\infty} |\mu_{p_n}(h)| \leq \|h\|_\infty$, cette forme est continue, et c'est donc une mesure. Il reste à vérifier que c'est une probabilité, pour obtenir, d'après le Lemme 6.3.3, la convergence en loi de $(\mu_{p_n})_{n \geq 0}$ vers μ .

Comme la suite $(X_n)_{n \geq 0}$ est tendue, on peut trouver $K > 0$ et une fonction h continue à support compact dans $[-(K+1), K+1]$ telle que $0 \leq h \leq 1$ et $\mu_{p_n}(h) > 1 - \varepsilon, \forall n \geq 0$. En particulier, $\mu(h) = \lim_{n \rightarrow +\infty} \mu_{p_n}(h) \geq 1 - \varepsilon$. Alors, $\mu(\mathbb{R}) \geq \mu([-(K+1), K+1]) \geq \mu(h) \geq 1 - \varepsilon$. Enfin, si g_k est une fonction à support compact $[-(k+1), k+1]$ et égale à 1 sur $[-k, k]$, à valeurs comprises entre 0 et 1, on a $\forall k \geq 0, \mu_{p_n}(g_k) \leq 1$ et donc $\mu(g_k) \leq 1$. Comme $g_k \uparrow \mathbb{I}_{\mathbb{R}}$, le théorème de convergence monotone donne $\mu(\mathbb{R}) \leq 1$. \square

On sait que les lois unidimensionnelles peuvent être représentées à l'aide des fonctions de répartition $F_n(x) = \mathbf{P}(X_n \leq x)$. La proposition suivante montre comment ces fonctions peuvent servir à vérifier une convergence en loi.

6.3.6 Proposition. *Soient (X_n) une suite de variables aléatoires réelles et X une variable aléatoire réelle. On note F_n (resp. F) la fonction de répartition de X_n (resp. X). Alors, sont équivalentes*

- (i) $\mu_{X_n} \Rightarrow \mu_X$
- (ii) $F_n(t) \rightarrow F(t)$ en tout point t où F est continue.

PREUVE : Supposons (i). Soit t un point de continuité pour F . Soit $\varepsilon > 0$. On peut trouver $\alpha > 0$ tel que $\mathbf{P}(X \in [t - \alpha, t + \alpha]) < \varepsilon$.

Soit alors $\psi_-(t)$ la fonction égale à 1 jusqu'en $t - \alpha$, à 0 après t et raccordée linéairement ; de même, $\psi_+(t)$ est la fonction égale à 1 jusqu'en t , à 0 après $t + \alpha$, et raccordée linéairement. On a clairement $\psi_-(t) \leq \mathbb{I}_{]-\infty, t]} \leq \psi_+(t)$ et $|\psi_+(t) - \psi_-(t)| \leq \mathbb{I}_{[t - \alpha, t + \alpha]}$. De ce fait, $\mathbf{E}(\psi_-(X_n)) \leq F_n(t) \leq \mathbf{E}(\psi_+(X_n))$ et les deux termes extrêmes tendent vers $\mathbf{E}(\psi_-(X))$ et $\mathbf{E}(\psi_+(X))$ par hypothèse. De plus, $\mathbf{E}(|\psi_+(X) - \psi_-(X)|) \leq \mathbf{P}(X \in [t - \alpha, t + \alpha]) < \varepsilon$. Enfin, $|\mathbf{E}(\psi_+(X)) - F(t)| \leq \mathbf{E}(|\psi_+(X) - \mathbb{I}_{]-\infty, t]}(X)|) \leq \mathbf{E}(|\psi_+(X) - \psi_-(X)|) < \varepsilon$ et de même $|\mathbf{E}(\psi_-(X)) - F(t)| < \varepsilon$. Finalement, la seule valeur d'adhérence de la suite $F_n(t)$ est $F(t)$, d'où le résultat cherché.

Réciproquement, supposons (ii). Soit f une fonction continue à support compact. Comme l'ensemble des points de discontinuité de F (croissante) est dénombrable, son complémentaire est dense dans \mathbb{R} . Soit $\varepsilon > 0$. On peut clairement trouver une fonction en escalier $\tilde{f} = \sum_{k=0}^n \alpha_k \mathbb{I}_{]t_k, t_{k+1}]}$ telle que $\|f - \tilde{f}\|_\infty < \varepsilon$ et telle que les points t_k soient des points de continuité de F . Notons qu'alors $\mathbf{E}(\tilde{f}(X_n)) \rightarrow \mathbf{E}(\tilde{f}(X))$ (car $\mathbf{E}(\mathbb{I}_{]t_k, t_{k+1}]}(X_n)) = F_n(t_{k+1}) - F_n(t_k)$). De plus, $|\mathbf{E}(f(X_n)) - \mathbf{E}(f(X))| \leq 2\varepsilon + |\mathbf{E}(\tilde{f}(X_n)) - \mathbf{E}(\tilde{f}(X))|$ d'où le résultat cherché. \square

Le résultat suivant est souvent utile en pratique pour achever la preuve d'une convergence en loi.

6.3.7 Corollaire. *Soient $(X_n)_{n \geq 0}$ une suite de variables aléatoires réelles tendue et X une variable aléatoire réelle. Si toutes les sous-suites de $(X_n)_{n \geq 0}$ qui convergent en loi convergent vers X , alors la suite $(X_n)_{n \geq 0}$ converge en loi vers X .*

PREUVE : Supposons que $(X_n)_{n \geq 0}$ ne converge pas en loi vers X . D'après la Proposition 6.3.6, on peut alors trouver $x \in \mathbb{R}$ tel que F est continue en x et $F_{X_n} \not\rightarrow F(x)$, où F est la fonction de répartition de X . Donc, il existe $\varepsilon > 0$ et une suite $(n_p)_{p \geq 0}$ d'entiers tels que $|F_{X_{n_p}}(x) - F(x)| > \varepsilon, \forall p$. La suite $(X_{n_p})_{p \geq 0}$ étant tendue, on peut trouver une sous-suite $(X_{n_{p_k}})_{k \geq 0}$ qui converge en loi, vers X par hypothèse. On a alors $F_{X_{n_{p_k}}}(x) \rightarrow F(x)$ puisque F est continue en x (Proposition 6.3.6), ce qui est une contradiction. \square

Le résultat suivant relie les fonctions caractéristiques à la convergence en loi.

6.3.8 Théorème. (Lévy) *Soit $(X_n)_{n \geq 0}$ une suite de variables réelles et X une variable réelle. Les propositions suivantes sont équivalentes :*

- (i) X_n converge en loi vers X
- (ii) Pour tout $t \in \mathbb{R}$, $\varphi_n(t) = \varphi_{X_n}(t) \rightarrow \varphi_X(t) = \varphi(t)$.

PREUVE : (i) \Rightarrow (ii) vient de la définition de la convergence en loi car $x \mapsto e^{itx}$ est une fonction continue et bornée sur \mathbb{R} .

Supposons maintenant (ii). Le lemme suivant montre que la suite $(X_n)_{n \geq 0}$ est tendue.

6.3.9 Lemme. Soit $(X_n)_{n \geq 0}$ une suite de v.a.r. et X une v.a.r. telles que pour tout $t \in \mathbb{R}$, $\varphi_n(t) \rightarrow \varphi(t)$. Alors $(X_n)_{n \geq 0}$ est tendue.

PREUVE : soit $\varepsilon > 0$. On a par le théorème de Fubini, pour tout $u > 0$,

$$\begin{aligned} \frac{1}{u} \int_{-u}^u (1 - \varphi_n(t)) dt &= \int_{-\infty}^{+\infty} \left[\frac{1}{u} \int_{-u}^u (1 - e^{itx}) dt \right] \mu_{X_n}(dx) = 2 \int_{-\infty}^{+\infty} \left(1 - \frac{\sin ux}{ux}\right) \mu_{X_n}(dx) \\ &\geq 2 \int_{|x| \geq \frac{2}{u}} \left(1 - \frac{1}{|ux|}\right) \mu_{X_n}(dx) \geq \mathbf{P}(|X_n| \geq \frac{2}{u}). \end{aligned} \quad (6.3.1)$$

Par ailleurs, le théorème de convergence dominée permet d'affirmer que $\frac{1}{u} \int_{-u}^u (1 - \varphi_n(t)) dt$ converge vers $\frac{1}{u} \int_{-u}^u (1 - \varphi(t)) dt$ et comme φ est continue en 0 et $\varphi(0) = 1$, on peut choisir u pour que $\frac{1}{u} \int_{-u}^u (1 - \varphi(t)) dt < \varepsilon$. On obtient alors $C > 0$ et $N \in \mathbb{N}$ tels que pour $n \geq N$, $\mathbf{P}(|X_n| > C) < 2\varepsilon$. Comme la suite finie de variables $(X_k)_{k \leq N}$ est tendue, quitte à agrandir C , l'inégalité est valable pour tout $n \geq 0$ et la suite (X_n) est donc tendue. \square

Revenons à la démonstration de (ii) \Rightarrow (i). Soit $(X_{n_p})_{p \geq 0}$ une sous-suite de $(X_n)_{n \geq 0}$ qui converge en loi vers la loi μ . On a alors d'après (i) \Rightarrow (ii), $\varphi_{X_{n_p}}(t) \rightarrow \hat{\mu}(t), \forall t \in \mathbb{R}$, où $\hat{\mu}$ désigne la fonction caractéristique de la loi μ .

Par l'hypothèse (ii), on a donc $\hat{\mu}(t) = \varphi(t), \forall t \in \mathbb{R}$ et donc $(X_{n_p})_{p \geq 0}$ converge en loi vers X . D'après le Corollaire 6.3.7, $(X_n)_{n \geq 0}$ converge en loi vers X . \square

En fait, le théorème précédent prend la forme plus précise suivante.

◇ Exercice 6.4

Soit (X_n) une suite de v.a.r. telle que $\forall t \in \mathbb{R}, \varphi_n(t) \rightarrow \varphi(t)$, où φ est une fonction continue en 0. Alors, φ est une fonction caractéristique et (X_n) converge en loi vers la loi de fonction caractéristique φ .

Exercice 6.5

Démontrer, à l'aide des fonctions caractéristiques, que la convergence en probabilités implique la convergence en loi.

◇ Exercice 6.6

la loi μ_X d'une variable aléatoire est dite à réseau s'il existe $a \in \mathbb{R}$ et $b > 0$ tels que $\mu_X(a + b\mathbb{Z}) = 1$.

a) Montrer que X admet une loi à réseau si et seulement si il existe $t \in \mathbb{R}^*$ tel que $|\varphi_X(t)| = 1$.

b) Montrer que si $|\varphi_X(t)| = |\varphi_X(t')| = 1$ pour deux réels t et t' tels que $t/t' \notin \mathbb{Q}$, X est constante.

Les opérations sur les variables aléatoires réelles ne sont pas systématiquement continues pour la convergence en loi. Ainsi, on n'a pas en général la convergence en loi de $X_n Y_n$ vers XY si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$ en loi. Citons néanmoins l'important

6.3.10 Théorème. (Slutsky) *On a les propriétés suivantes :*

a) Si $V_n \xrightarrow{\mathcal{L}} V$ et $W_n \xrightarrow{\mathcal{L}} 0$, alors $V_n + W_n \xrightarrow{\mathcal{L}} V$

b) Si $X_n \xrightarrow{\mathcal{L}} X$ et si $Y_n \xrightarrow{\mathcal{L}} a$, alors $X_n Y_n \xrightarrow{\mathcal{L}} aX$.

PREUVE :

a) Observons d'abord que $W_n \xrightarrow{\mathcal{L}} 0 \Rightarrow W_n \rightarrow 0$ en probabilités.

Soit $v \in \mathbb{R}$, et soient $w' < v < w''$ tels que $\mathbf{P}(V = w') = \mathbf{P}(V = w'') = 0$.

Si $w' < v - \varepsilon < v < v + \varepsilon < w''$, alors

$$\mathbf{P}(V_n \leq w') - \mathbf{P}(|W_n| \geq \varepsilon) \leq \mathbf{P}(V_n + W_n \leq v) \leq \mathbf{P}(V_n \leq w'') + \mathbf{P}(|W_n| \geq \varepsilon),$$

ce qui, puisque $V_n \xrightarrow{\mathcal{L}} V$ et $W_n \rightarrow 0$ en probabilités, donne à la limite

$$\mathbf{P}(V \leq w') \leq \liminf_n \mathbf{P}(V_n + W_n \leq v) \leq \limsup_n \mathbf{P}(V_n + W_n \leq v) \leq \mathbf{P}(V \leq w'').$$

Si alors $\mathbf{P}(V = v) = 0$, F_V est continue en v , et on peut trouver des $w' < v < w''$ arbitrairement proches de v , et aux propriétés requises. Donc $V_n + W_n \xrightarrow{\mathcal{L}} V$ puisque les fonctions caractéristiques convergent aux points de continuité de celle de V .

b) Notons que $X_n Y_n = X_n(Y_n - a) + aX_n$, et que si $Y_n \xrightarrow{\mathcal{L}} a$, alors $Y_n - a \xrightarrow{\mathcal{L}} 0$ (puisque c'est une convergence en probabilités), et enfin que si $X_n \xrightarrow{\mathcal{L}} X$, alors $aX_n \xrightarrow{\mathcal{L}} aX$ (distinguer ici le cas $a = 0$, et raisonner à l'aide des fonctions de répartition).

Par a), si nous montrons que $X_n(Y_n - a) \xrightarrow{\mathcal{L}} 0$, alors la seconde partie s'ensuivra. Notons donc $\delta_n = Y_n - a$, et nos hypothèses sont à présent que $\delta_n \xrightarrow{\mathcal{L}} 0$ et $X_n \xrightarrow{\mathcal{L}} X$. La conclusion recherchée est $X_n \delta_n \xrightarrow{\mathcal{L}} 0$.

Soient $\varepsilon > 0$ et K tel que $\mathbf{P}(|X| > K) \leq \varepsilon$. Choisissons $\lambda > 0$ tel que $\mathbf{P}(|X| = \frac{\varepsilon}{\lambda}) = 0$ et $\frac{\varepsilon}{\lambda} > K$. Notons que comme la loi de X ne charge pas les points $\pm \frac{\varepsilon}{\lambda}$, on a d'après la Proposition 6.3.6, $\mathbf{P}(|X_n| \geq \frac{\varepsilon}{\lambda}) \rightarrow \mathbf{P}(|X| \geq \frac{\varepsilon}{\lambda}) \leq \mathbf{P}(|X| > K) \leq \varepsilon$.

Alors, on peut écrire

$$\begin{aligned} \limsup_n \mathbf{P}(|\delta_n X_n| \geq \varepsilon) &\leq \limsup_n \mathbf{P}(|\delta_n| \geq \lambda) + \limsup_n \mathbf{P}(|X_n| \geq \varepsilon/\lambda) \\ &\leq 0 + \varepsilon. \end{aligned}$$

Donc, on a $\limsup_n \mathbf{P}(|\delta_n X_n| \geq \varepsilon) = 0$, soit $\delta_n X_n \xrightarrow{\mathbf{P}} 0$. □

◇ **Exercice 6.7**

- 1) Soit (X_n) une suite de variables aléatoires réelles.
- a) Montrer que si (X_n) est une suite tendue, alors (φ_{X_n}) est uniformément équi-continue.
- b) Montrer que si (X_n) converge en loi vers X , alors $\varphi_{X_n}(t)$ converge uniformément vers $\varphi_X(t)$, sur tout intervalle borné de \mathbb{R} .
- 2) Soient (a_n) et (b_n) deux suites réelles avec $a_n > 0$, pour tout n . Soient X et Y deux variables aléatoires non constantes et (X_n) une suite de variables aléatoires telle que X_n converge en loi vers X et $a_n X_n + b_n$ converge en loi vers Y .
On note φ_n, φ et ψ les fonctions caractéristiques de X_n, X et Y .
- a) Montrer que $|\varphi_n(a_n t)|$ converge uniformément, sur tout intervalle borné, vers $|\psi(t)|$.
- b) Dédire que 0 n'est pas valeur d'adhérence de la suite (a_n) .
- c) En "échangeant" les rôles de φ et ψ , montrer que $+\infty$ n'est pas valeur d'adhérence de la suite a_n .
- d) Montrer que (a_n) converge vers $a > 0$.
- e) Montrer que e^{itb_n} converge vers $\frac{\psi(t)}{\varphi(at)}$ dans un voisinage de 0.
- f) En considérant $\int_0^t e^{isb_n} ds$, déduire que (b_n) converge.

Comme signalé plus haut, la convergence en loi n'est pas une convergence spatiale, et ceci complique un certain nombre de manipulations. En particulier, pour des suites de v.a.r. vérifiant des convergences spatiales (du type convergence L^1 ou convergence p.s.), on peut être amené naturellement à utiliser des théorèmes de passage à la limite (convergence dominée ou monotone etc.). Il existe un moyen très efficace pour transformer une convergence en loi en une convergence de type spatial, pourvu qu'on accepte de changer d'espace. C'est l'objet du théorème suivant.

6.3.11 Théorème. (Skorokhod) *Soit (X_n) une suite de variables aléatoires réelles qui converge en loi vers une variable X . Il existe alors une suite de variables aléatoires (Y_n) et une variable Y sur l'espace $([0, 1], \mathcal{B}([0, 1]), \lambda)$ telles que les lois de Y_n et X_n (resp. de Y et de X) coïncident, et $Y_n \rightarrow Y, \lambda$ -p.s.*

PREUVE : considérons $F_n(t)$ la fonction de répartition de X_n et posons $Y_n(x) = \inf\{t \in \mathbb{R}, F_n(t) \geq x\}$. D'après l'exercice 6.1, Y_n est une variable aléatoire réelle sur $[0, 1]$ de même loi que X_n . On définit F et Y de façon similaire.

Soit alors $0 < x < 1$. Soit $\varepsilon > 0$. Puisque les points de discontinuité de F sont en nombre au plus dénombrable, on peut choisir s point de continuité de F tel que $Y(x) - \varepsilon < s < Y(x)$. On a $F(s) < x$. Comme $F_n(s) \rightarrow F(s)$, pour n assez grand on a $F_n(s) < x$ et donc $Y(x) - \varepsilon < s < Y_n(x)$. De ce fait $\underline{\lim} Y_n(x) \geq Y(x)$.

Soit maintenant $1 > x' > x$. Soit t un point de continuité de F tel que $Y(x') < t < Y(x') + \varepsilon$. Comme F est continue à droite, on a $x' \leq F(Y(x'))$ et donc on a puisque F est croissante $x < x' \leq F(Y(x')) \leq F(t)$ et donc, pour n assez grand, $x \leq F_n(t)$ et donc

$$Y_n(x) \leq t < Y(x') + \varepsilon.$$

D'où, $\overline{\lim} Y_n(x) \leq Y(x')$ si $x < x'$. Finalement, pour tout $x < x' < 1$, on a

$$Y(x) \leq \underline{\lim} Y_n(x) \leq \overline{\lim} Y_n(x) \leq Y(x').$$

De ce fait, $Y_n(x) \rightarrow Y(x)$ si Y est continue en x , ce qui est vrai λ -p.s. puisque Y est croissante. □

Exercice 6.8

Montrer que si $(X_n)_{n \geq 0}$ converge en loi vers X , $\mathbf{E}(|X|) \leq \underline{\lim} \mathbf{E}(|X_n|)$.

6.4 Le théorème limite central

6.4.1 Théorème. (Laplace-Gauss) Soit (X_n) une suite i.i.d. dans $L^2(\mathbf{P})$ ayant une moyenne c et une variance $\sigma^2 > 0$. Notons $S_N = \sum_{k=1}^N X_k$. Alors

$$\frac{S_N - Nc}{\sigma\sqrt{N}} \xrightarrow{\mathcal{L}} Y,$$

où Y a pour loi la loi normale centrée réduite $\mathcal{N}(0, 1)$, de densité

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

PREUVE : Notons $Y_N = \frac{S_N - Nc}{\sigma\sqrt{N}}$. Quitte à remplacer X_k par $(X_k - c) / \sigma$, on peut supposer que $c = 0$ et $\sigma = 1$. Notons $\phi_X = \phi_{X_k}$.

Nous allons montrer que $\phi_{Y_N} \rightarrow \phi_Y$ puis conclure en invoquant le théorème de continuité.

On sait que $\phi_Y(t) = e^{-\frac{t^2}{2}}$.

Par l'indépendance, on a $\phi_{Y_N}(t) = \phi_X(\frac{t}{\sqrt{N}})^N$. D'après (6.2.7), on a

$$\phi_{Y_N}(t) - e^{-\frac{t^2}{2}} = \left(1 - \frac{t^2}{2N} + o\left(\frac{1}{N}\right)\right)^N - \left(e^{-\frac{t^2}{2N}}\right)^N.$$

Le lemme suivant nous sera utile.

6.4.2 Lemme. Si $z_1, \dots, z_n, y_1, \dots, y_n$ sont des nombres complexes de module inférieur ou égal à 1, on a

$$\left| \prod_{i=1}^n z_i - \prod_{i=1}^n y_i \right| \leq \sum_{i=1}^n |z_i - y_i|.$$

PREUVE du lemme : Par une récurrence immédiate, en écrivant que

$$\prod_{i=1}^n z_i - \prod_{i=1}^n y_i = \left(\prod_{i=1}^{n-1} z_i - \prod_{i=1}^{n-1} y_i\right) \cdot z_n + \prod_{i=1}^{n-1} y_i \cdot (z_n - y_n). \square$$

Ceci s'applique à l'expression précédente de $|\phi_{Y_N}(t) - e^{-\frac{t^2}{2}}|$ pour conclure que l'on a $\phi_{Y_N}(t) - e^{-\frac{t^2}{2}} \rightarrow 0$. D'après le Théorème 6.3.8, on obtient la convergence cherchée.

Notons que puisque la loi normale ne charge pas les points, pour tout $x \in \mathbb{R}$,

$$\mathbf{P}\left(\frac{S_N - Nc}{\sigma\sqrt{n}} \leq x\right) \rightarrow \int_{-\infty}^x \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt. \quad \square$$

Citons sans preuve le

6.4.3 Théorème. (Théorème de Berry-Essen) Sous les hypothèses du théorème limite central, si en plus $\mathbf{E}(|X_1|^3) = \rho < \infty$, alors si $x \in \mathbb{R}$,

$$\left| \mathbf{P} \left(\frac{S_N - Nc}{\sigma\sqrt{n}} \leq x \right) - \int_{-\infty}^x \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \right| \leq \frac{3\rho}{\sqrt{n}\sigma^3}.$$

Exercice 6.9

Que X ait une loi a) $\mathcal{P}(\lambda)$ ou b) $\Gamma(1, \lambda)$ alors $Y_\lambda = (X - \mathbf{E}(X))\text{Var}(X)^{-1/2}$ converge en loi vers $\mathcal{N}(0, 1)$.

Montrer que

$$\lim_n e^{-n} \left(1 + n + \frac{n^2}{2!} + \dots + \frac{n^n}{n!} \right) = \frac{1}{2}.$$

Exercice 6.10

Trouver une suite de fonctions caractéristiques qui converge partout ponctuellement mais telle que la limite ne soit pas une fonction caractéristique (indication : lois uniformes sur $[-n, +n]$).

Exercice 6.11

Montrer que si $x \geq 0$, et $n \rightarrow \infty$, alors

$$\sum_{|k-n/2| \leq x\sqrt{n}/2} C_n^k \sim 2^n \int_{-x}^x e^{-u^2/2} du / \sqrt{2\pi};$$

$$\sum_{|k-n| \leq x\sqrt{n}} n^k / k! \sim e^n \int_{-x}^x e^{-u^2/2} du / \sqrt{2\pi}.$$

◇ **Exercice 6.12**

Soit $(X_i)_{i \geq 1}$ une suite iid de var équidistribuées sur $\{-1, +1\}$. Montrer qu'on a la convergence en loi

$$\sqrt{\frac{3}{n^3}} \sum_{k=1}^n kX_k \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

◇ **Exercice 6.13**

Soit (X_n) une suite de variables aléatoires telle que $\sup_n \mathbf{E}(X_n^2) < \infty$.

On suppose que (X_n) converge en loi vers X .

Montrer que $\mathbf{E}(X_n) \rightarrow \mathbf{E}(X)$.

◇ Exercice 6.14

On considère la densité de probabilités

$$h_\alpha(t) = \frac{\alpha}{\sqrt{2\pi}} \frac{1}{t^{\frac{3}{2}}} e^{-\frac{\alpha^2}{2t}} \mathbb{I}_{]0, \infty[}(t),$$

où $\alpha > 0$.

a) Soient $\alpha > 0$ et $\beta > 0$ donnés, et $X \sim h_\alpha, Y \sim h_\beta$ deux variables aléatoires indépendantes. Déterminer la loi de $X + Y$ (on pourra utiliser des transformées de Laplace).

b) Soit (X_n) une suite de variables aléatoires indépendantes de loi possédant la densité h_α . Déterminer la loi de

$$\frac{X_1 + \dots + X_n}{n^2}.$$

Y-a-t-il une contradiction avec la loi des grands nombres ?

c) Étudier la convergence de la suite

$$Y_n = \frac{1}{n^2} \max_{k \leq n} X_k$$

(on pourra regarder la fonction de répartition de Y_n)

◇ Exercice 6.15

(Formule de Stirling)

Soit $(X_i)_{i \geq 1}$ une suite i.i.d. de loi commune la loi $\mathcal{P}(1)$. Notons $S_n = X_1 + \dots + X_n$.

a) Montrer que $\mathbf{E} \left[\left(\frac{S_n - n}{\sqrt{n}} \right)^- \right] = e^{-n} \sum_{k=0}^n \binom{n-k}{\sqrt{n}} \frac{n^k}{k!} = \frac{n^{n+\frac{1}{2}} e^{-n}}{n!}$.

b) Montrer que $\left(\frac{S_n - n}{\sqrt{n}} \right)^- \xrightarrow{\mathcal{L}} N^-$, où N^- désigne la partie négative d'une v.a. N de loi $\mathcal{N}(0, 1)$.

c) Montrer que $\mathbf{E} \left[\left(\frac{S_n - n}{\sqrt{n}} \right)^- \right] \rightarrow_n \mathbf{E}(N^-) = \frac{1}{\sqrt{2\pi}}$.

d) Conclure que $n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$.

Chapitre 7

Théorèmes limites dans \mathbb{R}^k

7.1 Convergence faible et suites tendues

Nous allons tout d'abord prolonger à \mathbb{R}^k l'étude qui avait été faite à la section 6.3. Plusieurs définitions équivalentes sont possibles pour définir la convergence en loi. Pour des raisons qui apparaîtront par la suite, nous allons privilégier une définition à l'aide des fonctions de répartition, prolongeant ainsi le résultat de la Proposition 6.3.6.

7.1.1 Définition. *La fonction de répartition d'un vecteur aléatoire $\mathcal{X} = (X_1, \dots, X_d)$ est par définition l'application $F_{\mathcal{X}} : \mathbb{R}^d \rightarrow [0, 1]$ définie par*

$$F_{\mathcal{X}}(t_1, \dots, t_d) = \mathbf{P}(X_1 \leq t_1, \dots, X_d \leq t_d).$$

Dans le cas d'un vecteur aléatoire, contrairement au cas $d = 1$ (Lemme 6.1.2), la caractérisation des applications qui sont des fonctions de répartition de vecteurs aléatoires en dimension $d > 1$ est complexe. Elle nécessite d'ajouter des conditions supplémentaires de "positivité sur les contours". Nous n'insisterons pas sur ce point. Néanmoins, pour comprendre l'existence de ces conditions, on peut penser au cas \mathbb{R}^2 , $\mathcal{X} = (X_1, X_2)$, et soient $x_1 < y_1$ et $x_2 < y_2$. Alors $0 \leq \mathbf{P}((x_1 < X_1 \leq y_1) \wedge (x_2 < X_2 \leq y_2))$, et donc

$$F_{\mathcal{X}}(y_1, y_2) + F_{\mathcal{X}}(x_1, x_2) - F_{\mathcal{X}}(x_1, y_2) - F_{\mathcal{X}}(y_1, x_2) \geq 0.$$

Soient μ et μ_n , $n \geq 1$, des mesures de probabilité sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Soient X et X_n , $n \geq 1$, des vecteurs aléatoires à valeurs dans \mathbb{R}^k . Soient enfin F et F_n , $n \geq 1$, des fonctions de répartition de lois de probabilité, ou de vecteurs aléatoires, sur \mathbb{R}^k .

7.1.2 Définition. *Nous dirons que F_n converge faiblement vers F , noté $F_n \Rightarrow F$, si $F_n(x) \rightarrow F(x)$ en tout point de continuité x de F .*

Nous dirons que μ_n converge faiblement vers μ , et nous noterons $\mu_n \Rightarrow \mu$, si les fonctions de répartition associées convergent faiblement.

Nous dirons que X_n converge en loi vers X , noté $X_n \xrightarrow{\mathcal{L}} X$, si $\mu_{X_n} \Rightarrow \mu_X$.

Comme pour la droite réelle, nous caractérisons la convergence faible :

7.1.3 Proposition. *(Conditions équivalentes pour la convergence faible dans \mathbb{R}^k)*
Avec les mêmes notations, $\mu_n \Rightarrow \mu$ si et seulement si l'une des conditions suivantes est satisfaite :

- (i) $\lim_n E_{\mu_n}(f) = E_{\mu}(f)$ pour toute $f : \mathbb{R}^k \rightarrow \mathbb{R}$ continue bornée ;
- (ii) $\limsup_n \mu_n(C) \leq \mu(C)$ pour tout $C \subset \mathbb{R}^k$ fermé ;
- (iii) $\liminf_n \mu_n(G) \geq \mu(G)$ pour tout $G \subset \mathbb{R}^k$ ouvert ;
- (iv) $\lim_n \mu_n(A) = \mu(A)$ pour tout $A \in \mathcal{B}(\mathbb{R}^k)$ de frontière μ -négligeable .

PREUVE : Montrons d'abord que les quatre conditions exprimées sont équivalentes entre elles.

1. *Preuve de (i) \Rightarrow (ii)* : Soit C un ensemble fermé non vide et posons $\text{dist}(x, C) = \inf\{|x - y| : y \in C\}$. On a $x \mapsto \text{dist}(x, C)$ continue. Posons, pour $j > 0$,

$$\phi_j(t) = \begin{cases} 1 & \text{si } t \leq 0; \\ 1 - jt & \text{si } 0 \leq t \leq \frac{1}{j}; \\ 0 & \text{si } \frac{1}{j} \leq t. \end{cases}$$

Puis notons $f_j(x) = \phi_j(\text{dist}(x, C))$. La fonction f_j est continue bornée par 1, et $\lim_{j \rightarrow +\infty} f_j(x) = \mathbb{I}_C(x)$ car C est fermé. Par (i), on a $\limsup_n \mu_n(C) \leq \lim_n E_{\mu_n}(f_j) = E_{\mu}(f_j)$, et par convergence dominée, $\lim_{+\infty} E_{\mu}(f_j) = \mu(C)$.

2. *Preuve de (ii) \Leftrightarrow (iii)* : Poser $C = \mathbb{R}^k \setminus G$.

3. *Preuve de ((ii) \wedge (iii)) \Rightarrow (iv)* : En combinant les deux, on a

$$\mu(\overset{\circ}{A}) \leq \liminf_n \mu_n(\overset{\circ}{A}) \leq \liminf_n \mu_n(A) \leq \limsup_n \mu_n(A) \leq \limsup_n \mu_n(\bar{A}) \leq \mu(\bar{A}),$$

et puisque $\mu(\partial A) = 0$, on a $\mu(\overset{\circ}{A}) = \mu(\bar{A}) = \mu(A)$. De ce fait, $\lim_n \mu_n(A)$ existe et vaut $\mu(A)$.

4. *Preuve de (iv) \Rightarrow (i)* : Supposons f continue et $|f| \leq K$ (f bornée). Etant donné $\varepsilon > 0$, choisissons $\alpha_0 < \alpha_1 < \dots < \alpha_l$ tels que $\alpha_0 < -K < K < \alpha_l$ et que $\alpha_{i+1} - \alpha_i < \varepsilon$, et que $\mu(\{f = \alpha_i\}) = 0$ pour chaque i .

Notons alors $A_i = \{\alpha_{i-1} < f \leq \alpha_i\}$. Alors par continuité de f ,

$$\{\alpha_{i-1} < f < \alpha_i\} \subset \overset{\circ}{A}_i \subset \bar{A}_i \subset \{\alpha_{i-1} \leq f \leq \alpha_i\},$$

et donc $\partial A_i \subset \{f = \alpha_{i-1}\} \cup \{f = \alpha_i\}$. Et donc $\mu(\partial A_i) = 0$. D'autre part, que ν désigne μ_n ou μ ,

$$\left| \int f d\nu - \sum_{i=1}^l \alpha_i \nu(A_i) \right| \leq \varepsilon, \text{ et } \sum_{i=1}^l \alpha_i \mu_n(A_i) \rightarrow \sum_{i=1}^l \alpha_i \mu(A_i),$$

d'après (iv). Puisque ε est arbitraire, on en déduit (i).

Ainsi nous avons montré (i) \Rightarrow (ii) \Leftrightarrow (iii) \Rightarrow (iv) \Rightarrow (i), et reste donc à montrer par exemple que (iv) \Rightarrow ($\mu_n \Rightarrow \mu$) et que ($\mu_n \Rightarrow \mu$) \Rightarrow (iii).

5. *Preuve de (iv) \Rightarrow ($\mu_n \Rightarrow \mu$)* : Soit $x = (x_1, \dots, x_k) \in \mathbb{R}^k$, et soient F_n et F les fonctions de répartition respectives de μ_n et μ . On a donc $\mu_n(S_x) = F_n(x)$ et $\mu(S_x) = F(x)$ si

$$S_x = \{y \in \mathbb{R}^k : y_i \leq x_i, 1 \leq i \leq k\}.$$

La continuité de F en x équivaut à $\mu(\partial S_x) = 0$, et donc (iv) \Rightarrow ($F_n \Rightarrow F$).

6. *Preuve de $(\mu_n \Rightarrow \mu) \Rightarrow (iii)$* : Comme seul un ensemble au plus dénombrable d'hyperplans parallèles peuvent éventuellement ne pas être μ -négligeables, il existe un ensemble D dense dans \mathbb{R} tel que pour chaque i , et chaque $t \in D$, $\mu(\{x : x_i = t\}) = 0$.

Soit \mathcal{A} la classe des pavés $A = \prod_{i=1}^k]a_i, b_i]$ avec $a_i, b_i \in D$. Par construction, on a $\mu_n(A) \rightarrow \mu(A)$, puisque par hypothèse, $F_n \rightarrow F$ là où F est continue, et que tant $\mu_n(A)$ que $\mu(A)$ s'expriment à l'aide des valeurs de F_n et F en des points de continuité de F .

Et donc, si $B \in \tilde{\mathcal{A}}$, la classe des unions finies disjointes d'éléments de \mathcal{A} , on obtient $\mu_n(B) \rightarrow \mu(B)$. Maintenant par densité de D , si $G \subset \mathbb{R}^k$ est ouvert, il s'écrit comme union croissante d'éléments de $\tilde{\mathcal{A}}$, disons $G = \cup_n \uparrow B_n$, et alors

$$\liminf_n \mu_n(G) \geq \liminf_n \mu_n(B_q) = \mu(B_q),$$

pour tout q , et donc $\mu(G) = \sup_q \mu(B_q) \leq \liminf_n \mu_n(G)$. □

7.1.4 Proposition. (Transfert et convergence faible dans \mathbb{R}^k) Soit $h : \mathbb{R}^k \rightarrow \mathbb{R}^j$ une application. Alors l'ensemble D_h de ses points de discontinuité est mesurable. Supposons h borélienne, et d'autre part que $\mu_n \Rightarrow \mu$ dans $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Alors si $\mu(D_h) = 0$, on a

$$(\mu_n)_h \Rightarrow \mu_h.$$

PREUVE : Montrons d'abord que $D_h \in \mathcal{B}(\mathbb{R}^k)$: Soit

$$A(\varepsilon, \delta) = \{x \in \mathbb{R}^k : \exists y, z \in \mathbb{R}^k : |x - y| < \delta \wedge |x - z| < \delta \wedge |h(y) - h(z)| > \varepsilon\}.$$

Alors que h soit mesurable ou pas, $A(\varepsilon, \delta)$ est ouvert et

$$D_h = \cup_{\varepsilon \in \mathbb{Q}^+} \cap_{\delta \in \mathbb{Q}^+} A(\varepsilon, \delta) \in \mathcal{B}(\mathbb{R}^k).$$

Soit $C \subset \mathbb{R}^j$ fermé. Alors $\overline{h^{-1}C} \subset D_h \cup h^{-1}C$. Si $\mu_n \Rightarrow \mu$, par (ii),

$$\limsup_n \mu_n(h^{-1}C) \leq \limsup_n \mu_n(\overline{h^{-1}C}) \leq \mu(\overline{h^{-1}C}) \leq \mu(D_h \cup h^{-1}C) = \mu(h^{-1}C),$$

et donc encore par (ii), $(\mu_n)_h \Rightarrow \mu_h$. □

Notons que dans le cas de la droite réelle, nous avons utilisé, pour montrer le résultat correspondant, le Théorème 6.3.11 : il reste valable pour \mathbb{R}^k , mais sa preuve est plus complexe. Citons-le sans preuve :

7.1.5 Théorème. (Skorokhod version \mathbb{R}^k) Si $X_n, n \geq 1$ et X sont des vecteurs aléatoires à valeurs dans \mathbb{R}^k , et si $X_n \xrightarrow{\mathcal{L}} X$, alors il existe un espace probabilisé $(\Omega, \mathcal{B}, \mathbf{P})$, une suite de vecteurs aléatoires $Y_n : \Omega \rightarrow \mathbb{R}^k$ et un autre vecteur aléatoire $Y : \Omega \rightarrow \mathbb{R}^k$, tels que Y_n converge vers Y \mathbf{P} -p.s., et qu'en outre

$$\mathbf{P}_{Y_n} = \mathbf{P}_{X_n} \text{ et } \mathbf{P}_Y = \mathbf{P}_X.$$

À l'instar de la dimension 1, on donne la définition suivante.

7.1.6 Définition. Une suite de probabilités $(\mu_n)_{n \geq 1}$ sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ est **tendue** si, quel que soit $\varepsilon > 0$, il existe un rectangle $A \in \mathcal{B}(\mathbb{R}^k)$ tel que

$$\inf_n \mu_n(A) \geq 1 - \varepsilon.$$

Une adaptation, que nous admettrons, de la preuve correspondante pour la droite réelle, donne les résultats suivants (le lecteur pourra se reporter à [6]).

7.1.7 Proposition. *a) De toute suite tendue on peut extraire une sous-suite qui converge faiblement.*

b) Si une suite tendue est telle que toutes ses suites extraites qui convergent convergent vers la même limite, alors elle converge faiblement, vers la limite commune.

7.2 Fonctions caractéristiques

Soient $t = (t_1, \dots, t_k) \in \mathbb{R}^k$ et $X = (X_1, \dots, X_k)$ un vecteur aléatoire. Alors on pose

$$\langle t, X \rangle = \sum_{i=1}^k t_i X_i.$$

7.2.1 Définition. La fonction caractéristique de X (ou de sa loi μ_X) est définie par

$$\phi_X(t) = E(e^{i\langle t, X \rangle}) = \int_{\mathbb{R}^k} e^{i\langle t, x \rangle} d\mu(x) = \phi_\mu(t) \text{ si } \mu_X = \mu.$$

Pour l'essentiel, les propriétés de ϕ_X sont identiques à celles du cas de la droite ($k = 1$). Citons notamment l'uniforme continuité, et surtout la formule d'inversion (Théorème 1.15).

7.2.2 Théorème. (Formule d'inversion dans \mathbb{R}^k) *Si $A = \prod_{i=1}^k]a_i, b_i]$ est un rectangle borné, et si $\mu(\partial A) = 0$, alors*

$$\mu(A) = \lim_{T \rightarrow \infty} \frac{1}{(2\pi)^k} \int_{[-T, T]^k} \prod_{u=1}^k \frac{e^{-it_u a_u} - e^{-it_u b_u}}{it_u} \phi_\mu(t) dt,$$

où $dt = dt_1 \dots dt_k$.

Et donc la fonction caractéristique caractérise bien la loi.

Pour $t \in \mathbb{R}^k$, notons $h_t : \mathbb{R}^k \rightarrow \mathbb{R}$ définie par $h_t(x) = \langle t, x \rangle$.

7.2.3 Corollaire. (Caractérisation de μ sur les demi-plans) *Une mesure de probabilité μ sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ est entièrement déterminée par ses valeurs $\mu(\{x : \langle t, x \rangle \leq \alpha\})$, $t \in \mathbb{R}^k$, $\alpha \in \mathbb{R}$ (l'ensemble $\{x : \langle t, x \rangle \leq \alpha\}$ est ce que l'on appelle un demi-plan dans \mathbb{R}^k).*

PREUVE : On a $\mu(\{x : \langle t, x \rangle \leq \alpha\}) = \mu_{h_t}(] - \infty, \alpha])$. Par changement de variables (transfert),

$$\phi_{\mu_{h_t}}(s) = \phi_\mu(st_1, \dots, st_k),$$

et donc la donnée de μ sur les demi-plans détermine ϕ_μ , qui à son tour détermine μ . \square

7.2.4 Théorème. (Théorème de continuité dans \mathbb{R}^k) *Soient μ_n , $n \geq 1$ et μ des probabilités sur $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. Alors*

$$\mu_n \Rightarrow \mu \Leftrightarrow \phi_{\mu_n} \rightarrow \phi_\mu \text{ simplement.}$$

PREUVE : Si $\mu_n \Rightarrow \mu$, alors $\phi_{\mu_n}(t) = E_{\mu_n}(e^{i\langle t, x \rangle}) \rightarrow E_{\mu}(e^{i\langle t, x \rangle}) = \phi_{\mu}(t)$, d'après le (i) de la caractérisation de la convergence faible, puisque $x \mapsto e^{i\langle t, x \rangle}$ est continue bornée.

Réciproquement, si la convergence simple a lieu, alors pour chaque $t \in \mathbb{R}^k$, et chaque $s \in \mathbb{R}$,

$$\phi_{(\mu_n)_{h_t}}(s) \rightarrow \phi_{(\mu)_{h_t}}(s),$$

et donc par le théorème de continuité sur la droite, $(\mu_n)_{h_t} \Rightarrow \mu_{h_t}$.

Choisissons $t = e_u := (0, \dots, 0, \overbrace{1}^{u^{\text{ième}}}, 0, \dots, 0)$. La suite $((\mu_n)_{h_{e_u}})_{n \geq 1}$ converge faiblement, et donc est tendue. Pour $\varepsilon > 0$ donné, il existe $a_u < b_u$ tels que

$$\inf_n (\mu_n)_{h_{e_u}}([a_u, b_u]) \geq 1 - \frac{\varepsilon}{k}.$$

Alors en posant $A = \prod_{u=1}^k]a_u, b_u]$, on a $A^c \subset \cup_{u=1}^k \mathbb{R}^{u-1} \times]a_u, b_u]^c \times \mathbb{R}^{k-u}$, et donc

$$\begin{aligned} \sup_n \mu_n(A^c) &\leq \sum_{u=1}^k \sup_n \mu_n(\mathbb{R}^{u-1} \times]a_u, b_u]^c \times \mathbb{R}^{k-u}) \\ &= \sum_{u=1}^k \sup_n (\mu_n)_{h_{e_u}}(]a_u, b_u]^c) < k \times \frac{\varepsilon}{k} = \varepsilon, \end{aligned}$$

et donc la suite (μ_n) est tendue.

Supposons qu'une sous-suite $(\mu_{n_i})_{i \geq 1}$ converge faiblement vers ν . Par la convergence simple des fonctions caractéristiques, on a $\phi_{\nu} = \lim_n \phi_{\mu_{n_i}}$. Par ailleurs, le sens direct déjà traité donne $\lim_n \phi_{\mu_{n_i}} = \phi_{\nu}$. Donc $\phi_{\nu} = \phi_{\mu}$ et donc $\nu = \mu$. Par conséquent, toute sous-suite convergente converge vers μ . Donc $\mu_n \Rightarrow \mu$. \square

On peut aussi mentionner le théorème suivant, dont nous nous limiterons à ébaucher la démonstration.

7.2.5 Théorème. (Théorème de Cramér-Wold) *Une suite de vecteurs aléatoires dans \mathbb{R}^k $X_n = (X_{n,1}, \dots, X_{n,k})$ converge en loi vers le vecteur $X = (X_1, \dots, X_k)$ si et seulement si pour tout $t = (t_1, \dots, t_k) \in \mathbb{R}^k$,*

$$\sum_{u=1}^k t_u X_{n,u} \xrightarrow{\mathcal{L}} \sum_{u=1}^k t_u X_u.$$

PREUVE : Considérer les applications continues h_t pour la condition nécessaire. Pour la condition suffisante, chaque convergence en loi sur la droite (t fixé) entraîne la convergence simple des fonctions caractéristiques (théorème de continuité sur la droite), et donc pour chaque $s \in \mathbb{R}$,

$$\mathbf{E}(e^{is(\sum_{u=1}^k t_u X_{n,u})}) \rightarrow \mathbf{E}(e^{is(\sum_{u=1}^k t_u X_u)}).$$

Faire $s = 1$ pour en déduire la convergence simple des fonctions caractéristiques ϕ_{X_n} dans \mathbb{R}^k vers ϕ_X , et conclure avec la version \mathbb{R}^k du théorème de continuité. \square

7.3 Quelques prolongements

On regroupe ici quelques premiers résultats concernant les suites de mesures de probabilité sur un espace polonais (E, \mathcal{E}) (i.e. un espace métrique complet séparable muni de sa tribu borélienne \mathcal{E}). Ce type de situation se produira couramment dans l'approfondissement de l'étude des processus stochastiques en temps continu (mouvement brownien en premier lieu) pour laquelle on sera naturellement amené à travailler sur un espace canonique de fonctions continues.

(E, \mathcal{E}) désigne par la suite un espace polonais fixé.

7.3.1 Définition. On dit qu'une mesure μ sur E est **tendue** si pour tout $\varepsilon > 0$, on peut trouver un compact $K \subset E$ tel que $\mu(K^c) < \varepsilon$.

On a alors

7.3.2 Proposition. Toute mesure de probabilité sur (E, \mathcal{E}) est tendue.

PREUVE : Notons $(x_n)_{n \geq 0}$ une suite de E dense dans E . Soit $k > 0$. On a $\bigcup_{n \geq 0} \bar{B}(x_n, \frac{1}{k}) = E$, et de ce fait on peut choisir n_k tel que

$$\mu \left(\bigcup_{n=0}^{n_k} \bar{B}(x_n, \frac{1}{k}) \right) > 1 - \frac{\varepsilon}{2^k}.$$

Posons $K = \bigcap_{k \geq 1} \bigcup_{n=0}^{n_k} \bar{B}(x_n, \frac{1}{k})$. On a $\mu(K^c) \leq \sum_{k \geq 1} \mu \left(\left\{ \bigcup_{n=0}^{n_k} \bar{B}(x_n, \frac{1}{k}) \right\}^c \right) \leq \varepsilon$.

Montrons que K est compact. C'est clairement un fermé comme intersection de fermés. De plus, soit (a_n) une suite dans K (notons que $K \neq \emptyset$ puisque $x_0 \in K$). Appliquons alors un procédé d'extraction diagonal : il existe une sous-suite de (a_n) dans une boule $\bar{B}(x_{m_1}, 1)$ ($0 \leq m_1 \leq n_1$) ; on en choisit un qu'on note $a_{\varphi(1)}$. Puis, il existe une infinité de termes (i.e. une sous-suite) de la sous-suite précédente dans une boule $\bar{B}(x_{m_2}, \frac{1}{2})$ ($0 \leq m_2 \leq n_2$) ; on en choisit un noté $a_{\varphi(2)}$ et il est loisible d'exiger $\varphi(2) > \varphi(1)$. Et ainsi de suite. La sous-suite de (a_n) ainsi construite est clairement de Cauchy et donc convergente. \square

Soit $(\mu_n)_{n \in \mathbb{N}}$ une suite de mesures de probabilité sur E . Les définitions suivantes prolongent naturellement les définitions 7.1.6 et 4.1.1d :

7.3.3 Définition. 1) On dit que (μ_n) est une suite tendue si, pour tout $\varepsilon > 0$, il existe un compact $K \subset E$ tel que $\forall n \geq 0, \mu_n(K^c) < \varepsilon$.

2) Si μ est une mesure de probabilité sur E , on dira que (μ_n) converge étroitement vers μ si pour toute fonction f continue bornée de E dans \mathbb{R} , $\int_E f(x) \mu_n(dx) \rightarrow \int_E f(x) \mu(dx)$.

La difficulté d'étude de la convergence étroite provient de la structure en général complexe des fonctions continues sur E . Le cas particulier suivant est instructif. Des considérations générales sur la topologie de $\mathbb{R}^{\mathbb{N}}$ pourront être trouvées dans [13]. Rappelons uniquement que $f : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$ est continue en un point $(x_n^0)_{n \geq 0} \in \mathbb{R}^{\mathbb{N}}$ si et seulement si pour tout $\varepsilon > 0$, il existe $K > 0$ et $\alpha > 0$ tel que si $(y_n)_{n \geq 0}$ est tel que $|y_i - x_i^0| < \alpha, 0 \leq i \leq K$, alors $|f((y_n)_{n \geq 0}) - f((x_n^0)_{n \geq 0})| < \varepsilon$.

Comme en dimension finie, la tension apparaît comme une condition nécessaire de convergence.

7.3.4 Proposition. Dans un espace polonais, si $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ , $(\mu_n)_{n \in \mathbb{N}}$ est une suite tendue.

PREUVE : Commençons par rappeler un petit résultat de topologie des espaces métriques.

7.3.5 Lemme. Pour tous $\varepsilon > 0$, $\eta > 0$ et K un compact de (E, d) espace métrique, on peut trouver une fonction continue f à valeurs dans $[0, 1]$ et qui vaut 1 sur K et 0 si $d(x, K) > \eta$.

PREUVE : Il suffit de choisir $f(x) = (1 - \frac{d(x, K)}{\eta})^+$. \square

Fixons alors $n > 0$. On peut trouver K_n tel que $\mu(K_n) > 1 - \frac{\varepsilon}{2^n}$. Soit, grâce au lemme 7.3.5, f_n une fonction continue qui vaut 1 sur K_n et 0 si $d(x, K_n) > \frac{1}{n}$. On a

$$\int_E f_n d\mu_k \rightarrow \int_E f_n d\mu \geq \mu(K_n) > 1 - \frac{\varepsilon}{2^n}.$$

De ce fait, à partir d'un rang R , $\int_E f_n d\mu_k \geq 1 - \frac{\varepsilon}{2^n}$ et donc $\mu_k(K_n^*) \geq 1 - \frac{\varepsilon}{2^n}$ où l'on pose $K_n^* = \{x, d(x, K_n) \leq \frac{1}{n}\}$. K_n^* est recouvrable par un nombre fini de boules de rayon $\frac{2}{n}$: $B_1^n, \dots, B_{k_n}^n$. Quitte à ajouter des boules, on peut supposer que $\forall k, \mu_k(\bigcup_{i=1}^{k_n} B_i^n) > 1 - \frac{\varepsilon}{2^n}$. On considère alors $K = \bigcap_{n \geq 1} \bigcup_{i=1}^{k_n} B_i^n$ qui est compact (la démonstration est la même que dans la Proposition 7.3.2) et tel que $\mu_k(K^c) \leq \varepsilon, \forall k$. \square

7.3.6 Proposition. Soit $(\mu_n)_{n \geq 0}$ une suite de probabilités sur \mathbb{R}^N . Soit μ une probabilité sur \mathbb{R}^N . Si pour tout $k \geq 1$ et pour toute $f : \mathbb{R}^k \rightarrow \mathbb{R}$ continue et bornée, on a

$$\int_{\mathbb{R}^N} f(x_1, \dots, x_k) d\mu_n \rightarrow \int_{\mathbb{R}^N} f(x_1, \dots, x_k) d\mu \quad (7.3.1)$$

alors $(\mu_n)_{n \geq 0}$ converge étroitement vers μ .

PREUVE : Nous notons π_k la projection de \mathbb{R}^N dans \mathbb{R}^{k+1} qui à une suite $X = (x_n)_{n \geq 0}$ associe le $(k+1)$ -uplet (x_0, x_1, \dots, x_k) . Nous allons tout d'abord montrer le résultat suivant.

7.3.7 Lemme. Sous l'hypothèse (7.3.1) de la proposition 7.3.6, la suite de mesures $(\mu_n)_{n \geq 0}$ est tendue.

PREUVE : Soit $\varepsilon > 0$. Soit $k \geq 0$. Posons $\nu_n^k = \pi_k \cdot \mu_n$ et $\nu^k = \pi_k \cdot \mu$. Par l'hypothèse (7.3.1), $\nu_n^k \rightarrow \nu^k$ et donc $(\nu_n^k)_{n \geq 1}$ est une suite de mesures sur \mathbb{R}^{k+1} tendue. Il existe donc K_k compact de \mathbb{R}^{k+1} tel que

$$\nu_n^k(K_k) \geq 1 - \frac{\varepsilon}{2^k}, \forall n.$$

Posons $\tilde{K}_k = K_k \times \mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R} \times \dots \subset \mathbb{R}^N$ et $\tilde{K} = \bigcap_{k \geq 0} \tilde{K}_k$.

\tilde{K} est un compact. Il est en effet clairement fermé comme intersection de fermés. De plus, soit $X = ((x_n^k)_{n \geq 0})_{k \geq 1}$ une suite d'éléments de \tilde{K} . $(x_0^k)_{k \geq 1}$ est une suite de K_0 : elle admet donc une sous-suite $(x_0^{\varphi_1(k)})_{k \geq 1}$ convergente. Considérons alors $(x_0^{\varphi_1(k)}, x_1^{\varphi_1(k)})_{k \geq 1}$: c'est une suite de K_1 qui admet à ce titre une sous-suite $(x_0^{\varphi_2 \circ \varphi_1(k)}, x_1^{\varphi_2 \circ \varphi_1(k)})_{k \geq 1}$ convergente. Par procédé diagonal, on extrait ainsi une sous-suite $X_{\varphi_1(1)}, X_{\varphi_2 \circ \varphi_1(2)}, \dots$ convergente de $(X_k)_{k \geq 1}$. Par ailleurs, pour tout n ,

$$\mu_n(\tilde{K}^c) \leq \sum_{k \geq 1} \mu_n(\tilde{K}_k^c) = \sum_{k \geq 1} [1 - \mu_n(\tilde{K}_k)] = \sum_{k \geq 1} [1 - \nu_n^k(K_k)] \leq \sum_{k \geq 1} \frac{\varepsilon}{2^k} = \varepsilon. \quad \square$$

Terminons alors la preuve de la proposition 7.3.6. Soit $\varepsilon > 0$. D'après le lemme 7.3.7, (μ_n) est une suite tendue. Soit K un compact de \mathbb{R}^N tel que $\mu_n(K) \geq 1 - \varepsilon$ et $\mu(K) \geq 1 - \varepsilon$. Soit $f : \mathbb{R}^N \rightarrow \mathbb{R}$ continue et bornée. Sur K , f est uniformément continue : il existe $\alpha > 0$ et $N > 0$ tels que pour tous $X = (x_n)$ et $Y = (y_n)$ dans K , si $\max_{1 \leq i \leq N} |x_i - y_i| < \alpha$, on a $|f(X) - f(Y)| < \varepsilon$.

Posons alors $\varphi(X) = f(x_0, x_1, \dots, x_N, 0, 0, \dots)$ pour $X \in \mathbb{R}^N$: c'est une fonction continue sur \mathbb{R}^N . De plus, pour tout $X \in K$, $|f(X) - \varphi(X)| < \varepsilon$.

On a

$$\left| \int_{\mathbb{R}^N} f d\mu - \int_{\mathbb{R}^N} f d\mu_n \right| \leq \left| \int \varphi(x) \mu(dx) - \int \varphi(x) \mu_n(dx) \right| + \left| \int (\varphi - f)(x) \mu(dx) - \int (\varphi - f)(x) \mu_n(dx) \right|.$$

Or,

$$\begin{aligned} \left| \int_{\mathbb{R}^{\mathbb{N}}} (\varphi - f) d\mu_n \right| &\leq \left| \int_K (\varphi - f) d\mu_n \right| + \left| \int_{K^c} (\varphi - f) d\mu_n \right| \\ &\leq \varepsilon + 2\|f\|_{\infty} \mu_n(K^c) \\ &\leq \varepsilon + 2\|f\|_{\infty} \varepsilon \end{aligned}$$

et de même

$$\left| \int_{\mathbb{R}^{\mathbb{N}}} (\varphi - f) d\mu \right| \leq \varepsilon + 2\|f\|_{\infty} \varepsilon.$$

Finalement,

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{R}^{\mathbb{N}}} f d\mu_n = \int_{\mathbb{R}^{\mathbb{N}}} f d\mu. \quad \square$$

Enfin le résultat principal (Théorème de Prokhorov) qui prolonge les Propositions 6.3.5 et 7.1.7 et que nous admettrons dans le cas général (voir [6]).

7.3.8 Théorème. (Prokhorov) *Soit $(\mu_n)_{n \in \mathbb{N}}$ une suite de mesures tendues sur E . Alors, il existe une sous-suite qui converge étroitement.*

Néanmoins, on peut obtenir le cas particulier $E = \mathbb{R}^{\mathbb{N}}$ comme corollaire du cas $E = \mathbb{R}^n$.

PREUVE du Théorème 7.3.8 dans le cas $E = \mathbb{R}^{\mathbb{N}}$:

Soit donc $(\mu_n)_{n \in \mathbb{N}}$ une suite de mesures tendues sur $\mathbb{R}^{\mathbb{N}}$. Alors $(\pi_k \cdot \mu_n)$ est une suite tendue de mesures de probabilité sur \mathbb{R}^{k+1} (où comme d'habitude $\pi_k \cdot \mu_n(A) = \mu_n(\pi_k^{-1}(A))$, $A \in \mathcal{B}(\mathbb{R}^k)$). En effet, soit K un compact tel que $\mu_n(K) > 1 - \varepsilon$, $\forall n$. Notons $K' = \pi_k(K)$: c'est un compact de \mathbb{R}^{k+1} . Comme $\pi_k^{-1}(K') \supset K$, on a $\mu_n(\pi_k^{-1}(K')) > 1 - \varepsilon$, $\forall n$ i.e. $\pi_k \cdot \mu_n(K') > 1 - \varepsilon$, $\forall n$.

De $(\pi_0 \cdot \mu_n)_{n \geq 0}$ on peut extraire une sous-suite $(\pi_0 \cdot \mu_{k_n})_{n \geq 0}$ convergente vers $\tilde{\mu}_0$, mesure de probabilité sur \mathbb{R} .

De la suite $(\pi_1 \cdot \mu_{k_n})_{n \geq 1}$ on peut extraire une sous-suite $(\pi_1 \cdot \mu_{k_{n_1}})$ qui converge vers $\tilde{\mu}_1$, probabilité sur \mathbb{R}^2 etc... On construit ainsi une suite $(\mu_{\ell_n})_{n \geq 0}$ telle que $\forall k, \pi_k \mu_{\ell_n} \rightarrow \tilde{\mu}_k$, mesure de probabilité sur \mathbb{R}^{k+1} .

Par ailleurs, pour tout $k \geq 1$ fixé, notant $\pi : \mathbb{R}^{k+1} \rightarrow \mathbb{R}^k$ la projection excluant la dernière coordonnée, on a naturellement $\pi(\pi_{k+1} \mu_{\ell_n}) = \pi_k \mu_{\ell_n}$ d'où, passant à la limite par le théorème de transfert 7.1.4, $\pi \tilde{\mu}_{k+1} = \tilde{\mu}_k$ ce qui est la condition de compatibilité du Théorème (de Kolmogorov) 3.3.7. De ce fait, il existe une probabilité $\tilde{\mu}$ sur $\mathbb{R}^{\mathbb{N}}$ telle que $\forall k \geq 1, \pi_k \tilde{\mu} = \tilde{\mu}_k$. Par construction, pour tout $k \geq 1$, pour toute fonction continue bornée $f : \mathbb{R}^k \rightarrow \mathbb{R}$,

$$\int_{\mathbb{R}^{\mathbb{N}}} f(x_1, \dots, x_k) \mu_{\ell_n}(dx) \rightarrow \int_{\mathbb{R}^{\mathbb{N}}} f(x_1, \dots, x_k) \tilde{\mu}(dx).$$

La Proposition 7.3.6 permet alors de conclure que μ_{ℓ_n} converge vers $\tilde{\mu}$. □

Chapitre 8

Espérance conditionnelle

La notion d'espérance conditionnelle prolonge le formalisme précédemment introduit dans la section 2.4 en cherchant à intégrer la connaissance de la valeur d'une variable aléatoire, et plus seulement celle du fait qu'un événement particulier s'est réalisé.

8.1 Construction de l'espérance conditionnelle

Commençons par une situation simple mais exemplaire de ce qu'on cherche à faire. Soit X une variable aléatoire réelle intégrable et T une variable aléatoire à valeurs dans un ensemble dénombrable $E = \{t_0, t_1, \dots, t_n, \dots\}$. D'après la section 2.4, pour $k \in \mathbb{N}$, sur l'événement $(T = t_k)$ est définie une probabilité $Q_k = \mathbf{P}(\cdot / T = t_k)$ qui représente la mesure de probabilité \mathbf{P} tenant compte du fait que $(T = t_k)$ s'est réalisé. De ce fait, la valeur moyenne de X sachant que $T = t_k$ sera naturellement prise égale à la quantité $\mathbf{E}(X / T = t_k) = \int_{(T=t_k)} X(\omega) dQ_k(\omega)$, et on définira par suite l'espérance conditionnelle de X sachant T comme étant la *variable aléatoire* égale à

$$\mathbf{E}(X/T) = \sum_{k \geq 0} \mathbf{E}(X/T = t_k) \mathbb{I}_{(T=t_k)}.$$

Noter qu'avec cette définition $\mathbf{E}(X/T)$ apparaît comme une fonction de T , ou encore comme une variable $\sigma(T)$ -mesurable.

Pour chercher à prolonger ce résultat au cas de variables T à valeurs non dénombrables, on peut remarquer sur l'exemple précédent que pour n'importe quelle fonction bornée $f : E \rightarrow \mathbb{R}$, $\mathbf{E}(f(T)X) = \sum_{k \geq 0} f(t_k) \mathbf{E}(\mathbb{I}_{(T=t_k)} X)$. Or, comme pour n'importe quel $A \in \mathcal{F}$, $\mathbf{E}(\mathbb{I}_{(T=t_k)} \mathbb{I}_A) = \mathbf{P}((T = t_k) \cap A) = \mathbf{P}(T = t_k) Q_k(A) = \mathbf{E}(\mathbb{I}_{(T=t_k)} Q_k(A))$, on en déduit par les arguments habituels que $\mathbf{E}(\mathbb{I}_{(T=t_k)} X) = \mathbf{E}(\mathbb{I}_{(T=t_k)} \int_{\Omega} X dQ_k)$, et donc que

$$\mathbf{E}(f(T)X) = \mathbf{E}\left(\sum_{k \geq 0} f(t_k) \mathbb{I}_{(T=t_k)} \int_{\Omega} X dQ_k\right) = \mathbf{E}(f(T) \mathbf{E}(X/T)).$$

Ceci conduit naturellement à poser la définition suivante.

8.1.1 Théorème et Définition. Soient X une variable aléatoire réelle admettant un moment d'ordre 1 et T une variable aléatoire. On appelle **espérance conditionnelle** de X sachant T l'unique variable aléatoire Y mesurable par rapport à $\sigma(T)$ telle que pour toute variable bornée Z mesurable par rapport à $\sigma(T)$, on ait

$$\mathbf{E}(X.Z) = \mathbf{E}(Y.Z). \tag{8.1.1}$$

Cette variable est notée $\mathbf{E}(X/T)$ ou $\mathbf{E}^{\sigma(T)}(X)$.

PREUVE :

1. *Unicité.*

Supposons que Y et Y' satisfassent la Définition 8.1.1. La variable aléatoire

$Z = \frac{Y - Y'}{|Y - Y'|} \mathbb{1}_{Y \neq Y'}$ est $\sigma(T)$ -mesurable et bornée. De ce fait, en appliquant (8.1.1), $0 = \mathbf{E}((Y - Y')Z) = \mathbf{E}(|Y - Y'| \mathbb{1}_{Y \neq Y'})$ d'où $Y = Y'$ -p.s.

2. *Existence.*

Commençons par supposer que $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$. Posons $\mathcal{H} = L^2(\Omega, \sigma(T), \mathbf{P})$, il s'agit d'un sous-espace vectoriel fermé de l'espace de Hilbert $L^2(\Omega, \mathcal{F}, \mathbf{P})$, et de ce fait, on peut définir Y la projection orthogonale de X sur \mathcal{H} . C'est une variable $\sigma(T)$ -mesurable et elle satisfait, pour toute variable Z , $\sigma(T)$ -mesurable et bornée (et donc dans \mathcal{H}), $\mathbf{E}(XZ) = \mathbf{E}(YZ)$ soit (8.1.1). D'où l'existence cherchée dans ce cas. Noter que par propriété de la projection orthogonale, l'espérance conditionnelle sur L^2 est ainsi linéaire.

En prenant $Z = 1$ dans (8.1.1), on montre que $\mathbf{E}(\mathbf{E}(X/T)) = \mathbf{E}(X)$. De plus, si $X \geq 0$, $\mathbf{E}(X/T) \geq 0$ (prendre $Z = \mathbb{1}_{\mathbf{E}(X/T) \leq 0}$ dans (8.1.1)). De ce fait, pour $X \in L^2$, on a $\mathbf{E}(|X|/T) \geq |\mathbf{E}(X/T)|$ et donc, prenant l'espérance, $\|\mathbf{E}(X/T)\|_1 \leq \|X\|_1$.

Achevons la preuve du Théorème 8.1.1. Supposons que X soit dans L^1 . On peut trouver une suite X_n de variables de L^2 telle que $\|X_n - X\|_1 \rightarrow 0$. On a par conséquent $\|\mathbf{E}(X_n/T) - \mathbf{E}(X_p/T)\|_1 \leq \|X_n - X_p\|_1$ et donc $\mathbf{E}(X_n/T)$ est une suite de Cauchy dans $L^1(\Omega, \sigma(T), \mathbf{P})$ qui converge vers une variable $\sigma(T)$ -mesurable notée $\mathbf{E}(X/T)$. Pour tout Z , $\sigma(T)$ -mesurable et borné, on a $\mathbf{E}(X_n.Z) = \mathbf{E}(\mathbf{E}(X_n/T).Z)$ et donc par passage à la limite L^1 , $\mathbf{E}(X.Z) = \mathbf{E}(\mathbf{E}(X/T).Z)$. \square

8.2 Propriétés de l'espérance conditionnelle

Regardons d'abord les principales propriétés de l'espérance conditionnelle.

8.2.1 Proposition. *Soit T une variable aléatoire et soit X une v.a.r. intégrable. On a :*

- (i) $\mathbf{E}(\cdot/T)$ est un opérateur linéaire positif continu sur L^1
- (ii) Si X est $\sigma(T)$ -mesurable, $\mathbf{E}(X/T) = X$, p.s.
- (iii) Si φ est convexe sur \mathbb{R} , $\varphi(\mathbf{E}(X/T)) \leq \mathbf{E}(\varphi(X)/T)$ p.s.
- (iv) $\|\mathbf{E}(X/T)\|_1 \leq \|X\|_1$
- (v) Si $X_n \rightarrow X$ p.s., et $\forall n \geq 0, |X_n| \leq Y \in L^1$, $\mathbf{E}(X_n/T) \rightarrow \mathbf{E}(X/T)$ dans L^1 .

PREUVE :

Les propriétés (i),(ii) et (iv) se prouvent aisément à partir de la preuve du théorème précédent .

On pourrait prouver (iii) en refaisant une démonstration du style de celle de l'inégalité de Jensen. On peut aussi utiliser ce qui sera démontré au chapitre suivant : l'espérance conditionnelle $\mathbf{E}(\varphi(X)/T)$ est obtenue comme intégrale de φ par rapport à la loi conditionnelle. La propriété (iii) est alors une conséquence de l'inégalité de Jensen (c'est-à-dire le Théorème 1.6).

Enfin, supposons satisfaites les hypothèses de (v). Notons que par convergence dominée, $X_n \rightarrow X$ dans L^1 . Or, par (iv),

$$\|\mathbf{E}(X_n/T) - \mathbf{E}(X_p/T)\|_1 \leq \mathbf{E}(|X_n - X_p|) = \|X_n - X_p\|_1.$$

De ce fait, $(\mathbf{E}(X_n/T))_{n \geq 1}$ est une suite de Cauchy dans L^1 qui converge vers une v.a. Z .

On a, pour toute fonction f mesurable bornée $\mathbf{E}(f(T)X_n) = \mathbf{E}(f(T)\mathbf{E}(X_n/T))$. On a $f(T)X_n \rightarrow f(T)X$ p.s. et de plus $|f(T)X_n| \leq \|f\|_\infty |Y|$. Par le théorème de convergence dominée, on a donc $\mathbf{E}(f(T)X_n) \rightarrow \mathbf{E}(f(T)X)$.

Enfin, $\|f(T)\mathbf{E}(X_n/T) - f(T)Z\|_1 \leq \|f\|_\infty \|\mathbf{E}(X_n/T) - Z\|_1 \rightarrow 0$. On déduit de ce qui précède que $\mathbf{E}(f(T)Z) = \mathbf{E}(f(T)X)$ et donc $Z = \mathbf{E}(X/T)$. \square

Il sera souvent utile de considérer l'information apportée dans le conditionnement sous la forme d'une sous-tribu \mathcal{B} de \mathcal{F} , le cas précédent correspondant à $\mathcal{B} = \sigma(T)$. La Définition 8.1.1 s'étend alors naturellement :

8.2.2 Définition. Soient X une variable aléatoire réelle intégrable, et \mathcal{B} une sous-tribu de \mathcal{F} . On appelle espérance conditionnelle de X sachant \mathcal{B} l'unique variable aléatoire Y, \mathcal{B} -mesurable, telle que pour tout Z, \mathcal{B} -mesurable bornée, $\mathbf{E}(X.Z) = \mathbf{E}(Y.Z)$.

Une propriété utile qui se déduit de la définition précédente est l'emboîtement des conditionnements successifs.

8.2.3 Proposition. Soient X une variable aléatoire réelle intégrable et $\mathcal{B}' \subset \mathcal{B}$ deux sous-tribus de \mathcal{F} . Alors $\mathbf{E}(X/\mathcal{B}') = \mathbf{E}[\mathbf{E}(X/\mathcal{B})/\mathcal{B}']$.

PREUVE : On remarque que $\mathbf{E}(\mathbf{E}(X/\mathcal{B})/\mathcal{B}')$ est une variable aléatoire \mathcal{B}' -mesurable et que, pour toute variable aléatoire Y, \mathcal{B}' -mesurable bornée, on a

$$\mathbf{E}(Y.\mathbf{E}[\mathbf{E}(X/\mathcal{B})/\mathcal{B}']) = \mathbf{E}(Y.\mathbf{E}(X/\mathcal{B})) = \mathbf{E}(Y.X),$$

la dernière égalité résultant du fait que Y est aussi \mathcal{B} -mesurable. \square

Notons enfin, pour clore ce paragraphe sur l'espérance conditionnelle, que si X et T sont indépendantes, la relation (8.1.1) est évidemment vérifiée en prenant Y égale à la constante $\mathbf{E}(X)$, et donc, vu l'unicité de l'espérance conditionnelle, on peut énoncer le corollaire suivant :

8.2.4 Corollaire. Soient X une v.a.r. intégrable, et T une variable aléatoire telles que X et T soient indépendantes. Alors, $\mathbf{E}(X/T) = \mathbf{E}(X)$ p.s.

◇ Exercice 8.1

Soit $(\mathcal{B}_i)_{i \in I}$ la famille de toutes les sous-tribus de \mathcal{F} et X une v.a. intégrable. Montrer que la famille $(\mathbf{E}^{\mathcal{B}_i}(X))_{i \in I}$ est équi-intégrable.

Exercice 8.2

Soit X une variable aléatoire réelle de carré intégrable. On pose $\text{Var}(X/\mathcal{G}) = \mathbf{E}[(X - \mathbf{E}(X/\mathcal{G}))^2/\mathcal{G}]$. Montrer que

$$\text{Var}(X) = \mathbf{E}[\text{Var}(X/\mathcal{G})] + \text{Var}(\mathbf{E}(X/\mathcal{G})).$$

Exercice 8.3

a) 1) Soit (X, Y, Z) tel que (X, Z) a même loi que (Y, Z) . Montrer que pour toute $f \geq 0$ et mesurable, $\mathbf{E}(f(X)/Z) = \mathbf{E}(f(Y)/Z)$.

2) On pose $h_1(X) = \mathbf{E}(g(Z)/X)$ et $h_2(Y) = \mathbf{E}(g(Z)/Y)$ pour g mesurable positive donnée. Montrer que $h_1 = h_2, \mu$ -pp, où μ désigne la loi de X .

b) Soient T_1, \dots, T_n des variables aléatoires réelles intégrables indépendantes et de même loi.

On pose $T = T_1 + \dots + T_n$.

b1) Montrer que $\mathbf{E}(T_1/T) = T/n$.

b2) Montrer que $\mathbf{E}(T/T_1) = T_1 + (n-1)\mathbf{E}(T_1)$.

Exercice 8.4

Soit (X_n) une suite de variables aléatoires réelles indépendantes à valeurs dans \mathbb{R}^d . On pose $S_0 = 0, S_n = X_1 + \dots + X_n, \mathcal{F}_n = \sigma(S_k, k \leq n)$. Montrer que pour toute $f : \mathbb{R}^d \rightarrow \mathbb{R}$ borélienne bornée,

$$\mathbf{E}(f(S_n)/\mathcal{F}_{n-1}) = \mathbf{E}(f(S_n)/S_{n-1})$$

et calculer cette quantité.

◇ Exercice 8.5

Soient $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ trois sous-tribus de \mathcal{F} . On dit que \mathcal{F}_1 et \mathcal{F}_2 sont conditionnellement indépendantes sachant \mathcal{F}_3 si $\forall (A, B) \in \mathcal{F}_1 \times \mathcal{F}_2$,

$$\mathbf{E}(\mathbb{I}_A \mathbb{I}_B / \mathcal{F}_3) = \mathbf{E}(\mathbb{I}_A / \mathcal{F}_3) \mathbf{E}(\mathbb{I}_B / \mathcal{F}_3).$$

Montrer que cette condition équivaut à : pour toute variable Y, \mathcal{F}_2 -mesurable bornée, on a $\mathbf{E}(Y/\mathcal{F}_3) = \mathbf{E}(Y/\mathcal{F}_{13})$, où $\mathcal{F}_{13} = \sigma(\mathcal{F}_1, \mathcal{F}_3)$.

Chapitre 9

Lois conditionnelles

De même que la loi d'une variable aléatoire X contient l'information probabiliste requise pour l'utilisation de cette variable, on a naturellement besoin d'une notion correspondante pour l'information qu'apporte cette variable lorsqu'on connaît en plus une autre variable aléatoire T . C'est le rôle que va jouer la loi conditionnelle.

9.1 Lois conditionnelles

Soient X une variable aléatoire à valeurs dans (E, \mathcal{E}) et T une variable aléatoire. Le problème qu'on se pose est le suivant : construire une mesure ν_T sur (E, \mathcal{E}) telle que pour toute fonction $f : E \rightarrow \mathbb{R}$ mesurable bornée, on ait $\mathbf{E}(f(X)/T) = \int_E f(x) \nu_T(dx)$. Naturellement, une telle relation va imposer à ν_T d'être une mesure aléatoire, et même plus précisément que l'aléa se lise à travers T (puisque par définition, $\mathbf{E}(f(X)/T)$ est une variable $\sigma(T)$ -mesurable).

L'idée qui vient naturellement est de définir la mesure ν_T par $\nu_T(A) = \mathbf{E}(\mathbb{I}_A(X)/T)$. Le problème est que la relation précédente, une égalité entre variables aléatoires, n'a de sens que p.s. et donc, sauf dans le cas particulier où X prend un nombre dénombrable de valeurs, il n'est pas du tout évident que $\nu_T(\omega)$ ainsi définie soit une probabilité sur (E, \mathcal{E}) . De ce fait, on a en fait besoin de trouver ce qu'on nomme une *probabilité conditionnelle régulière* qui permette de définir une véritable mesure aléatoire.

La définition formelle suivante fixe le cadre.

9.1.1 Définition. Soit (E, \mathcal{E}) un espace mesurable et T une variable aléatoire à valeurs dans (F, \mathcal{F}) . On appelle *T-probabilité de transition sur E* une application ν définie sur $\Omega \times \mathcal{E}$ et à valeurs dans $\rightarrow [0, +\infty]$ telle que :

- (i) Pour tout $A \in \mathcal{E}$, $\omega \mapsto \nu(\omega, A)$ est une variable aléatoire $\sigma(T)$ -mesurable ;
- (ii) P-p.s., $\nu(\omega, dx)$ est une mesure de probabilité sur (E, \mathcal{E})

Pour une variable aléatoire X à valeurs dans un espace mesurable quelconque, on ne peut pas trouver en général de loi conditionnelle régulière. Le résultat suivant est de ce fait très important.

9.1.2 Théorème et Définition. Soient X une variable aléatoire à valeurs dans (E, \mathcal{E}) un espace métrique complet séparable muni de sa tribu borélienne, et T une variable aléatoire. Il existe une *T-probabilité de transition* ν_T sur E telle que pour toute $f : E \rightarrow \mathbb{R}$ mesurable bornée,

$$\mathbf{E}(f(X)/T)(\omega) = \int_E f(x) \nu_T(\omega, dx) \text{ p.s..}$$

On l'appelle **loi conditionnelle régulière de X sachant T**.

PREUVE : Nous nous contenterons de faire la démonstration pour $E = \mathbb{R}$. Le cas général s'obtient en étendant le résultat à $\overline{\mathbb{R}}$ puis en utilisant le fait qu'un espace polonais est un espace mesurable isomorphe à une partie borélienne de $[0,1]$. Le lecteur que tout cela excite pourra trouver des détails dans [12].

Soit donc $E = \mathbb{R}$ et X une variable aléatoire réelle. Pour $r \in \mathbb{Q}$, définissons

$$F(r, \omega) = E(\mathbb{I}_{X < r} / T)(\omega) \text{ p.s.}$$

(i) Soient $r < r'$ des rationnels. On a p.s. $F(r, \omega) \leq F(r', \omega)$. Soit $N_{r,r'}$ l'ensemble négligeable sur lequel l'inégalité n'est pas vérifiée. Posons $N = \bigcup_{(r,r') \in \mathbb{Q}^2} N_{r,r'}$. On a $P(N) = 0$,

et pour $\omega \in N^c, \forall r \leq r', F(r, \omega) \leq F(r', \omega)$.

(ii) Pour $r \in \mathbb{Q}$ fixé, $F(r - \frac{1}{k}, \omega) \uparrow F(r, \omega)$ p.s. puisque $\mathbb{I}_{X < r - \frac{1}{k}} \uparrow \mathbb{I}_{X < r}$ p.s.

Soit N'_r l'ensemble négligeable sur lequel on n'a pas convergence. Posons $N' = \bigcup_{r \in \mathbb{Q}} N'_r$.

On a $P(N') = 0$ et $\forall \omega \in N'^c, F(r - \frac{1}{k}, \omega) \uparrow F(r, \omega), \forall r \in \mathbb{Q}$ donc $F(., \omega)$ est continue à gauche sur \mathbb{Q} .

(iii) On a $F(r, \omega) \rightarrow 1$ p.s. quand $r \rightarrow +\infty$ puisque $\mathbb{I}_{X < r} \rightarrow 1$ p.s. quand $r \rightarrow +\infty$. Soit \tilde{N} l'ensemble négligeable où la convergence n'est pas vraie. Pour $\omega \in \tilde{N}^c, F(r, \omega) \rightarrow 1$ quand $r \rightarrow +\infty$.

(iv) De même on a un \tilde{N}' négligeable tel que si $\omega \in \tilde{N}'^c, F(r, \omega) \rightarrow 0$ quand $r \rightarrow -\infty$.

Soit alors $\underline{N} = N \cup N' \cup \tilde{N} \cup \tilde{N}'$. On a $P(\underline{N}) = 0$ et pour $\omega \in \underline{N}^c$, posons pour $x \in \mathbb{R}$, $\tilde{F}(x, \omega) = \sup_{r < x} F(r, \omega)$. Pour $\omega \in \underline{N}^c, \tilde{F}(., \omega)$ est croissante, continue à gauche et on a

$$\lim_{x \rightarrow +\infty} \tilde{F}(x, \omega) = 1 \text{ et } \lim_{x \rightarrow -\infty} \tilde{F}(x, \omega) = 0.$$

Pour $\omega \in \underline{N}^c, \tilde{F}(., \omega)$ est donc la fonction de répartition continue à gauche d'une mesure de probabilité sur \mathbb{R} , que nous notons $\nu_T(dx, \omega)$, soit la condition (ii) de la Définition 9.1.1.

Montrons que la condition (i) est également satisfaite. Soit

$$\mathcal{L} = \{A \in \mathcal{E} / \nu_T(A, \omega) \text{ est } \sigma(T) \text{ - mesurable et p.s. } \nu_T(A) = E(\mathbb{I}_A(X) / T)\}.$$

\mathcal{L} est clairement un λ -système (voir l'exercice 2.1) qui contient par construction tous les intervalles $] -\infty, r], r \in \mathbb{Q}$ qui forment clairement un π -système. Par le $\pi\lambda$ théorème, \mathcal{L} contient la tribu borélienne $\mathcal{B}(\mathbb{R})$.

L'égalité p.s. $E(f(X) / T) = \int_E f(x) \nu_T(dx)$ pour f mesurable bornée est alors obtenue par un raisonnement classique. \square

La démonstration précédente dans le cas $E = \mathbb{R}$ permet d'obtenir une preuve du Théorème de Kolmogorov 3.3.7 dans le cas $E = \mathbb{R}^{\mathbb{N}}$.

9.1.3 Corollaire. (Théorème 3.3.7) *Soit $(\mu_n)_{n \geq 0}$ une famille de probabilités sur les espaces produit $(\mathbb{R}^{n+1}, \mathcal{B}(\mathbb{R}^{n+1}))$ qui satisfait la condition de compatibilité*

$$\mu_n(dx_0, \dots, dx_n) = \int_{\mathbb{R}} \mu_{n+1}(dx_0, \dots, dx_n, dx_{n+1}).$$

Alors il existe une unique probabilité \mathbf{P} sur l'espace canonique $\mathbb{R}^{\mathbb{N}}$ telle que sous \mathbf{P} le processus canonique \tilde{X} admette les lois μ_n comme répartitions finies.

PREUVE : Il suffit de démontrer qu'on peut trouver un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ et une suite de variables aléatoires réelles $(X_n)_{n \geq 0}$ telle que pour chaque $n \geq 0$, la loi de (X_0, \dots, X_n) est donnée par μ_n . La probabilité cherchée est alors la loi du processus.

D'après l'Exercice 3.6, sur l'espace $([0, 1], \mathcal{B}([0, 1]), \mathbf{P})$ où \mathbf{P} est la mesure de Lebesgue, il existe une suite de variables aléatoires indépendantes $(U_n)_{n \geq 0}$ de loi uniforme sur $[0, 1]$. Supposons que (X_0, \dots, X_{n-1}) soit construit tel que (X_0, \dots, X_k) soit $\sigma(U_0, \dots, U_k)$ -mesurable et suive la loi μ_k pour tout $k \leq n-1$, et construisons X_n . L'idée va être de se débrouiller pour que cette variable admette la bonne loi conditionnelle sachant (X_0, \dots, X_{n-1}) . Considérons l'espace $E = \mathbb{R}^{n+1}$ muni de la probabilité $Q = \mu_n$, et (Y_0, \dots, Y_n) le $n+1$ -uplet des projections canoniques. Le vecteur aléatoire (Y_0, \dots, Y_n) admet donc la loi μ_n . Notons que par la condition de compatibilité, (Y_0, \dots, Y_{n-1}) admet la loi μ_{n-1} .

D'après le Théorème 9.1.2, il existe une loi conditionnelle régulière de Y_n sachant (Y_0, \dots, Y_{n-1}) , dont on note $F_n(t; Y_0, \dots, Y_{n-1})$ la fonction de répartition.

Pour (y_0, \dots, y_{n-1}) donné, soit

$$G_n(y; y_0, \dots, y_{n-1}) = \inf\{t \in [0, 1], F_n(t; y_0, \dots, y_{n-1}) \geq y\}$$

l'inverse généralisée de F_n . On sait (cf. la démonstration du Théorème 6.3.11), qu'alors $G_n(U_n; y_0, \dots, y_{n-1})$ suit la loi $F_n(dz; y_0, \dots, y_{n-1})$. Posons $X_n = G_n(U_n; X_0, \dots, X_{n-1})$: c'est bien une variable aléatoire grâce à la mesurabilité de $(t, y_0, \dots, y_{n-1}) \mapsto F_n(t, y_0, \dots, y_{n-1})$ qui résulte du choix d'une version régulière. Il est en outre clair que X_n est $\sigma(U_0, \dots, U_n)$ -mesurable.

Examinons maintenant la loi du vecteur (X_0, \dots, X_n) . Soit $\varphi : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ mesurable bornée.

Comme U_n et (X_0, \dots, X_{n-1}) sont indépendants, on a

$$\begin{aligned} \mathbf{E}(\varphi(X_0, \dots, X_n)) &= \mathbf{E}(\varphi(X_0, \dots, X_{n-1}, G_n(U_n; X_0, \dots, X_{n-1}))) \\ &= \mathbf{E}\left(\int_{\mathbb{R}} \varphi(X_0, \dots, X_{n-1}, z) F_n(dz; X_0, \dots, X_{n-1})\right) \\ &= \mathbf{E}_Q\left(\int_{\mathbb{R}} \varphi(Y_0, \dots, Y_{n-1}, z) F_n(dz; Y_0, \dots, Y_{n-1})\right) \\ &= \mathbf{E}_Q(\mathbf{E}_Q(\varphi(Y_0, \dots, Y_{n-1}, Y_n) / Y_0, \dots, Y_{n-1})) \\ &= \mathbf{E}_Q(\varphi(Y_0, \dots, Y_n)) = \int_{\mathbb{R}^{n+1}} \varphi(y_0, \dots, y_n) \mu_n(dy_0, \dots, dy_n) \end{aligned}$$

donc (X_0, \dots, X_n) suit la loi μ_n □

9.2 Cas particuliers

L'exemple le plus important où la définition 9.1.2 s'applique est celui où le vecteur (X, T) est à valeurs dans \mathbb{R}^{n+d} et y admet une densité $\lambda(x_1, \dots, x_n; t_1, \dots, t_d)$. La loi conditionnelle se calcule alors très aisément :

9.2.1 Proposition. Soit (X, T) un vecteur aléatoire à valeurs dans \mathbb{R}^{n+d} dont la loi admet une densité $\lambda(x_1, \dots, x_n; t_1, \dots, t_d)$. Alors la loi conditionnelle de X sachant T est p.s. la

loi sur \mathbb{R}^n admettant la densité $\alpha_T(x_1, \dots, x_n) = \frac{\lambda(x_1, \dots, x_n; T)}{\int_{\mathbb{R}^n} \lambda(u_1, \dots, u_n; T) du_1 \dots du_n}$.

PREUVE : Naturellement, la loi aléatoire à densité précédente est bien une T -probabilité de transition sur \mathbb{R}^n .

Considérons $\rho(t) = \int_{\mathbb{R}^n} \lambda(u_1, \dots, u_n; t) du_1 \dots du_n$ la densité de la loi de T .

Pour f (resp. g) boréliennes bornées sur \mathbb{R}^n (resp. \mathbb{R}^d), on peut écrire grâce au théorème de Fubini qui s'applique puisque f et g sont bornées,

$$\begin{aligned} \mathbf{E}(f(X)g(T)) &= \int_{\mathbb{R}^n \times \mathbb{R}^d} f(x)g(t)\lambda(x,t) dx dt \\ &= \int_{\mathbb{R}^d} g(t) \left[\int_{\mathbb{R}^n} f(x) \frac{\lambda(x,t)}{\rho(t)} dx \right] \rho(t) dt = \int_{\mathbb{R}^d} g(t) \left[\int_{\mathbb{R}^n} f(x) \alpha_t(x) dx \right] \rho(t) dt, \end{aligned}$$

où $\rho(t) = \int_{\mathbb{R}^n} \lambda(u_1, \dots, u_n; t) du_1 \dots du_n$ est la densité de la loi de T . De ce fait, on a $\mathbf{E}(f(X)g(T)) = \mathbf{E}(g(T) \left[\int_{\mathbb{R}^n} f(x) \alpha_T(x) dx \right])$ et p.s. $\mathbf{E}(f(X)/T) = \int_{\mathbb{R}^n} f(x) \alpha_T(x) dx$. \square

Exercice 9.1

Soient T_1, T_2, \dots des variables aléatoires de loi exponentielle de paramètre λ indépendantes. On pose $T = T_1 + \dots + T_n$.

a) Déterminer la loi conditionnelle de T_1 sachant T et calculer $\mathbf{E}(T_1/T)$ (voir l'Exercice 8.8.3).

b) Calculer $\mathbf{E}(T_1^2/T)$ et $\mathbf{E}(T_1 T_2/T)$.

◇ Exercice 9.2

Soient X et Y deux variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. Déterminer la loi conditionnelle de $(X - Y)^+$ sachant Y .

Exercice 9.3

Soit X à valeurs dans \mathbb{R}^m tel que $X = \varphi(Y) + Z$ où Y et Z sont indépendantes. Montrer que la loi conditionnelle de X sachant Y est $\mu_{\varphi(Y)}(dz)$ où μ_a est la mesure sur \mathbb{R}^m définie par

$$\int_{\mathbb{R}^m} f(z) \mu_a(dz) = \int_{\mathbb{R}^m} f(a+z) \mu_Z(dz).$$

Chapitre 10

Vecteurs gaussiens

Dans ce chapitre, nous allons étudier plus en détail le cas particulier des vecteurs gaussiens. Son importance provient du fait que l'indépendance des composantes d'un vecteur gaussien s'exprime de façon très simple.

10.1 Distributions normales dans \mathbb{R}^k

Rappelons qu'une variable aléatoire réelle X suit une loi normale centrée réduite si sa densité est donnée par

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

On a alors

10.1.1 Définition. On appelle **variable gaussienne** toute variable aléatoire Y telle qu'il existe une variable X normale centrée réduite et deux réels m et $\sigma \geq 0$, tels que

$$Y = \sigma X + m.$$

On notera la loi d'une telle variable $\mathcal{N}(m, \sigma^2)$.

On peut noter en particulier qu'une variable constante c est gaussienne, de loi $\mathcal{N}(c, 0)$.

10.1.2 Proposition. La fonction caractéristique de Y de loi $\mathcal{N}(m, \sigma^2)$ est donnée par

$$\varphi_Y(t) = \exp\left(-\frac{1}{2} \frac{\sigma^2 t^2}{2} + itm\right).$$

PREUVE : On sait que la fonction caractéristique de X de loi $\mathcal{N}(0, 1)$ est $\varphi_X(t) = e^{-\frac{t^2}{2}}$. Soit Y de loi $\mathcal{N}(m, \sigma^2)$. Par définition, il existe une variable X normale centrée réduite telle que $Y = \sigma X + m$. On a alors

$$\varphi_Y(t) = \mathbf{E}(e^{itY}) = \mathbf{E}(e^{i(t\sigma)X} e^{itm}) = e^{itm} \varphi_X(t\sigma) = \exp\left(-\frac{\sigma^2 t^2}{2} + itm\right). \quad \square$$

On pose alors la définition suivante en dimension k .

10.1.3 Définition. On dit qu'un vecteur aléatoire $X = \begin{pmatrix} X_1 \\ \dots \\ X_k \end{pmatrix}$ est gaussien si, quelle que soit la matrice $A \in \mathcal{M}_{\mathbb{R}}(1, k)$, la variable aléatoire réelle $Y = AX$ est gaussienne.

La définition précédente permet alors d'obtenir immédiatement la stabilité du caractère gaussien par transformation linéaire.

10.1.4 Propriété. Si $X = \begin{pmatrix} X_1 \\ \dots \\ X_k \end{pmatrix}$ est gaussien, et $A \in \mathcal{M}_{\mathbb{R}}(n, k)$, alors AX est un vecteur gaussien de \mathbb{R}^n .

Soit $X = \begin{pmatrix} X_1 \\ \dots \\ X_k \end{pmatrix}$ un vecteur gaussien. On définit comme d'habitude le vecteur espérance $\mathbf{E}(X) = \begin{pmatrix} \mathbf{E}(X_1) \\ \mathbf{E}(X_2) \\ \dots \\ \mathbf{E}(X_k) \end{pmatrix}$ et la matrice de dispersion (ou de covariance) par

$$D_X = (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq k}.$$

Il est pratique d'écrire que $D_X = \mathbf{E}(XX')$ (où X' désigne le transposé de X , qui est donc un vecteur ligne). Il est ainsi clair que D_X est une matrice symétrique positive.

On a alors immédiatement

10.1.5 Propriété. Soit X un vecteur gaussien de \mathbb{R}^k . Soient $A \in \mathcal{M}_{\mathbb{R}}(n, k)$ et $Y = AX$. Alors

$$\mathbf{E}(Y) = A\mathbf{E}(X), D_Y = AD_X A'.$$

Nous allons voir que de la même façon qu'en dimension 1, la loi d'une variable gaussienne est parfaitement déterminée par la donnée de l'espérance et de la variance, en dimension k elle est caractérisée par celle du vecteur espérance et de la matrice de dispersion.

Supposons d'abord que $\mathbf{E}(X) = 0$. La fonction caractéristique de X est donnée par $\varphi_X(\mathbf{t}) =$

$\mathbf{E}(e^{i\langle \mathbf{t}, X \rangle})$ où $\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ \dots \\ t_k \end{pmatrix}$. Fixons $\mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \\ \dots \\ t_k \end{pmatrix}$. $\langle \mathbf{t}, X \rangle = t_1 X_1 + \dots + t_n X_n$ est une variable

réelle gaussienne de loi $\mathcal{N}(0, \sigma^2)$ où

$$\sigma^2 = \text{Var}(t_1 X_1 + \dots + t_n X_n) = \mathbf{E}((t_1 X_1 + \dots + t_n X_n)^2).$$

On vérifie immédiatement $\sigma^2 = \mathbf{t}' D_X \mathbf{t}$. Par conséquent, $\varphi_X(\mathbf{t}) = \exp(-\frac{\mathbf{t}' D_X \mathbf{t}}{2})$. Si $\mathbf{E}(X)$ est quelconque, on se ramène au cas précédent en posant $Y = X - \mathbf{E}(X)$ et on obtient $\varphi_Y(\mathbf{t}) = \exp(-\frac{\mathbf{t}' D_X \mathbf{t}}{2} + i\langle \mathbf{t}, \mathbf{E}(X) \rangle)$. En résumé,

10.1.6 Proposition. La loi d'un vecteur gaussien X de \mathbb{R}^k est entièrement déterminée par son espérance $\mathbf{E}(X)$ et sa matrice de dispersion D_X . Sa fonction caractéristique est donnée par

$$\varphi_X(\mathbf{t}) = \exp(-\frac{1}{2} \mathbf{t}' D_X \mathbf{t} + i\langle \mathbf{t}, \mathbf{E}(X) \rangle).$$

La loi d'un vecteur gaussien de \mathbb{R}^k d'espérance $\mathbf{E}(X)$ et de dispersion D_X est notée $\mathcal{N}_k(\mathbf{E}(X), D_X)$.

On déduit de la forme précédente de la fonction caractéristique la propriété fondamentale suivante concernant l'indépendance.

10.1.7 Proposition. Soit $X = \begin{pmatrix} X_1 \\ \dots \\ X_k \end{pmatrix}$ un vecteur gaussien de \mathbb{R}^k . Les variables aléatoires réelles X_1, X_2, \dots, X_k sont indépendantes si et seulement si D_X est une matrice diagonale.

PREUVE : Si X_1, X_2, \dots, X_k sont indépendantes, $\text{Cov}(X_i, X_j) = 0$ si $i \neq j$ et donc D_X est diagonale.

Réciproquement, supposons que D_X soit diagonale de coefficients $\sigma_1^2, \dots, \sigma_k^2$. Alors,

$$\begin{aligned} \mathbf{E}(e^{i(t_1 X_1 + \dots + t_k X_k)}) &= \varphi_X(\mathbf{t}) \\ &= \exp \frac{1}{2} (-\sum_{k=1}^n t_k^2 \sigma_k^2 + i \sum_{k=1}^n t_k \mathbf{E}(X_k)) \\ &= \prod_{k=1}^n \exp(-\frac{t_k^2 \sigma_k^2}{2} + i t_k \mathbf{E}(X_k)) = \prod_{k=1}^n \varphi_{X_k}(t_k). \end{aligned}$$

Ceci prouve l'indépendance cherchée. \square

La proposition précédente montre que dans le cadre des vecteurs gaussiens centrés, l'indépendance stochastique des composantes est équivalente à l'orthogonalité dans L^2 .

Une autre propriété fondamentale des vecteurs gaussiens est leur stabilité par convergence en loi (et donc *a fortiori* pour les autres types de convergence). Nous aurons besoin d'abord d'un petit résultat d'algèbre linéaire.

10.1.8 Lemme. Soit $(A_n)_{n \in \mathbb{N}}$ une suite de matrices symétriques de $\mathcal{M}_k(\mathbb{R})$ telle que pour tout $\mathbf{t} \in \mathbb{R}^k$, la suite $(\mathbf{t}' A_n \mathbf{t})_{n \in \mathbb{N}}$ converge dans \mathbb{R} . Alors $(A_n)_{n \in \mathbb{N}}$ converge dans $\mathcal{M}_k(\mathbb{R})$.

PREUVE : si $(\mathbf{s}, \mathbf{t}) \in \mathbb{R}^k \times \mathbb{R}^k$, comme A_n est symétrique, on a

$$\mathbf{s}' A_n \mathbf{t} = \frac{1}{4} [(\mathbf{s} + \mathbf{t})' A_n (\mathbf{s} + \mathbf{t}) - (\mathbf{s} - \mathbf{t})' A_n (\mathbf{s} - \mathbf{t})].$$

De ce fait, $(\mathbf{s}' A_n \mathbf{t})_{n \in \mathbb{N}}$ est une suite convergente dans \mathbb{R} . Soient $1 \leq i \leq k, 1 \leq j \leq k$. Notant $a_{i,j}^n$ le terme de la i -ième ligne et j -ième colonne de A_n , on a

$$a_{i,j}^n = \mathbf{t}_i' A_n \mathbf{t}_j$$

où $\mathbf{t}_r = \begin{pmatrix} 0 \\ \dots \\ 0 \\ 1 \\ \dots \\ 0 \end{pmatrix}$ avec le 1 à la r -ième place. De ce fait, $(a_{i,j}^n)_{n \in \mathbb{N}}$ converge. \square

10.1.9 Théorème. Soit $(X_n)_{n \geq 0}$ une suite de vecteurs aléatoires gaussiens de \mathbb{R}^k . On note $M_n = \mathbf{E}(X_n)$ et D_n la matrice de dispersion de X_n , et on suppose que $X_n \xrightarrow{\mathcal{L}} X$ lorsque $n \rightarrow \infty$. Alors X est un vecteur gaussien.

PREUVE : Puisque pour tout $t \in \mathbb{R}^k$,

$$\varphi_{X_n}(t) = e^{-\frac{1}{2} t' D_n t + i t' M_n}$$

il est clair qu'il suffit de prouver que $(D_n)_{n \in \mathbb{N}}$ et $(M_n)_{n \in \mathbb{N}}$ sont des suites convergentes dans $\mathcal{M}_k(\mathbb{R})$ et \mathbb{R}^k respectivement.

Par hypothèse, $\forall \mathbf{t} \in \mathbb{R}^k$, $\varphi_{X_n}(\mathbf{t}) \rightarrow \varphi(\mathbf{t})$. Alors

$$|\varphi_{X_n}(\mathbf{t})| = e^{-\frac{1}{2}\mathbf{t}'D_n\mathbf{t}} \rightarrow |\varphi(\mathbf{t})|.$$

Notons que $|\varphi(\mathbf{t})| \neq 0$. En effet, puisque $\varphi(0) = 1$ et φ est continue, on peut trouver $\alpha > 0$ tel que si $\|\mathbf{t}\| \leq \alpha$, $|\varphi(\mathbf{t})| \neq 0$. Soit alors $\mathbf{t} \in \mathbb{R}^k$, $\mathbf{t} \neq 0$. On a

$$e^{-\frac{1}{2}\mathbf{t}'D_n\mathbf{t}} = \left(e^{-\frac{1}{2} \frac{\alpha'}{\|\mathbf{t}\|} D_n \frac{\alpha}{\|\mathbf{t}\|}} \right) \frac{\|\mathbf{t}\|^2}{\alpha^2} \rightarrow \left| \varphi\left(\frac{\alpha\mathbf{t}}{\|\mathbf{t}\|}\right) \right| \frac{\|\mathbf{t}\|^2}{\alpha^2} \neq 0.$$

De ce fait, $(\mathbf{t}'D_n\mathbf{t})_{n \in \mathbb{N}}$ est une suite convergente vers $-2 \ln |\varphi(\mathbf{t})|$. En appliquant le lemme 10.1.8, on obtient la convergence de $(D_n)_{n \in \mathbb{N}}$ vers $D \in \mathcal{M}_k(\mathbb{R})$.

On déduit alors que pour tout $\mathbf{t} \in \mathbb{R}^k$,

$$e^{i\mathbf{t}'M_n} \rightarrow e^{\frac{1}{2}\mathbf{t}'D\mathbf{t}} \varphi(\mathbf{t}) = \psi(\mathbf{t}).$$

Posons $M_n = \begin{pmatrix} M_1^n \\ \dots \\ M_k^n \end{pmatrix}$. Il faut maintenant montrer que $\forall 1 \leq j \leq k$, $(M_j^n)_{n \geq 0}$ converge dans

\mathbb{R} . Pour $1 \leq j \leq k$, considérons $\mathbf{t}_j = \begin{pmatrix} 0 \\ \dots \\ 1 \\ \dots \\ 0 \end{pmatrix}$ (le 1 est à la j -ième place).

On a, pour $\lambda \in \mathbb{R}$,

$$e^{i\lambda M_j^n} = e^{i(\lambda\mathbf{t}_j)'M_n} \xrightarrow{n \rightarrow \infty} \psi(\lambda\mathbf{t}_j).$$

Le théorème de convergence dominée permet alors d'obtenir pour tout $r \geq 0$,

$$\int_0^r e^{i\lambda M_j^n} d\lambda \rightarrow \int_0^r \psi(\lambda\mathbf{t}_j) d\lambda. \quad (10.1.1)$$

Distinguons alors deux cas.

1. *Premier cas* : 0 n'est pas valeur d'adhérence de la suite M_j^n .

Alors, au moins pour n assez grand,

$$\int_0^r e^{i\lambda M_j^n} d\lambda = \frac{e^{irM_j^n} - 1}{iM_j^n}.$$

Comme $\psi(0) = 1$ et ψ est continue en 0, il existe $r_0 > 0$ tel que $\int_0^{r_0} \psi(\lambda\mathbf{t}_j) d\lambda \neq 0$.

Comme $e^{ir_0 M_j^n} = e^{ir_0 \mathbf{t}_j' M_j^n} \rightarrow \psi(r_0 \mathbf{t}_j)$, on déduit de (10.1.1) la convergence de M_j^n vers $\frac{\psi(r_0 \mathbf{t}_j) - 1}{i \int_0^{r_0} \psi(\lambda\mathbf{t}_j) d\lambda}$.

2. *Second cas* : 0 est valeur d'adhérence de la suite M_j^n . Soit une sous-suite $M_j^{n_k} \rightarrow 0$.

On a alors $e^{i\lambda M_j^{n_k}} \rightarrow 1, \forall \lambda$ et donc $\psi(\lambda\mathbf{t}_j) = 1, \forall \lambda$. De ce fait, $e^{i\lambda M_j^n} \rightarrow 1, \forall \lambda$. Supposons que $(M_j^n)_{n \geq 0}$ ait une autre valeur propre que 0 : $M_j^{n_k} \rightarrow m \in \overline{\mathbb{R}} \setminus \{0\}$. Alors,

$$\int_0^r e^{i\lambda M_j^{n_k}} d\lambda = \frac{e^{irM_j^{n_k}} - 1}{iM_j^{n_k}} \rightarrow 0$$

donc $\int_0^r \psi(\lambda\mathbf{t}_j) d\lambda = 0, \forall r$, ce qui contredit $\psi(0) = 1$ et ψ continue en 0. Donc 0 est seule valeur propre de $(M_j^n)_{n \geq 0}$ dans $\overline{\mathbb{R}}$ et donc $M_j^n \rightarrow 0$. \square

◇ Exercice 10.1

Soit $(B_t)_{t \in \mathbb{R}^+}$ un processus stochastique indexé par \mathbb{R}^+ vérifiant les conditions suivantes :

- (i) $\forall t_1 < t_2 < \dots < t_n, (B_{t_1}, \dots, B_{t_n})$ est un vecteur gaussien centré ;
- (ii) $\forall (s, t), \mathbf{E}(B_s B_t) = s \wedge t$;
- (iii) p.s., $t \mapsto B_t(\omega)$ est une fonction continue.

Ce processus est l'exemple le plus important de processus à temps continu. On l'appelle le **mouvement brownien**.

a) Montrer que (B_t) est à accroissements indépendants et stationnaires i.e. $\forall t_1 < t_2 < \dots < t_n, B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$ sont des variables indépendantes et $\forall s < t, B_t - B_s$ a même loi que B_{t-s} .

b) Soit $t > 0$. On pose $G_t^n = \sum_{k=0}^{n-1} (B_{\frac{k+1}{n}t} - B_{\frac{k}{n}t})^2$. Montrer de deux façons différentes qu'il existe une suite n_p telle que $G_t^{n_p} \rightarrow t$ p.s.

c) Soit $t > 0$. Montrer que p.s., $s \mapsto B_s(\omega)$ n'est pas à variations finies sur l'intervalle $[0, t]$.

◇ Exercice 10.2

Soient X et Y deux variables aléatoires indépendantes, de même loi centrée, de variance 1.

a) On suppose X et Y gaussiennes. Trouver une condition nécessaire et suffisante sur les réels a, b, c, d pour que $aX + bY$ et $cX + dY$ soient indépendantes.

Montrer en particulier que $X + Y$ et $X - Y$ sont indépendantes.

b) On suppose que $X + Y$ et $X - Y$ sont indépendantes.

b1) Montrer que ϕ , fonction caractéristique de X est telle que

$$\phi(2t) = \phi(t)^3 \phi(-t).$$

b2) Montrer que ϕ ne s'annule pas.

b3) Montrer que $\phi(t) = \phi(-t)$ pour tout t (considérer $\rho(t) = \phi(t)/\phi(-t)$).

b4) Dédurre que X et Y sont gaussiennes.

◇ Exercice 10.3

Soit $(U_n)_{n \geq 0}$ une suite de v.a.r. indépendantes de loi commune la loi $\mathcal{N}(m, \sigma^2)$, $\sigma > 0$.

Soit $a > 0$ donné. On considère $X_0 = 0$ et on suppose que $X_{n+1} = aX_n + U_n$ pour chaque $n \geq 0$.

a) (X_1, \dots, X_n) est-il un vecteur gaussien ?

b) Calculer $\mathbf{E}(X_n)$ et $\text{Var}(X_n)$.

c) c1) On suppose que $a < 1$. Montrer que $(X_n)_{n \geq 1}$ converge en loi.

c2) On suppose que $a > 1$. Montrer que $\frac{X_n}{a^{n-1}}$ converge dans L^2 vers une v.a.r. X dont on déterminera la loi.

c3) On suppose $a = 1$. Que peut-on dire de X_n ?

◇ **Exercice 10.4**

Soit $(X_1, X_2, X_3) \sim \mathcal{N}_3(0, I_3)$ un vecteur Gaussien centré de \mathbb{R}^3 de matrice de dispersion la matrice identité d'ordre 3. Soient $U = X_1 - X_2 + X_3$, $T_1 = X_1 + X_2$, $T_2 = X_2 + X_3$, $T_3 = X_1 - X_3$.

- a) Quelle est la loi de U ?
 b) Montrer que U est indépendante de (T_1, T_2, T_3) .

◇ **Exercice 10.5**

Soit $(X_i)_{i \geq 1}$, une suite de variables indépendantes de loi normale centrée réduite. On pose $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ et $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

- a) Quelle est la loi de \bar{X} ?
 b) Montrer que \bar{X} et S^2 sont indépendantes.

10.2 Le Théorème Limite Central dans \mathbb{R}^k

10.2.1 Théorème. (TLC dans \mathbb{R}^k) Soit $(X_n)_{n \geq 1} = ((X_{n1}, \dots, X_{nk}))_{n \geq 1}$ une suite de vecteurs aléatoires indépendants et de même loi dans \mathbb{R}^k , tels que $\max_u E(X_{nu}^2) < +\infty$.

Notons $c = (c_1, \dots, c_k) = (E(X_{n1}), \dots, E(X_{nk}))$ le vecteur d'espérances de la loi commune, et soit $\Sigma = (\sigma_{uv})$ la dispersion associée : $\sigma_{uv} = E((X_{nu} - c_u)(X_{nv} - c_v))$.

Alors

$$\frac{X_1 + \dots + X_n - nc}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(c, \Sigma),$$

où $\mathcal{N}(c, \Sigma)$ désigne la loi de la variable Gaussienne d'espérance c et de matrice de dispersion Σ .

PREUVE : Soit Y un vecteur Gaussien centré de covariance Σ . Soit $\mathbf{t} = (t_1, \dots, t_k)$ et notons $Z_n = \sum_{1 \leq u \leq k} t_u (X_{nu} - c_u)$, et $Z = \sum_{1 \leq u \leq k} t_u Y_u$.

Par le Théorème de Cramér-Wold (Théorème 7.2.5), il suffit de montrer la convergence en loi des variables réelles $\frac{Z_1 + \dots + Z_n}{\sqrt{n}} \xrightarrow{\mathcal{L}} Y$, ce qui découle directement des hypothèses et du TLC sur la droite. □

10.3 Vecteurs gaussiens et conditionnement

La caractérisation de l'indépendance des composantes d'un vecteur gaussien va être très utile pour calculer simplement des lois conditionnelles.

Commençons par un lemme qui fait apparaître une telle relation d'indépendance.

10.3.1 Lemme. Soient X et Y à valeurs dans \mathbb{R}^k et \mathbb{R}^p respectivement. On suppose que (X, Y) est un vecteur gaussien de moyenne $\begin{pmatrix} M_X \\ M_Y \end{pmatrix}$ et de covariance $\begin{pmatrix} R_X & R_{XY} \\ R_{YX} & R_Y \end{pmatrix}$. On suppose R_Y inversible. Il existe A telle que $X - AY$ et Y soient indépendantes.

PREUVE : Commençons par remarquer que

$$\begin{pmatrix} X - AY \\ Y \end{pmatrix} = \begin{pmatrix} I_k & -A \\ 0 & I_p \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

et donc que le vecteur $\begin{pmatrix} X - AY \\ Y \end{pmatrix}$ est gaussien. Sa matrice de dispersion est donnée par

$$\begin{aligned} \begin{pmatrix} I_k & -A \\ 0 & I_p \end{pmatrix} \begin{pmatrix} R_X & R_{XY} \\ R_{YX} & R_Y \end{pmatrix} \begin{pmatrix} I_k & 0 \\ -A^* & I_p \end{pmatrix} &= \begin{pmatrix} I_k & -A \\ 0 & I_p \end{pmatrix} \begin{pmatrix} R_X - R_{XY}A^* & R_{XY} \\ R_{YX} - R_YA^* & R_Y \end{pmatrix} \\ &= \begin{pmatrix} R_X - R_{XY}A^* - AR_{YX} + AR_YA^* & R_{XY} - AR_Y \\ R_{YX} - R_YA^* & R_Y \end{pmatrix}. \end{aligned}$$

Il y a indépendance d'après la proposition 10.1.7 si $R_{XY} - AR_Y = 0$ (ce qui implique $R_{YX} - R_YA^* = 0$ puisque les matrices de dispersion sont symétriques), et on prend $A = R_{XY}R_Y^{-1}$ pour avoir le résultat qu'on cherchait. \square

On a alors la proposition :

10.3.2 Proposition. *Sous les conditions du lemme 10.3.1, la loi conditionnelle de X sachant Y est la loi gaussienne de moyenne $\mathbf{E}(X/Y) = M_X + R_{XY}R_Y^{-1}(Y - M_Y)$ et de dispersion $R_X - R_{XY}R_Y^{-1}R_{YX}$.*

PREUVE : Prenons $A = R_{XY}R_Y^{-1}$ satisfaisant aux conclusions du lemme 10.3.1. Comme $X = (X - AY) + AY$ et que $X - AY$ et AY sont indépendantes, en appliquant le résultat de l'exercice 9.3, on déduit que la loi conditionnelle de X sachant Y est la translatée par AY de la loi de $X - AY$. D'après la preuve du lemme 10.3.1 cette dernière est une loi normale de dispersion $R_X - R_{XY}A^* - AR_{YX} + AR_YA^*$ et de moyenne $M_X - AM_Y$. Remplaçant A par sa valeur $R_{XY}R_Y^{-1}$, on obtient que le vecteur $X - AY$ est gaussien et suit la loi de probabilités $\mathcal{N}(M_X - R_{XY}R_Y^{-1}M_Y, R_X - R_{XY}R_Y^{-1}R_{YX})$. \square

Une application intéressante est donnée ci-après. Il s'agit d'un cas élémentaire de filtrage linéaire.

Soit $(Z_n)_{n \geq 0}$ une suite de variables aléatoires réelles indépendantes. On suppose que pour tout n , $Z_n \sim \mathcal{N}(0, c_n^2)$ où $c_n > 0$. Soit X une variable aléatoire réelle de loi $\mathcal{N}(0, \sigma^2)$ ($\sigma > 0$), indépendante de la suite $(Z_n)_{n \geq 0}$. On pose $Y_n = X + Z_n$, $\mathcal{G}_n = \sigma(Y_1, \dots, Y_n)$, $\hat{X}_n = \mathbf{E}(X/\mathcal{G}_n)$. On pose $Y^n = (Y_1, \dots, Y_n)'$.

10.3.3 Lemme. *Soient $\sigma > 0, a \in \mathbb{R}^d$ et C une matrice $d \times d$ définie positive. Alors*

$$(C + \sigma^2 aa')^{-1} = C^{-1} - \frac{C^{-1}aa'C^{-1}}{\sigma^{-2} + \langle C^{-1}a, a \rangle}.$$

PREUVE : Il suffit de calculer

$$\begin{aligned} &(C + \sigma^2 aa')(C^{-1} - \frac{C^{-1}aa'C^{-1}}{\sigma^{-2} + \langle C^{-1}a, a \rangle}) \\ &= I + \sigma^2 aa'C^{-1} - \frac{1}{\sigma^{-2} + \langle C^{-1}a, a \rangle} (aa'C^{-1} + \sigma^2 aa'C^{-1}aa'C^{-1}) \\ &= I + \sigma^2 aa'C^{-1} - \frac{1 + \sigma^2 \langle C^{-1}a, a \rangle}{\sigma^{-2} + \langle C^{-1}a, a \rangle} aa'C^{-1} = I. \end{aligned} \quad \square$$

On peut alors obtenir comme suit \hat{X}_n et $\mathbf{E}((X - \hat{X}_n)^2)$.

10.3.4 Proposition. *La loi conditionnelle de X sachant $Y^n = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix}$ est une loi normale*

de moyenne $\hat{X}_n = \frac{\sum_{k=1}^n c_k^{-2} Y_k}{\sigma^{-2} + \sum_{k=1}^n c_k^{-2}}$ et de variance $\mathbf{E}((X - \hat{X}_n)^2) = \frac{1}{\sigma^{-2} + \sum_{k=1}^n c_k^{-2}}$.

PREUVE : Par la Proposition 10.3.2, la loi conditionnelle de X sachant Y^n est gaussienne de moyenne $\hat{X}_n = M_X + R_{XY^n} R_{Y^n}^{-1} (Y^n - M_{Y^n})$ et de variance $\mathbf{E}((X - \hat{X}_n)^2) = R_X - R_{XY^n} R_{Y^n}^{-1} R_{Y^n X}$. Ici, $M_X = 0$, $M_{Y^n} = (0, 0, \dots, 0)$,

$$\begin{aligned} R_{Y^n} &= \mathbf{E}(Y^n (Y^n)') = (\mathbf{E}[(X + Z_i)(X + Z_j)])_{1 \leq i, j \leq n} = (\mathbf{E}(X^2) + \mathbf{E}(Z_i Z_j))_{1 \leq i, j \leq n} \\ &= C_n + \sigma^2 e e' \end{aligned}$$

où e est le n -vecteur ne contenant que des 1, et

$$C_n = \begin{pmatrix} c_1^2 & & & \\ & c_2^2 & & 0 \\ & & \ddots & \\ & 0 & & c_n^2 \end{pmatrix}, R_{XY^n} = \mathbf{E}\left(X \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}\right) = \begin{pmatrix} \mathbf{E}(XY_1) \\ \mathbf{E}(XY_2) \\ \dots \\ \mathbf{E}(XY_n) \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ \sigma^2 \\ \dots \\ \sigma^2 \end{pmatrix} = \sigma^2 e.$$

Remplaçant ces valeurs dans les expressions précédentes, et utilisant la relation (obtenue du Lemme 10.3.3)

$$R_{Y^n}^{-1} = C_n^{-1} - \frac{C_n^{-1} e e' C_n^{-1}}{\sigma^{-2} + \langle C_n^{-1} e, e \rangle},$$

on obtient

$$\hat{X}_n = \frac{\langle C_n^{-1} Y^n, e \rangle}{\sigma^{-2} + \langle C_n^{-1} e, e \rangle} = \frac{\sum_{k=1}^n c_k^{-2} Y_k}{\sigma^{-2} + \sum_{k=1}^n c_k^{-2}}$$

et

$$\mathbf{E}((X - \hat{X}_n)^2) = \sigma^2 - \sigma^2 e (C_n^{-1} - \frac{C_n^{-1} e e' C_n^{-1}}{\sigma^{-2} + \langle C_n^{-1} e, e \rangle}) (\sigma^2 e)',$$

soit encore, en remarquant que $e' (C_n^{-1} e e' C_n^{-1}) e = \langle C_n^{-1} e, e \rangle^2$ et en réduisant au même dénominateur

$$\frac{1}{\sigma^{-2} + \langle C_n^{-1} e, e \rangle} = \frac{1}{\sigma^{-2} + \sum_{k=1}^n c_k^{-2}}. \quad \square$$

10.3.5 Corollaire. $\hat{X}_n \rightarrow X$ dans L^2 si et seulement si $\sum_{n \geq 1} c_n^{-2} = +\infty$.

PREUVE : Comme

$$\|X - \hat{X}_n\|_2^2 = \mathbf{E}((X - \hat{X}_n)^2) = \frac{1}{\sigma^{-2} + \sum_{k=1}^n c_k^{-2}}$$

le résultat est immédiat. □

Chapitre 11

Simulation

Le but du présent chapitre est de donner quelques informations sur une technique qui a pris beaucoup d'importance dans les applications des probabilités : la simulation par ordinateur d'une loi de variable aléatoire ou de l'utilisation de cette loi, par exemple à travers un calcul d'espérance. En effet, si l'on ne veut pas se limiter à des résultats théoriques, il est nécessaire de mettre en œuvre des méthodes pratiques qui permettent un usage de quantités dépendant du hasard. Le lecteur désireux d'approfondir les méthodes qui seront juste effleurées ici pourra se reporter à [7] ou [5].

La plupart des ordinateurs possèdent des générateurs de nombres au hasard sous forme d'une fonction *random*. Naturellement, un ordinateur ne peut fournir que des nombres obtenus de façon déterministe, puisqu'ils sont résultat d'un algorithme programmé. Néanmoins, certaines procédures fournissent des réels de l'intervalle $[0, 1]$ de telle sorte qu'ils satisfont très raisonnablement aux tests statistiques vérifiés par des nombres obtenus par tirage selon la loi uniforme sur $[0, 1]$. Nous considérerons ainsi comme point de départ qu'on sait simuler le tirage d'une variable U de loi uniforme sur $[0, 1]$, et plus généralement d'une suite de variables aléatoires $(U_n)_{n \geq 1}$ indépendantes et de même loi uniforme - puisque chaque appel à une fonction *random* fournit le tirage d'une variable "aléatoire" de loi uniforme indépendante des appels précédents. Dans la suite, U (resp. $(U_n)_{n \geq 1}$) désignera ainsi une variable aléatoire de loi uniforme sur $[0, 1]$ (resp. une suite de variables aléatoires indépendantes et de loi uniforme sur $[0, 1]$).

11.1 Quelques méthodes classiques

Nous présentons ici trois grandes méthodes pour simuler des lois de variables à partir de U ou $(U_n)_{n \geq 1}$.

11.1.1 Variables aléatoires discrètes

Soient p_1, p_2, \dots, p_n des réels positifs tels que $\sum_{i=1}^n p_i = 1$. On pose en outre $p_0 = 0$.

On considère la loi de probabilités

$$\mu = \sum_{i=1}^n p_i \delta_{x_i}$$

qu'on désire simuler. Posons

$$X^U = \sum_{i=1}^n x_i \mathbb{I}_{\sum_{j=0}^{i-1} p_j \leq U < \sum_{j=0}^i p_j}.$$

On vérifie immédiatement que $\mathbf{P}(X^U = x_i) = p_i$, c'est-à-dire que X^U suit la loi μ .

11.1.2 Méthode de l'inversion

Soit μ une loi de probabilités sur \mathbb{R} , de fonction de répartition F . On pose

$$G(x) = \inf\{t \in \mathbb{R}, F(t) \geq x\} \text{ (pseudo-inverse de } F\text{)}.$$

On sait alors (voir Exercice 6.6.1) que la variable aléatoire $X = G(U)$ suit la loi μ .

Exemple : *Simulation de la loi exponentielle de paramètre $\lambda > 0$.*

Pour cette loi, on sait que $F(t) = 1 - e^{-\lambda t}$ pour $t \geq 0$, et donc $G(x) = -\frac{1}{\lambda} \ln(1-x)$ pour $0 \leq x \leq 1$. Si U est une variable de loi uniforme sur $[0, 1]$, noter que $1-U$ suit encore la même loi. On pose donc

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

La variable X suit une loi exponentielle de paramètre λ (densité $\lambda e^{-\lambda x} \mathbb{I}_{x>0}$).

11.1.3 Méthode du rejet

Soit μ une loi de probabilités admettant une densité f à support compact sur $[a, b]$ telle qu'il existe $K \in]1, +\infty[$ avec $\max_{x \in [a, b]} f(x) \leq K$.

On va utiliser pour la simuler deux suites indépendantes (U_n) et (V_n) de variables indépendantes de même loi uniforme sur $[0, 1]$. Posons

$$T_K = \inf\{n \geq 1, KV_n < f(a + (b-a)U_n)\}$$

et

$$X = a + (b-a)U_{T_K}.$$

En pratique, cela signifie donc qu'on regarde le point $(a + (b-a)U_n, KV_n)$ du rectangle $[a, b] \times [0, K]$ et qu'on "accepte" le point s'il se trouve dans la partie située sous la courbe de f . On a alors

11.1.1 Proposition. *Sous les conditions précédentes, X et T_K sont des variables indépendantes telles que*

1. *La variable X suit la loi μ*
2. *T_K suit la loi géométrique de paramètre $\frac{1}{K}$*

$$\mathbf{P}(T_K = n) = \frac{1}{K(b-a)} \left(1 - \frac{1}{K(b-a)}\right)^{n-1}.$$

En particulier, $T_K < +\infty$ p.s. et $E(T_K) = K(b-a)$.

PREUVE : posons pour simplifier $\tilde{U}_n = a + (b-a)U_n$ et $\tilde{V}_n = KV_n$. Ce sont deux suites indépendantes de variables aléatoires indépendantes de lois respectives uniformes sur $[a, b]$ et $[0, K]$.

Soient A un borélien de $[a, b]$ et $n \in \mathbb{N}$. On a

$$\begin{aligned} \mathbf{P}(X \in A, T_K = n) &= \mathbf{P}(\tilde{U}_n \in A, T = n) \\ &= \mathbf{P}(\tilde{U}_n \in A, \tilde{V}_1 \geq f(\tilde{U}_1), \dots, \tilde{V}_{n-1} \\ &\geq f(\tilde{U}_{n-1}), \tilde{V}_n < f(\tilde{U}_n)) \\ &= \prod_{i=1}^{n-1} \mathbf{P}(\tilde{V}_i \geq f(\tilde{U}_i)) \mathbf{P}[\tilde{U}_n \in A, \tilde{V}_n < f(\tilde{U}_n)]. \end{aligned}$$

On a

$$\begin{aligned}\mathbf{P}(\tilde{V}_i \geq f(\tilde{U}_i)) &= [K(b-a)]^{-1} \int_{[a,b] \times [0,K]} \mathbb{I}_{y \geq f(x)} dx dy \\ &= [K(b-a)]^{-1} \int_{[a,b]} dx (K - f(x)) \\ &= 1 - [K(b-a)]^{-1}\end{aligned}$$

et

$$\begin{aligned}\mathbf{P}(\tilde{U}_n \in A, \tilde{V}_n < f(\tilde{U}_n)) &= [K(b-a)]^{-1} \int_{[a,b] \times [0,K]} \mathbb{I}_{x \in A} \mathbb{I}_{y < f(x)} dx dy \\ &= [K(b-a)]^{-1} \int_{[a,b]} f(x) \mathbb{I}_{x \in A} dx \\ &= \mu(A) [K(b-a)]^{-1}.\end{aligned}$$

De ce fait,

$$\mathbf{P}(X \in A, T_K = n) = \mu(A) [K(b-a)]^{-1} (1 - [K(b-a)]^{-1})^{n-1}. \quad \square$$

11.2 Simulation de lois particulières

Les méthodes que nous avons présentées précédemment sont générales. Néanmoins, pour simuler certaines lois usuelles, il est plus efficace d'utiliser des techniques spécifiques à ces cas particuliers.

11.2.1 Loi binomiale $\mathcal{B}(n, p)$

On sait qu'une telle loi est celle de la somme de n variables indépendantes de loi de Bernoulli de paramètre p . Partant donc de la donnée de n variables indépendantes U_1, \dots, U_n de loi uniforme sur $[0, 1]$, on pose

$$X = \sum_{i=1}^n \mathbb{I}_{U_i < p}.$$

11.2.2 Loi de Poisson de paramètre $\lambda > 0$

La méthode repose sur un résultat préliminaire.

11.2.1 Lemme. Soient $n \in \mathbb{N}^*$ et X_1, \dots, X_{n+1} , $n+1$ variables aléatoires indépendantes de loi exponentielle de paramètre $\lambda > 0$. Alors

$$\mathbf{P}(X_1 + \dots + X_n \leq 1 < X_1 + \dots + X_{n+1}) = e^{-\lambda} \frac{\lambda^n}{n!}.$$

PREUVE : La loi de (X_1, \dots, X_{n+1}) admet sur \mathbb{R}^{n+1} la densité

$$\lambda^{n+1} e^{-\lambda(x_1 + \dots + x_{n+1})} \mathbb{I}_{x_1 > 0, \dots, x_{n+1} > 0} dx_1 \dots dx_{n+1}.$$

On a

$$\begin{aligned}\mathbf{P}(X_1 + \dots + X_n \leq 1 < X_1 + \dots + X_{n+1}) &= \lambda^{n+1} \int_{\mathbb{R}^{n+1}} dx_1 \dots dx_{n+1} e^{-\lambda(x_1 + \dots + x_{n+1})} \mathbb{I}_{x_1 > 0, \dots, x_{n+1} > 0} \mathbb{I}_{x_1 + \dots + x_n \leq 1 < x_1 + \dots + x_{n+1}} \\ &= \lambda^n \int_{\mathbb{R}^n} dx_1 \dots dx_n e^{-\lambda(x_1 + \dots + x_n)} \mathbb{I}_{x_1 > 0, \dots, x_n > 0} \left[\int_{1 - (x_1 + \dots + x_n)}^{\infty} \lambda e^{-\lambda x_{n+1}} dx_{n+1} \right] \\ &= \lambda^n \int_{\mathbb{R}^n} dx_1 \dots dx_n e^{-\lambda(x_1 + \dots + x_n)} \mathbb{I}_{x_1 > 0, \dots, x_n > 0} e^{-\lambda} e^{\lambda(x_1 + \dots + x_n)} \\ &= \lambda^n e^{-\lambda} \int_{\mathbb{R}^n} dx_1 \dots dx_n \mathbb{I}_{x_1 > 0, \dots, x_n > 0, x_1 + \dots + x_n \leq 1}.\end{aligned}$$

Or, par le changement de variable $u_1 = x_1, u_2 = x_1 + x_2, \dots, u_n = x_1 + x_2 + \dots + x_n$, on obtient

$$\int_{\mathbb{R}^n} dx_1 \dots dx_n \mathbb{I}_{x_1 > 0, \dots, x_n > 0, x_1 + \dots + x_n \leq 1} = \int_{\mathbb{R}^n} du_1 \dots du_n \mathbb{I}_{0 < u_1 < u_2 < \dots < u_n < 1} = \frac{1}{n!}. \quad \square$$

Partant d'une suite $(U_i)_{i \geq 1}$ de variables aléatoires indépendantes de loi uniforme sur $[0, 1]$, on obtient donc la simulation suivante pour une loi de Poisson :

$$X = \inf\{n \in \mathbb{N}, T_1 + \dots + T_n \leq 1 < T_1 + \dots + T_{n+1}\}$$

où $T_i = -\frac{1}{\lambda} \ln U_i$, soit encore

$$X = \inf\{n \in \mathbb{N}, \prod_{k=1}^{n+1} U_k \leq e^{-\frac{1}{\lambda}} < \prod_{k=1}^n U_k\}.$$

11.2.3 Variable normale réduite

On a d'abord le résultat suivant.

11.2.2 Lemme. Soient R et Θ deux variables aléatoires réelles indépendantes de lois respectives exponentielle de paramètre $\frac{1}{2}$ et uniforme sur $[0, 2\pi]$. Alors le vecteur aléatoire de \mathbb{R}^2

$$\tilde{X} = (\sqrt{R} \cos \Theta, \sqrt{R} \sin \Theta)$$

suit une loi $\mathcal{N}(0, I_2)$.

PREUVE : Soit f borélienne bornée de \mathbb{R}^2 dans \mathbb{R} . On a

$$\begin{aligned} E(f(\tilde{X})) &= E(f(\sqrt{R} \cos \Theta, \sqrt{R} \sin \Theta)) \\ &= \frac{1}{2\pi} \frac{1}{2} \int_{\mathbb{R}^2} f(\sqrt{r} \cos \theta, \sqrt{r} \sin \theta) e^{-\frac{1}{2}r} \mathbb{I}_{r > 0, \theta \in [0, 2\pi]} dr d\theta. \end{aligned}$$

Effectuant le changement de variables

$$\begin{cases} x = \sqrt{r} \cos \theta \\ y = \sqrt{r} \sin \theta \end{cases},$$

on obtient

$$E(f(\tilde{X})) = \frac{1}{2\pi} \int_{\mathbb{R}^2} f(x, y) e^{-\frac{x^2+y^2}{2}} dx dy. \quad \square$$

Soient donc U_1 et U_2 deux variables aléatoires indépendantes de lois uniformes sur $[0, 1]$. Appliquant le Lemme 11.2.2, on obtient que

$$X = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$$

suit une loi normale centrée réduite. Naturellement, on a même plus puisque

$$Y = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$$

suit aussi une telle loi, et est indépendante de X .

11.3 Calculs d'espérances

Il arrive fréquemment dans la pratique que pour résoudre un problème, on soit en fait ramené au calcul de l'espérance d'une variable aléatoire X . Nous supposons dans cette section que la variable X est "obtenue" sous la forme d'une simulation $f(U_1, \dots, U_n)$ où les U_i sont comme précédemment des variables aléatoires indépendantes de même loi uniforme sur $[0, 1]$. On suppose en outre que $\mathbf{E}(X)$ existe, et nous allons présenter deux méthodes pour le calcul de cette espérance.

11.3.1 Méthode de Monte-Carlo

Elle repose sur des tirages indépendants. On considère une suite $(U_i)_{i \geq 0}$ de variables indépendantes de loi uniforme sur $[0, 1]$, et on pose, pour $k \geq 0$,

$$X_k = f(U_{kn}, \dots, U_{kn+(n-1)}).$$

La suite $(X_k)_{k \in \mathbb{N}}$ est alors composée de variables aléatoires indépendantes de même loi que X (l'indépendance provenant du fait que les indices intervenant dans les variables U_i sont à chaque fois différents). Par la loi forte des grands nombres, $\frac{1}{n} \sum_{k=1}^n X_k \rightarrow \mathbf{E}(X)$ p.s. et dans L^1 .

Si X est de carré intégrable, l'inégalité de Bienaymé-Chebyshev donne une estimation de la probabilité que l'approximation soit plus ou moins bonne sous la forme

$$\mathbf{P}\left(\left| \mathbf{E}(X) - \frac{1}{n} \sum_{k=1}^n X_k \right| > \varepsilon\right) \leq \frac{\text{Var}(X)}{n\varepsilon^2}.$$

11.3.2 Suites à discrédance faible

Pour le calcul numérique de l'espérance de X , on peut cependant avoir intérêt à utiliser une suite de points déterministes convenablement choisie plutôt que de faire des simulations à partir de variables aléatoires. Un choix efficace est celui des suites dites à discrédance faible. Pour simplifier la présentation, nous allons supposer ici que $X = f(U)$ et que $\mathbf{E}(X)$ est donc de la forme $\int_0^1 f(x) dx$. Il est cependant possible d'étendre la méthode au cas de la dimension d (voir [7]) mais c'est surtout quand d est petit que cette méthode s'avère efficace car les constantes apparaissant dans les majorations peuvent exploser rapidement avec d . En grande dimension, il est ainsi préférable d'utiliser des suites de variables aléatoires uniformes pour lesquelles le comportement asymptotique est indépendant de la dimension. Nous ne donnons ici que quelques résultats et renvoyons à [21] pour des prolongements.

11.3.1 Définition. Pour une suite $(x_k)_{k \geq 1}$, on appelle **discrédance** de rang r la quantité

$$D_r = \sup_{y \in [0, 1]} \left| \frac{1}{r} \sum_{k=1}^r \mathbb{I}_{[0, y]}(x_k) - y \right|.$$

On dit que la suite est équirépartie si D_r tend vers 0 quand $r \rightarrow +\infty$.

On vérifie aisément que la condition $D_r \rightarrow 0$ est équivalente à la convergence étroite des mesures $\frac{1}{r} \sum_{k=1}^r \delta_{x_k}$ vers la mesure de Lebesgue sur $[0, 1]$. La remarque importante est qu'il est possible de trouver des suites pour lesquelles la convergence de $\frac{1}{n} \sum_{k=1}^n \delta_{x_k}$ vers la mesure de Lebesgue est plus rapide que celle de $\frac{1}{r} \sum_{k=1}^r \delta_{U_k}$ si (U_1, \dots, U_r) sont des

variables aléatoires indépendantes de loi uniforme sur $[0,1]$. On a bien sûr intérêt à choisir des suites pour lesquelles la discrédance est la plus faible possible.

Un exemple de telle suite sont les suites de Van der Corput. Soit $p \in \mathbb{N}^*$ un nombre premier. Pour $r \geq 1$, il existe une unique décomposition (dite p -adique) de r sous la forme

$$r = a_0^{r,p} + a_1^{r,p} p + \dots + a_{k_{r,p}}^{r,p} p^{k_{r,p}}$$

où $k_{r,p}$ est un entier non nul, $0 \leq a_k^{r,p} \leq p-1$, $a_{k_{r,p}}^{r,p} \neq 0$. On pose alors

$$\tilde{x}_r = \frac{a_0^{r,p}}{p} + \dots + \frac{a_{k_{r,p}}^{r,p}}{p^{k_{r,p}+1}}.$$

11.3.2 Proposition. *Pour le choix précédent de \tilde{x}_k , pour toute fonction f à variation finie sur $[0,1]$, il existe une constante K_f telle que*

$$\left| \frac{1}{r} \sum_{j=1}^r f(\tilde{x}_j) - \int_0^1 f(x) dx \right| < \frac{K_f \ln r}{r}.$$

Cette vitesse de convergence est donc meilleure que celle qu'on aurait obtenue avec une suite de variables aléatoires indépendantes de loi uniforme sur $[0,1]$, pour laquelle on peut montrer un comportement de l'ordre de $\sqrt{\frac{\ln \ln r}{r}}$.

Chapitre 12

Martingales à temps discret

12.1 Généralités

On fixe un espace de probabilités filtré $(\Omega, (\mathcal{F}_n)_n, \mathcal{F}, \mathbf{P})$. On suppose que \mathcal{F} contient ses ensembles négligeables mais les tribus \mathcal{F}_n ne sont pas nécessairement supposées complètes. On sait que la famille d'ensembles $\bigcup_{n \geq 0} \mathcal{F}_n$ n'est pas une tribu. On définit donc $\mathcal{F}_\infty = \sigma(\bigcup_{n \geq 0} \mathcal{F}_n)$. \mathcal{F}_∞ est une tribu, en général plus petite que \mathcal{F} , qui jouera un rôle important dans la suite.

• Un processus réel adapté $(X_n)_{n \geq 0}$ est appelé une martingale (resp. une surmartingale, une sous-martingale) s'il est intégrable et si, pour tout $n \in \mathbb{N}$,

$$\mathbf{E}(X_{n+1} | \mathcal{F}_n) = X_n \quad \text{p.s. (resp. } \leq, \geq \text{)}.$$

Donc $(X_n)_{n \geq 0}$ est une martingale (resp. une surmartingale, une sous-martingale) si et seulement si pour tout $A \in \mathcal{F}_n$ on a

$$\int_A X_{n+1} d\mathbf{P} = \int_A X_n d\mathbf{P} \quad (\text{resp. } \leq, \geq)$$

Notons que $(X_n)_{n \geq 0}$ est une surmartingale si et seulement si $(-X_n)_{n \geq 0}$ est une sous-martingale (et réciproquement) et que $(X_n)_{n \geq 0}$ est une martingale si et seulement si $(X_n)_{n \geq 0}$ est à la fois une surmartingale et une sous-martingale.

Les définitions ci-dessus sont encore valables si on remplace “ X_n intégrable” par “ X_n positif”. On parlera alors de martingale (resp. surmartingale, sous-martingale) positive. Il reste entendu qu'une martingale (resp. surmartingale, sous-martingale) sans adjectif est nécessairement intégrable.

Exemple Soit $X \in L^1$ ou ≥ 0 . On vérifie, grâce aux propriétés de l'espérance conditionnelle, que $X_n = \mathbf{E}(X | \mathcal{F}_n)$ est une martingale.

Enfin un processus $X = (\Omega, \mathcal{F}, (X_n)_{n \geq 0}, \mathbf{P})$ sera dit une martingale (resp. une surmartingale, une sous-martingale) si c'est une martingale (resp. une surmartingale, une sous-martingale) par rapport à la filtration $\mathcal{F}_n^0 = \sigma(X_0, \dots, X_n)$.

• Il est clair que le processus $(X_n)_{n \geq 0}$ est une martingale (resp. une surmartingale, une sous-martingale) si et seulement si $\mathbf{E}(X_{n+1} - X_n | \mathcal{F}_n) = 0$, (resp. \leq, \geq). Donc si $(Y_n)_{n \geq 0}$ est un processus adapté (intégrable ou positif), alors $X_n = Y_0 + Y_1 + \dots + Y_n$ définit une martingale (resp. une surmartingale, une sous-martingale) si et seulement si $\mathbf{E}(Y_{n+1} | \mathcal{F}_n) = 0$, (resp. $\leq 0, \geq 0$). En particulier si $(Y_n)_{n \geq 0}$ est une suite de v.a. indépendantes intégrables ou positives et si on pose $S_0 = 0, S_n = Y_0 + Y_1 + \dots + Y_n$ pour $n \geq 1$, alors $(S_n)_{n \geq 0}$ est une martingale (resp. une surmartingale, une sous-martingale) pour la filtration $\sigma(Y_0, Y_1, \dots, Y_n)$ si, pour tout $n \geq 1, \mathbf{E}(Y_n) = 0$, (resp. $\leq 0, \geq 0$).

12.2 Premières propriétés

- (i) Si $(X_n)_{n \geq 0}$ est une martingale (resp. une surmartingale, une sous-martingale), la suite $(\mathbf{E}(X_n))_{n \geq 0}$ est constante (resp. décroissante, croissante).
- (ii) Si $(X_n)_{n \geq 0}$ est une martingale (resp. une surmartingale, une sous-martingale), on a, pour $m < n$, $\mathbf{E}(X_n | \mathcal{F}_m) = X_m$ p.s. (resp. \leq, \geq).
- (iii) Si $(X_n)_{n \geq 0}$ et $(Y_n)_{n \geq 0}$ sont des surmartingales, il en est de même de $X_n + Y_n$ et de $X_n \wedge Y_n$.
- (iv) Si $(X_n)_{n \geq 0}$ est une martingale (resp. une surmartingale) et f est une fonction concave (resp. concave croissante) telle que $f(X_n)$ est intégrable ou positive pour tout $n \geq 0$, alors $Y_n = f(X_n)$ est une surmartingale (ceci résulte de l'inégalité de Jensen). En particulier si $(X_n)_{n \geq 0}$ est une martingale, $|X_n|$ et X_n^2 sont des sous-martingales.
- (v) Si $(X_n)_{n \geq 0}$ est une martingale (resp. une surmartingale, une sous-martingale), il en est de même du processus stoppé $X_n^v = X_{n \wedge v}$, où v est un temps d'arrêt de la filtration $\{\mathcal{F}_n\}_n$. En effet, d'après la définition de temps d'arrêt, on a $\{v \geq n+1\} = \{v \leq n\}^C \in \mathcal{F}_n$ et donc

$$\begin{aligned} \mathbf{E}(X_{n+1}^v - X_n^v | \mathcal{F}_n) &= \mathbf{E}((X_{n+1} - X_n)1_{\{v \geq n+1\}} | \mathcal{F}_n) = \\ &= 1_{\{v \geq n+1\}} \mathbf{E}(X_{n+1} - X_n | \mathcal{F}_n). \end{aligned} \quad (12.2.1)$$

- On dit qu'un processus adapté $(A_n)_{n \geq 0}$ est un processus croissant prévisible si $A_0 = 0$, $A_n \leq A_{n+1}$ et si A_{n+1} est \mathcal{F}_n mesurable. Soit $(X_n)_{n \geq 0}$ une sous-martingale intégrable. On définit

$$A_0 = 0, \quad A_{n+1} = A_n + \mathbf{E}(X_{n+1} - X_n | \mathcal{F}_n).$$

Par construction, $(A_n)_n$ est un processus croissant prévisible intégrable et $M_n = X_n - A_n$ vérifie $\mathbf{E}(M_{n+1} - M_n | \mathcal{F}_n) = 0$ (puisque A_{n+1} est \mathcal{F}_n mesurable!). La décomposition est de plus unique car si $X_n = M_n + A_n = M'_n + A'_n$, on a $A_0 = A'_0 = 0$ et

$$A'_{n+1} - A'_n = X_{n+1} - X_n - (M_{n+1} - M_n)$$

d'où, en conditionnant par \mathcal{F}_n , $A'_{n+1} - A'_n = \mathbf{E}(X_{n+1} | \mathcal{F}_n) - X_n = A_{n+1} - A_n$ d'où $A'_n = A_n$ puis $M_{n+1} = M_n$. On en déduit :

12.2.1 Théorème. (Décomposition de Doob).

Toute sous-martingale intégrable X_n s'écrit, de façon unique, sous la forme $X_n = M_n + A_n$ avec $(M_n)_{n \geq 0}$ martingale intégrable et $(A_n)_{n \geq 0}$ processus croissant prévisible intégrable.

Un tel processus croissant $(A_n)_{n \geq 0}$ s'appelle le compensateur de la sous-martingale $(X_n)_{n \geq 0}$.

12.3 Le théorème d'arrêt

- Soit $(X_n)_{n \geq 0}$ une surmartingale. On a donc, pour $m < n$, $\mathbf{E}(X_n | \mathcal{F}_m) \leq X_m$ p.s. Cette propriété reste-t-elle vraie si l'on remplace m et n par des temps d'arrêt i.e. : si v_1 et v_2 sont deux temps d'arrêt avec $v_1 \leq v_2$, a-t-on $\mathbf{E}(X_{v_2} | \mathcal{F}_{v_1}) \leq X_{v_1}$ ou encore

$$\text{pour tout } A \in \mathcal{F}_{v_1}, \quad \int_A X_{v_1} d\mathbf{P} \geq \int_A X_{v_2} d\mathbf{P}. \quad (12.3.1)$$

Supposons d'abord que $v_1 \leq v_2 = k \in \mathbb{N}$. Vu que $A \cap \{v_1 = j\} \in \mathcal{F}_j$, on a, pour $j \leq k$ $\int_{A \cap \{v_1 = j\}} X_j d\mathbf{P} \geq \int_{A \cap \{v_1 = j\}} X_k d\mathbf{P}$. Sommant en j , on obtient,

$$\int_A X_{v_1} d\mathbf{P} = \sum_{j=0}^k \int_{A \cap \{v_1 = j\}} X_j d\mathbf{P} \geq \sum_{j=0}^k \int_{A \cap \{v_1 = j\}} X_k d\mathbf{P} = \int_A X_{v_2} d\mathbf{P}$$

Supposons maintenant que $v_1 \leq v_2 \leq k \in \mathbb{N}$. Alors, appliquant 12.3.(v), pour $A \in \mathcal{F}_{v_1}$,

$$\int_A X_{v_1} d\mathbf{P} = \int_A X_{v_1}^{v_2} d\mathbf{P} \geq \int_A X_k^{v_2} d\mathbf{P} = \int_A X_{v_2} d\mathbf{P}.$$

On a montré,

12.3.1 Théorème. Soient $(X_n)_{n \geq 0}$ une martingale (resp. une surmartingale, une sous-martingale) intégrable ou positive et v_1 et v_2 deux temps d'arrêt bornés avec $v_1 \leq v_2$; alors on a $\mathbf{E}(X_{v_2} | \mathcal{F}_{v_1}) = X_{v_1}$, (resp. \leq, \geq).

Prenant l'espérance, on obtient,

12.3.2 Corollaire. Soient $(X_n)_{n \geq 0}$ une martingale (resp. une surmartingale, une sous-martingale) intégrable ou positive et v un temps d'arrêt borné, alors $\mathbf{E}(X_v) = \mathbf{E}(X_0)$, (resp. \leq, \geq).

• Dans ces énoncés, la condition v borné est essentielle et ne peut être levée sans hypothèses supplémentaires. Considérons par exemple la marche aléatoire $S_0 = 0, S_n = X_1 + X_2 + \dots + X_n$ où les $(X_n)_{n \geq 0}$ sont des v.a. indépendantes telles que $\mathbf{P}(X_k = 1) = \mathbf{P}(X_k = -1) = \frac{1}{2}$. On sait que cette marche est récurrente et donc $v = \inf(n; S_n = 1)$ vérifie $\mathbf{P}(v < +\infty) = 1$. Par ailleurs $(S_n)_{n \geq 0}$ est une martingale (12.2). On a d'une part $\mathbf{E}(S_0) = \mathbf{E}(S_n) = 0$ et d'autre part $S_v = 1$ p.s. et donc $\mathbf{E}(S_v) = 1$. La conclusion du Corollaire 12.3.2 n'est donc pas satisfaite; v n'est pas borné car $\mathbf{P}(v > n) \geq (\frac{1}{2})^n > 0$.

12.4 Inégalités maximales

• Ce sont des inégalités concernant les v.a.

$$\max_{0 \leq k \leq n} X_k, \quad \max_{0 \leq k \leq n} |X_k|, \quad \sup_n X_n, \quad \sup_n |X_n|.$$

Donnons deux exemples simples mais caractéristiques.

Soit $(X_n)_{n \geq 0}$ une surmartingale positive. Posons, pour $a > 0, v = \inf(n, X_n > a)$. Rappelons que $v = +\infty$ si $() = \emptyset$. On a alors $\{\sup_k X_k > a\} = \{v < +\infty\}$ et $X_v > a$ sur $\{v < +\infty\}$ et donc $X_{n \wedge v} \geq a 1_{\{v \leq n\}}$. On en déduit $\mathbf{E}(X_0) \geq \mathbf{E}(X_{n \wedge v}) \geq a \mathbf{P}(v \leq n)$ et, faisant tendre n vers $+\infty, a \mathbf{P}(v < +\infty) \leq \mathbf{E}(X_0)$.

De même, si $(X_n)_{n \geq 0}$ est une sous-martingale positive, on a (avec le même v), $\{\max_{0 \leq k \leq n} X_k > a\} = \{v \leq n\}$. Pour $k \leq n$,

$$\int_{\{v=k\}} X_k d\mathbf{P} \leq \int_{\{v=k\}} X_n d\mathbf{P},$$

d'où

$$\begin{aligned} a \mathbf{P}(v \leq n) &\leq \int_{\{v \leq n\}} X_v d\mathbf{P} = \sum_{k=0}^n \int_{\{v=k\}} X_k d\mathbf{P} \leq \\ &\leq \sum_{k=0}^n \int_{\{v=k\}} X_n d\mathbf{P} = \int_{\{v \leq n\}} X_n d\mathbf{P} \leq \mathbf{E}(X_n). \end{aligned} \tag{12.4.1}$$

On a montré,

12.4.1 Théorème. On a, pour $a > 0$,

- (i) si $(X_n)_{n \geq 0}$ est une surmartingale positive, $\mathbf{P}(\sup_k X_k > a) \leq \frac{1}{a} \mathbf{E}(X_0)$;
- (ii) si $(X_n)_{n \geq 0}$ est une sous-martingale positive,

$$\mathbf{P}(\max_{0 \leq k \leq n} X_k > a) \leq \int_{\{\max_{0 \leq k \leq n} X_k > a\}} X_n d\mathbf{P} \leq \frac{1}{a} \mathbf{E}(X_n).$$

A partir de ce théorème, on peut montrer (voir [2])

12.4.2 Théorème. (*Inégalité de Doob*) Soient $p > 1$ et X_n une martingale telle que $\sup_n \mathbf{E}|X_n|^p < +\infty$. Alors $\sup_n |X_n|^p$ est intégrable et,

$$\left\| \sup_n |X_n| \right\|_p \leq \frac{p}{p-1} \sup_n \|X_n\|_p.$$

Le cas le plus important est $p = 2$ pour lequel on a $\mathbf{E}[\sup_n X_n^2] \leq 4 \sup_n \mathbf{E}(X_n^2)$.

12.5 Martingales de carré intégrable

• Soit $(M_n)_{n \geq 0}$ une martingale telle que $\mathbf{E}(M_n^2) < +\infty$. Remarquons que $(M_n)_{n \geq 0}$ est une sous-martingale et donc $\mathbf{E}(M_n^2)$ est croissante. On note $(\langle M \rangle_n)_{n \geq 0}$ le processus croissant $(A_n)_{n \geq 0}$ intervenant dans la décomposition de Doob de la sous-martingale $(M_n^2)_{n \geq 0}$ (voir 12.4). Donc $\langle M \rangle_{n+1}$ est \mathcal{F}_n mesurable et $(M_n^2 - \langle M \rangle_n)$ est une martingale intégrable. Puisque $\mathbf{E}(M_{n+p}M_n) = \mathbf{E}(M_n \mathbf{E}(M_{n+p} | \mathcal{F}_n)) = \mathbf{E}(M_n^2)$, on a

$$\mathbf{E}((M_{n+p} - M_n)^2) = \mathbf{E}(M_{n+p}^2) - \mathbf{E}(M_n^2) = \mathbf{E}(\langle M \rangle_{n+p}) - \mathbf{E}(\langle M \rangle_n) \quad (12.5.1)$$

• Supposons maintenant que la martingale $(M_n)_n$ soit bornée dans L^2 i.e. que $\sup_n \mathbf{E}(M_n^2) < +\infty$. En fait, vu que $\mathbf{E}(M_n^2)$ est croissante, $\mathbf{E}(M_n^2) \nearrow m^*$. On a alors, d'après (12.5.1), $\mathbf{E}(M_{n+p} - M_n)^2 \leq m^* - \mathbf{E}(M_n^2)$ d'où $\sup_p \mathbf{E}(M_{n+p} - M_n)^2 \rightarrow_n 0$, ce qui montre que $(M_n)_{n \geq 0}$ est une suite de Cauchy dans L^2 et donc converge dans L^2 . Montrons que $(M_n)_{n \geq 0}$ converge p.s.. Soit $V_n = \sup_{i,j \geq n} |M_i - M_j|$. Evidemment $V_n \searrow V$ et il suffit de montrer que $V = 0$ p.s. car, alors, M_n est p.s. une suite de Cauchy et donc converge p.s.. Utilisant le théorème 12.4.2 et l'inégalité de Markov, on a, pour tout $\rho > 0$,

$$\begin{aligned} \mathbf{P}(V_n \geq \rho) &= \mathbf{P}\left(\sup_{i,j \geq n} |M_i - M_j| \geq \rho\right) \leq \mathbf{P}\left(\sup_{i \geq n} |M_i - M_n| \geq \frac{\rho}{2}\right) \leq \\ &\leq \frac{16}{\rho^2} \sup_{i \geq n} \mathbf{E}(M_i - M_n)^2 \rightarrow 0. \end{aligned}$$

Ceci implique $\mathbf{P}(V > 0) = 0$ et le :

12.5.1 Théorème. Soit M_n une martingale telle que $\sup_n \mathbf{E}(M_n^2) < +\infty$. Alors M_n converge p.s. et dans L^2 .

12.6 Théorèmes de convergence

12.6.1 Proposition. Soient $(X_n)_{n \geq 0}$ une sous-martingale bornée en module par K . Alors $(X_n)_{n \geq 0}$ converge p.s.

PREUVE :

Soient $(X_n)_{n \geq 0}$ une sous-martingale bornée en module par K et $X_n = M_n + A_n$ sa décomposition de Doob. On a $\mathbf{E}(A_n) = \mathbf{E}(X_n) - \mathbf{E}(M_0) \leq 2K$ et donc $\mathbf{E}(A_\infty) \leq 2K$ d'où $A_\infty < +\infty$ p.s. On pose $\tau_p = \inf\{n; A_{n+1} > p\}$: τ_p est un temps d'arrêt et $A_n^{\tau_p} \leq p$. Puisque $|M_{\tau_p \wedge n}| \leq |X_{\tau_p \wedge n} - A_{\tau_p \wedge n}| \leq K + p$, M^{τ_p} est une martingale bornée donc de carré intégrable

qui converge p.s. On en déduit la convergence p.s. de $(X_n^{\tau_p})_{n \geq 0}$ puis celle de $(X_n)_{n \geq 0}$ sur $(\tau_p = +\infty)$. Enfin $(X_n)_{n \geq 0}$ converge p.s. puisque, A_∞ étant p.s. finie,

$$\bigcup_{p=1}^{\infty} (\tau_p = +\infty) = \bigcup_{p=1}^{\infty} (A_\infty < p) = (A_\infty < +\infty). \square$$

• On déduit de cette proposition un théorème est très agréable à utiliser car ses hypothèses sont très simples à vérifier. Rappelons qu'une v.a positive peut prendre des valeurs infinies avec une probabilité positive.

12.6.2 Corollaire. Soit $(X_n)_{n \geq 0}$ une surmartingale positive. Alors elle converge p.s. vers une v.a. X_∞ et l'on a $X_n \geq \mathbf{E}(X_\infty | \mathcal{F}_n)$. De plus, si $\mathbf{P}(X_0 < +\infty) = 1$, $\mathbf{P}(X_\infty < +\infty) = 1$.

PREUVE :

Si $(X_n)_{n \geq 0}$ est une sur-martingale positive, la fonction exponentielle étant convexe et croissante, $Y_n = e^{-X_n}$ est une sous-martingale telle que $0 \leq Y_n \leq 1$. Donc, d'après la proposition précédente $(Y_n)_{n \geq 0}$ et $(X_n)_{n \geq 0}$ convergent p.s. De plus, le lemme de Fatou donne

$$\mathbf{E}(X_\infty / \mathcal{F}_k) = \mathbf{E}(\liminf X_n / \mathcal{F}_k) \leq \liminf \mathbf{E}(X_n / \mathcal{F}_k) \leq X_k.$$

Ceci implique que pour tout $M \geq 0$, $\mathbf{E}(X_\infty 1_{(X_0 \leq M)}) = \mathbf{E}(X_0 1_{(X_0 \leq M)}) \leq M$ et donc $X_\infty < +\infty$ sur $(X_0 < +\infty)$. \square

Bien entendu, une martingale positive est une surmartingale positive et le théorème précédent donne également la convergence dans ce cas.

12.6.3 Théorème. Soit $(X_n)_{n \geq 0}$ une sous-martingale telle que $\sup_n \mathbf{E}|X_n| < +\infty$. Alors $(X_n)_{n \geq 0}$ converge p.s.

PREUVE :

Soit $X_n^+ = M_n + A_n$ la décomposition de Doob de la sous-martingale $(X_n^+)_{n \geq 0}$. On a

$$\mathbf{E}(A_\infty) = \sup_{n \geq 0} \mathbf{E}(A_n) \leq \mathbf{E}(|M_0|) + \sup_{n \geq 0} \mathbf{E}(|X_n|) < +\infty.$$

Si $Y_n = M_n + \mathbf{E}(A_\infty / \mathcal{F}_n)$, vu que $Y_n \geq X_n^+ \geq X_n$, $(Y_n)_{n \geq 0}$ est une martingale positive et $Z_n = Y_n - X_n$ est une surmartingale positive. Evidemment, $X_n = Y_n - Z_n$ qui donc converge p.s. puisque d'après le Théorème, $\mathbf{E}(Y_\infty) < +\infty$ et $\mathbf{E}(Z_\infty) < +\infty$ et donc $Y_\infty < +\infty$ et $Z_\infty < +\infty$ p.s. \square

A propos de ce théorème, notons que, si $(X_n)_{n \geq 0}$ est une sous-martingale intégrable, on a, puisque $|x| = 2x^+ - x$, $\mathbf{E}|X_n| = 2\mathbf{E}(X_n^+) - \mathbf{E}(X_n) \leq 2\mathbf{E}(X_n^+) - \mathbf{E}(X_0)$. Donc $\sup_n \mathbf{E}|X_n| < +\infty$ si et seulement si $\sup_n \mathbf{E}X_n^+ < +\infty$. De même pour une surmartingale intégrable, $\sup_n \mathbf{E}|X_n| < +\infty$ si et seulement si $\sup_n \mathbf{E}X_n^- < +\infty$.

• Il faut noter que, contrairement au cas L^2 , la condition $\sup_n \mathbf{E}|X_n| < +\infty$ n'implique pas la convergence dans L^1 comme le montre l'exemple suivant. On choisit $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, $P =$ mesure de Lebesgue, $\mathcal{F}_n = \sigma([k2^{-n}, (k+1)2^{-n}], k = 0, \dots, 2^n - 1)$, $X_n = 1_{[0, 2^{-n}[}$. On vérifie facilement que X_n est une \mathcal{F}_n martingale, que X_n converge vers 0, que $\mathbf{E}|X_n| = \mathbf{E}(X_n) = 1$ et donc que X_n ne converge pas vers 0 dans L^1 .

• Considérons notre premier exemple de martingale i.e. $X_n = \mathbf{E}(X | \mathcal{F}_n)$, $X \in L^1$. Alors

$$\mathbf{E}|X_n| = \mathbf{E}|\mathbf{E}(X | \mathcal{F}_n)| \leq \mathbf{E}(\mathbf{E}(|X| | \mathcal{F}_n)) = \mathbf{E}(\mathbf{E}(|X|)) = \mathbf{E}|X|.$$

Donc (Th. 12.6.3) $X_n \rightarrow X_\infty$ p.s. Etudions la convergence dans L^1 . Ecrivait $X = X^+ - X^-$, on peut supposer $X \geq 0$. Soit $a > 0$. D'après le théorème de convergence dominée $\|X - X \wedge a\|_1 \rightarrow 0$ lorsque $a \rightarrow +\infty$. On vient de voir que $\mathbf{E}(X \wedge a | \mathcal{F}_n)$ converge p.s. mais, étant borné, $\mathbf{E}(X \wedge a | \mathcal{F}_n)$ converge aussi dans L^1 . Il suffit alors d'écrire que

$$\begin{aligned} & \|\mathbf{E}(X | \mathcal{F}_n) - \mathbf{E}(X | \mathcal{F}_m)\|_1 \leq \\ & \leq \|\mathbf{E}(X | \mathcal{F}_n) - \mathbf{E}(X \wedge a | \mathcal{F}_n)\|_1 + \|\mathbf{E}(X \wedge a | \mathcal{F}_n) - \mathbf{E}(X \wedge a | \mathcal{F}_m)\|_1 + \\ & \quad + \|\mathbf{E}(X \wedge a | \mathcal{F}_m) - \mathbf{E}(X | \mathcal{F}_m)\|_1 \leq \\ & \leq 2\|X - X \wedge a\|_1 + \|\mathbf{E}(X \wedge a | \mathcal{F}_n) - \mathbf{E}(X \wedge a | \mathcal{F}_m)\|_1 \end{aligned}$$

pour en déduire que $\mathbf{E}(X | \mathcal{F}_n)$ est une suite de Cauchy dans L^1 et donc converge dans L^1 . Par définition de X_n ,

$$\text{pour tout } A \in \mathcal{F}_n, \int_A X_n d\mathbf{P} = \int_A X d\mathbf{P},$$

d'où, puisque X_n converge dans L^1 ,

$$\text{pour tout } A \in \mathcal{F}_n, \int_A X_\infty d\mathbf{P} = \int_A X d\mathbf{P}. \quad (12.6.1)$$

Il est facile de voir par un argument classique que (12.6.1) est valable pour tout $A \in \mathcal{F}_\infty = \sigma(\bigcup_{n \geq 0} \mathcal{F}_n)$. Ce qui montre que $X_\infty = \mathbf{E}(X | \mathcal{F}_\infty)$. En résumé,

12.6.4 Théorème. Soit $X \in L^1$. La martingale $X_n = \mathbf{E}(X | \mathcal{F}_n)$ converge p.s. et dans L^1 vers $\mathbf{E}(X | \mathcal{F}_\infty)$.

• Enfin, au sujet des martingales inverses, nous mentionnerons seulement,

12.6.5 Théorème. Soient $X \in L^1$ et \mathcal{G}_n une suite décroissante de tribus avec $\mathcal{G}_\infty = \bigcap_n \mathcal{G}_n$. Alors la suite $X_n = \mathbf{E}(X | \mathcal{G}_n)$ converge p.s. et dans L^1 vers $\mathbf{E}(X | \mathcal{G}_\infty)$.

12.7 Martingales régulières

• Une martingale de la forme $X_n = \mathbf{E}(X | \mathcal{F}_n)$, $X \in L^1$, est appelée régulière. On vient de voir qu'une telle martingale convergeait p.s. et dans L^1 vers $X_\infty = \mathbf{E}(X | \mathcal{F}_\infty)$. Réciproquement, soit X_n une martingale convergeant dans L^1 vers une v.a. X . Puisque $|X_n| \leq |X| + |X - X_n|$, on a $\sup_n \mathbf{E}|X_n| < +\infty$. Donc (Th. 12.6.3), X_n converge p.s. vers X . On a alors, pour tout p , pour tout $A \in \mathcal{F}_n$,

$$\int_A X_{n+p} d\mathbf{P} = \int_A X_n d\mathbf{P},$$

et, faisant tendre p vers $+\infty$, vu que X_{n+p} converge dans L^1 ,

$$\text{pour tout } A \in \mathcal{F}_n, \int_A X d\mathbf{P} = \int_A X_n d\mathbf{P},$$

ce qui signifie que $X_n = \mathbf{E}(X | \mathcal{F}_n)$. En conclusion,

12.7.1 Théorème. Une martingale est régulière si et seulement si elle converge dans L^1 .

• Soit X_n une martingale régulière de limite X_∞ , la martingale est fermée en ce sens que $(X_n, n \in \bar{\mathbb{N}})$ est une martingale i.e. vérifie la définition pour tout $m \leq n \leq +\infty$. Alors, si ν est un temps d'arrêt à valeurs $\bar{\mathbb{N}}$, X_ν est parfaitement définie par $X_\nu = X_\infty$ sur $\{\nu = +\infty\}$. Dans ce cas, le théorème d'arrêt prend une forme particulièrement agréable,

12.7.2 Proposition. Soit $X_n = \mathbf{E}(X | \mathcal{F}_n)$, $X \in L^1$, une martingale régulière. Alors,

(i) si ν est un temps d'arrêt, X_ν est intégrable et $X_\nu = \mathbf{E}(X | \mathcal{F}_\nu)$,

(ii) si $\nu_1 \leq \nu_2$ sont deux temps d'arrêt, $\mathbf{E}(X_{\nu_2} | \mathcal{F}_{\nu_1}) = X_{\nu_1}$.

Exercices

Exercice 12.1

- a) Soit $X = (X_n)_{n \geq 0}$ une surmartingale telle que $\mathbf{E}(X_n)$ soit constante. Montrer que $(X_n)_{n \geq 0}$ est une martingale.
- b) Soit $(X_n)_{n \geq 0}$ un processus intégrable adapté à la filtration $(\mathcal{F}_n)_{n \geq 0}$. Montrer que $(X_n)_{n \geq 0}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale si et seulement si, pour tout temps d'arrêt borné τ de $(\mathcal{F}_n)_{n \geq 0}$, on a $\mathbf{E}(X_\tau) = \mathbf{E}(X_0)$.

Exercice 12.2

- Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbf{P})$ un espace de probabilité filtré sur lequel on considère deux martingales $(X_n)_{n \geq 0}$ et $(Y_n)_{n \geq 0}$ de carré intégrable.
- a) Montrer que, pour $m \leq n$, on a $\mathbf{E}(X_m Y_n | \mathcal{F}_m) = X_m Y_m$ p.s.
- b) Montrer que $\mathbf{E}(X_n Y_n) - \mathbf{E}(X_0 Y_0) = \sum_{k=1}^n \mathbf{E}((X_k - X_{k-1})(Y_k - Y_{k-1}))$.

Exercice 12.3

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbf{P})$ un espace de probabilité filtré sur lequel on considère une martingale réelle $(M_n)_{n \geq 0}$ telle que, pour tout $n \geq 0$, $|M_n| \leq K$. On pose

$$X_n = \sum_{k=1}^n \frac{1}{k} (M_k - M_{k-1}).$$

Montrer que $(X_n)_{n \geq 1}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale qui converge p.s. et dans L^2 .

Exercice 12.4

Soit $(Y_n)_{n \geq 1}$ une suite de v.a. indépendantes de même loi normale $N(0, \sigma^2)$, où $\sigma > 0$. On pose $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$ et $X_n = Y_1 + \dots + Y_n$. On rappelle que

$$\mathbf{E}(e^{uY_1}) = e^{u^2 \sigma^2 / 2}. \quad (12.7.1)$$

- 1) Soit $Z_n^u = \exp(uX_n - nu^2 \sigma^2 / 2)$. Montrer que, pour tout $u \in \mathbb{R}$, $(Z_n^u)_{n \geq 1}$ est une $(\mathcal{F}_n)_{n \geq 1}$ -martingale.
- 2) Montrer que, pour tout $u \in \mathbb{R}$, $(Z_n^u)_{n \geq 1}$ converge p.s. vers une v.a. Z_∞^u finie. Que vaut cette limite? Pour quelles valeurs de $u \in \mathbb{R}$, $(Z_n^u)_{n \geq 1}$ est-elle une martingale régulière?

Exercice 12.5

On dit qu'un processus $(M_n)_{n \geq 0}$ est à *accroissements indépendants* si, pour tout n , la v.a. $M_{n+1} - M_n$ est indépendante de la tribu $\mathcal{F}_n = \sigma(M_0, \dots, M_n)$.

1) Soit $(M_n)_{n \geq 0}$ une martingale de carré intégrable et à accroissements indépendants. On pose $\sigma_0^2 = \text{Var}(M_0)$ et, pour $n \geq 1$, $\sigma_k = \text{Var}(M_k - M_{k-1})$.

1a) Montrer que $\text{Var}(M_n) = \sum_{k=0}^n \sigma_k^2$.

1b) Soit $(\langle M \rangle_n)_{n \geq 0}$ le processus croissant associé à $(M_n)_{n \geq 0}$. Calculer $\langle M \rangle_n$.

2) Soit $(M_n)_{n \geq 0}$ une martingale gaussienne (on rappelle que le processus $(M_n)_{n \geq 0}$ est gaussien si, pour tout n , le vecteur (M_0, \dots, M_n) est gaussien).

2a) Montrer que $(M_n)_{n \geq 0}$ est à accroissements indépendants.

2b) Montrer que, pour tout $\theta \in \mathbb{R}$ fixé, le processus

$$Z_n^\theta = e^{\theta M_n - \frac{\theta^2}{2} \langle M \rangle_n} \quad (12.7.2)$$

est une martingale. Cette martingale converge-t-elle p.s. ?

Exercice 12.6

Soit $(X_n)_{n \geq 0}$ une suite de v.a. à valeurs $[0, 1]$. On pose $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$. On suppose que $X_0 = a \in [0, 1]$ et que

$$\mathbf{P}\left(X_{n+1} = \frac{X_n}{2} \mid \mathcal{F}_n\right) = 1 - X_n, \quad \mathbf{P}\left(X_{n+1} = \frac{1 + X_n}{2} \mid \mathcal{F}_n\right) = X_n.$$

1) Montrer que $(X_n)_{n \geq 0}$ est une martingale qui converge p.s. et dans L^2 vers une v.a. Z .

2) Montrer que $\mathbf{E}((X_{n+1} - X_n)^2) = \frac{1}{4} \mathbf{E}(X_n(1 - X_n))$.

3) Calculer $\mathbf{E}(Z(1 - Z))$. Quelle est la loi de Z ?

Exercice 12.7

A l'instant 1 une urne contient une boule blanche et une boule rouge. On tire une boule et on la remplace par deux boules de la même couleur que celle tirée, ce qui donne la nouvelle composition de l'urne à l'instant 2, et ainsi de suite suivant le même procédé.

On note Y_n et $X_n = Y_n/(n+1)$ le nombre et la proportion de boules blanches dans l'urne à l'instant n . On pose $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$.

1) Montrer que $(X_n)_{n \geq 1}$ est une martingale qui converge p.s. vers une v.a. U et que l'on a, pour tout $k \geq 1$, $\lim_{n \rightarrow +\infty} \mathbf{E}(X_n^k) = \mathbf{E}(U^k)$.

2) On fixe $k \geq 1$. On pose, pour $n \geq 1$,

$$Z_n = \frac{Y_n(Y_n+1)\dots(Y_n+k-1)}{(n+1)(n+2)\dots(n+k)}.$$

Montrer que $(Z_n)_{n \geq 1}$ est une martingale. Quelle est sa limite? En déduire la valeur de $\mathbf{E}(U^k)$.

3) Soit X une v.a. réelle telle que $|X| \leq M < +\infty$ p.s. Montrer que sa fonction caractéristique se développe en série entière

$$\phi(t) = \sum_{k=0}^{\infty} \frac{\phi^{(k)}(0)}{k!} t^k \quad (12.7.3)$$

pour tout $t \in \mathbb{R}$.

4) Quelle est la loi de U ?

Exercice 12.8

(Une preuve de la loi 0-1 de Kolmogorov par les martingales) Soit $(Y_n)_{n \geq 1}$ une suite de v.a. indépendantes. On définit

$$\begin{aligned} \mathcal{F}_n &= \sigma(Y_1, \dots, Y_n), & \mathcal{F}_\infty &= \sigma\left(\bigcup_{n \geq 1} \mathcal{F}_n\right) \\ \mathcal{F}^n &= \sigma(Y_n, Y_{n+1}, \dots), & \mathcal{F}^\infty &= \bigcap_{n \geq 1} \mathcal{F}^n. \end{aligned}$$

1) Soit $A \in \mathcal{F}^\infty$. Montrer que $\mathbf{P}(A) = 0$ ou 1. (Suggestion : $Z_n = \mathbf{E}^{\mathcal{F}_n}(1_A)$ est une martingale ...).

2) Montrer que, si X est une v.a.r. \mathcal{F}^∞ -mesurable, $X = a$ p.s.

Exercice 12.9

(Une preuve par les martingales inverses de la loi forte des grands nombres) Soit $(Y_n)_{n \geq 1}$ une suite de v.a. réelles, indépendantes, de même loi et intégrables. On pose $S_0 = 0$, $S_n = Y_1 + \dots + Y_n$ et $\mathcal{G}_n = \sigma(Y_1 + \dots + Y_n, Y_{n+1}, Y_{n+2}, \dots)$.

1) Montrer que, pour $1 \leq m \leq n$, $\mathbf{E}^{\mathcal{G}_n}(Y_m) = \mathbf{E}^{\mathcal{G}_n}(Y_1)$ p.s. et en déduire que $\mathbf{E}^{\mathcal{G}_n}(Y_1) = S_n/n$ p.s.

2) Que peut-on dire de la convergence de $Z_n = \mathbf{E}^{\mathcal{G}_n}(Y_1)$?

3) Montrer que $X = \mathbf{E}(Y_1)$ p.s.

Exercice 12.10

Soient $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, (M_n)_{n \geq 0}, \mathbf{P})$ une martingale de carré intégrable et $(\langle M \rangle_n)_{n \geq 0}$ le processus croissant associé. On note $\langle M \rangle_\infty = \lim_{n \rightarrow +\infty} \uparrow \langle M \rangle_n$.

- 1) Soit τ un temps d'arrêt. Montrer que $(\langle M \rangle_{n \wedge \tau})_{n \geq 0}$ est le processus croissant associé à la martingale $(M_n^\tau)_{n \geq 0}$ (on rappelle que $M_n^\tau = M_{n \wedge \tau}$).
- 2) Soit $a > 0$. Montrer que $\tau_a = \inf\{n, \langle M \rangle_{n+1} > a\}$ est un temps d'arrêt.
- 3) Montrer que la martingale $(M_n^{\tau_a})_{n \geq 0}$ converge p.s. et dans L^2 .
- 4) Montrer que, sur $\{\langle M \rangle_\infty < +\infty\}$, $(M_n)_{n \geq 0}$ converge p.s.

Exercice 12.11

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, (M_n)_{n \geq 0}, \mathbf{P})$ une martingale de carré intégrable. On note $A_n = \langle M \rangle_n$ le processus croissant associé et $A_\infty = \lim_{n \rightarrow \infty} \uparrow A_n$. On a vu dans l'Exercice 12.10 que cette martingale converge p.s. sur l'événement $\{A_\infty < +\infty\}$. Dans cet exercice, on précise ce qui se passe sur $\{A_\infty = +\infty\}$. On pose

$$X_0 = 0, \quad X_n = \sum_{k=1}^n \frac{M_k - M_{k-1}}{1 + A_k}, \quad n \geq 1.$$

- 1) Montrer que $(X_n)_{n \geq 0}$ est une martingale de carré intégrable et que

$$\mathbf{E}^{\mathcal{F}_{n-1}}[(X_n - X_{n-1})^2] \leq \frac{1}{1 + A_{n-1}} - \frac{1}{1 + A_n}.$$

En déduire que $\langle X \rangle_n \leq 1$ pour tout n et que $(X_n)_{n \geq 0}$ converge p.s.

- 2a) Soient $(a_n)_{n \geq 0}$ une suite de nombres strictement positifs telle que $\lim_{n \rightarrow +\infty} \uparrow a_n = +\infty$ et $(u_n)_{n \geq 0}$ une suite de nombres réels convergeant vers u . Montrer que

$$\frac{1}{a_n} \sum_{k=1}^n (a_k - a_{k-1}) u_k \xrightarrow[n \rightarrow \infty]{} u.$$

- 2b) (Lemme de Kronecker). Soient $(a_n)_{n \geq 0}$ une suite de nombres strictement positifs telle que $\lim_{n \rightarrow +\infty} \uparrow a_n = +\infty$ et $(x_n)_{n \geq 0}$ une suite de nombres réels. On pose $s_n = x_1 + \dots + x_n$. Montrer que si la série de terme général x_n/s_n converge, alors

$$\lim_{n \rightarrow +\infty} \frac{s_n}{a_n} = 0.$$

- 3) Montrer que, sur $\{A_\infty = \infty\}$, $M_n/A_n \rightarrow_{n \rightarrow \infty} 0$ p.s.

Exercice 12.12

Soit $(Z_n)_{n \geq 1}$ une suite de v.a. indépendantes telles que $\mathbf{P}(Z_i = 1) = \mathbf{P}(Z_i = -1) = \frac{1}{2}$ pour $i = 1, 2, \dots$. On pose $S_0 = 0$, $S_n = Z_1 + \dots + Z_n$, $\mathcal{F}_0 = \{\Omega, \emptyset\}$ et $\mathcal{F}_n = \sigma(Z_1, \dots, Z_n)$. Soient a un entier > 0 et $\tau = \inf\{n \geq 0, S_n = a\}$ le premier temps de passage par a .

a) Montrer que

$$X_n^\theta = \frac{e^{\theta S_n}}{(\cosh \theta)^n}$$

est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale. Montrer que, si $\theta \geq 0$, $(X_{n \wedge \tau}^\theta)_{n \geq 0}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale bornée.

b1) Montrer que, pour tout $\theta \geq 0$, $(X_{n \wedge \tau}^\theta)_{n \geq 0}$ converge p.s. et dans L^2 vers la v.a.

$$W^{\text{th}} = \frac{e^{\theta a}}{(\cosh \theta)^\tau} 1_{\{\tau < +\infty\}} \tag{12.7.4}$$

b2) Montrer que $\mathbf{P}\{\tau < +\infty\} = 1$ et que, pour tout $\theta \geq 0$,

$$\mathbf{E}[(\cosh \theta)^{-\tau}] = e^{-\theta a}.$$

Exercice 12.13

Soit, comme dans l'Exercice 12.12, $(Z_n)_{n \geq 1}$ une suite de v.a. indépendantes telles que $\mathbf{P}(Z_i = 1) = \mathbf{P}(Z_i = -1) = \frac{1}{2}$ pour $i = 1, 2, \dots$. On pose $S_0 = 0$, $S_n = Z_1 + \dots + Z_n$, $\mathcal{F}_0 = \{\Omega, \emptyset\}$ et $\mathcal{F}_n = \sigma(Z_1, \dots, Z_n)$. Soient a un entier > 0 et λ un réel tel que $0 < \lambda < \pi/(2a)$. Soit $\tau = \inf\{n \geq 0, |S_n| = a\}$ le temps de sortie de $] -a, a[$.

a) Montrer que $X_n = (\cos \lambda)^{-n} \cos(\lambda S_n)$ est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale.

b) Montrer que

$$1 = \mathbf{E}(X_{n \wedge \tau}) \geq \cos(\lambda a) \mathbf{E}((\cos \lambda)^{-n \wedge \tau}).$$

c) En déduire que $\mathbf{E}((\cos \lambda)^{-\tau}) \leq (\cos(\lambda a))^{-1}$ puis que τ est p.s. fini et que la martingale $(X_{n \wedge \tau})_{n \geq 0}$ est régulière.

d) Que vaut $\mathbf{E}((\cos \lambda)^{-\tau})$? A-t-on $\tau \in L^p, p \geq 1$?

Exercice 12.14

Soit $(S_n)_{n \geq 0}$ une marche aléatoire simple sur \mathbb{Z} : $S_0 = 0$, $S_n = U_1 + \dots + U_n$, où les v.a. U_i sont indépendantes et de même loi et telles que $0 < \mathbf{P}\{U_i = 1\} = p < 1, \mathbf{P}\{U_i = -1\} = 1 - p := q$.

a) Soit $Z_n = (\frac{q}{p})^{S_n}$. Montrer que $(Z_n)_{n \geq 0}$ est une martingale positive.

b)

Déduire d'une inégalité maximale appliquée à la martingale $(Z_n)_{n \geq 0}$ que

$$\mathbf{P}\left\{\sup_{n \geq 0} S_n \geq k\right\} \leq \left(\frac{p}{q}\right)^k$$

et que, lorsque $q > p$,

$$\mathbf{E}\left(\sup_{n \geq 0} S_n\right) \leq \frac{p}{q - p}$$

Exercice 12.15

On considère une suite $(X_n)_{n \geq 1}$ de v.a. réelles indépendantes de même loi normale $N(m, \sigma^2)$ avec $m < 0$ et on pose $S_0 = 0$,

$$S_n = X_1 + \cdots + X_n$$

et

$$\mathcal{B}_n = \sigma(S_0, \dots, S_n) \quad W = \sup_{n \geq 0} S_n.$$

Le but de cet exercice est de montrer certaines propriétés de la v.a. W .

- 1) Montrer que $\mathbf{P}(W < +\infty) = 1$.
- 2) On rappelle que, pour λ réel, $\mathbf{E}(e^{\lambda X_1}) = e^{\lambda^2 \sigma^2 / 2 + \lambda m}$. Que vaut $\mathbf{E}(e^{\lambda S_{n+1}} | \mathcal{B}_n)$?
- 3) Montrer qu'il existe un $\lambda_0 > 0$ unique tel que $(e^{\lambda_0 S_n})_{n \geq 0}$ soit une martingale.
- 4) Montrer que, pour tout $a > 1$, on a

$$\mathbf{P}(e^{\lambda_0 W} > a) \leq \frac{1}{a}$$

et que, pour $t > 0$, $\mathbf{P}(W > t) \leq e^{-\lambda_0 t}$.

- 5) Montrer que

$$\mathbf{E}(e^{\lambda W}) = 1 + \lambda \int_0^{+\infty} e^{\lambda t} \mathbf{P}(W > t) dt \quad (12.7.5)$$

et en déduire que, pour tout $\lambda < \lambda_0$, $\mathbf{E}(e^{\lambda W}) < +\infty$. En particulier la v.a. W a des moments de tous les ordres.

Exercice 12.16

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, (X_n)_{n \geq 0}, \mathbf{P})$ une sous-martingale telle que $\sup_{n \geq 0} \mathbf{E}|X_n| < +\infty$.

- 1) Montrer que, n étant fixé, la suite $(\mathbf{E}^{\mathcal{F}_p}(X_n^+))_{p \geq n}$ est croissante en p .
- 2) On pose $M_n = \lim_{p \rightarrow \infty} \uparrow \mathbf{E}^{\mathcal{F}_p}(X_n^+)$. Montrer que $(M_n)_{n \geq 0}$ est une martingale positive intégrable.
- 3) On pose $Y_n = M_n - X_n$. Montrer que $(Y_n)_{n \geq 0}$ est une surmartingale positive intégrable.

On conclut donc que toute sous-martingale bornée dans L^1 s'écrit comme la différence d'une martingale et d'une surmartingale positives intégrables (décomposition de Krickeberg).

Exercice 12.17

Soit $(Y_n)_{n \geq 0}$ une suite de v.a. indépendantes et de même loi telles que $\mathbf{P}(Y_k = 1) = \mathbf{P}(Y_k = -1) = \frac{1}{2}$. On pose $\mathcal{B}_0 = \{\emptyset, \Omega\}$, $\mathcal{B}_n = \sigma(Y_1, \dots, Y_n)$ et $S_0 = 0, S_n = Y_1 + \dots + Y_n, n \geq 1$. On note sign la fonction

$$\text{sign}(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ -1 & \text{si } x < 0 \end{cases}$$

et on considère le processus défini par

$$M_0 = 0, \quad M_n = \sum_{k=1}^n \text{sign}(S_{k-1}) Y_k, \quad n = 1, 2, \dots$$

- 1) Quel est le compensateur de la sous-martingale $(S_n^2)_{n \geq 0}$?
- 2) Montrer que $(M_n)_{n \geq 0}$ est une martingale et calculer le compensateur de $(M_n^2)_{n \geq 0}$.
- 3) Quelle est la décomposition de Doob de $(|S_n|)_{n \geq 0}$? En déduire que M_n est mesurable par rapport à la tribu $\sigma(|S_1|, \dots, |S_n|)$.

Exercice 12.18

Soient p et q deux probabilités sur un espace discret E telles que $p \neq q$ et $q(x) > 0$ pour tout $x \in E$. Soit $(X_n)_{n \geq 1}$ une suite de v.a. indépendantes à valeurs E et de même loi q . Montrer que la suite

$$Y_n = \prod_{k=1}^n \frac{p(X_k)}{q(X_k)}$$

est une martingale positive dont la limite p.s. est 0. La martingale est-elle régulière ? (Suggestion : calculer la moyenne de $\sqrt{Y_n}$.)

Exercice 12.19

Soit $(Y_n)_{n \geq 1}$ une suite de v.a. positives indépendantes d'espérance 1 et $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma(Y_k; k \leq n)$. On pose $X_0 = 1$ et $X_n = \prod_{k=1}^n Y_k$.

- 1) Montrer que $(X_n)_{n \geq 0}$ est une martingale pour la filtration $(\mathcal{F}_n)_{n \geq 0}$ et en déduire que $(\sqrt{X_n})_{n \geq 0}$ est une surmartingale.
- 2) On suppose que $\prod_{k=1}^{\infty} \mathbf{E}(\sqrt{Y_k}) = 0$. Étudier la convergence et la limite de $(\sqrt{X_n})_{n \geq 0}$ et puis de $(X_n)_{n \geq 0}$. La martingale $(X_n)_{n \geq 0}$ est-elle régulière ?
- 3) On suppose que $\prod_{k=1}^{\infty} \mathbf{E}(\sqrt{Y_k}) > 0$. Montrer que $(\sqrt{X_n})_{n \geq 0}$ est une suite de Cauchy dans L^2 et en déduire que la martingale $(X_n)_{n \geq 0}$ est régulière.

Exercice 12.20

Le but de cet exercice est de fournir une version du Lemme de Borel-Cantelli pour une famille de v.a. non nécessairement indépendantes, ce qui est souvent utile. On rappelle que, pour une martingale $(M_n)_{n \geq 0}$, les trois conditions

$$\sup_{n \geq 0} \mathbf{E}(|M_n|) < +\infty, \quad \sup_{n \geq 0} \mathbf{E}(M_n^+) < +\infty, \quad \sup_{n \geq 0} \mathbf{E}(M_n^-) < +\infty$$

sont équivalentes.

Soient $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbf{P})$ un espace de probabilités filtré et $(Y_n)_{n \geq 1}$ une suite de v.a. positives, intégrables, adaptées (pas nécessairement indépendantes).

1a) On pose $X_0 = 0, X_n = Y_1 + \dots + Y_n$. Montrer que $(X_n)_{n \geq 0}$ est une sous-martingale et déterminer le processus croissant associé $(A_n)_{n \geq 1}$.

1b) Montrer que, pour tout $a > 0$, $\tau_a = \inf\{n, A_{n+1} > a\}$ est un temps d'arrêt.

1c) On pose $Z_n = X_n - A_n$. Montrer que, pour tout n , $Z_{n \wedge \tau_a}^- \leq a$. En déduire que $(Z_{n \wedge \tau_a})_{n \geq 0}$ converge p.s.

1d) Montrer que $\{\lim_{n \rightarrow \infty} \uparrow A_n < +\infty\} \subset \{\lim_{n \rightarrow \infty} \uparrow X_n < +\infty\}$ p.s. (suggestion : se placer sur $\{\tau_a = +\infty\}$).

2) On suppose, de plus, que $\sup_{n \geq 1} Y_n \in L^1$. Montrer que

$$\{\lim_{n \rightarrow \infty} \uparrow X_n < +\infty\} = \{\lim_{n \rightarrow \infty} \uparrow A_n < +\infty\} \quad \text{p.s.}$$

(on pourra introduire le temps d'arrêt $\sigma_a = \inf(n, X_n > a)$ et considérer $Z_{n \wedge \sigma_a}^+$).

3) Soit $(B_n)_{n \geq 1}$ une suite d'événements adaptés. Montrer que

$$\left\{ \sum_{n \geq 1} \mathbf{P}^{\mathcal{F}_{n-1}}(B_n) < +\infty \right\} = \left\{ \sum_{n \geq 1} 1_{B_n} < +\infty \right\} \quad \text{p.s..}$$

Exercice 12.21

Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles, de carré intégrable, adaptées, définies sur $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbf{P})$, et $(\sigma_n^2)_{n \geq 0}$ une suite de nombres positifs. On suppose que p.s., pour tout $n \geq 1$,

$$\mathbf{E}^{\mathcal{F}_{n-1}}(X_n) = 0, \quad \mathbf{E}^{\mathcal{F}_{n-1}}(X_n^2) = \sigma_n^2 \text{ p.s.}$$

(cette condition est satisfaite, par exemple, si les v.a. $(X_n)_{n \geq 1}$ sont indépendantes centrées et de variance finie). On pose $S_0 = 0$, $A_0 = 0$ et, pour $n \geq 1$, $S_n = X_1 + \dots + X_n$, $A_n = \sigma_1^2 + \dots + \sigma_n^2$, $V_n = S_n^2 - A_n$.

- 1) Montrer que $(S_n)_{n \geq 0}$ et $(V_n)_{n \geq 0}$ sont des martingales intégrables.
- 2) Montrer que, si $\sum_{k=1}^{\infty} \sigma_k^2 < +\infty$, $(S_n)_{n \geq 0}$ converge p.s. et dans L^2 .
- 3) On suppose que $(S_n)_{n \geq 0}$ converge p.s. et qu'il existe une constante M telle que, pour tout $n \geq 1$, $|X_n| \leq M$ p.s. Pour $a > 0$, on pose $\tau_a = \inf\{n \geq 0; |S_n| > a\}$.
- 3a) Montrer que, pour tout n , $\mathbf{E}(S_{n \wedge \tau_a}^2) = \mathbf{E}(A_{n \wedge \tau_a})$.
- 3b) Montrer qu'il existe $a > 0$ tel que $\mathbf{P}(\tau_a = +\infty) > 0$.
- 3c) En déduire que $\sum_{k=1}^{\infty} \sigma_k^2 < +\infty$.

Soit $(X_n)_{n \geq 1}$ une suite de v.a. réelles, *centrées*, indépendantes, de carré intégrable. Il est clair qu'une telle suite vérifie les hypothèses de cet exercice. On a donc retrouvé les résultats classiques :

- (i) si $\sum_{k=1}^{\infty} \text{Var}(X_k) < +\infty$, alors $\sum_{k=1}^{\infty} X_k$ converge p.s.,
- (ii) si les v.a. X_k sont uniformément bornées, alors $\sum_{k=1}^{\infty} X_k$ converge p.s. si et seulement si $\sum_{k=1}^{\infty} \text{Var}(X_k) < +\infty$.

Exercice 12.22

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, (M_n)_{n \geq 0}, \mathbf{P})$ une martingale de carré intégrable et notons $A_n = \langle M \rangle_n$ le processus croissant associé. On pose $\tau_a = \inf\{n \geq 0; A_{n+1} > a^2\}$.

- 1) Montrer que τ_a est un temps d'arrêt.
- 2) Montrer que $\mathbf{P}(\sup_{n \geq 0} |M_{n \wedge \tau_a}| > a) \leq a^{-2} \mathbf{E}(A_\infty \wedge a^2)$.
- 3) Montrer que

$$\mathbf{P}\left(\sup_{n \geq 0} |M_n| > a\right) \leq \mathbf{P}(A_\infty > a^2) + \mathbf{P}\left(\sup_{n \geq 0} |M_{n \wedge \tau_a}| > a\right). \quad (12.7.6)$$

- 4) Soit X une v.a. positive. Montrer, en appliquant le théorème de Fubini, les deux relations

$$\int_0^\lambda \mathbf{P}(X > t) dt = \mathbf{E}(X \wedge \lambda) \text{ pour tout } \lambda \in [0, +\infty],$$

$$\int_0^\infty a^{-2} \mathbf{E}(X \wedge a^2) da = 2\mathbf{E}(\sqrt{X}).$$

- 5) Montrer que $\mathbf{E}(\sup_{n \geq 0} |M_n|) \leq 3\mathbf{E}(\sqrt{A_\infty})$, (on pourra intégrer (12.7.6) par rapport à a de 0 à $+\infty \dots$).

6) Soit $(Y_n)_{n \geq 1}$ une suite v.a. centrées, indépendantes, de même loi et de carré intégrable. On pose $S_0 = 0$, $\mathcal{F}_0 = \{\Omega, \emptyset\}$ et, pour $n \geq 1$, $S_n = Y_1 + \dots + Y_n$, $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$. Montrer que si τ est un temps d'arrêt tel que $\mathbf{E}(\sqrt{\tau}) < +\infty$, alors $\mathbf{E}(S_\tau) = 0$.

Il est intéressant de comparer le résultat de la question 6) avec celui de la partie A de l'Exercice 12.24 pour $m = 0$. On en déduit que le temps de passage τ de la partie A4) de l'Exercice 12.24 est de plus tel que $\mathbf{E}(\sqrt{\tau}) = +\infty$.

Exercice 12.23

Sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ on considère des v.a. $X_n^i, n = 1, 2, \dots, i = 1, 2$, indépendantes et de Bernoulli $B(1, \frac{1}{2})$. On pose $S_n^i = \sum_{k=1}^n X_k^i$, $\nu_i = \inf\{n; S_n^i = a\}$ où a est un entier ≥ 1 . On pose $\nu = \nu_1 \wedge \nu_2$.

- 1) Montrer que $\mathbf{P}(\nu_i < +\infty) = 1, i = 1, 2$.
- 2) On pose, pour $i = 1, 2$ et tout $n \geq 0$,

$$M_n^i = 2S_n^i - n, \quad M_n^{i,j} = (2S_n^i - n)(2S_n^j - n) - n\delta_{i,j}$$

où $\delta_{i,j} = 1$ si $i = j$ et $= 0$ sinon. Montrer que $(M_n^i)_n$ et $(M_n^{i,j})_n$ sont des martingales par rapport à la filtration

$$\mathcal{F}_n = \sigma(X_k^i, i = 1, 2, k \leq n).$$

- 3) Montrer que $\mathbf{E}(\nu) \leq 2a$.
- 4) Montrer que $\mathbf{E}(M_\nu^{i,j}) = 0$.
- 5) Montrer que $\mathbf{E}(|S_\nu^1 - S_\nu^2|) \leq \sqrt{a}$ (suggestion : considérer la martingale $M_n^{1,1} - 2M_n^{1,2} + M_n^{2,2}$).

Exercice 12.24

(Identités de Wald) Soit $(Y_n)_{n \geq 1}$ une suite de v.a.r. indépendantes, intégrables, de même loi. On pose $m = \mathbf{E}(Y_1)$, $S_0 = 0$, $\mathcal{F}_0 = \{\Omega, \emptyset\}$ et, pour $n \geq 1$, $S_n = Y_1 + \dots + Y_n$, $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$. Soit ν un temps d'arrêt intégrable.

A1) On pose $X_n = S_n - nm$. Montrer que $(X_n)_{n \geq 0}$ est une martingale.

A2) Montrer que, pour tout n , $\mathbf{E}(S_{n \wedge \nu}) = m\mathbf{E}(n \wedge \nu)$.

A3) Montrer que S_ν est intégrable et que $\mathbf{E}(S_\nu) = m\mathbf{E}(\nu)$ (considérer d'abord le cas $Y_n \geq 0$).

A4) Supposons $\mathbf{P}(Y_n = -1) = \mathbf{P}(Y_n = 1) = \frac{1}{2}$, pour tout n et $\tau = \inf\{n; S_n \geq a\}$, où a est un entier ≥ 1 . Dans l'Exercice 12.12, on a montré que $\tau < +\infty$ p.s. Montrer que τ n'est pas intégrable.

B) On suppose de plus que $\mathbf{E}(Y_1^2) < +\infty$ et on note $\sigma^2 = \text{Var}(Y_1)$. On suppose d'abord que $m = 0$ et on pose $Z_n = S_n^2 - n\sigma^2$.

B1) Montrer que $(Z_n)_{n \geq 0}$ est une \mathcal{F}_n -martingale.

B2) Montrer que, pour tout $j < k$, $\mathbf{E}[Y_j 1_{\{j \leq \nu\}} Y_k 1_{\{k \leq \nu\}}] = 0$ puis que $\mathbf{E}[\sum_{k=1}^{\infty} Y_k^2 1_{\{k \leq \nu\}}] < +\infty$ (suggestion : on utilise A).

B3) Montrer que $(S_{n \wedge \nu})_{n \geq 0}$ est une suite de Cauchy dans L^2 . En déduire que $S_{n \wedge \nu} \rightarrow_{n \rightarrow \infty} S_\nu$ dans L^2 .

B4) Montrer que $\mathbf{E}(S_\nu^2) = \sigma^2 \mathbf{E}(\nu)$.

B5) On ne suppose plus $m = 0$. Montrer que $\mathbf{E}((S_\nu - m\nu)^2) = \sigma^2 \mathbf{E}(\nu)$.

Exercice 12.25

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbf{P})$ un espace de probabilité filtré et ν une mesure finie sur $\mathcal{F} = \mathcal{F}_\infty$. On suppose que, pour tout $n \geq 0$, \mathbf{P} domine ν sur \mathcal{F}_n et on note X_n la densité de Radon-Nikodym : X_n est donc \mathcal{F}_n -mesurable et

$$\nu(A) = \int_A X_n d\mathbf{P}$$

pour tout $A \in \mathcal{F}_n$ (en particulier $X_n \geq 0$).

a) Montrer que $(X_n)_{n \geq 0}$ est une martingale.

b) Montrer que $(X_n)_{n \geq 0}$ converge p.s. vers une variable intégrable X .

c) Montrer que si \mathbf{P} domine ν sur \mathcal{F}_∞ , X est la densité de Radon-Nikodym correspondante.

d) On suppose que les deux mesures ν et \mathbf{P} sont étrangères sur \mathcal{F}_∞ , c'est à dire qu'il existe $S \in \mathcal{F}_\infty$ tel que $\mathbf{P}(S) = 1$ et $\nu(S) = 0$. Montrer qu'alors $X = 0$ p.s.

Exercice 12.26

Soit $(X_n)_{n \geq 0}$ une martingale intégrable définie sur un espace $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbf{P})$ et soit ν un temps d'arrêt vérifiant

$$\mathbf{P}(\nu < +\infty) = 1, \quad \mathbf{E}(|X_\nu|) < +\infty, \quad \int_{\{\nu > n\}} |X_n| d\mathbf{P} \xrightarrow{n \rightarrow \infty} 0.$$

1) Montrer que

$$\int_{\{\nu > n\}} |X_\nu| d\mathbf{P} \xrightarrow{n \rightarrow \infty} 0$$

2) Montrer que $\mathbf{E}(|X_{\nu \wedge n} - X_\nu|) \rightarrow 0$.

3) En déduire que $\mathbf{E}(X_\nu) = \mathbf{E}(X_0)$.

Exercice 12.27

(Un résultat d'arrêt pour un temps d'arrêt non borné) On considère une surmartingale intégrable $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, (X_n)_{n \geq 0}, \mathbf{P})$. On suppose qu'il existe une constante M telle que, pour tout $n \geq 1$,

$$\mathbf{E}^{\mathcal{F}_{n-1}}(|X_n - X_{n-1}|) \leq M \quad \text{p.s.}$$

1) Montrer que, si $(V_n)_{n \geq 1}$ est un processus positif tel que, pour tout $n \geq 0$, V_n soit \mathcal{F}_{n-1} -mesurable, on a

$$\mathbf{E}\left(\sum_{n=1}^{\infty} V_n |X_n - X_{n-1}|\right) \leq M \mathbf{E}\left(\sum_{n=1}^{\infty} V_n\right).$$

2) Soit ν un temps d'arrêt *intégrable*. On rappelle que $\mathbf{E}(\nu) = \sum_{n \geq 1} \mathbf{P}\{\nu \geq n\}$.

2a) Déduire de 1) que $\mathbf{E}(\sum_{n \geq 1} 1_{\{\nu \geq n\}} |X_n - X_{n-1}|) < +\infty$.

2b) Que vaut $\sum_{n \geq 1} 1_{\{\nu \geq n\}} (X_n - X_{n-1})$? En déduire que X_ν est intégrable.

3) Montrer que $(X_{\nu \wedge p})_{p \geq 0}$ tend vers X_ν dans L^1 lorsque $p \rightarrow \infty$

4) En déduire que, si $\nu_1 \leq \nu_2$ sont deux temps d'arrêt avec ν_2 intégrable, on a

$$\mathbf{E}(X_{\nu_2} | \mathcal{F}_{\nu_1}) \leq X_{\nu_1}$$

(on peut se servir du fait suivant : si $A \in \mathcal{F}_{\nu_1}$, alors $A \cap \{\nu_1 \leq k\} \in \mathcal{F}_{\nu_1 \wedge k}$, après l'avoir prouvé ...).

Chapitre 13

Processus de Poisson

Nous allons nous intéresser dans ce chapitre au processus à temps continu le plus élémentaire, appelé processus de Poisson pour des raisons qui apparaîtront bientôt. Ce processus est un exemple de processus ponctuel qui modélise des phénomènes d'arrivées aléatoires dans le temps. L'exemple classiquement proposé est celui des instants de désintégration d'un matériau radioactif.

13.1 Processus ponctuel

Un **processus ponctuel** sur \mathbb{R}^+ se fait à travers la donnée d'une suite $(T_n)_{n \geq 1}$ *strictement croissante* de variables aléatoires à valeurs dans \mathbb{R}^+ , représentant les instants (aléatoires) où se produisent successivement les phénomènes étudiés (arrivées). On suppose en outre que T_n converge presque sûrement vers $+\infty$.

Par convention, on pose $T_0 = 0$. Naturellement, la donnée des $(T_i)_{i \geq 1}$ est équivalente à celle des $(\tau_i)_{i \geq 1}$, où $\tau_i = T_i - T_{i-1}$ sont les intervalles entre les arrivées. On a en effet

$$T_n = \sum_{j=1}^n \tau_j.$$

La condition de stricte croissance de la suite $(T_n)_{n \geq 0}$ revient à supposer que $\tau_i > 0$ pour tout i .

On appelle **fonction de comptage** la fonction (aléatoire) N_t définie sur \mathbb{R}^+ par

$$N_t = \sup\{n \geq 0, T_n \leq t\}.$$

Le nombre N_t désigne donc le nombre d'arrivées qui se sont produites avant l'instant t . On a naturellement $N_0 = 0$ puisque $T_1 > 0$ p.s. et $N_t < +\infty$ p.s. puisque $T_n \rightarrow +\infty$.

13.2 Le processus de Poisson

Nous allons maintenant nous concentrer sur un cas particulier.

13.2.1 Définition. On appelle **processus de Poisson** un processus ponctuel $(T_n)_{n \geq 1}$ pour lequel la fonction de comptage satisfait aux deux conditions suivantes :

- (i) quels que soient $n \in \mathbb{N}$ et des réels $0 \leq t_0 < t_1 < \dots < t_n$, les variables $(N_{t_j} - N_{t_{j-1}})_{1 \leq j \leq n}$ sont indépendantes (accroissements indépendants) ;
- (ii) si $s < t$, alors $N_t - N_s$ a même loi que N_{t-s} (stationnarité).

La proposition suivante explique d'où vient le nom processus de Poisson.

13.2.2 Proposition. Si $(T_n)_{n \geq 1}$ est un processus de Poisson, il existe une constante $\theta > 0$ telle que pour tout $s \in \mathbb{R}^+$, N_s suit une loi de Poisson de paramètre θs . Ce paramètre θ s'appelle l'**intensité** du processus, et il exprime le nombre moyen d'arrivées par unité de temps.

PREUVE : Notons pour $0 \leq u \leq 1$, $f_s(u) = \mathbf{E}(u^{N_s})$ la fonction génératrice de la variable entière N_s . Comme pour tous réels $0 \leq s < t$, on a $N_t = N_s + (N_t - N_s)$, que les variables N_s et $N_t - N_s$ sont indépendantes, et que la loi de $N_t - N_s$ est celle de N_{t-s} , on a $f_t(u) = f_s(u)f_{t-s}(u)$ soit encore, pour tous s et t réels positifs

$$f_{t+s}(u) = f_t(u)f_s(u). \quad (13.2.1)$$

La fonction $s \mapsto f_s(u)$ étant clairement décroissante (puisque N_s est croissante), un raisonnement classique (en raisonnant d'abord sur les entiers, puis sur les rationnels) montre que la solution de (13.2.1) est de la forme $f_s(u) = \exp(-s\theta(u))$ où $\theta(u)$ est un réel positif. On a

$$\begin{aligned} \theta(u) &= \lim_{h \downarrow 0} \frac{1 - \exp(-h\theta(u))}{h} \\ &= \lim_{h \downarrow 0} \frac{1 - \mathbf{E}(u^{N_h})}{h} \\ &= \lim_{h \downarrow 0} \sum_{k \geq 1} \frac{1}{h} \mathbf{P}(N_h = k)(1 - u^k) \\ &= \lim_{h \downarrow 0} \left[\frac{\mathbf{P}(N_h = 1)}{h} (1 - u) + \sum_{k \geq 2} \frac{1}{h} \mathbf{P}(N_h = k)(1 - u^k) \right] \end{aligned} \quad (13.2.2)$$

Notons que $0 \leq \sum_{k \geq 2} \frac{1}{h} \mathbf{P}(N_h = k)(1 - u^k) \leq \frac{\mathbf{P}(N_h \geq 2)}{h}$. On a

$$\bigcup_{n \geq 1} (N_{nh} - N_{(n-1)h} \geq 2, N_{(n-1)h} = 0) \subset (T_2 < T_1 + h), \quad (13.2.3)$$

l'union étant disjointe. Notons que $\mathbf{P}(T_2 < T_1 + h) \downarrow 0$ quand $h \rightarrow 0$ puisque $T_1 < T_2$ p.s. Calculant la probabilité de la réunion dans (13.2.3), on obtient en utilisant l'indépendance de $N_{nh} - N_{(n-1)h}$ et de $N_{(n-1)h}$, ainsi que le fait que $N_{nh} - N_{(n-1)h}$ a même loi que N_h ,

$$\sum_{n \geq 1} \mathbf{P}(N_h \geq 2) \mathbf{P}(N_{(n-1)h} = 0). \quad (13.2.4)$$

On a, $\mathbf{P}(N_{(n-1)h} = 0) = f_{(n-1)h}(0) = \exp(-\theta(0)(n-1)h)$, et donc (13.2.4) devient

$$\sum_{n \geq 1} \exp(-\theta(0)(n-1)h) \mathbf{P}(N_h \geq 2) = (1 - \exp(-\theta(0)h))^{-1} \mathbf{P}(N_h \geq 2).$$

Utilisant alors l'inclusion de (13.2.3), on obtient

$$0 \leq \mathbf{P}(N_h \geq 2) \leq (1 - \exp(-\theta(0)h)) \mathbf{P}(T_2 < T_1 + h)$$

et donc $\lim_{h \downarrow 0} \frac{\mathbf{P}(N_h \geq 2)}{h} = 0$.

De ce fait, de (13.2.2), on tire

$$\theta(u) = \left[\lim_{h \downarrow 0} \frac{\mathbf{P}(N_h = 1)}{h} \right] (1 - u).$$

Posant $\theta = \lim_{h \downarrow 0} \frac{\mathbf{P}(N_h = 1)}{h}$, on obtient $f_s(u) = \exp(-\theta s(1 - u))$ et l'on reconnaît la fonction génératrice d'une loi de Poisson de paramètre θs . \square

Nous allons maintenant nous intéresser à la première arrivée après un temps déterministe s fixé.

13.2.3 Proposition. Soit $(T_n)_{n \geq 1}$ un processus de Poisson de paramètre θ .

Soit $s \geq 0$. Le temps d'attente après s du premier point du processus est $T_{N_s+1} - s$ et suit une loi exponentielle de paramètre θ . En particulier, T_1 suit une loi exponentielle de paramètre θ .

PREUVE : On a clairement

$$(T_{N_s+1} - s > t) = (T_{N_s+1} > s+t) = (N_{s+t} - N_s = 0).$$

Comme $N_{s+t} - N_s$ suit la même loi que N_t , c'est-à-dire, d'après la Proposition 13.2.2 une loi de Poisson de paramètre θt , on obtient que $\mathbf{P}(T_{N_s+1} - s > t) = e^{-\theta t}$ d'où l'on déduit que $T_{N_s+1} - s$ suit une loi exponentielle de paramètre θ . \square

13.3 Quelques propriétés du processus de Poisson

Introduisons la notion d'échantillon uniforme réordonné. Supposons que X_1, \dots, X_n soient n variables aléatoires indépendantes de loi uniforme sur l'intervalle $]0, t[$. Notons que pour chaque $1 \leq i \leq n$, $1 \leq j \leq n$, $i \neq j$, l'événement $(X_i = X_j)$ est de probabilité nulle, ce qui fait que p.s., les $(X_i)_{1 \leq i \leq n}$ prennent des valeurs différentes : ils sont donc ordonnables en une suite strictement croissante. On pose ainsi

$$X_1^* = \min\{X_i, 1 \leq i \leq n\}, X_2^* = \min\{\{X_i, 1 \leq i \leq n\} \setminus \{X_1^*\}\} \dots$$

et p.s. $X_1^* < X_2^* < X_3^* < \dots < X_n^*$. Un calcul simple montre que la loi de (X_1^*, \dots, X_n^*) admet sur \mathbb{R}^n la densité

$$d_n(s_1, \dots, s_n) = \frac{n!}{t^n} \mathbb{I}_{0 < s_1 < s_2 < \dots < s_n < t}.$$

Cette loi μ_n s'appelle la loi d'un n -échantillon uniforme réordonné. On a le résultat suivant :

13.3.1 Proposition. On considère $(T_n)_{n \geq 1}$ un processus de Poisson. Soient $t > 0$ et $n > 0$ fixés. La loi conditionnelle de (T_1, \dots, T_n) sachant $N_t = n$ est celle d'un n -échantillon uniforme réordonné.

PREUVE : Fixons des réels t_i et h_i tels que

$$0 < t_1 < t_1 + h_1 < t_2 < t_2 + h_2 < \dots < t_n < t_n + h_n < t.$$

On a

$$\begin{aligned} & \mathbf{P}(t_1 < T_1 < t_1 + h_1, t_2 < T_2 < t_2 + h_2, \dots, t_n < T_n < t_n + h_n / N_t = n) \\ &= \mathbf{P}(t_1 < T_1 < t_1 + h_1, t_2 < T_2 < t_2 + h_2, \dots, t_n < T_n < t_n + h_n; N_t = n) / \mathbf{P}(N_t = n). \end{aligned}$$

Or,

$$\begin{aligned} & \mathbf{P}(t_1 < T_1 < t_1 + h_1, t_2 < T_2 < t_2 + h_2, \dots, t_n < T_n < t_n + h_n; N_t = n) \\ &= \mathbf{P}(N_{t_1} = 0, N_{t_1+h_1} - N_{t_1} = 1, N_{t_2} - N_{t_1+h_1} = 0, \dots, N_{t_n+h_n} - N_{t_n} = 1, N_t - N_{t_n+h_n} = 0). \end{aligned} \tag{13.3.1}$$

Cette dernière probabilité peut être calculée sans problème puisqu'elle ne fait intervenir que des accroissements indépendants de la fonction de comptage. Noter que si $u < v$,

$$\mathbf{P}(N_v - N_u = 0) = e^{-\theta(v-u)}, \mathbf{P}(N_v - N_u = 1) = e^{-\theta(v-u)} \theta(v-u).$$

L'expression (13.3.1) est donc égale à

$$e^{-\theta t_1} e^{-\theta h_1} (\theta h_1) e^{-\theta(t_2 - (t_1 + h_1))} e^{-\theta h_2} (\theta h_2) \dots e^{-\theta(t - (t_n + h_n))} = e^{-\theta t} \prod_{i=1}^n h_i \theta^n.$$

Comme $\mathbf{P}(N_t = n) = e^{-\theta t} \theta^n t^n / n!$, on obtient

$$\mathbf{P}(t_1 < T_1 < t_1 + h_1, t_2 < T_2 < t_2 + h_2, \dots, t_n < T_n < t_n + h_n / N_t = n) = \frac{n!}{t^n} \prod_{i=1}^n h_i$$

qui est bien l'expression de la mesure μ_n de $\prod_{i=1}^n]t_i, t_i + h_i[$. □

Un des intérêts de faire l'hypothèse que les temps d'arrivée suivent un processus de Poisson est la très grande stabilité du modèle qui permet aisément d'obtenir des estimations de probabilités. L'exercice suivant étudie un exemple d'étude d'encombrement d'un serveur.

Exercice 13.1

Des clients se présentent aux guichets d'un service (par exemple à la Poste) à des instants donnés par un processus de Poisson de paramètre θ . On suppose que le n -ième client a besoin d'un temps de service donné par une variable Z_n , et on suppose que les Z_n sont indépendantes entre elles ainsi que du flot des arrivées, et de même loi sur \mathbb{R}^+ , admettant une espérance finie (ce qui est raisonnable !). On suppose en outre formellement qu'il y a une infinité de guichets potentiels, ce qui fait que tout client est immédiatement reçu. On se pose la question du nombre v_t de guichets occupés à un instant t , ceci afin naturellement de planifier le nombre raisonnable de guichets à ouvrir.

- Soit t fixé. Calculer la probabilité p_t pour qu'un client arrivé à un temps aléatoire uniformément distribué entre 0 et t soit encore présent au temps t .
- Déduire que conditionnellement à $N_t = n$, v_t suit une loi binomiale de paramètres n et p_t .
- Déduire que v_t suit une loi de Poisson de paramètre $\theta t p_t$. Quelle est la limite en loi de v_t quand t tend vers $+\infty$?

Une propriété centrale des processus de Poisson est qu'ils repartent d'un temps s déterministe pris comme nouvelle origine des temps comme un nouveau processus de Poisson. En effet, si l'on considère $(T_n^s = T_{N_s + n} - s)_{n \geq 0}$, il est facile de vérifier que les conditions de la définition sont satisfaites. Mais en fait, nous allons montrer une propriété beaucoup plus forte, dite propriété de Markov forte : le résultat reste valable si s est remplacé par une variable aléatoire d'un type particulier, un *temps d'arrêt*, que nous allons définir d'une façon qui prolonge le cas discret de la Définition 3.3.3.

13.3.2 Définition. On appelle **temps d'arrêt** du processus de Poisson $(T_n)_{n \geq 1}$ une variable aléatoire positive S telle que pour tout $t \geq 0$, $(S \leq t)$ soit $\sigma(N_s, s \leq t)$ mesurable.

On note $\mathcal{F}_t = \sigma(N_s, s \leq t)$ la filtration naturelle du processus. La propriété d'accroissements indépendants permet alors d'écrire que si $t < u < v$, $N_v - N_u$ est indépendant de \mathcal{F}_t . Quand S est un temps d'arrêt, on note

$$\mathcal{F}_S = \{A \text{ tels que } \forall s \geq 0, A \cap (S \leq s) \in \mathcal{F}_s\}.$$

On vérifie aisément qu'il s'agit d'une tribu dite tribu du passé de S (faire le parallèle avec la Proposition 3.3.5). Par ailleurs, il est clair que si S et T sont deux temps d'arrêt tels que $S \leq T$, $\mathcal{F}_S \subset \mathcal{F}_T$.

On a alors la proposition :

13.3.3 Proposition. Soit $(T_n)_{n \geq 1}$ un processus de Poisson de paramètre θ et soit S un temps d'arrêt. Alors, le processus $(\tilde{T}_n = T_{N_S+n} - S)_{n \geq 1}$ est un processus de Poisson de paramètre θ , indépendant de \mathcal{F}_S .

PREUVE : notons \tilde{N} la fonction de comptage du processus (\tilde{T}_n) . La variable aléatoire \tilde{N}_s compte donc le nombre d'arrivées \tilde{T}_n entre 0 et s , c'est-à-dire le nombre de T_i qui arrivent entre S et $S+s$. On a donc $\tilde{N}_s = N_{S+s} - N_S$.

Commençons par démontrer le résultat quand S est à valeurs dans une suite croissante de réels $(s_j, j \geq 1)$. Soient donc des temps $0 < t_1 < t_2 < \dots < t_\ell$, A un événement de \mathcal{F}_S et n_1, \dots, n_ℓ des entiers. On a

$$\begin{aligned} & \mathbf{P}(A, \tilde{N}_{t_1} = n_1, \tilde{N}_{t_2} - \tilde{N}_{t_1} = n_2, \dots, \tilde{N}_{t_\ell} - \tilde{N}_{t_{\ell-1}} = n_\ell) \\ &= \mathbf{P}(A, N_{S+t_1} - N_S = n_1, N_{S+t_2} - N_{S+t_1} = n_2, \dots, N_{S+t_\ell} - N_{S+t_{\ell-1}} = n_\ell) \\ &= \sum_{j \geq 1} \mathbf{P}(S = s_j, A, N_{s_j+t_1} - N_{s_j} = n_1, N_{s_j+t_2} - N_{s_j+t_1} = n_2, \dots, N_{s_j+t_\ell} - N_{s_j+t_{\ell-1}} = n_\ell) \end{aligned} \quad (13.3.2)$$

Mais comme $A \in \mathcal{F}_S$, $A \cap (S = s_j) \in \mathcal{F}_{s_j}$ et donc les accroissements de N présents dans (13.3.2) sont indépendants de \mathcal{F}_{s_j} . De ce fait, (13.3.2) devient

$$\begin{aligned} & \sum_{j \geq 1} \mathbf{P}(S = s_j, A) \mathbf{P}(N_{s_j+t_1} - N_{s_j} = n_1, N_{s_j+t_2} - N_{s_j+t_1} = n_2, \dots, N_{s_j+t_\ell} - N_{s_j+t_{\ell-1}} = n_\ell) \\ &= \sum_{j \geq 1} \mathbf{P}(S = s_j, A) \mathbf{P}(N_{s_j+t_1} - N_{s_j} = n_1) \dots \mathbf{P}(N_{s_j+t_\ell} - N_{s_j+t_{\ell-1}} = n_\ell) \\ & \quad (\text{en utilisant l'indépendance des accroissements}) \\ &= \sum_{j \geq 1} \mathbf{P}(S = s_j, A) \mathbf{P}(N_{t_1} = n_1) \mathbf{P}(N_{t_2-t_1} = n_2) \dots \mathbf{P}(N_{t_\ell-t_{\ell-1}}) \\ &= \mathbf{P}(A) \mathbf{P}(N_{t_1} = n_1) \mathbf{P}(N_{t_2-t_1} = n_2) \dots \mathbf{P}(N_{t_\ell-t_{\ell-1}}) \end{aligned}$$

et ceci établit donc le résultat cherché.

Supposons maintenant que S soit un temps d'arrêt quelconque. Pour $n \geq 1$, on définit

$$S_n = \sum_{k \geq 0} \frac{k+1}{n} \mathbb{I}_{\frac{k}{n} \leq S < \frac{k+1}{n}}.$$

Il est clair que $S^n \downarrow S$. On a donc si $u < t$, en utilisant la continuité à droite de $r \mapsto N_r$, $\mathbb{I}_{N_{S_n+t} - N_{S_n+u} = k} \rightarrow \mathbb{I}_{N_{S+t} - N_{S+u} = k}$. De ce fait,

$$\begin{aligned} & \lim_{n \rightarrow +\infty} \mathbf{P}(A, N_{S_n+t_1} - N_{S_n} = n_1, N_{S_n+t_2} - N_{S_n+t_1} = n_2, \dots, N_{S_n+t_\ell} - N_{S_n+t_{\ell-1}} = n_\ell) \\ &= \mathbf{P}(A, N_{S+t_1} - N_S = n_1, N_{S+t_2} - N_{S+t_1} = n_2, \dots, N_{S+t_\ell} - N_{S+t_{\ell-1}} = n_\ell). \end{aligned}$$

Notons que S_n est un temps d'arrêt. En effet, si $s \geq 0$ on vérifie les égalités suivantes

$$(S_n \leq s) = \left(\frac{[ns] + 1}{n} \leq s \right) = \left(S \leq \frac{[ns] - 1}{n} \right) \in \mathcal{F}_S$$

puisque $\frac{[ns]-1}{n} \leq s$ (où $[a]$ est la partie entière du réel a). Puisque $S_n \geq S$, on a : $\mathcal{F}_S \subset \mathcal{F}_{S_n}$. Soit $A \in \mathcal{F}_S$. Comme $A \in \mathcal{F}_{S_n}$, et que S_n est à valeurs discrètes,

$$\begin{aligned} & \mathbf{P}(A, N_{S_n+t_1} - N_{S_n} = n_1, N_{S_n+t_2} - N_{S_n+t_1} = n_2, \dots, N_{S_n+t_\ell} - N_{S_n+t_{\ell-1}} = n_\ell) \\ &= \mathbf{P}(A) \mathbf{P}(N_{t_1} = n_1) \mathbf{P}(N_{t_2-t_1} = n_2) \dots \mathbf{P}(N_{t_\ell-t_{\ell-1}}). \end{aligned}$$

Passant à la limite, on obtient donc que

$$\begin{aligned} & \mathbf{P}(A, N_{S+t_1} - N_S = n_1, N_{S+t_2} - N_{S+t_1} = n_2, \dots, N_{S+t_\ell} - N_{S+t_{\ell-1}} = n_\ell) \\ &= \mathbf{P}(A) \mathbf{P}(N_{t_1} = n_1) \mathbf{P}(N_{t_2-t_1} = n_2) \dots \mathbf{P}(N_{t_\ell-t_{\ell-1}}) \end{aligned}$$

et c'est ce qu'on voulait démontrer. □

On obtient alors immédiatement la loi des intervalles entre les arrivées.

13.3.4 Corollaire. Si $(T_n)_{n \geq 1}$ est un processus de Poisson de paramètre θ , les variables $\tau_i = T_i - T_{i-1}$ sont indépendantes et suivent une loi exponentielle de paramètre θ .

PREUVE : On sait déjà d'après la Proposition 13.2.3 que T_1 suit une loi exponentielle de paramètre θ . Comme par la Proposition 13.3.3 appliquée avec le temps d'arrêt T_1 , $(T_n - T_1)_{n \geq 2}$ est aussi un processus de Poisson de paramètre θ , dont le premier temps d'arrivée est $T_2 - T_1 = \tau_2$, cette variable suit une loi exponentielle de paramètre θ , et est indépendante de $T_1 = \tau_1$. On poursuit ainsi par récurrence, et on obtient le résultat cherché. \square

Une très curieuse application de la Proposition 13.3.4 est le *paradoxe de l'autobus*. Supposons que les temps de passages d'autobus à un arrêt donné soient représentés par un processus de Poisson d'intensité θ . L'écart entre deux autobus est alors donné par les variables τ_i , qui suivent une loi exponentielle de paramètre θ . En moyenne cet écart vaut donc $\mathbf{E}(\tau_i) = \frac{1}{\theta}$. De plus, si un passager arrive à un temps s , son temps d'attente suit une loi exponentielle de paramètre θ . Or, si les temps de passage étaient régulièrement espacés de $\frac{1}{\theta}$, le passager arrivant à un instant aléatoire uniformément distribué entre deux autobus attendrait en moyenne $\frac{1}{2\theta}$. Son temps d'attente est donc en moyenne doublé.

13.4 Intégrale stochastique

Nous allons ici définir une intégrale par rapport à un processus de Poisson et étudier ses propriétés. Rappelons d'abord quelques notions sur l'intégrale de Stieltjès. Quand $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction croissante et continue à droite, on peut définir une mesure sur \mathbb{R} , notée df , par l'égalité

$$df([a, b[) = f(b) - f(a).$$

Pour g mesurable bornée, on peut alors donner un sens à l'intégrale $\int_a^b g(s) ds$. Bien entendu, ceci s'étend sans problème au cas où f est réelle à variations bornées (puisqu'elle s'écrit alors comme différence de deux fonctions croissantes), et par extension quand elle est complexe à variations bornées.

Quand f est de classe C^1 , on a $df(ds) = f'(s)ds$ (où ds désigne la mesure de Lebesgue). Quand f est une fonction croissante, nulle en zéro, continue à droite, en escalier dont les sauts valent 1 (ce qui sera le cas des trajectoires d'un processus ponctuel), on vérifie aisément que

$$df(ds) = \sum_{k=1}^{\infty} \delta_{t_k}(ds)$$

où les t_n sont les instants de sauts de f . De plus, si φ est une fonction de \mathbb{N} dans \mathbb{R} (ou dans \mathbb{C}),

$$d[\varphi \circ f](ds) = [\varphi \circ f(s) - \varphi \circ f(s-)]df(ds) \quad (13.4.1)$$

où $f(s-)$ désigne la limite à gauche de f en s .

Si $(N_t)_{t \geq 0}$ est un processus ponctuel, par définition même, les trajectoires $t \mapsto N_t(\omega)$ sont p.s. continues à droite et croissantes. De ce fait, pour tout processus $C_t(\omega)$ tel que p.s. $t \mapsto C_t(\omega)$ soit mesurable et bornée, on sait définir la variable aléatoire

$$I_{a,b}(C) = \int_a^b C_s dN_s$$

à comprendre comme

$$I_{a,b}(C)(\omega) = \int_a^b C_s(\omega) dN_s(\omega) \text{ p.s.}$$

Cette variable aléatoire est dite intégrale stochastique de G par rapport à N . Nous allons voir que si N est un processus de Poisson, pour des processus G bien choisis, l'intégrale stochastique possède d'intéressantes propriétés en tant que variable aléatoire. Commençons par prolonger la définition donnée au Chapitre 12.

13.4.1 Définition. Soit $(M_t)_{t \geq 0}$ un processus stochastique indexé par \mathbb{R}^+ . Notons

$$\mathcal{F}_t = \sigma(M_s, s \leq t).$$

On dit que $(M_t)_{t \geq 0}$ est une martingale si, pour tous $0 \leq s \leq t$,

$$\mathbf{E}(M_t / \mathcal{F}_s) = M_s.$$

Notons qu'une martingale est par définition même un processus **adapté** au sens où pour tout t , M_t est \mathcal{F}_t -mesurable. On étendra par ailleurs naturellement la définition précédente en disant que $(M_t)_{0 \leq t \leq T}$ est une martingale si $\mathbf{E}(M_t / \mathcal{F}_s) = M_s$ est valable dès que $0 \leq s \leq t \leq T$.

Une question importante est de savoir, étant donné un processus de Poisson $(N_t)_{t \geq 0}$, comment il est possible de définir une martingale $(M_t)_{t \geq 0}$ à l'aide de ce processus. Une réponse est donnée par le théorème suivant. À partir d'ici, $(N_t)_{t \geq 0}$ est donc un processus de Poisson de paramètre θ et on pose $\tilde{N}_t = N_t - \theta t$. On pose $\mathcal{F}_t = \sigma(N_s, s \leq t) = \sigma(\tilde{N}_s, s \leq t)$, et les martingales et processus adaptés sont considérés par rapport à cette filtration.

13.4.2 Théorème. Soit $(C_t)_{0 \leq t \leq T}$ un processus adapté, borné et tel que $t \mapsto C_t(\omega)$ soit p.s. continu à gauche. Alors, posant $M_t = \int_0^t C_s d\tilde{N}_s$, $(M_t)_{0 \leq t \leq T}$ est une martingale.

PREUVE : Commençons par supposer que $C_t = \mathbb{I}_{]a,b]}(t) \mathbb{I}_A$ où $A \in \mathcal{F}_a$. Soient $0 \leq s \leq t \leq T$. On a

$$M_t - M_s = \int_s^t C_u d\tilde{N}_u = \mathbb{I}_A \cdot \int_0^t \mathbb{I}_{[s,t] \cap]a,b]}(u) d\tilde{N}_u$$

et, en distinguant les cas de positionnement de a et b par rapport à s et t , on vérifie aisément que $\mathbf{E}(M_t - M_s / \mathcal{F}_s) = 0$. Par exemple, si $a \leq s \leq t \leq b$, $\mathbb{I}_{[s,t] \cap]a,b]} = \mathbb{I}_{[s,t]}$ et donc

$$\int_0^t \mathbb{I}_{[s,t] \cap]a,b]}(u) d\tilde{N}_u = \int_0^t \mathbb{I}_{[s,t]}(u) d\tilde{N}_u = \tilde{N}_t - \tilde{N}_s = (N_t - N_s) - \theta(t - s).$$

Mais $N_t - N_s$ est indépendante de \mathcal{F}_s et $A \in \mathcal{F}_s$. Donc $\mathbf{E}(M_t - M_s / \mathcal{F}_s) = \mathbb{I}_A (\mathbf{E}(N_t - N_s) - \theta(t - s)) = 0$.

Notons que l'ensemble des $]a, b] \times A$, $A \in \mathcal{F}_a$ est clairement un π -système noté \mathcal{L} .

Soit maintenant \mathcal{H} la classe des processus $(C_t)_{0 \leq t \leq T}$ bornés tels que $M_t = \int_0^t C_s d\tilde{N}_s$ soit une martingale, c'est-à-dire

$$\forall 0 \leq s \leq t \leq T, \mathbf{E}(\int_0^t C_u d\tilde{N}_u / \mathcal{F}_s) = \int_0^s C_u d\tilde{N}_u.$$

Il est clair que \mathcal{H} est un espace vectoriel qui contient les constantes, et est stable par limite monotone uniformément bornée. Comme \mathcal{H} contient le π -système \mathcal{L} , d'après le Théorème 2.1.4, il contient les processus bornés $\sigma(\mathcal{L})$ -mesurables.

Il reste donc à vérifier que C_t pris comme dans l'énoncé est $\sigma(\mathcal{L})$ -mesurable. Or, si $(C_t)_{t \geq 0}$ est continu à gauche et adapté, on a

$$C_t = \lim_{n \rightarrow +\infty} \sum_{k=0}^{\infty} C_n^k \mathbb{I}_{] \frac{k}{n}, \frac{k+1}{n}]}(t)$$

et comme C_t est adapté, $C_{\frac{k}{n}}$ est $\mathcal{F}_{\frac{k}{n}}$ -mesurable, et $\sum_{k=0}^{\infty} C_{\frac{k}{n}} \mathbb{I}_{] \frac{k}{n}, \frac{k+1}{n}]}(t)$ est $\sigma(\mathcal{L})$ -mesurable, ainsi que $(t, \omega) \mapsto C_t(\omega)$. \square

Nous allons tirer du Théorème 13.4.2, une importante caractérisation des processus de Poisson parmi les processus ponctuels. Énonçons d'abord un lemme technique sur l'intégrale de Stieltjès.

13.4.3 Lemme. Soient f et g deux fonctions à variations bornées de \mathbb{R}^+ dans \mathbb{R} , continues à droite. Alors

$$f(t)g(t) = f(s)g(s) + \int_s^t f(u)dg(u) + \int_s^t g(u-)df(u). \quad (13.4.2)$$

PREUVE : D'après le théorème de Fubini,

$$(f(t) - f(s))(g(t) - g(s)) = \int_s^t df(u) \int_s^t dg(v) = \int \int_{]s,t]^2} df(u)dg(v). \quad (13.4.3)$$

Posons $A = \{(u, v) \in]s, t]^2, u \leq v\}$ et $B = \{(u, v) \in]s, t]^2, u > v\}$. On a

$$\begin{aligned} \int \int_A df(u)dg(v) &= \int_{]s,t]} \left(\int_{]s,v]} df(u) \right) dg(v) \\ &= \int_{]s,t]} (f(v) - f(s)) dg(v) \\ &= \int_{]s,t]} f(v) dg(v) - f(s)(g(t) - g(s)). \end{aligned}$$

De même,

$$\begin{aligned} \int \int_B df(u)dg(v) &= \int_{]s,t]} \left(\int_{]s,u[} df(u) \right) dg(v) \\ &= \int_{]s,t]} (g(u-) - g(s)) df(u) \\ &= \int_{]s,t]} g(u-) df(u) - g(s)(f(t) - f(s)). \end{aligned}$$

En sommant, on obtient 13.4.2. \square

On peut alors énoncer le théorème de caractérisation suivant, dû à Watanabe dans sa version la plus générale.

13.4.4 Théorème. Soit $(N_t)_{t \geq 0}$ un processus ponctuel. Sont équivalentes :

- (i) $(N_t)_{t \geq 0}$ est un processus de Poisson de paramètre θ ;
- (ii) $(N_t - \theta t)_{t \geq 0}$ est une martingale.

PREUVE : (i) \Rightarrow (ii) : Supposons que $(N_t)_{t \geq 0}$ soit un processus de Poisson de paramètre θ . Soient $s \leq t$. On a $N_t - \theta t = [(N_t - N_s) - \theta(t - s)] + [N_s - \theta s]$. Or, $N_t - N_s$ est indépendante de \mathcal{F}_s , et de loi de Poisson de paramètre $\theta(t - s)$. On a donc

$$\mathbf{E}(N_t - N_s / \mathcal{F}_s) = \mathbf{E}(N_t - N_s) = \theta(t - s)$$

et donc

$$\mathbf{E}((N_t - N_s) - \theta(t - s) / \mathcal{F}_s) = 0.$$

Comme, de plus, $N_s - \theta s$ est \mathcal{F}_s -mesurable,

$$\mathbf{E}(N_t - \theta t / \mathcal{F}_s) = N_s - \theta s.$$

(ii) \Rightarrow (i) : Supposons maintenant que $(N_t - \theta t)_{t \geq 0}$ soit une martingale. Appliquant la formule (13.4.2) à $u \mapsto e^{i\alpha N_u}$ et $u \mapsto e^{-\theta u(e^{i\alpha} - 1)}$, on obtient

$$e^{i\alpha N_t} e^{-\theta t(e^{i\alpha} - 1)} = e^{i\alpha N_s} e^{-\theta s(e^{i\alpha} - 1)} - \int_s^t e^{i\alpha N_u} \theta (e^{i\alpha} - 1) e^{-\theta u(e^{i\alpha} - 1)} du - \int_s^t e^{-\theta u(e^{i\alpha} - 1)} e^{i\alpha N_{u-}} dN_u$$

ce qui se réécrit

$$e^{i\alpha N_t} = e^{i\alpha N_s} e^{\theta(t-s)(e^{i\alpha} - 1)} - \int_s^t (e^{i\alpha} - 1) e^{i\alpha N_{u-}} e^{-\theta(t-u)(e^{i\alpha} - 1)} (dN_u - \theta du)$$

où on a utilisé $d[e^{i\alpha N_u}] = (e^{i\alpha} - 1) e^{i\alpha N_{u-}} dN_u$ (voir (13.4.1)).

L'application $u \mapsto (e^{i\alpha} - 1) e^{i\alpha N_{u-}} e^{-\theta(t-u)(e^{i\alpha} - 1)}$ étant bornée et continue à gauche,

$$\left(\int_s^r (e^{i\alpha} - 1) e^{i\alpha N_{u-}} e^{-\theta(t-u)(e^{i\alpha} - 1)} (dN_u - \theta du) \right)_{s \leq r \leq t}$$

est une martingale par le Théorème 13.4.2. De ce fait,

$$\mathbf{E}(e^{i\alpha N_t} / \mathcal{F}_s) = e^{i\alpha N_s} e^{\theta(t-s)(e^{i\alpha} - 1)}. \quad (13.4.4)$$

Soient alors $0 \leq t_0 < t_1 < \dots < t_n$ et $\alpha_1, \dots, \alpha_n$ des réels, on a

$$\begin{aligned} & \mathbf{E}(e^{i\alpha_1 N_{t_1}} e^{i\alpha_2 (N_{t_2} - N_{t_1})} \dots e^{i\alpha_n (N_{t_n} - N_{t_{n-1}})}) \\ &= \mathbf{E}(e^{i\alpha_1 N_{t_1}} e^{i\alpha_2 (N_{t_2} - N_{t_1})} \dots e^{i\alpha_n (N_{t_{n-1}} - N_{t_{n-2}})} \mathbf{E}(e^{i\alpha_n (N_{t_n} - N_{t_{n-1}})} / \mathcal{F}_{t_{n-1}})) \\ &= \mathbf{E}(e^{i\alpha_1 N_{t_1}} e^{i\alpha_2 (N_{t_2} - N_{t_1})} \dots e^{i\alpha_n (N_{t_{n-1}} - N_{t_{n-2}})}) e^{\theta(t_n - t_{n-1})(e^{i\alpha_n} - 1)} \end{aligned}$$

grâce à (13.4.4). Par conditionnements successifs, on obtient (avec $t_0 = 0$)

$$\mathbf{E}\left(\prod_{k=0}^{n-1} e^{i\alpha_k (N_{t_{k+1}} - N_{t_k})}\right) = \prod_{k=0}^{n-1} e^{\theta(t_{k+1} - t_k)(e^{i\alpha_k} - 1)}$$

ce qui prouve simultanément que les accroissements de N sont indépendants et qu'ils suivent la loi de Poisson attendue. \square

Bibliographie

- [1] J. AUSLANDER, *Minimal flows and their Extensions*. North-Holland Mathematics Studies, 153. North-Holland Publishing Co., Amsterdam, 1988.
- [2] P.BALDI, L.MAZLIAK & P.PRIOURET , *Martingales et Chaînes de Markov*. 3ème édition. Hermann, Paris, 2006
- [3] P. BARBE & M. LEDOUX, *Probabilité*, Belin, Editions espaces 34, Paris, 1998.
- [4] M.BARBUT, B.LOCKER & L.MAZLIAK, *Paul Lévy - Maurice Fréchet. 50 years of mathematical correspondence*, Springer, 2014.
- [5] N.BARTOLI & P.DEL MORAL, *Simulation et Algorithmes stochastiques*, CEPADUES, Toulouse, 2001.
- [6] P. BILLINGSLEY, *Probability and Measure*, 3rd Edition, Wiley Series in Probability and Mathematical Statistics, New-York, 1995.
- [7] N.BOULEAU AND D.LÉPINGLE, *Numerical Methods for Stochastic Processes*, Wiley Series in Probability and Mathematical Statistics, New-York, 1994.
- [8] P.BRÉMAUD, *Point Processes and Queues*, Springer-Verlag, New-York, 1981.
- [9] M.BRIANE & G.PAGÈS, *Intégration*, 4ème Edition, Vuibert, Paris, 2006.
- [10] S.CALLENS, *Les maîtres de l'erreur*, PUF, Paris, 1997.
- [11] E. CHARPENTIER, A. LESNE & N. NIKOLSKI, *L'héritage de Kolmogorov en mathématiques*, Belin, Paris, 2004.
- [12] E.ÇINLAR, *Probability and Stochastics*. Springer, 2011.
- [13] J.DIXMIER, *Topologie générale*, PUF, Paris, 1982
- [14] N. DUNFORD AND J. T. SCHWARTZ, *Linear Operators, Part I, General Theory*. Wiley Classics Library, New-York, 1988
- [15] R. DURRETT, *Probability : Theory and Examples*, Second Edition , Duxbury Press, 1996.
- [16] I.HACKING, *The Emergence of Probability*, Cambridge University Press, 1975.
- [17] Y.LACROIX ET L.MAZLIAK, *Probabilités*. Coll. Mathématiques à l'Université. Ellipses, Paris, 2006.
- [18] E. LESIGNE, *Pile ou Face. Une introduction aux théorèmes limites du calcul des probabilités*, Ellipses, Paris, 2001.
- [19] F. LE LIONNAIS, *Les grands courants de la pensée mathématique*, Hermann, Paris, 1998.
- [20] L.MAZLIAK, *Calcul de probabilités*, Collection Livrets d'Exercice, Hermann, Paris, 1997
- [21] G.PAGÈS AND Y.J.XIAO, Sequences with low Discrepancy and Pseudo-Random Numbers : Theoretical Remarks and Numerical Tests, *Journal of Statistical Computation and Simulation*, 56, pp.163-183, 1997

- [22] B. PASCAL, *Pensées*, texte établi par Louis Lafuma, Seuil, Paris, 1962.
- [23] W. RUDIN, *Real and Complex analysis*, McGraw Hill. , New-York, 1986.
- [24] J. VON PLATO, *Creating modern probability*, Cambridge University Press, Cambridge, 1994.