Introduction aux probabilités

Sacha Braun

October 3, 2025

Ce document est le support écrit du cours d'introduction aux probabilités de la CPES2 Maths / Physique de l'université PSL. Son contenu et son organisation reprennent en partie ceux des notes du cours de 2023/2024 Probabilités 2 de la L2 MIDO de Dauphine par Julien Poisat et José Trashoras, et sont adaptés du cours d'Elise Devey, lui-même adapté du cours de Roman Gambelin. Le cours a été réadapté pour correspondre aux attentes du programme de L2 MIDO de 2025/2026. Il a pour objectif premier de préparer aux enseignements d'intégration, probabilité et statistique de la L3. Il est accompagné de cinq fiches de TD qui viennent compléter les résultats qui seront traités ici.

Table des matières

Un peu de contexte...

Bien que l'existence du dé à six faces (et donc des jeux de hasard) soit attestée depuis le troisième milénaire avant J.C. en Mésopotamie, il faudra attendre le XVI$^{\text{ème}}$ siècle pour que Gerolamo Cardano s'intéresse le premier aux probabilités qui en découlent. Ses écrits constituent le premier ouvrage répertorié traitant de calculs probabilistes. Ce ne sera qu'au siècle suivant qu'une première ébauche de théorie mathématique verra le jour avec les échanges de Blaise Pascal et Pierre de Fermat sur les lancers du même dé (on leur doit notamment le concept d'espérance). La théorie des probabilités est donc une discipline relativement récente dans l'histoire des mathématiques (en comparaison avec l'algèbre ou la géométrie par exemple). Son formalisme actuel, quant à lui, l'est encore plus. D'abord cantonnée aux phénomènes aléatoires finis ou dénombrables (empruntant principalement à la combinatoire), le développement de l'analyse aux XVIII$^{\text{ème}}$ et XIX$^{\text{ème}}$ permit son extension à des phénomènes continus. Cependant, jusqu'au début du XX$^{\text{ème}}$, ces deux branches seront étudiées séparément et le concept même de probabilité reposera sur des définitions formulées dans le language courant. Cette dernière période voit naître une volonté générale d'unification des mathématiques et plusieurs avancées majeures dans nombre de ses disciplines. Parmi elles, les travaux de Henri Lebesgue, René Fréchet, Émile Borel et d'autres sur la théorie de la mesure (dont la paternité est attribuée à Lebesgue) vont révolutionner plusieurs champs des mathématiques, dont le principe d'intégration. C'est dans ce contexte qu'Andrey Kolmogorov, dans son ouvrage de 1933 "Fondations de la Théorie des Probabilités", va établir une base axiomatique des probabilités. Fondée sur la théorie de la mesure, cette dernière est aujourd'hui indisputée parmi les académiciens.

Ce raccrochement formel à une branche fondamentale des mathématiques a permis un développement considérable de la théorie dans la seconde moitié du XX$^{\text{ème}}$ siècle et, surtout, son extension à la plupart des autres disciplines. Dans la recherche contemporaine, on retrouve ainsi l'analyse stochastique, la géométrie aléatoire ou encore la théorie des nombres probabiliste. Ce fleurissement s'est produit conjointement avec une explosion du nombre de champs d'application. Parmi ceux-ci:

- En physique, on retrouve les systèmes de particules en interaction, comme les molécules d'un gaz, ou la mécanique quantique, où l'aléatoire intervient de manière fondamentale. Les probabilités apparaissent aussi dans l'étude de systèmes déterministes mais chaotiques (c'est-à-dire très sensibles aux conditions initiales) et dont l'évolution exacte est donc impossible ou très coûteuse à calculer (les phénomènes météorologiques en sont l'exemple type).
- En biologie, les dynamiques de populations peuvent être modélisées par des processus aléatoires dits "de branchement" et les réseaux de neurones par des systèmes d'équations différentielles stochastiques.
- En finance, l'évolution du prix d'un actif est modélisée par des processus appelés "diffusions". Ces derniers, empruntés à la physique, sont dérivés du mouvement brownien (du botaniste Robert Brown) qui représentait initialement le mouvement d'une particule (en l'occurence de pollen) dans un fluide.
- En informatique, de nombreux algorithmes, dits "de Monte-Carlo", reposent sur la génération de nombres aléatoires. Ils produisent des résultats aléatoires dont la variabilité est contrôlée et souvent compensée par un temps d'éxécution plus court à précision égale.

L'objectif de ce cours est de donner une première introduction à la théorie des probabilités dans sa formulation moderne (basée sur les axiomes de Kolmogorov). Outre son importance mathématique et ses applications sus-citées, son étude a un intérêt qui dépasse le cadre seul des sciences. Les raisonnements probabilistes sont monnaie courante, qu'ils soient quantifiés ou non (par exemple, la lecture de statistiques). Une méconnaissance de la théorie peut donc amener à de mauvaises interprétations ou décisions. En particulier, de nombreux termes techniques ont également un sens dans la vie courante. De par son histoire et son rattachement tardif à une même branche des mathématiques, ceci est d'autant plus vrai pour la théorie des probabilités. Si cette proximité peut aider à comprendre les motivations derrière certaines notions abstraites, elle peut aussi s'avérer source de confusion lorsque l'on s'accroche à un usage personnel, non-expert, et souvent vague ou ambigu, du vocabulaire. En ce sens, l'apprentissage des probabilités de manière axiomatique, à l'image du reste de la science déductive à laquelle elle s'intègre, nous permet de corriger nos biais et d'avoir une base solide quant au sens qu'on donne aux mots. Nous terminons cette brève introduction sur ceux de William Feller, probabiliste clé du siècle dernier, sur ce caractère pluriel des mathématiques qui fait leur attrait:

> "In each field we must carefully distinguish three aspects of the theory: (a) the formal logical content, (b) the intuitive background, and (c) the applications. The character, and the charm, of the whole structure cannot be appreciated without considering all three aspects in their proper relation."
>
> (W. Feller — An Introduction to Probability Theory and its Applications, 1950)

Chapitre 1. Fondamentaux

Ce chapitre a pour but d'introduire les bases de la théorie moderne des probabilités. Comme mentionné dans l'introduction, celle-ci repose sur la théorie de la mesure, dont les notions nécessaires seront donc introduites conjointement. Plusieurs résultats seront admis, un cours général de théorie de la mesure étant hors programme et de niveau L3.

1.1 Introduction informelle aux probabilités sur des univers quelconques

Le programme de terminale et de L1 couvre les probabilités sur des univers finis. Ici, nous allons étendre ces notions aux univers quelconques. Au cours de cette section et au travers de deux exemples, nous motiverons l'introduction de la propritété d'additivité dénombrable pour la mesure de probabilité ainsi que de la notion de tribu.

On suppose disposer d'une pièce de monnaie, éventuellement truquée, avec une probilité $p \in (0, T)$ d'obtenir pile.

1.1.1 Renforcement de l'additivité "finie" en additivité dénombrable

Dans le cas où l'univers est fini, se donner une probabilité $\mathbf{P}$ sur Ω revient à définir $\mathbf{P}$ sur les événements élémentaires (i.e. les singletons). Ainsi, puisque tout événement $A \in \mathscr{P}(\Omega)$ est fini, la propriété d'additivité donne :

$$\mathbf{P}(A) = \sum_{x \in A} \mathbf{P}(\{x\})$$

.

Si l'on se contente de définir une probailité $\mathbf{P}$ par ses valeurs sur les singletons, alors la propritété d'additivité ne permet de calculer $\mathbf{P}(A)$ que pour des parties A finies. Or, si l'univers est infini, il existe nécessairement des parties $A \in \mathscr{P}(\Omega)$ qui sont infinies. Dans le cas des univers dénombrables, on résout ce problème en imposant à $\mathbf{P}$ une propriété d'additivité plus forte : la propriété d'additivité dénombrable. Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements incompatibles, alors on a :

$$\mathbf{P}(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n=0}^{+\infty} \mathbf{P}(A_n)$$

Exemple (Cas d'un univers dénombrable)**.** *On lance la pièce de monnaie jusqu'à l'obtention du premier pile et l'on s'intéresse au nombre de lancers effectués, l'univers naturellement associé à cette expérience aléatoire est $\Omega = \mathbb{N}^*$.*

Les événements élémentaires sont les $\{k\}$: " k lancers ont été effectués". Ils ont pour probabilité :

$$\mathbf{P}(\{k\}) = p(1-p)^{k-1}$$

Puisque Ω est dénombrable, la propritété d'additivité dénombrable permet alors de calculer la probabilité de tout événement $A \subset \mathscr{P}(\Omega)$. Si l'on considère par exemple A : "on a effectué un nombre pair de lancers", on a alors

$$\mathbf{P}(A) = \mathbf{P}(\bigcup_{k\in\mathbb{N}^*} \{2k\}) = \sum_{k=1}^{+\infty} \mathbf{P}(\{2k\}) = \sum_{k=1}^{+\infty} p(1-p)^{2k+1} = \frac{p(1-p)}{1-(1-p)^2}$$

1.1.2 Introduction à la notion de tribu

La situation où Ω n'est pas dénombrable reste problématique. Dans ce cas, la connaissance des $\mathbf{P}(\{x\})$ pour $x \in \Omega$ ne permet d'obtenir $\mathbf{P}(A)$ que pour des parties A au plus dénombrables, ce qui représente une infime partie de $\mathscr{P}(\Omega)$. De plus, il est insuffisant de ne travailler que sur les parties au plus dénombrables, comme le témoigne l'exemple ci-dessous.

Exemple (Cas d'un univers infini non dénombrable)**.** *Deux joueurs jouent à pile ou face, l'un pariant systématiquement sur pile et l'autre sur face. À chaque lancer, le perdant donne 1 euro au gagnant. Les joueurs partent avec chacun une somme initiale et la partie s'arrête lorsque l'un des joueurs est ruiné. Pour modéliser ce jeu, pile sera représenté par 0 et face par 1. Le déroulement d'une partie peut être décrit par la liste des résultats des lancers, c'est à dire par un élément de $\{0,1\}^n$. La longueur n de la partie étant aléatoire, il est impossible de prendre $\{0,1\}^n$ comme univers. Pour surmonter cette difficulté, on envisage une expérience aléatoire plus abstraite : la pièce est lancée une infinité de fois. L'univers choisi est donc $\Omega = \{0,1\}^{\mathbb{N}}$.*

Ainsi, un élément de Ω est une suite infinie de 0 et de 1; et cet événement isolé est observé avec probabilité nulle. Par additivité dénombrable, il s'ensuit que toute partie au plus dénombrable est de probabilité nulle. Si l'on souhaite munir Ω d'une probabilité $\mathbf{P}$ modélisant le jeu, il faut s'intéresser au comportement de $\mathbf{P}$ sur des parties non dénombrables de Ω. Cependant, il a été démontré qu'il n'existe pas de probabilité définie sur $\mathscr{P}(\Omega)$ tout entier permettant de modéliser ce jeu. Il faut donc définir $\mathbf{P}$ sur un sous-ensemble strict $\mathscr{A}$ de $\mathscr{P}(\Omega)$. Ainsi, un événement sont définis comme les parties de Ω appartenant à $\mathscr{A}$.

1.2 Espace probabilisé

1.2.1 Notion de tribu

À chaque théorie ses espaces. L'espace de base en probabilités est l'espace probabilisé. Ce dernier consiste en un ensemble appelé univers des possibles, un ensemble de parties (sous-ensembles) de ce dernier qui représente les événements observables et une application qui attribue à chacun de ces événements une valeur dans $[0,1]$ (appelée probabilité de l'événement). Ainsi, une probabilité est une application qui associe une valeur réelle positive à certains sous-ensembles d'un ensemble. C'est le cas, plus généralement, de ce qu'on appelle une mesure. On impose certaines conditions de stabilité sur l'ensemble des parties mesurables d'un ensemble, appelé tribu.

Définition 1.2.1 (Tribu). *Soit Ω un ensemble quelconque. On appelle tribu sur Ω un ensemble $\mathscr{A}$ de parties de Ω qui satisfait les propriétés suivantes:*

- $\Omega \in \mathscr{A}$.
- *Pour tout $A \in \mathscr{A}$, $A^c := \Omega \backslash A \in \mathscr{A}$ (stabilité par complémentarité).*
- *Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathscr{A}$, $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}$ (stabilité par union dénombrable).*

Remarque. *Les tribus sont également appelées σ-algèbres, d'où l'utilisation standard d'un $\mathscr{A}$ caligraphié pour en désigner une. En anglais, le nom de "σ-field" leur est parfois préféré. L'utilisation d'un $\mathscr{F}$ au lieu d'un $\mathscr{A}$ est donc aussi fréquente. Les termes "algèbre" ou "field" renvoient à la notion de stabilité par certaines opérations (dans le cas d'algèbres d'ensembles, il s'agit de l'union finie et du passage au complémentaire). Le σ (s en grec) est souvent employé pour désigner l'union dénombrable (l'union étant associée à une somme dans la représentation ensembliste de l'algèbre de Boole) et indique donc ici la stabilité par celle-ci.*

On appelle espace mesurable la donnée $(\Omega, \mathscr{A})$ d'un ensemble Ω muni d'une tribu $\mathscr{A}$ sur celui-ci. Une première observation est que puisque $\Omega \in \mathscr{A}$ et $\mathscr{A}$ est stable par complémentarité, $\emptyset \in \mathscr{A}$.

Exemple (Jeu du pile ou face infini). *Pour établir la tribu naturellement associé au jeu décrit dans la section précédente, il convient de se poser la question : quelles parties de Ω constituent naturellement des événements vis-à-vis de cette expérience aléatoire?*

Pour tout $n \in \mathbb{N}^*$, introduisons les événements, contraire l'un de l'autres, associés au n-ème lancer :

$$P_n = \text{"pile au n-ème lancer"} ;$$
$$F_n = \text{"face au n-ème lancer"}.$$

À partir de ces événements, il est naturel d'envisager des réunions, intersections, passage à l'événement contraire.

Exemple. *$\mathscr{A} := \{\emptyset, \{1\}, \{2,5\}, \{3,4\}, \{1,2,5\}, \{1,3,4\}, \{2,3,4,5\}, \Omega\}$ est une tribu sur $\Omega := \{1,2,3,4,5\}$.*

Exemple. *La tribu $\{\emptyset, \Omega\}$ sur un ensemble Ω est appelée tribu triviale (ou grossière).*

Notation. *Pour un ensemble Ω, on note $\mathscr{P}(\Omega)$ l'ensemble des parties de Ω.*

Exemple. *La tribu $\mathscr{P}(\Omega)$ sur un ensemble Ω est appelée tribu discrète.*

Exercice 1.2.1. *Soit $\mathscr{A}$ une tribu sur un ensemble Ω et $A, B \in \mathscr{A}$. Montrer que $A\backslash B \in \mathscr{A}$ et $A\Delta B := (A \cup B)\backslash(A \cap B) \in \mathscr{A}$. On dit que $\mathscr{A}$ est stable par complémentarité relative et différence symétrique.*

En pratique, lorsqu'on travaille avec un ensemble Ω dénombrable, on considère toujours sa tribu discrète.

Exercice 1.2.2. *Soit $(\mathscr{A}_i)_{i\in I}$ une famille quelconque de tribus sur un même ensemble Ω. Montrer que $\bigcap_{i\in I} \mathscr{A}_i$ est une tribu sur Ω.*

Définition 1.2.2 (Tribu engendrée). *Soit Ω un ensemble et $\mathscr{F}$ une famille de parties de Ω. On appelle tribu sur Ω engendrée par $\mathscr{F}$, notée $\sigma(\mathscr{F})$, la plus petite tribu (au sens de l'inclusion) incluant $\mathscr{F}$.*

Dans la définition précédente, $\sigma(\mathscr{F})$ est bien définie car toute intersection de tribus est une tribu (cf. dernier exercice) et on a donc trivialement:

$$\sigma(\mathscr{F}) = \bigcap_{\substack{\mathscr{A} \text{ tribu sur } \Omega \\ \mathscr{F} \subset \mathscr{A}}} \mathscr{A}$$

Notons que l'ensemble des tribus sur Ω est bien défini en tant que sous-ensemble de $\mathscr{P}(\mathscr{P}(\Omega))$.

Exemple. *Soit $\Omega := \{a,b,c,d\}$. Alors $\sigma(\{\{a\}\}) = \{\emptyset, \{a\}, \{b,c,d\}, \Omega\}$.*

La tribu engendrée correspond à la plus petite tribu à considérer lorsqu'on s'intéresse uniquement à une famille d'événements donnée. Dans l'exemple ci-dessus, on souhaite uniquement mesurer le singleton $\{a\}$. Les éléments $\{b, c, d\}$ sont donc en un sens indistinguables, et on pourrait par exemple les remplacer par un élément unique a^c.

Exercice 1.2.3. *Soit Ω un ensemble. Montrer que si Ω est (au plus) dénombrable alors $\sigma(\{\{\omega\} : \omega \in \Omega\}) = \mathscr{P}(\Omega)$. On suppose maintenant Ω non dénombrable. Montrer que l'ensemble suivant est une tribu sur Ω :*

$$\{A \subset \Omega : A \text{ est dénombrable ou } A^c \text{ est dénombrable}\}$$

L'égalité $\sigma(\{\{\omega\} : \omega \in \Omega\}) = \mathscr{P}(\Omega)$ est-elle toujours vraie ?

La tribu suivante est fondamentale et sera le cadre dans lequel se place la plupart du cours.

Définition 1.2.3 (Tribu borélienne)**.** *Soit I un intervalle de $\mathbb{R}$. On appelle tribu borélienne sur I, notée $\mathscr{B}(I)$, la tribu suivante :*

$$\sigma(\{]a, b[: a, b \in I\})$$

Les éléments de $\mathscr{B}(I)$ sont appelés les (ensembles) boréliens de I.

Remarque. *Plus généralement, la tribu borélienne d'un espace topologique est la tribu engendrée par les ouverts de ce dernier. On verra notamment dans la section 3.1 la tribu borélienne d'un espace euclidien de dimension quelconque.*

Exemple. *Par stabilité par union dénombrable, on a que, pour tout $a \in \mathbb{R}$, $]-\infty, a[,]a, +\infty[\in \mathscr{B}(\mathbb{R})$.*

Par définition, une tribu est stable par union dénombrable. En fait, on peut montrer qu'elle est également stable par intersection dénombrable.

Proposition 1.2.4 (Stabilité par intersection dénombrable)**.** *Soit $\mathscr{A}$ une tribu sur un ensemble Ω et $(A_n)_{n\in\mathbb{N}}$ une suite d'élements de $\mathscr{A}$. Alors:*

$$\bigcap_{n\in\mathbb{N}} A_n \in \mathscr{A}$$

Preuve

Pour tout $n \in \mathbb{N}$, on a, par hypothèse, $A_n \in \mathscr{A}$, et donc, par stabilité par complémentarité, $A_n^c \in \mathscr{A}$. Par stabilité par union dénombrable, on en déduit que

$$\bigcup_{n\in\mathbb{N}} A_n^c \in \mathscr{A}$$

et par stabilité par complémentarité encore

$$\left(\bigcup_{n\in\mathbb{N}} A_n^c\right)^c \in \mathscr{A}$$

Or, par la loi de De Morgan, on a:

$$\left(\bigcup_{n\in\mathbb{N}} A_n^c\right)^c = \bigcap_{n\in\mathbb{N}} A_n$$

Remarque. *La propriété de stabilité par intersection dénombrable peut être utilisée pour définir les tribus, en lieu et place de la stabilité par union dénombrable. Cette dernière est en effet démontrable à partir de la première, en reprenant la preuve ci-dessus en sens inverse.*

Exemple. *Pour tout $a \in \mathbb{R}$, on a $\{a\} \in \mathscr{B}(\mathbb{R})$. En effet, $\{a\} = \bigcap_{n\in\mathbb{N}^*}]a - 1/n, a + 1/n[$. Par stabilité par union finie, on en déduit que $\mathscr{B}(\mathbb{R})$ contient tous les intervalles fermés et semi-ouverts (qu'ils soient bornés ou non). En posant, $\mathscr{F} := \sigma(\{[a, b] : a, b \in \mathbb{R}, a < b\})$, on a donc $\mathscr{F} \subset \mathscr{B}(\mathbb{R})$. En effet, $\mathscr{F}$ est, par définition, la plus petite tribu sur $\mathbb{R}$ contenant les intervalles fermés. En particulier, $\mathscr{F} = \mathscr{F} \cap \mathscr{B}(\mathbb{R}) \subset \mathscr{B}(\mathbb{R})$. De même, pour $a, b \in \mathbb{R}$ avec $a < b$, on a $]a, b[= \bigcup_{n>2/(b-a)} [a + 1/n, b - 1/n]$. Par le même raisonnement, on en déduit que $\mathscr{B}(\mathbb{R}) \subset \mathscr{F}$ et donc $\mathscr{B}(\mathbb{R}) = \mathscr{F}$. Comme $\mathscr{B}(\mathbb{R})$ contient tous les singletons, par stabiltié par union dénombrable, $\mathscr{B}(\mathbb{R})$ contient également toutes les parties dénombrables de $\mathbb{R}$. Il est toutefois facilement vérifiable que $\mathscr{B}(\mathbb{R})$ n'est pas engendrée par celles-ci.*

Exercice 1.2.4. *Montrer que $\mathscr{B}(\mathbb{R})$ est engendrée par chacune des familles suivantes:*

- $\{[a, b[: a, b \in \mathbb{R}, a < b\}$

- $\{]a,b] : a,b \in \mathbb{R}, a < b\}$
- $\{]-\infty,a[: a \in \mathbb{R}\}$
- $\{]-\infty,a] : a \in \mathbb{R}\}$
- $\{]a,+\infty[: a \in \mathbb{R}\}$
- $\{[a,+\infty[: a \in \mathbb{R}\}$

Définition 1.2.5 (Tribu trace). *Soient $(\Omega, \mathscr{A})$ un espace mesurable et $\Omega' \subset \Omega$. On définit la tribu trace $\mathscr{A}'$ de $\mathscr{A}$ sur Ω' comme la tribu suivante sur Ω' :*

$$\mathscr{A}' := \{A \cap \Omega' : A \in \mathscr{A}\}$$

Exemple. *Soit E un sous-ensemble dénombrable de $\mathbb{R}$. La trace de $\mathscr{B}(\mathbb{R})$ sur E est $\mathscr{P}(E)$.*

Proposition 1.2.6. *Soient $(\Omega, \mathscr{A})$ un espace mesurable, $\Omega' \subset \Omega$ et $\mathscr{A}'$ la tribu trace de $\mathscr{A}$ sur Ω'. Alors :*

- *Si $\mathscr{A} = \sigma(\mathscr{F})$ pour une famille $\mathscr{F}$ de parties de A, $\mathscr{A}' = \sigma(\{A \cap \Omega' : A \in \mathscr{F}\})$.*
- *Si $\Omega' \in \mathscr{A}$, $\mathscr{A}' \subset \mathscr{A}$.*

Preuve

Voir TD1.

Exemple. *Soit I un intervalle de $\mathbb{R}$. On rappelle que $I \in \mathscr{B}(\mathbb{R})$. Ainsi, par la proposition 1.2.6, la tribu trace de $\mathscr{B}(\mathbb{R})$ sur I est $\mathscr{B}(I)$ et $\mathscr{B}(I) \subset \mathscr{B}(\mathbb{R})$.*

Nous sommes maintenant en position de définir les mesures.

1.2.2 Notion de mesure

Définition 1.2.7 (Mesure). *Soit $(\Omega, \mathscr{A})$ un espace mesurable. On appelle mesure sur $(\Omega, \mathscr{A})$ toute application μ de $\mathscr{A}$ dans $\bar{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{+\infty\}$ telle que :*

- *μ est σ-additive, c'est-à-dire que pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathscr{A}$ deux à deux disjoints,*

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n)$$

- *Il existe $A \in \mathscr{A}$ tel que $\mu(A) < +\infty$ (i.e. $\mu \not\equiv +\infty$).*

Remarque. *Il est courant de voir une mesure comme une application attribuant de la masse à des objets (géométriques par exemple). Avec cette vision, la σ-additivité se lit ainsi : si on décompose un objet en un ensemble au plus dénombrable de ses parties, alors sa masse totale est égale à la somme des masses de ses parties. Cette interprétation est en un sens canonique dans la théorie : si μ est une mesure sur un espace mesurable $(\Omega, \mathscr{A})$ et $B \in \mathscr{A}$, on appelle $\mu(B)$ la masse de B pour la mesure μ. De même, la valeur $\mu(\Omega) \in \bar{\mathbb{R}}_+$ est appelée la masse (totale) de μ.*

La donnée $(\Omega, \mathscr{A}, \mu)$ d'un ensemble Ω, d'une tribu $\mathscr{A}$ sur Ω et d'une mesure μ sur $(\Omega, \mathscr{A})$ est appelée espace mesuré.

Notation. *Pour $m, n \in \mathbb{Z}$, on note $[\![m,n]\!]$ l'ensemble des entiers relatifs de m à n compris (en particulier $[\![m,n]\!] = \emptyset$ si $n < m$). De plus, pour $n \in \mathbb{N}$, on écrit $[\![n]\!] := [\![1;n]\!]$.*

Exercice 1.2.5. *Soit $(\Omega, \mathscr{A}, \mu)$ un espace mesuré. Montrer que $\mu(\emptyset) = 0$. En déduire que pour tout $n \in \mathbb{N}^*$ et toute suite finie $(A_k)_{k\in[\![n]\!]}$ d'éléments de $\mathscr{A}$:*

$$\mu\left(\bigcup_{k\in[\![n]\!]} A_k\right) = \sum_{k=1}^{n} \mu(A_k)$$

Exemple. *Soit $(\Omega, \mathscr{A})$ un espace mesurable. La mesure sur $(\Omega, \mathscr{A})$ qui a pour image $\{0\}$ est appelée mesure triviale.*

Notation. *Soient Ω un ensemble et $A \subset \Omega$, on note $\mathbb{1}_A$ la fonction suivante, appelée fonction indicatrice de A :*

$$\begin{aligned} \mathbb{1}_A : \Omega &\to \{0,1\} \\ \omega &\mapsto \begin{cases} 1, & si\ \omega \in A \\ 0, & si\ \omega \notin A \end{cases} \end{aligned}$$

Exercice 1.2.6. *Soient Ω un ensemble et $A, B \subset \Omega$. Montrer les égalités suivantes :*

- $\mathbb{1}_A \mathbb{1}_B = \min\{\mathbb{1}_A, \mathbb{1}_B\} = \mathbb{1}_{A \cap B}$
- $1 - \mathbb{1}_A = \mathbb{1}_{A^c}$
- $\mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \mathbb{1}_B = \max\{\mathbb{1}_A, \mathbb{1}_B\} = \mathbb{1}_{A \cup B}$
- $(\mathbb{1}_A - \mathbb{1}_B)^2 = \mathbb{1}_{A \Delta B}$

Exemple. *Soit Ω un ensemble, $\omega \in \Omega$ et $\mathscr{A}$ une tribu sur Ω telle que $\{\omega\} \in \mathscr{A}$. La mesure $\delta_\omega : \mathscr{A} \to \bar{\mathbb{R}}_+, A \mapsto \mathbb{1}_A(\omega)$ sur $(\Omega, \mathscr{A})$ est appelée mesure de Dirac en ω (souvent abréviée Dirac en ω).*

Proposition 1.2.8. *Soient $(\mu_i)_{i \in I}$ une famille au plus dénombrable de mesures sur un même espace mesurable $(\Omega, \mathscr{A})$ et $(\alpha_i)_{i \in I}$ une famille de réels positifs. Alors la fonction $\sum_{i \in I} \alpha_i \mu_i$ est une mesure sur $(\Omega, \mathscr{A})$.*

Preuve

Posons

$$\nu := \sum_{i \in I} \alpha_i \mu_i$$

Pour tout $A \in \mathscr{A}$, $\nu(A)$ est une somme (infinie) de termes positifs et est donc bien défini dans $\bar{\mathbb{R}}_+$. Autrement dit, ν est bien définie comme fonction de $\mathscr{A}$ dans $\bar{\mathbb{R}}_+$. De plus, étant donnée une suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de $\mathscr{A}$ deux à deux disjoints, on a :

$$\nu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{i \in I} \alpha_i \mu_i \left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{i \in I} \sum_{n \in \mathbb{N}} \alpha_i \mu_i(A_n) = \sum_{n \in \mathbb{N}} \sum_{i \in I} \alpha_i \mu_i(A_k) = \sum_{n \in \mathbb{N}} \nu(A_n)$$

où la seconde égalité est par σ-additivité de μ_i pour $i \in I$ et la troisième égalité car $\mu_i(A_n) \in \bar{\mathbb{R}}_+$ pour tous $i \in I$ et $n \in \mathbb{N}$, ce qui permet d'inverser l'ordre de sommation. Enfin $\nu(\emptyset) = 0$ trivialement et donc ν est bien une mesure.

Exemple. *Soit $(\Omega, \mathscr{A})$ un espace mesurable tel que $\{\{\omega\} : \omega \in \Omega\} \subset \mathscr{A}$. On appelle mesure de comptage sur $(\Omega, \mathscr{A})$ la mesure suivante:*

$$\begin{aligned} \sum_{\omega \in \Omega} \delta_\omega : \mathscr{A} &\to \bar{\mathbb{R}}_+ \\ A &\mapsto \sum_{\omega \in \Omega} \mathbb{1}_A(\omega) = |A| \end{aligned}$$

La mesure suivante, indissociable de la tribu borélienne, est fondamentale dans le traitement des probabilités continues et, plus généralement, de l'analyse moderne.

Théorème 1.2.9 (Mesure de Lebesgue). *Il existe une unique mesure λ sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, appelée mesure de Lebesgue (sur $\mathbb{R}$), telle que, pour tous $a, b \in \mathbb{R}$ avec $a < b$, on a:*

$$\lambda(]a, b[) = b - a$$

Preuve

Admis. La preuve repose sur un resultat fondamental de théorie de la mesure appelé théorème d'extension de Carathéodory.

Comme on le verra dans les chapitres suivants, il existe également une mesure de Lebesgue sur $\mathbb{R}^n$, pour $n \in \mathbb{N}^*$ quelconque, qui formalise la notion de volume. En dimension $n = 1$, le "volume" correspond à une longueur. De la même manière qu'enlever un point unique à un objet en trois dimensions ne change pas son volume, la mesure de Lebesgue d'un intervalle auquel on enlève un nombre fini (en fait dénombrable) de points ne change pas. Avant de revenir à ce sujet, on montre d'abord des propriétés importantes des mesures.

Proposition 1.2.10. *Soient $(\Omega, \mathscr{A}, \mu)$ un espace mesuré et $(A_n)_{n\in\mathbb{N}}$ une suite croissante d'ensembles mesurables (i.e. pour tout $n \in \mathbb{N}$, $A_n \subset A_{n+1}$). Alors :*

$$\lim_{n\to+\infty} \mu(A_n) = \sup_{n\in\mathbb{N}} \mu(A_n) = \mu\left(\bigcup_{n\in\mathbb{N}} A_n\right)$$

Preuve

Montrons tout d'abord que la suite réelle $(\mu(A_n))_{n\in\mathbb{N}}$ est croissante. Soit $n \in \mathbb{N}$, montrons que $\mu(A_{n+1}) \geq \mu(A_n)$. En effet, on a :

$$\mu(A_{n+1}) = \mu(A_n \cup (A_{n+1}\backslash A_n)) = \mu(A_n) + \mu(A_{n+1}\backslash A_n) \geq \mu(A_n)$$

où la seconde égalité est car A_n et $A_{n+1}\backslash A_n$ sont disjoints et μ est σ-additive et la dernière car μ est positive.

Ainsi, par croissance de la suite, la limite $\lim_{n\to+\infty} \mu(A_n)$ est bien définie. Par croissance encore, on a de plus que $\lim_{n\to+\infty} \mu(A_n) = \sup_{n\in\mathbb{N}} \mu(A_n)$. Pour $n \in \mathbb{N}$, on écrit :

$$A'_n := A_n \backslash \left(\bigcup_{k\in[\![0,n-1]\!]} A_k\right)$$

Par construction, les éléments de la suite $(A'_n)_{n\in\mathbb{N}}$ sont deux à deux incompatibles et tels que, pour tout $n \in \mathbb{N}$:

$$A_n = \bigcup_{k\in[\![0;n]\!]} A'_k$$

En particulier :

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} A'_n$$

Par σ-additivité de μ, on peut donc écrire :

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} A'_n\right) = \sum_{n\in\mathbb{N}} \mu(A'_n) = \lim_{N\to+\infty} \sum_{n=0}^{N} \mu(A'_n) = \lim_{N\to+\infty} \mu\left(\bigcup_{n\in[\![0;N]\!]} A'_n\right) = \lim_{N\to+\infty} \mu(A_N)$$

La proposition ci-dessus nous permet d'étendre des propriétés connues en univers fini à des univers quelconques.

Théorème 1.2.11 (σ-sous-additivité)**.** *Soient $(\Omega, \mathscr{A}, \mu)$ un espace mesuré et $(A_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $\mathscr{A}$. Alors :*

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} \mu(A_n)$$

Lemme 1.2.12. *Soit $(\Omega, \mathscr{A}, \mu)$ un espace mesuré. Alors, pour tout $n \in \mathbb{N}^*$ et famille $(A_k)_{k\in[\![n]\!]}$ d'éléments de $\mathscr{A}$:*

$$\mu\left(\bigcup_{k\in[\![n]\!]} A_k\right) \leq \sum_{k\in[\![n]\!]} \mu(A_k)$$

Preuve

La propriété est trivialement vérifiée pour $n = 1$. Pour les rangs suivant, on procède par récurrence. Pour $n = 2$, on a :

$$\begin{aligned}\mu(A_1 \cup A_2) &= \mu((A_1\backslash A_2) \cup (A_2\backslash A_1) \cup (A_1 \cap A_2)) = \mu(A_1\backslash A_2) + \mu(A_2\backslash A_1) + \mu(A_1 \cap A_2) \\ &\leq \mu(A_1\backslash A_2) + \mu(A_1 \cap A_2) + \mu(A_2\backslash A_1) + \mu(A_1 \cap A_2) = \mu(A_2) + \mu(A_2)\end{aligned}$$

où la seconde et dernière égalité sont par σ-additivité de μ et l'inégalité par positivité de μ. Supposons donc la propriété vraie jusqu'à un rang $n \in \mathbb{N}^*$ arbitraire. En appliquant l'hypothèse deux fois, on obtient alors :

$$\mu\left(\bigcup_{k\in[\![n+1]\!]} A_k\right) = \mu\left(\left(\bigcup_{k\in[\![n]\!]} A_k\right) \cup A_{n+1}\right) \leq \mu\left(\bigcup_{k\in[\![n]\!]} A_k\right) + \mu(A_{n+1}) \leq \sum_{k\in[\![n+1]\!]} \mu(A_k)$$

ce qui conclut la récurrence.

Preuve

[Preuve du théorème 1.2.11] Pour $n \in \mathbb{N}$, posons :

$$F_n := \bigcup_{k \in [\![0;n]\!]} A_n$$

La suite $(F_n)_{n\in\mathbb{N}}$ est trivialement croissante au sens de l'inclusion. De plus :

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} F_n$$

On a donc :

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} F_n\right) = \lim_{n\to+\infty} \mu(F_n) = \lim_{n\to+\infty} \mu\left(\bigcup_{k\in[\![0;n]\!]} A_k\right) \leq \lim_{n\to+\infty} \sum_{k\in[\![0;n]\!]} \mu(A_n) = \sum_{n\in\mathbb{N}} \mu(A_n)$$

où la seconde égalité est par la proposition 1.2.10 et la quatrième par le lemme 1.2.12.

On introduit maintenant un peu de vocabulaire relatif aux mesures.

Définition 1.2.13 (Restriction d'une mesure)**.** *Soient $(\Omega, \mathscr{A}, \mu)$ un espace mesuré et $E \in \mathscr{A}$. On appelle restriction de μ à E, notée $\mu|_E$, la mesure suivante sur $(\Omega, \mathscr{A}, \mu)$:*

$$\begin{aligned} \mathscr{A} &\to \bar{\mathbb{R}}_+ \\ A &\mapsto \mu(A \cap E) \end{aligned}$$

Définition 1.2.14. *[Mesures finies et de probabilité] Soit $(\Omega, \mathscr{A}, \mu)$ un espace mesuré. μ est dite finie si elle est de masse finie ($\mu(\Omega) < +\infty$). Si, de plus, on a $\mu(\Omega) = 1$, on dit que μ est une mesure de probabilité (ou simplement probabilité).*

Remarque. *Les mesures de probabilité sont parfois appelées lois ou distributions, notamment dans certains contextes. On parle par exemple de la loi ou distribution d'une variable aléatoire, mais jamais de sa mesure de probabilité.*

Un espace mesuré $(\Omega, \mathscr{A}, \mathbf{P})$ est appelé espace probabilisé (ou de probabilité) si $\mathbf{P}$ est une mesure de probabilité. Dans ce contexte, les éléments de $\mathscr{A}$ sont apppelés événements. Plus particulièrement, les singletons sont appelés événements élémentaires et $\emptyset$ et Ω sont appelés les événements triviaux. Un événement de probabilité 1 est qualifié de "presque sûr" (on dit notammment qu'il se réalise presque sûrement). Avec ce vocabulaire, les conditions de stabilité d'une tribu se relisent ainsi:
- Si A est un événement, alors "A ne se réalise pas" est un événement.
- Si $(A_i)_{i\in I}$ est une famille dénombrable d'événements, alors "au moins un des $(A_i)_{i\in I}$ se réalise" est un événement.

Exemple. *Trivialement, toute mesure de Dirac est une mesure de probabilité. Elle représente un phénomène déterministe.*

Exemple. *Soit $p \in [0,1]$. La mesure $p\delta_1 + (1-p)\delta_0$ sur $(\{0,1\}, \mathscr{P}(\{0,1\}))$ est une probabilité. Elle est appellée loi de Bernoulli de paramètre p et notée $\mathscr{B}(p)$. Dans le cas $p \in \{0,1\}$, on obtient une Dirac. La loi est utilisée pour modéliser un phénomène où on s'intéresse à la réalisation d'un unique événement (qui arrive avec probabilité p). Prenons l'exemple du lancer d'une pièce de monnaie. On associe le résultat pile à 1 et face à 0. Ce dernier est donc modélisé par une loi de Bernoulli de paramètre $\frac{1}{2}$.*

Exemple. *Soit Ω un ensemble fini. On appelle distribution uniforme sur Ω la mesure de probabilité suivante sur $(\Omega, \mathscr{P}(\Omega))$:*

$$\frac{1}{|\Omega|} \sum_{\omega\in\Omega} \delta_\omega$$

On note cette dernière $\mathscr{U}(\Omega)$. Cette loi modélise le cas fini où tous les événements élémentaires ont la même chance de se produire. La probabilité qu'un événement se produise est ainsi son cardinal divisé par celui de l'univers Ω. L'exemple typique est le tirage d'une carte "au hasard" dans un jeu.

Exemple. *Soit $a, b \in \mathbb{R}$ tels que $a < b$. On appelle distribution uniforme sur $[a,b]$ la mesure de probabilité sur $([a,b], \mathscr{B}([a,b]))$ suivante :*

$$\begin{aligned} \mathscr{B}([a,b]) &\to [0,1] \\ A &\mapsto \frac{\lambda|_{[a,b]}(A)}{b-a} \end{aligned}$$

On note cette dernière $\mathscr{U}([a,b])$. Elle correspond au tirage d'un point "au hasard" dans un intervalle borné. La probabilité qu'il soit tiré dans un sous-intervalle est la proportion de la longueur de celui-ci dans l'intervalle total. Plus généralement, pour $B \in \mathscr{B}(\mathbb{R})$ tel que $\lambda(B) > 0$, on peut définir la distribution uniforme $\mathscr{U}(B)$ sur B par

$$\mathscr{U}(B) := \frac{1}{\lambda(B)} \lambda|_B$$

Exemple. *Soit $(p_n)_{n\in\mathbb{N}}$ une suite positive de somme 1. La mesure suivante définit une probabilité sur $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$:*

$$\sum_{n\in\mathbb{N}} p_n \delta_n$$

En fait, les probabilités sur $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$ sont en bijection avec les suites positives de somme 1. En effet, étant donné une probabilité μ sur $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$, on peut définir la suite $(p'_n)_{n\in\mathbb{N}}$ définie, pour $n \in \mathbb{N}$, par :

$$p'_n := \mu(\{n\})$$

Un exemple fondamental dans la modélisation de phénomènes stochastiques est la loi de Poisson de paramètre $\alpha \in \mathbb{R}_+^*$, notée $\mathscr{P}(\alpha)$. Celle-ci est donnée par la suite $\left(e^{-\alpha}\frac{\alpha^n}{n!}\right)_{n\in\mathbb{N}}$. En terme d'application, la loi de Poisson est utilisée pour modéliser le nombre d'arrivées dans une unité de temps lorsque les arrivées sont considérées indépendantes les unes des autres (par exemple, le nombre de clients visitant un magasin en une heure). Le paramètre représente ici l'intensité (dans le même exemple, plus α est grand, plus le nombre de clients dans une heure sera élevé en moyenne).

Lorsqu'on travaille avec un ensemble dénombrable muni de sa tribu discrète, il suffit, par σ-additivité, de vérifier l'égalité sur chacun des singletons pour comparer deux probabilités. Dans le cas où l'espace mesurable est non dénombrable (comme $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$), il paraît impossible de vérifier l'égalité sur tous les événements de la tribu. En fait, le lemme des classes monotones nous dit qu'il est suffisant de s'intéresser à une sous-famille minimum. Avant d'énoncer ce dernier, nous devons introduire une nouvelle notion de stabilité.

Définition 1.2.15 (π-système). *Soit Ω un ensemble et $\mathscr{S} \subset \mathscr{P}(\Omega)$. On dit que $\mathscr{S}$ est un π-système sur Ω si il est stable par intersection finie, c'est-à-dire :*

$$\forall A, B \in \mathscr{S} : A \cap B \in \mathscr{S}$$

Exemple. *L'ensemble des intervalles bornés et fermés de $\mathbb{R}$ est un π-système sur $\mathbb{R}$.*

Théorème 1.2.16 (Lemme des classes monotones). *Soient $(\Omega, \mathscr{A})$ un espace mesurable, $\mathbf{P}_1$ et $\mathbf{P}_2$ deux probabilités sur $(\Omega, \mathscr{A})$ et $\mathscr{S}$ un π-système sur Ω tel que $\sigma(\mathscr{S}) = \mathscr{A}$. Si, pour tous $A \in \mathscr{S}$, $\mathbf{P}_1(A) = \mathbf{P}_2(A)$, alors $\mathbf{P}_1 = \mathbf{P}_2$.*

Preuve

Admis.

Exemple. *Il est facile de vérifier que l'ensemble $\mathscr{S} := \{(-\infty, x] : x \in \mathbb{R}\} \subset \mathscr{B}(\mathbb{R})$ est un π-système. Or, on sait qu'il engendre $\mathscr{B}(\mathbb{R})$. En conséquence, si deux probabilités $\mathbf{P}_1$ et $\mathbf{P}_2$ sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ satisfont $\mathbf{P}_1((-\infty, x]) = \mathbf{P}_2((-\infty, x])$ pour tout $x \in \mathbb{R}$, alors elles sont égales.*

1.3 Événements et calculs élémentaires

Nous pouvons maintenant définir quelques propriétés élémentaires des événements et formules sur leurs probabilités. La première telle propriété est l'incompatibilité. Moralement, une famille d'événements est dite incompatible si ils ne peuvent se produire tous en même temps, quelque soit la probabilité qu'on regarde sur l'espace.

Définition 1.3.1 (Incompatibilité). *Soit $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé et $(A_i)_{i\in I}$ une famille d'événements.*

- *On dit que les événements sont incompatibles si:*

$$\bigcap_{i\in I} A_i = \emptyset$$

- *On dit que les événements sont deux à deux incompatibles si, pour tous $i, j \in I$ tels que $i \neq j$, on a:*

$$A_i \cap A_j = \emptyset$$

Dans la définition ci-dessus, on voit que la notion d'incompatiblité ne dépend pas de la probabilité donnée sur l'espace (c'est une notion purement ensembliste). Une propriété légèrement plus faible, mais plus pertinente dans le contexte des probabilités, est celle que nous appellerons ici incompatibilité mod 0 (ou "à un événement de probabilité nulle près"). En reprenant les mêmes notations:

- On dit que les événements sont incompatibles mod 0 si

$$\mathbf{P}\left(\bigcap_{i\in I} A_i\right) = 0$$

- On dit que les événements sont deux à deux incompatibles mod 0 si, pour tous $i, j \in I$ tels que $i \neq j$, on a

$$\mathbf{P}(A_i \cap A_j) = 0$$

Moralement, une famille d'événements est incompatible mod 0 si la probabilité qu'ils se réalisent tous en même temps est nulle. L'incompatibilité implique l'incompatiblité mod 0 car $\mathbf{P}(\emptyset) = 0$. La réciproque est fausse car deux événements d'intersection non-vide mais de probabilité nulle sont incompatibles mod 0 mais pas incompatibles. Il s'agit donc d'une notion probabiliste. En pratique, le terme incompatible est des fois utilisé pour désigner une situation où des événements sont uniquement incompatibles mod 0 et le terme incompatible mod 0 n'est jamais utilisé. Dans ce cours, nous ferons toujours la distinction et nous nous pencherons sur cette seconde notion que le temps de quelques exercices. Son intérêt principal est de réaliser que, en fixant une probabilité, la plupart des résultats demandant l'incompatibilité d'événements sont en fait aussi vrais si on suppose uniquement l'incompatibilité mod 0.

Exemple. *Trivialement, si A est un événement, alors A et son complémentaire A^c sont incompatibles.*

Exercice 1.3.1. *Quelle notion d'incompatiblité (générale ou deux à deux) est plus forte ?*

Exemple. *On lance un dé à six faces et considère les événements suivants:*

- *A: On obtient un 1 ou un 2.*
- *B: On obtient un 3 ou un 4.*
- *C: On obtient un 5 ou un 6.*
- *D: On obtient un nombre inférieur ou égal à 3.*
- *E: On obtient un nombre strictement supérieur à 3.*

Cette expérience aléatoire peut être modélisée par l'espace probabilisé $(\Omega, \mathscr{A}, \mathbf{P})$, où $\Omega := \{1,2,3,4,5,6\}$, $\mathscr{A} := \mathscr{P}(\Omega)$ et $\mathbf{P} := \mathscr{U}(\Omega)$. Avec ce formalisme, on a $A := \{1,2\}$, $B := \{2,3\}$, $C := \{5,6\}$, $D := \{1,2,3\}$ et $E := \{4,5,6\}$. On vérifie facilement que les événements A, B et C sont deux à deux incompatibles. De même, les événements A, D et E sont incompatibles. En revanche, on a $A \cap D = A \neq \emptyset$. A, D et E ne sont donc pas deux à deux compatibles.

Exercice 1.3.2. *On considère l'espace probabilisé $([0,1], \mathscr{B}([0,1]), \mathscr{U}([0,1]))$. Donner une famille d'événements deux à deux incompatibles puis donner une famille d'événements incompatibles qui ne sont pas deux à deux incompatibles. Refaire le même exercice mais en remplaçant incompatible par incompatible mod 0 et en considérant uniquement des événements compatibles (qui ne sont pas incompatibles).*

On donne dans la proposition suivante quelques propriétés élémentaires d'une probabilité. Celles-ci, dont la plupart ont déjà été explicitement formulées, découlent directement des propriétés d'une mesure.

Proposition 1.3.2. *Soit $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé. Alors :*

- $\mathbf{P}(\emptyset) = 0$.
- *Pour tout $A \in \mathscr{A}$, on a :*

$$\mathbf{P}(A) \in [0,1]$$

- *Pour tout $A \in \mathscr{A}$, on a :*

$$\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$$

- *Pour tous $A, B \in \mathscr{A}$, on a :*

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$$

- *Pour tous $A, B \in \mathscr{A}$ tels que $A \subset B$, on a :*

$$\mathbf{P}(A) \leq \mathbf{P}(B)$$

- $\mathbf{P}$ *est sous-additive, c'est-à-dire que, pour toute famille $(A_i)_{i \in I}$ au plus dénombrable d'événements, on a :*

$$\mathbf{P}\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} \mathbf{P}(A_i)$$

- *Pour toute famille $(A_i)_{i\in I}$ au plus dénombrable d'événements deux à deux incompatibles, on a :*

$$\mathbf{P}\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} \mathbf{P}(A_i)$$

Preuve

Ce qui n'a pas déjà été prouvé est trivial.

Exemple. *En reprenant l'exemple précédent, on a:*

$$\mathbf{P}(A\cup B\cup C) = \mathbf{P}(A)+\mathbf{P}(B)+\mathbf{P}(C) = \frac{2}{6}+\frac{2}{6}+\frac{2}{6} = 1$$

Ceci est consistant avec la définition d'une mesure de probabilité. En effet, on a $A\cup B\cup C = \Omega$, or $\mathbf{P}(\Omega) = 1$. Une telle partition de Ω (faite d'événements deux à deux incompatibles) est appelée système complet (ou exhaustif) d'événements. De même, on a vu que $A\cap C = A$ (i.e. $A\subset C$) et on vérifie bien que:

$$\mathbf{P}(A) = \frac{1}{3} \leq \frac{1}{2} = \mathbf{P}(D)$$

On peut s'amuser à vérifier les autres formules de la proposition précédente avec ce modèle.

La seconde propriété importante que peut posséder une famille d'événements est l'indépendance. Moralement, des événements sont indépendants si la réalisation de l'un n'impacte pas les chances de réalisation des autres.

Définition 1.3.3 (Indépendance)**.** *Soit $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé et $(A_i)_{i\in I}$ une famille d'événements.*

- *Les événements sont dits mutuellement indépendants (ou simplement indépendants) si pour tout $J\subset I$ fini:*

$$\mathbf{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbf{P}(A_j)$$

- *Les événements sont dits deux à deux indépendants si, pour tous $i,j\in I$ tels que $i\neq j$:*

$$\mathbf{P}(A_i\cap A_j) = \mathbf{P}(A_i)\mathbf{P}(A_j)$$

Remarque. *L'indépendance est en général notée $\perp\!\!\!\perp$ (par exemple "A et B sont indépendants" s'écrit "$A\perp\!\!\!\perp B$"), et plus rarement $\perp$. Ce dernier symbole est ambigu car il désigne également l'orthogonalité, notion distincte et aussi présente en probabilités.*

Exercice 1.3.3. *Combien de conditions faut-il vérifier pour s'assurer de l'indépendance de $n\geq 2$ événements ?*

En principe, la notion d'indépendance qui nous intéresse est la première. Il convient toutefois de remarquer qu'elles ne sont pas équivalentes. En particulier, comme nous le montre l'exemple suivant, la seconde est plus faible.

Exemple. *On lance deux fois une pièce de monnaie et on considère les événements suivants:*

- *A: On obtient pile au premier lancer.*
- *B: On obtient face au deuxième lancer.*
- *C: On obtient la même chose aux deux lancers.*

Cette expérience aléatoire peut être modélisée par l'espace probabilisé $(\Omega, \mathscr{A}, \mathbf{P})$, où $\Omega := \{0,1\}^2$ (où 0 représente pile et 1 face par exemple), $\mathscr{A} := \mathscr{P}(\Omega)$ et $\mathbf{P} := \mathscr{U}(\Omega)$ (pourquoi ?). Avec ce formalisme, on a $A := \{(0,0),(0,1)\}$, $B := \{(0,1),(1,1)\}$, $C := \{(0,0),(1,1)\}$. Par définition de la distribution uniforme, on trouve facilement que:

$$\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = \frac{1}{2}$$
$$\mathbf{P}(A\cap B) = \mathbf{P}(A\cap C) = \mathbf{P}(B\cap C) = \frac{1}{4}$$

Les événements A, B et C sont donc deux à deux indépendants. En revanche, on a:

$$\mathbf{P}(A\cap B\cap C) = \mathbf{P}(\emptyset) = 0 \neq \frac{1}{8} = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$$

On en conclut que A, B et C ne sont pas indépendants.

Exercice 1.3.4. *Montrer qu'une famille d'événements incompatibles mod 0 de probabilités strictement positives n'est pas indépendante.*

Exercice 1.3.5. *Soient A et B deux événements. Montrer que si A est indépendant de B, alors A est indépendant de B^c.*

Comme déjà énoncé, si deux événements ne sont pas indépendants, alors la réalisation de l'un influence la probabilité de l'autre. C'est pour tenir compte du gain d'information que donne la connaissance de la réalisation de l'un que l'on introduit le concept de probabilité conditionnelle.

Définition 1.3.4 (Probabilités conditionnelles)**.** *Soit $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé et $B \in \mathscr{A}$ tel que $\mathbf{P}(B) > 0$. On appelle probabilité conditionnelle à B la probabilité suivante sur $(\Omega, \mathscr{A})$:*

$$\begin{aligned} \mathscr{A} &\to [0,1] \\ A &\mapsto \mathbf{P}(A|B) := \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} \end{aligned}$$

Exercice 1.3.6. *Dans la définition ci-dessus, vérifier que la probabilité conditionelle à B est bien une probabilité sur $(\Omega, \mathscr{A})$.*

On remarque dans la définition ci-dessus que la probabilité conditionelle à B est un facteur (normalisant) de la restriction de $\mathbf{P}$ à B. Ainsi, il est possible de définir la probabilité conditionelle à B sur l'espace mesurable $(B, \mathscr{B})$ où $\mathscr{B}$ est la trace de $\mathscr{A}$ sur B.

Exemple. *Dans l'exemple précédent du dé à six faces, avec les mêmes événements, la probabilité d'obtenir un 1 ou un 2 (événement A) sachant qu'on obtient un nombre inférieur ou égal à 3 (événement D) est:*

$$\mathbf{P}(A|D) = \frac{\mathbf{P}(A \cap D)}{\mathbf{P}(D)} = \frac{|A \cap D|}{|D|} = \frac{|\{1,2\}|}{|\{1,2,3\}|} = \frac{2}{3}$$

Exercice 1.3.7. *Soit $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé et $A, B \in \mathscr{A}$ tels que $\mathbf{P}(B) > 0$. Montrer que A et B sont indépendants si et seulement si $\mathbf{P}(A|B) = \mathbf{P}(A)$.*

Les probabilités conditionnelles permettent de simplifier le calcul de la probabilité d'un événement particulier B si l'on connaît un système complet d'événements pour lesquels les probabilité conditionnelles de B par rapport à chacun d'eux sont connues. Cette relation est donnée par la formule des probabilités totales.

Théorème 1.3.5 (Formule des probabilités totales)**.** *Soit $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé et $(B_i)_{i \in I}$ un système complet d'événements au plus dénombrable. Pour tout événement $A \in \mathscr{A}$, on a:*

$$\mathbf{P}(A) = \sum_{i \in I} \mathbf{P}(A \cap B_i)$$

Si de plus, pour tout $i \in I$, B_i est de probabilité non nulle, alors, pour tout $A \in \mathscr{A}$:

$$\mathbf{P}(A) = \sum_{i \in I} \mathbf{P}(A|B_i)\mathbf{P}(B_i)$$

Preuve

Par définition d'un système complet, on a $\bigcup_{i \in I} B_i = \Omega$ et on peut donc écrire:

$$A = \bigcup_{i \in I} (A \cap B_i)$$

Puisque, pour $i \in I$, $A \cap B_i \subset B_i$ et que, pour $i, j \in I$, $B_i \cap B_j = \emptyset$, on a, pour $i, j \in A$, $(A \cap B_i) \cap (A \cap B_j) = \emptyset$. C'est-à-dire $(A \cap B_i)_{i \in I}$ est une famille d'événements deux à deux incompatibles. En utilisant le dernier point de la proposition 1.3.2, on obtient:

$$\mathbf{P}(A) = \sum_{i \in I} \mathbf{P}(A \cap B_i)$$

Supposons maintenant que, pour tout $i \in I$, $\mathbf{P}(B_i) > 0$. On peut alors définir $\mathbf{P}(A|B_i)$. On remarque (par définition) que, pour $i \in I$, $\mathbf{P}(A \cap B_i) = \mathbf{P}(A|B_i)\mathbf{P}(B_i)$, ce qui conclut la preuve.

Remarque. *Dans la formule des probabilités totales, on peut remplacer système complet par système complet mod 0 (c'est-à-dire dont les événements constitutifs sont deux à deux incompatibles mod 0).*

Exercice 1.3.8 (Loi de succession de Laplace). *On dispose de N+1 urnes, numérotées de 0 à N. L'urne k contient k boules blanches et N-k boules noires.*

On choisit au hasard une urne, puis on réalise des tirages avec remise. Lors de n premiers tirages on a obtenu n boules blanches.

1. *Quelle est la probabilité $p_{N,n}$ de tirer une boule blanche au n+1-ième tirage ?*

2. *Déterminer $\lim_{N\to+\infty} p_{N,n}$*

Ce calcul fut proposé par Laplace pour déterminer la probabilité que le soleil se lève demain ...

Proposition 1.3.6 (Formule de Bayes). *Soient $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé et $A, B \in \mathscr{A}$ deux événements de probabilités non nulles. Alors:*

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

Preuve

Immédiat en remarquant que $\mathbf{P}(B|A)\mathbf{P}(A) = \mathbf{P}(A \cap B)$.

Remarque. *La formule de Bayes est originellement apparue dans des notes non publiées de Thomas Bayes (*XVIII$^{\text{ème}}$ *siècle), à qui elle doit son nom, sur l'évaluation des risques. Bien que triviale, elle a un intérêt épistémologique considérable et est à la source de la statistique dite bayésienne, qui a une interprétation des probabilités différente de celle de l'approche fréquentiste.*

Exemple. *Un virus touche 1% de la population. Un laboratoire fournit un test pour lequel 95% des malades sont positifs et 95% des personnes saines sont négatives. On prend une personne au hasard dans la population et on la teste. On s'intéresse à la probabilité qu'elle soit malade si son test est positif. Cette situation peut être modélisée par l'espace probabilisé $(\Omega, \mathscr{P}(\Omega), \mathbf{P})$ avec $\Omega := \{0,1\}^2$, où la première composante est égale à 1 si la personne est malade et la seconde est égale à 1 si son test est positif. La probabilité $\mathbf{P}$ est elle à déterminer. On écrit $M := \{(1,0),(1,1)\}$ l'événement "la personne est malade", $S := \{(0,0),(0,1)\}$ l'événement "la personne est saine", $P := \{(0,1),(1,1)\}$ l'événement "le test est positif" et $N := \{(0,0),(1,0)\}$ l'événement "le test est négatif". Premièrement, on sait que 1% de la population est malade, ainsi:*

$$\mathbf{P}(M) = 0,01 \qquad \mathbf{P}(S) = \mathbf{P}(M^c) = 1 - \mathbf{P}(M) = 0,99$$

De plus, en traduisant les résultats du laboratoire en termes probabilistes, on obtient:

$$\mathbf{P}(P|M) = 0.95 \qquad \mathbf{P}(N|S) = 0.95$$

Puisque $P = N^c$ et que $\mathbf{P}(\cdot|S)$ est une probabilité, on a:

$$\mathbf{P}(P|S) = \mathbf{P}(N^c|S) = 1 - \mathbf{P}(N|S) = 0.05$$

Par la formule des probabilités totales, on obtient:

$$\mathbf{P}(P) = \mathbf{P}(P|M)\mathbf{P}(M) + \mathbf{P}(P|S)P(S) = 0,95 \times 0,01 + 0,05 \times 0,99 = 0,059$$

Par la formule de Bayes, on en déduit:

$$\mathbf{P}(M|P) = \frac{\mathbf{P}(P|M)\mathbf{P}(M)}{\mathbf{P}(P)} = \frac{0,95 \times 0,01}{0,059} = \frac{95}{590} \approx 0,161$$

En mots, si la personne est testée positive, ses chances d'être malade sont d'approximativement 16,1%. C'est donc un test très sensible. Cette valeur peut sembler surprenante au premier abord. En effet, la plupart des gens malades sont testés positifs, la plupart des gens sains sont testés négatifs, et pourtant, la plupart des gens testés positifs sont négatifs. Ceci est dû à la particulièrement faible probabilité d'être malade.

Dans le cas spécifique d'une probabilité, la proposition 1.2.10 admet également une version pour suites décroissantes. Ensemble, les deux versions sont souvent désignées sous le nom de théorème de la limite monotone en probabilités élémentaires. Après avoir énoncé ce dernier, on donnera un exemple de son utilisation dans un cas concret de modélisation puis dans la preuve d'un autre résultat fondamental en probabilités.

Proposition 1.3.7 (Théorème de la limite monotone). *Soit $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé et $(A_n)_{n\in\mathbb{N}}$ une suite d'événements.*

- *Si $(A_n)_{n\in\mathbb{N}}$ est croissante (i.e. pour tout $n \in \mathbb{N}$, $A_n \subset A_{n+1}$), on a :*

$$\lim_{n\to+\infty} \mathbf{P}(A_n) = \sup_{n\in\mathbb{N}} \mathbf{P}(A_n) = \mathbf{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right)$$

- *Si $(A_n)_{n\in\mathbb{N}}$ est décroissante (i.e. pour tout $n \in \mathbb{N}$, $A_{n+1} \subset A_n$), on a :*

$$\lim_{n\to+\infty} \mathbf{P}(A_n) = \inf_{n\in\mathbb{N}} \mathbf{P}(A_n) = \mathbf{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right)$$

Preuve

Le résultat pour $(A_n)_{n\in\mathbb{N}}$ croissante a déjà été démontré dans la proposition 1.2.10. Supposons donc $(A_n)_{n\in\mathbb{N}}$ décroissante. Par la loi de De Morgan, on a :

$$\mathbf{P}\left(\left(\bigcap_{n\in\mathbb{N}} A_n\right)^c\right) = \mathbf{P}\left(\bigcup_{n\in\mathbb{N}} A_n^c\right)$$

Puisque $(A_n)_{n\in\mathbb{N}}$ est décroissante, $(A_n^c)_{n\in\mathbb{N}}$ est croissante. En utilisant la proposition 1.2.10, on obtient :

$$\mathbf{P}\left(\left(\bigcap_{n\in\mathbb{N}} A_n\right)^c\right) = \lim_{n\to+\infty} \mathbf{P}(A_n^c) = \lim_{n\to+\infty} (1 - \mathbf{P}(A_n)) = 1 - \lim_{n\to+\infty} \mathbf{P}(A_n)$$

Puisque

$$\mathbf{P}\left(\left(\bigcap_{n\in\mathbb{N}} A_n\right)^c\right) = 1 - \mathbf{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right)$$

on en déduit :

$$\mathbf{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty} \mathbf{P}(A_n)$$

Enfin, par décroissance de $(A_n)_{n\in\mathbb{N}}$ (au sens de l'inclusion), la suite $(\mathbf{P}(A_n))_{n\in\mathbb{N}}$ est décroissante (au sens des réels) et donc $\lim_{n\to+\infty} \mathbf{P}(A_n) = \inf_{n\in\mathbb{N}} \mathbf{P}(A_n)$.

Avant de passer à l'exemple suivant, nous invitons le lecteur non familier avec la cardinalité des ensembles infinis à lire l'annexe B.

Exemple. *On lance un dé à six faces une infinité de fois et on s'intéresse à la probabilité d'obtenir au moins une fois un 6. Pour modéliser cette expérience, il est naturel de considérer l'univers $\Omega := [\![6]\!]^{\mathbb{N}^*}$ (l'ensemble des suites indexées par $\mathbb{N}^*$ à valeurs dans $[\![6]\!]$). Il est clair que Ω n'est pas dénombrable et il n'est donc pas judicieux de l'équiper de sa tribu discrète (confère annexe B). En effet, d'après le théorème de Cantor, on a $Card(\mathscr{P}(\Omega)) > Card(\Omega) = Card(\mathbb{R})$. On souhaite au minimum pouvoir mesurer les ensembles $E_{n,k} := \{(x_i)_{i\in\mathbb{N}^*} \in \Omega : x_n = k\}$ pour $n \in \mathbb{N}^*$ et $k \in [\![6]\!]$ ($E_{n,k}$ est l'événement "le $n^{ième}$ lancé est un k") et on considère donc la tribu $\mathscr{A} = \sigma(\{E_{n,k} : n \in \mathbb{N}^*, k \in [\![6]\!]\})$. Soit $\mathscr{S}$ l'ensemble des parties de Ω qui peuvent s'écrire comme intersections finies des $E_{n,k}$ ($\mathscr{S}$ est le π-système engendré par la famille $(E_{n,k})_{n\in\mathbb{N}^*,k\in[\![6]\!]}$). Par définition, $\mathscr{S}$ est un π-système qui engendre $\mathscr{A}$. En effet, $E_{n,k} \in \mathscr{A}$ pour tous $n \in \mathbb{N}^*, k \in [\![6]\!]$ (par définition de $\mathscr{A}$) et, par stabilité par intersection dénombrable d'une tribu, $\mathscr{S} \subset \mathscr{A}$. Par le lemme des classes monotones, il est donc suffisant de définir une probabilité sur les éléments de $\mathscr{S}$ pour la caractériser. Pour tout événement $F \in \mathscr{S}$ non vide, il existe $N \in \mathbb{N}^*$, $n_1, \dots, n_N \in \mathbb{N}^*$ distincts et $k_1, \dots, k_n \in [\![6]\!]$ tels que $F = \bigcap_{i=1}^N E_{n_i,k_i}$. On définit la probabilité $\mathbf{P}$ sur $(\Omega, \mathscr{A})$ qui à un tel $F \in \mathscr{S}$ associe la probabilité:*

$$\mathbf{P}(F) = \left(\frac{1}{6}\right)^N$$

Cette probabilité est appropriée car la loi d'un lancer est uniforme et les lancers sont supposés indépendants. Maintenant que le problème est bien formulé mathématiquement, on peut poser la probabilité de l'événement E: "le 6 n'est jamais obtenu". On obtient jamais 6 si, pour chaque $n \in \mathbb{N}^*$, on n'obtient pas 6. Ainsi, l'événement E s'écrit:

$$E := \bigcap_{n\in\mathbb{N}^*} (E_{n,6})^c$$

Pour $N \in \mathbb{N}^*$, notons $E_N := \bigcap_{n=1}^N (E_{n,6})^c$. La suite $(E_N)_{N\in\mathbb{N}^*}$ est décroissante au sens de l'inclusion et on a

$$E = \bigcap_{N\in\mathbb{N}^*} E_N$$

Par le théorème de la limite monotone, on en déduit:

$$\mathbf{P}(E) = \mathbf{P}\left(\bigcap_{N\in\mathbb{N}^*} E_N\right) = \lim_{N\to+\infty} \mathbf{P}(E_N)$$

Or, pour $N \in \mathbb{N}^*$:

$$\mathbf{P}(E_N) = \mathbf{P}\left(\bigcap_{n=1}^{N} (E_{n,6})^c\right) = \prod_{n=1}^{N} \mathbf{P}((E_{n,6})^c) = \prod_{n=1}^{N} \frac{5}{6} = \left(\frac{5}{6}\right)^N$$

où la seconde égalité est par indépendance de la famille $((E_{n,6})^c)_{n\in[\![N]\!]}$ (ne pas obtenir un 6 à un lancer n'influence pas la probabilité de ne pas en obtenir aux autres). En particulier, $\lim_{N\to+\infty} \mathbf{P}(E_N) = 0$. On en déduit $\mathbf{P}(E) = 0$ et donc $\mathbf{P}(E^c) = 1$. C'est-à-dire on obtient presque sûrement au moins un 6.

Théorème 1.3.8 (Inégalité de Boole)**.** *Soient $(\Omega, \mathscr{A}, \mathbf{P})$ un espace mesuré et $(A_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $\mathscr{A}$. Alors :*

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

Preuve

Ce résultat est exactement le théorème 1.2.11 appliqué à une mesure de probabilité.

Remarque. *Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements négligeables (i.e. $\forall n \in \mathbb{N}, \mathbf{P}(A_n) = 0$), alors $\bigcup_{n\in\mathbb{N}} A_n$ est négligeable.*

Exercice 1.3.9. *Considérons le jeu de pile ou face infini. Montrer que l'événement "Obtenir un nombre fini de piles" est négligeable.*

(indice : introduire l'événement B_p "N'obtenir que des faces à partir du rang p")

Définition 1.3.9 (Limite supérieure)**.** *Soit $(\Omega, \mathscr{A})$ un espace mesurable et $(A_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $\mathscr{A}$. On appelle limite supérieure de $(A_n)_{n\in\mathbb{N}}$, notée $\overline{\lim}_{n\to\infty} A_n$, l'élément de $\mathscr{A}$ suivant:*

$$\overline{\lim_{n\to\infty}} A_n := \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements, alors $\overline{\lim}_{n\to\infty} A_n$ correspond est l'événement "une infinité de A_n se réalisent".

Le résultat suivant est d'une grande utilité car il donne une condition suffisante pour s'assurer que seulement un nombre fini d'événements se réalisent parmis une suite donnée.

Corollaire 1.3.10 (Lemme de Borel-Cantelli)**.** *Soit $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé et $(E_n)_{n\in\mathbb{N}}$ une suite d'événements telle que :*

$$\sum_{n\in\mathbb{N}} \mathbf{P}(E_n) < +\infty$$

Alors, presque sûrement, seulement un nombre fini de ces événements se réalisent. C'est-à-dire :

$$\mathbf{P}\left(\overline{\lim_{n\to\infty}} E_n\right) = 0$$

Preuve

Pour $n \in \mathbb{N}$, posons $U_n = \bigcup_{k=n}^{\infty} E_k$. La suite $(U_n)_{n\in\mathbb{N}}$ est décroissante et on a, par définition :

$$\overline{\lim_{n\to\infty}} E_n = \bigcap_{n=0}^{\infty} U_n$$

Par le théorème de la limite monotone, on a donc :

$$\mathbf{P}\left(\overline{\lim_{n\to\infty}} E_n\right) = \lim_{n\to+\infty} \mathbf{P}(U_n)$$

Or, par σ-sous-additivité de $\mathbf{P}$, on a, pour tout $n \in \mathbb{N}$:

$$\mathbf{P}(U_n) = \mathbf{P}\left(\bigcup_{k=n}^{\infty} E_k\right) \leq \sum_{k=n}^{+\infty} \mathbf{P}(E_k)$$

Par convergence de la série de terme général $\mathbf{P}(E_k)$ $(k \in \mathbb{N})$, la suite des restes $\left(\sum_{k=n}^{+\infty} \mathbf{P}(E_k)\right)_{n\in\mathbb{N}}$ converge nécessairement vers 0. Par l'inégalité ci-dessus, on en conclut:

$$\lim_{n\to+\infty} \mathbf{P}(U_n) = 0$$

Nous verrons des applications du lemme de Borel-Cantelli plus tard. Pour l'instant, nous nous contenterons de donner une réciproque partielle à ce dernier, illustrée d'un exemple.

Corollaire 1.3.11 (Second lemme de Borel-Cantelli)**.** *Soit $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé et $(E_n)_{n\in\mathbb{N}}$ une famille d'événements indépendants tels que:*

$$\sum_{n\in\mathbb{N}} \mathbf{P}(E_n) = +\infty$$

Alors, presque sûrement, une infinité de ces événements se réalisent. C'est-à-dire:

$$\mathbf{P}\left(\overline{\lim_{n\to\infty}} E_n\right) = 1$$

Preuve

On veut montrer que $1 - \mathbf{P}\left(\overline{\lim_{n\to\infty}} E_n\right) = 0$. On a:

$$1 - \mathbf{P}\left(\overline{\lim_{n\to\infty}} E_n\right) = 1 - \mathbf{P}\left(\bigcap_{n=0}^{\infty}\bigcup_{k=n}^{\infty} E_k\right) = \mathbf{P}\left(\left(\bigcap_{n=0}^{\infty}\bigcup_{k=n}^{\infty} E_k\right)^c\right)$$

Par la loi de De Morgan, on obtient:

$$1 - \mathbf{P}\left(\overline{\lim_{n\to\infty}} E_n\right) = \mathbf{P}\left(\bigcup_{n=0}^{\infty}\bigcap_{k=n}^{\infty} E_k^c\right)$$

Pour $n \in \mathbb{N}$, posons $I_n := \bigcap_{k=n}^{\infty} E_k^c$. La suite $(I_n)_{n\in\mathbb{N}}$ est croissante et donc, par le théorème de la limite monotone, on en déduit:

$$1 - \mathbf{P}\left(\overline{\lim_{n\to\infty}} E_n\right) = \lim_{n\to\infty} \mathbf{P}(I_n)$$

Or, pour $n \in \mathbb{N}$, on a par indépendance de la famille $(E_k)_{k\in\mathbb{N}}$ (et donc $(E_k^c)_{k\in\mathbb{N}}$):

$$\mathbf{P}(I_n) = \mathbf{P}\left(\bigcap_{k=n}^{\infty} E_k^c\right) = \prod_{k=n}^{+\infty} \mathbf{P}(E_k^c) = \prod_{k=n}^{+\infty} (1 - \mathbf{P}(E_k))$$

En utilisant l'inégalité $1 - x \leq e^{-x}$ pour tout $x \in \mathbb{R}$, on obtient, pour $n \in \mathbb{N}$:

$$\mathbf{P}(I_n) \leq \prod_{k=n}^{+\infty} e^{-\mathbf{P}(E_k)} = e^{-\sum_{k=n}^{+\infty} \mathbf{P}(E_k)} = 0$$

où la derniere égalité est par hypothèse sur la suite $(\mathbf{P}(E_k))_{k\in\mathbb{N}}$. On en déduit $\mathbf{P}\left(\overline{\lim_{n\to\infty}} E_n\right) = 1$.

Exemple. *Dans le dernier exemple, avec les mêmes notations, on considère l'événement $\overline{\lim_{n\to+\infty}} E_{n,6}$. La famille $(E_{n,6})_{n\in\mathbb{N}^*}$ est indépendante et la série de terme général $\mathbf{P}(E_{n,6})$ $(n \in \mathbb{N}^*)$ est clairement divergente. Par le second lemme de Borel-Cantelli, on en déduit:*

$$\mathbf{P}\left(\overline{\lim_{n\to+\infty}} E_{n,6}\right) = 1$$

Autrement dit, on obtient presque sûrement une infinité de 6.

Exercice 1.3.10. *Soit $n \in \mathbb{N}^*$. Dans l'exemple précédent, montrer que l'événement "n 6 sont obtenus consécutivement" se réalise presque sûrement une infinité de fois.*

Chapitre 2. Variable aléatoire

2.1 Fonction mesurable et variable aléatoire

Nous avons vu qu'un espace probabilisé représente (ou peut représenter) un phénomène aléatoire. En particulier, il donne une probabilité à l'ensemble des réalisations possibles de ce dernier. Dans certaines situations, un phénomène aléatoire peut découler d'un autre, les réalisations de l'un pouvant être fonction des réalisations de l'autre. De même, étant donné un espace probabilisé, il est possible d'en créer un autre en appliquant une transformation aux réalisations du premier. Prenons (encore) l'exemple du lancer de dé à 6 faces, modélisé par l'espace probabilisé $(\llbracket n \rrbracket, \mathscr{P}(\llbracket n \rrbracket), \mathbf{P} := \mathscr{U}(\llbracket n \rrbracket))$. Informellement, on note X le résultat que l'on obtient. On s'intéresse à la distribution de $f(X)$ ou $f : \llbracket 6 \rrbracket \to \{0,1\}, x \mapsto \mathbb{1}_{\llbracket 3 \rrbracket}(x)$ (en mots, $f(X)$ vaut 1 si le résultat du dé est inférieur ou égal à 3, 0 sinon). Clairement, $f(X)$ ne peut prendre que deux valeurs (0 et 1). De plus, l'événement A:"$f(X) = 1$" peut se reformuler "$X \in \{1,2,3\}$" et s'écrire mathématiquement $A := \{1,2,3\} = f^{-1}(\{1\})$. Ainsi, la probabilité que "$f(X) = 1$" est $\mathbf{P}(\llbracket 3 \rrbracket) = \frac{1}{2}$. On voit que le résultat de l'expérience aléatoire $f(X)$ où X est un lancer de dé est modélisé par l'espace probabilisé $(\{0,1\}, \mathscr{P}(\{0,1\}), \mathbf{P}' := \mathscr{B}(1/2))$. On observe ainsi que la loi $\mathbf{P}'$ est donnée par:

$$\mathbf{P}'(\{0\}) = \mathbf{P}(f^{-1}(\{0\})) \qquad \mathbf{P}'(\{1\}) = \mathbf{P}(f^{-1}(\{1\}))$$

C'est-à-dire, $\mathbf{P}' = \mathbf{P} \circ f^{-1}$. Le but des variables aléatoires est de donner un sens rigoureux à X et de formaliser et systématiser les calculs que l'on vient de faire. Pour ce faire, la première étape est de définir les transformations f autorisées. Comme nous venons de le voir, les antécédents des événements de l'espace d'arrivée (ou "construit") par f doivent être eux même mesurables. Une application entre deux espaces mesurables qui satisfait cette condition est dite mesurable.

Définition 2.1.1 (Fonction mesurable). *Soient $(\Omega_1, \mathscr{A}_1)$ et $(\Omega_2, \mathscr{A}_2)$ deux espaces mesurables. On dit qu'une fonction $f : \Omega_1 \to \Omega_2$ est mesurable si, pour tout $A \in \mathscr{A}_2$, $f^{-1}(A) \in \mathscr{A}_1$. Si de plus $(\Omega_2, \mathscr{A}_2) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, on dit que f est une fonction mesurable réelle.*

Exercice 2.1.1. *Soient $(\Omega, \mathscr{P}(\Omega))$ et $(\mathbb{X}, \mathscr{F})$ deux espaces mesurables. Montrer que toute fonction $f : \Omega \to \mathbb{X}$ est mesurable.*

Exercice 2.1.2. *Soient $(\Omega, \mathscr{A})$ et $(\mathbb{X}, \{\emptyset, \mathbb{X}\})$ deux espaces mesurables. Montrer que toute fonction $f : \Omega \to \mathbb{X}$ est mesurable.*

Exemple. *Soit $(\Omega, \mathscr{A})$ un espace mesurable et $A \in \mathscr{A}$. Alors la fonction suivante est mesurable pour la tribu borélienne:*

$$\mathbb{1}_A : \Omega \to \mathbb{R}$$
$$x \mapsto \begin{cases} 1, & \text{si } x \in A \\ 0, & \text{si } x \notin A \end{cases}$$

En effet, soit $B \in \mathscr{B}(\mathbb{R})$. Si $\{0,1\} \subset B$ alors $\mathbb{1}_A^{-1}(B) := \{x \in \Omega : \mathbb{1}_A(x) \in B\} = \Omega \in \mathscr{A}$. Sinon, si $1 \in B$ mais $0 \notin B$, alors $\mathbb{1}_A^{-1}(B) = A \in \mathscr{A}$. De même, si $1 \notin B$ mais $0 \in B$, alors $\mathbb{1}_A^{-1}(B) = A^c$. Or, par stabilité par complémentarité d'une tribu, $A^c \in \mathscr{A}$ donc $\mathbb{1}_A^{-1}(B) \in \mathscr{A}$. Enfin, si $1 \notin B$ et $0 \notin B$, alors $\mathbb{1}_A^{-1}(B) = \emptyset \in \mathscr{A}$. On a donc montré que, pour tout $B \in \mathscr{B}(\mathbb{R})$, $\mathbb{1}_A^{-1}(B) \in \mathscr{A}$, c'est-à-dire $\mathbb{1}_A$ est mesurable. Avec les mêmes arguments, on peut montrer que $\mathbb{1}_A$ est mesurable lorsqu'on considère $(\{0,1\}, \mathscr{P}(\{0,1\}))$ comme espace d'arrivée au lieu de $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Exercice 2.1.3. *Soient $(\Omega, \mathscr{A})$ et $(\mathbb{X}, \mathscr{F})$ deux espaces mesurables et $\mathscr{F}'$ une tribu sur $\mathbb{X}$ telle que $\mathscr{F}' \subset \mathscr{F}$. Montrer que toute fonction mesurable f de $(\Omega, \mathscr{A})$ dans $(\mathbb{X}, \mathscr{F})$ l'est également pour la tribu $\mathscr{F}'$. Fixons une telle fonction f et supposons maintenant que $\mathscr{F}'$ est une tribu sur un sous-ensemble $\mathbb{X}'$ de $\mathbb{X}$ tel que $f(\Omega) \subset \mathbb{X}'$ qui satisfait toujours $\mathscr{F}' \subset \mathscr{F}$. Montrer que f reste mesurable si on remplace son espace d'arrivée par $(\mathbb{X}', \mathscr{F}')$.*

De l'exercice précédent et la proposition 1.2.6, on a en particulier qu'une fonction mesurable avec espace d'arrivée $(\mathbb{X}, \mathscr{F})$ et image inclue dans un sous-ensemble $\mathbb{X}' \in \mathscr{F}$ est également mesurable pour l'espace d'arrivée $(\mathbb{X}', \mathscr{F}')$, où $\mathscr{F}'$ est la tribu trace de $\mathscr{F}$ sur $\mathbb{X}'$.

Proposition 2.1.2. *Soient $(\Omega_i, \mathscr{A}_i)_{i=1}^3$ une famille d'ensemble mesurables et $f : \Omega_1 \to \Omega_2$ et $g : \Omega_2 \to \Omega_3$ deux fonctions mesurables. Alors la fonction composée $g \circ f : \Omega_1 \to \Omega_3$ est mesurable.*

Preuve

Soit $A \in \mathscr{A}_3$. Par mesurabilité de g, on a $g^{-1}(A) \in \mathscr{A}_2$. Or $(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A))$. Par mesurabilité de f, on en en déduit $(g \circ f)^{-1}(A) \in \mathscr{A}_1$, c'est-à-dire $g \circ f$ est mesurable.

Remarque. *Par récurrence, toute composition de fonctions mesurables est mesurable.*

Définition 2.1.3 (Convergence simple). *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions réelles définies sur un même ensemble $\mathbb{X}$. On dit que la suite converge simplement vers une fonction $f : \mathbb{X} \to \mathbb{R}$ si, pour tout $x \in \mathbb{X}$, $\lim_{n\to+\infty} f_n(x) = f(x)$. On note alors $\lim_{n\to+\infty} f_n = f$.*

Proposition 2.1.4. *Soit $(\Omega, \mathscr{A})$ un espace mesurable et $\mathscr{M}(\Omega, \mathbb{R})$ l'espace des fonctions mesurables de $(\Omega, \mathscr{A})$ dans $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. On a:*

- *Pour tous $\alpha \in \mathbb{R}$ et $f \in \mathscr{M}(\Omega, \mathbb{R})$, $\alpha f \in \mathscr{M}(\Omega, \mathbb{R})$.*
- *Pour tous $f, g \in \mathscr{M}(\Omega, \mathbb{R})$, $f + g \in \mathscr{M}(\Omega, \mathbb{R})$.*
- *Pour tous $f, g \in \mathscr{M}(\Omega, \mathbb{R})$, $fg \in \mathscr{M}(\Omega, \mathbb{R})$.*
- *Pour tous $f, g \in \mathscr{M}(\Omega, \mathbb{R})$, $\max\{f, g\}, \min\{f, g\} \in \mathscr{M}(\Omega, \mathbb{R})$.*
- *Pour toute suite $(f_n)_{n\in\mathbb{N}}$ à valeurs dans $\mathscr{M}(\Omega, \mathbb{R})$ qui converge simplement vers une fonction f, $f \in \mathscr{M}(\Omega, \mathbb{R})$.*

Preuve

Admis. Les preuves se trouvent facilement sur internet et peuvent être faites en exercice.

Exemple. *Soit $(\Omega, \mathscr{A})$ un espace mesurable. Par les deux premiers points de la proposition précédente, on a que, pour $n \in \mathbb{N}^*$, les fonctions du type $\sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$ où, pour $i \in [\![n]\!]$, $\alpha_i \in \mathbb{R}$ et $A_i \in \mathscr{A}$, sont mesurables. Ces fonctions sont appelées fonctions étagées. Lorsque les $(\alpha_i)_{i\in[\![n]\!]}$ et $(A_i)_{i\in[\![n]\!]}$ sont deux à deux disjoints et que les $(\alpha_i)_{i\in[\![n]\!]}$ sont non nuls, l'écriture $\sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$ est appelée la forme (ou décomposition) canonique de la foncton étagée. Il est facile de vérifier que toute foncton étagée non nulle admet une unique forme canonique. En appliquant l'avant-dernier point de la proposition 2.1.4, on en déduit que les fonctions du type $\sum_{i\in I} \alpha_i \mathbb{1}_{A_i}$, avec I au plus dénombrable, $(\alpha_i)_{i\in I}$ une famille à valeurs dans $\mathbb{R}_+$ (ou $\mathbb{R}_-$) et $(A_i)_{i\in I}$ une famille d'éléments de $\mathscr{A}$ deux à deux disjoints, sont mesurables.*

Exercice 2.1.4. *Montrer que toute fonction mesurable réelle f d'image finie $F \subset \mathbb{R}$ est une fonction étagée.*

On sait donc que les fonctions étagées et leurs limites (lorsque bien définies) sont mesurables. Le théorème suivant nous dit que la plupart des fonctions réelles traitées jusqu'ici en L1 et L2 sont en fait mesurables. On rappelle qu'une fonction $f : \mathbb{R} \to \mathbb{R}$ est dite continue par morceaux si il existe une suite $(t_z)_{z\in\mathbb{Z}}$ de réels croissante avec $\lim_{z\to-\infty} t_z = -\infty$ et $\lim_{z\to+\infty} t_z = +\infty$ telle que, pour tout $z \in \mathbb{Z}$, f est continue sur l'intervalle $]t_z, t_{z+1}[$, c'est-à-dire f est continue en x pour tout $x \in]t_z, t_{z+1}[$. En mots, une fonction est continue par morceaux si on peut diviser $\mathbb{R}$ en segments ouverts consécutifs sur lesquels elle est continue. Lorsqu'on traite de fonctions continues par morceaux, on ne s'intéresse en général pas aux valeurs aux bornes des intervalles et on n'impose donc aucune condition dessus (elles sont parfois définies avec la contrainte supplémentaire d'être continues à droite ou à gauche en tout point).

Théorème 2.1.5. *Toute fonction de $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ dans $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ continue par morceaux est mesurable.*

Exercice 2.1.5. *Soient E, F des ensembles et $f : E \to F$. Montrer que, pour toute famille $(A_i)_{i\in I}$ de parties de E :*

- $f^{-1}\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} f^{-1}(A_i)$
- $f^{-1}\left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} f^{-1}(A_i)$
- $f^{-1}(A_0 \backslash A_1) = f^{-1}(A_0) \backslash f^{-1}(A_1)$

Lemme 2.1.6. *Soit $(\Omega, \mathscr{A})$ un espace mesurable, $\mathbb{X}$ un ensemble et $f : \Omega \to \mathbb{X}$. L'ensemble $\mathscr{F} := \{E \subset \mathbb{X} : f^{-1}(E) \in \mathscr{A}\}$ est une tribu sur $\mathbb{X}$.*

Preuve

Puisque f est définie sur Ω, il est immédiat que $f^{-1}(\mathbb{X}) = \Omega \in \mathscr{A}$, c'est-à-dire $\mathbb{X} \in \mathscr{F}$. Soit $E \in \mathscr{F}$. On a $f^{-1}(\mathbb{X}\backslash E) = f^{-1}(\mathbb{X}) \backslash f^{-1}(E) = \Omega \backslash f^{-1}(E) \in \mathscr{A}$ par stabilité par complémentarité de $\mathscr{A}$. Donc $E^c := \mathbb{X}\backslash E \in \mathscr{F}$. Soit $(E_n)_{n\in\mathbb{N}}$ suite à valeurs dans $\mathscr{F}$. On a $f^{-1}\left(\bigcup_{n\in\mathbb{N}} E_n\right) = \bigcup_{n\in\mathbb{N}} f^{-1}(E_n) \in \mathscr{A}$ par stabilité par union dénombrable de $\mathscr{A}$, c'est-à-dire $\bigcup_{n\in\mathbb{N}} E_n \in \mathscr{F}$. On en déduit le résultat souhaité.

Remarque. *La tribu $\mathscr{F}$ dans le lemme précédent est appelée tribu induite par f sur $\mathbb{R}$. C'est la plus grande tribu sur le codomaine de f qui rend f mesurable. De manière similaire, si f est une fonction définie sur un ensemble Ω à valeurs dans un espace mesurable $(\mathbb{X}, \mathscr{F})$, l'ensemble $\mathscr{A} := \{f^{-1}(E) : E \in \mathscr{F}\}$ est une tribu sur Ω. Appelée tribu induite par f sur Ω, c'est la plus petite tribu sur Ω qui rend f mesurable.*

Preuve

[Preuve du théorème 2.1.5] On montre d'abord le résultat pour les fonctions continues sur intervalle ouvert et nulles ailleurs. Soient $f : \mathbb{R} \to \mathbb{R}$ continue sur un intervalle ouvert I et nulle ailleurs et $\mathscr{F} := \{E \subset \mathbb{R} : f^{-1}(E) \in \mathscr{B}(\mathbb{R})\}$. Par le lemme 2.1.6, $\mathscr{F}$ est une tribu sur $\mathbb{R}$. Trivialement, f est mesurable pour $\mathscr{F}$. On rappelle que f est continue sur I si et seulement si, pour tout intervalle ouvert I', $f^{-1}(I') \cap I$ est ouvert (réunion dénombrable d'intervalles ouverts). Soit I' un intervalle ouvert. Si $0 \notin I'$, alors $f^{-1}(I') \subset I$ et donc $f^{-1}(I') = f^{-1}(I') \cap I$ est ouvert. Comme toute réunion dénombrable d'intervalles ouverts est dans $\mathscr{B}(\mathbb{R})$, on en déduit que $f^{-1}(I') \in \mathscr{B}(\mathbb{R})$. Sinon, si $0 \in I'$, alors $f^{-1}(I') = \left(f^{-1}(I') \cap I\right) \cup I^c$. $f^{-1}(I') \cap I$ étant ouvert, on a $f^{-1}(I') \cap I \in \mathscr{B}(\mathbb{R})$ par la remarque précédente. De même, $I \in \mathscr{B}(\mathbb{R})$ et donc par stabilité par complémentarité d'une tribu, $I^c \in \mathscr{B}(\mathbb{R})$. Par stabilité par l'union, on en déduit que $f^{-1}(I') \in \mathscr{B}(\mathbb{R})$. On vient de montrer que, pour tout intervalle ouvert I', $f^{-1}(I') \in \mathscr{B}(\mathbb{R})$, c'est-à-dire $\mathscr{F}$ contient tous les intervalles ouverts. On a donc $\mathscr{B}(\mathbb{R}) := \sigma(\{]a, b[: a, b \in \mathbb{R}\}) \subset \mathscr{F}$ et, par l'exercice 2.1.3, on en déduit que f est mesurable pour $\mathscr{B}(\mathbb{R})$. Supposons maintenant f continue par morceaux. Il existe alors une suite $(t_z)_{z\in\mathbb{R}}$ à valeurs dans $\mathbb{R}$, croissante avec $\lim_{z\to-\infty} t_z = -\infty$ et $\lim_{z\to+\infty} t_z = +\infty$ telle que, pour $z \in \mathbb{Z}$, f est continue sur $]t_z, t_{z+1}[$. Pour $n \in \mathbb{N}$, posons:

$$f_n := \sum_{z=-n}^{n} \left(f\mathbb{1}_{]t_z,t_{z+1}[} + f(t_z)\mathbb{1}_{\{t_z\}}\right)$$

On a déjà montré que, pour $z \in \mathbb{Z}$, $f\mathbb{1}_{]t_z,t_{z+1}[}$ est mesurable (c'est une fonction continue sur l'intervalle ouvert $]t_z, t_{z+1}[$ et nulle ailleurs). De même, par le premier exemple de la section (et car $\mathscr{B}(\mathbb{R})$ contient les singletons) et le premier point de la proposition 2.1.4, pour $z \in \mathbb{Z}$, $f(t_z)\mathbb{1}_{\{t_z\}}$ est mesurable. Par le second point de la proposition 2.1.4, on sait qu'une somme finie de fonctions mesurables est mesurable. Pour $n \in \mathbb{N}$, f_n est donc mesurable. De plus, la suite $(f_n)_{n\in\mathbb{N}}$ converge simplement vers f. Par le dernier point de la proposiiton 2.1.4, on en déduit que f est mesurable, ce qui conclut la preuve.

Exemple. *Les fonctions réelles continues en particulier sont mesurables.*

Exemple. *La fonction $f : \mathbb{R} \to \mathbb{R}, x \mapsto \frac{1}{x}\mathbb{1}_{]-\infty,0[\cup]0,+\infty[}(x)$ est mesurable.*

Nous avons maintenant tous les éléments nécessaire pour définir rigoureusement la notion de variable aléatoire.

Définition 2.1.7 (Variable aléatoire)**.** *On appelle variable aléatoire toute fonction mesurable d'un espace probabilisé dans un espace mesurable. Lorsque ce dernier est $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, on parle plus précisément de variable aléatoire réelle.*

Une variable aléatoire est donc une fonction mesurable. Par la remarque qui suit la proposition 2.1.2, toute composition d'une variable aléatoire par des fonctions mesurables est elle même une variable aléatoire. Cette propriété fait des variables aléatoires un outil particulièrement pratique car elles permettent d'encoder le résultat d'une expérience aléatoire dans une variable formelle et d'appliquer des transformations à cette dernière. Le pouvoir de ce formalisme est que les transformations appliquées à la variable transfèrent la probabilité de départ sur un nouvel espace d'intérêt (qui contient l'image de la composition des transformations). En effet, comme nous l'avons vu dans l'exemple introductif de cette section, la donnée d'un espace probabilisé et d'une variable aléatoire sur celui-ci induit une probabilité sur l'espace d'arrivée de la variable aléatoire. On appelle cette dernière sa loi ou distribution.

Définition 2.1.8 (Loi d'une variable aléatoire)**.** *Soient $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé, $(\mathbb{X}, \mathscr{F})$ un espace mesurable et $X : \Omega \to \mathbb{X}$ une variable aléatoire. On appelle loi de X la mesure de probabilité $\mathbf{P}_X$ sur $(\mathbb{X}, \mathscr{F})$ donnée par:*

$$\begin{aligned} \mathbf{P}_X : \mathscr{F} &\to [0,1] \\ A &\mapsto \mathbf{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbf{P}(\{X^{-1}(A)\}) \end{aligned}$$

Remarque. *Plus généralement, pour une fonction mesurable f d'un espace mesuré $(\Omega, \mathscr{A}, \mu)$ dans un espace mesurable $(\mathbb{X}, \mathscr{F})$, on peut définir la mesure μ_f induite par f sur $(\mathbb{X}, \mathscr{F})$ de la même façon. On appelle alors μ_f la mesure image de μ par f et on la note usuellement $f_*\mu$ ou $\mu \circ f^{-1}$.*

Comme l'illustre l'exemple introductif, la mesurabilité de X dans la définition précédente sert à assurer que $\{\omega \in \Omega : X(\omega) \in A\}$ est bien un événement de $\mathscr{A}$ pour tout $A \in \mathscr{F}$. Par souci de parcimonie, on note le plus souvent $\{X \in A\}$ pour $\{\omega \in \Omega : X(\omega) \in A\}$ et $\mathbf{P}(X \in A)$ pour $\mathbf{P}(\{\omega \in \Omega : X(\omega) \in A\})$.

Exercice 2.1.6. *Vérifier que l'application $\mathbf{P}_X$ dans la définition précédente est une probabilité.*

En général, lorsqu'on s'intéresse à une loi donnée, on commence par définir formellement une variable aléatoire suivant cette loi puis on écrit nos calculs avec (pour une variable aléatoire X suivant une loi μ, on écrit souvent $X \sim \mu$). Cette pratique permet de rendre ces derniers plus simples à suivre mais implique l'existence d'un espace probabilisé initial. Celui-ci, habituellement noté $(\Omega, \mathscr{A}, \mathbf{P})$ ou $(\Omega, \mathscr{F}, \mathbf{P})$, est rarement précisé explicitement et peut être vu comme un résidu nécessaire de l'usage de variables aléatoires. On peut toutefois lui donner une interprétation intéressante. Cet espace initial peut en effet être vu comme l'ensemble des événements, potentiellement inconnus, influant sur le résultat de notre expérience aléatoire. Avant d'illustrer la définition d'une loi par un exemple, nous

faisons une observation sur l'espace d'arrivée des variables aléatoires. Cela vient du fait qu'une variable aléatoire X avec pour espace d'arrivée $(\mathbb{X}, \mathscr{F})$ et image inclue dans un sous-ensemble $\mathbb{X}' \subset \mathscr{F}$ est également mesurable pour l'espace d'arrivée $(\mathbb{X}', \mathscr{F}')$, où $\mathscr{F}'$ est la tribu-trace de $\mathscr{F}$ sur $\mathbb{X}'$. Ainsi, la loi $\mathbf{P}_X$ de X peut être définie sur $(\mathbb{X}', \mathscr{F}')$ ou $(\mathbb{X}, \mathscr{F})$. Les espaces probabilisés $(\mathbb{X}', \mathscr{F}', \mathbf{P}_X)$ et $(\mathbb{X}, \mathscr{F}, \mathbf{P}_X)$ sont alors isomorphiques mod 0 (voir la remarque suivant l'exercice **??**). En particulier, lorsque X est une variable aléatoire réelle d'image finie (ou dénombrable) $X(\Omega) \subset \mathbb{R}$, on peut considérer l'espace probabilisé (d'arrivée) $(X(\Omega), \mathscr{P}(X(\Omega)), \mathbf{P}_X)$ au lieu de $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbf{P}_X)$. En fait, dans ce cas, on peut même restreindre $\mathbf{P}_X$ à l'espace $(\Xi, \mathscr{P}(\Xi))$ où $\Xi := \{x \in X(\Omega) : \mathbf{P}(X = x) > 0\}$. Ces espaces étant fonctionnellement équivalents, on omettra régulièrement d'en expliciter un.

Exemple. *Soit $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé et $A \in \mathscr{A}$. On considère la variable aléatoire $X := \mathbb{1}_A$. La loi de X est alors $\mathbf{P}_X = \mathbf{P}(A)\delta_1 + (1 - \mathbf{P}(A))\delta_0 = \mathscr{B}(\mathbf{P}(A))$. En mots, X suit une loi de Bernoulli de paramètre $\mathbf{P}(A)$.*

Exercice 2.1.7. *On considère l'espace probabilisé $([0, n], \mathscr{B}([0, n]), \mathscr{U}([0, n]))$ avec $n \in \mathbb{N}^*$. Donner la loi de $X := \sum_{k=1}^{n} k\mathbb{1}_{[k-1,k)}$.*

Une fonction importante lorsqu'on traite de variables aléatoires réelles est la fonction de répartition. Celle-ci détermine leur loi.

Définition 2.1.9 (Fonction de répartition)**.** *Soit X une variable aléatoire réelle. On appelle fonction de répartition de X, notée F_X, la fonction suivante:*

$$\begin{aligned} F_X : \mathbb{R} &\to [0, 1] \\ x &\mapsto \mathbf{P}(X \leq x) \end{aligned}$$

Corollaire 2.1.10. *Soient X et Y deux variables aléatoires réelles de fonctions de répartition respectives F_X et F_Y. Si $F_X(x) = F_Y(x)$ pour tout $x \in \mathbb{R}$, alors X et Y ont même loi, c'est-à-dire $\mathbf{P}_X = \mathbf{P}_Y$.*

Preuve

On a vu dans le dernier exemple de la section 1.2 que $\mathbf{P}_X = \mathbf{P}_Y$ si et seulement si $\mathbf{P}_X(]-\infty, x]) = \mathbf{P}_Y(]-\infty, x])$ pour tout $x \in \mathbb{R}$. Or, par définition, on a, pour $x \in \mathbb{R}$, $\mathbf{P}_X(]-\infty, x]) = \mathbf{P}(X \leq x)$ et $\mathbf{P}_Y(]-\infty, x]) = \mathbf{P}(Y \leq x)$.

Exemple. *Soit $X \sim \mathscr{B}(p)$ avec $p \in [0, 1]$. Alors F_X est donnée par :*

$$\begin{aligned} F_X : \mathbb{R} &\to [0, 1] \\ x &\mapsto (1 - p)\mathbb{1}_{[0,1[}(x) + \mathbb{1}_{[1,+\infty[}(x) \end{aligned}$$

Exercice 2.1.8. *Soit $X \sim \mathscr{U}(\llbracket n \rrbracket)$ avec $n \in \mathbb{N}^*$. Donner la fonction de répartition de X.*

Exemple. *Soit $X \sim \mathscr{U}([0, 1])$. La fonction de répartition F_X est donnée par :*

$$\begin{aligned} F_X : \mathbb{R} &\to [0, 1] \\ x &\mapsto x\mathbb{1}_{[0,1[}(x) + \mathbb{1}_{[1,+\infty[}(x) \end{aligned}$$

Exercice 2.1.9. *Donner la fonction de répartion d'une variable aléatoire de loi $\mathscr{U}([a, b])$ avec $a, b \in \mathbb{R}$ tels que $a < b$. On suppose maintenant que $a \in \mathbb{R}_+^*$. En déduire que $Y := aX + b$, où $X \sim \mathscr{U}([0, 1])$, a pour loi $\mathscr{U}([b, a + b])$.*

2.2 Intégration par une mesure et lois à densité

2.2.1 Définition d'une intégrale "au sens de Lebesgue"

L'intégration par une mesure généralise l'intégration de Riemann et les sommes indexées, qui sont des cas particuliers de celle-ci. En particulier, elle permet de définir des intégrales (dans un sens large) de fonctions réelles sur des espaces mesurés quelconques et est ainsi à la base du concept d'espérance en probabilités.

Définition 2.2.1 (Parties positive et négative)**.** *Soit f une fonction réelle mesurable sur un espace $(\mathbb{X}, \mathscr{F})$. On appelle partie positive de f, notée f_+, et partie négative de f, notée f_-, les fonctions suivantes:*

$$\begin{aligned} f_+ : \mathbb{X} &\to \mathbb{R} \\ x &\mapsto \max\{0, f(x)\} \\ f_- : \mathbb{X} &\to \mathbb{R} \\ x &\mapsto -\min\{0, f(x)\} \end{aligned}$$

f_+ et f_- sont des fonctions réelles positives telles que $f = f_+ - f_-$. De plus, par le quatrième point de la proposition 2.1.4, elles sont mesurables.

Définition 2.2.2 (Intégration par une mesure). *Soit $(\mathbb{X}, \mathscr{F}, \mu)$ un espace mesuré et f une fonction mesurable réelle sur $(\mathbb{X}, \mathscr{F})$. L'intégrale de f par rapport à μ, notée $\int f d\mu$, $\int_{\mathbb{X}} f d\mu$, $\int_{\mathbb{X}} f(x) d\mu(x)$, $\int_{\mathbb{X}} f(x) \mu(dx)$ ou encore $\mu(f)$, est définie comme suit:*

- *Si f est une fonction étagée positive, qu'on peut donc écrire $f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$ $(\alpha_i \in \mathbb{R}_+)$, alors*

$$\int f d\mu := \sum_{i=1}^{n} \alpha_i \mu(A_i) \in \bar{\mathbb{R}}_+$$

- *Si f est positive, alors*

$$\int f d\mu := \sup \left\{ \int g d\mu : g \leq f, g \text{ étagée positive} \right\} \in \bar{\mathbb{R}}_+$$

- *Si f est de signe quelconque et $\int |f| d\mu < +\infty$ (f est dite intégrable — "au sens de Lebesgue"), alors*

$$\int f d\mu := \int f_+ d\mu - \int f_- d\mu \in \mathbb{R}$$

  où f_+ et f_- désignent respectivement les parties positive et négative de f.

Remarque. *Dans la définition précédente, l'indice au bas du signe intégral parfois utilisé dans les notations de l'intégrale de f par rapport à μ désigne le domaine d'intégration. Lorsque celui-ci est clair, il est courant de l'omettre.*

Remarque. *L'intégrale par une mesure est parfois appelée intégrale de Lebesgue (à qui on doit son invention), bien que ce terme soit le plus souvent réservé à l'intégrale par la mesure de Lebesgue. Pour distinguer les deux, on parle fréquemment d'intégrale "au sens de Lebesgue" pour désigner la première.*

Exemple. *Soit $(\mathbb{X}, \mathscr{F}, \mu)$ un espace mesuré. Alors $\int 0 d\mu = 0$. En effet, $\mathbb{1}_\emptyset \equiv 0$ et, par définition de l'intégrale d'une fonction étagée positive :*

$$\int \mathbb{1}_\emptyset d\mu = \mu(\emptyset) = 0$$

Une première propriété fondamentale de l'intégrale par une mesure est sa monotonie.

Proposition 2.2.3 (Monotonie de l'intégrale). *Soient f et g deux fonctions mesurables réelles sur un espace mesuré $(\mathbb{X}, \mathscr{F}, \mu)$ positives ou intégrables et telles que $f \leq g$. Alors :*

$$\int f d\mu \leq \int g d\mu$$

Preuve

L'inégalité est facilement vérifiée pour des fonctions f et g positives (noter que ceci est vrai dès lors que f est positive). Supposons f et g intégrable. Puisque $f \leq g$, on a $f_+ \leq g_+$ et $f_- \geq g_-$. On a alors :

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu \leq \int g_+ d\mu - \int g_- d\mu = \int g d\mu$$

où l'inégalité ci-dessus est une conséquence de la monotonie de l'intégrale pour les fonctions positives et les égalités sont par définition de l'intégrale d'une fonction intégrable. Enfin, si g est positive mais non intégrable et que f est seulement intégrable, alors l'inégalité est trivialement vraie car $\int g d\mu = +\infty$.

On rappelle que l'intégrale de Riemann peut être définie sur différents segments de $\mathbb{R}$. De façon similaire, l'intégrale par une mesure définie sur un espace mesurable $(\mathbb{X}, \mathscr{F})$ peut se faire sur n'importe quelle partie mesurable $B \in \mathscr{F}$ de l'espace. On parle de restriction de l'intégrale à B.

Définition 2.2.4 (Restriction d'une intégrale). *Soit $(\mathbb{X}, \mathscr{F}, \mu)$ un espace mesuré, f intégrable pour μ (respectivement positive) et $B \in \mathscr{F}$. La fonction $f\mathbb{1}_B$ est alors trivialement intégrable (respectivement positive) et on appelle restriction de son intégrale à B la valeur suivante:*

$$\int_B f d\mu := \int f \mathbb{1}_B d\mu$$

$\int_B f d\mu$ est aussi appelée l'intégrale de f par μ sur B.

Exemple. *Dans la définition précédente, le cas particulier $f \equiv 1$ donne $\int_B d\mu = \mu(B)$.*

Exercice 2.2.1. *Soient $(\mathbb{X}, \mathscr{F}, \mu)$ un espace mesuré, $B \in \mathscr{F}$ et f une fonction mesurable réelle sur $(\mathbb{X}, \mathscr{F})$ positive ou telle que $f\mathbb{1}_B$ est intégrable par μ. Montrer l'égalité suivante :*

$$\int f d\mu|_B = \int_B f d\mu$$

On pourra commencer par montrer le résultat pour les fonctions étagées positives.

Proposition 2.2.5. *Soient $(\mathbb{X}, \mathscr{F}, \mu)$ un espace mesuré, f une fonction réelle mesurable sur $(\mathbb{X}, \mathscr{F})$ et $B \in \mathscr{F}$ tel que $\mu(B) = 0$. Alors:*

$$\int_B f d\mu = 0$$

Preuve

Supposons f positive. Alors, par définition de l'intégrale par μ d'une fonction positive:

$$\begin{aligned}\int_B f d\mu = \int \mathbb{1}_B f d\mu &= \sup\left\{ \int g d\mu : g \leq \mathbb{1}_B f, g \text{ fonction étagée positive} \right\} \\ &= \sup\left\{ \int \mathbb{1}_B g d\mu : g \leq \mathbb{1}_B f, g \text{ fonction étagée positive} \right\}\end{aligned}$$

où la dernière égalité est car $g \leq \mathbb{1}_B f$ implique que $g = \mathbb{1}_B g$ pour toute fonction positive g. Soit g une fonction étagée positive satisfaisant $g \leq \mathbb{1}_B f$ et de forme canonique $g = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$. On a :

$$\mathbb{1}_B g = \mathbb{1}_B \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i} = \sum_{i=1}^n \alpha_i \mathbb{1}_B \mathbb{1}_{A_i} = \sum_{i=1}^n \alpha_i \mathbb{1}_{B \cap A_i}$$

Par définition de l'intégrale par μ d'une fonciton étagée et par hypothèse sur B, on en déduit:

$$\int \mathbb{1}_B g d\mu = \sum_{i=1}^n \alpha_i \mu(B \cap A_i) \leq \sum_{i=1}^n \alpha_i \mu(B) = 0$$

En particulier, $\int_B f d\mu = \sup\{0\} = 0$. Supposons maintenant f de signe quelconque. Par le résultat déjà montré pour les fonctions positives, $\int |\mathbb{1}_B f| d\mu = \int \mathbb{1}_B |f| d\mu = 0 < +\infty$. C'est-à-dire, $\mathbb{1}_B f$ est intégrable par μ. Par définition de l'intégrale d'une fonction intégrable, on a alors :

$$\int_B f d\mu = \int \mathbb{1}_B f d\mu = \int (\mathbb{1}_B f)_+ d\mu - \int (\mathbb{1}_B f)_- d\mu = \int \mathbb{1}_B f_+ d\mu - \int \mathbb{1}_B f_- d\mu = \int_B f_+ d\mu - \int_B f_- d\mu = 0$$

où la dernière égalité est car f_+ et f_- sont des fonctions positives, ce qui nous renvoie au point précédent.

2.2.2 Théorèmes fondamentaux de convergence

Tout comme pour l'intégrale de Riemann ou les séries, il existe des résultats qui nous permettent d'intervenir la limite et le signe intégral lorsqu'on traite de suite de fonctions. L'un des deux plus importants est le suivant.

Définition 2.2.6 (Suite croissante de fonctions)**.** *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions réelles définies sur un même ensemble $\mathbb{X}$. On dit que $(f_n)_{n\in\mathbb{N}}$ est croissante point par point si $f_n(x) \leq f_{n+1}(x)$ pour tous $x \in \mathbb{X}$ et $n \in \mathbb{N}$.*

Théorème 2.2.7 (Convergence monotone)**.** *Soient $(\mathbb{X}, \mathscr{F}, \mu)$ un espace mesuré et $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions réelles positives mesurables sur $(\mathbb{X}, \mathscr{F})$ croissante point par point qui converge simplement vers une fonction f. Alors:*

$$\int f d\mu = \lim_{n\to+\infty} \int f_n d\mu$$

Remarque. *Le théorème de convergence monotone est parfois connu sous le nom de théorème de Beppo Levi, mathématicien italien qui démontra le résultat cinq ans après la publication de Lebesgue sur l'intégrale éponyme.*

Prouvons maintenant le théorème.

Lemme 2.2.8. *Soit s une fonction étagée positive sur un espace mesurée $(\mathbb{X}, \mathscr{F}, \mu)$ et $c \in \mathbb{R}_+$. Alors :*

$$\int cs d\mu = c \int s d\mu$$

Preuve

Si $c = 0$ ou $s \equiv 0$, le résultat est trivial. Supposons donc $s \not\equiv 0$ et $c > 0$ et écrivons $s = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$ la forme canonique de s. cs est également une fonction étagée positive de forme canonique $cs = \sum_{i=1}^{n} c\alpha_i \mathbb{1}_{A_i}$. Par définiton de l'intégrale d'une fonction étagée positive, on a alors :

$$\int cs d\mu = \sum_{i=1}^{n} c\alpha_i \mu(B_i) = c \sum_{i=1}^{n} \alpha_i \mu(B_i) = c \int s d\mu$$

Lemme 2.2.9. *Soit s une fonction étagée positive sur un espace mesuré $(\mathbb{X}, \mathscr{F}, \mu)$, $E \in \mathscr{F}$ et $(E_n)_{n\in\mathbb{N}}$ une suite croissante d'éléments de $\mathscr{F}$ telle que $\bigcup_{n\in\mathbb{N}} E_n = E$. Alors :*

$$\lim_{n\to+\infty} \int_{E_n} s d\mu = \int_E s d\mu$$

Preuve

Soit $s = \sum_{i=1}^{N} \alpha_i \mathbb{1}_{B_i}$ la décomposition canonique de s. Pour $n \in \mathbb{N}$, on a :

$$\int_{E_n} s d\mu = \int s \mathbb{1}_{E_N} d\mu = \int \left(\sum_{i=1}^{N} \alpha_i \mathbb{1}_{B_i \cap E_n} \right) d\mu = \sum_{i=1}^{N} \alpha_i \mu(B_i \cap E_n) = \sum_{i=1}^{N} \alpha_i \mu|_{B_i}(E_n) = \nu(E_n)$$

où ν est la mesure définie par $\nu := \sum_{i=1}^{N} \alpha_i \mu|_{B_i}$. Par le lemme 1.2.10, on obtient :

$$\lim_{n\to+\infty} \int_{E_n} s d\mu = \lim_{n\to+\infty} \nu(E_n) = \nu(E) = \sum_{i=1}^{N} \alpha_i \mu(B_i \cap E) = \int s \mathbb{1}_E d\mu = \int_E s d\mu$$

Preuve

[Preuve du théorème 2.2.7] Notons $I := \int f d\mu$ et, pour $n \in \mathbb{N}$, $I_n := \int f_n d\mu$. Par monotonie de l'intégrale et car $f_n \leq f$ pour tout $n \in \mathbb{N}$, on a, pour $n \in \mathbb{N}$:

$$I_n \leq I$$

Par monotonie encore et croissance point par point de la suite $(f_n)_{n\in\mathbb{N}}$, la suite $(I_n)_{n\in\mathbb{N}}$ est croissante. On en déduit que $\lim_{n\to+\infty} I_n$ est bien définie dans $\bar{\mathbb{R}}_+$ et satisfait $\lim_{n\to+\infty} I_n \leq I$. Considérons une fonction étagée positive s sur $(\mathbb{X}, \mathscr{F})$ telle que $s \leq f$ et fixons $c \in]0, 1[$. Pour $n \in \mathbb{N}$, on définit :

$$E_n := \{x \in \mathbb{X} : f_n(x) \geq cs(x)\}$$

Par construction, on a :

$$I_n \geq \int_{E_n} f_n d\mu \geq \int_{E_n} cs d\mu = c \int_{E_n} s d\mu$$

où l'égalité est par le lemme 2.2.8. De plus, la suite $(E_n)_{n\in\mathbb{N}}$ est croissante au sens de l'inclusion et $\bigcup_{n\in\mathbb{N}} E_n = \mathbb{X}$. Par le lemme 2.2.9, on obtient :

$$\lim_{n\to+\infty} \int_{E_n} s d\mu = \int_{\mathbb{X}} s d\mu = \int s d\mu$$

Ainsi :

$$\lim_{n\to+\infty} I_n \geq c \int s d\mu$$

Puisque c est pris arbitrairement dans $(0, 1)$, on en déduit que :

$$\lim_{n\to+\infty} I_n \geq \sup_{c\in(0,1)} c \int s d\mu = \int s d\mu$$

Comme cette dernière inégalité est vraie pour toute fonction étagée positive s sur $(\mathbb{X}, \mathscr{F})$ telle que $s \leq f$, on en déduit par définition de l'intégrale d'une fonction mesurable positive que $\lim_{n\to+\infty} I_n \geq I$ et donc $\lim_{n\to+\infty} I_n = I$.

L'autre théorème fondamental de convergence est énoncé ci-dessous.

Théorème 2.2.10 (Convergence dominée). *Soit $(\mathbb{X}, \mathscr{F}, \mu)$ un espace mesure et $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions réelles mesurables sur $(\mathbb{X}, \mathscr{F})$ qui converge simplement vers une fonction f. Si il existe une fonction réelle g sur $\mathbb{X}$ intégrable par rapport à μ et $N \in \mathbb{N}$ tels que, pour tous $n \geq N$, $|f_n| \leq g$, alors f est intégrable par rapport à μ et on a:*

$$\int f d\mu = \lim_{n\to+\infty} \int f_n d\mu$$

Preuve

Admis. La preuve est relativement courte mais requiert un résultat hors programme (le lemme de Fatou).

Remarque. *Dans le théorème de convergence dominée, la suite $(f_n)_{n\in\mathbb{N}}$ est dite éventuellement "dominée" par g.*

2.2.3 Propriétés de l'intégrale "au sens de Lebesgue"

Cette partie s'intéresse aux différentes propriétés er résultats sur l'intégrale définie.

Le théorème de convergence monotone, en association avec le lemme suivant, est un outil essentiel pour montrer ces propriétés.

Lemme 2.2.11. *Soit $(\mathbb{X}, \mathscr{F})$ un espace mesurable et f une fonction réelle positive mesurable sur $(\mathbb{X}, \mathscr{F})$. Il existe une suite $(s_n)_{n\in\mathbb{N}}$ de fonctions étagées positives sur $(\mathbb{X}, \mathscr{F})$ croissantes point par point qui converge simplement vers f.*

Preuve

Admis. La preuve est assez longue mais tous les résultats nécessaires ont déjà été énoncés. Les plus intéressés la trouveront facilement sur internet.

Méthodologie : Lorsqu'on souhaite montrer un résultat sur l'intégrale par une mesure, on procède en général de la sorte. On commence par montrer le résultat pour les fonctions étagées positives. On le montre ensuite pour les fonctions mesurables positives en prenant une suite de fonctions étagées positives croissante point par point qui converge simplement vers notre fonction et on passe à la limite par le théorème de convergence monotone. Enfin, on le montre trivialement pour les fonctions intégrables à partir du résultat prouvé pour les fonctions positives (en décomposant la fonction en sa partie positive et négative, qui sont toutes deux des fonctions positives). La preuve de la linéarité de l'intégrale est une première illustration de la mécanique décrite ci-dessus.

Proposition 2.2.12 (Linéarité de l'intégrale). *Soient f et g deux fonctions mesurables réelles sur un espace mesuré $(\mathbb{X}, \mathscr{F}, \mu)$ et $c \in \mathbb{R}$. Si f et g sont positives et $c \in \mathbb{R}_+$ ou si f et g sont intégrables, alors :*

$$\int (f + cg) d\mu = \int f d\mu + c \int g d\mu$$

Preuve

On considère d'abord $c \in \mathbb{R}_+$. Supposons que f et g sont des fonctions étagées positives et écrivons $f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$ et $g = \sum_{j=1}^{m} \beta_j \mathbb{1}_{B_i}$ leurs formes canoniques. Par le lemme 2.2.8, il est suffisant de montrer que $\int (f + g) d\mu = \int f d\mu + \int g d\mu$. En écrivant $\alpha_0 = \beta_0 = 0$, $A_0 = \left(\bigcup_{i\in[\![n]\!]} A_i\right)^c$ et $B_0 = \left(\bigcup_{j\in[\![m]\!]B_j}\right)^c$, on peut réécrire $f = \sum_{i=0}^{n} \alpha_i \mathbb{1}_{A_i}$ et $g = \sum_{j=0}^{n} \beta_j \mathbb{1}_{B_i}$. $f + g$ est une fonction étagée positive qui peut alors s'écrire :

$$f + g = \sum_{i=0}^{n} \sum_{j=0}^{m} (\alpha_i + \beta_j) \mathbb{1}_{A_i \cap B_j}$$

Par définition de l'intégrale d'une fonction étagée positive, on obtient :

$$\int (f + g) d\mu = \sum_{i=0}^{n} \sum_{j=0}^{m} (\alpha_i + \beta_j) \mu(A_i \cap B_j) = \sum_{i=0}^{n} \alpha_i \sum_{j=0}^{m} \mu(A_i \cap B_j) + \sum_{j=0}^{m} \beta_j \sum_{i=0}^{n} \mu(A_i \cap B_j)$$

Puisque $(A_i)_{i\in[\![0;n]\!]}$ et $(B_j)_{j\in[\![0;m]\!]}$ sont des partitions de $\mathbb{X}$, la σ-additivté de μ nous donne :

$$\int (f + g) d\mu = \sum_{i=0}^{n} \alpha_i \mu(A_i) + \sum_{j=0}^{m} \beta_j \mu(B_j) = \int f d\mu + \int g d\mu$$

où la seconde égalité est encore un fois par définition de l'intégrale d'une fonciton étagée positive. On a montré que l'intégrale est linéaire pour les fonctions étagées positives. Supposons maintenant simplement f et g mesurables positives. Par le lemme 2.2.11, il existe des suites $(f_n)_{n\in\mathbb{N}}$ et $(g_n)_{n\in\mathbb{N}}$ de fonctions étagées positives croissantes point

par point qui convergent simplement vers f et g respectivement. On suppose dans un premier temps que $c \in \mathbb{R}_+$. On vérifie facilement que la suite $(f_n + cg_n)_{n\in\mathbb{N}}$ est croissante point par point et converge simplement vers $f + cg$. On a alors :

$$\int (f+cg)d\mu = \lim_{n\to+\infty} \int (f_n + cg_n)d\mu = \lim_{n\to+\infty} \left(\int f_n d\mu + c\int g_n d\mu \right) = \lim_{n\to+\infty} \int f_n d\mu + c \lim_{n\to+\infty} \int g_n d\mu = \int f d\mu + c\int g d\mu$$

où la première et dernière égalités sont par le théorème de convergence monotone, la deuxième par le résultat déjà montré pour des fonctions étagées positives et la troisième par linéarité de la limite. Supposons maintenant $c \in \mathbb{R}$ quelconque et f et g intégrables. On observe d'abord que $f + cg$ est intégrable. En effet :

$$\int |f + cg|d\mu \leq \int (|f| + |c||g|)d\mu = \int |f|d\mu + |c| \int |g|d\mu < +\infty$$

où la première inégalité est par inégalité triangulaire sur $\mathbb{R}$ et monotonie de l'intégrale, l'égalité est par le résultat déjà montré pour les fonctons mesurables positives et une constante positive et la dernière inégalité est par intégrabilité de f et g. On montre d'abord que l'intégrale d'une fonction intégrable est homogène, i.e. $\int cgd\mu = c\int gd\mu$. Si $c = 0$ le résultat est trivial. Supposons $c > 0$. Alors :

$$\int cgd\mu = \int (cg)_+ d\mu - \int (cg)_- d\mu = \int cg_+ d\mu - cg_- d\mu = c\int g_+ d\mu - c\int g_- d\mu = c\left(\int g_+ d\mu - \int g_- d\mu \right) = c\int g d\mu$$

où les première et dernière égalité sont par définition de l'intégrale d'une fonction intégrable, la seconde par positivité de c et la troisième par la linéarité le résultat déjà montré pour les fonctions et une constante c positives. Si $c < 0$, on remarque que $(cg)_+ = (-c)g_-$ et $(cg)_- = (-c)g_+$. On obtient alors, de façon équivalente :

$$\int cgd\mu = \int (cg)_+ d\mu - \int (cg)_- d\mu = \int (-c)g_- d\mu + \int (-c)g_+ d\mu = (-c)\int g_- d\mu - (-c)\int g_+ d\mu = c\left(\int g_+ d\mu - \int g_- d\mu \right) = c\int g d\mu$$

Il suffit maintenant de montrer que $\int (f + g)d\mu = \int f d\mu + \int g d\mu$. Observons que $f + g = (f_+ + g_+) - (f_- + g_-)$ mais également $f + g = (f + g)_+ - (f + g)_-$. Ainsi :

$$(f_+ + g_+) + (f + g)_- = (f_- + g_-) + (f + g)_+$$

En intégrant et utilisant le résultat déjà montré pour des fonctions positives avec $c = 1$, on obtient :

$$\int (f_+ + g_+)d\mu + \int (f + g)_- d\mu = \int (f_- + g_-)d\mu + \int (f + g)_+ d\mu$$

Ce qui nous donne :

$$\int (f + g)d\mu = \int (f + g)_+ d\mu - \int (f + g)_- d\mu = \int (f_+ + g_+)d\mu - \int (g_- + f_-)d\mu$$

On conclut ensuite en appliquant une seconde fois le résultat déjà montré pour les fonctions positives :

$$\int (f_+ + g_+)d\mu - \int (g_- + f_-)d\mu = \int f_+ d\mu - \int f_- d\mu + \int g_+ d\mu - \int g_- d\mu = \int f d\mu + \int g d\mu$$

Notons que la proposition précédente reste vraie si une des deux fonctions est positive (ou négative) et l'autre intégrable, quelque soit le signe de la constante c choisie. Les soucis peuvent arriver si aucune des deux fonctions n'est intégrable et que leur intégrales sont soit de signes différents avec un c positif, soit de même signe avec un c négatif. On se retrouve alors avec une forme $(+\infty) + (-\infty)$ qui n'est pas définie.

Exemple. *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions mesurables positives sur un même espace mesuré $(\mathbb{X}, \mathscr{F}, \mu)$ telle que $\sum_{n\in\mathbb{N}} f_n \leq \infty$. Pour $n \in \mathbb{N}$, posons $F_n := \sum_{k=0}^n f_k$. Pour $n \in \mathbb{N}$, F_n est une fonction mesurable réelle (par sommme finie de fonctions mesurables réelles positives) positive (par somme de fonctions positives). De plus, par positivité des $(f_n)_{n\in\mathbb{N}}$, la suite $(F_n)_{n\in\mathbb{N}}$ est croissante point par point et admet pour limite $F := \sum_{n\in\mathbb{N}} f_n$. F est mesurable (par limite de fonctions mesurables) et bien définie dans $\mathbb{R}_+$. Son intégrale est donc bien définie dans $\bar{\mathbb{R}}_+$. Le théorème de convergence monotone nous dit alors :*

$$\int F d\mu = \int \left(\lim_{n\in\mathbb{N}} F_n \right) d\mu = \lim_{n\in\mathbb{N}} \int F_n d\mu$$

Or, par linéarité de l'intégrale et pour $n \in \mathbb{N}$:

$$\int F_n d\mu = \int \left(\sum_{k=0}^n f_k \right) d\mu = \sum_{k=0}^n \int f_k d\mu$$

Donc, par positivité des $(f_n)_{n\in\mathbb{N}}$ et monotonie de l'intégrale :

$$\lim_{n\in\mathbb{N}} \int F_n d\mu = \sum_{n\in\mathbb{N}} \int f_n d\mu$$

Autrement dit, on a :

$$\int \left(\sum_{n\in\mathbb{N}} f_n \right) d\mu = \sum_{n\in\mathbb{N}} \int f_n d\mu$$

Le théorème de convergence monotone combiné avec la linéarité de l'intégrale nous apprennent ainsi qu'on peut intervertir une somme infinie (donc quelconque) et une intégrale si les fonctions intégrées sont positives.

Dans la propositon 2.2.12 et dans le cas où $c = 1$, on parle d'additivité de l'intégrale. On va voir plusieurs conséquences immédiates de cette dernière. Une première est la relation de Chasles donnée dans l'exercice suivant.

Exercice 2.2.2. *Soient $(\mathbb{X}, \mathscr{F}, \mu)$ un espace mesuré, $A, B \in \mathscr{F}$ disjoints et f une fonction réelle mesurable sur $(\mathbb{X}, \mathscr{F})$ positive ou intégrable par μ. Montrer que $\int_{A\cup B} f d\mu = \int_A f d\mu + \int_B f d\mu$. Cette égalité est appelée relation de Chasles.*

Une seconde est que l'intégrale de la valeur absolue d'une fonction intégrable f sur un espace mesuré $(\mathbb{X}, \mathscr{F}, \mu)$ peut s'écrire :

$$\int |f| d\mu = \int f_+ d\mu + \int f_- d\mu$$

En effet, $|f| = f_+ + f_-$. De cette observation, on peut facilement déduire l'inégalité triangulaire pour les intégrales.

Proposition 2.2.13 (Inégalité triangulaire)**.** *Soit f une fonction mesurable sur un espace mesuré $(\mathbb{X}, \mathscr{F}, \mu)$ positive ou intégrable. Alors :*

$$\left| \int f d\mu \right| \leq \int |f| d\mu$$

Preuve

Si f est positive, on a égalité et donc le résultat est trivial. Supposons alors f intégrable quelconque. Par définition de l'intégrale d'une fonction intégrable, on a :

$$\left| \int f d\mu \right| = \left| \int f_+ d\mu - \int f_- d\mu \right| \leq \left| \int f_+ d\mu \right| + \left| \int f_- d\mu \right| = \int f_+ d\mu + \int f_- d\mu = \int |f| d\mu$$

où l'inégalité est par l'inégalité triangulaire classique sur $\mathbb{R}$, la seconde égalité par positivité de f_+ et f_- (et donc de $\int f_+ d\mu$ et $\int f_- d\mu$), la dernière par l'observation faite avant la proposition.

Une dernière conséquence de l'additivité, et plus précisément de la relation de Chasles, est la suivante. On a vu dans la propositon 2.2.5 que l'intégrale d'une fonction sur une partie négligeable est nulle. On peut en fait montrer que deux fonctions qui sont égales à un ensemble négligeable près (dites égales presque partout) ont la même intégrale.

Définition 2.2.14 (Égalité presque partout)**.** *Soit $(\mathbb{X}, \mathscr{F}, \mu)$ un espace mesuré et f, g deux fonctions réelles mesurables sur $(\mathbb{X}, \mathscr{F})$. On dit que f et g coïncident (ou sont égales) μ-presque partout si $\mu(\{x \in \mathbb{X} : f(x) \neq g(x)\}) = 0$. On utilise le plus souvent l'abréviation $f = g$ μ-p.p.. Lorsque μ est une probabilité, on dit également que f et g sont égales μ-presque sûrement (noté $f = g$ μ-p.s.).*

Remarque. *Dans la définition précédente, l'ensemble $\{x \in \mathbb{X} : f(x) \neq g(x)\}$ est bien mesurable. En effet, pour tout $r \in \mathbb{R}$, $\{x \in \mathbb{X} : f(x) < r < g(x)\} = \{x \in \mathbb{X} : f(x) < r\} \cap \{x \in \mathbb{X} : r < g(x)\} = f^{-1}(]-\infty, r[) \cap g^{-1}(]r, +\infty[) \in \mathscr{F}$. On peut ensuite écrire $\{x \in \mathbb{X} : f(x) < g(x)\} = \bigcup_{r\in\mathbb{Q}} \{x \in \mathbb{X} : f(x) < r < g(x)\}$. Par stabilité par union dénombrable d'une tribu, on a donc $\{x \in \mathbb{X} : f(x) < g(x)\} \in \mathscr{F}$ (on rappelle que l'ensemble $\mathbb{Q}$ des rationnels est dénombrable, confère annexe B si ce n'est pas clair). Par symétrie, $\{x \in \mathbb{X} : f(x) > g(x)\} \in \mathscr{F}$ et on en déduit $\{x \in \mathbb{X} : f(x) \neq g(x)\} = \{x \in \mathbb{X} : f(x) < g(x)\} \cup \{x \in \mathbb{X} : f(x) > g(x)\} \in \mathscr{F}$.*

Exemple. *Soit f une fonction réelle mesurable sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ et λ la mesure de Lebesgue. Car λ est diffuse, f et toute modification de f sur un nombre au plus dénombrable de points sont égales λ-p.p..*

Corollaire 2.2.15. *Soient $(\mathbb{X}, \mathscr{F}, \mu)$ un espace mesuré et f, g deux fonctions réelles mesurables sur $(\mathbb{X}, \mathscr{F})$ telles que $f = g$ μ-p.p.. Alors, si f ou g est positive ou intégrable par μ, on a:*

$$\int f d\mu = \int g d\mu$$

Preuve

Soit $B := \{x \in \mathbb{X} : g(x) \neq f(x)\} \in \mathscr{F}$. Par la relation de Chasles, on a :

$$\int f d\mu = \int_{\mathbb{X}} f d\mu = \int_{B \cup B^c} f d\mu = \int_B f d\mu + \int_{B^c} f d\mu = \int_B f d\mu + \int f \mathbb{1}_{B^c} d\mu$$

Or, par définition de B, $f|_{B^c} = g|_{B^c}$ et donc $f\mathbb{1}_{B^c} = g\mathbb{1}_{B^c}$. De plus, comme $f = g$ μ-p.p., $\mu(B) = 0$. La proposition 2.2.5 nous dit donc que $\int_B f d\mu = 0$. On obtient ainsi :

$$\int f d\mu = \int f \mathbb{1}_{B^c} d\mu = \int g \mathbb{1}_{B^c} d\mu$$

Par symétrie, on a $\int g d\mu = \int f \mathbb{1}_{B^c} d\mu = \int f d\mu$.

Exemple. *En reprenant l'exemple précédent, l'intégrale par λ d'une fonction intégrable f sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda)$ ne change pas si on modifie f sur un nombre au plus dénombrable de ses points. En particulier, l'intégrale par λ d'une fonction continue par morceaux ne dépend pas de ses valeurs à ses points de discontinuité.*

La proposition suivante nous dit que, pour intégrer une fonction par une somme (potentiellement infinie et pondérée) de mesures, il suffit de considérer la somme des intégrales par chacune de ces mesures.

Proposition 2.2.16. *Soit $(\mathbb{X}, \mathscr{F})$ un espace mesurable, $(\mu_i)_{i \in I}$ une famille au plus dénombrable de mesures sur $(\mathbb{X}, \mathscr{F})$, $(\alpha_i)_{i \in I}$ une famille de réels positifs, $\mu := \sum_{i \in I} \alpha_i \mu_i$ et f une fonction mesurable sur $(\mathbb{X}, \mathscr{F})$. Alors f est intégrable par rapport à μ si et seulement si $\sum_{i \in I} \alpha_i \int |f| d\mu_i < +\infty$. De plus, si f est positive ou intégrable par rapport à μ:*

$$\int f d\mu = \sum_{i \in I} \alpha_i \int f d\mu_i$$

Preuve

Supposons f positive. Soit s une fonction étagée positive sur $(\mathbb{X}, \mathscr{F})$ de forme canonique $\sum_{k=1}^n \beta_k \mathbb{1}_{A_k}$ et telle que $s \leq f$. Par définition de l'intégrale d'une fonction étagée, on a :

$$\int s d\mu = \sum_{k=1}^n \beta_k \mu(A_k) = \sum_{k=1}^n \beta_k \sum_{i \in I} \alpha_i \mu_i(A_k) = \sum_{i \in I} \alpha_i \sum_{k=1}^n \beta_k \mu_i(A_k) = \sum_{i \in I} \alpha_i \int s d\mu_i$$

où l'interversion des sommes est possible car les termes sont positifs. Maintenant, par le lemme 2.2.11, il existe une suite $(s_n)_{n \in \mathbb{N}}$ de fonctions étagées positives sur $(\mathbb{X}, \mathscr{F})$ croissantes point par point qui converge simplement vers f. Par le point qu'on vient de montrer, on a, pour tout $n \in \mathbb{N}$:

$$\int s_n d\mu = \sum_{i \in I} \alpha_i \int s_n d\mu_i$$

Par positivité des suite $(s_n)_{n \in \mathbb{N}}$ et $(\alpha_i)_{i \in I}$, $\alpha_i \int s_n d\mu_i \in \bar{\mathbb{R}}_+$ pour tous $n \in \mathbb{N}$ et $i \in I$. On peut donc intervertir somme et limite et on a :

$$\lim_{n \to +\infty} \sum_{i \in I} \alpha_i \int s_n d\mu_i = \sum_{i \in I} \lim_{n \to +\infty} \alpha_i \int s_n d\mu_i = \sum_{i \in I} \alpha_i \lim_{n \to +\infty} \int s_n d\mu_i$$

où la seconde égalité est par linéarité de la limite. Or, par le théorème de convergence monotone et hypothèse sur la suite $(s_n)_{n \in \mathbb{N}}$, $\lim_{n \to +\infty} \int s_n d\mu = \int f d\mu$ et, pour tout $i \in I$, $\lim_{n \to +\infty} \int s_n d\mu_i = \int f d\mu_i$. C'est-à-dire :

$$\int f d\mu = \sum_{i \in I} \alpha_i \int f d\mu_i$$

Supposons maintenant f intégrable par μ. Par le point précédent, il est immédiat que cette condition est équivalente à $\sum_{i \in I} \alpha_i \int |f| d\mu_i < +\infty$. On a alors :

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu = \sum_{i \in I} \alpha_i \int f_+ d\mu_i - \sum_{i \in I} \alpha_i \int f_- d\mu_i = \sum_{i \in I} \alpha_i \left(\int f_+ d\mu_i - \int f_- d\mu_i \right) = \sum_{i \in I} \alpha_i \int f d\mu_i$$

où les première et dernière égalités sont par définition de l'intégrale d'une fonction intégrable et la deuxième par le résultat déjà montré pour les fonctions positives.

Une application directe de la proposition précédente est l'intégration par une mesure atomique. Celle-ci se réduit en fait à une somme des images des atomes (i.e. les singletons de masse non nulle) par la fonction, pondérée par les masses de ces derniers. Dans le cas où la mesure est une probabilité, ces masses sont de somme 1 et on peut donc voir l'intégration comme une moyenne pondérée.

Corollaire 2.2.17 (Intégration par une mesure atomique). *Soit $(\mathbb{X}, \mathscr{F})$ un espace mesurable tel que $\mathscr{F}$ contient les singletons (i.e. $\{\{x\} : x \in \mathbb{X}\} \subset \mathscr{F}$). On considère la mesure atomique $\mu := \sum_{i\in I} \alpha_i \delta_{a_i}$ sur $(\mathbb{X}, \mathscr{F})$ où I est un ensemble au plus dénombrable, $(\alpha_i)_{i\in I}$ une famille de réels positifs et $(a_i)_{i\in I}$ une famille d'éléments distincts de $\mathbb{X}$. Une fonction réelle mesurable f sur $(\mathbb{X}, \mathscr{F})$ est intégrable par rapport à μ si et seulement si $\sum_{i\in I} \alpha_i |f(a_i)| < +\infty$. De plus, pour toute fonction réelle mesurable f sur $(\mathbb{X}, \mathscr{F})$ positive ou intégrable par rapport à μ, on a:*

$$\int f d\mu := \sum_{i\in I} \alpha_i f(a_i)$$

Preuve

Soit $f : \mathbb{X} \to \mathbb{R}$ mesurable. On suppose que f est positive ou intégrable par rapport μ. Par la proposition 2.2.16, il est suffisant de montrer que, pour tout $x \in \mathbb{X}$, $\int f d\delta_x = f(x)$. Soient $x \in \mathbb{X}$ et $g := f(x)\mathbb{1}_{\{x\}}$. g est une fonction étagée sur $(\mathbb{X}, \mathscr{F})$ et son intégrale est donnée par:

$$\int g d\delta_x = f(x)\delta_x(\{x\}) = f(x)$$

Posons $B := \{y \in \mathbb{X} : f(y) \neq g(y)\}$. On a $B \subset \mathbb{X}\backslash\{x\}$ et donc, par définition d'une Dirac, $\delta_x(B) = 0$, c'est-à-dire $f = g$ δ_x-p.p.. Par le corollaire 2.2.15, on en déduit:

$$\int f d\delta_x = \int g d\delta_x = f(x)$$

Par la condition d'intégrabilité donnée dans la proposition 2.2.16, on en déduit également que f est intégrable par rapport à μ si et seulement si $\sum_{i\in I} \alpha_i |f(a_i)| < +\infty$.

Exemple. *Soient $n \in \mathbb{N}^*$, μ la mesure de comptage sur $(\llbracket n \rrbracket, \mathscr{P}(\llbracket n \rrbracket))$ et $f : \llbracket n \rrbracket \to \llbracket n \rrbracket$ la fonction identité. On a alors:*

$$\int f d\mu = \sum_{i=1}^{n} f(i) = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

Exercice 2.2.3. *Soient $n \in \mathbb{N}^*$, μ la mesure de comptage sur $(\llbracket n \rrbracket, \mathscr{P}(\llbracket n \rrbracket))$ et f la fonction carrée. Justifier que $\int f d\mu$ est bien définie et donner sa valeur.*

Remarque. *Les intégrales de Riemann sont en fait des cas particuliers de l'intégrale par rapport à Lebesgue. Pour les curieux, cet énoncé est démontré en annexe D. Ainsi, toutes les fonctions intégrables au sens de Riemann le sont également pour la mesure de Lebesgue. L'implication réciproque est fausse.*

Exemple (Fonction de Dirichlet). *La fonction de Dirichlet, $\mathbb{1}_{\mathbb{Q}}$, n'est pas Riemann-intégrable (cf annexe D) mais est intégrable par la mesure de Lebesgue. En effet, $\mathbb{Q}$ étant dénombrable et la mesure de Lebesgue λ diffuse, on a $\lambda(\mathbb{Q}) = 0$ et donc $\mathbb{1}_{\mathbb{Q}}$ est égale $\lambda - p.p.$ à la fonction constante et égale à 0.On en déduit que son intégrale par la mesure de Lebesgue est bien définie et nulle (en particulier $\int_{[0,1]} \mathbb{1}_{\mathbb{Q}} d\lambda = 0$).*

2.2.4 Mesures à densité par rapport à Lebesgue

Nous allons maintenant voir qu'à partir de la mesure de Lebesgue nous pouvons définir toute une classe de mesures diffuses sur $\mathbb{R}$.

Définition 2.2.18 (Mesure à densité par rapport à Lebesgue). *Soit λ la mesure de Lebesgue. On dit d'une mesure μ sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ qu'elle admet une densité contre ou par rapport à (la mesure de) Lebesgue si il existe $g : \mathbb{R} \to \mathbb{R}_+$ mesurable telle que, pour tout $B \in \mathscr{B}(\mathbb{R})$:*

$$\mu(B) = \int_B g d\lambda$$

Remarque. *Dans la définition précédente, g est aussi appelée la dérivée de Radon-Nikodym de μ par rapport à λ et peut être notée $\frac{d\mu}{d\lambda}$. Cette appelation (et notation) est justifée par l'heuristique qui suit : si on ignore le signe intégral dans l'égalité précédente, on obtient l'égalité formelle*

$$d\mu = g d\lambda$$

En traitant $d\mu$ et $d\lambda$ comme des quantités petites mais strictement positives (principe qui est à la base de l'intégration de Riemann), on obtient alors g comme une fraction de $d\mu$ avec $d\lambda$.

Exemple. *Trivialement, la mesure de Lebesgue λ admet une densité contre elle même donnée par $\frac{d\lambda}{d\lambda} \equiv 1$.*

Exemple. *Soit λ la mesure de Lebesgue et $B \in \mathscr{B}(\mathbb{R})$. Alors $\frac{d\lambda|_B}{d\lambda} = \mathbb{1}_B$.*

Dans ce cours, lorsqu'on parlera de mesure à densité, on fera toujours (implicitement) référence à une mesure admettant une densité par rapport à Lebesgue. En particulier, on parlera de lois et variables aléatoires à densité. Dans ce dernier cas, on parle également de variables aléatoires continues (par opposition à discrètes qui correspond au cas atomique).

Exercice 2.2.4. *Soient μ une mesure sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ et $g : \mathbb{R} \to \mathbb{R}$ une fonction mesurable et positive. Montrer que l'application suivante est une mesure sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$:*

$$\begin{aligned} \mathscr{B}(\mathbb{R}) &\to \bar{\mathbb{R}} \\ B &\mapsto \int_B g d\mu \end{aligned}$$

Par l'exercice précédent, on déduit que toute fonction mesurable réelle g sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ positive λ-p.p. et telle que $\int g d\lambda = 1$ (où λ est la mesure de Lebesgue) définit une probabilité diffuse sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Une telle fonction est appelée densité de probabilité sur $\mathbb{R}$. Puisque l'intégrale de g par λ ne dépend pas des valeurs de g sur les ensembles négligeables de λ, la mesure sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ ayant pour densité g admet également pour densité toute fonction λ-p.p. égale à g. Puisque l'égalité presque partout est une relation d'équivalence, il est courant de voir la densité d'une mesure comme une classe d'équivalence.

Exemple. *La loi $\mu := \mathscr{U}([a,b])$ (pour $a, b \in \mathbb{R}$ avec $a < b$) admet une densité par rapport à Lebesgue donnée par:*

$$\frac{d\mu}{d\lambda} = \frac{1}{b-a}\mathbb{1}_{[a,b]}$$

De façon plus générale, pour tout $B \in \mathscr{B}(\mathbb{R})$ tel que $\lambda(B) \in \mathbb{R}_+^*$, $\mathscr{U}(B)$ admet la densité suivante contre Lebesgue:

$$\frac{d\mathscr{U}(B)}{d\lambda} = \frac{1}{\lambda(B)}\mathbb{1}_B$$

Exemple. *On appelle loi exponentielle de paramètre $\alpha \in \mathbb{R}_+^*$, notée $\mathscr{E}(\alpha)$, la probabilité sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ qui admet la densité suivante contre Lebesgue:*

$$\begin{aligned} \frac{d\mathscr{E}(\alpha)}{d\lambda} : \mathbb{R} &\to \mathbb{R}_+ \\ x &\mapsto \mathbb{1}_{\mathbb{R}_+}(x)\alpha e^{-\alpha x} \end{aligned}$$

Exemple. *Soient $m \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$. On appelle loi normale d'espérence m et de variance σ^2, notée $\mathscr{N}(m, \sigma^2)$, la probabilité sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ qui admet la densité suivante contre Lebesgue:*

$$\begin{aligned} \frac{d\mathscr{N}(m,\sigma)}{d\lambda} : \mathbb{R} &\to \mathbb{R}_+ \\ x &\mapsto \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \end{aligned}$$

Notons qu'une loi à densité sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ peut également être définie sur le support de sa densité. Par exemple, $\mathscr{E}(\alpha)$ (pour $\alpha \in \mathbb{R}_+^*$) peut être définie sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, $(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+))$ ou $(\mathbb{R}_+^*, \mathscr{B}(\mathbb{R}_+^*))$. En effet, les trois espaces (munis de $\mathscr{E}(\alpha)$) sont isomorphiques mod 0.

Exercice 2.2.5. *Soient μ une probabilité sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ admettant une densité g continue par morceaux et X une variable aléatoire réelle de loi μ et de fonction de répartition F_X. Montrer que, en tout point de continuité $x \in \mathbb{R}$ de g, F_X est dérivable en x et on a $F_X'(x) = g(x)$.*

La proposition suivante nous dit que pour intégrer une fonction f par une mesure à densité g (par rapport à Lebesgue), il suffit en fait d'intégrer la fonction produit fg par Lebesgue.

Proposition 2.2.19. *Soit λ la mesure de Lebesgue, μ une mesure sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ qui admet une densité g par rapport à Lebesgue et f une fonction réelle mesurable sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Alors f est intégrable par μ si et seulement si $\int |f| g d\lambda < +\infty$. De plus, si f est positive ou intégrable par μ, alors:*

$$\int f d\mu = \int f g d\lambda$$

Preuve

La stratégie est toujours la même, on commence par montrer le résultat pour les fonctions étagées puis on étend grâce au théorème de convergence monotone. Si $f := \mathbb{1}_B$ pour un $B \in \mathscr{B}(\mathbb{R})$, alors:

$$\int f d\mu = \int \mathbb{1}_B d\mu = \int_B d\mu = \mu(B) = \int_B g d\lambda = \int \mathbb{1}_B g d\lambda = \int f g d\lambda$$

Par linéarité de l'intégrale, on en déduit que l'égalité a lieu pour toutes les fonctions étagées positives. Supposons maintenant f positive. Par le lemme 2.2.11, il existe une suite $(s_n)_{n\in\mathbb{N}}$ de fonctions étagées positives croissante point par point qui converge simplement vers f. Par le théorème de convergence monotone, on a alors:

$$\int f d\mu = \lim_{n\to+\infty} \int s_n d\mu$$

Or, par le point précédent, pour tout $n \in \mathbb{N}$, $\int s_n d\mu = \int s_n g d\lambda$. On a donc:

$$\int f d\mu = \lim_{n\to+\infty} \int s_n g d\lambda$$

Puisque la suite $(s_n)_{n\in\mathbb{N}}$ converge simplement vers f, la suite $(s_n g)_{n\in\mathbb{N}}$ converge simplement vers fg. De même, puisque $(s_n)_{n\in\mathbb{N}}$ est positive et croissante point par point et que g est positive, la suite $(s_n g)_{n\in\mathbb{N}}$ est positive et croissante point par point. En appliquant le théorème de convergence monotone une nouvelle fois, on en déduit:

$$\int f d\mu = \int f g d\lambda$$

Prenons maintenant f mesurable de signe quelconque. Par définition, f est intégrable si et seulement si $\int |f| d\mu < +\infty$. Car $|f|$ est positive, par le point que l'on vient de montrer, cette dernière condition est équivalente à $\int |f| g d\lambda < +\infty$. Supposons donc que $\int |f| g d\lambda < +\infty$. On a alors:

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu = \int f_+ g d\lambda - \int f_- g d\lambda = \int (f_+ - f_-) g d\lambda = \int f g d\lambda.$$

où la première égalité est par définition de l'intégrale d'une fonction intégrable, la seconde par le résultat déjà montré pour les fonctions positives et la troisième par linéarité de l'intégrale.

2.3 Espérance et théorème de transfert

Nous pouvons maitenant définir la notion d'espérance d'une variable aléatoire réelle. Moralement, celle-ci est sa moyenne théorique. Mathématiquement, c'est son intégrale par la mesure de probabilité de l'espace probabilisé sur lequel elle est définie. Cet espace étant rarement explicité, cette définition peut sembler abstraite et peu pratique. On va voir en fait, avec le théorème de transfert, que la loi d'une variable aléatoire détermine son espérance.

Définition 2.3.1 (Espérance d'une variable aléatoire réelle)**.** *Soit X une variable aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbf{P})$. Si X est positive ou intégrable par rapport à $\mathbf{P}$, on appelle espérance de X la valeur suivante:*

$$\mathbf{E}[X] := \int_\Omega X d\mathbf{P}$$

$\mathbf{E}$ est appelé (l'opérateur) espérance de $(\Omega, \mathscr{A}, \mathbf{P})$ (ou simplement $\mathbf{P}$).

Exercice 2.3.1. *Soient $(\Omega, \mathscr{A}, \mathbf{P})$ un espace probabilisé et $A \in \mathscr{A}$. Montrer que $\mathbf{E}[\mathbb{1}_A] = \mathbf{P}(A)$.*

Théorème 2.3.2 (Théorème de transfert)**.** *Soit X une variable aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbf{P})$ et $\varphi : \mathbb{R} \to \mathbb{R}$ mesurable. Alors $\varphi \circ X$ est intégrable par rapport à $\mathbf{P}$ si et seulement si φ est intégrable par rapport à la loi $\mathbf{P}_X$ de X. De plus, si φ est positive ou intégrable par rapport $\mathbf{P}_X$, alors :*

$$\mathbf{E}[\varphi(X)] = \int_\Omega (\varphi \circ X) d\mathbf{P} = \int_{\mathbb{R}} \varphi d\mathbf{P}_X$$

Preuve

Comme pour la plupart des résultats d'intégration, on va d'abord montrer l'égalité pour les fonctons étagées puis généraliser aux fonctions positives par le théorème de convergence monotone. Si $\varphi = \mathbb{1}_B$ pour un $B \in \mathscr{B}(\mathbb{R})$, alors :

$$\mathbf{E}[\varphi(X)] = \int_\Omega \mathbb{1}_B(X(\omega)) \mathbf{P}(d\omega) = \int_\Omega \mathbb{1}_{X^{-1}(B)}(\omega) \mathbf{P}(d\omega) = \mathbf{P}(X^{-1}(B)) = \mathbf{P}_X(B) = \int_{\mathbb{R}} \mathbb{1}_B d\mathbf{P}_X = \int_{\mathbb{R}} \varphi d\mathbf{P}_X$$

Par linéarité de l'intégrale, on en déduit que l'égalité est vérifiée pour toutes les fonctions étagées positives. Supposons maintenant φ positive. Alors, par le lemme 2.2.11, il existe une suite $(s_n)_{n\in\mathbb{N}}$ de fonctions étagées positives croissantes point par point qui converge simplement vers φ. On remarque que la suite $(s_n \circ X)_{n\in\mathbb{N}}$ est également une suite de fonctions étagées (car d'images finies et mesurables par composition de fonctions mesurables) positives et croissante point par point qui converge simplement vers $\varphi \circ X$. En utilisant le théorème de convergence monotone deux fois, on obtient :

$$\mathbf{E}[\varphi(X)] = \int_\Omega (\varphi \circ X) d\mathbf{P} = \lim_{n\to+\infty} \int_\Omega (s_n \circ X) d\mathbf{P} = \lim_{n\to+\infty} \int_{\mathbb{R}} s_n d\mathbf{P}_X = \int_{\mathbb{R}} \varphi d\mathbf{P}_X$$

où la troisième égalité est par le point précédent. Prenons maintenant φ mesurable quelconque. Par définition, $\varphi \circ X$ est intégrable si et seulement si $\mathbf{E}[|\varphi(X)|] = \int |\varphi \circ X| d\mathbf{P} = \int (|\varphi| \circ X) d\mathbf{P} < +\infty$. Comme $|\varphi|$ est positive, on obtient en appliquant le point précédent que cette condition est équivalente à $\int_{\mathbb{R}} |\varphi| d\mathbf{P}_X < +\infty$, c'est-à-dire φ est intégrable par rapport à $\mathbf{P}_X$. Supposons donc que φ est intégrable par rapport à $\mathbf{P}_X$. On a alors :

$$\mathbf{E}[\varphi(X)] = \mathbf{E}[\varphi_+(X) - \varphi_-(X)] = \mathbf{E}[\varphi_+(X)] - \mathbf{E}[\varphi_-(X)] = \int_{\mathbb{R}} \varphi_+ d\mathbf{P}_X - \int_{\mathbb{R}} \varphi_+ d\mathbf{P}_X = \int_{\mathbb{R}} (\varphi_+ - \varphi_-) d\mathbf{P}_X = \int_{\mathbb{R}} \varphi d\mathbf{P}_X$$

où la seconde et avant-dernière égalité sont par linéarité de l'intégrale et la troisième est par le point précédent.

Remarque. *Le théorème de transfert n'est pas spécifique aux probabilités. En effet, sa preuve n'utilise à aucun moment le fait que $\mathbf{P}$ ou $\mathbf{P}_X$ soient des probabilités en particulier. Dans le cas général, avec f une fonction mesurable réelle définie sur un espace mesuré $(\mathbb{X}, \mathscr{F}, \mu)$ telle que $\varphi \circ f$ est positive ou intégrable, il se réécrit :*

$$\int_{\mathbb{X}} (\varphi \circ f) d\mu = \int_{\mathbb{R}} \varphi d(f_* \mu)$$

où $f_*\mu := \mu \circ f^{-1}$ est la mesure image de μ par f sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Exemple. *Soit X une variable aléatoire réelle de loi $\mathbf{P}_X$. Par le théorème de transfert, on a $\mathbf{E}[\mathbb{1}_B(X)] = \mathbf{P}_X(B)$ pour tout $B \in \mathscr{B}(\mathbb{R})$.*

Exemple. *Soit $X \sim \mathbf{P}_X := \mathscr{U}([a,b])$ avec $a, b \in \mathbb{R}$ tels que $a < b$. En appliquant le théorème de transfert avec φ la fonction identité, on obtient :*

$$\mathbf{E}[X] = \int_{\mathbb{R}} x \mathbf{P}_X(dx) = \int_{\mathbb{R}} x \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) \lambda(dx) = \frac{1}{b-a} \int_{[a,b]} x \lambda(dx) = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left[\frac{1}{2} x^2\right]_{x=a}^{x=b} = \frac{b+a}{2}$$

où la seconde égalité est par la proposition 2.2.19, la troisième par linéarité de l'intégrale et définition de sa restriction et la quatrième par la proposition 4.1.

Exemple. *Soit $X \sim \mathscr{E}(\alpha)$ avec $\alpha \in \mathbb{R}_+^*$. Le théorème de transfert nous donne :*

$$\mathbf{E}[X] = \int_{\mathbb{R}} x \mathbb{1}_{\mathbb{R}_+}(x) \alpha e^{-\alpha x} \lambda(dx) = \alpha \int_{\mathbb{R}_+} x e^{-\alpha x} \lambda(dx) = \alpha \int_0^{+\infty} x e^{-\alpha x} dx$$

Une intégration par parties nous donne ensuite :

$$\mathbf{E}[X] = \alpha \left(\left[x \frac{1}{-\alpha} e^{-\alpha x} \right]_{x=0}^{x\to+\infty} - \int_0^{+\infty} \frac{1}{-\alpha} e^{-\alpha x} dx \right) = \int_0^{+\infty} e^{-\alpha x} dx - \left[x e^{-\alpha x} \right]_{x=0}^{x\to+\infty} = \left[\frac{1}{-\alpha} e^{-\alpha x} \right]_{x=0}^{x\to+\infty} = \frac{1}{\alpha}$$

Exercice 2.3.2. *Soient $\alpha \in \mathbb{R}_+^*$, $m \in \mathbb{R}$ X une variable aléatoire réelle de loi $\mathbf{P}_X$ admettant la densité suivante par rapport à Lebesgue :*

$$x \mapsto \frac{1}{\pi\alpha \left(1 + \left(\frac{x-m}{\alpha}\right)^2\right)}$$

X admet-elle une espérance ? $\mathbf{P}_X$ est appelée la loi de Cauchy de position m et d'échelle α et est notée $\mathscr{C}(m, \alpha)$.

Exemple. *Soit $X \sim \mathscr{U}([\![n]\!])$ avec $n \in \mathbb{N}^*$. En appliquant le théorème de transfert, on a :*

$$\mathbf{E}[X] = \int_{\mathbb{R}} x \mathbf{P}_X(dx) = \int_{\mathbb{R}} x \left(\frac{1}{n} \sum_{k=1}^n \delta_k \right)(dx) = \frac{1}{n} \sum_{k=1}^n k = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}$$

où la troisième égalité est par le corollaire 2.2.17.

Exercice 2.3.3. *Soit X une variable aléatoire réelle de loi $\mathbf{P}_X := \mathscr{P}(\alpha)$ avec $\alpha \in \mathbb{R}_+^*$. Calculer l'espérance de X.*

L'intégrale (par une mesure) étant linéaire, l'espérance d'une somme de variables aléatoires est égale à la somme des espérances des variables aléatoires.

Exemple. *Soit X et Y deux variables aléatoires réelles définies sur un même espace $(\Omega, \mathscr{A}, \mathbf{P})$ d'espérance $\mathbf{E}$ et de lois respectives $\mathscr{E}(\alpha)$ et $\mathscr{U}(\llbracket n \rrbracket)$, avec $\alpha \in \mathbb{R}_+^*$ et $n \in \mathbb{N}^*$. Alors $Z = X + Y$ est une variable aléatoire réelle d'espérance*

$$\mathbf{E}[Z] = \mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y] = \frac{1}{\alpha} + \frac{n+1}{2} = \frac{\alpha(n+1)+2}{2\alpha}$$

Exercice 2.3.4. *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles définies sur un même espace $(\Omega, \mathscr{A}, \mathbf{P})$ d'espérance $\mathbf{E}$ telle que $\sum_{n\in\mathbb{N}} \mathbf{E}[|X_n|] < +\infty$. Justifier que $Z := \sum_{n\in\mathbb{N}} X_n$ est une variable aléatoire réelle intégrable d'espérance $\mathbf{E}[Z] = \sum_{n\in\mathbb{N}} \mathbf{E}[X_n]$.*

Pour étudier la loi d'une variable aléatoire, on va souvent s'intéresser à des statistiques de celle-ci. Dans notre contexte, on appelle statistique (théorique) de la variable aléatoire X, ou de sa loi $\mathbf{P}_X$, une valeur de la forme $\mathbf{E}[f(X)]$ (avec f une fonction mesurable telle que l'espérance est bien définie). L'espérance, par exemple, est une statistique. Une classe importante de statistiques d'une variable aléatoire est ses moments.

Définition 2.3.3 (Moments d'une variable aléatoire)**.** *Soit X une variable aléatoire réelle et $n \in \mathbb{N}$. Si $\mathbf{E}[|X|^n] < +\infty$, on dit que X admet un moment d'ordre n, qui est la valeur $\mathbf{E}[X^n] \in \mathbb{R}$.*

Lorsqu'une variable aléatoire réelle admet un moment d'ordre 1, on dit aussi qu'elle est d'espérance finie.

Exercice 2.3.5. *Soit X une variable aléatoire réelle telle que $\mathbf{E}[|X|^m] < +\infty$ avec $m \in \mathbb{R}_+^*$. Montrer que, pour tout $a \in [0, m]$, $\mathbf{E}[|X|^a] < +\infty$. En particulier, X admet un moment d'ordre $n \in \mathbb{N}$ pour tout $n \leq m$. On pourra utiliser l'inégalité $x^a \leq \max\{1, x^m\} \leq 1 + x^m$ pour tout $x \in \mathbb{R}_+$ et $a \in [0, m]$. Réciproquement, montrer que si $\mathbf{E}[|X|^m] = +\infty$, alors $\mathbf{E}[|X|^a] = +\infty$ pour tout $a \in [m, +\infty[$.*

Une autre statistique importante, fonction de ses moments, est la variance. Celle-ci mesure la dispersion moyenne (quadratique) d'une variable aléatoire autour de sa moyenne.

Définition 2.3.4 (Variance d'une variable aléatoire réelle)**.** *Soit X une variable aléatoire réelle d'espérance finie. On appelle variance de X la valeur suivante :*

$$Var(X) := \mathbf{E}\big[(X - \mathbf{E}[X])^2\big] \in \bar{\mathbb{R}}_+$$

Exercice 2.3.6. *Soit X une variable aléatoire réelle d'espérance finie. Montrer que :*

$$Var(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2$$

Cette identité remarquable est parfois connue sous le nom de "théorème de König-Huygens". En déduire que $Var(X)$ est finie si et seulement si X admet un moment d'ordre 2.

Exercice 2.3.7. *Soit X une variable aléatoire réelle d'espérance finie. Montrer que, pour tous $a, b \in \mathbb{R}$:*

$$Var(aX + b) = a^2 Var(X)$$

Par l'exercice ci-dessus, on voit immédiatement que, contrairement à l'espérance, la variance n'est pas linéaire.

Exemple. *Soit $X \sim \mathscr{U}([a, b])$ avec $a, b \in \mathbb{R}$ tels que $a < b$. En appliquant le théorème de transfert avec $\varphi : x \mapsto x^2$, on obtient :*

$$\mathbf{E}\big[X^2\big] = \int_{\mathbb{R}} x^2 \mathbf{P}_X(dx) = \int_{\mathbb{R}} x^2 \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) \lambda(dx) = \frac{1}{b-a} \int_{[a,b]} x^2 \lambda(dx) = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \left[\frac{1}{3} x^3\right]_{x=a}^{x=b} = \frac{b^3 - a^3}{3(b-a)}$$

En utilisant le théorème de König-Huygens et l'espérance de X déjà calculée dans un exemple précédent, on déduit :

$$Var(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{4(b-a)(a^2+ab+b^2) - 3(b-a)(a+b)^2}{12(b-a)} = \frac{(b-a)^2}{12}$$

Exercice 2.3.8. *Soit $X \sim \mathscr{E}(\alpha)$ avec $\alpha \in \mathbb{R}_+^*$. Calculer la variance de X.*

Exercice 2.3.9. *Soit $X \sim \mathscr{U}(\llbracket n \rrbracket)$ avec $n \in \mathbb{N}^*$. Calculer la variance de X.*

Nous donnons maintenant deux inégalités qui font intervenir l'espérance et la variance d'une variable aléatoire. La première inégalité, dite de Markov, majore la probabilité qu'une variable aléatoire positive prenne une valeur arbitrairement grande.

Proposition 2.3.5 (Inégalité de Markov). *Soit X une variable aléatoire positive. Alors, pour tout $a \in \mathbb{R}_+^*$:*

$$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}[X]}{a}$$

Preuve

Soit $a \in \mathbb{R}_+^*$. Si X n'est pas d'espérance finie alors l'inégalité est trivialement vérifiée car le terme de droite est égal à $+\infty$. Supposons donc $\mathbf{E}[X] \in \mathbb{R}_+$. On a alors :

$$\mathbf{E}[X] = \int_{\mathbb{R}} x\mathbf{P}_X(dx) = \int_{]-\infty,a[} x\mathbf{P}_X(dx) + \int_{[a,+\infty[} x\mathbf{P}_X(dx) \geq \int_{[a,+\infty[} x\mathbf{P}_X(dx)$$

où la première égalité est par le théorème de transfert, la seconde par la relation de Chasles et la dernière inégalité par positivité de X (en effet $\int_{]-\infty,a[} x\mathbf{P}_X(dx) = \mathbf{E}[\mathbb{1}_{]-\infty,a[}(X)X] \geq 0$). De plus, pour tout $x \in [a,+\infty[$, $x \geq a$ (trivialement). Par monotonicité de l'intégrale, on a donc :

$$\int_{[a,+\infty[} x\mathbf{P}_X(dx) \geq \int_{[a,+\infty[} a\mathbf{P}_X(dx) = a\int_{[a,+\infty[} \mathbf{P}_X(dx) = a\mathbf{P}_X([a,+\infty[) = a\mathbf{P}(X \geq a)$$

où la première égalité est par linéarité de l'intégrale et la dernière par définition de la loi d'une variable aléatoire.

Exemple. *Soit $Z = X + Y$, où X et Y sont des variables aléatoires réelles définies sur un même espace et de lois respectives $\mathscr{E}(\alpha)$ et $\mathscr{U}([\![n]\!])$, avec $\alpha \in \mathbb{R}_+^*$ et $n \in \mathbb{N}^*$. Alors, pour tout $a \in \mathbb{R}_+^*$:*

$$\mathbf{P}(Z \geq a) \leq \frac{\alpha(n+1)+2}{2\alpha a}$$

Exercice 2.3.10. *Soit X une variable aléatoire positive telle que $\mathbf{E}[X] = 0$. Montrer que X est presque sûrement nulle. On pourra utiliser le théorème de la limite monotone.*

La seconde inégalité, dite de Bienaymé-Tchebichev, majore la probabilité qu'une variable aléatoire s'éloigne trop de sa moyenne.

Corollaire 2.3.6 (Inégalité de Bienaymé-Tchebychev). *Soit X une variable aléatoire réelle d'espérance finie. Alors, pour tout $a \in \mathbb{R}_+^*$:*

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq a) \leq \frac{Var(X)}{a^2}$$

Preuve

Soit $a \in \mathbb{R}_+^*$. Si $Var(X) = +\infty$, l'inégalité est trivialement vérifiée. Sinon, en appliquant l'inégalité de Markov à la variable positive $(X - \mathbf{E}[X])^2$, on obtient :

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq a) = \mathbf{P}((X - \mathbf{E}[X])^2 \geq a^2) \leq \frac{\mathbf{E}[(X - \mathbf{E}[X])^2]}{a^2} = \frac{Var(X)}{a^2}$$

Exemple. *Soit $X \sim \mathscr{E}(\alpha)$ avec $\alpha \in \mathbb{R}_+^*$. On admet que $Var(X) = \frac{1}{\alpha^2}$. On a alors, pour tout $a \in \mathbb{R}_+^*$:*

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq a) \leq \frac{1}{(\alpha a)^2}$$

Avec ces deux inégalités, on voit que, sans connaître la loi d'une variable aléatoire, on peut déjà restreindre les probabilités de certains événements à partir de la simple donnée de quelques statistiques (en l'occurence l'espérance et la variance). De plus, ces deux inégalités s'appliquent à toutes les lois (en particulier diffuses et atomiques) et leur preuve est relativement simple une fois que l'on est familiarisé avec le formalisme et les résultats fondamentaux de la théorie de la mesure. On peut ainsi contempler l'intérêt de cette approche unificatrice et axiomatique des probabilités.

Théorème 2.3.7 (Loi faible des grands nombres). *Soit (X_n) une suite de variables aléatoires réelles indépendantes, non constantes, de même loi. On suppose que pour tout n, X_n admet un moment d'ordre 2. On note m l'espérance commune et σ l'écart-type commun. On suppose $S_n = X_1 + ... + X_n$ et $\bar{X}_n = \frac{1}{n}S_n$ la fréquence statistique. Alors, pour tout $\epsilon \geq$)*

$$\mathbf{P}(|\bar{X}_n - m| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

. *En particulier, pour tout* $\epsilon \geq 0$

$$\mathbf{P}(|\bar{X}_n - m| \geq \epsilon) \underset{n\to\infty}{\to} 0.$$

Exemple. *Soit* $(A_n)_{n\in\mathbb{N}}$ *une suite d'événements indépendants, ayant tous la même probabilité* p*, inconnue. On cherche une estimation de* p*. On a*

$$\mathbf{P}(|\bar{X}_n - p| \geq \epsilon) \leq \frac{pq}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$$

Ainsi, $\bar{X}_n$ *serait une estimation de* p *à* ϵ *près avec une probabilité d'erreur inférieur à* 10^{-m} *près dès que* $\frac{10^{-m}}{4\epsilon^2} \leq n$*.*

Chapitre 3. Probabilités multidimensionelles

Dans le premier chapitre, nous avons vu qu'un espace probabilisé $(\Omega, \mathscr{A}, \mathbf{P})$ modélise un phénomène aléatoire. Une variable aléatoire réelle X sur $(\Omega, \mathscr{A}, \mathbf{P})$ peut alors être interprétée comme une observable numérique de ce dernier. Celle-ci transfère la probabilité $\mathbf{P}$ sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Cette probabilité "image" est appelée la loi de X. Dans ce chapitre, nous allons nous intéresser à ce qu'il se passe lorsqu'on observe plusieurs valeurs numériques provenant d'une même réalisation d'un phénomène. Plus formellement, prenons une seconde variable aléatoire Y sur $(\Omega, \mathscr{A}, \mathbf{P})$ et une fonction $\varphi : \mathbb{R}^2 \to \mathbb{R}$ mesurable (pour une tribu sur $\mathbb{R}^2$ à préciser). Une question naturelle est alors si $\varphi(X, Y)$ est une variable aléatoire réelle et, dans ce cas, si sa loi est déterminée par celles de X et Y. Nous pouvons déjà répondre partiellement par un exemple illustratif très simple. On possède une urne contenant 4 jetons numérotés de 1 à 4 et on en tire deux "au hasard" (uniformément) et sans remise. On note X le résultat du premier tirage et Y celui du second. Cette exéprience est naturellement modélisée par l'espace probabilisé $(\Omega, \mathscr{A}, \mathbf{P})$ où $\Omega := \{1,2,3,4\}^2$, $\mathscr{A} := \mathscr{P}(\Omega)$ et $\mathbf{P}$ est donnée, pour $(m,n) \in \Omega$, par:

$$\mathbf{P}(\{(m,n)\}) = \begin{cases} \frac{1}{12}, & \text{si } m \neq n \\ 0, & \text{si } m = n \end{cases}$$

X et Y sont alors des variables aléatoires réelles sur $(\Omega, \mathscr{A}, \mathbf{P})$ données par $X((m,n)) = m$ et $Y((m,n)) = n$ pour $(m,n) \in \Omega$. On vérifie facilement que X et Y ont toutes deux pour loi $\mathscr{U}(\{1,2,3,4\})$. En effet, pour $m, n \in \{1,2,3,4\}$:

$$\mathbf{P}(X = m) = \mathbf{P}((X = m) \wedge (Y \in \{1,2,3,4\})) = \mathbf{P}(\{m\} \times \{1,2,3,4\}) = \frac{3}{12} = \frac{1}{4}$$
$$\mathbf{P}(Y = n) = \mathbf{P}((X \in \{1,2,3,4\}) \wedge (Y = n)) = \mathbf{P}(\{1,2,3,4\} \times \{n\}) = \frac{3}{12} = \frac{1}{4}$$

En posant la variable aléatoire réelle $Z := XY$, on observe alors que:

$$\mathbf{P}(Z = 1) = \mathbf{P}((X = 1) \wedge (Y = 1)) = \mathbf{P}(\{X = 1\} \cap \{Y = 1\}) = \mathbf{P}(\{(1,1)\}) = 0$$

On s'intéresse maintenant au cas d'un tirage avec remise. Ce changement peut être pris en compte en considérant la probabilité $\mathbf{P}' = \mathscr{U}(\Omega)$ sur $(\Omega, \mathscr{A})$. On vérifie immédiatement que X et Y ont toujours pour loi $\mathscr{U}(\{1,2,3,4\})$ sous $\mathbf{P}'$. Pour autant:

$$\mathbf{P}'(Z = 1) = \mathbf{P}'(\{(1,1)\}) = \frac{1}{16} \neq 0 = \mathbf{P}(Z = 1)$$

Ainsi, la loi de Z n'est pas déterminée par celles de X et Y. Cette différence s'explique par la dépendance entre X et Y dans le tirage sans remise. Pour déterminer la loi de Z, il est de fait nécessaire de connaître la loi du couple (X, Y), aussi appelée loi jointe (qui est ici $\mathbf{P}$ ou $\mathbf{P}'$ car (X, Y) est la fonction identité sur Ω). On peut toutefois observer que, lors du tirage avec remise, on a $\mathbf{P}'((X = m) \wedge (Y = n)) = \mathbf{P}'(X = m)\mathbf{P}'(Y = n)$ pour tous $(m,n) \in \Omega$. On dit que les variables X et Y sont indépendantes (sous $\mathbf{P}'$). Sous cette hypothèse, connaître les lois de X et Y est suffisant pour déterminer leur loi jointe. L'espace $(\Omega, \mathscr{A}, \mathbf{P}')$ peut alors être noté $(\{1,2,3,4\}^2, \mathscr{P}(\{1,2,3,4\})^{\otimes 2}, \mathscr{U}(\{1,2,3,4\})^{\otimes 2})$ et est dit espace produit de $(\{1,2,3,4\}, \mathscr{P}(\{1,2,3,4\}), \mathscr{U}(\{1,2,3,4\}))$ avec lui même. Le but de ce chapitre est de formaliser et étudier ces notions d'espace produit, de loi jointe et de dépendance pour des espaces probabilisés et variables aléatoires réelles quelconques.

3.1 Espace produit et théorème de Fubini-Tonelli

3.1.1 Tribu produit

Comme nous l'avons vu dans l'exemple introductif de ce chapitre, étant donné un espace probabilisé fini $(\Omega, \mathscr{A}, \mathbf{P})$, il est possible de considérer son espace produit $(\Omega^2, \mathscr{A}^{\otimes 2}, \mathbf{P}^{\otimes 2})$ qui modélise deux réalisations indépendantes du même phénomène aléatoire. De manière plus générale, étant donné $n \in \mathbb{N}^*$ espaces probabilisés $(\Omega_1, \mathscr{A}_1, \mathbf{P}_1), \ldots, (\Omega_n, \mathscr{A}_n, \mathbf{P}_n)$, on peut définir l'espace produit $(\Omega_1 \times \cdots \times \Omega_n, \mathscr{A}_1 \otimes \cdots \otimes \mathscr{A}_n, \mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_n)$ qui modélise une réalisation indépendante et simultanée de chacun des phénonèmes modélisés par les $(\Omega_i, \mathscr{A}_i, \mathbf{P}_i)$ $(i \in [\![n]\!])$. En fait, cette construction théorique n'est pas spécifique aux espaces probabilisés et on peut, par exemple, considérer l'espace produit $(\mathbb{R}^n, \mathscr{B}(\mathbb{R})^{\otimes n}, \lambda^{\otimes n})$, où λ est la mesure de Lebesgue sur $\mathbb{R}$. Ce dernier jouera le même rôle fondamental en dimension n que $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda)$ en dimension 1. Avant de définir formellement le produit d'espaces mesurés, on va s'intéresser au produit d'espaces mesurables et, en particulier, au produit de tribus. Dans cette section, $n \in \mathbb{N}^*$ est un entier naturel non nul, $((E_i, \mathscr{E}_i))_{i \in [\![n]\!]}$ une famille d'espaces mesurables quelconques et $E := \prod_{i=1}^n E_i$ le produit cartésien des $(E_i)_{i \in [\![n]\!]}$.

Définition 3.1.1 (Tribu produit). *On appelle tribu produit (des $(\mathscr{E}_i)_{i \in [\![n]\!]}$), notée $\bigotimes_{i \in [\![n]\!]} \mathscr{E}_i$, la tribu suivante sur E:*

$$\bigotimes_{i \in [\![n]\!]} \mathscr{E}_i := \mathscr{E}_1 \otimes \cdots \otimes \mathscr{E}_n := \sigma(\mathscr{R})$$

où

$$\mathscr{R} := \left\{ B_1 \times \cdots \times B_n : (B_i)_{i\in[\![n]\!]} \in \prod_{i=1}^{n} \mathscr{E}_i \right\}$$

Remarque. *Les éléments $\mathscr{R}$ dans la définition précédente sont appelés pavés ou plus rarement rectangles. Ce dernier terme est notamment celui utilisé en anglais, parfois avec la précision "measurable rectangle". On évitera ici celui-ci afin de ne pas le confondre avec les rectangles de $\mathbb{R}^n$, qui auront aussi un rôle important.*

Notation. *Pour un espace mesurable $(\mathbb{X}, \mathscr{F})$, on note $\mathscr{F}^{\otimes n} := \mathscr{F} \otimes \cdots \otimes \mathscr{F}$ la tribu produit sur $\mathbb{X}^n$.*

On rappelle qu'on peut naturellement identifier $E_1 \times E_2 \times E_3$ (dont les éléments sont de la forme (x, y, z)) avec $E_1 \times (E_2 \times E_3)$ ou $(E_1 \times E_2) \times E_3$ (dont les éléments sont de la forme $(x, (y, z))$ ou $((x, y), z)$). De la même manière, la tribu produit $\mathscr{E}_1 \otimes \mathscr{E}_2 \otimes \mathscr{E}_3$ s'identifie avec $\mathscr{E}_1 \otimes (\mathscr{E}_2 \otimes \mathscr{E}_3)$ ou $(\mathscr{E}_1 \otimes \mathscr{E}_2) \otimes \mathscr{E}_3$. On considèrera qu'elles sont égales. En jargon algébrique, l'opération $\otimes$ est dite associative. Tout comme le produit cartésien, elle n'est en revanche pas commutative, c'est-à-dire que l'ordre des tribus dans le produit est important (i.e. $\mathscr{E}_1 \otimes \mathscr{E}_2 \neq \mathscr{E}_2 \otimes \mathscr{E}_1$ si $\mathscr{E}_1 \neq \mathscr{E}_2$). Dans le reste de la section, on notera $\mathscr{E} := \bigotimes_{i\in[\![n]\!]} \mathscr{E}_i$. L'espace $(E, \mathscr{E})$ est appelé espace (mesurable) produit (des $((E_i, \mathscr{E}_i))_{i\in[\![n]\!]}$). Si la tribu produit rend tous les pavés (c'est-à-dire les produits d'ensembles mesurables) mesurables, tous ses éléments ne sont pas des pavés.

Exemple. *Soient $E_1 := \{0, 1\}$ et $E_2 := \{a, b\}$ munis respectivement de leurs tribus discrètes $\mathscr{E}_1$ et $\mathscr{E}_2$. On pose $E := E_1 \times E_2 := \{(0, a), (0, b), (1, a), (1, b)\}$ et $\mathscr{E} := \mathscr{E}_1 \otimes \mathscr{E}_2$. On a $\{(0, a)\} := \{0\} \times \{a\}$, c'est-à-dire $\{(0, a)\}$ est un pavé et donc dans $\mathscr{E}$. De même, par stabilité par complémentarité, $\{(0, a)\}^c \in \mathscr{E}$. Pourtant $\{(0, a)\}^c := E\backslash\{(0, a)\} = \{(0, b), (1, a), (1, b)\}$ n'est pas un pavé.*

Exercice 3.1.1. *Si les ensembles $(E_i)_{i\in[\![n]\!]}$ sont dénombrables et $(\mathscr{E}_i)_{i\in[\![n]\!]}$ leurs tribus discrètes respectives, alors :*

$$\mathscr{E} = \mathscr{P}(E)$$

L'inclusion $\subset$ est triviale, $\mathscr{P}(E)$ étant la plus grande tribu sur E.

Pour $i \in [\![n]\!]$, on définit la projection de E sur E_i comme l'application suivante :

$$\begin{aligned} \pi_i : E &\to E_i \\ (x_1, \ldots, x_n) &\mapsto x_i \end{aligned}$$

Proposition 3.1.2 (Tribu engendrée par les projections)**.** *$\mathscr{E}$ est la tribu engendrée par la famille $(\pi_i)_{i\in[\![n]\!]}$, c'est-à-dire la plus petite tribu qui rend π_i mesurable pour $i \in [\![n]\!]$.*

Preuve

Soit $\mathscr{G} := \left\{\pi_i^{-1}(B_i) : i \in [\![n]\!], B_i \in \mathscr{E}_i\right\}$. On veut montrer que $\sigma(\mathscr{R}) = \sigma(\mathscr{G})$. Pour tous $(B_i)_{i\in[\![n]\!]} \in \prod_{i=1}^n \mathscr{E}_i$, on a, pour $i \in [\![n]\!]$:

$$\pi_i^{-1}(B_i) = E_1 \times \cdots \times E_{i-1} \times B_i \times E_{i+1} \times \cdots \times E_n \in \mathscr{R}$$

On en déduit que, pour $i \in [\![n]\!]$, π_i est mesurable, c'est-à-dire $\sigma(\mathscr{G}) \subset \sigma(\mathscr{R})$. Or, pour tous $(B_i)_{i\in[\![n]\!]} \in \prod_{i=1}^n \mathscr{E}_i$:

$$B_1 \times \cdots \times B_n = \bigcap_{i=1}^{n} \pi_i^{-1}(B_i) \in \sigma(\mathscr{G})$$

C'est-à-dire, $\mathscr{R} \subset \sigma(\mathscr{G})$ et donc $\sigma(\mathscr{R}) \subset \sigma(\sigma(\mathscr{G})) = \sigma(\mathscr{G})$. On en déduit le résultat souhaité.

Étant donné une fonction f d'un ensemble $\mathbb{X}$ dans E, les fonctions $f_i := \pi_i \circ f : \mathbb{X} \to E_i$, pour $i \in [\![n]\!]$, sont appelées les (fonctions) coordonnées de f. Pour $x \in \mathbb{X}$, $f(x)$ peut alors s'écrire $f(x) := (f_1(x), \ldots, f_n(x)) \in E$. Un corollaire simple de la proposition précédente est qu'une telle fonction est mesurable (lorsqu'on équipe $\mathbb{X}$ d'une tribu) si et seulement si ses coordonnées le sont.

Corollaire 3.1.3. *Soient $(\mathbb{X}, \mathscr{F})$ un espace mesurable et $f : \mathbb{X} \to E$. f est mesurable si et seulement si, pour $i \in [\![n]\!]$, $\pi_i \circ f$ est mesurable.*

Preuve

Soit $f : \mathbb{X} \to E$. Si f est mesurable, alors par la proposition 3.1.2 et par composition de fonctions mesurables, $\pi_i \circ f$ est mesurable pour tout $i \in [\![n]\!]$. Supposons que $\pi_i \circ f$ est mesurable pour tout $i \in [\![n]\!]$. On rappelle que pour montrer la mesurabilité d'une fonction, il est suffisant de montrer que l'image réciproque de tout ensemble d'une famille génératrice soit mesurable. Soit $P := B_1 \times \cdots \times B_n \in \mathscr{R}$. On a:

$$P = B_1 \times \cdots \times B_n = \bigcap_{i\in[\![n]\!]} \pi_i^{-1}(B_i)$$

Donc, par préservation des opérations ensemblistes par l'image réciproque:

$$f^{-1}(P) = f^{-1}\left(\bigcap_{i\in[\![n]\!]} \pi_i^{-1}(B_i)\right) = \bigcap_{i\in[\![n]\!]} f^{-1}\left(\pi_i^{-1}(B_i)\right) = \bigcap_{i\in[\![n]\!]} (\pi_i \circ f)^{-1}(B_i)$$

Par mesurabilité de $\pi_i \circ f$, $(\pi_i \circ f)^{-1}(B_i) \in \mathscr{F}$ pour $i \in [\![n]\!]$. Par stabilité par intersection finie de $\mathscr{F}$, on en déduit que $f^{-1}(P) \in \mathscr{F}$, c'est-à-dire f est mesurable.

On rappelle que la tribu Borélienne sur $\mathbb{R}$ est engendrée par les intervalles ouverts et contient tous les ouverts de $\mathbb{R}$ (car un ouvert sur $\mathbb{R}$ est une réunion dénombrable d'intervalles ouverts). En particulier, elle est engendrée par les ouverts de $\mathbb{R}$. On peut en fait définir les boréliens sur $\mathbb{R}^n$ en se basant sur cette propriété.

Définition 3.1.4 (Tribu borélienne)**.** *On appelle tribu borélienne de $\mathbb{R}^n$, notée $\mathscr{B}(\mathbb{R}^n)$, la tribu sur $\mathbb{R}^n$ engendrée par les ouverts de $\mathbb{R}^n$.*

Remarque. *De manière plus générale, la tribu borélienne peut être définie identiquement sur n'importe quel espace topologique (c'est-à-dire la donnée d'un ensemble et de ses ouverts). En pratique, les tribus considérées dans la plupart des domaines de recherche mathématique sont toujours boréliennes pour une topologie donnée.*

Une question naturelle est de se demander si les espaces mesurables $(\mathbb{R}^n, \mathscr{B}(\mathbb{R})^{\otimes n})$ et $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ sont égaux. La réponse est positive.

Proposition 3.1.5. *$\mathscr{B}(\mathbb{R}^n) = \mathscr{B}(\mathbb{R})^{\otimes n}$.*

La preuve de cette proposition, non détaillée dans ce cours, repose sur deux résultats :

- $\mathscr{E} := \mathscr{B}(\mathbb{R})^{\otimes n} = \sigma(\mathscr{H})$ où $\mathscr{H} := \{]a_1, b_1[\times\cdots\times]a_n, b_n[: a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}\}$ est l'ensemble des hyperrectangles ouverts dans $\mathbb{R}^n$;
- Tout ouvert de $\mathbb{R}^n$ peut s'écrire comme réunion dénombrables d'hyperrectangles ouverts de $\mathbb{R}^n$. Il généralise la propriété d'un ouvert sur $\mathbb{R}$ de s'écrire comme réunion dénombrable d'intervalles ouverts.

Par défaut, on considèrera toujours la tribu $\mathscr{B}(\mathbb{R}^n)$ sur $\mathbb{R}^n$.

On rappelle (théorème 2.1.5) que toute fonction de $\mathbb{R}$ dans $\mathbb{R}$ continue par morceaux est mesurable. Pour une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$, il n'y a pas de notion équivalente. En utilisant les mêmes arguments que ceux de la preuve du théorème 2.1.5, on peut en revanche obtenir le résultat qui suit. On rappelle d'abord quelques définitions équivalentes de la continuité sur $\mathbb{R}^n$. Une fonction $f : \mathbb{R}^n \to \mathbb{R}$ est continue en un point $x \in \mathbb{R}^n$ si l'une des deux conditions équivalentes suivantes est satisfaite:

- Pour tout $\varepsilon > 0$, il existe un $\delta > 0$ tel que, pour tout $y \in \mathbb{R}^n$ tel que $\|y - x\| < \delta$, $|f(y) - f(x)| < \varepsilon$.
- Pour toute suite $(x_n)_{n\in\mathbb{N}}$ convergeant vers x (i.e. telle que $\lim_{n\to+\infty} \|x_n - x\| = 0$), $\lim_{n\to+\infty} f(x_n) = f(x)$.

où $\|\cdot\|$ est n'importe quelle norme sur $\mathbb{R}^n$. On dit de plus que f est continue sur un ouvert O de $\mathbb{R}^n$ si l'une des deux conditions équivalentes suivantes est satisfaite:

- f est continue en tout point x de O.
- $f^{-1}(I) \cap O$ est ouvert (dans $\mathbb{R}^n$) pour tout intervalle ouvert I de $\mathbb{R}$.

Proposition 3.1.6. *Toute fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ continue sur des ouverts (de $\mathbb{R}^n$) et nulle ailleurs est mesurable.*

Preuve

Soit $f : \mathbb{R}^n \to \mathbb{R}$ continue sur des ouverts $(O_i)_{i\in I}$ et nulle ailleurs. On rappelle que, par le lemme 2.1.6, $\mathscr{F} := \left\{A \in \mathbb{R} : f^{-1}(A) \in \mathscr{B}(\mathbb{R}^n)\right\}$ est la plus grande tribu sur $\mathbb{R}$ qui rend f mesurable. Puisque f est mesurable pour toute sous-tribu de $\mathscr{F}$, on souhaite montrer que $\mathscr{B}(\mathbb{R}) \subset \mathscr{F}$. Soit $O := \bigcup_{i\in I} O_i$. Car toute réunion d'ouverts est ouverte, O est ouvert. De plus, f est continue sur O. C'est-à-dire, pour tout intervalle ouvert $I \subset \mathbb{R}$, $f^{-1}(I) \cap O \subset \mathbb{R}^n$ est ouvert. Soit $I \subset \mathbb{R}$ une intervalle ouvert. Si $0 \notin I$, alors $f^{-1}(I) \subset O$ et donc $f^{-1}(I) = f^{-1}(I) \cap O$ est ouvert (par continuité de f sur O). En particulier, $f^{-1}(I) \in \mathscr{B}(\mathbb{R}^n)$ (car, par définition, $\mathscr{B}(\mathbb{R}^n)$ est engendrée par les ouverts). Sinon, si $0 \in I$, alors $f^{-1}(I) = (f^{-1}(I) \cap O) \cup O^c$. Comme O est ouvert, $O \in \mathscr{B}(\mathbb{R}^n)$ et donc, par stabilité par complémentarité de $\mathscr{B}(\mathbb{R}^n)$, $O^c \in \mathscr{B}(\mathbb{R}^n)$. De même, $f^{-1}(I) \cap O$ est ouvert (par continuité de f) et donc dans $\mathscr{B}(\mathbb{R}^n)$. Par stabilité par union finie de $\mathscr{B}(\mathbb{R}^n)$, on en déduit que $f^{-1}(I) \in \mathscr{B}(\mathbb{R}^n)$. On vient de montrer que l'image réciproque de tout intervale ouvert par f est dans $\mathscr{B}(\mathbb{R}^n)$. C'est-à-dire, $\mathscr{F}$ contient tous les intervalles ouverts. Puisque ceux-ci engendrent $\mathscr{B}(\mathbb{R})$, on a $\mathscr{B}(\mathbb{R}) \subset \mathscr{F}$, ce qui montre que f est mesurable.

Corollaire 3.1.7. *Soit $m \in \mathbb{N}^*$. Toute fonction $f : \mathbb{R}^n \to \mathbb{R}^m$ continue sur des ouverts et nulle ailleurs est mesurable.*

Preuve

Soit $f : \mathbb{R}^n \to \mathbb{R}^m$. f est continue sur un ouvert O si et seulement si, pour $i \in [\![m]\!]$, $f_i := \pi \circ f : \mathbb{R}^n \to \mathbb{R}$ est continue sur O. Supposons donc f continue sur des ouverts et nulle ailleurs. Alors, par la proposition 3.1.6, f_i est mesurable pour $i \in [\![m]\!]$. Par le corollaire 3.1.3, on en déduit que f est mesurable.

Exemple. *La fonction $f : \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto x^2 + y^2$ est continue et donc mesurable. On en déduit que le cercle $S_1 := \{(x, y) \in \mathbb{R}^2, x^2 + y^2 = 1\}$ est mesurable (i.e. $S_1 \in \mathscr{B}(\mathbb{R}^2)$). En effet, on a $S_1 = f^{-1}(\{1\})$ et $\{1\} \in \mathscr{B}(\mathbb{R})$.*

Comme $\mathscr{B}(\mathbb{R}^n)$ contient les singletons, $\mathscr{P}(F) \subset \mathscr{B}(\mathbb{R}^n)$ pour tout sous-ensemble $F \subset \mathbb{R}^n$ dénombrable. On en déduit que, identiquement à ce qui se passe en dimension 1, toute fonction mesurable dans $(F, \mathscr{P}(F))$ l'est également dans $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ (et réciproquement si son image dans $\mathbb{R}^n$ est inclue dans F). $\mathscr{P}(F)$ est en fait la tribu trace de $\mathscr{B}(\mathbb{R}^n)$ sur F.

Exercice 3.1.2. *Soit $(\mathbb{X}, \mathscr{F})$ un espace mesurable et f une fonction mesurable de $(\mathbb{X}, \mathscr{F})$ dans $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ d'image finie. Montrer que les coordonnées de f sont des fonctions étagées.*

3.1.2 Mesure produit et intégration par cette mesure

Nous pouvons maintenant définir le produit de mesures et, plus généralement, d'espaces mesurés. Celui-ci n'est proprement défini que pour des mesures et des espaces espaces σ-finis (cf. définition **??**). Dans la suite de la section, pour $i \in [\![n]\!]$, μ_i est une mesure σ-finie sur $(E_i, \mathscr{E}_i)$.

Théorème 3.1.8 (Mesure produit). *Il existe une unique mesure sur $(E, \mathscr{E})$, notée $\bigotimes_{i\in[\![n]\!]} \mu_i := \mu_1 \otimes \cdots \otimes \mu_n$, telle que, pour tous $(B_i)_{i\in[\![n]\!]} \in \prod_{i=1}^n \mathscr{E}_i$:*

$$\left(\bigotimes_{i\in[\![n]\!]} \mu_i\right)(B_1 \times \cdots \times B_n) = \prod_{i=1}^n \mu_i(B_i)$$

Preuve

Admis. Tout comme l'existence de la mesure de Lebesgue, l'existence d'une mesure produit découle du théorème d'extension de Carathéodory. L'unicité est une conséquence du théorème **??**.

Comme pour les tribus, le produit de mesures est associatif mais non commutatif. Dans la suite, on écrira $\mu := \bigotimes_{i\in[\![n]\!]} \mu_i$. μ et $(E, \mathscr{E}, \mu)$ sont respectivement appelés mesure produit (des $(\mu_i)_{i\in[\![n]\!]}$) et espace (mesuré) produit (des $((E_i, \mathscr{E}_i, \mu_i))_{i\in[\![n]\!]}$).

Notation. *Soit $(\mathbb{X}, \mathscr{F}, \nu)$ un espace mesuré σ-fini. On note $\nu^{\otimes n} := \nu \otimes \cdots \otimes \nu$ la mesure produit sur $(\mathbb{X}^n, \mathscr{F}^{\otimes n})$.*

Exercice 3.1.3. *Montrer que $(E, \mathscr{E}, \mu)$ est σ-fini.*

Exemple (Produit de mesures de comptage). *Si les $((E_i, \mathscr{E}_i))_{i\in[\![n]\!]}$ sont dénombrables et discrets et $(\mu_i)_{i\in[\![n]\!]}$ leurs mesures de comptage respectives, alors μ est la mesure de comptage sur $(E, \mathscr{E})$. On rappelle que, dans ce cas, $\mathscr{E} := \mathscr{P}(E)$. De plus, les singletons de E sont des pavés (car, pour $i \in [\![n]\!]$, les singletons de E_i sont dans $\mathscr{E}_i := \mathscr{P}(E_i)$). Le théorème 3.1.8 nous dit alors que, pour tout singleton $\{(e_1, \ldots, e_n)\} = \{e_1\} \times \cdots \times \{e_n\} \in \prod_{i=1}^n \mathscr{E}_i \subset \mathscr{E}$ (où $e_i \in E_i$ pour $i \in [\![n]\!]$), on a:*

$$\mu(\{(e_1, \ldots, e_n)\}) = \prod_{i=1}^n \mu_i(\{e_i\}) = \prod_{i=1}^n |\{e_i\}| = \prod_{i=1}^n 1 = 1$$

μ donne ainsi masse 1 à tous les singletons de E. On en déduit que c'est nécessairement la mesure de comptage.

Exemple (Espace probabilisé produit). *Si les $((E_i, \mathscr{E}_i, \mu_i))_{i\in[\![n]\!]}$ sont des espaces probabilisés, alors $(E, \mathscr{E}, \mu)$ est également un espace probabilisé. En effet, c'est un espace mesuré et on vérifie facilement que μ est de masse 1:*

$$\mu(E) = \mu\left(\prod_{i=1}^n E_i\right) = \prod_{i=1}^n \mu_i(E_i) = \prod_{i=1}^n 1 = 1$$

Il modélise un phénomène aléatoire lui même composé de plusieurs phénomènes aléatoires indépendants (modélisés par les espaces $((E_i, \mathscr{E}_i, \mu_i))_{i\in[\![n]\!]}$). On a déjà vu de tels cas, notamment avec les lancers successifs (et indépendants) de dés ou de pièces de monnaie. Par exemple, si l'espace probabilisé $(\{0, 1\}, \mathscr{P}(\{0, 1\}), \mathscr{U}(\{0, 1\}))$ représente le lancer d'une pièce de monnaie, alors l'espace produit $(\{0, 1\}^n, \mathscr{P}(\{0, 1\})^{\otimes n}, \mathscr{U}(\{0, 1\})^{\otimes n}) = (\{0, 1\}^n, \mathscr{P}(\{0, 1\}^n), \mathscr{U}(\{0, 1\}^n))$ représente la répétition de cette expérience n fois de manière indépendante.

Exercice 3.1.4. *On suppose que, pour $i \in [\![n]\!]$, E_i est fini, $\mathscr{E}_i := \mathscr{P}(E_i)$ et $\mu_i := \mathscr{U}(E_i)$. Montrer que $\mu = \mathscr{U}(E)$.*

On rappelle que la mesure de Lebesgue sur $\mathbb{R}$ correspond à la notion de longueur. De manière plus générale, il est possible de définir son équivalent sur $\mathbb{R}^n$. Ce dernier correspond alors à la notion de n-volume. En particulier, le 2-volume d'une surface est son aire et le 3-volume d'un solide son volume (le 1-volume d'un segment est donc sa longueur).

Définition 3.1.9 (Mesure de Lebesgue sur $\mathbb{R}^n$)**.** *On appelle mesure de Lebesgue sur $\mathbb{R}^n$ la mesure $\lambda_n := \lambda^{\otimes n}$ sur $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$, où λ est la mesure de Lebesgue sur $\mathbb{R}$.*

Remarque. *La mesure de Lebesgue sur $\mathbb{R}^n$ est bien définie en tant que mesure produit car $\mathscr{B}(\mathbb{R}^n) = \mathscr{B}(\mathbb{R})^{\otimes n}$.*

Exemple. *Soient $a, b, c, d \in \mathbb{R}$ et considérons le rectangle $[a, b] \times [c, d] \in \mathscr{B}(\mathbb{R}^2)$. Son aire (ou 2-volume) est donnée par :*

$$\lambda_2([a, b] \times [c, d]) = \lambda([a, b])\lambda([c, d]) = (b - a)(d - c)$$

Couramment, on parlera simplement de volume pour désigner le n-volume. Il faut toutefois faire attention que, si le n-volume d'un objet est fini, alors son k-volume est nul pour tout $k \in \mathbb{N}$ tel que $k > n$. De même, si le n-volume d'un objet est strictement positif, alors k-volume est infini pour tout $k \in [\![n - 1]\!]$. Le volume n'a donc de sens que lorsqu'on fixe une dimension.

Exemple. *En reprenant l'exemple précédent, on peut identifier le rectangle $[a, b] \times [c, d]$ dans $\mathbb{R}^2$ au rectangle $[a, b] \times [c, d] \times \{0\}$ dans $\mathbb{R}^3$. Toutefois, le volume de ce dernier dans $\mathbb{R}^3$ est nul :*

$$\lambda_3([a, b] \times [c, d] \times \{0\}) = \lambda([a, b])\lambda([c, d])\lambda(\{0\}) = 0$$

où la seconde égalité est car $\lambda(\{0\}) = 0$.

On sait donc comment mesurer le (n-)volume d'un (hyper-)rectangle. Le théorème 3.1.8 ne nous informe toutefois pas directement sur la mesure des ensembles qui ne sont pas des pavés. Le théorème de Fubini-Tonelli nous dit comment mesurer ces ensembles et, de manière plus générale, comment calculer l'intégrale d'une fonction mesurable réelle sur un espace produit.

Avant d'énoncer celui-ci, il convient de donner deux résultats de mesurabilité qui sont nécessaires à la bonne définition des termes du théorème:

- Une fonction g mesurable sur E est mesurable en chacune de ses variables (lorsqu'on fixe les autres).

  Mathématiquement, soient $(\mathbb{X}, \mathscr{F})$ et $g : E \to \mathbb{X}$ mesurable. Alors, pour tous $i \in [\![n]\!]$ et $(x_j)_{j \in [\![n]\!] \setminus \{i\}} \in \prod_{i \in [\![n]\!] \setminus \{i\}} E_j$, l'application suivante est mesurable:

  $$\begin{aligned} E_i &\to \mathbb{X} \\ x &\mapsto g(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) \end{aligned}$$

  Attention, la réciproque est fausse, une fonction mesurable en chacune de ses variables n'est pas nécessairement mesurable. Contre-exemple :

  $$f : \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto \mathbb{1}_{\{y\}}(x)\mathbb{1}_E(x)$$

  avec $E \subset \mathbb{R}$ non mesurable (i.e. tel que $E \notin \mathscr{B}(\mathbb{R})$).

- Intégrer une fonction mesurable positive sur E sur seulement une de ses variables nous donne encore une fonction mesurable (à valeurs dans $\bar{\mathbb{R}}_+$).

  Mathématiquement, soit f une fonction mesurable réelle positive sur $(E, \mathscr{E})$. Pour tous $i \in [\![n]\!]$, la fonction suivante est mesurable pour la tribu $\bigotimes_{j \in [\![n]\!] \setminus \{i\}} \mathscr{E}_j$ sur $\prod_{j \in [\![n]\!] \setminus \{i\}} E_j$:

  $$\begin{aligned} \prod_{j \in [\![n]\!] \setminus \{i\}} E_j &\to \bar{\mathbb{R}}_+ \\ (x_j)_{j \in [\![n]\!] \setminus \{i\}} &\mapsto \int_{E_i} f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)\mu_i(dx) \end{aligned}$$

Théorème 3.1.10 (Fubini-Tonelli)**.** *Soit f une fonction mesurable réelle sur $(E, \mathscr{E}, \mu)$. Alors, pour toute permutation π sur $[\![n]\!]$ (i.e. $\pi : [\![n]\!] \to [\![n]\!]$ bijective), on a:*

$$\int_E |f| d\mu = \int_{E_{\pi(n)}} \cdots \left(\int_{E_{\pi(1)}} |f(x_1, \dots, x_n)| \mu_{\pi(1)}(dx_{\pi(1)}) \right) \cdots \mu_{\pi(n)}(dx_{\pi(n)})$$

Si de plus $\int_E |f| d\mu < +\infty$ (i.e. f est intégrable), alors pour toute permutation π sur $[\![n]\!]$:

$$\int_E f d\mu = \int_{E_{\pi(n)}} \cdots \left(\int_{E_{\pi(1)}} f(x_1, \ldots, x_n) \mu_{\pi(1)}(dx_{\pi(1)}) \right) \cdots \mu_{\pi(n)}(dx_{\pi(n)})$$

Remarque. *Le théorème de Fubini-Tonelli doit son nom aux mathématiciens italiens Guido Fubini et Leonida Tonelli. Fubini démontra d'abord le résultat pour les fonctions intégrables (seconde équation dans le théorème) en 1907. Tonelli le démontra ensuite pour les fonctions positives mesurables (première équation) en 1909. Le théorème est parfois simplement référé sous le nom de théorème de Fubini.*

Le théorème de Fubini-Tonelli nous dit que, pour intégrer une fonction mesurable positive sur E par la mesure produit μ, alors il suffit de l'intégrer par chacune des mesures μ_i ($i \in [\![n]\!]$) successivement et que l'ordre n'a pas d'importance. Les deux résultats de mesurabilité nous assurent que cette intégrale multiple est bien définie. Les fonctions intégrables sur E sont ainsi les fonctions mesurables dont l'intégrale multiple (au sens ci-dessus) de leur valeur absolue est finie. En décomposant une telle fonction en sa partie positive et sa partie négative, on peut déduire son intégrale par μ par le théorème pour les fonctions positives. Celle-ci est en effet l'intégrale multiple de f par les mesures μ_i. Encore une fois, l'ordre n'a pas d'importance.

Exemple. *On sait que l'aire d'un disque de rayon $r \in \mathbb{R}_+^*$ est πr^2. Le disque de centre 0 et de rayon r est $D_r := \{(x,y) \in \mathbb{R}^2 : x^2+y^2 \leq r^2\}$. Clairement, D_r n'est pas un pavé. On peut alors calculer son aire $\lambda_2(D_r)$ par le théorème de Fubini-Tonelli, on observant que $\mathbb{1}_{D_r}(x,y) = \mathbb{1}_{[-r,r]}(y)\mathbb{1}_{\left[-\sqrt{r^2-y^2}, \sqrt{r^2-y^2}\right]}(x)$ pour $(x,y) \in \mathbb{R}^2$:*

$$\lambda_2(D_r) = \int_{\mathbb{R}^2} \mathbb{1}_{D_r}(x,y)\lambda_2(d(x,y)) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{[-r,r]}(y)\mathbb{1}_{\left[-\sqrt{r^2-y^2}, \sqrt{r^2-y^2}\right]}(x)\lambda(dx)\lambda(dy) = \int_{[-r,r]} 2\sqrt{r^2-y^2}\lambda(dy)$$

En procédant au changement de variable $u = \frac{y}{r}$, on obtient:

$$\lambda_2(D_r) = 2r \int_{-r}^{r} \sqrt{1 - \left(\frac{y}{r}\right)^2} dy = 2r^2 \int_{-1}^{1} \sqrt{1-u^2} du$$

À partir de ce point, il s'agit d'un exercice d'analyse classique. On peut vérifier que $\int_{-1}^{1} \sqrt{1-u^2} du = \frac{\pi}{2}$.

Exercice 3.1.5. *Soit $D := \{(x,x) : x \in \mathbb{R}\} \subset \mathbb{R}^2$. Montrer que $D \in \mathscr{B}(\mathbb{R}^2)$. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ mesurable et positive. Montrer que la fonction suivante est mesurable:*

$$\begin{aligned} g : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto f(x,x) \end{aligned}$$

Si f est λ_2-intégrable, g est-elle nécessairement λ-intégrable ?

Le théorème de Fubini-Tonelli généralise plusieurs résultats d'analyse. En effet, les inversions de sommes infinies et intégrales en sont des cas particuliers.

Exemple (Sommes multiples). *Soient $(E_1, \mathscr{P}(E_1), \mu_1)$ et $(E_1, \mathscr{P}(E_2), \mu_2)$ deux ensembles dénombrables munis de leurs tribus discrètes et mesures de comptage respectives et $f : E_1 \times E_2 \to \mathbb{R}$ mesurable pour la tribu produit. Le théorème de Fubini nous dit alors que, si f est positive ou $\sum_{(x,y) \in E_1 \times E_2} |f(x,y)| < +\infty$, alors:*

$$\sum_{(x,y) \in E_1 \times E_2} f(x,y) = \sum_{x \in E_1} \sum_{y \in E_2} f(x,y) = \sum_{y \in E_2} \sum_{x \in E_1} f(x,y)$$

Le résultat s'étend évidemment au cas de sommes multiples quelconques.

Exemple (Inversion série-intégrale). *Soient $(f_k)_{k \in \mathbb{N}}$ une suite de fonctions mesurables réelles sur un espace mesuré $(\mathbb{X}, \mathscr{F}, \nu)$. Définissons la fonction suivante:*

$$\begin{aligned} f : \mathbb{N} \times \mathbb{X} &\to \mathbb{R} \\ (k,x) &\mapsto f_k(x) \end{aligned}$$

On considère l'espace produit $(\mathbb{N} \times \mathbb{X}, \mathscr{P}(\mathbb{N}) \otimes \mathscr{F}, \chi \otimes \nu)$ où χ est la mesure de comptage sur $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$. Montrons que f est mesurable de $(\mathbb{N} \times \mathbb{X}, \mathscr{P}(\mathbb{N}) \otimes \mathscr{F})$ dans $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Soit $B \in \mathscr{B}(\mathbb{R})$. Pour $k \in \mathbb{N}$, on a, par mesurabilité de f_k, $A_k := f_k^{-1}(B) \in \mathscr{F}$. Or:

$$f^{-1}(B) = \bigcup_{k \in \mathbb{N}} \{k\} \times A_k$$

Pour $k \in \mathbb{N}$, $\{k\} \times A_k$ est un pavé et donc dans $\mathscr{P}(\mathbb{N}) \otimes \mathscr{F}$. Par stabilité par union dénombrable de $\mathscr{P}(\mathbb{N}) \otimes \mathscr{F}$, on en déduit que $f^{-1}(B) \in \mathscr{P}(\mathbb{N}) \otimes \mathscr{F}$, c'est-à-dire f est mesurable. Par le théorème de Fubini-Tonelli :

$$\int |f| d(\chi \otimes \nu) = \sum_{k \in \mathbb{N}} \int |f_k| d\nu = \int \sum_{k \in \mathbb{N}} |f_k| d\nu$$

De plus, si la valeur ci-dessus est finie (i.e. f est $(\chi \otimes \nu)$-intégrable), alors :

$$\sum_{k \in \mathbb{N}} \int f_k d\nu = \int \sum_{k \in \mathbb{N}} f_k d\nu$$

On rappelle qu'une mesure à densité sur $\mathbb{R}$ définit la masse d'un borélien $B \in \mathscr{B}(\mathbb{R})$ comme l'intégrale d'une fonction mesurable $g : \mathbb{R} \to \mathbb{R}_+$ (unique, à égalité μ-p.p., pour tous les boréliens) sur B par la mesure de Lebesgue sur $\mathbb{R}$. La fonction g est alors appelée densité de μ (par rapport à Lebesgue). Dans cette section, nous avons étendu les notions de borélien, mesure de Lebesgue et fonction mesurable sur $\mathbb{R}^n$. On peut donc naturellement définir les mesures à densité sur $\mathbb{R}^n$.

Définition 3.1.11 (Mesure à densité sur $\mathbb{R}^n$)**.** *On dit qu'une mesure ν sur $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ admet une densité (par rapport à Lebesgue) si il existe une fonction mesurable réelle positive g sur $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ telle que, pour $B \in \mathscr{B}(\mathbb{R}^n)$, on a :*

$$\nu(B) = \int_B g d\lambda_n$$

où λ_n est la mesure de Lebesgue sur $\mathbb{R}^n$.

Tout comme pour les mesures à densité sur $\mathbb{R}$, dans la définition précédente, g est appelée densité de ν (par rapport à Lebesgue) et notée $g := \frac{d\nu}{d\lambda_n}$.

Exemple. *Soit $B \in \mathscr{B}(\mathbb{R}^n)$. La restriction $\lambda_n|_B$ admet trivialement la densité suivante $\frac{d\lambda_n|_B}{d\lambda_n} = \mathbb{1}_B$. En particulier, $\frac{d\lambda_n}{d\lambda_n} \equiv 1$.*

3.2 Vecteur aléatoire

Dans cette section, n est un entier supérieur ou égal à 2, $((E_i, \mathscr{E}_i))_{i \in [\![n]\!]}$ une famille d'espaces mesurables et $(E, \mathscr{E})$ leur espace produit. Considérons une famille $(X_i)_{i \in [\![n]\!]}$ de variables aléatoires définies sur un même espace $(\Omega, \mathscr{A}, \mathbf{P})$ et telle que, pour $i \in [\![n]\!]$, X_i est à valeurs dans $(E_i, \mathscr{E}_i)$. Par le corollaire 3.1.3, la fonction $X := (X_1, \dots, X_n)$ de Ω dans E est alors une variable aléatoire. Réciproquement, étant donné une variable aléatoire $X' := (X'_1, \dots, X'_n)$ de $(\Omega, \mathscr{A}, \mathbf{P})$ dans $(E, \mathscr{E})$, alors, par le même corollaire, X'_i est une variable aléatoire de $(\Omega, \mathscr{A}, \mathbf{P})$ dans $(E_i, \mathscr{E}_i)$ pour $i \in [\![n]\!]$. Ainsi, on peut voir toute collection finie de variables aléatoires (sur un même espace) comme une unique variable aléatoire dans un espace produit. Ceci répond à la première question de l'exemple introductif du chapitre: étant donné une fonction mesurable réelle φ sur $(E, \mathscr{E})$, $\varphi(X_1, \dots, X_n)$ est une variable aléatoire réelle. En effet, $\varphi(X_1, \dots, X_n) = \varphi \circ X : \Omega \to \mathbb{R}$ est mesurable par composition de fonctions mesurables. Comme nous l'avons vu dans le même exemple, la loi de X n'est pas (sans hypothèse supplémentaire) déterminée par les lois respectives des $(X_i)_{i \in [\![n]\!]}$. Ces dernières sont appelées lois marginales de X.

3.2.1 Lois jointe et marginales de variables aléatoires

Définition 3.2.1 (Lois jointe et marginales)**.** *Soit $X := (X_1, \dots, X_n)$ une variable aléatoire dans $(E, \mathscr{E})$. On appelle lois marginales de X (ou des $(X_i)_{i \in [\![n]\!]}$) les lois respectives des variables aléatoires $(X_i)_{i \in [\![n]\!]}$. La loi de X sur $(E, \mathscr{E})$ est également appelée loi jointe (de X ou des $(X_i)_{i \in [\![n]\!]}$).*

Exemple. *Soit $(\Omega, \mathscr{A}, \mathbf{P})$ l'espace probabilisé donné par $\Omega := \mathbb{R}$, $\mathscr{A} := \mathscr{B}(\mathbb{R})$ et $\mathbf{P} := \mathscr{U}([0,1])$. On considère les variables aléatoires réelles $X := \mathbb{1}_{[0,2/3]}$ et $Y := \mathbb{1}_{[1/3,1]}$ sur $(\Omega, \mathscr{A}, \mathbf{P})$ de lois respectives $\mathbf{P}_X$ et $\mathbf{P}_Y$ et on pose $Z := (X, Y)$. On vérifie facilement que $\mathbf{P}_X = \mathbf{P}_Y = \mathscr{B}(2/3)$. Ce sont les lois marginales (en l'occurence égales) de Z. Puisque $X(\Omega) = Y(\Omega) = \{0, 1\}$, $Z(\Omega) = \{0, 1\}^2 \subset \mathbb{R}^2$. La loi jointe $\mathbf{P}_{(X,Y)} = \mathbf{P}_Z$ sur $(\{0,1\}^2, \mathscr{P}(\{0,1\}^2))$ (ou de façon équivalente $(\mathbb{R}^2, \mathscr{B}(\mathbb{R}^2))$) de X et Y est donnée par:*

$$\mathbf{P}_Z(\{(0,0)\}) = \mathbf{P}((X,Y) = (0,0)) = \mathbf{P}((X = 0) \wedge (Y = 0)) = \mathbf{P}(]2/3, 1] \cap [0, 1/3[) = \mathbf{P}(\emptyset) = 0$$

$$\mathbf{P}_Z(\{(0,1)\}) = \mathbf{P}((X,Y) = (0,1)) = \mathbf{P}((X = 0) \wedge (Y = 1)) = \mathbf{P}(]2/3, 1] \cap [1/3, 1]) = \mathbf{P}(]2/3, 1]) = \frac{1}{3}$$

$$\mathbf{P}_Z(\{(1,0)\}) = \mathbf{P}((X,Y) = (1,0)) = \mathbf{P}((X = 1) \wedge (Y = 0)) = \mathbf{P}([0, 2/3] \cap [0, 1/3[) = \mathbf{P}([0, 1/3[) = 1/3$$

$$\mathbf{P}_Z(\{(1,1)\}) = \mathbf{P}((X,Y) = (1,1)) = \mathbf{P}((X = 1) \wedge (Y = 1)) = \mathbf{P}([0, 2/3] \cap [1/3, 2/3]) = \mathbf{P}([1/3, 2/3]) = \frac{1}{3}$$

On a ainsi $\mathbf{P}_Z = \mathscr{U}(\{(0,1), (1,0), (1,1)\})$.

Étant donné une variable aléatoire $X := (X_1, \dots, X_n)$ dans $(E, \mathscr{E})$. La loi (marginale) de X_i (pour $i \in [\![n]\!]$), est donnée, pour $B \in \mathscr{E}_i$, par:

$$\mathbf{P}(X_i \in B_i) = \mathbf{P}(X \in E_1 \times \cdots \times E_{i-1} \times B_i \times E_{i+1} \times \cdots \times E_n)$$

On peut donc toujours obtenir les lois marginales à partir de la loi jointe.

Exemple (Lois discrètes). *Soit $X := (X_1, \dots, X_n)$ une variable aléatoire dans $(E, \mathscr{E})$ de loi jointe $\mathbf{P}_X$. Supposons que, pour $i \in [\![n]\!]$, E_i soit dénombrable et $\mathscr{E} := \mathscr{P}(E)$. Pour $i \in [\![n]\!]$, la marginale $\mathbf{P}_{X_i}$ de X_i est alors donnée, pour $x \in \mathscr{E}_i$, par:*

$$\begin{aligned}\mathbf{P}_{X_i}(\{x\}) = \sum_{\substack{(x_1,\dots,x_n)\in E \\ x_i = x}} \mathbf{P}_X(\{(x_1,\dots,x_n)\}) &= \sum_{(x_1,\dots,x_n)\in E} \mathbf{P}_X(\{(x_1,\dots,x_n)\})\mathbb{1}_x(x_i)\mu(\{(x_1,\dots,x_n)\}) \\ &= \sum_{x_1\in E_1} \cdots \left(\sum_{x_n \in E_n} \mathbf{P}_X(\{(x_1,\dots,x_n)\})\mathbb{1}_x(x_i)\mu_n(\{x_n\}) \right) \cdots \mu_1(\{x_1\}) \\ &= \sum_{x_1\in E_1} \cdots \sum_{x_{i-1}\in E_{i-1}} \sum_{x_{i+1}\in E_{i+1}} \cdots \sum_{x_n \in E_n} \mathbf{P}(\{(x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_n)\})\end{aligned}$$

où la seconde égalité est en voyant la loi jointe P_X comme une fonction sur l'espace produit, que l'on intègre par la mesure de comptage μ sur $(E, \mathscr{E})$. La troisième inégalité est par l'application du theorème de Fubini-Tonelli (la mesure de comptage étant une mesure produit).

Exemple (Lois à densité). *Soit $X := (X_1, \dots, X_n)$ une variable aléatoire dans $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ de loi $\mathbf{P}_X$ admettant la densité g contre Lebesgue. Alors, pour $i \in [\![n]\!]$, la loi marginale $\mathbf{P}_{X_i}$ de X_i est donnée, pour $B \in \mathscr{B}(\mathbb{R})$, par:*

$$\mathbf{P}_{X_i}(B) = \mathbf{P}_X(\mathbb{R}^{i-1} \times B \times \mathbb{R}^{n-i}) = \int_{\mathbb{R}^n} \mathbb{1}_B(x_i) g(x_1,\dots,x_n)\lambda_n(dx)$$

où on écrit $\lambda_n(dx)$ pour $\lambda_n(d(x_1,\dots,x_n))$. Par le théorème de Fubini-Tonelli, on obtient:

$$\mathbf{P}_{X_i}(B) = \int_B g_i(x)\lambda(dx)$$

où, pour $x \in \mathbb{R}$, g_i est donnée par:

$$\begin{aligned} g_i : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \int_{\mathbb{R}^{n-1}} g(y_1,\dots,y_{i-1},x,y_i,\dots,y_{n-1})\lambda_{n-1}(dy)\end{aligned}$$

On rappelle que g_i est mesurable. De plus, elle est positive par intégration de fonction positive. On en déduit que $\mathbf{P}_{X_i}$ admet la densité g_i par rapport à Lebesgue.

Exercice 3.2.1. *Soit μ la loi admettant la densité g définie comme suit :*

$$\begin{aligned} g : \mathbb{R}^3 &\to \mathbb{R} \\ (x,y,z) &\mapsto \begin{cases} Cxyz, & \text{si } 0 \le x \le y \le z \le 1 \\ 0, & \text{sinon} \end{cases}.\end{aligned}$$

Soit $X := (X_1, X_2, X_3)$ une variable aléatoire de loi μ. Déterminer les lois marginales de X.

Remarquons que le théorème de transfert s'applique toujours pour des variables aléatoires $X := (X_1, \dots, X_n)$ dans $(E, \mathscr{E})$ et fonctions mesurables réelles φ sur $(E, \mathscr{E})$. En effet, par compositions de fonctions mesurable, $\varphi(X)$ est une variable aléatoire réelle de loi $\varphi_* \mathbf{P}_X$ (où $\mathbf{P}_X$ est la loi jointe de X) et on a alors:

$$\mathbf{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x_1,\dots,x_n)\mathbf{P}_X(d(x_1,\dots,x_n))$$

Exemple. *Soient $(X, Y) \sim \mathscr{U}([0,1]^2)$, $Z := XY$ et f une fonction mesurable réelle sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. On a:*

$$\mathbf{E}[f(Z)] = \mathbf{E}[f(XY)] = \int_{\mathbb{R}^2} f(xy)\mathbb{1}_{[0,1]^2}(x,y)\lambda_2(d(x,y)) = \int_{]0,1[}\int_{]0,1[} f(xy)\lambda(dx)\lambda(dy)$$

où la deuxième égalité est par le théorème de transfert et la troisième par le théorème de Fubini-Tonelli et car la loi de (X, Y) est diffuse. En faisant le changement de variable $z = xy$, on obtient:

$$\mathbf{E}[f(Z)] = \int_{]0,1[}\int_{]0,y[} f(z)\frac{1}{y}\lambda(dz)\lambda(dy) = \int_{]0,1[} f(z)\int_{]z,1[}\frac{1}{y}\lambda(dy)\lambda(dz) = \int_{\mathbb{R}} f(z)(-\log(z)\mathbb{1}_{]0,1[}(z))\lambda(dz)$$

où la seconde égalité est par Fubini-Tonelli encore. On en déduit que la loi de Z a pour densité $z \mapsto -\log(z)\mathbb{1}_{]0,1[}(z)$.

3.2.2 Indépendance de variables aléatoires

On peut maintenant formaliser la notion d'indépendance de variables aléatoires. En mots, des variables aléatoires sont indépendantes si tous les événements qui ne dépendent que de leurs valeurs respectives le sont. Tout comme pour l'indépendance d'événements, on peut considérer l'indépendance d'une famille quelconque de variables aléatoires.

Définition 3.2.2 (Indépendance de variables aléatoires)**.** *Soit $(X_i)_{i\in I}$ une famille de variables aléatoires sur un même espace probabilisé telle que, pour $i \in I$, X_i est à valeurs dans un espace mesurable $(\mathbb{X}_i, \mathscr{F}_i)$. On dit que la famille est indépendante (ou que les variables aléatoires le sont) si, pour tous $(B_i)_{i\in I} \in \prod_{i\in I} \mathscr{F}_i$, la famille d'événements $(\{X_i \in B_i\})_{i\in I}$ l'est.*

On peut observer qu'une famille de variables aléatoires est indépendantes si et seulement si toute sous-famille finie l'est. Ainsi, on se contentera d'énoncer les résultats d'indépendance pour n variables aléatoires. On vérifie facilement que la notion d'indépendance énoncée ci-dessus est équivalente à celle intuitée dans l'exemple introductif du chapitre et la section 3.1.

Proposition 3.2.3. *Soit $X := (X_1, \dots, X_n)$ une variable aléatoire dans $(E, \mathscr{E})$. Les coordonnées de X sont indépendantes si et seulement si l'une des conditions équivalentes suivantes est satisfaite:*

- *Pour toutes fonctions mesurables réelles (toutes) positives ou bornées $(\varphi_i)_{i\in[\![n]\!]}$, avec $\varphi_i : E_i \to \mathbb{R}$ pour $i \in [\![n]\!]$:*

$$\mathbf{E}\left[\prod_{i=1}^{n} \varphi_i(X_i)\right] = \prod_{i=1}^{n} \mathbf{E}[\varphi_i(X_i)]$$

- *La loi jointe $\mathbf{P}_X$ de X est le produit de ses lois marginales $(\mathbf{P}_{X_i})_{i\in[\![n]\!]}$, c'est-à-dire:*

$$\mathbf{P}_X = \bigotimes_{i\in[\![n]\!]} \mathbf{P}_{X_i}$$

Preuve

Supposons que $\mathbf{P}_X$ est produit des lois marginales de X et prenons $(\varphi_i)_{i\in[\![n]\!]}$ où, pour $i \in [\![n]\!]$, φ_i est une fonction mesurable réelle positives (resp. bornées) sur $(E_i, \mathscr{E}_i)$. On a alors par le théorème de transfert:

$$\mathbf{E}\left[\prod_{i=1}^{n} \varphi_i(X_i)\right] = \int_E \left(\prod_{i=1}^{n} \varphi_i(x_i)\right) \mathbf{P}_X(d(x_1, \dots, x_n)) = \int_E \left(\prod_{i=1}^{n} \varphi_i(x_i)\right) \left(\bigotimes_{i\in[\![n]\!]} \mathbf{P}_{X_i}\right)(d(x_1, \dots, x_n))$$

où la seconde égalité est par hypothèse. Si les $(\varphi_i)_{i\in[\![n]\!]}$ sont prises bornées, alors, pour $i \in [\![n]\!]$, φ_i est trivialement $\mathbf{P}_{X_i}$-intégrable ($\mathbf{E}[|\varphi_i(X_i)|] \leq \mathbf{E}[M] = M \in \mathbb{R}_+^*$ où M est une borne de $|\varphi_i|$). Le théorème de Fubini-Tonelli nous donne ensuite:

$$\mathbf{E}\left[\prod_{i=1}^{n} \varphi_i(X_i)\right] = \int_{E_n} \cdots \int_{E_1} \left(\prod_{i=1}^{n} \varphi_i(x_i)\right) \mathbf{P}_{X_1}(dx_1) \cdots \mathbf{P}_{X_n}(dx_n) = \prod_{i=1}^{n} \int_{E_i} \varphi_i(x_i) \mathbf{P}_{X_i}(dx_i) = \prod_{i=1}^{n} \mathbf{E}[\varphi(X_i)]$$

où la dernière égalité est par le théorème de transfert encore. Supposons maintenant seulement que:

$$\mathbf{E}\left[\prod_{i=1}^{n} \varphi_i(X_i)\right] = \prod_{i=1}^{n} \mathbf{E}[\varphi_i(X_i)]$$

pour toute famille $(\varphi_i)_{i\in[\![n]\!]}$ comme au dessus. Pour $i \in [\![n]\!]$, prenons $B_i \in \mathscr{E}_i$ et $\varphi_i := \mathbb{1}_{B_i}$. φ_i est clairement mesurable et positive pour $i \in [\![n]\!]$. On a alors immédiatement:

$$\mathbf{P}\left(\bigcap_{i\in[\![n]\!]} \{X_i \in B_i\}\right) = \mathbf{E}\left[\prod_{i=1}^{n} \mathbb{1}_{B_i}(X_i)\right] = \mathbf{E}\left[\prod_{i=1}^{n} \varphi_i(X_i)\right] = \prod_{i=1}^{n} \mathbf{E}[\varphi_i(X_i)] = \prod_{i=1}^{n} \mathbf{E}[\mathbb{1}_{B_i}(X_i)] = \prod_{i=1}^{n} \mathbf{P}(\{X_i \in B_i\})$$

C'est-à-dire, les coordonnées de X sont indépendantes. Supposons maintenant uniquement que les coordonnées de X sont indépendantes. Par définition, pour tout pavé $B_1 \times \cdots \times B_n \in \mathscr{E}$:

$$\mathbf{P}\left(\bigcap_{i\in[\![n]\!]} \{X_i \in B_i\}\right) = \prod_{i=1}^{n} \mathbf{P}(X_i \in B_i) = \prod_{i=1}^{n} \mathbf{P}_{X_i}(B_i)$$

Or:

$$\mathbf{P}\left(\bigcap_{i\in[\![n]\!]} \{X_i \in B_i\}\right) = \mathbf{P}(X \in B_1 \times \cdots \times B_n) = \mathbf{P}_X(B_1 \times \cdots \times B_n)$$

Par le théorème 3.1.8 (ou le lemme des classes monotones, en remarquant que les pavés forment un π-système qui engendre $\mathscr{E}$), $\bigotimes_{i\in[\![n]\!]} \mathbf{P}_{X_i}$ est l'unique mesure sur $(E,\mathscr{E})$ telle que:

$$\left(\bigotimes_{i\in[\![n]\!]} \mathbf{P}_{X_i}\right)(B_1\times\cdots\times B_n) = \prod_{i=1}^{n} \mathbf{P}_{X_i}(B_i)$$

pour tout pavé $B_1\times\cdots\times B_n\in\mathscr{E}$. On en déduit que $\mathbf{P}_X = \bigotimes_{i\in[\![n]\!]} \mathbf{P}_{X_i}$, ce qui conclut la preuve.

Remarque. *Dans la proposition ci-dessus, les espérances de la première condition sont bien définies. En effet, pour $i\in[\![n]\!]$, $\varphi_i(X_i)$ est une variable aléatoire (positive) par composition de fonctions mesurables. De même, $\prod_{i=1}^n \varphi_i(X_i)$ est une variable aléatoire (également positive) par produit fini de fonctions mesurables. La mesurabilité et la positivité garantissent l'existence des espérances. Le résultat tient en fait toujours à partir du moment où les $(\varphi_i)_{i\in[\![n]\!]}$ sont telles que les espérances sont bien définies.*

Appliquons cette notion d'indépendance pour établir une propriété sur des variables aléatoires suivant la loi de Poisson.

Exercice 3.2.2. *Si X et Y sont deux variables aléatoires indépendantes, suivant des lois de Poisson de paramètre λ et μ respectivement, alors $X+Y$ suit une loi de Poisson de paramètre $\lambda+\mu$.*

3.2.3 Notion de vecteur aléatoire et quelques statistiques associées

Dans la suite du cours, on va s'intéresser au cas où $(E,\mathscr{E}) = (\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$. On rappelle que l'espérance et la variance, deux notions probabilites essentielles, ne sont définies que pour des variables aléatoires réelles. On va voir ici que la structure algébrique et topologique riche des espaces euclidiens nous permet en fait d'étendre ces notions à la dimension n et, en particulier, de quantifier certaines dépendances de variables aléatoires.

Définition 3.2.4 (Vecteur aléatoire)**.** *On appelle vecteur aléatoire (réel) une variable aléatoire dans $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$.*

Remarque. *Le nom de vecteur aléatoire est justifié par l'observation suivante. On rappelle d'abord que, par les deux premiers points de la proposition 2.1.4, l'ensemble des fonctions mesurables réelles $\mathscr{M}(\Omega,\mathbb{R})$ définies sur un même espace $(\Omega,\mathscr{A})$ forme un espace vectoriel (et même une algèbre par le troisième point), où l'élément nul est la fonction constante et égale à 0. De plus, par le corollaire 3.1.3, une fonction $f := (f_1,\dots,f_n)$ de Ω dans $\mathbb{R}^n$ est mesurable si et seulement si $f_i\in\mathscr{M}(\Omega,\mathbb{R})$ pour tout $i\in[\![n]\!]$. Soient $f := (f_1,\dots,f_n)$, $g := (g_1,\dots,g_n)$ deux fonction mesurables de Ω dans $\mathbb{R}^n$ et $\alpha\in\mathbb{R}$. Par la structure vectorielle de $\mathscr{M}(\Omega,\mathbb{R})$, $f_i+\alpha g_i\in\mathscr{M}(\Omega,\mathbb{R})$ pour tout $i\in[\![n]\!]$. Par le corollaire 3.1.3 encore, on en déduit que $f+\alpha g := (f_1+\alpha g_1,\dots,f_n+\alpha g_n)$ est mesurable. C'est-à-dire, l'ensemble $\mathscr{M}(\Omega,\mathbb{R}^n)$ des fonctions mesurables de $(\Omega,\mathscr{A})$ dans $(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n))$ forme un espace vectoriel pour l'addition comme définie ci-dessus. En particulier, l'ensemble des vecteurs aléatoires sur un espace probabilisé est bien un espace vectoriel. Par le même raisonnement, on en déduit que, muni de la multiplication $fg := (f_1g_1,\dots,f_ng_n)$, $\mathscr{M}(\Omega,\mathbb{R}^n)$ est une algèbre.*

Définition 3.2.5 (Espérance d'un vecteur aléatoire)**.** *Soit $X := (X_1,\dots,X_n)$ un vecteur aléatoire. Si, pour tout $i\in[\![n]\!]$, X_i est positive ou intégrable, alors on définit l'espérance de X comme le n-uplet suivant:*

$$\mathbf{E}[X] = \mathbf{E}[(X_1,\dots,X_n)] := (\mathbf{E}[X_1],\dots,\mathbf{E}[X_n])\in\bar{\mathbb{R}}^n$$

Notation. *Étant donnés deux espaces vectoriels E et F, on note $\mathscr{L}(E,F)$ l'ensemble des applications linéaires de E dans F. Lorsque ces derniers sont de dimensions finies, $\mathscr{L}(E,F)$ est identifiable aux matrices de taille $\dim(F)\times dim(E)$. En particulier, on dénotera par la même lettre une application linéaire et sa matrice dans les bases canoniques.*

De la définition, l'espérance d'un vecteur aléatoire X dont les coordonnées sont intégrables est un vecteur de $\mathbb{R}^n$. On parle de vecteur aléatoire intégrable. Par linéarité de l'espérance (en dimension 1), on en déduit que, pour un tel vecteur et pour tous $m\in\mathbb{N}^*$ et $A\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$, $\mathbf{E}[A(X)] = A(\mathbf{E}[X])$. De même, pour deux vecteurs aléatoires X et Y dont les coordonnées respectives sont positives ou intégrables, on a $\mathbf{E}[X+Y] = \mathbf{E}[X]+\mathbf{E}[Y]$. En mots, l'espérance en dimension n est également linéaire.

Notation. *Dans la suite du cours, on verra souvent les vecteurs de $\mathbb{R}^n$ et les vecteurs aléatoires comme des vecteurs colonnes. Avec cette identification, et celle d'une application linéaire à sa matrice, on écrit notamment $\mathbf{E}[AX] = A\mathbf{E}[X]$. De plus, pour un vecteur $v\in\mathbb{R}^n$ et une matrice A, on écrit v^* et A^* les transposés de v et A.*

Avant de nous pencher sur l'extension de la notion de variance à des vecteurs aléatoires, nous énonçons une inégalité utile.

Exercice 3.2.3. *Soient $p,q\in\mathbb{R}_+^*$ tels que $\frac{1}{p}+\frac{1}{q}=1$. Montrer que, pour tous $a,b\in\mathbb{R}_+$:*

$$ab\le\frac{a^p}{p}+\frac{b^q}{q}$$

L'inégalité est triviale si $a=0$ ou $b=0$. On pourra traiter le cas restant en utilisant la convexité de la fonction exponentielle. Ce résultat est appelé inégalité de Young.

Proposition 3.2.6 (Inégalité de Hölder). *Soient X et Y deux variables aléatoires réelles définies sur un même espace et $p, q \in \mathbb{R}_+^*$ deux réels tels que $\frac{1}{p} + \frac{1}{q} = 1$. Alors:*

$$\mathbf{E}[|XY|] \leq \mathbf{E}[|X|^p]^{\frac{1}{p}} \mathbf{E}[|Y|^q]^{\frac{1}{q}}$$

Preuve

Posons $A := \mathbf{E}[|X|^p]^{\frac{1}{p}}$ et $B := \mathbf{E}[|Y|^q]^{\frac{1}{q}}$. Si $A = 0$, alors $AB = 0$ (on rappelle que, par convention, $0 \times (+\infty) = 0$) et $|X|$ est presquement sûrement nulle. On en déduit que $|XY|$ est également presque sûrement nulle et donc d'espérance nulle. De même, si $A = +\infty$ et $B > 0$, alors $AB = +\infty$ et l'inégalité est trivialement vérifiée. Par symmétrie, on démontre pareillement les cas avec $B = 0$ ou $B = +\infty$. Supposons donc $A, B \in \mathbb{R}_+^*$ et posons $\bar{X} := A^{-1}|X|$ et $\bar{Y} = B^{-1}|Y|$. $\bar{X}$ et $\bar{Y}$ sont des variables aléatoires réelles positives telles que:

$$\mathbf{E}\left[\left|\bar{X}\right|^p\right] = \mathbf{E}\left[\left|\bar{Y}\right|^q\right] = 1$$

Par l'inégalité de Young, on a:

$$\bar{X}\bar{Y} \leq \frac{1}{p}\bar{X}^p + \frac{1}{q}\bar{Y}^q$$

En passant à l'espérance, on obtient:

$$\mathbf{E}\left[\bar{Y}\bar{X}\right] \leq \frac{1}{p}\mathbf{E}\left[\bar{X}^p\right] + \frac{1}{q}E\left[\bar{Y}^q\right] = \frac{1}{p} + \frac{1}{q} = 1$$

On en déduit le résultat souhaité.

Remarque. *L'inégalité de Hölder dans le cas $p = q = \frac{1}{2}$ correspond à l'inégalité de Cauchy-Schwartz.*

Définition 3.2.7 (Covariance). *Soient X et Y deux variables aléatoires définies sur le même espace et de variances finies. On définit alors leur covariance, notée $Cov(X, Y)$ comme la valeur suivante:*

$$Cov(X, Y) := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \in \mathbb{R}$$

Remarque. *Par l'inégalité de Cauchy-Schwarz, l'espérance est bien définie et finie dans la définition ci-dessus.*

Exercice 3.2.4. *Soient X et Y deux variables aléatoires définies sur le même espace et admettant des moments d'ordre 2. Montrer que:*

$$Cov(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$$

Remarque. *La covariance entre deux variables aléatoires X et Y de carré intégrable (i.e. admettant des moments d'ordre 2) désigne leur degré de dépendance affine. Afin de détailler ce que l'on entend par là, nous nous permettons une digression hors programme sur la notion de régression linéaire. Il est possible de définir une "pseudo" métrique d_Q naturelle sur l'ensemble des variables aléatoires admettant des moments d'ordre 2, appelée risque ou distance quadratique. Celle-ci est définie, pour deux telles variables aléatoires X et Y, de la sorte:*

$$d_Q(X, Y) = \sqrt{\mathbf{E}[(X - Y)^2]}$$

On observe que deux variables aléatoires à distance quadratique nulle sont égales presque sûrement (d'où le qualificatif de "pseudo": on a pas égalité au sens propre). Elle peut être pensée comme une généralisation de la distance euclidienne à des espaces de dimension infinie. En effet, si Ω est un ensemble fini et $\mathbf{P} = \mathscr{U}(\Omega)$, alors X et Y sont identifiables à des vecteurs de $\mathbb{R}^{|\Omega|}$ et $d_Q(X, Y)$ correspond à leur distance euclidienne (à un facteur $|\Omega|^{-1/2}$ près). Imaginons qu'on souhaite approximer Y par une fonction affine de X. On va alors considérer la fonction affine A qui minimise la distance quadratique entre Y et $A(X)$. Autrement dit, on va considérer le problème suivante:

$$\arg\min_{a,b \in \mathbb{R}} d_Q(Y, aX + b)$$

On a ainsi un problème d'optimisation à deux variables réelles. On peut montrer (cf annexe E) que la "meilleure"(au sens de la distance quadratique) approximation affine de Y par X est :

$$\hat{Y} := \frac{Cov(X, Y)}{Var(X)}X + \mathbf{E}[Y] - \frac{Cov(X, Y)}{Var(X)}\mathbf{E}[X] = \mathbf{E}[Y] + \frac{Cov(X, Y)}{Var(X)}(X - \mathbf{E}[X])$$

Cette méthode est appelée la régression linéaire. Par métonymie, il est également courant d'appeler $\hat{Y}$ la régression linéaire de Y par X. On observe que $\hat{Y}$ peut se décomposer en l'espérance de Y plus une variable centrée (i.e. d'espérance nulle) dont on récupère et pondère la variabilité. En particulier, $\mathbf{E}[Y] = \mathbf{E}[\hat{Y}]$. On montre aussi facilement que $Cov(X,Y) = Cov(X,\hat{Y})$. Notamment, si $Cov(X,Y) = 0$, alors la meilleure approximation affine de Y par X est la variable constance et égale à $\mathbf{E}[Y]$ (on ignore complètement X). En d'autres termes, Y et X ne sont absolument pas affinement corrélés. Un cas particulier de ce phénomène est si X et Y sont indépendants. En effet, on a alors:

$$Cov(X,Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] = \mathbf{E}[X]\mathbf{E}[Y] - \mathbf{E}[X]\mathbf{E}[Y] = 0$$

où la seconde égalité est par la proposition 3.2.3.

Il convient de remarquer que si X et Y sont indépendants, alors leur covariance est nulle. Comme nous le montre l'exemple suivant, la réciproque est fausse.

Exemple. *Soient X et Y deux variables aléatoires réelles données par :*

$$\mathbf{P}(X = -1) = \mathbf{P}(X = 1) = \frac{1}{2}$$
$$\mathbf{P}(Y = 0 | X = -1) = 1$$
$$\mathbf{P}(Y = -1 | X = 1) = \mathbf{P}(Y = 1 | X = 1) = \frac{1}{2}$$

Clairement, X et Y ne sont pas indépendantes puisque connaître Y nous donne X presque sûrement et ni X ni Y n'est presque sûrement constante. Pourtant leur covariance est nulle. La loi jointe du couple (X,Y) est donnée par :

$$\mathbf{P}((X,Y) = (0,-1)) = \frac{1}{2}$$
$$\mathbf{P}((X,Y) = (-1,1)) = \mathbf{P}((X,Y) = (1,1)) = \frac{1}{4}$$

À partir de celle-ci, on peut calculer l'espérance du produit :

$$\mathbf{E}[XY] = \mathbf{P}((X,Y) = (1,1)) - \mathbf{P}((X,Y) = (-1,1)) = \frac{1}{4} - \frac{1}{4} = 0$$

De même, on vérifie facilement que X et Y sont des variables centrées (i.e. d'espérance nulle). Ainsi :

$$Cov(X,Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] = 0$$

Plus généralement, si une variable aléatoire Y est centrée quelque soit la valeur d'une autre variable aléatoire X, alors elles auront une covariance nulle. Dans le cas d'un couple à densité g, cela revient à dire que $\int_{\mathbb{R}} g(x,y)\lambda(dy) = 0$ pour λ-presque tout $x \in \mathbb{R}$.

Exercice 3.2.5. *Soient X, Y, Z trois variables aléatoires de carrés intégrables définies sur un même espace et $a \in \mathbb{R}$. Montrer les égalités suivantes:*

- *$Cov(aX+Z,Y) = aCov(X,Y) + Cov(Z,Y)$*
- *$Cov(X,Y) = Cov(Y,X)$*
- *$Cov(a,X) = 0$*
- *$Cov(X,X) = Var(X)$*
- *$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$*

Définition 3.2.8 (Matrice de dispersion)**.** *Soit $X := (X_1, \dots, X_n)$ un vecteur aléatoire carré intégrable (i.e. dont les coordonnées sont carrés intégrables). On appelle matrice de dispersion (ou de covariance) de X la matrice carrée $D_X := (Cov(X_i, X_j))_{(i,j)\in[\![n]\!]^2}$.*

Remarque. *La matrice de dispersion est la généralisation la plus naturelle de la variance pour un vecteur aléatoire. Une autre écriture existe et met cette remarque en évidence. En définissant une matrice aléatoire $X := (X_{(i,j)})_{(i,j)\in[\![k]\!]\times[\![l]\!]}$ comme un vecteur aléatoire réel de taille kl dont on arrange les coordonnées dans un tableau de dimensions $k \times l$, l'espérance $\mathbf{E}[X]$ d'une telle matrice est alors simplement la matrice des espérances des coordonnées, i.e. $\left(\mathbf{E}[X_{(i,j)}]\right)_{(i,j)\in[\![k]\!]\times[\![l]\!]}$. Par linéarité de l'espérance pour un vecteur aléatoire intégrable, si la matrice aléatoire X est intégrable (i.e. si toutes ses coordonnées le sont), alors pour tous $A \in \mathscr{L}(\mathbb{R}^l, \mathbb{R}^k)$ et $B \in \mathscr{L}(\mathbb{R}^k, \mathbb{R}^l)$, on a*

$\mathbf{E}[AX] = A\mathbf{E}[X]$ *et* $\mathbf{E}[XB] = \mathbf{E}[X]B$. *Revenons au cas où* $X := (X_1, \dots, X_n)$ *est un vecteur aléatoire. On peut alors définir la matrice aléatoire suivante :*

$$(X - \mathbf{E}[X])(X - \mathbf{E}[X])^* = \begin{pmatrix} X_1 - \mathbf{E}[X_1] \\ \vdots \\ X_n - \mathbf{E}[X_n] \end{pmatrix} \begin{pmatrix} X_1 - \mathbf{E}[X_1] & \cdots & X_n - \mathbf{E}[X_n] \end{pmatrix} = ((X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j]))_{(i,j)\in[\![n]\!]^2}$$

En particulier :

$$\mathbf{E}[(X - \mathbf{E}[X])(X - \mathbf{E}[X])^*] = (\mathbf{E}[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])])_{(i,j)\in[\![n]\!]^2} = (Cov(X_i, X_j))_{(i,j)\in[\![n]\!]^2} = D_X$$

En plus d'être concise, l'écriture matricielle permet de simplifier de nombreuses preuves.

Exemple. *En reprenant encore le premier exemple du chapitre, on calcule facilement :*

$$\mathbf{E}[XY] = \frac{1}{3}$$

et on en déduit

$$Cov(X, Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] = \frac{1}{3} - \frac{2}{3}\frac{2}{3} = -\frac{1}{9}$$

De même, X *et* Y *ayant toutes deux loi* $\mathscr{B}(2/3)$*, on a :*

$$Var(X) = Var(Y) = \frac{2}{3}\left(1 - \frac{2}{3}\right) = \frac{2}{9}$$

On en déduit la matrice de dispersion D_Z *du couple* $Z := (X, Y)$ *:*

$$D_Z := \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(Y,X) & Var(Y) \end{pmatrix} = \begin{pmatrix} \frac{2}{9} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{2}{9} \end{pmatrix} = \frac{1}{9}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

Car la covariance de deux variables indépendantes est nulle si elles sont indépendantes, il est immédiat que la matrice de dispersion d'un vecteur aléatoire est diagonale si ses coordonnées sont indépendantes.

3.2.4 Fonction de répartition et fonction caractéristique d'un vecteur aléatoire

- Fonction de répartition

Définition 3.2.9 (Fonction de répartition)**.** *Soit* $X := (X_1, \dots, X_n)$ *un vecteur aléatoire. On appelle fonction de répartition de* X*, notée* F_X*, la fonction suivante:*

$$F_X : \mathbb{R}^n \to \mathbb{R}$$
$$(x_1, \dots, x_n) \mapsto \mathbf{P}((X_1 \leq x_1) \wedge \cdots \wedge (X_n \leq x_n))$$

La fonction de répartition a la même utilité qu'en dimension 1, en ce sens qu'elle caractérise la loi. Ainsi, lorsqu'on souhaite donner la loi d'un vecteur aléatoire, il est toujours suffisant de donner sa fonction de répartition.

Proposition 3.2.10. *Deux vecteurs aléatoires ont même loi si et seulement si ils ont la même fonction de répartition.*

Preuve

Admis. C'est une généralisation du corollaire 2.1.10.

Notation. *Soit* U *un ouvert de* $\mathbb{R}^n$*. Étant donnés une fonction* $f : U \to \mathbb{R}$*,* $i \in [\![n]\!]$ *et* $x \in \mathbb{R}^n$*, on note* $\partial_i f(x)$ *la dérivée partielle en sa* $i^{ème}$ *variable de* f *en* x*, si elle existe. Si cette dernière est définie pour tout* $x \in U$*, on note* $\partial_i f : U \to \mathbb{R}, x \mapsto \partial_i f(x)$*. Plus généralement, étant donnés* $k \in \mathbb{N}^*$*,* $i_1, \dots, i_k \in [\![n]\!]$ *et* $x \in U$*, on note* $\partial_{i_1 \dots i_k} f(x) := \partial_{i_k} \cdots \partial_{i_1} f(x)$ *si le terme de droite est bien défini. De même, si celui-ci est défini pour tout* $x \in \mathbb{R}^n$*, on note* $\partial_{i_1 \dots i_k} f : U \to \mathbb{R}, x \mapsto \partial_{i_1 \dots i_k} f(x)$*.*

Exemple. *Soit* X *un vecteur aléatoire de fonction de répartition* F_X *dont la loi admet une densité* f_X*. On a alors:*

$$F_X(x_1, \dots, x_n) = \int_{]-\infty, x_1] \times \cdots \times]-\infty, x_n]} f_X d\lambda_n = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_X(y_1, \dots, y_n) dy_1 \cdots dy_n$$

où la dernière égalité est par Fubini-Tonelli (le théorème s'applique car f_X *est positive et même trivialement intégrable). En tout point* $x \in \mathbb{R}^n$ *de continuité de* f*, on en déduit:*

$$\partial_{1 \cdots n} F_X(x) = f_X(x)$$

Notons que, par le théorème de Fubini-Tonelli, on peut intégrer dans l'ordre qu'on veut et ainsi l'ordre de dérivation n'a pas d'importance.

À partir de la fonction de répartition F_X d'une vecteur aléatoire $X := (X_1, \dots, X_n)$, on peut retrouver les lois marginales de X. En effet, celles-ci sont données par les fonctions de répartitions respectives $(F_{X_i})_{i\in[\![n]\!]}$ de ses coordonnées et on a, pour $i \in [\![n]\!]$:

$$\begin{aligned} F_{X_i}(x) &= \mathbf{P}(X_i \in]-\infty, x]) = \mathbf{P}_X(\mathbb{R}^{i-1}\times]-\infty, x] \times \mathbb{R}^{n-i}) \\ &= \lim_{N\to+\infty} \mathbf{P}_X\big(]-\infty, N]^{i-1}\times]-\infty, x]\times]-\infty, N]^{n-i}\big) = \lim_{N\to+\infty} F_X(N, \dots, N, x, N, \dots, N) \end{aligned}$$

où la troisième égalité est par le théorème de la limite monotone. Par le même raisonnement, il est possible d'obtenir la fonction de répartition de tout sous-vecteur de X. Il suffit de passer à la limite uniquement sur les indices des composantes qu'on souhaite ignorer.

Exemple. *En reprenant le premier exemple de la section, la fonction de répartition F_Z de Z est donnée par:*

$$\begin{aligned} F_Z : \mathbb{R} &\to [0,1] \\ (x,y) &\mapsto \begin{cases} 0, & \text{si } (x<0 \text{ ou } y<0) \text{ ou } (x\in[0,1[\text{ et } y\in[0,1[) \\ \frac{1}{3}, & \text{si } (x\in[0,1[\text{ et } y\geq 1) \text{ ou } (x\geq 1 \text{ et } y\in[0,1[) \\ 1, & \text{si } x\geq 1 \text{ et } y\geq 1 \end{cases} \end{aligned}$$

En passant à la limite, on retrouve bien les fonctions de répartition respectives F_X et F_Y de X et Y. En effet, pour $(x,y)\in\mathbb{R}^2$, on a:

$$\begin{aligned} F_X(x) &= \lim_{N\to+\infty} F_Z(x,N) = \frac{1}{3}\mathbb{1}_{[0,1[}(x) + \mathbb{1}_{[1,+\infty[}(x) \\ F_Y(y) &= \lim_{N\to+\infty} F_Z(N,y) = \frac{1}{3}\mathbb{1}_{[0,1[}(y) + \mathbb{1}_{[1,+\infty[}(y) \end{aligned}$$

C'est-à-dire, X et Y suivent suivent tous deux une loi de Bernoulli de paramètre $2/3$.

Corollaire 3.2.11. *Soit $X := (X_1, \dots, X_n)$ un vecteur aléatoire de fonction de répartiton F_X. Pour $i \in [\![n]\!]$, on note F_{X_i} la fonction de répartition de X_i. Les coordonnées de X sont indépendantes si et seulement si, pour tout $(x_1, \dots, x_n)\in\mathbb{R}^n$:*

$$F_X(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i)$$

Preuve

Si les coordonnées sont indépendantes, il est immédiat que l'équation est vérifiée. En effet, pour tous $(x_1, \dots, x_n) \in \mathbb{R}^n$:

$$F_X(x_1, \dots, x_n) = \mathbf{P}\left(\bigcap_{i\in[\![n]\!]} \{X_i \leq x_i\}\right) = \prod_{i=1}^n \mathbf{P}(X_i \leq x_i) = \prod_{i=1}^n F_{X_i}(x_i)$$

où la seconde égalité est par définition de l'indépendance de variables aléatoires. De plus, par la proposition 3.2.3, la loi $\mathbf{P}_X$ de X est donnée par $\mathbf{P}_X = \bigotimes_{i\in[\![n]\!]} \mathbf{P}_{X_i}$, où, pour $i \in [\![n]\!]$, $\mathbf{P}_{X_i}$ est la loi de X_i. Supposons maintenant seulement que, pour $x \in \mathbb{R}^n$:

$$F_X(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i)$$

On a vu que cette équation est satisfaite si $\mathbf{P}_X = \bigotimes_{i\in[\![n]\!]} \mathbf{P}_{X_i}$. Or, par la proposition 3.2.10, la fonction de répartion caractérise la loi. On en déduit que nécessairement $\mathbf{P}_X = \bigotimes_{i\in[\![n]\!]} \mathbf{P}_{X_i}$. Par la proposition 3.2.3 encore, on en conclut que les coordonnées de X sont indépendantes.

Exemple. *En reprenant le dernier exemple, on vérifie facilement que X et Y ne sont pas indépendants. En effet, pour $(x,y)\in\mathbb{R}^2$, on a:*

$$\begin{aligned} F_X(x)F_Y(y) &= \left(\frac{1}{3}\mathbb{1}_{[0,1[}(x) + \mathbb{1}_{[1,+\infty[}(x)\right)\left(\frac{1}{3}\mathbb{1}_{[0,1[}(y) + \mathbb{1}_{[1,+\infty[}(y)\right) \\ &= \frac{1}{9}\mathbb{1}_{[0,1[^2}(x,y) + \frac{1}{3}\left(\mathbb{1}_{[0,1[\times[1,+\infty[}(x,y) + \mathbb{1}_{[1,+\infty[\times[0,1[}(x,y)\right) + \mathbb{1}_{[1,+\infty[^2}(x,y) \end{aligned}$$

En particulier, $F_X(1/2)F_Y(1/2) = 1/9$. Or $F_Z(1/2,1/2) = 0$.

Exercice 3.2.6. *Soit $X := (X_1, \dots, X_n)$ un vecteur aléatoire dont les coordonnées sont indépendantes, suivant la même loi. On pose $U = sup(X_1, ..., X_n)$ et $U = inf(X_1, ..., X_n)$.*
Calculer F_U et F_V en fonction de la fonction de répartition F de X_1.

- Fonction de caractéristique

On rappelle que le plan complexe $\mathbb{C}$ peut être identifié au plan euclidien $\mathbb{R}^2$ (en traitant un nombre complexe $a+ib \in \mathbb{C}$ comme un vecteur $(a, b) \in \mathbb{R}^2$). Avec cette identification, le module de $\mathbb{C}$ définit une distance sur $\mathbb{C}$ qui correspond à la distance euclidienne de $\mathbb{R}^2$. On peut alors définir les ouverts de $\mathbb{C}$ (à partir des boules ouvertes pour le module) et donc la tribu borélienne $\mathscr{B}(\mathbb{C})$ sur $\mathbb{C}$ engendrée par les ouverts de $\mathbb{C}$. Ceux-ci correspondent aux ouverts de $\mathbb{R}^2$ lorsque ce dernier est identifié à $\mathbb{C}$. Avec cette même identificaiton et par le corollaire 3.1.3, une fonction d'un espace mesurable $(\Omega, \mathscr{A})$ dans $(\mathbb{C}, \mathscr{B}(\mathbb{C}))$ est mesurable si et seulement si ses parties réelle et imaginaire le sont. Mathématiquement, une fonction $f := f_r + if_i : \Omega \to \mathbb{C}$, où f_r et f_i sont (respectivement) les parties réelle et imaginaire de f, est mesurable si et seulement si $f_r \in \mathscr{M}(\Omega, \mathbb{R})$ et $f_i \in \mathscr{M}(\Omega, \mathbb{R})$. f est appelée une fonction mesurable complexe. On note $\mathscr{M}(\Omega, \mathbb{C})$ l'ensemble des fonctions mesurables complexes sur $(\Omega, \mathscr{A})$. Par l'observation que l'ont vient de faire et car $\mathscr{M}(\Omega, \mathbb{R})$ est une espace vectoriel, $\mathscr{M}(\Omega, \mathbb{C})$ est également un espace vectoriel. Si on muni $(\Omega, \mathscr{A})$ d'une probabilité, alors on peut voir les éléments de $\mathscr{M}(\Omega, \mathbb{C})$ comme des variables aléatoires complexes.

Définition 3.2.12 (Intégrale d'une fonction à valeurs complexe)**.** *Soit $(\mathbb{X}, \mathscr{F}, \mu)$ un espace mesuré et $f := f_r + if_i$ mesurable complexe sur $(\mathbb{X}, \mathscr{F})$. On dit que f est μ-intégrable si:*

$$\int |f| d\mu < +\infty$$

où $|f|$ est le module de f. Si f est μ-intégrable, alors on définit son intégrale par μ la sorte:

$$\int f d\mu = \int f_r d\mu + i \int f_i d\mu \in \mathbb{C}$$

Remarque. *Dans la définition ci-dessus, f est μ-intégrable (en tant que fonction mesurable complexe) si et seulement si f_r et f_i le sont (en tant que fonctions mesurables réelles). En effet:*

$$\int |f| d\mu = \int \sqrt{f_r^2 + f_i^2} d\mu \leq \int \left(\sqrt{f_r^2} + \sqrt{f_i^2} \right) d\mu = \int |f_r| d\mu + \int |f_i| d\mu$$

où la première égalité est par définition du module, l'inégalité car la fonction racine est concave donc sous-additive et par croissance de l'intégrale (d'une fonction mesurable réelle) et la dernière égalité par linéarité de l'intégrale (d'une fonction mesurable réelle). Réciproquement:

$$\int |f_r| d\mu + \int |f_i| d\mu = \int (|f_r| + |f_i|) d\mu \leq \int \sqrt{2} \sqrt{f_r^2 + f_i^2} d\mu = \sqrt{2} \int |f| d\mu$$

Cette équivalence provient en fait de l'équivalence des normes (en l'occurence 1 et 2) en dimension finie, une fois qu'on a fait l'identification de $\mathbb{C}$ avec $\mathbb{R}^2$.

On peut montrer facilement que tous les résultats que l'on a vu pour l'intégrale d'une fonction mesurable réelle sont toujours valides pour celle d'une fonction mesurale complexe, lorsque la transposition a un sens. En particulier, l'intégrale d'une fonction mesurable complexe est linéaire, satisfait l'inégalité triangulaire (avec le module au lieu de la valeur absolue) et la relation de Chasles dénombrable. Avec l'identification de $\mathbb{C}$ à $\mathbb{R}^2$, on peut définir l'espérance d'une variable aléatoire complexe $X := X_r + iX_i$ par:

$$\mathbf{E}[X] = \mathbf{E}[X_r] + i\mathbf{E}[X_i]$$

Cette définition laisse l'espérance linéaire, comme dans le cas réel.

Définition 3.2.13 (Fonction caractéristique)**.** *Soit X une variable aléatoire réelle. On appelle fonction caractéristique de X, notée φ_X, la fonction suivante:*

$$\begin{aligned} \varphi_X : \mathbb{R} &\to \mathbb{C} \\ t &\mapsto \mathbf{E}\left[e^{itX}\right] \end{aligned}$$

Remarque. *Dans la définition ci-dessus, φ_X est bien définie sur tout $\mathbb{R}$. En effet, pour tous $t, x \in \mathbb{R}$, $\left|e^{itx}\right| = 1$. Donc $\mathbf{E}[|e^{itX}|] = 1 < +\infty$, c'est-à-dire e^{itX} est intégrable.*

Soit X une variable aléatoire réelle de loi $\mathbf{P}_X$ et de fonction caractéristique φ_X. On a, pour $t \in \mathbb{R}$:

$$\varphi_X(t) := \mathbf{E}\left[e^{itX}\right] := \mathbf{E}[\cos(tX)] + i\mathbf{E}[\sin(tX)] = \int_\Omega \cos(tX)d\mathbf{P} + i\int_\Omega \sin(tX)d\mathbf{P} = \int_{\mathbb{R}} \cos(tx)\mathbf{P}_X(dx) + i\int_{\mathbb{R}} \sin(tx)\mathbf{P}_X(dx)$$

où la dernière égalité est par le théorème de transfert. Sans perte de généralité, on peut voir les intégrales dans le terme de droite ci-dessus comme des intégrales de fonctions mesurables complexe. Comme on l'a mentionné, celles-ci sont également linéaires. On a donc:

$$\varphi_X(t) = \int_{\mathbb{R}} (\cos(tx) + i\sin(tx))\mathbf{P}_X(dx) = \int_{\mathbb{R}} e^{itx}\mathbf{P}_X(dx)$$

En d'autres termes, le théorème de transfert s'applique toujours pour l'espérance de fonctions (mesurables) complexes de variables aléatoires réelles. On peut ainsi résoudre l'intégrale en traitant i comme une constante réelle. Dans certains cas, cela simplifie le calcul. Attention, on ne peut toutefois pas faire de changement de variable complexe (c'est en fait possible mais requiert des notions d'analyse complexe qu'on abordera pas ici).

Exemple. *Soit X une variable aléatoire de loi $\mathscr{E}(\alpha)$ avec $\alpha \in \mathbb{R}_+^*$. La fonction caractéristique φ_X de X est donnée, pour $t \in \mathbb{R}$, par:*

$$\varphi_X(t) := \mathbf{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx}\mathbb{1}_{\mathbb{R}_+}(x)\alpha e^{-\alpha x}\lambda(dx) = \alpha\int_{\mathbb{R}_+} e^{x(it-\alpha)}\lambda(dx) = \frac{\alpha}{it-\alpha}\left[e^{x(it-\alpha)}\right]_{x=0}^{x\to+\infty}$$

Or, $\lim_{x\to+\infty} e^{x(it-\alpha)} = 0$. En effet:

$$\left|e^{x(it-\alpha)} - 0\right| = \left|e^{x(it-\alpha)}\right| = \left|e^{itx}e^{-\alpha x}\right| = \left|e^{itx}\right|\left|e^{-\alpha x}\right| = e^{-\alpha x} \underset{x\to+\infty}{\longrightarrow} 0$$

On en conclut:

$$\varphi_X(t) = \frac{\alpha}{it-\alpha}$$

Exercice 3.2.7. *Soit X une variable aléatoire de loi $\mathscr{G}(p)$ avec $p \in]0,1]$. Donner la fonction caractéristique φ_X de X. On admettra que la formule pour la somme d'une série géométrique s'étend à toute série géométrique de raison complexe de module strictement inférieur à 1.*

Exercice 3.2.8. *Soit X une variable aléatoire de loi $\mathscr{L}(m,\alpha)$ avec $m \in \mathbb{R}$ et $\alpha \in \mathbb{R}_+^*$. Donner la fonction caractéristique φ_X de X.*

Le premier intérêt de la fonction caractéristique est que, tout comme la fonction de répartition, elle caractérise la loi.

Proposition 3.2.14. *Soient X et Y deux variables aléatoires réelles de fonctions caractéristiques respectives φ_X et φ_Y. X et Y ont même loi si et seulement si $\varphi_X = \varphi_Y$.*

Preuve

Admis. Un résultat, appelé théorème de Lévy (un des nombreux), donne en fait une formule pour obtenir la fonction de répartition à partir de la fonction caractéristique. Comme la fonction de répartition caractérise la loi, la fonction caractéristique aussi.

On a vu que la fonction de répartition se généralise aux vecteurs aléatoires ; c'est en fait aussi le cas de la fonction caractéristique et de ses propriétés.

Définition 3.2.15 (Fonction caractéristique d'un vecteur aléatoire)**.** *Soit $X := (X_1, \dots, X_n)$ un vecteur aléatoire. On appelle fonction caractéristique de X, notée φ_X, la fonction suivante:*

$$\varphi_X : \mathbb{R}^n \to \mathbb{C}$$
$$t := (t_1, \dots, t_n) \mapsto \mathbf{E}[e^{it^*X}] = \mathbf{E}[e^{i\sum_{k=1}^n t_k X_k}]$$

Proposition 3.2.16. *Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de fonction caractéristique φ_X, $A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$ et $v := (v_1, \dots, v_n) \in \mathbb{R}^n$. Alors, la fonction caractéristique φ_{AX+v} de $AX + v$ est donnée, pour $t := (t_1, \dots, t_n) \in \mathbb{R}^n$, par:*

$$\varphi_{AX+v}(t) = e^{it^*v}\varphi_X(A^*t)$$

Preuve

Pour $t \in \mathbb{R}^n$, on a:

$$\varphi_{AX+v}(t) = \mathbf{E}\left[e^{it^*(AX+v)}\right] = \mathbf{E}\left[e^{it^*AX}e^{it^*v}\right] = e^{it^*v}\mathbf{E}\left[e^{it^*AX}\right] = e^{it^*v}\mathbf{E}\left[e^{i(A^*t)^*X}\right] = e^{it^*v}\varphi_X(A^*t)$$

où la troisième égalité est par linéarité de l'espérance de variables aléatoires complexes intégrables.

Proposition 3.2.17. *Deux vecteurs aléatoires X et Y ont la même loi si et seulement si ils ont la même fonction caractéristique.*

Preuve

Admis.

Comme c'est le cas pour la fonction de répartition, la fonction caractéristique peut être utilisée pour montrer l'indépendance des coordonnées d'un vecteur aléatoire.

Proposition 3.2.18. *Soient $X := (X_1, \dots, X_n)$ un vecteur aléatoire de fonction caractéristique φ_X et, pour $k \in [\![n]\!]$, φ_{X_k} la fonction caractéristique de X_k. Les coordonnées de X sont indépendantes si et seulement si, pour tous $t := (t_1, \dots, t_n) \in \mathbb{R}^n$:*

$$\varphi_X(t) = \prod_{k=1}^{n} \varphi_{X_k}(t_k)$$

Preuve

Supposons les coordonnées de X indépendantes et prenons $t := (t_1, \dots, t_n) \in \mathbb{R}^n$. On a:

$$\varphi_Z(t) := \mathbf{E}\left[\prod_{k=1}^{n} e^{it_k X_k}\right] = \mathbf{E}\left[\prod_{k=1}^{n} (\cos(t_k X_k) + i \sin(t_k X_k))\right]$$

On voit ensuite qu'on peut décomposer le produit dans l'espérance en une somme dont chaque terme est un produit de n cosinus et sinus de variables indépendantes, avec un 1, i, -1 ou un $-i$ en facteur selon la parité du nombre de sinus dans le terme. Par linéarité de l'espérance d'une variable aléatoire complexe, on peut sortir ce facteur de chacune des espérances de ces termes. Par la proposition 3.2.3, l'espérance de chacun de ces termes (sans leur facteur) est égal au produit des espérances des sinus et cosinus qui le compose. En refactorisant les termes, on obtient:

$$\varphi_Z(t) = \prod_{k=1}^{n} \mathbf{E}[\cos(t_k X_k) + i \sin(t_k X_k)] = \prod_{k=1}^{n} \mathbf{E}[e^{it_k X_k}] = \prod_{k=1}^{n} \varphi_{X_k}(t_k)$$

Supposons maintenant seulement que $\varphi_Z = \prod_{k=1}^{n} \varphi_{X_k}$. Alors, comme on vient de le montrer, φ_Z est la fonction caractéristique d'un vecteur aléatoire dont les coordonnées sont indépendantes. Par la proposition 3.2.17, la fonction caractéristique caractérise la loi. On en conclut que les coordonnées de X sont nécessairement indépendantes.

Exemple. *Soit $X := (X_1, \dots, X_n)$ un vecteur aléatoire de fonction caractéristique φ_X donnée, pour $t := (t_1, \dots, t_n)$, par:*

$$\varphi_X(t) = e^{-\frac{1}{2} t^* t} = e^{-\frac{1}{2} \sum_{k=1}^{n} t_k^2}$$

On reconnaît que φ_X est le produit de fonctions caractéristiques de variables aléatoires de loi $\mathcal{N}(0,1)$. Par les propositions 3.2.18 et 3.2.17, on en déduit que les coordonnées de X sont indépendantes et que, pour $k \in [\![n]\!]$, X_k a loi $\mathcal{N}(0,1)$. Autrement dit, la loi de X est $\mathbf{P}_X = \mathcal{N}(0,1)^{\otimes n}$. On dit que X est un vecteur gaussien standard.

Un corollaire immédiat de ce résultat est le calcul de la loi d'une somme finie variables aléatoires réelles indépendantes. Dans ce domaine, la fonction caractéristique se révèle bien plus utile que la fonction de répartition et permet souvent de simplifier les calculs.

Corollaire 3.2.19. *Soit $X := (X_1, \dots, X_n)$ un vecteur aléatoire de coordonnées indépendantes et $Z := \sum_{i=1}^{n} X_i$. La fonction caractéristique φ_Z de Z est donnée, pour $t \in \mathbb{R}$, par:*

$$\varphi_Z(t) = \prod_{k=1}^{n} \varphi_{X_k}(t)$$

où, pour $k \in [\![n]\!]$, φ_{X_k} est la fonction caractéristique de X_k.

Preuve

Soit $t \in \mathbb{R}$. On a:

$$\varphi_Z(t) := \mathbf{E}\left[e^{itZ}\right] = \mathbf{E}\left[e^{i \sum_{k=1}^{n} tX_k}\right] = \varphi_X(t, \dots, t)$$

où φ_X est la fonction caractéristique du vecteur X. Par indépendance de ses coordonnées, la proposition 3.2.18 nous donne:

$$\varphi_Z(t) = \prod_{k=1}^{n} \varphi_{X_k}(t)$$

Ainsi, pour donner la fonction caractéristique d'une somme de variables aléatoires indépendantes, il suffit de calculer leurs fonctions caractéristiques respectives et de les multiplier. En guise d'illustration de la puissance de cette technique, on recalcule dans l'exemple suivant la loi d'une somme de variables aléatoires binomiales indépendantes grâce aux fonctions caractéristiques. En TD, on verra d'autres exemples qui peuvent difficilement être couverts sans passer par celles-ci.

Exemple. *Soit X une variable aléatoire de loi $\mathscr{B}(n,p)$ avec $p \in [0,1]$. La fonction caractéristique φ_X de X est donnée, pour $t \in \mathbb{R}$, par:*

$$\varphi_X(t) := \mathbf{E}[e^{itX}] = \sum_{k=0}^{n} e^{itk}\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^{n}\binom{n}{k}(e^{it}p)^k(1-p)^{n-k} = \left(e^{it}p+1-p\right)^n = \left(p(e^{it}-1)+1\right)^n$$

Soient maintenant $(X_1,\dots,X_n)$ un vecteur aléatoire de coordonnées indépendantes tel que, pour $k \in [\![n]\!]$, X_k a loi $\mathscr{B}(n_k,p)$ avec $n_k \in \mathbb{N}$, et $Z := \sum_{k=1}^n X_k$. Par le corollaire 3.2.19, la fonction caractéristique φ_Z de Z est donnée, pour $t \in \mathbb{R}$, par:

$$\varphi_Z(t) = \prod_{k=1}^{n}\varphi_{X_k}(t) = \prod_{k=1}^{n}\left(p(e^{it}-1)+1\right)^{n_k} = \left(p(e^{it}-1)+1\right)^{\sum_{k=1}^n n_k}$$

où, pour $k \in [\![n]\!]$, φ_{X_k} est la fonction caractéristique de X_k. On reconnaît que φ_Z est la fonction caractéristique d'une variable aléatoire de loi $\mathscr{B}(N,p)$, avec $N := \sum_{k=1}^n n_k$. Comme la fonction caractéristique caractérise la loi, on en déduit que la loi de Z est $\mathbf{P}_Z = \mathscr{B}(N,p)$.

3.3 Extension aux suites de variables aléatoires

3.3.1 Théorèmes limites fondamentaux

Soit $(\Omega,\mathscr{A},\mathbf{P})$ un espace de probabilité et $(X_n)_{n\geq 1}$ une suite de variables aléatoires définies sur Ω et à valeurs dans $\mathbb{R}$.

Dans ce qui suit, on parle d'une suite $(X_n)_{n\geq 1}$ i.i.d pour dire que les variables aléatoires sont indépendantes et identiquement distribuées (i.e qu'elles ont toutes même loi).

Définition 3.3.1 (Convergence presque sûre)**.** *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles. On dit que $(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers une variable aléatoire réelle X si:*

$$\mathbf{P}\left(\lim_{n\to+\infty} X_n = X\right) := \mathbf{P}\left(\left\{\omega\in\Omega : \lim_{n\to+\infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

On note cette convergence:

$$X_n \xrightarrow[n\to+\infty]{p.s.} X$$

Exemple. *Soient $(\Omega,\mathscr{A},\mathbf{P}) = (\mathbb{R},\mathscr{B}(\mathbb{R}),\mathscr{U}([0,1]))$. Pour $n\in\mathbb{N}$, on pose $X_n := \mathbb{1}_{[0,\frac{1}{n}]}$. La suite $(X_n)_{n\in\mathbb{N}}$ converge alors presque sûrement vers 0. En effet, on a:*

$$\mathbf{P}\left(\lim_{n+\infty} X_n = 0\right) = \mathbf{P}\left(\left\{\omega\in\Omega : \lim_{n\to+\infty} X_n(\omega) = 0\right\}\right) = \mathbf{P}\left(\left\{x\in\mathbb{R} : \mathbb{1}_{\{0\}}(x) = 0\right\}\right) = \mathbf{P}(]0,1]) = 1$$

Définition 3.3.2 (Convergence en loi)**.** *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles. On dit que $(X_n)_{n\in\mathbb{N}}$ converge en loi (ou en distribution ou faiblement) vers une variable aléatoire X si, pour toute fonction continue bornée $\varphi : \mathbb{R}\to\mathbb{R}$:*

$$\lim_{n\to+\infty}\mathbf{E}[\varphi(X_n)] = \mathbf{E}[\varphi(X)]$$

On note cette en général cette convergence

$$X_n \xrightarrow[n\to+\infty]{d} X$$

Remarque. *Dans la notation ci-dessus, le d, pour distribution, est parfois remplacé par un L calligraphié ($\mathscr{L}$ ou $\mathscr{L}$), pour loi (ou "law" en anglais). Il est également courant de voir le mot "law" écrit au lieu de d.*

Définition 3.3.3 (Convergence en loi)**.** *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles. On dit que $(X_n)_{n\in\mathbb{N}}$ **converge en loi** (ou en distribution, ou faiblement) vers une variable aléatoire X si, pour toute fonction continue bornée $\varphi : \mathbb{R}\to\mathbb{R}$:*

$$\lim_{n\to\infty}\mathbb{E}[\varphi(X_n)] = \mathbb{E}[\varphi(X)].$$

On note alors $X_n \xrightarrow{\mathscr{L}} X$.

Définition 3.3.4 (Convergence en probabilité). *On dit que* $(X_n)_{n\in\mathbb{N}}$ ***converge en probabilité*** *vers* X *si, pour tout* $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

On note $X_n \xrightarrow{\mathbb{P}} X$.

Définition 3.3.5 (Convergence en moyenne quadratique). *On dit que* $(X_n)_{n\in\mathbb{N}}$ ***converge en moyenne quadratique*** *(ou en* L^2*) vers* X *si*

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^2] = 0.$$

On note $X_n \xrightarrow{L^2} X$.

Théorème 3.3.6. *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires réelles et* X *une variable aléatoire réelle. On a les implications suivantes:*

- *Si* $(X_n)_{n\in\mathbb{N}}$ *converge vers* X *en moyenne quadratique, alors elle converge également en probabilité.*
- *Si* $(X_n)_{n\in\mathbb{N}}$ *converge vers* X *en probabilité, alors elle converge également en distribution.*
- *Si* $(X_n)_{n\in\mathbb{N}}$ *converge vers* X *presque sûrement, alors elle converge également en probabilité.*

Preuve

Admis.

Relations entre les convergences :

$$X_n \xrightarrow{L^2} X \;\Rightarrow\; X_n \xrightarrow{\mathbb{P}} X \;\Rightarrow\; X_n \xrightarrow{\mathscr{L}} X,$$

et

$$X_n \xrightarrow{p.s.} X \;\Rightarrow\; X_n \xrightarrow{\mathbb{P}} X$$

les réciproques étant fausses en général.

Exemple. *Soit* $(X_n)_{n\in\mathbb{N}}$ *définie par :*

$$X_n = \begin{cases} 1 & \text{avec probabilité } \frac{1}{n}, \\ 0 & \text{avec probabilité } 1 - \frac{1}{n}. \end{cases}$$

et soit $X = 0$.

- ***Convergence en loi :*** $X_n \xrightarrow{\mathscr{L}} 0$.
- ***Convergence en probabilité :***

  $$\mathbb{P}(|X_n - 0| > \varepsilon) = \tfrac{1}{n} \to 0,$$

  donc $X_n \xrightarrow{\mathbb{P}} 0$.
- ***Convergence en moyenne quadratique :***

  $$\mathbb{E}[|X_n - 0|^2] = \tfrac{1}{n} \to 0,$$

  donc $X_n \xrightarrow{L^2} 0$.

Exemple. ***Contre-exemple :*** *Si* X_n *est uniforme sur* $\{0, 1, \ldots, n\}$, *alors* $X_n \xrightarrow{\mathscr{L}} +\infty$ *mais ne converge pas en probabilité ni en* L^2 *vers une variable finie.*

Exemple (Convergence p.s. mais pas en L^2). *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires définie par*

$$X_n = \begin{cases} n & \text{avec probabilité } \frac{1}{n}, \\ 0 & \text{avec probabilité } 1 - \frac{1}{n}. \end{cases}$$

et posons $X = 0$.

- ***Convergence presque sûre :*** *On a*

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq 0) = \sum_{n=1}^{\infty} \tfrac{1}{n} = +\infty,$$

donc par le lemme de Borel–Cantelli, $X_n \neq 0$ infiniment souvent avec probabilité 1. Cependant, quand $X_n \neq 0$, on a $X_n = n \to \infty$, et donc $X_n \to 0$ ***ne se produit pas presque sûrement****.*

Correction : Pour obtenir un vrai exemple de convergence ***presque sûre*** *mais pas en L^2, on prend plutôt :*

$$X_n = \begin{cases} \sqrt{n} & \text{avec probabilité } \frac{1}{n}, \\ 0 & \text{avec probabilité } 1 - \frac{1}{n}. \end{cases}$$

Alors :

- ***Presque sûrement :*** *$X_n \to 0$, car*

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq 0) = \sum_{n=1}^{\infty} \tfrac{1}{n} = +\infty,$$

et pourtant, comme $X_n = \sqrt{n}$ de manière très rare, on peut montrer (via Borel–Cantelli) que la suite prend la valeur non nulle seulement un nombre fini de fois presque sûrement. Donc $X_n \to 0$ p.s.

- ***Mais pas en L^2 :***

$$\mathbb{E}[|X_n|^2] = \tfrac{1}{n} \cdot n = 1,$$

donc X_n ne converge pas vers 0 en L^2.

Théorème 3.3.7 (Loi (forte) des grands nombres)**.** *Soit $(X_k)_{k\in[\![n]\!]}$ une suite de variables aléatoires réelles i.i.d. positives ou intégrables. Alors:*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{p.s.} m$$

où $m := \mathbf{E}[X_1] \in \mathbb{R} \cup \{+\infty\}$.

Preuve

Admis.

Théorème 3.3.8 (Théorème central limite)**.** *Soit $(X_k)_{k\in\mathbb{N}}$ une suite de variables aléatoires réelles i.i.d. de carrés intégrables, $m := \mathbf{E}[X_1]$ et $\sigma := \sqrt{Var(X_1)}$. Alors:*

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} \frac{X_k - m}{\sigma} \xrightarrow[n\to+\infty]{d} Z$$

où Z est une variable aléatoire de loi $\mathcal{N}(0,1)$.

Preuve

Admis.

Annexes

A Lois usuelles et variables aléatoires associées

Lois à support fini

- Loi mesure de Dirac δ_ω centrée en $\omega \in \Omega$ est une probabilité sur un espace mesurable $(\Omega, \mathscr{A})$ tel que $\{\omega\} \in \mathscr{A}$.
- La loi uniforme $\mathscr{U}(\Omega)$ sur un ensemble Ω fini est la probabilité suivante sur $(\Omega, \mathscr{P}(\Omega))$:

$$\mathscr{U}(\Omega) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} \delta_\omega$$

- La loi de Bernoulli $\mathscr{B}(p)$ de paramètre $p \in [0,1]$ est la probabilité suivante sur $(\{0,1\}, \mathscr{P}(\{0,1\}))$:

$$\mathscr{B}(p) = p\delta_1 + (1-p)\delta_0$$

- La loi binomiale $\mathscr{B}(n,p)$ de paramètres $(n,p) \in \mathbb{N} \times [0,1]$ est la probabilité suivante sur $(\llbracket 0;n \rrbracket, \mathscr{P}(\llbracket 0;n \rrbracket))$:

$$\mathscr{B}(n,p) = \sum_{k \in \llbracket 0;n \rrbracket} \binom{n}{k} p^k (1-p)^{n-k} \delta_k$$

Lois discrètes à support infini

- La loi de Poisson $\mathscr{P}(\alpha)$ de paramètre $\alpha \in \mathbb{R}_+^*$ est la probabilité suivante sur $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$:

$$\mathscr{P}(\alpha) = \sum_{k \in \mathbb{N}} e^{-\alpha} \frac{\alpha^k}{k!} \delta_k$$

- La loi géométrique $\mathscr{G}(p)$ de paramètre $p \in]0,1]$ est la probabilité suivante sur $(\mathbb{N}^*, \mathscr{P}(\mathbb{N}^*))$:

$$\mathscr{G}(p) = \sum_{k \in \mathbb{N}^*} p(1-p)^{k-1} \delta_k$$

Lois à densité sur $\mathbb{R}$

Soit λ la mesure de Lebesgue sur $\mathbb{R}$.

- La loi uniforme $\mathscr{U}(B)$ sur un borélien $B \in \mathscr{B}(\mathbb{R})$ tel que $\lambda(B) > 0$ a pour densité:

$$\frac{d\mathscr{U}(B)}{d\lambda} = \frac{1}{\lambda(B)} \mathbb{1}_B$$

- La loi exponentielle $\mathscr{E}(\alpha)$ de paramètre $\alpha \in \mathbb{R}_+^*$ a pour densité:

$$\frac{d\mathscr{E}(\alpha)}{d\lambda} : x \mapsto \mathbb{1}_{\mathbb{R}_+}(x) \alpha e^{-\alpha x}$$

- La loi normale (ou gaussienne) $\mathscr{N}(m, \sigma^2)$ de paramètres $(m, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ a pour densité:

$$\frac{d\mathscr{N}(\mu, \sigma^2)}{d\lambda} : x \mapsto \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{x-m}{\sigma} \right)^2 \right\}$$

  On définit appelle également loi normale dégénérée $\mathscr{N}(m,0) \coloneqq \delta_m$.

- La loi gamma $\mathscr{G}(\alpha, \beta)$ de paramètres $(\alpha, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ a pour densité:

$$\frac{d\mathscr{G}(\alpha, \beta)}{d\lambda} : x \mapsto \mathbb{1}_{\mathbb{R}_+^*}(x) \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

  où

$$\Gamma(\alpha) \coloneqq \int_{\mathbb{R}_+^*} t^{\alpha-1} e^{-t} \lambda(dt) \in \mathbb{R}_+^*$$

- La loi de Laplace $\mathscr{L}(m,\alpha)$ de paramètres $(m,\alpha)\in\mathbb{R}\times\mathbb{R}_+^*$ a pour densité:

$$\frac{d\mathscr{L}(m,\alpha)}{d\lambda}:x\mapsto\frac{1}{2\alpha}\exp\left\{-\frac{|x-m|}{\alpha}\right\}$$

- La loi Beta $Beta(a,b)$ de paramètres $(a,b)\in\mathbb{R}_+^*\times\mathbb{R}_+^*$ a pour densité:

$$\frac{dBeta(a,b)}{d\lambda}:x\mapsto\mathbb{1}_{]0,1[}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$$

- La loi de Cauchy $\mathscr{C}(m,\alpha)$ de paramètres $(m,\alpha)\in\mathbb{R}\times\mathbb{R}_+^*$ a pour densité:

$$\frac{d\mathscr{C}(m,\alpha)}{d\lambda}:x\mapsto\frac{1}{\pi\alpha\left(1+\left(\frac{x-m}{\alpha}\right)^2\right)}$$

Tableau des moments de lois usuelles

Moments des lois usuelles			
	Valeurs prises $X(\Omega)$	Espérance $\mathbf{E}(X)$	Variance $\mathbb{V}(X)$
Loi de Bernoulli $\mathscr{B}(p)$	$\{0,1\}$	p	$p(1-p)$
Loi binomiale $\mathscr{B}(n,p)$	$\{0,...,n\}$	np	$np(1-p)$
Loi de Poisson $\mathscr{P}(\alpha)$	$\mathbb{N}$	α	α
Loi géométrique $\mathscr{G}(p)$	$\mathbb{N}^*$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Loi uniforme $\mathscr{U}([a,b])$	$[a,b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Loi exponentielle $\mathscr{E}(\alpha)$	$\mathbb{R}_+$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Loi normale $\mathscr{N}(m,\sigma^2)$	$\mathbb{R}$	m	σ^2

B Cardinalité des ensembles

Pour citer Wikipédia, "la cardinalité est une notion de taille pour les ensembles". Dans le cas d'un ensemble fini, c'est-à-dire dont les éléments peuvent être énumérés dans une liste finie, le cardinal de l'ensemble peut être défini comme la taille de cette liste (il est alors identifiable à un élément de $\mathbb{N}$). On note le plus souvent $|\Omega|$ ou $Card(\Omega)$ le cardinal d'un ensemble Ω. Il est facile de voir que deux ensembles finis ont même cardinal si et seulement si il sont en bijection. De même, pour deux ensembles finis E et F, $|E| \leq |F|$ si et seulement si il existe une injection de E dans F. Dans le cas d'ensembles quelconques, potentiellement infinis (i.e. non finis), définir rigoureusement la cardinalité s'avère moins évident et demande des notions avancées de théorie des ensembles. La définition formelle part toutefois des observations faites ci-dessus dans le cas fini. On se contentera donc ici de donner une définition purement fonctionnelle de la cardinalité basée sur ces dernières. Pour référence, tout ce qui suit est valable pour la théorie des ensembles de Zermelo-Fraenkel avec l'axiome du choix (abrégée ZFC). En particulier, sans ce dernier, les résultats énoncés sont faux dans leur généralité et la notion de cardinal n'est pas toujours bien définie.

Définition 2.1. *Soient E et F deux ensembles.*

- *On dit que E et F sont équipotents (ont même cardinal), noté $|E| = |F|$, si il existe une bijection de E dans F.*
- *E a un cardinal plus petit ou égal à celui de F, noté $|E| \leq |F|$, si il existe une injection de E dans F.*
- *Si $|E| \neq |F|$ et $E \leq F$, alors E a un cardinal strictement plus petit que celui de F et on note $|E| < |F|$.*

Remarque. *Dans la définition ci-dessus, si $|E| \leq |F|$ et $|E| \geq |F|$, alors $|E| = |F|$. En effet, ceci revient à dire que, si il existe une injection de E dans F et une injection de F dans E, alors E et F sont en bijection. C'est un résultat classique de ZFC connu sous le nom de théorème de Cantor-Bernstein. De même, comme la composition de deux fonctions bijectives est bijective, si $|E| = |F|$ et qu'un ensemble G est tel que $|G| = |F|$, alors $|E| = |G|$. En mots, l'équipotence est une relation transitive (c'est en fait une relation d'équivalence). Enfin, il est intéressant de noter que l'existence d'une injection de E dans F revient à l'existence d'une surjection de F dans E.*

Remarque. *Un autre résultat important de ZFC est que les cardinaux de deux ensembles sont nécessairement comparables (i.e. il existe toujours une injection de l'un dans l'autre). Autrement dit, les cardinaux sont ordonnés.*

Un ensemble qui peut s'injecter dans $\mathbb{N}$ est dit dénombrable (ou au plus dénombrable). Dit autrement, un ensemble est dénombrable si on peut énumérer ses éléments. Les ensembles en bijection avec $\mathbb{N}$ sont les uniques ensembles infinis dénombrables. Par exemple, l'ensemble des entiers relatifs $\mathbb{Z}$ est (infini) dénombrable (il suffit d'énumérer les nombres par éloignement de 0: 0,1,-1,2,-2,3,-3,...). De même, il est assez facile de montrer que $\mathbb{N}^2$ est dénombrable (si on représente $\mathbb{N}^2$ dans un tableau dont les lignes et les colonnes sont indexées par $\mathbb{N}$, on peut passer successivement sur chacun des éléments en "zigzagant" sur le tableau à partir de la coordonnée $(0,0)$). On en déduit que l'ensemble $\mathbb{Q}$ des rationnels (fractions de deux entiers relatifs) est dénombrable. Cet argument s'adapte à $\mathbb{N}^n$ (pour $n \in \mathbb{N}^*$) et, de manière plus générale, pour tout ensemble infini Ω et $n \in \mathbb{N}^*$, $|\Omega| = |\Omega^n|$. En particulier, pour $n \in \mathbb{N}^*$, $\mathbb{R}$ et $\mathbb{R}^n$ sont équipotents. On peut se demander si $\mathbb{R}$ est dénombrable et si il existe des ensembles de cardinal strictement supérieur à celui de $\mathbb{R}$. Nous allons voir que les réponses à ces questions sont respectivement non et oui.

Proposition 2.2. *Le cardinal de $\mathbb{N}$ est strictement plus petit que celui de l'ensemble de ses parties, symboliquement:*

$$|\mathbb{N}| < |\mathscr{P}(\mathbb{N})|$$

Lemme 2.3. *Soit Ω un ensemble. Alors $\mathscr{P}(\Omega)$ et $\{0,1\}^\Omega$ sont équipotents.*

Preuve

Par définition, on cherche une bijection de $\mathscr{P}(\Omega)$ dans $\{0,1\}^\Omega$, l'ensemble des fonctions sur Ω à valeurs dans $\{0,1\}$. On définit:

$$\begin{aligned} B : \mathscr{P}(\Omega) &\to \{0,1\}^\Omega \\ E &\mapsto (\Omega \to \{0,1\}, \omega \mapsto \mathbb{1}_E(\Omega)) \end{aligned}$$

En mots, B est la fonction sur $\mathscr{P}(\Omega)$ qui à un ensemble $E \in \mathscr{P}(\Omega)$ associe la fonction sur Ω qui renvoie 1 pour les éléments de Ω dans E et 0 pour les autres. On vérifie immédiatement que B est une bijection. En effet, étant donné une fonction f de Ω dans $\{0,1\}$, on peut définir $B^{-1}(f) := \{\omega \in \Omega : f(\omega) = 1\} \in \mathscr{P}(\Omega)$.

Preuve

[Preuve de la proposition 2.2] Par le lemme 2.3, $\mathscr{P}(\mathbb{N})$ et $\{0,1\}^\mathbb{N}$ sont équipotents. On va donc chercher à montrer qu'il n'existe aucune bijection de $\mathbb{N}$ dans $\{0,1\}^\mathbb{N}$. On va utiliser un argument par l'absurde appelé la diagonale de Cantor. Une fonction sur $\mathbb{N}$ est une suite. Supposons donc qu'il existe une suite $B := (b^n)_{n\in\mathbb{N}}$, où pour $n \in \mathbb{N}$,

$b^n := (b^n_k)_{k\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$, qui énumère de façon unique les éléments de $\{0,1\}^{\mathbb{N}}$ (l'espace des suites à valeurs dans $\mathbb{N}$). On va montrer qu'il existe une suite $c := (c_k)_{k\in\mathbb{N}}$ à valeurs dans $\{0,1\}$ qui n'est pas dans B. Pour tout $k \in \mathbb{N}$, on pose

$$c_k := \begin{cases} 1, & \text{si } b^k_k = 0 \\ 0, & \text{si } b^k_k = 1 \end{cases}$$

En mots, c est la suite obtenue en prenant les opposés des éléments diagonaux du tableau infini où, pour $n \in \mathbb{N}$, la $n^{\text{ème}}$ ligne correspond à la suite b^n, d'où le nom de l'argument. Il est alors facile de vérifier qu'il n'existe aucun $n \in \mathbb{N}$ tel que $b^n = c$. Suppons qu'il existe un tel $n \in \mathbb{N}$. Alors, pour tout $k \in \mathbb{N}$, $b^n_k = c_k$. Or, par définition, $b^n_n \neq c_n$. On a donc trouvé une suite c qui n'est pas dans B. Par contradiction, on en déduit que $|\mathbb{N}| \neq \left|\{0,1\}^{\mathbb{N}}\right|$. Pour montrer que $|\mathbb{N}| \leq \left|\{0,1\}^{\mathbb{N}}\right|$, et donc $|\mathbb{N}| < \left|\{0,1\}^{\mathbb{N}}\right|$, il suffit de trouver une injection de $\mathbb{N}$ dans $\{0,1\}^{\mathbb{N}}$. On peut par exemple prendre la fonction qui à $n \in \mathbb{N}$ associe la suite dans $\{0,1\}^{\mathbb{N}}$ dont le $n^{\text{ème}}$ élément est égal à 1 et les autres à 0.

Proposition 2.4. *$\mathbb{R}$ et $\mathscr{P}(\mathbb{N})$ sont équipotents. En particulier, $\mathbb{R}$ n'est pas dénombrable.*

Preuve

Par le lemme 2.3, il équivalent de montrer que $\mathbb{R}$ et $\{0,1\}^{\mathbb{N}}$ sont équipotents. Pour tout réel a, il existe un entier $N \in \mathbb{N}$ et une suite $(a_k)_{k\in\mathbb{N}}$ à valeurs dans $\{0,1\}$ tels que a peut s'écrire $a = a_0 \dots a_N, a_{N+1}a_{N+2}a_{N+3} \dots$ en base 2. Pour $x := (x_k)_{k\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$, on note $N_x := \inf\{n \in \mathbb{N} : x_k = 0\}$ l'indice du premier 0 de la suite x et on définit le réel $S(x) \in \mathbb{R}$ par:

$$S(x) := \sum_{k\in\mathbb{N}} x_{N_x+1+k} 2^{N_x-k}$$

En mots, $S(x)$ est le réel qui peut s'écrire $x_{N_x+1} \dots x_{2N_x+2}, x_{2N_x+3}x_{2N_x+4}x_{2N_x+5} \dots$ en base 2. Pour tout réel $a \in \mathbb{R}$, il existe donc une suite $x \in \{0,1\}^{\mathbb{N}}$ telle que $S(x) = a$. De plus, $S(x)$ est bien définie pour toute suite $x \in \{0,1\}^{\mathbb{N}}$. L'application $S : \{0,1\}^{\mathbb{N}} \to \mathbb{R}, x \mapsto S(x)$ est donc une surjection et on a $\left|\{0,1\}^{\mathbb{N}}\right| \geq |\mathbb{R}|$. On va maintenant chercher à injecter $\{0,1\}^{\mathbb{N}}$ dans $\mathbb{R}$. Certains réels admettent deux écritures en base 2. Par exemple, $0,5$ (en base 10) peut s'écrire $0,1$ ou $0,01111\dots$ en base 2. Ainsi $S((0,0,1,0,0,0,0,\dots)) = S((0,0,0,1,1,1,1,\dots)) = 0.5$ et S n'est pas bijective. Les suites se terminant par une infinité consécutive de 1s sont les seules pour lesquelles leur image par S admet un autre antécédent (obtenu en remplaçant le dernier 0 par un 1 et les 1s suivant par des 0s). En se servant de cette observation, il est ainsi possible de modifier S pour la rendre injective. Pour $x := (x_k)_{k\in\mathbb{N}}$, définissons le réel $I(x) \in \mathbb{R}$ par:

$$I(x) := S((x_0, 0, x_1, 0, x_2, 0, x_3, 0, \dots))$$

Par la remarque précédente, il est alors clair que la fonction $I : \{0,1\} \to \mathbb{R}, x \mapsto I(x)$ est injective et on a $\left|\{0,1\}^{\mathbb{N}}\right| \leq |\mathbb{R}|$. On en déduit $\left|\{0,1\}^{\mathbb{N}}\right| = |\mathbb{R}|$. Par la proposition 2.2, $\mathbb{R}$ n'est pas dénombrable.

L'existence d'un ensemble E tel que $|\mathbb{N}| < |E| < |\mathbb{R}|$ est une question ouverte. Sa non-existence est une conjecture connue sous le nom d'hypothèse du continu (on appelle cardinal du continu le cardinal de $\mathbb{R}$). Le théorème suivant nous dit que la proposition 2.2 est en fait valable pour tout ensemble (et non seulement $\mathbb{N}$).

Théorème 2.5 (Théorème de Cantor)**.** *Pour tout ensemble Ω:*

$$|\Omega| < |\mathscr{P}(\Omega)|$$

Preuve

Soit Ω un ensemble. Si Ω est fini alors il existe $n \in \mathbb{N}$ tel que $|\Omega| = n$ et l'inégalité est trivialement vérifiée car $|\mathscr{P}(\Omega)| = 2^n > n$. Supposons donc Ω infini. Par le lemme 2.3 encore, il est équivalent de montrer que $|\Omega| < \left|\{0,1\}^{\Omega}\right|$. On va voir que l'argument de la diagonale de Cantor s'adapte parfaitement à un ensemble quelconque. Supposons qu'il existe une bijection $B : \Omega \to \{0,1\}^{\Omega}$. Construisons alors la fonction $c : \Omega \to \{0,1\}$ définie, pour $\omega \in \Omega$, par:

$$c(\omega) := \begin{cases} 1, & \text{si } B(\omega)(\omega) = 0 \\ 0, & \text{si } B(\omega)(\omega) = 1 \end{cases}$$

On vérifie alors qu'il n'existe aucun $\omega \in \Omega$ tel que $B(\omega) = c$. En effet, suppons qu'il existe un tel ω. Cela signifie que, pour tout $\omega' \in \Omega$, $B(\omega)(\omega') = c(\omega')$. Or, par construction $B(\omega)(\omega) \neq c(\omega)$. On en conclut que c n'est pas dans l'image de B et donc B n'est pas bijective. On a alors une contradiction et on en conclut que $|\Omega| \neq \left|\{0,1\}^{\Omega}\right|$. Considérons maintenant la fonction $I : \Omega \to \{0,1\}^{\Omega}, \omega \mapsto \mathbb{1}_{\{\omega\}}$. Il est clair que I est injective et donc $|\Omega| < \left|\{0,1\}^{\Omega}\right|$.

L'existence de deux ensembles infinis Ω et Γ tels que $|\Omega| < |\Gamma| < |\mathscr{P}(\Omega)|$ est également une question ouverte (qui comprend donc l'hypothèse du continu). Leur non-existence est une conjecture appelée hypothèse du continu généralisée. Une première conséquence du théorème de Cantor est qu'il n'existe pas d'ensemble "maximal" au sens de la cardinalité. En particulier, il existe une infinité de cardinaux d'ensembles infinis. En pratique, la quasi-exclusivité

des calculs en mathématique se font sur des ensembles dont la cardinalité est au plus celle du continu. C'est notamment le cas en théorie de la mesure où on équipe $\mathbb{R}$ de la tribu borélienne et non de sa tribu discrète. On peut en fait montrer que $|\mathscr{B}(\mathbb{R})| = |\mathbb{R}|$. Pour être plus précis, il est possible, au prix d'efforts techniques, d'étendre la mesure de Lebesgue à une tribu équipotente à $\mathscr{P}(\mathbb{R})$ tout en conservant ses propriétés fondamentales (en particulier l'invariance par translation), mais cette extension ne consiste qu'à ajouter des ensembles de mesure nulle. Elle a ainsi un intérêt très limité et n'est en général pas considérée.

C Dénombrabilité

Définition 3.1. *Un ensemble E est dit **dénombrable** si $|E| \leq |\mathbb{N}|$, c'est-à-dire s'il existe une injection de E dans $\mathbb{N}$. On distingue deux cas :*

- *Si E est fini, on dit que E est **fini dénombrable**.*
- *Si E est infini et qu'il existe une bijection $E \simeq \mathbb{N}$, on dit que E est **infini dénombrable**.*

Ainsi, un ensemble est dénombrable si et seulement si ses éléments peuvent être énumérés (dans une liste finie ou infinie).

Proposition 3.2.

1. *Toute partie d'un ensemble dénombrable est dénombrable.*
2. *Toute union finie d'ensembles dénombrables est dénombrable.*
3. *L'union dénombrable d'ensembles dénombrables est dénombrable.*
4. *Si A et B sont dénombrables, alors $A \times B$ est dénombrable. Par récurrence, A^k est dénombrable pour tout $k \in \mathbb{N}^*$.*
5. *Tout ensemble infini dénombrable est équipotent à $\mathbb{N}$.*

Exemple. *L'ensemble $\mathbb{N}$ est dénombrable par définition.*

Exemple. *L'ensemble $\mathbb{Z}$ est dénombrable : on peut énumérer ses éléments en partant de 0 et en alternant positifs et négatifs :*

$$0, 1, -1, 2, -2, 3, -3, \dots$$

Exemple. *L'ensemble $\mathbb{N}^2$ est dénombrable : on peut représenter ses éléments dans un tableau et les parcourir en diagonale (argument du zigzag).*

Exemple. *Par récurrence, $\mathbb{N}^k$ est dénombrable pour tout $k \in \mathbb{N}^*$. En particulier, l'ensemble des rationnels $\mathbb{Q}$, qui s'identifie à une sous-partie de $\mathbb{Z} \times \mathbb{N}^*$, est dénombrable.*

Exemple. *En revanche, $\mathbb{R}$ n'est pas dénombrable (preuve par l'argument diagonal de Cantor).*

D Lien entre Intégrales de Lebesgue et Riemann

On va voir que les intégrales de Riemann sont des cas particuliers de l'intégrale par rapport à Lebesgue. Soient $a, b \in \mathbb{R}$ avec $a < b$ et $\Pi := \bigcup_{n\in\mathbb{N}^*} \Pi_n$ où, pour $n \in \mathbb{N}^*$, on pose $\Pi_n := \{t_0, \dots, t_n \in [a, b] : a = t_0 \leq \cdots \leq t_n = b\}$ l'ensemble des (bornes de) partitions de $[a, b]$ en n intervalles. Pour $n \in \mathbb{N}^*$ et $P := (t_0, \dots, t_n) \in \Pi_n$, on écrit:

$$L(P, f) = \sum_{i=0}^{n-1} \min_{x\in[t_i,t_{i+1}]} f(x)(t_{i+1} - t_i) \qquad U(P, f) = \sum_{i=0}^{n-1} \max_{x\in[t_i,t_{i+1}]} f(x)(t_{i+1} - t_i)$$

$L(P, f)$ est appelée une somme de Riemann inférieure (le L est pour "lower") et $U(P, f)$ une somme de Riemann supérieure (le U est pour "upper"). On rappelle qu'une fonction réelle f est Riemann-intégrable sur $[a, b]$ si et seulement si elle est bornée sur $[a, b]$ et:

$$\sup_{P\in\Pi} L(P, f) = \inf_{P\in\Pi} U(P, f)$$

Auquel cas, on pose $\int_a^b f(x)dx := \sup_{P\in\Pi} L(P, f)$. C'est par exemple le cas des fonctions continues par morceaux. Pour $n \in \mathbb{N}^*$, une partition $P = (t_0, \dots, t_n) \in \Pi_n$ et une fonction mesurable réelle f, on définit la fonction étagée suivante:

$$s(P, f) = \sum_{i=0}^{n-1} \min_{x\in[t_i,t_{i+1}]} f(x)\mathbb{1}_{[t_i,t_{i+1}[}$$

Pour une telle fonction, on a $\int_a^b s(P,f)(x)dx = L(P,f)$. Un résultat en théorie de l'intégration de Riemann nous dit que si une fonction réelle positive f est Riemann-intégrable sur $[a,b]$, alors il existe une suite emboitée de partitions $(P_n)_{n\in\mathbb{N}}$ telle que $s(P_n,f)$ converge simplement vers f et $\lim_{n\to+\infty} L(P_n,f) = \int_a^b f(x)dx$ (le point important étant que les partitions sont emboitées, c'est-à-dire P_n" $\subset$ "P_{n+1} pour tout $n \in \mathbb{N}$).

Proposition 4.1 (Intégration par Lebesgue d'une fonction Riemann-intégrable)**.** *Soient $a,b \in \mathbb{R}$ avec $a<b$, $f:\mathbb{R}\to\mathbb{R}$ une fonction Riemann-intégrable sur $[a,b]$ et λ la mesure de Lebesgue. Alors f est intégrable au sens de Lebesgue et*

$$\int_{[a,b]} f d\lambda = \int_a^b f(x)dx$$

Preuve

Pour $n \in \mathbb{N}^*$ et une telle partition $P \in \Pi_n$, l'intégrale de Riemann de $s(P,f)$ est donnée par:

$$\int_a^b s(P,f)(x)dx = \sum_{i=0}^{n-1} \min_{x\in[t_i,t_{i+1}]} f(x)(t_{i+1}-t_i) = \sum_{i=0}^{n-1} \min_{x\in[t_i,t_{i+1}]} f(x)\lambda([t_i,t_{i+1}]) = \int s(P,f)d\lambda = \int_{[a,b]} s(P,f)d\lambda$$

où λ et la mesure de Lebesgue et la dernière égalité est car $s(P,f)$ est nulle sur $[a,b]^c$. Puisqu'on s'intéresse à l'intégrale de f sur $[a,b]$, on peut supposer f nulle sur $[a,b]^c$ sans perte de généralité. Premièrement, puisque f est bornée, il existe $M \in \mathbb{R}_+^*$ tel que $|f| \le M$. Par monotonicité de l'intégrale, on a donc:

$$\int |f| d\lambda = \int_{[a,b]} |f| d\lambda \le \int_{[a,b]} M d\lambda = M\lambda([a,b]) = M(b-a) < +\infty$$

Donc f est intégrable. Supposons f positive. Par Riemann-intégrabilité de f, il existe une suite $(P_n)_{n\in\mathbb{N}^*}$ de partitions emboitées telles que f est limite simple de $(s(P_n,f))_{n\in\mathbb{N}}$ et $\lim_{n\to+\infty} L(P_n,f) = \int_a^b f(x)dx)$. Puisque les partitions sont emboitées et f est positive, on a que $(s(P_n,f))_{n\in\mathbb{N}}$ est une suite de fonctions étagées positives croissantes point par point. En appliquant le théorème de convergence monotone, on obtient:

$$\int_a^b f(x)dx = \lim_{n\to+\infty} L(P_n,f) = \lim_{n\to+\infty} \int_a^b s(P_n,f)(x)dx = \lim_{n\to+\infty} \int_{[a,b]} s(P_n,f)d\lambda = \int_{[a,b]} f d\lambda$$

Supposons maintenant f seulement intégrable. On a alors:

$$\int_a^b f(x)dx = \int_a^b (f_+(x)-f_-(x))dx = \int_a^b f_+(x)dx - \int_a^b f_-(x)dx = \int_{[a,b]} f_+ d\lambda - \int_{[a,b]} f_- d\lambda = \int_{[a,b]} (f_+ - f_-)d\lambda = \int_{[a,b]} f d\lambda$$

où la seconde égalité est par linéarité de l'intégrale de Riemann, la troisième par le point précédent et la quatrième par linéarité de l'intégrale par une mesure.

On rappelle qu'il est possible d'étendre l'intégrale de Riemann à des intervalles ou fonctions non bornées. On parle alors d'intégrale "impropre". Pour $a,b \in \bar{\mathbb{R}}$ et une fonction $f:]a,b[\to\mathbb{R}_+$ Riemann-intégrable sur tout segment $[c,d]$ avec $c,d\in\mathbb{R}$ tels que $a<c<d<b$, on définit $\int_a^b f(x)dx := \lim_{d\to b^-}\lim_{c\to a^+}\int_c^d f(x)dx \in \bar{\mathbb{R}}_+$. Si f n'est plus à valeurs dans $\mathbb{R}_+$ mais $\mathbb{R}$ et que $\int_a^b f_+(x)dx$ et $\int_a^b f_-(x)dx$ sont finies, on pose alors $\int_a^b f(x)dx := \int_a^b f_+(x)dx - \int_a^b f_-(x)dx$, où les intégrales sont au sens impropre. Dans les deux cas, on parlera ici de l'intégrale de Riemann de f sur le l'intervalle ouvert $]a,b[$. Notons que la notion d'intégrale imporpre est consistente avec celle d'intégrale définie : l'intégrale sur $]a,b[$ d'une fonction f Riemann-intégrale sur $[a,b]$ est égale à l'intégrale de f sur $[a,b]$. Encore une fois, l'intégrale impropre d'une fonction sur un intervalle ouvert est égale à l'intégrale de cette fonction par la mesure de Lebesgue sur ce même intervalle.

Corollaire 4.2. *Soient $a,b\in\bar{\mathbb{R}}$ tels que $a<b$ et f une fonction réelle sur $\mathbb{R}$ (ou $]a,b[$) telle que $\int_a^b f(x)dx$ est bien définie. Alors $\int_{]a,b[} f d\lambda$ est également bien définie et on a :*

$$\int_{]a,b[} f d\lambda = \int_a^b f(x)dx$$

En particulier, si $a=-\infty$ et $b=+\infty$:

$$\int f d\lambda = \int_{\mathbb{R}} f d\lambda = \int_{-\infty}^{+\infty} f(x)dx$$

Preuve

Supposons d'abord f positive. Pour tout $n > 2/(b-a)$, la proposition 4.1 nous donne :

$$\int_{[a+1/n,b-1/n]} f d\lambda = \int_{a+1/n}^{b-1/n} f(x)dx$$

En passant à la limite, on a donc :

$$\lim_{n\to+\infty} \int_{[a+1/n,b-1/n]} f d\lambda = \lim_{n\to+\infty} \int_{a+1/n}^{b-1/n} f(x)dx = \int_a^b f(x)dx$$

Or, on observe que la suite $\left(\mathbb{1}_{[a+1/n,b-1/n]} f\right)_{n>2/(b-a)}$ est positive, croissante point par point et converge simplement vers $\mathbb{1}_{]a,b[} f$. En appliquant le théorème de convergence monotone, on a alors :

$$\int_{]a,b[} f d\lambda = \int_a^b f(x)dx$$

Supposons maintenant f quelconque à valeurs dans $\mathbb{R}$ telle que $\int_a^b f_+(x)dx$ et $\int_a^b f_-(x)dx$ sont bien définies et finies. Par le point précédent, $\int_{]a,b[} f_+ d\lambda$ et $\int_{]a,b[} f_- d\lambda$ sont également bien définies et finies. C'est-à-dire, $\mathbb{1}_{]a,b[} f$ est λ-intégrable et on a alors :

$$\int_{]a,b[} f d\lambda = \int_{]a,b[} f_+ d\lambda - \int_{]a,b[} f_- d\lambda = \int_a^b f_+(x)dx - \int_a^b f_-(x)dx = \int_a^b f(x)dx$$

Remarque. *De même que les intégrales de Riemann sur $[a,b]$ et $]a,b[$ d'une fonction Riemann-intégrable sur $[a,b]$ sont égales, l'intégrale par Lebesgue d'une fonction intégrable par rapport à Lebesgue sur $]a,b[$ ou $[a,b]$ sont égales. En effet, la relation de Chasles, la diffusivité de λ et la proposition 2.2.5 nous donnent :*

$$\int_{[a,b]} f d\lambda = \int_{]a,b[} f d\lambda + \int_{\{a,b\}} f d\lambda = \int_{]a,b[} f d\lambda$$

Alternativement, on observe que $f\mathbb{1}_{[a,b]}$ et $f\mathbb{1}_{]a,b[}$ sont égales λ-p.p. et utilise immédiatement le corollaire 2.2.15.

Ainsi, toutes les fonctions intégrables au sens de Riemann le sont également pour la mesure de Lebesgue. L'implication réciproque est fausse. En effet, l'intégrale de Lebesgue permet d'intégrer un nombre bien plus important de fonctions. Par exemple, la fonction $\mathbb{1}_{\mathbb{Q}}$, appelée fonction de Dirichlet, a toutes ses sommes supérieures de Riemann égales à 1 et ses sommes inférieures de Riemann égales à 0 sur le segment $[0,1]$. Elle n'est donc pas Riemann-intégrable sur $[0,1]$. Pour autant, $\mathbb{Q}$ étant dénombrable et la mesure de Lebesgue λ diffuse, on a $\lambda(\mathbb{Q}) = 0$ et donc $\mathbb{1}_{\mathbb{Q}}$ est égale $\lambda - p.p.$ à la fonction constante et égale à 0. On en déduit que son intégrale par la mesure de Lebesgue est bien définie et nulle (en particulier $\int_{[0,1]} \mathbb{1}_{\mathbb{Q}} d\lambda = 0$).

E Espaces de Lebesgue

Dans cette annexe on s'intéresse à la structure vectorielle de certains espaces de variables aléatoire. On rappelle d'abord quelques notions d'algèbre linéaire. On se placera ensuite dans le cadre plus général des espaces mesurés avant de donner des applications en théorie des probabilités.

Définition 5.1 (Espace vectoriel réel). *Un espace vectoriel réel est un ensemble E muni de deux lois, $\cdot + \cdot : E^2 \to E$ et $\cdot \times \cdot : \mathbb{R} \times E \to E$, telles que:*

- *$(E,+)$ est un groupe abélien.*
- *$\times$ est associative à droite et à gauche, distributive et telle que $1 \in \mathbb{R}$ est un élément neutre à gauche.*

En mots, un espace vectoriel réel est un ensemble dont on peut additionner les éléments, appelés vecteurs, et les multiplier par des réels (appelés scalaires dans le cas général). La première condition de la définition garantie de plus l'existence d'un élément neutre pour l'addition (souvent noté 0_E). Dans la suite, on appellera simplement espace vectoriel un espace vectoriel réel.

Définition 5.2 (Dimension d'espace vectoriel). *Soit E un espace vectoriel. On dit qu'une famille $F := \{v_i\}_{i\in I}$ d'éléments de E est génératrice si tout vecteur $v \in E$ peut s'écrire comme combinaison linéaire finie d'éléments de F. Une famille génératrice dont les vecteurs sont linéairement indépendants est appelée une base. Si B est une base de E, alors on appelle dimension de E, notée $\dim E$, le cardinal de B. On peut montrer que cette définition ne dépend pas de la base choisie.*

Les espaces euclidiens ($\mathbb{R}^n$ avec $n \in \mathbb{N}^*$) sont l'exemple le plus élémentaire d'espaces vectoriels. Pour $n \in \mathbb{N}^*$, il est facile de voir que $\dim \mathbb{R}^n = n$. L'espace $\mathbb{R}^\Omega$ des fonctions d'un ensemble Ω dans $\mathbb{R}$ est également un espace vectoriel, lorsque muni de l'addition point par point, où l'élément neutre pour l'addition est la fonction constante est égale à 0. Si Ω est un ensemble infini, alors la dimension de $\mathbb{R}^\Omega$ est nécessairement infinie. Inversement, si Ω est fini, alors $\mathbb{R}^\Omega$ est identifiable à $\mathbb{R}^{|\Omega|}$. En analyse, il est souvent très utile d'avoir une notion de taille des vecteurs. On introduit alors une norme sur l'espace vectoriel.

Définition 5.3 (Norme). *Une norme est une application $\|\cdot\|_E$ d'espace vectoriel E dans $\mathbb{R}_+$ telle que:*

- *Pour tout $x \in E$, si $\|x\|_E = 0$ alors $x = 0_E$ (séparation).*
- *Pour tout $(\alpha, x) \in \mathbb{R} \times E$, $\|\alpha x\|_E = |\alpha| \|x\|_E$ (homogénéité).*
- *Pour tous $(x, y) \in E^2$, $\|x + y\|_E \leq \|x\|_E + \|y\|_E$ (sous-additivité).*

Sur $\mathbb{R}^n$, on peut définir naturellement toute une famille de norme, indexée par $[1, +\infty[$. Pour $x := (x_1, \dots, x_n) \in \mathbb{R}^n$, on définit ainsi la norme "p" $\|x\|_p$ de x par:

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

En dimension finie, toutes les normes sont équivalentes. C'est-à-dire que, pour deux normes $\|\cdot\|$ et $\|\cdot\|'$ sur $\mathbb{R}^n$, il existe $m, M \in \mathbb{R}_+^*$ tels que:

$$m\|\cdot\|' \leq \|\cdot\| \leq M\|\cdot\|'$$

Parmi ces normes, la norme 2, dite euclidienne, tient une place particulière car elle peut être dérivée d'un produit scalaire.

Définition 5.4 (Produit scalaire). *Soit E un espace vectoriel réel. On appelle produit scalaire sur E une application $\langle \cdot, \cdot \rangle_E : E^2 \to \mathbb{R}$ telle que:*

- *$\langle \cdot, \cdot \rangle_E$ est bilinéaire symétrique.*
- *$\langle \cdot, \cdot \rangle_E$ est définie, i.e. pour tout $x \in E$, si $\langle x, x \rangle_E = 0$ alors $x = 0_E$.*
- *$\langle \cdot, \cdot \rangle_E$ est positive, i.e. pour tout $x \in E$, $\langle x, x \rangle_E \geq 0$.*

On vérifie facilement qu'étant donné un produit scalaire $\langle \cdot, \cdot \rangle_E$ sur un espace vectoriel E, l'application $E \to \mathbb{R}_+, x \mapsto \sqrt{\langle x, x \rangle_E}$ est une norme sur E. De plus, pour $x, y \in E$, on a l'inégalité suivante, dite "de Cauchy-Schwartz":

$$|\langle x, y \rangle_E| \leq \|x\|_E \|y\|_E$$

Le produit scalaire "canonique" sur $\mathbb{R}^n$ est défini, pour $x := (x_1, \dots, x_n), y := (y_1, \dots, y_n) \in \mathbb{R}^n$, par:

$$\langle x, y \rangle_{\mathbb{R}^n} := \sum_{i=1}^{n} x_i y_i$$

Trivialement, on a $\langle x, x \rangle_{\mathbb{R}^n} = \|x\|_2^2$. Le produit scalaire permet d'introduire la notion d'orthogonalité sur un espace vectoriel.

Définition 5.5 (Orthogonalité). *Soit E un espace vectoriel muni d'un produit scalaire $\langle \cdot, \cdot \rangle$. Deux vecteurs $v, w \in E$ sont dits orthogonaux, noté $v \perp w$, si $\langle v, w \rangle = 0$. Si v est orthogonal à tous les vecteurs d'un sous-espace $W \subset E$, alors on dit que v est orthogonal à W, noté $v \perp W$. Enfin deux sous-espaces $V, W \subset E$ sont dits orthogonaux, noté $V \perp W$, si pour tous $v \in V$ et $w \in W$, $v \perp w$.*

On va voir qu'il est possible d'étendre les normes "p" et le produit scalaire canonique de $\mathbb{R}^n$ à des espaces de dimension infinie. On rappelle qu'une somme peut être vue comme une intégrale par une mesure de comptage. Le cadre "naturel" pour étendre la structure des espaces euclidiens est en fait celui des fonctions mesurables réelles sur un espace mesuré. Soit $(\mathbb{X}, \mathscr{F})$ un espace mesurable. Dans le cours, on a vu que l'ensemble $\mathscr{M}(\mathbb{X}, \mathbb{R})$ des fonctions mesurables réelles sur $(\mathbb{X}, \mathscr{F})$ est un sous-espace vectoriel de $\mathbb{R}^{\mathbb{X}}$. Munissons $(\mathbb{X}, \mathscr{F})$ d'une mesure μ. Pour $p \in [1, +\infty[$, on définit l'ensemble suivant:

$$\mathscr{L}^p(\mathbb{X}, \mu) := \left\{ f \in \mathscr{M}(\mathbb{X}, \mathbb{R}) : \int |f|^p d\mu < +\infty \right\}$$

En mots, $\mathscr{L}^p(\mathbb{X}, \mu)$ est l'ensemble des fonctions mesurables réelles f sur $(\mathbb{X}, \mathscr{F})$ telles que $|f|^p$ est μ-intégrable. On peut montrer que c'est en fait un sous-espace vectoriel de $\mathscr{M}(\mathbb{X}, \mathbb{R})$.

Proposition 5.6 (Inégalité de Minkowski). *Soient $p \in [1, +\infty[$ et $f, g \in \mathscr{L}^p(\mathbb{X}, \mu)$. Alors:*

$$\left(\int |f+g|^p d\mu\right)^{\frac{1}{p}} \leq \left(\int |f|^p d\mu\right)^{\frac{1}{p}} + \left(\int |g|^p d\mu\right)^{\frac{1}{p}}$$

Preuve

Dans le cas $p = 1$, il s'agit simplement de l'inégalité triangulaire (pour la valeur absolue). Supposons donc $p > 1$. Par convexité de la fonction $\mathbb{R} \to \mathbb{R}, x \mapsto |x|^p$, on a:

$$|f+g|^p = 2^p \left|\frac{1}{2}f + \frac{1}{2}g\right|^p \leq 2^{p-1}(|f|^p + |g|^p)$$

Maintenant:

$$\int |f+g|^p d\mu = \int |f+g||f+g|^{p-1} d\mu \leq \int (|f|+|g|)|f+g|^{p-1} d\mu = \int |f||f+g|^{p-1} d\mu + \int |g||f+g|^{p-1} d\mu$$

En appliquant l'inégalité de Hölder (on admettra qu'elle reste valide pour des mesures qui ne sont pas des probabilités, la preuve étant identique), on obtient:

$$\begin{aligned}\int |f+g|^p d\mu &\leq \left(\left(\int |f|^p d\mu\right)^{\frac{1}{p}} + \left(\int |g|^p d\mu\right)^{\frac{1}{p}}\right)\left(\int |f+g|^{(p-1)\frac{p}{p-1}}\right)^{\frac{p-1}{p}} \\ &= \left(\left(\int |f|^p d\mu\right)^{\frac{1}{p}} + \left(\int |g|^p d\mu\right)^{\frac{1}{p}}\right)\left(\int |f+g|^p d\mu\right)^{1-\frac{1}{p}}\end{aligned}$$

C'est-à-dire:

$$\left(\int |f+g|^p d\mu\right)^{\frac{1}{p}} \leq \left(\int |f|^p d\mu\right)^{\frac{1}{p}} + \left(\int |g|^p d\mu\right)^{\frac{1}{p}}$$

De l'inégalité de Minkowski, on déduit que, si $f, g \in \mathscr{L}^p(\mathbb{X}, \mu)$ et $\alpha \in \mathbb{R}$, alors $f + \alpha g \in \mathscr{L}^p(\mathbb{X}, \mu)$. Il est donc clair que $\mathscr{L}^p(\mathbb{X}, \mu)$ est un sous-espace vectoriel de $\mathscr{M}(\mathbb{X}, \mathbb{R})$. En fait, l'application

$$\mathscr{L}^p(\mathbb{X}, \mu) \to \mathbb{R}_+, f \mapsto \left(\int |f|^p d\mu\right)^{\frac{1}{p}}$$

définit presque une norme sur $\mathscr{L}^p(\mathbb{X}, \mu)$. La seule condition qui n'est pas vérifiée est la séparation des points (on parle de "pseudonorme"). En effet, une fonction mesurable d'image nulle pour l'application ci-dessus est seulement nulle μ-presque partout. Ainsi, deux fonctions dont la différence est d'image nulle pour cette même applications sont seulement égales μ-presque partout. On rappelle qu'en pratique, on ne s'intéresse uniquement à ce qui se passe sur des ensembles non négligeables. Ainsi, deux fonctions mesurables égales μ-presque partout seront considérées comme fonctionnellement indistinguables lorsqu'on intègre par μ. En particulier, lorsque μ est une probabilité, on obtient des variables aléatoires égales presque sûrement. Une manière de contourner ce problème de séparation est alors de prendre le quotient $L^p(\mathbb{X}, \mu)$ de $\mathscr{L}^p(\mathbb{X}, \mu)$ par la relation d'équivalence $f \sim g$ si $f = g$ μ-presque partout. Celui-ci est défini rigoureusement de la sorte:

$$L^p(\Omega, \mu) := \{\{g \in \mathscr{L}^p(\Omega, \mu) : g = f \ \mu\text{-p.p.}\} : f \in \mathscr{L}^p(\Omega, \mu)\}$$

Un élément de $\mathscr{L}^p(\mathbb{X}, \mu)$ est donc un ensemble (appelé classe) de fonction mesurables égales μ-p.p.. Pour deux fonctions f, g dans une même classe, on a, pour toute fonction $F : \mathbb{R} \times \mathbb{X} \to \mathbb{R}$ mesurable:

$$\int_{\mathbb{X}} F(f(x), x)\mu(dx) = \int_{\{x \in \mathbb{X} : f(x) = g(x)\}} F(f(x), x)\mu(dx) = \int_{\{x \in \mathbb{X} : f(x) = g(x)\}} F(g(x), x)\mu(dx) = \int_{\mathbb{X}} F(g(x), x)\mu(dx)$$

Avec un léger abus de notation, pour $f \in \mathscr{L}^p(\mathbb{X}, \mu)$, on note également f sa classe d'équivalence (i.e. l'ensemble $E \in L^p(\mathbb{X}, \mathbf{P})$ tel que $f \in E$) dans $L^p(\mathbb{X}, \mathbf{P})$. Cet usage ne pose pas de problème car, par l'observation ci-dessus, tout calcul intégral fait avec f peut être fait en substituant f par un autre élément de sa classe. Ainsi, on peut maintenant définir la norme $\|\|_{L^p}$ sur $L^p(\mathbb{X}, \mu)$, définit, pour $f \in L^p(\mathbb{X}, \mu)$, par:

$$\|f\|_{L^p} := \left(\int |f|^p d\mu\right)^{\frac{1}{p}}$$

Les espaces vectoriels normés $(L^p(\mathbb{X}, \mu), \|\cdot\|_{L^p})$ (pour $p \in [1, +\infty[$) sont appelés espaces de Lebesgue. Lorsque $p = 2$, ils sont une généralisation des espaces euclidiens en dimension infinie. En particulier, si $(\mathbb{X}, \mathscr{F}) = (\llbracket n \rrbracket, \mathscr{P}(\llbracket n \rrbracket))$ (pour

$n \in \mathbb{N}^*$) et que μ est la mesure de comptage, alors $(L^p([\![n]\!], \mu), \|\cdot\|_{L^p})$ peut être identifié à $(\mathbb{R}^n, \|\cdot\|_p)$. Tout comme en dimension 1, le cas $p = 2$ est particulier car il permet de définir un produit scalaire. On vérifie en effet facilement que l'application suivante est un produit scalaire sur $L^2(\mathbb{X}, \mu)$ qui engendre $\|\cdot\|_{L^2}$:

$$L^2(\mathbb{X}, \mu) \times L^2(\mathbb{X}, \mu) \to \mathbb{R}$$
$$(f, g) \mapsto (f, g)_{L^2} := \int f g d\mu$$

L'application est bien définie par l'inégalité de Hölder. De plus, par linéarité de l'intégrale, elle est bilinéaire et, pour $f \in L^2(\mathbb{X}, \mu)$, $(f, f)_{L^2} = \|f\|_{L^2}^2$. On rappelle que les espaces euclidiens sont complets. On va voir que c'est également le cas des espaces de Lebesgue. On rappelle d'abord les définitions.

Définition 5.7 (Espace métrique). *On appelle espace métrique une paire (E, d) où E est un ensemble et d une application de E^2 dans $\mathbb{R}$ telle que:*

- *Pour tous $x, y \in E$, $d(x, y) = 0$ si et seulement si $x = y$ (séparation).*
- *Pour tous $x, y \in E$, $d(x, y) = d(y, x)$ (symétrie).*
- *Pour tous $x, y, z \in E$, $d(x, z) \leq d(x, y) + d(y, z)$ (inégalité triangulaire).*

L'application d est appelée une distance.

Une suite $(x_n)_{n\in\mathbb{N}}$ à valeurs dans un espace métrique (E, d) converge vers un point $x \in E$ si $\lim_{n\to+\infty} d(x_n, x) = 0$. Par la propriété de séparation et l'inégalité triangulaire, la limite d'une suite, si elle existe, est unique. On vérifie facilement qu'un espace vectoriel normé $(E, \|\cdot\|)$ définit un espace métrique en prenant comme distance $d : E^2 \to \mathbb{R}_+, (x, y) \mapsto \|x - y\|$. Ainsi, une suite $(x_n)_{n\in\mathbb{N}}$ à valeur dans un tel espace E converge vers $x \in E$ si $\lim_{n\to+\infty} \|x_n - x\| = 0$.

Définition 5.8 (Suite de Cauchy). *Soit (E, d) un espace métrique et $(x_n)_{n\in\mathbb{N}}$ une suite à valeurs dans E. $(x_n)_{n\in\mathbb{N}}$ est dite de Cauchy si, pour tout $\varepsilon > 0$, il existe $N \in \mathbb{N}$ tel que, pour tous $m, n \geq N$, $d(x_n, x_m) < \varepsilon$.*

En mots, une suite de Cauchy est une suite dont les éléments se rapprochent de plus en plus.

Définition 5.9 (Complétude). *Un espace métrique $(E, \|\cdot\|)$ est dit complet (et sa distance complète) si toutes ses suites de Cauchy convergent.*

Un espace complet est donc un espace où, pour toute suite dont les points se rapprochent de plus en plus, il existe un point dans l'espace qui est la limite de cette suite. L'intervale $]0, 1]$ muni de la distance euclidienne, par exemple, n'est pas complet. En effet, la suite $(1/n)_{n\in\mathbb{N}^*}$ est de Cauchy mais n'admet aucune limite dans $]0, 1]$. Comme on l'a déjà mentionné, les espaces vectoriels normés de dimensions finies sont complets (en tant qu'espaces métriques). De manière générale, ce n'est plus vrai en dimension infinie. On introduit alors une terminologie particulière pour les espaces vectoriels normés qui satisfont toujours cette propriété.

Définition 5.10 (Espaces de Banach et Hilbert). *Un espace vectoriel normé est dit "de Banach" si il est complet. Si, de plus, sa norme est induite par un produit scalaire, alors on parle plus précisément d'espace "de Hilbert".*

Remarque. *Par abréviation, on appelle souvent un espace de Banach (resp. de Hilbert) simplement "un Banach" (resp. "un Hilbert"). En ce qui concerne les espaces de Hilbert, il est aussi fréquent de trouver l'adjectif hilbertien. Dans ce contexte, les espaces vectoriels munis de produits scalaires (pas nécessairement complets) sont dits "préhilbertiens".*

Théorème 5.11. *Les espaces de Lebesgue sont des espaces de Banach. En particulier, $L^2(\mathbb{X}, \mu)$ est hilbertien.*

Preuve

Admis.

Les espaces de Banach et de Hilbert sont fondamentaux en analyse. En effet, la complétude est propriété qui est nécessaire à de nombreux résultats. Parmis eux, on peut noter l'existence de projections orthogonales. En mots, la projection orthogonale d'un vecteur v d'un espace vectoriel normé E sur un sous-espace W est l'unique vecteur de W qui minimise la distance à v. Autrement dit, c'est le vecteur de W le plus proche de v. En particulier, la projection orthogonal sur W d'un vecteur $w \in W$ est lui même.

Définition 5.12 (Projection orthogonale). *Soit $(E, \langle\cdot,\cdot\rangle)$ un espace préhilbertien de norme $\|\cdot\|$, $v \in E$ et W un sous-espace de E. On dit qu'un vecteur $h \in W$ est une projection orthogonale de v sur W si il minimise la distance à v. Mathématiquement, pour tout $w \in W$:*

$$\|v - h\| \leq \|v - w\|$$

Proposition 5.13. *Soit $(E, \langle\cdot,\cdot\rangle)$ un espace préhilbertien, $v \in E$ et W un sous-espace de E et $h \in W$. Alors h est une projection orthogonale de v sur W si et seulement si $v - h \perp W$. De plus, si h est une projection orthogonale de v sur W, alors c'est l'unique.*

Lemme 5.14 (Théorème de Pythagore)**.** *Soit $(E, \langle\cdot,\cdot\rangle)$ un espace préhilbertien de norme $\|\cdot\|$ et $v, w \in E$ tels que $v \perp w$. Alors:*

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2$$

Preuve

[Preuve du lemme 5.14] On a:

$$\|v + w\|^2 = \langle v + w, v + w\rangle = \langle v, v\rangle + \langle w, v\rangle + \langle v, w\rangle + \langle w, w\rangle = \|v\|^2 + \|w\|^2 + 2\langle v, w\rangle = \|v\|^2 + \|w\|^2$$

où la dernière égalité est car $v \perp w$ et donc $\langle v, w\rangle = 0$, par définition.

Preuve

[Preuve de la proposition 5.13] Supposons d'abord que h est une projection orthogonale de v sur W. Soient $w \in W$ et $t \in \mathbb{R}$. Puisque $h + tw \in W$ et par définition d'une projection orthogonale, on a:

$$\|v - h\| \leq \|v - (h + tw)\|$$

C'est-à-dire:

$$\|v - (h + tw)\| - \|v - h\| = 2t\langle h - v, w\rangle + t^2\|w\|^2 \geq 0$$

Si $\langle h - v, w\rangle \neq 0$, alors en prenant $t := -\frac{\langle h-v,w\rangle}{\|w\|^2}$, on a $2t\langle h - v, w\rangle + t^2\|w\|^2 < 0$, ce qui est impossible. On en déduit que nécessairement $\langle h - v, w\rangle = 0$. Comme w est pris arbitrairement dans W, on a $v - h \perp W$. Supposons maintenant seulement que $h - v \perp W$ et prenons $w \in W$. On a $h - w \in W$ et donc $v - h \perp h - w$. Par le théorème de Pythagore, on obtient:

$$\|v - w\|^2 = \|v - h\|^2 + \|h - w\|^2$$

En particulier, $\|v - h\|^2 \leq \|v - w\|^2$. Donc h est une projection orthogonale de v sur W. Par la meme égalité, on voit que $\|v - h\| = \|v - w\|$ si et seulement si $\|h - w\| = 0$, c'est-à-dire si $h = w$ (par propriété de séparation de la norme), ce qui prouve l'unicité.

Comme nous l'explique la proposition précédente, la projection orthogonale d'un vecteur v sur un sous-espace W, si elle existe, est unique. On peut donc la définir comme fonction de v et W et on la note généralement $\pi_W(v)$. On va maintenant des hypothèses suffisantes pour l'existence, c'est-à-dire pour la bonne défintion de $\pi_W(v)$.

Définition 5.15 (Sous-espace fermé)**.** *Soit $(E, \|\cdot\|)$ un espace vectoriel normé. On dit qu'un sous-espace vectoriel $V \subset E$ est fermé si il est fermé dans E pour la topologie induite par $\|\cdot\|$. C'est-à-dire, toute suite à valeurs dans V qui converge dans E a sa limite dans V.*

En dimension finie, tous les sous-espaces vectoriels sont trivialement fermés. En particulier, les sous-espaces vectoriels de dimension finie d'un espace vectoriel normé de dimension quelconque sont fermés. En dimension infinie cette propriété n'est plus vraie. Par l'exemple, l'ensemble des polynômes sur $[0, 1]$ est un sous-espace de $L^2([0, 1], \lambda)$, où λ est la mesure de Lebesgue. On peut montrer que toute fonction $f \in L^2([0, 1], \lambda)$ peut être approximée par une suite de polynômes $(f_n)_{n\in\mathbb{N}}$ pour la norme $\|\cdot\|_{L^2}$, c'est-à-dire $\lim_{n\to+\infty} \|f_n - f\|_{L^2} = 0$. Puisque $L^2([0, 1], \lambda)$ contient évidemment d'autres fonctions que des polynômes, ceux-ci ne forment pas un sous-espace fermé.

Théorème 5.16. *Soit $(H, \langle\cdot,\cdot\rangle)$ un espace de Hilbert et W un sous-espace fermé de H. L'application suivante est bien définie:*

$$\begin{aligned} \pi_W : H &\to W \\ v &\mapsto \pi_W(v) \end{aligned}$$

Preuve

Soit $v \in H$. On a déjà montré que si une projection orthogonale de v sur W existe, alors elle est unique. Montrons donc qu'elle existe. Soit $\delta := \inf_{w\in W} \|x - w\|$ la distance de x à W. Par définition de l'infimum, il existe un suite $(w_k)_{k\in\mathbb{N}^*}$ de vecteurs de W tels que, pour $k \in \mathbb{N}^*$:

$$\|v - w_k\| \leq \delta + \frac{1}{k}$$

Soient $m, n \in \mathbb{N}^*$. On a:

$$\|w_m - w_n\|^2 = \|(w_m - x) + (x - w_n)\|^2 = \|w_m - x\|^2 + \|x - w_n\|^2 + 2\langle w_m - x, x - w_n\rangle$$

Par hypothèse sur $(w_k)_{k\in\mathbb{N}^*}$, on a donc:

$$\|w_m - w_n\|^2 \leq \frac{1}{m^2} + \frac{1}{n^2} + 2\langle w_m - x, x - w_n\rangle$$

L'inégaltié de Cauchy-Schwartz et la même hypothèse nous donnent ensuite:

$$\|w_m - w_n\|^2 \leq \frac{1}{m} + \frac{1}{n} + 2\|w_m - w\|\|x - w_n\| \leq \frac{1}{m^2} + \frac{1}{n^2} + 2\frac{1}{m}\frac{1}{n}$$

En particulier, on observe que la suite $(w_k)_{k\in\mathbb{N}}$ est de Cauchy. Comme H est complet, elle admet une limite $w \in H$. De plus, comme W est fermé, $w \in W$. Il est ensuite clair que $\|v - w\| = \delta$, c'est-à-dire w est l'unique projection orthogonale de v sur W. En effet, pour $k \in \mathbb{N}^*$:

$$\delta \leq \|v - w\| = \|v - w_k + w_k - w\| \leq \|v - w_k\| + \|w_k - w\| \leq \delta + \frac{1}{k} + \|w_k - w\|$$

En passant à la limite, on obtient le résultat souhaité.

Voyons maintenant une application à la théorie des probabilités. Dans notre contexte $(\mathbb{X}, \mathscr{F}, \mu) = (\Omega, \mathscr{A}, \mathbf{P})$. Pour $p \in [1, +\infty[$, l'espace $L^p(\Omega, \mathbf{P})$ est alors l'ensembles des variables aléatoires réelles X sur $(\Omega, \mathscr{A}, \mathbf{P})$ telles que $\|X\|_{L^p} := \int |X|^p d\mathbf{P} = \mathbf{E}[|X|^p] < +\infty$, où on a identifié les variables aléatoires égales presque sûrement. Par l'exercice 2.3.5, on a $L^p(\Omega, \mathbf{P}) \subset L^q(\Omega, \mathbf{P})$ pour $p, q \in [1, +\infty[$ tels que $p \leq q$. Dans ce contexte, l'espace $L^2(\Omega, \mathbf{P})$ correspond aux variables aléatoires de carrés intégrables (ou de manière équivalente, de variances finies). Soient $X, Y \in L^2(\Omega, \mathbf{P})$ tels que X est de variance non nulle (donc non presque sûrement constante). Le produit scalaire de X et Y est donné par $(X, Y)_{L^2} := \int XY d\mathbf{P} = \mathbf{E}[XY]$ et leur distance par $\|X - Y\|_{L^2} := \int (X - Y)^2 d\mathbf{P} = \mathbf{E}[(X - Y)^2]$. En particulier, on reconnaît la distance quadratique définie dans la section 3.2. Dans cette même section, on a défini la régression linéaire $\hat{Y}$ de Y par X comme la meilleure approximation affine de Y par X pour la distance quadratique. Autrement dit, $\hat{Y} = aX + b$ où

$$(a, b) := \underset{(x,y)\in\mathbb{R}^2}{\arg\min} \|Y - (xX + y)\|_{L^2}$$

On a vu que a et b peuvent être déterminés par un problème d'optimisation convexe. En passant par les espaces de Lebesgue, on peut en fait se réduire à un problème algébrique beaucoup plus simple en termes d'hypothèses à vérifier. Remarquons que l'ensemble $W := \{xX + y : (x, y) \in \mathbb{R}^2\} = Span(X, 1)$ est un sous-espace vectoriel de dimension finie de $L^2(\Omega, \mathbf{P})$ (X et 1 ne sont colinéaires que si X est presque sûrement constante, ce qu'on exclut par hypothèse, et donc $\dim(W) = 2$). En particulier, il est fermé. De plus, par définition, $\hat{Y}$ est la projection orthogonale de Y sur W, formellement $\hat{Y} := \pi_W(Y)$. Or, $L^2(\Omega, \mathbf{P})$ est un espace de Hilbert. Donc, par le théorème 5.16, $\hat{Y}$ est bien défini et est l'unique vecteur de W qui satisfait $Y - \hat{Y} \perp W$, c'est-à-dire:

$$\begin{aligned}
&\forall\, Z \in W, \quad \mathbf{E}[(Y - \hat{Y})Z] = 0 \\
\iff &\forall\, (x, y) \in \mathbb{R}^2, \quad \mathbf{E}[(Y - (aX + b))(xX + y)] = 0 \\
\iff &\forall\, (x, y) \in \mathbb{R}^2, \quad x\mathbf{E}[YX] + y\mathbf{E}[Y] - ax\mathbf{E}[X^2] - bx\mathbf{E}[X] + ay\mathbf{E}[X] + by = 0 \\
\iff &\forall\, (x, y) \in \mathbb{R}^2, \quad x\big(\mathbf{E}[YX] - a\mathbf{E}[X^2] - b\mathbf{E}[X]\big) + y(\mathbf{E}[Y] + a\mathbf{E}[X] + b) = 0 \\
\iff &\begin{cases} \mathbf{E}[YX] - a\mathbf{E}[X^2] - b\mathbf{E}[X] & = 0 \\ \mathbf{E}[Y] + a\mathbf{E}[X] + b & = 0 \end{cases}
\end{aligned}$$

En résolvant le système linéaire à deux équations et deux inconnues, on retrouve bien:

$$(a, b) = \left(\frac{Cov(X, Y)}{Var(X)}, \mathbf{E}[Y] - \frac{Cov(X, Y)}{Var(X)}\mathbf{E}[X]\right)$$

Cette technique simple et efficace nous permet également de calculer des régressions linéaires multivariées. Soient $n \in \mathbb{N}^*$ et $(X_i)_{i\in[\![n]\!]}$ une suite libre (au sens linéaire) de variables aléatoires et de variances strictement positives. On cherche à donner une approximation affine $\hat{Y} := a_1X_1 + \cdots + a_nX_n + b$ de Y par la famille $(X_i)_{i\in[\![n]\!]}$. En considérant $a := (a_1, \ldots, a_n)$ et $X := (X_1, \ldots, X_n)$ comme des vecteurs colonnes, on a $Y = a^*X + b$. On pose alors

$W \coloneqq Span(X_1, \dots, X_n, 1)$ et le théorème 5.16 nous donne:

$$
\begin{aligned}
& Y - \hat{Y} \perp W \\
\iff & \forall\, Z \in W, \quad \mathbf{E}\Big[(Y - \hat{Y})Z\Big] = 0 \\
\iff & \forall\, (x_1, \dots, x_n, y) \in \mathbb{R}^{n+1}, \quad \mathbf{E}\left[\left(Y - b - \sum_{i=1}^{n} a_i X_i\right)\left(y + \sum_{j=1}^{n} x_j X_j\right)\right] \\
\iff & \forall\, (x_1, \dots, x_n, y) \in \mathbb{R}^{n+1}, \quad y\mathbf{E}[Y] - by - \sum_{i=1}^{n} a_i y \mathbf{E}[X_i] + \sum_{j=1}^{n} x_j \mathbf{E}[Y X_j] - \sum_{j=1}^{n} b x_j \mathbf{E}[X_j] - \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \mathbf{E}[X_i X_j] = 0 \\
\iff & \forall\, (x_1, \dots, x_n, y) \in \mathbb{R}^{n+1}, \quad y\left(\mathbf{E}[Y] - b - \sum_{i=1}^{n} a_i \mathbf{E}[X_i]\right) + \sum_{j=1}^{n} x_j \left(\mathbf{E}[Y X_j] - b\mathbf{E}[X_j] - \sum_{i=1}^{n} a_i \mathbf{E}[X_i X_j]\right) = 0
\end{aligned}
$$

On obtient donc le sytème linéaire suivant:

$$
\begin{aligned}
& \begin{cases} \mathbf{E}[Y] - b - \sum_{i=1}^{n} a_i \mathbf{E}[X_i] & = 0 \\ \mathbf{E}[Y X_1] - b\mathbf{E}[X_1] - \sum_{i=1}^{n} a_i \mathbf{E}[X_i X_1] & = 0 \\ & \vdots \\ \mathbf{E}[Y X_n] - b\mathbf{E}[X_n] - \sum_{i=1}^{n} a_i \mathbf{E}[X_i X_n] & = 0 \end{cases} \\
\iff & \begin{cases} b & = \mathbf{E}[Y] - \sum_{i=1}^{n} a_i \mathbf{E}[X_i] \\ Cov(Y, X_1) & = \sum_{i=1}^{n} a_i Cov(X_i, X_1) \\ & \vdots \\ Cov(Y, X_n) & = \sum_{i=1}^{n} a_i Cov(X_i, X_n) \end{cases} \\
\iff & \begin{cases} b & = \mathbf{E}[Y] - a^* \mathbf{E}[X] \\ Cov(Y, X) & = a^* D_X \end{cases}
\end{aligned}
$$

où $Cov(Y, X) \coloneqq (Cov(Y, X_i))_{i \in [\![n]\!]} \in \mathbb{R}^n$ (vu comme un vecteur ligne) et D_X est la matrice de dispersion du vecteur aléatoire X. En particulier, on a $Cov(Y, X)^* = D_X^* a$. Par la proposition ???, D_X est définie-positive et symétrique, donc inversible d'inverse symétrique. On a donc $Cov(Y, X)^* = D_X a$ et, en calculant l'inverse D_X^{-1} de D_X, on obtient $a = D_X^{-1} Cov(Y, X)^*$. $\hat{Y}$ est donc donné par:

$$\hat{Y} = \mathbf{E}[Y] + Cov(Y, X) D_X^{-1} (X - \mathbf{E}[X])$$

On peut remarquer que, si $n = 1$, on retrouve la formule donnée en section 3.2

F Un exemple d'ensemble de $\mathbb{R}$ non mesurable : autour de l'ensemble de Cantor

Lorsqu'on introduit la notion de mesure (en particulier la mesure de Lebesgue), on aimerait pouvoir mesurer tous les ensembles de $\mathbb{R}$ de manière intuitive : longueur pour les intervalles, aire pour des parties de $\mathbb{R}^2$, etc. Pourtant, une découverte marquante du début du XXe siècle montre que ce n'est pas possible : il existe des ensembles dits **non mesurables**.

Un détour par l'ensemble de Cantor Un objet fascinant et assez célèbre est l'**ensemble de Cantor**, introduit par Georg Cantor à la fin du XIXe siècle. On le construit à partir de l'intervalle $[0, 1]$ en retirant successivement le tiers central :

- Étape 1 : on enlève $(\frac{1}{3}, \frac{2}{3})$ et on garde deux intervalles $[0, \frac{1}{3}]$ et $[\frac{2}{3}, 1]$.
- Étape 2 : à chaque intervalle restant, on enlève à nouveau son tiers central, etc.
- En continuant ainsi à l'infini, on obtient un ensemble fermé, parfait, sans intérieur, appelé **ensemble de Cantor**.

Malgré le fait qu'on enlève à chaque fois des morceaux de plus en plus nombreux, on peut montrer que la somme des longueurs retirées est 1, de sorte que l'ensemble final a *mesure nulle*. Autrement dit, le Cantor est "énorme" du point de vue de la cardinalité (il est équipotent à $\mathbb{R}$!) mais "petit" du point de vue de la mesure.

Image 1: Construction de l'ensemble de Cantor
Source: `https://fr.wikipedia.org/wiki/Ensemble_de_Cantor`.

De Cantor à Vitali : l'existence d'ensembles non mesurables La véritable surprise est venue en 1905, quand Giuseppe Vitali a démontré, à partir de l'axiome du choix, l'existence d'ensembles de réels **non mesurables**. L'idée (schématiquement) est de prendre une classe de représentants des réels modulo $\mathbb{Q}$ dans $[0, 1]$. Autrement dit, on choisit un seul représentant par classe d'équivalence pour la relation $x \sim y \iff x - y \in \mathbb{Q}$. Cet ensemble, appelé **ensemble de Vitali**, n'est pas mesurable au sens de Lebesgue.

Ce résultat a été vécu comme un petit choc : on découvrait que l'axiome du choix, introduit par Zermelo en 1904, impliquait l'existence d'ensembles « pathologiques » qui défiaient l'intuition géométrique.

Cantor, Lebesgue et la modernité des mathématiques Historiquement, Georg Cantor (1845–1918) est surtout connu pour avoir introduit la notion de cardinal infini et la diagonale qui prouve que $\mathbb{R}$ est non dénombrable. Henri Lebesgue (1875–1941), quant à lui, a développé la théorie de l'intégrale qui porte son nom, justement pour pallier les limites de l'intégrale de Riemann. Vitali, en construisant son ensemble non mesurable, a mis en évidence la tension entre ces nouvelles notions.

Aujourd'hui, l'ensemble de Cantor reste un exemple incontournable : il illustre la coexistence paradoxale entre structure (fermé, parfait, non dénombrable) et petitesse (mesure nulle). L'ensemble de Vitali, lui, illustre le fait que la mesure de Lebesgue ne peut pas être définie sur *tous* les sous-ensembles de $\mathbb{R}$ sans restrictions.

Remarque. *Il existe d'autres exemples célèbres d'ensembles non mesurables, comme les ensembles de Bernstein ou les constructions liées au paradoxe de Banach–Tarski (où une boule peut être « découpée » et réassemblée en deux boules de même volume !). Ces paradoxes, aussi étranges soient-ils, sont les enfants naturels de l'axiome du choix.*