M1

Yves Lacroix
Laurent Mazliak

Probabilités

Variables aléatoires, convergences, conditionnement
Processus à temps discret

2e édition

ellipses

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 systématiquement 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 seizi^{e} 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^{e} 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^{e} 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^{e} 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^{e} siècle que la mathématisation des notions de probabilités s'accentue 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, 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 [23]) 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 3.21). 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^{e} 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^{e} siècle, ces résultats vont être précisés et augmentés. Les mathématiciens russes de l'école de Saint-Petersbourg, à commencer par son fondateur Pafnuty Lvovich Chebyshev (1821-1894), puis Andrei Andreevich Markov (1856-1922) et Aleksandr Mikhailovitch Lyapunov (1857-1918), jouent un rôle prépondérant en fournissant des démonstrations de la loi (faible) des grands nombres pour laquelle Chebyschev énonça sa célèbre inégalité qui avait été aussi découverte et démontrée antérieurement par le Français Irenée 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 Andrei Nikolaievitch 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 20^{e} siècle. Initiée par les travaux d'Andrei Andreevitch Markov et d'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 !) Andrei Nikolaievitch Kolmogorov, Alexandr Yakovlevitch Khinchin (1894-1959), Serguei 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 et les travaux de Norbert Wiener (1894-1964), le premier à définir en 1923 un modèle mathématique complet (processus de Wiener) pour le phénomène physique découvert en 1827 par le naturaliste écossais Robert Brown et étudié physiquement en 1905 par Albert Einstein. Pour être tout à fait exact, un premier modèle, plus élémentaire, avait été proposé en 1900 par le mathématicien français Louis Bachelier (1870-1946) pour modéliser les fluctuations d'un cours de bourse mais avait été relativement peu remarqué. 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 les a vues intervenir dans de multiples champs du savoir : en physique naturellement, mais aussi en biologie, en économie, en recherche opérationelle 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 [19], [32], [13], [11], [4], [26], [9], [12], [24].

Première partie

Variables aléatoires

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 [10]).
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

$$\begin{array}{ll} \text{(i)} & \emptyset \in \mathcal{E}. \\ \text{(ii)} & A \in \mathcal{E} \Rightarrow A^c \in \mathcal{E}. \\ \text{(iii)} & (A_n)_{n\in\mathbb{N}} \in \mathcal{E} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{E}. \end{array}$$

Le couple $(E,\mathcal{E})$ est dit **ensemble mesurable**.
Quand la condition (iii) est remplacée par la condition beaucoup plus faible

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

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 $f : E \to F$ 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 $\mu:\mathcal{E}\to\mathbb{R}^+\cup\{+\infty\}$ telle que pour toute suite (A_n) d'éléments de E disjoints deux à deux on ait

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

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 qui sera essentiellement démontré dans le chapitre suivant. Une mesure μ sur $(E,\mathcal{E})$ permet de définir l'**intégrale** par rapport à μ d'une fonction mesurable positive $f:(E,\mathcal{E})\to(\mathbb{R},\mathcal{B}(\mathbb{R}))$. 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, on peut écrire $f = f^+ - f^-$ 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 $(X, \mathcal{X}, \mu)$ 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'Egorov)

L'ensemble des fonctions intégrables sur $(E, \mathcal{E}, \mu)$ (ou plus exactement l'ensemble des classes d'équivalence de fonctions intégrables pour la relation d'égalité μ- presque partout, abus qui sera systématiquement sous-entendu par la suite) 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

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

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 $(E, \mathcal{E}, \mu)$ un espace mesuré tel que $\mu(E) = 1$. Soit $\phi : \mathbb{R} \to \mathbb{R}$ une fonction convexe. Pour toute f fonction intégrable telle que $\phi \circ f$ soit intégrable, on a*

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

Exercice 1.2

Soit μ une mesure positive sur $(X,\mathcal{B})$, $f : X \to \mathbb{C}$ 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 $L^p(\mu) \subset L^s(\mu) \cap L^r(\mu)$.

e) Supposons qu'il existe $r > 0$ tel que $\|f\|_r < \infty$. Montrer qu'alors $\|f\|_p \to \|f\|_\infty$.

♢ 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

$$\sqrt{1+A^2} \leq \int_X (\sqrt{1+h^2}) d\mu \leq 1 + A.$$

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 $\|F\|_p^p = -p\int_0^\infty F^{p-1}(x)xF'(x)dx$. Remarquer que $xF' = f - F$, 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 $(E,\mathcal{E},\mu)$. 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 :

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

♢ Exercice 1.7

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

a) Démontrer que $\|f\|_r \leq \|f\|_s$ si $0 < r < s \leq \infty$.

b) A quelles conditions peut-on avoir $0 < r < s \leq \infty$ et $\|f\|_r = \|f\|_s < \infty$?

c) En supposant que $\|f\|_r < \infty$ pour un certain $r > 0$, montrer que

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

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'intervertir 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 $\bigcup_n \uparrow A_n = E$ et $\mu(A_n)<+\infty$).*
Soit $f : E\times F\to\mathbb{R}^+, \mathcal{E}\otimes\mathcal{F}$-mesurable, alors

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

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 $f : E\times F\to\mathbb{R}, \mathcal{E}\otimes\mathcal{F}$-mesurable telle qu'il existe $G : F\to\mathbb{R}$, intégrable sur $(F,\mathcal{F},\nu)$ pour laquelle ν-p.p.,*

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

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

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

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

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

$\hat{\mu}$ 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*

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

On verra au Chapitre 6 une démonstration élémentaire de ce théorème. 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

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

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

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

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 $(\Omega, \mathcal{F})$ telle que $\mathbf{P}(\Omega) = 1$.

Rappelons que les éléments de $\mathcal{F}$ s'appelent 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 $\mathcal{M}$ de parties de Ω est dit une **classe monotone** s'il est stable par réunion dńombrable 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. □

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 $(\Omega, \mathcal{F})$.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 exigente que pour une tribu).

a) Montrer que toute intersection de λ-systèmes est un λ-système.

b) Montrer qu'un π-système qui est aussi un λ-système est une tribu.

c) 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 14.4.

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 $\mathbb{I}_A \in \mathcal{H}$. 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 $\cap_n A_n = \emptyset$, alors $\mathbf{P}(A_n) \to 0$.*

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

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

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

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) :

$$\begin{array}{lll} (i) & : & \mathbf{P}^*(\emptyset) = 0; \\ (ii) & : & \mathbf{P}^* \geq 0 \text{ (positivité)}; \\ (iii) & : & E \subset F \Rightarrow \mathbf{P}^*(E) \leq \mathbf{P}^*(F) \text{ (monotonie)}; \end{array}$$

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

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

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

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

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

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

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

Par récurrence, si $A_1, \ldots A_n \in \mathcal{M}$, 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

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

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 $E = \Omega$. □

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. □

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 $\mathbf{P}(A) \leq \sum_n \mathbf{P}(A_n \cap A) \leq \sum \mathbf{P}(A_n)$, et en passant à l'inf sur ces recouvrements, on obtient $\mathbf{P}(A) \leq \mathbf{P}^*(A)$. □

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 P^* mesure nulle.

2.3 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 dé- finition, 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 $\lambda(A_n) \to \emptyset$. On peut alors poursuivre la contruction 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 $\lambda(A_n) \to 0$. 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 $K_n = K_1^0 \cap \ldots \cap K_n^0$. 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 $n \geq n_0$,

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

et donc on a montré que $\lambda(A_n) \to 0$.
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 $\lambda(I) = \textit{longueur}(I)$ 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 $\mathbb{R}$.
Une autre notion qui prendra une importance capitale dans l'étude des processus stochastiques est celle de complétion d'une mesure.

2.3.1 Définition. *Un espace de probabilités $(\Omega, \mathcal{F}, P)$ est dit* **complet** *si pour tout $B \in \mathcal{F}$ tel que $P(B) = 0$, $A \subset B$ implique que $A \in \mathcal{F}$ (et donc, naturellement, que $P(A) = 0$).*

Signalons sans démonstration le théorème suivant :

2.3.2 Théorème. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilités. Il existe une extension $\mathcal{F}^*$ de la tribu $\mathcal{F}$ et une extension P^* de la mesure P à $\mathcal{F}^*$ telles que P et P^* coïncident sur $\mathcal{F}$ et $(\Omega, \mathcal{F}^*, P^*)$ est complet.*

2.4 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.

2.4.1 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

$$\mathbf{P}(A) = \lim_{N\to+\infty} \downarrow \mathbf{P}\Big(\bigcup_{n\geq N} A_n\Big) \leq \lim_{N\to+\infty} \sum_{n\geq N} \mathbf{P}(A_n) = 0. \qquad \square$$

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

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

2.5 Probabilité conditionnelle et indépendance

Commençons par quelques rappels de notions élémentaires.

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

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

2.5.2 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 $\mathcal{G}$ est indépendant de tout élément de $\mathcal{H}$.

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.

2.5.3 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 $A_j \in \mathcal{B}_j$, $A_i \perp\!\!\!\perp A_j$.

2.5.4 Proposition. *(Borel-Cantelli) Si la suite d'événements (A_n) est indépendante, alors, si $\sum_n \mathbf{P}(A_n) = \infty$ et si A est l'événement "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 $x \geq 0$, $1 - x \leq \mathrm{e}^{-x}$, et donc

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

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 $\mathbf{P}(A) = 1$.
□

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

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

c) Montrer que P est σ-additive sur $\mathcal{A}$ (utiliser le Lemme 2.2.1).

d) Soit $\mathcal{T} = \sigma(\mathcal{A})$. On note $\mathbf{P}_\infty$ 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 $\mathbf{P}_\infty$-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 lors de l'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 $(\Omega, \mathcal{F})$ 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 $(X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ 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'ambiguité 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

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

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

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

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

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

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

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

L'espérance d'un vecteur aléatoire est définie par

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

Pour les vecteurs aléatoires, nous bénéficions des mêmes théorèmes généraux d'interversion 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 par composante, à savoir pour chaque X_i.

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

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

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

$$\mathrm{Cov}(Y,Z) = \mathbf{E}\left((Y-\mathbf{E}(Y))(Z-\mathbf{E}(Z))\right).$$

♢ Exercice 3.1

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

a) Quelle est la dispersion du vecteur aléatoire AX ?

b) Soit X^* le vecteur transposé du vecteur X. Montrer que la variable aléatoire réelle X^*AX admet une espérance égale à $\mathrm{tr}(A)$.

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* $\varepsilon > 0$*,*

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