CHAPITRES D'INTÉGRATION ET DE PROBABILITÉS

J. Féjoz
Université Paris-Dauphine
jacques.fejoz@dauphine.fr

2014

Avertissement

Nous avons pris le parti de développer de front les points de vue "intégrale de Lebesgue" et "espérance". Aussi, l'accent dans ce cours est mis sur les exercices, qui pourront être tantôt traités en cours ou en petite classe, ou laissés à la recherche personnelle. Des textes alternatifs sont : les livres de Rudin [Rud76, Rud80, Rud87], de Briane-Pagès [BP06], de Sinaï [KS07] ou de Jacod-Protter [JP03]. Des références plus avancées sont [Bil95, BJ06, Rud76, Rud80, Sko05, Tao11].

Le cours paraîtra inévitablement abstrait sans un important travail d'appropriation, qui passe par la lecture et la relecture du cours, ainsi que la résolution des exercices. Il faut savoir redire précisément les définitions et les énoncés des propositions et théorèmes, ainsi qu'avoir au moins une idée des démonstrations, sans quoi l'on ne peut pas prétendre comprendre les énoncés, ni aller plus loin dans des cours ultérieurs, même plus "concrets".

Un glossaire à la fin rappelle quelques notions nécessaires et signalées dans le corps du texte avec un astérisque *. Par ailleurs, les chapitres et les paragraphes les plus avancés, qui peuvent être omis en première lecture, sont marqués d'une étoile ★.

Les exercices sont le plus souvent des applications immédiates du cours. Il faut aussi s'entraîner à faire *certains des exercices plus ambitieux, comportant plusieurs questions interdépendantes qui ne sont pas directement reliées au chapitre en cours* ; c'est le seul test véritable pour voir si l'on a compris. Les solutions des exercices ne sont pas incluses dans le polycopié parce qu'il est beaucoup plus important de chercher à résoudre les exercices, même sans y parvenir complètement, que de voir une solution imaginée par quelqu'un d'autre. Des exercices inclus uniquement dans la version électronique de ce cours, souvent plus longs, sont conventionnellement appelés *problèmes*.

L'examen a pour objectif de vérifier cette capacité à résoudre des exercices simples mais variés (qui ne sont donc pas tous dans le poly !), et s'assurer que des cours plus poussés seront profitables. Essayer de s'y préparer directement en apprenant rapidement un petit nombre de résultats clef, ou en apprenant à résoudre un petit nombre d'exercices type, serait une perversion de l'idée d'examen.

Le jeu en est à la fois intéressant et difficile. J'espère que le lecteur y trouvera plaisir... Bon travail !

J. F.

Remerciements

J'ai copieusement utilisé des cours polycopiés existants : de H. Doss [Dos11] et de F. Le Gall [LG06], un recueil d'exercices de l'Université P. & M. Curie ; et les livres suivants : de Briane-Pagès [BP06], de Jeulin-Brancovan [BJ06], de Rudin [Rud76, Rud87] et de Tao [Tao11]. J'ai essayé de mentionner quand un exercice provient d'une source particulière. Un grand merci à A.-M. Boussion, R. Butez, B. Haspot, J. Lehec, J. Ripa, É. Séré et F. Simenhaus pour les corrections qu'ils m'ont suggérées... et merci d'avance à ceux qui m'en suggèreront de nouvelles.

Table des matières

Table des figures

Notations

$f(\cdot, y)$ Application partielle $x \mapsto f(x,y)$ de f, page 99

$\bar{A} = A^c = \complement A = \Omega \backslash A$ Complémentaire de A (dans Ω), page 22

$\bar{\mathbb{R}}$ Droite numérique achevée $\mathbb{R} \cup \{-\infty, +\infty\}$, page 39

$\bar{\mathcal{A}}$ Tribu complétée de $\mathcal{A}$, page 33

$\#\Omega$ Cardinal d'un ensemble Ω, page 36

K_X Matrice de covariance de $X \in L^2_{\mathbb{R}^d}(\Omega)$, page 180

$A \Delta B$ Différence symétrique $(A \backslash B) \cup (B \backslash A)$, page 207

F' Dual d'un espace vectoriel topologique F, page 147

σ_X Écart-type de X, page 179

$\mathcal{L}^1$ Ensemble des fonctions intégrables, page 58

C_x, C^y Sections de C, page 99

$\mathcal{E}_+$ Ensemble des fonctions mesurables étagées positives, page 52

$\mathbb{N}$ Ensemble des entiers naturels $\{0, 1, 2, ...\}$, page 36

$\mathbb{N}_*$ Ensemble des entiers naturels non nuls, page 122

$\mathcal{R}$ Ensemble des fonctions Riemann-intégrables, page 16

F_σ Union dénombrable de fermés, page 263

G_δ Intersection dénombrable d'ouverts, page 263

Esc Ensemble des fonctions en escalier, page 15

$\ell^1(\mathbb{N})$ Espace des suites réelles sommables, page 145

$\ell^\infty(\mathbb{N})$ Espace des suites réelles bornées, page 147

$c_0(\mathbb{N})$ Espace des suites de limite nulle, page 149

$c_c(\mathbb{N})$ Espace des suites à support compact, page 149

$L^1(E, \mathcal{A}, \mu)$ Premier espace de Lebesgue, page 144

$L^1_{loc}(\mathbb{R}^n)$ Espace des fonctions localement intégrable, page 152

$L^p(E, \mathcal{A}, \mu)$ p-ième espace de Lebesgue, page 146

$\underline{\int}_a^b f\,dx$ Intégrale inférieure de Darboux de f, page 15

$\overline{\int}_a^b f\,dx$ Intégrale supérieure de Darboux de f, page 15

$\int f\,d\mu$ Intégrale de Lebesgue de f, page 50

$\int f\,dF$ Intégrale de Lebesgue-Stieltjes, page 79

$\int_a^b f\,dx$ Intégrale de Lebesgue de f sur $]a,b[$, page 50

$E(f)$ Espérance de f, page 50

$J\varphi$ Déterminant jacobien de φ, page 115

$\lim \uparrow$ Limite d'une suite croissante, page 32

$N(\mu, K)$ Loi gaussienne d'espérance μ et de covariance K, page 195

$f\mu$ Mesure de densité f par rapport à μ, page 58

$\mu * \nu$ Mesure convolée de μ et ν, page 188

λ Mesure de Lebesgue sur $\mathbb{R}$, page 32

$\varphi(\mu)$ Mesure image de μ par φ, page 46

$\mu \otimes \nu$ Mesure produit, page 101

μ_F Mesure de Lebesgue-Stieltjes, page 79

$\nu \ll \mu$ ν absolument continue par rapport à μ, page 239

$\nu \perp \mu$ ν étrangère à μ, page 239

μ-p.p. μ presque partout, page 56

μ-p.s. μ presque sûrement, page 56

$f * g$ Produit de convolution de f et g, page 151

$\sigma(\mathcal{C})$ Tribu engendrée par $\mathcal{C}$, page 26

$\check{f}$ Transformée de Fourier inverse de f, page 158

$\bar{\mathcal{F}}(f)$ Transformée de Fourier inverse de f, page 158

$\hat{f}$ Transformée de Fourier de f, page 155

$\mathcal{F}(f)$ Transformée de Fourier de f, page 155

Φ_X Fonction caractéristique de X, page 155

$\mathcal{B}(E)$ Tribu borélienne de E, page 26

$\bar{\mathcal{A}}$ Tribu complétée de $\mathcal{A}$, page 33

$\bar{\mathcal{B}}(\mathbb{R}^d)$ Tribu de Lebesgue, page 33

$f_*(\mathcal{A})$ Tribu image de $\mathcal{A}$ par f, page 44

$\mathcal{A} \otimes \mathcal{B}$ Tribu produit, page 99

$\sigma(f) = f^{-1}(\mathcal{B})$ Tribu engendrée par f, page 46

$f^{-1}(\mathcal{B})$ Tribu réciproque de $\mathcal{B}$ par f, page 46

$\mathcal{M}(\mu^*)$ Tribu des parties μ^*-mesurables, page 82

$\cup \uparrow$ Union d'une famille croissante d'ensembles, page 55

$\mathrm{var}(X)$ Variance de X, page 179

Tous les types d'énoncés sont numérotés consécutivement, sauf les exercices et les problèmes, qui ont leur propre numérotation.

Chapitre 1

Rappel sur l'intégrale de Riemann

La notion d'intrégrale des fonctions définies sur un intervalle compact a été introduite par Riemann [Rie54]. [1] En toute rigueur, cette notion d'intégrale n'est pas nécessaire pour la construction de l'intégrale de Lebesgue. Mais elle est fondamentale pour en saisir la beauté et la puissance. La présentation adoptée ici est due à Darboux. [2]

Soit $[a, b]$ un intervalle compact non vide de $\mathbb{R}$. Une *subdivision* $D = \{x_0 = a < x_1 < ... < x_N = b\}$ de $[a, b]$ est un ensemble fini de points de $[a, b]$, qui contient a et b, et dont on numérote bien sûr les points de façon croissante. Une fonction $g : [a, b] \to \mathbb{R}$ est *en escalier** et l'on note $g \in \mathrm{Esc}$ s'il existe une subdivision comme ci-dessus et des réels $y_1, ..., y_N$ tels que

$$g|_{]x_{i-1},x_i[} \equiv y_i \quad (\forall i = 1, ..., N);$$

les valeurs que g prend en les points x_i peuvent être encore différentes. On pose alors

$$I(g) = \sum_{1 \leqslant i \leqslant N} y_i(x_i - x_{i-1}).$$

Soit maintenant $f : [a, b] \to [m, M] \subset \mathbb{R}$ une fonction bornée. Les *intégrales inférieure* et *supérieure* (de Darboux) de f sont

$$\underline{\int_a^b} f\,dx = \sup_{g \in \mathrm{Esc},\, g \leqslant f} I(g) \quad \text{et} \quad \overline{\int_a^b} f\,dx = \inf_{g \in \mathrm{Esc},\, g \geqslant f} I(g).$$

Comme

$$m(b-a) \leqslant \underline{\int_a^b} f\,dx \leqslant \overline{\int_a^b} f\,dx \leqslant M(b-a),$$

ces intégrales inférieure et supérieures sont dans un intervalle borné. La question de savoir quand elles coïncident est délicate... Donc on en fait une définition !

1. Bernhard RIEMANN, mathématicien allemand (1826–1866), dont les contributions furent fondamentales en analyse, en théorie des nombres et en géométrie différentielle.

2. Jean-Gaston DARBOUX, mathématicien français (1842-1917), spécialiste de géométrie différentielle et d'analyse

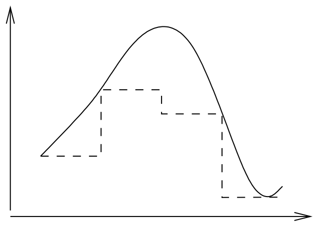


FIGURE 1 – Une fonction f et une fonction $g \in \mathrm{Esc}$, $g \leqslant f$

Définition 1.1. La fonction f est *Riemann-intégrable*, ce que l'on note $f \in \mathcal{R}$, si

$$\underline{\int_a^b} f\,dx = \overline{\int_a^b} f\,dx,$$

auquel cas on note $\int_a^b f\,dx$ cette valeur commune. [3]

Rappelons quelques critères d'intégrabilité, dont on trouvera une démonstration dans [Rud76].

Théorème 1.2. *Dans chacun des cas suivants, f est Riemann-intégrable :*
- *f est continue*
- *f est monotone*
- *f possède un nombre fini de discontinuités*
- *f est le produit de deux fonctions Riemann-intégrables*
- *f est de la forme $f = \phi \circ g$, où g est bornée et ϕ continue.*

Des nombreuses propriétés de l'intégrale de Riemann, rappelons simplement la suivante.

Théorème 1.3 (fondamental de l'analyse [Rud76])**.** *1. Si f est continue sur $[a,b]$, la fonction $F : x \mapsto \int_a^x f(t)\,dt$ est dérivable et $F' = f$ sur $[a,b]$.*

2. Si F est dérivable sur $[a,b]$ et si F' est Riemann-intégrable,

$$\int_a^b F'(x)\,dx = F(b) - F(a).$$

Exercice 1.1 (Somme de Riemann d'une fonction Lipschitzienne)**.**
Montrer que, si $f : [0,1] \to \mathbb{R}$ est une fonction lipschitzienne,

$$\left| \int_0^1 f(x)\,dx - \frac{1}{n} \sum_{1 \leqslant k \leqslant n} f\left(\frac{k}{n}\right) \right| \leqslant \frac{\mathrm{Lip}\, f}{2n},$$

où $\mathrm{Lip}\, f$ est la constante de Lipschitz de f. Cette inégalité s'améliore-t-elle si f est plus régulière ?

3. Le symbole $\int$ d'intégration est une forme italique alongée du caractère typgraphique ancien "s long". Il a été utilisé dès le XVIIe siècle par Leibniz, comme abbréviation de *summa*, qui veut dire "somme" en latin.

Exercice 1.2 (Intégration par parties)**.**
Soient f et g deux fonctions dérivables sur $[a, b]$ telles que f' et g' soient Riemann-intégrables. Montrer que

$$\int_a^b f(x)g'(x)\,dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\,dx.$$

(Indication : appliquer le théorème fondamental à $h := fg$, en remarquant que h' est Riemann-intégrable.)

L'exercice suivant montre comment ramener certains calculs fastidieux, voire impossibles en général, au simple calcul d'une intégrale.

Problème 1.1 (Courbes rectifiables)**.**
Soient $\gamma : T = [a, b] \to \mathbb{R}^n$ une courbe continue et $\tau = \{t_0 = a \leqslant t_1 \leqslant \cdots \leqslant t_{n-1} \leqslant t_n = b\}$ une subdivision de $[a, b]$. La *variation totale* de γ relativement à τ est

$$V_\tau(\gamma) = \sum_{i=1}^{n} \|\gamma(t_i) - \gamma(t_{i-1})\| \in [0, +\infty[,$$

où $\|\cdot\|$ est la norme euclidienne de $\mathbb{R}^n$. Autrement dit, $V_\tau(\gamma)$ est la longueur du polygone de sommets successifs $\gamma(t_0), ..., \gamma(t_n)$. Quand on choisit des subdivisions contenant de plus en plus de points de $[a, b]$, ce polygone approche la courbe γ de plus en plus près. La *longueur* de γ (voir la figure 2) est

$$L(\gamma) = \sup_\tau V_\tau(\gamma) \in [0, +\infty].$$

La courbe γ est *rectifiable* si elle est de longueur finie.

1. Montrer que

$$\gamma : [0, 1] \to \mathbb{R}, \quad t \mapsto \begin{cases} t^2 \cos^2(\pi/t^2) & \text{si } t \neq 0 \\ 0 \quad \text{si } t = 0 \end{cases}$$

n'est pas rectifiable.

2. Montrer que si γ est continûment dérivable, γ est rectifiable et

$$L(\gamma) = \int_a^b \|\gamma'(t)\|\,dt.$$

Cependant, l'une des premières surprises quand on découvre la théorie de l'intégration est que beaucoup (en fait, la plupart) des intégrales ne se calculent pas explicitement. Il s'avère que ce n'est forcément pas dû à notre médiocrité en calcul : c'est au contraire le fait de profonds théorèmes d'impossibilité. L'exercice suivant en décrit un très simple, dû à Newton [4] dans son étude de l'équation de Kepler. [5]

4. Isaac Newton, physicien et mathématicien anglais (1642–1727), probablement le savant ayant eu la plus grande influence de tous les temps sur la Science moderne.

5. Johannes Kepler, mathématicien et astronome allemand (1571–1630), célèbre pour avoir découvert les trois lois régissant le mouvement de révolution des planètes autour du Soleil.

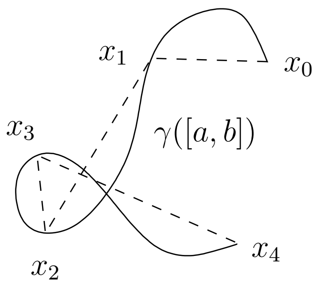


FIGURE 2 – La longueur de γ est la limite supérieure des courbes polygonales inscrites

Problème ★ 1.2 (Courbes quarrables [Arn99]).
Soit C une *courbe algébrique ovale*, c'est-à-dire une courbe convexe du plan d'équation $P(x, y) = 0$, pour un certain polynôme non nul P. On s'intéresse au problème suivant (figure 4). Soit une droite L d'équation $ax + by + c = 0$. Soit A l'aire de la région du plan délimitée par C et L, par exemple du côté où $ax + by + c > 0$. Une telle fonction A de a, b et c s'appelle une *intégrale abélienne.*[6] On dit que C est *algébriquement quarrable* si A est une fonction algébrique de L, c'est-à-dire qu'il existe un polynôme non nul Q de quatre variables tel que

$$Q(A, a, b, c) = 0 \quad (\forall a, b, c).$$

1. Montrer que l'ovale $y^2 = x^2 - x^3$ (figure 3) est algébriquement quarrable ; on pourra remarquer, en considérant l'intersection de la courbe avec la droite $y = tx$, que le courbe admet le paramétrage

$$x = 1 - t^2, \quad y = t - t^3,$$

puis montrer que $\int y\,dx$ est un polynôme en t, puis conclure.

2. Montrer que, si C est lisse (c'est-à-dire qu'elle est partout localement le graphe d'une fonction, de x ou de y, de classe C^∞), C n'est pas algébriquement quarrable.

3. Montrer par un argument analogue que si C est lisse la longueur d'arc de C n'est pas algébrique.

4. Montrer qu'une ellipse n'est pas algébriquement quarrable. En déduire, en utilisant la deuxième loi de Kepler (exercice 15.1), que la position d'une planète autour du Soleil est une fonction non algébrique (on dit *transcendante*) du temps.

6. Niels Henrik ABEL, mathématicien norvégien (1802–1829), a démontré que les équations polynomiales de degré 5 ne sont généralement pas résolubles par radicaux.

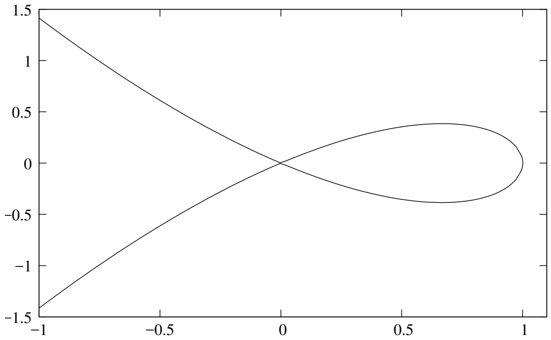


FIGURE 3 – L'ovale algébrique $y^2 = x^2 - x^3$

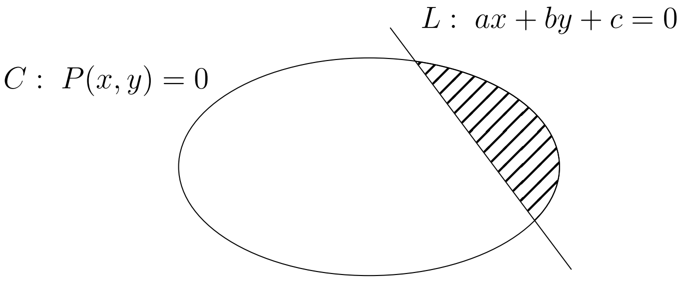


FIGURE 4 – Une ovale non quarrable

FIGURE 5 – Bernhard RIEMANN (1826–1866)

Chapitre 2

L'idée de mesure et les expériences aléatoires

L'idée de départ de la théorie de la mesure est d'assigner, à chaque partie d'un ensemble donné Ω, un nombre réel positif (la mesure de la partie), de manière à satisfaire certaines propriétés naturelles, notamment d'additivité. Ceci généralise les notions classiques de longueur d'une courbe, d'aire d'une surface ou de volume d'un solide.

Pour des raisons profondes, il n'est généralement pas possible de définir la mesure de toute partie de Ω : on doit se restreindre à une certaine classe* (*tribu*) de parties de Ω (dites *mesurables*). Nous reviendrons sur le problème de la mesure dans $\mathbb{R}^n$ au chapitre 10.

Il se trouve que le concept de probabilité est mathématiquement un cas particulier de celui de mesure, correspondant au cas où la mesure de Ω (la "masse totale") vaut 1. À la suite des travaux de Borel [1] et de Fréchet [2] qui ont laissé entrevoir les applications de la théorie de la mesure au calcul des probabilités, puis de ceux de Lévy, [3] qui a compris le rôle fondamental de la notion de convergence en loi des suites de variables aléatoires, Kolmogorov [4] a dégagé le modèle mathématique d'une expérience aléatoire (le futile tirage d'une boule de loto, par exemple), qui comporte les trois éléments principaux suivants (dont la détermination, d'ailleurs, est souvent problématique) [KS92].

L'espace des états Ω C'est l'ensemble des résultats possibles de l'expérience aléatoire. Un élément $\omega \in \Omega$ s'appelle un *résultat d'expérience*, ou une *issue du hasard*.

---

1. Émile Borel, mathématicien français (1871–1956)
2. Maurice Fréchet, mathématicien français (1878–1973)
3. Paul Pierre Lévy, mathématicien français (1886–1971), célèbre pour ses travaus en probabilités, et notamment pour avoir introduit les martingales et les processus de Lévy.
4. Andreï Nikolaevich Kolmogorov, mathématicien russe de premier plan (1903–1987), célèbre pour ses travaux fondamentaux notamment en probabilité, topologie, logique, turbulence et mécanique.

Les événements Ce sont des parties de $A \subset \Omega$. [5]

Il convient de distinguer un résultat d'expérience $\omega \in \Omega$ unique (qui s'identifie à l'événement *élémentaire* $\{\omega\} \subset \Omega$) d'un événement plus général (par exemple $\{\omega_1, \omega_2\}$) qui n'est pas forcément un singleton.*

On note $\mathcal{A}$ la classe de tous les événements considérés. Cela fait partie des données de l'expérience aléatoire qu'on étudie, de se limiter à une certaine classe d'événements $\mathcal{A} \subset \mathcal{P}(\Omega)$.

Si $A, B \in \mathcal{A}$ sont deux événements, il est souhaitable que l'événement $A \cup B$ (A ou B) soit aussi dans $\mathcal{A}$, puisque si l'on connaît la probabilité de A et celle de B, l'additivité de P permet de connaître la probabilité de $A \cup B$. Les axiomes que l'on souhaite imposer à $\mathcal{A}$ conduisent à *définir* la notion même de tribu comme dans le chapitre 3.

La probabilité À chaque événement A, on associe un nombre, noté $P(A)$ et appelé *probabilité de* A, compris entre 0 et 1. Ce nombre mesure le degré de vraisemblance de A, au sens suivant :

Si l'on répète l'expérience aléatoire une infinité de fois, $P(A)$ est la limite de la fréquence avec laquelle A est réalisé.

On en déduit immédiatement certaines propriétés naturelles de la fonction $P : \mathcal{A} \to [0, 1]$:

1. $0 \leqslant P(A) \leqslant 1$
2. $P(\Omega) = 1$ et $P(\varnothing) = 0$
3. $P(A^c) = 1 - P(A)$
4. $P(A \cup B) = P(A) + P(B)$ si $A \cap B = \varnothing$ (additivité).

Dans le chapitre 4, nous définirons une mesure, abstraitement, par une série d'axiomes de ce type.

Exemple 2.1. Si l'on tire successivement deux dés à 6 faces, l'espace des états est $\Omega = \{1, ..., 6\}^2$. L'événement "deux tirages pairs", par exemple, est la partie $A = \{2, 4, 6\}^2$. Si les dés sont symétriques, la probabilité est uniforme : pour tout $\omega \in \Omega$, $P(\{\omega\}) = 1/\sharp\Omega = 1/36$ et, pour tout événement $A \subset \Omega$,

$$P(A) = \frac{\sharp A}{36}.$$

Une quatrième notion, également importante quoique moins élémentaire, est la suivante.

Variable aléatoire C'est une application $X : \Omega \to E$ sur l'espace des états, à valeurs dans un ensemble E. Souvent, $E = \mathbb{R}$ ou $\mathbb{R}^d$, auquel cas on parle respectivement de variable aléatoire *réelle* ou *vectorielle*. On lui impose en tout cas une

5. Remarque sur l'orthographe et la prononciation du mot *événement*, tirée du *Trésor de la langue française* : L'accent aigu sur la deuxième syllabe ne signifie pas qu'il faut prononcer [e]. Il correspond à un [ε], soit un e moyen, voyelle à timbre non tranché, normal en syllabe inaccentuée.

propriété de compatibilité avec les tribus de Ω et de E : si $B \subset E$ est un événement mesurable de E, on veut que l'événement $\{X \in B\} := X^{-1}(B) \subset \Omega$ soit lui-même un événement mesurable.

Exercice 2.1 (Opérateurs à noyau de transition (S. Attal))**.**
Soit Ω un ensemble dénombrable et ν un *noyau de transition*, c'est-à-dire une application $E \times E \to [0, 1]$ telle que, pour tout x, $\sum_{y \in \Omega} \nu(x, y) = 1$; c'est dire que, pour tout x, $y \mapsto \nu(x, y)$ définit une probabilité sur Ω.

Soient $L^\infty(E)$ l'ensemble des fonctions réelles bornées sur Ω, et $T : L^\infty(E) \circlearrowleft$ l'opérateur associé au noau ν, défini par

$$Tf(x) = \sum_{y \in \Omega} f(y) \nu(x, y)$$

(quantité qu'on notera bientôt $\int_\Omega f(y)\, \nu(x, dy)$).

On dit que T est *déterministe* si et seulement si, pour tout x, la probabilité $\nu(x, \cdot)$ vaut 1 sur un unique point $\varphi(x) \in E$, et 0 ailleurs. En termes de T, cela signifie que $Tf = f \circ \varphi$.

1. Montrer que T est déterministe si et seulement si c'est un morphisme d'algèbre.

2. Montrer que si T est inversible et si son inverse possède un noyau de transition, T est déterministe.

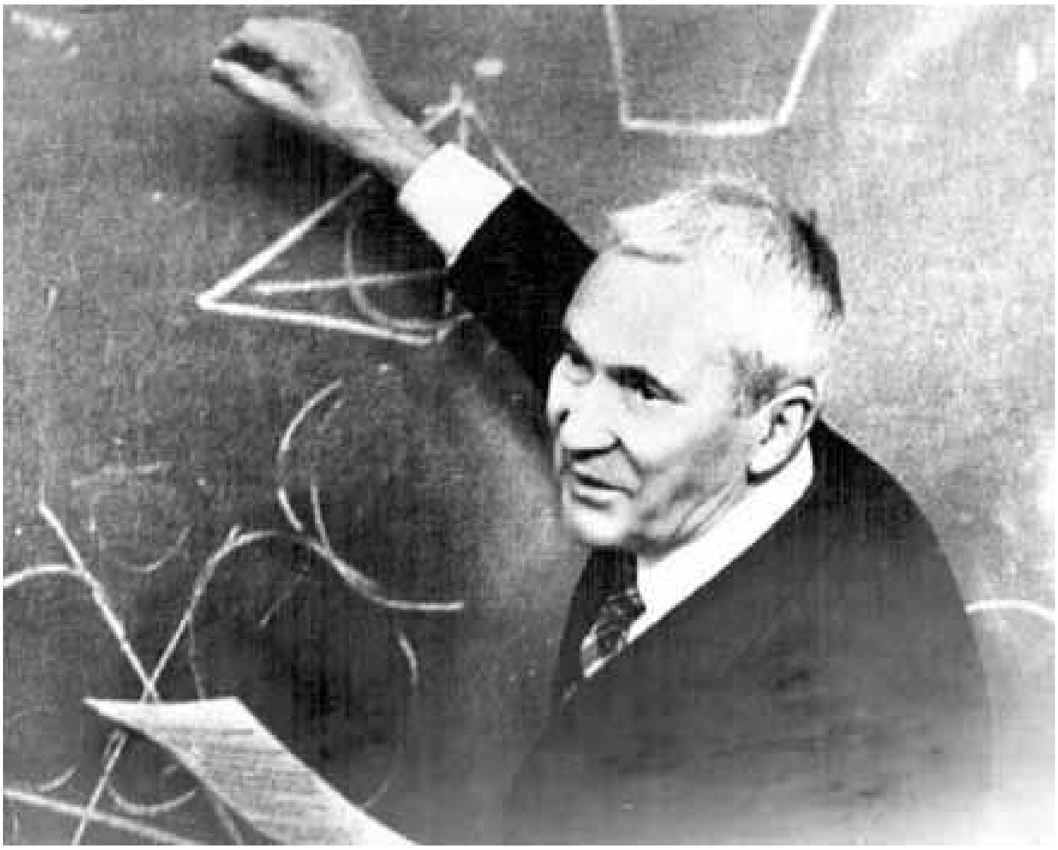


FIGURE 6 – Andreï Nikolaevich KOLMOGOROV (1903–1987), dont les cours furent aussi incompréhensibles que ses idées mathématiques furent profondes

Chapitre 3

Tribus

L'idée de tribu est due à Borel, Lebesgue, Fréchet, Kolmogorov, et bien d'autres. C'est une remarque fondamentale que d'une part, on ne *pourrait* pas, en général, assigner une mesure (avec des propriétés raisonables qui seront énoncées dans le prochain chapitre) à n'importe quelle partie d'un ensemble E ; et que d'autre part, on ne *voudrait* pas le faire, parce que la donnée de la tribu est une donnée à part entière des problèmes d'analyse ou des expériences aléatoires.

Définition 3.1. Une *tribu* ou σ*-algèbre* sur un ensemble E est une classe $\mathcal{A}$ de parties de E vérifiant les deux axiomes suivants :
- stabilité par complémentation : $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
- stabilité par union dénombrable* [1] : $A_1, A_2, ... \in \mathcal{A} \Rightarrow \cup A_i \in \mathcal{A}$.

Le couple $(E, \mathcal{A})$ est un *espace mesurable* ou *probabilisable.* Une partie $A \subset E$ est $\mathcal{A}$*-mesurable* si $A \in \mathcal{A}$ (on dit simplement *mesurable* s'il n'y a pas d'ambiguité).

Conséquences immédiates de la définition :
- $\varnothing \in \mathcal{A}$ (prendre l'union d'une famille vide de parties mesurables)
- $E \in \mathcal{A}$ par passage au complémentaire
- stabilité par intersection dénombrable : $A_1, A_2, ... \in \mathcal{A} \Rightarrow \cap A_i = (\cup A_i^c)^c \in \mathcal{A}$
- stabilité par différence : si $A, B \in \mathcal{A}$, $A \backslash B = A \cap B^c \in \mathcal{A}$.

Exercice 3.1 (Exemples élémentaires de tribus)**.**
Les classes suivantes sont-elles des tribus de E ?

1. $\mathcal{P}(E)$ (ensemble des parties de E)
2. $\{\varnothing, E\}$ (*tribu triviale*)
3. La classe des parties A de E qui sont dénombrables ou de complémentaire dénombrable
4. Dans le cas où $E = \mathbb{R}$, la classe des intervalles.

La dénombrabilité joue un rôle essentiel en théorie de la mesure. On peut en faire remonter la "cause" à la propriété suivante.

---

1. Nous convenons dans ce cours de qualifier un ensemble de *dénombrable* s'il s'injecte dans $\mathbb{N}$. En particulier, l'ensemble vide, ou n'importe quel ensemble fini, sont dénombrables.

Exercice 3.2 (Familles sommables)**.**
Soient I un ensemble quelconque et $(x_i)_{i\in I}$ une famille de nombre réels > 0. Par définition, la *somme* de (x_i) est

$$\sum_{i\in I} x_i := \sup_{J \text{ finie } \subset I} \sum_{j\in J} x_j \quad \in [0,+\infty].$$

Montrer que, si la somme de (x_i) est finie (on dit alors que la famille est *sommable*), I est dénombrable ; on pourra d'abord montrer que $J_n := \{j \in I,\ x_j \geqslant 1/n\}$ est fini pour tout $n \geqslant 1$.

Pour décrire des exemples de tribus plus intéressants, remarquons qu'une intersection quelconque (non vide) de tribus est encore une tribu (pourquoi ?). Cette idée conduit à la définition fondamentale suivante.

Définition 3.2. La *tribu engendrée* par une classe $\mathcal{C}$ de parties de Ω est

$$\sigma(\mathcal{C}) = \bigcap_{\mathcal{A} \text{ tribu de } \Omega\ \supset \mathcal{C}} \mathcal{A},$$

soit la plus petite tribu de Ω contenant $\mathcal{C}$.

Nous illustrons cette définition avec trois cas particuliers.

Tribu engendrée par une partition dénombrable

Soit $\mathcal{C}$ une partition dénombrable de E, c'est-à-dire une classe dénombrable de parties de E, telle que

1. $\mathcal{C}$ recouvre E : $\bigcup_{A\in\mathcal{C}} A = \Omega$
2. les éléments de $\mathcal{C}$ sont deux à deux disjoints : si $A \neq B$, $A \cap B = \varnothing$
3. si $A \in \mathcal{C}$, $A \neq \varnothing$ (non trivialité).

La tribu $\sigma(\mathcal{C})$ s'appelle une tribu *atomique*, d'*atomes* les parties A, $A \in \mathcal{C}$; elle est la classe des parties de Ω obtenues par réunion d'un nombre quelconque de parties $A \in \mathcal{C}$:

$$\sigma(\mathcal{C}) = \left\{ \bigcup_{A\in\mathcal{D}} A,\ \mathcal{D} \subset \mathcal{C} \right\}.$$

Donc, si $\operatorname{Card} \mathcal{C} < \infty$, $\operatorname{Card} \sigma(\mathcal{C}) = 2^{\operatorname{Card} \mathcal{C}}$.

Tribu borélienne d'un espace métrique

Rappelons que, si E est un espace métrique, sa *topologie* $\mathcal{O}$ est l'ensemble de ses ouverts.* Le couple $(E, \mathcal{O})$ est un *espace topologique.*

Définition 3.3. La *tribu borélienne* de $(E, \mathcal{O})$ est $\sigma(\mathcal{O})$. On la note $\mathcal{B}(E)$ ou $\mathcal{B}$. Un *borélien* est un événement $A \in \mathcal{B}$.

Exercice 3.3 (Tribu borélienne d'un espace séparable)**.**
Montrer que, si E est séparable*, $\sigma(\mathcal{O})$ est engendré par les boules ouvertes.

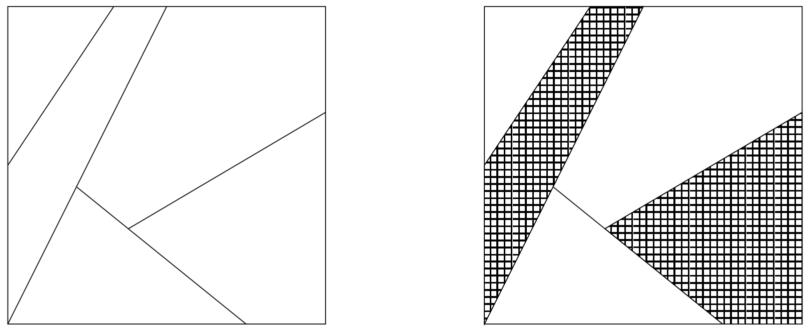


FIGURE 7 – À gauche, une partition d'un ensemble Ω, en 5 parties. À droite, un exemple de partie mesurable (grisée) pour la tribu engendrée par la partition.

Tribu produit

Soient $(E_1, \mathcal{A}_1)$ et $(E_2, \mathcal{A}_2)$ deux espaces mesurables. On veut munir le produit cartésien

$$E_1 \times E_2 = \{(x_1, x_2),\ x_1 \in E_1, x_2 \in E_2\}$$

d'une tribu.

Un *pavé* de $E_1 \times E_2$ est une partie de $E_1 \times E_2$ de la forme

$$A_1 \times A_2, \quad \text{avec } A_1 \subset E_1 \text{ et } A_2 \subset E_2.$$

(Par exemple, une partie en forme de "L" n'est pas un pavé.) Logiquement, le pavé est dit *mesurable* si A_1 et A_2 le sont.

La *tribu produit* sur $E_1 \times E_2$ est la tribu, notée $\mathcal{A}_1 \otimes \mathcal{A}_2$ engendrée par la classe des pavés mesurables.[2]

Autres exercices

Exercice 3.4 (Tribu engendrée par des intervalles)**.**
Calculer les tribus $\mathbb{R}$ engendrées par les ensembles $\mathcal{C}$ suivants :

$$\{[0,1],[0,2]\}, \quad \{[0,1],[2,3]\} \quad \text{et} \quad \{[0,2],[1,3]\};$$

on pourra montrer que chacune de ces tribus contient une partition $\mathcal{P}$ de $\mathbb{R}$, qui l'engendre.

Problème 3.1 (Exhaustion d'une algèbre par récurrence [Tao11])**.**
Une *algèbre*[3] sur E est une classe $\mathcal{A}$ de parties de E vérifiant les mêmes axiomes qu'une tribu, où la stabilité par union dénombrable est remplacée par la propriété plus faible de stabilité par union finie :

— $\varnothing \in \mathcal{A}$

2. Si l'on identifie une paire (A_1, A_2) à $A_1 \times A_2$ via l'application "produit cartésien" : $\mathcal{P}(E_1) \times \mathcal{P}(E_2) \to \mathcal{P}(E_1 \times E_2)$, $(A_1, A_2) \mapsto A_1 \times A_2$, on peut écrire (abusivement) : $\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\mathcal{A}_1 \times \mathcal{A}_2)$...

3. On précise parfois : *algèbre de Boole* (par opposition aux algèbres qui sont à la fois un corps et un anneau) *concrète* (par opposition aux algèbres de Boole abstraites).

— $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
— $A_1, A_2, ..., A_n \in \mathcal{A} \Rightarrow \cup_i A_i \in \mathcal{A}$ (pour tout entier n).

On définit de même l'algèbre $\alpha(\mathcal{C})$ engendrée par une classe de parties $\mathcal{C}$ comme la plus petite algèbre contenant $\mathcal{C}$. Définissons par récurrence les ensembles suivants :

— $\mathcal{A}_0 := \mathcal{C}$
— Pour tout $n \geqslant 1$, soit $\mathcal{A}_n$ la classe des parties obtenues comme union finie d'éléments de $\mathcal{A}_{n-1}$, ou comme le complémentaire d'une telle union.

Montrer que $\alpha(\mathcal{C}) = \cup_{n \geqslant 0} \mathcal{A}_n$.

Exercice 3.5 (Tribus et partitions)**.**

On rappelle que, dans ce cours, on appelle "dénombrable" un ensemble qui s'injecte dans $\mathbb{N}$ *; en particulier, cet ensemble peut être fini.*

1. Décrire la tribu engendrée par une partition $\mathcal{A}$ de E ; cettre tribu s'appelle la *tribu atomique* de E, d'atomes les éléments de $\mathcal{A}$.

Une tribu $\mathcal{E}$ définit naturellement une partition $\mathcal{A}_\mathcal{E}$ de E, dont les éléments sont les parties de la forme

$$\bar{x} = \bigcap_{x \in A \in \mathcal{E}} A, \quad x \in E.$$

2. Montrer que $\mathcal{A}_\mathcal{E}$ est bien une partition de E.

3. Montrer que si la tribu $\mathcal{E}$ est dénombrable la partition $\mathcal{A}_\mathcal{E}$ qui lui est associée engendre $\mathcal{E}$.

4. Montrer que si $\mathcal{E}$ est engendrée par une partition dénombrable $\mathcal{B}$ cette partition est $\mathcal{A}_\mathcal{E}$.

(On trouvera des exemples explicites dans l'exercice 3.4.)

5. Montrer que la tribu $\mathcal{P}(\mathbb{R})$ n'est engendrée par aucune partition de $\mathbb{R}$.

6. Montrer qu'une tribu infinie $\mathcal{E}$ n'est pas dénombrable et que donc la question (3) ne concerne que les tribus finies. (Indication : Raisonner par l'absurde et montrer que $\mathcal{E}$ serait en bijection avec l'ensemble des parties de la partition qui l'engendre.)

FIGURE 8 – Émile BOREL (1871–1956) et Maurice FRÉCHET (1878–1973)

Chapitre 4

Mesures et probabilités

Soit $(E, \mathcal{A})$ un espace mesurable.

Définition 4.1. Une *mesure* sur $(E, \mathcal{A})$ est une application $\mu : \mathcal{A} \to [0, +\infty]$ σ-additive [1] : pour toute famille dénombrable $A_1, A_2, \ldots$ de parties mesurables disjointes deux à deux,

$$\mu\left(\bigcup A_i\right) = \sum \mu(A_i).$$

De plus, on dit que

- μ est une *probabilité* si $\mu(E) = 1$.
- μ est *finie* si $\mu(E) < \infty$.

Un *espace mesuré* est un triplet $(E, \mathcal{A}, \mu)$; on l'appelle un *espace probabilisé* si μ est une probabilité.

En prenant une famille vide de parties mesurables (son union est $\varnothing$ et la somme des mesures est 0) on voit que la σ-additivité implique : $\mu(\varnothing) = 0$.

Exercice 4.1 (Exemples élémentaires de mesures)**.**
Les applications suivantes sont-elles des mesures ? Des probabilités ?

- La mesure de Dirac [2] δ_a en un point $a \in E$, définie sur $\mathcal{P}(E)$ par

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon.} \end{cases}$$

- La mesure de comptage, définie sur $\mathcal{P}(E)$ par

$$c(A) = \operatorname{Card} A.$$

- La fonction définie, sur la tribu de $\mathcal{P}(\mathbb{N})$, par

$$\mu(A) = \sum_{n \in A} 2^{-n}.$$

1. Le σ fait référence la dénombrabilité de la famille (A_j) à laquelle la propriété s'applique. On parle aussi d'*additivité dénombrable*.
2. Paul Adrien Maurice Dirac, physicien britannique (1902–1984), qui eut des contributions fondamentales en mécanique quantique et en électrodynamique quantique.

L'exemple de mesure le plus important dans ce cours est la mesure de Lebesgue[3] λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On verra ultérieurement que cette mesure est complètement caractérisée par le fait que, pour tout intervalle ouvert $]a, b[$ de $\mathbb{R}$ on a

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

Mais un coup d'oeil dans l'appendice A montre que l'attribution d'une mesure à des parties boréliennes n'a rien d'évident.

Par ailleurs, on se convaincra, s'il en est besoin, que les applications suivantes ne sont pas de même nature que des mesures (ce ne sont pas des applications d'une tribu vers $[0, +\infty]$)[4] :

— Le "symbole de Kronecker"

$$\delta_{ab} = \begin{cases} 1 & \text{si } a = b \\ 0 & \text{si } a \neq b, \end{cases}$$

défini pour $a, b \in E$.

— La fonction indicatrice

$$1_A : E \to \{0, 1\}, \quad x \mapsto \begin{cases} 0 & \text{si } x \in A \\ 1 & \text{sinon.} \end{cases}$$

d'une partie $A \in \mathcal{A}$ (on a cependant $1_A(x) = \delta_x(A)$).

Exercice 4.2 (Propriétés élémentaires des mesures)**.**
Démontrer les propriétés suivantes :

1. μ est croissante : si $A \subset B$, $\mu(A) \leqslant \mu(B)$.

2. Si $A \subset B$ et $\mu(A) < \infty$, $\mu(B \backslash A) = \mu(B) - \mu(A)$.

3. $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.

4. Si $A_1, A_2, ... \in \mathcal{A}$,

$$\mu\left(\cup_n A_n\right) \leqslant \sum_n \mu(A_n)$$

5. μ est *continue intérieurement* : si $A_n \subset A_{n+1}$ pour tout n,

$$\lim_n \uparrow \mu(A_n) = \mu\left(\cup_n A_n\right).$$

6. μ est *continue extérieurement* : si $A_n \supset A_{n+1}$ pour tout n, et si $\mu(A_0) < \infty$,

$$\lim_n \downarrow \mu(A_n) = \mu\left(\cap_n A_n\right).$$

---

3. Henri Lebesgue, mathématicien français (1875–1941), surtout célèbre pour sa théorie de l'intégration, mais dont l'influence fut considérable dans tout le domaine de l'analyse réelle.

4. En traitant δ_{ab} ou $\mathbb{1}_A$ comme des mesures dans une copie d'examen (ou simplement en se posant la question de savoir si ces applications sont σ-additives, ce qui n'a pas de sens), on crispe le correcteur avec probabilté totale.

(Ces deux propriétés de continuité sont les premières instances du théorème de convergence monotone, capital dans la théorie de l'intégration et dont on verra des versions plus générales ultérieurement.)

7. Donner un contre-exemple à la propriété de continuité extérieure quand on supprime l'hypothèse de finitude de $\mu(A_0)$.

8. λ étant la mesure de Lebesgue, montrer que, pour tous $a < b$,

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

et que

$$\lambda(\mathbb{R}) = \infty, \quad \lambda(\mathbb{Q}) = 0 \quad \text{et} \quad \lambda(\mathbb{R}\backslash\mathbb{Q}) = \infty;$$

pourquoi ceci fournit-il une nouvelle preuve du fait que $\mathbb{R}\backslash\mathbb{Q}$ est indénombrable ?

À peu de frais on peut prolonger une mesure à tous les ensembles obtenus en adjoignant, à une partie mesurable, une partie (éventuellement non-mesurable) contenue dans une partie mesurable de mesure nulle.

Définition 4.2. Soit $(E, \mathcal{A}, \mu)$ un espace mesuré. Une partie $A \subset E$ (pas forcément élément de $\mathcal{A}$) est *négligeable* si il existe $B \in \mathcal{A}$ telle que

$$A \subset B \quad \text{et} \quad \mu(B) = 0.$$

La *tribu complétée* de $\mathcal{A}$ est

$$\bar{\mathcal{A}} = \sigma(\mathcal{A} \cup \mathcal{N}),$$

où $\mathcal{N}$ est l'ensemble des parties négligeables de E.

Cas particulier : La tribu complétée $\bar{\mathcal{B}}(\mathbb{R}^d)$ de la tribu borélienne de $\mathbb{R}^d$ s'appelle la *tribu de Lebesgue* de $\mathbb{R}^d$.

La première instance où il est utile de compléter une tribu est dans la démonstration de la proposition 12.1, sur la comparaison entre intégrales de Riemann et de Lebesgue.

Exercice 4.3 (Tribu complétée)**.**
Montrer que $\bar{\mathcal{A}}$ est l'ensemble des parties A de E telles qu'il existe deux parties $B, B' \in \mathcal{A}$

— qui encadrent A : $B \subset A \subset B'$
— et telles que $\mu(B'\backslash B) = 0$.

Proposition 4.3. *Il existe une unique mesure sur* $(E, \bar{\mathcal{A}})$ *qui prolonge* μ.

Démonstration. Si $A \in \bar{\mathcal{A}}$, d'après l'exercice 4.3, il existe $B, B' \in \mathcal{A}$ tels que $B \subset A \subset B'$ et $\mu(B'\backslash B) = 0$. En particulier, $\mu(B) = \mu(B')$. Si l'on veut prolonger μ en une mesure sur $\bar{\mathcal{A}}$, la croissance de μ impose de poser

$$\mu(A) = \mu(B) = \mu(B').$$

Cette définition ne dépend pas du choix de B et B'. En effet, si $C, C' \in \mathcal{A}$ sont tels que $C \subset A \subset C'$ et $\mu(C'\backslash C) = 0$, on a à la fois

$$\begin{cases} C \subset A \subset B', & \text{donc } \mu(C) \leqslant \mu(B') \\ B \subset A \subset C', & \text{donc } \mu(B) \leqslant \mu(C'). \end{cases}$$

Or $\mu(B) = \mu(B')$ et $\mu(C) = \mu(C')$. Donc $\mu(B) = \mu(B') = \mu(C) = \mu(C')$.

Enfin, la fonction μ ainsi définie est bien une mesure. Le seul axiome d'une mesure à ne pas être immédiat est la σ-additivité. Si $A_n \in \bar{\mathcal{A}}$, il existe $B_n \in \mathcal{A}$ tel que $B_n \subset A_n$ et $\mu(A_n \backslash B_n) = 0$. On a

$$\begin{aligned}\mu\left(\bigcup A_n\right) &= \mu\left(\cup(A_n\backslash B_n)\bigcup(\cup B_n)\right)\\ &\leqslant \sum \mu(A_n\backslash B_n) + \mu(\cup B_n) = \mu\left(\bigcup B_n\right).\end{aligned}$$

Comme l'inégalité opposée est vraie aussi, $\cup A_n$ and $\cup B_n$ ont même mesure. Donc, si de plus les parties A_n sont deux à deux disjointes, il en est de même des B_n et

$$\begin{aligned}\mu\left(\bigcup A_n\right) &= \mu\left(\bigcup B_n\right)\\ &= \sum \mu(B_n) = \sum \mu(A_n).\end{aligned}$$

□

Autres exercices

Exercice 4.4 (L'ensemble triadique de Cantor).
Soient $I_0 = [0,1]$ l'intervalle unité, $I_1 = [0,1/3] \cup [2/3,1]$ l'ensemble obtenu à partir de I_0 en enlevant l'intérieur de la partie du milieu quand on découpe I_0 en trois parties égales, etc. Plus formellement, soit

$$I_n = \bigcup_{a_1,\dots,a_n \in \{0,2\}} \left[\sum_{i=1}^{n} \frac{a_i}{3^i}, \sum_{i=1}^{n} \frac{a_i}{3^i} + \frac{1}{3^n}\right].$$

Soit enfin $C = \cap_{n\geqslant 1} \downarrow I_n$ l'*ensemble triadique de Cantor.*[5] Montrer que C est borélien, compact, négligeable donc totalement discontinu[6] mais non dénombrable.

FIGURE 9 – $I_0, \dots, I_6$

Exercice 4.5 (Une mesure diffuse purement atomique, *partiel 2013*).
1. Monter que la tribu $\mathcal{A}$ engendrée par les singletons d'un ensemble E est la classe des parties qui sont soit dénombrables soit de complémentaire dénombrable.

2. Dans le cas particulier où E est l'ensemble $\mathbb{R}$, comparer $\mathcal{A}$ à la tribu borélienne de $\mathbb{R}$ et donner un exemple de partie qui ne soit pas dans la tribu $\mathcal{A}$.

5. Georg Ferdinand CANTOR, mathématicien allemand (1845-1918), fondateur de la théorie des ensembles et découvreur des nombres transfinis.

6. *Totalement discontinu* signifie qu'il ne contient aucun intervalle non trivial.

FIGURE 10 – Georg CANTOR (1845-1918)

On suppose dorénavant que E est (infini) non dénombrable. Pour toute partie $A \in \mathcal{A}$, on pose

$$P(A) = \begin{cases} 0 & \text{si } A \text{ est dénombrable} \\ 1 & \text{si } A^c \text{ est dénombrable.} \end{cases}$$

3. Montrer que, dans une famille (A_i) de parties mesurables disjointes deux à deux, il existe au plus une partie non dénombrable. En déduire que P est une probabilité.

4. Montrer que P est *diffuse*, i.e. $P(\{x\}) = 0$ pour tout $x \in E$.

5. Déterminer les *atomes* de P, c'est-à-dire les parties $A \in \mathcal{A}$ de mesure strictement positive et telles que pour toute partie $B \in \mathcal{A}$ incluse dans A on a $P(B) = 0$ ou $P(A \backslash B) = 0$ (ceci généralise la définition d'un élément atome ci-dessus). En déduire que P est *purement atomique*, c'est-à-dire que E est l'union d'atomes de P.

6. Soit Q une probabilité sur $(E, \mathcal{A})$. Montrer que, si A est l'ensemble des $x \in E$ tels que $Q(\{x\}) > 0$, il existe $c \in \mathbb{R}$ que l'on calculera tel que

$$Q = \sum_{a \in A} Q(\{a\})\delta_a + cP$$

(δ_a désignant la mesure de Dirac).

7. Montrer qu'une fonction mesurable $f : (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est constante en dehors d'une partie dénombrable ; on pourra d'abord considérer le cas d'une fonction étagée.

Exercice 4.6 (Deux exemples de mesures, *partiel 2011*).
Soient μ et ν les mesures sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définies par

$$\mu = \sum_{n \geqslant 1} e^{-n} \delta_{1/n} \quad \text{et} \quad \nu = \sum_{n \geqslant 1} e^{n} \delta_{1/n},$$

où δ_a est la mesure de Dirac en $a \in \mathbb{R}$

1. Vérifier que μ et ν sont bien des mesures.

2. Les mesures μ et ν sont-elles finies ? de probabilité ? σ-finies ?

3. Calculer $\mu(\{0\})$, $\mu([0, 1/k])$ $(k \geqslant 1)$ et $\lim_k \mu([0, 1/k])$. Comparer les résultats.

4. Faire de même avec ν.

Exercice 4.7 (Densité asymptotique des parties de $\mathbb{N}$).
Si A est une partie de $\mathbb{N}$, on pose

$$A_n = A \cap \{1, ..., n\}, \quad \text{et} \quad a_n = \# A_n \in \{0, ..., n\}.$$

De plus, on définit la *densité asymptotique de A*,

$$\mu(A) = \lim_{n \to +\infty} \frac{a_n}{n},$$

quand cette limite existe ; soit $\mathcal{L}$ l'ensemble des parties A possédant une telle limite.

1. Donner des exemples de parties A pour lesquelles $\mu(A)$ vaut respectivement 0, 1/2 et 1.

2. Calculer $\mu(A)$ lorsque A est l'ensemble des nombres premiers ; on pourra utiliser le théorème des nombres premiers, selon lequel

$$a_n \sim \frac{n}{\ln n}.$$

3. Montrer que $\mathcal{L} \subsetneq \mathcal{P}(\mathbb{N})$.

4. L'ensemble $\mathcal{L}$ est-il une tribu de $\mathbb{N}$?

5. La fonction μ est-elle σ-additive au sein de $\mathcal{L}$ (au sens que si $A_1, A_2, ... \in \mathcal{L}$ sont disjointes deux à deux et si $\cup_i A_i \in \mathcal{L}$, $\mu(\cup_i A_i) = \sum_i \mu(A_i)$), et peut-on la prolonger en une mesure sur la tribu engendrée par $\mathcal{L}$?

Chapitre 5

La tribu borélienne de $\mathbb{R}$, $\bar{\mathbb{R}}$ et $\mathbb{R}^d$

La tribu borélienne d'un espace métrique a été définie comme la tribu engendrée par la topologie (c'est-à-dire l'ensemble des ouverts). Nous revenons sur cette tribu borélienne dans les cas particulièrement importants de $\mathbb{R}$, de $\bar{\mathbb{R}}$ et de $\mathbb{R}^d$.

Nous aurons besoin de la description suivante des ouverts de $\mathbb{R}$.

Exercice 5.1 (Ouverts de $\mathbb{R}$)**.**
1. Montrer qu'un ouvert de $\mathbb{R}$ est une union dénombrable d'intervalles ouverts disjoints. En déduire que $\mathcal{B}(\mathbb{R}) = \sigma(\{]a,b[,\ a < b\})$

La force de la propriété précédente paraîtra plus grande au regard par exemple de l'impossibilité suivante.

2. ★ Montrer que l'intervalle semi-ouvert $[0,1[$ ne peut pas être l'union dénombrable d'intervalles *fermés* disjoints ; on pourra utiliser l'exercice 4.4.

Corollaire 5.1. *La tribu borélienne de $\mathbb{R}$ est engendrée par l'ensemble des intervalles de la forme*

$$]a,b[\,, \quad a,b \in \mathbb{R};$$

et par l'ensemble dénombrable des intervalles de la forme

$$]-\infty,a[\,, \quad a \in \mathbb{Q}.$$

Démonstration. Notons

$$\begin{cases} \mathcal{I} = \{]-\infty,a[,\ a \in \mathbb{Q}\} \\ \mathcal{J} = \{]a,b[,\ a,b \in \mathbb{R}\} \\ \mathcal{O} = \{O \text{ ouvert de } \mathbb{R}\}. \end{cases}$$

On veut montrer que $\sigma(\mathcal{I}) = \sigma(\mathcal{J}) = \sigma(\mathcal{O})$.

Montrons d'abord que $\sigma(\mathcal{J}) = \sigma(\mathcal{O})$. Comme $\mathcal{J} \subset \mathcal{O}$, $\sigma(\mathcal{J}) \subset \sigma(\mathcal{O})$. Il s'agit donc maintenant de démontrer que $\sigma(\mathcal{O}) \subset \sigma(\mathcal{J})$. Il suffit pour cela de démontrer que $\mathcal{O} \subset \sigma(\mathcal{J})$, puisque $\sigma(\mathcal{O})$ est la plus petite tribu contenant $\mathcal{O}$. Mais ceci découle du fait, démontré dans l'exercice 5.1, qu'un ouvert de $\mathbb{R}$ est l'union dénombrable d'intervalles ouverts, et que $\sigma(\mathcal{J})$ est stable par union dénombrable.

Montrons maintenant que $\sigma(\mathcal{I}) = \sigma(\mathcal{J})$. Pour tout $a \in \mathbb{Q}$,

$$]-\infty,a[= \cup_n]-n,a[\in \sigma(\mathcal{J}),$$

donc $\sigma(\mathcal{I}) \subset \sigma(\mathcal{J})$. Réciproquement, pour tous $a, b \in \mathbb{R}$, $]a, b[$ appartient à $\sigma(\mathcal{I})$. En effet, si (a_n) est une suite de rationnels qui tende vers a par valeurs supérieures, et (b_n) une suite de rationnels qui tende vers a par valeurs inférieures,

$$]a, b[=] - \infty, b[\setminus] - \infty, a] = \left(\bigcup_n] - \infty, b_n[\right) \setminus \left(\bigcup_n] - \infty, a_n[\right) \in \sigma(\mathcal{I}),$$

donc $\sigma(\mathcal{J}) \subset \sigma(\mathcal{I})$. □

Exercice 5.2 (Sur la tribu borélienne de $\mathbb{R}$)**.**
Montrer que la tribu borélienne de $\mathbb{R}$ est engendrée par chacune des classes d'intervalles suivantes :

1. $]a, b]$, $a, b \in \mathbb{Q}$
2. $]a, +\infty[$, $a \in \mathbb{Q}$
3. etc.

Remarque 5.2. La tribu borélienne d'un intervalle I de $\mathbb{R}$ est la tribu trace de la tribu borélienne de $\mathbb{R}$ sur I, au sens de l'exercice 6.7.

Remarque ★ 5.3 (Exhaustion de la tribu borélienne par récurrence transfinie). On a vu dans le problème 3.1 qu'une algèbre engendrée par une classe de parties peut être décrite par exhaustion dénombrable.

De même, à partir de la classe $\mathcal{O}$ des ouverts de $\mathbb{R}$, on aimerait pouvoir constuire la tribu borélienne par récurrence de la façon suivante :

— L'ensemble $\mathcal{B}_0 := \mathcal{O}$ est stable par union dénombrable (et même par union quelconque), mais pas par passage au complémentaire.
— L'ensemble $\mathcal{B}_1 := \mathcal{O} \cup \{B \subset \mathbb{R},\ B^c \in \mathcal{O}\}$ est stable par passage au complémentaire, mais plus par union dénombrable.
— L'ensemble $\mathcal{B}_2$ constitué des unions dénombrables d'éléments de $\mathcal{B}_1$ est stable par union dénombrable, mais plus par passage au complémentaire.
— Et ainsi de suite.

Par exemple,

$$\begin{cases}] - \infty, 0[\cup]0, +\infty[\in \mathcal{B}_0 \\ \{0\} = (] - \infty, 0[\cup]0, +\infty[)^c \in \mathcal{B}_1 \backslash \mathcal{B}_0 \\ \mathbb{Q} = \cup_{q \in \mathbb{Q}} \{q\} \in \mathcal{B}_2 \backslash \mathcal{B}_1 \\ \mathbb{Q}^c \in \mathcal{B}_3 \backslash \mathcal{B}_2 \\ \text{etc.} \end{cases}$$

On peut montrer par une sorte de procédé diagonal que cette récurrence ne converge pas avant l'étape “ω_1”, mais qu'on obtient donc ainsi la tribu borélienne au prix d'une *récurrence transfinie* (qui est une généralisation de la notion élémentaire de récurrence, utilisée par K. Gödel dans sa démonstration du théorème d'incomplétude de l'arithmétique formelle).[1]

---

1. Voir `http://fr.wikipedia.org/wiki/Hierarchie_de_Borel`

Rappelons que la *droite numérique achevée* $\bar{\mathbb{R}}$ est l'ensemble $\mathbb{R} \cup \{-\infty, +\infty\}$, muni de l'unique relation d'ordre qui prolonge celle de $\mathbb{R}$ et telle que

$$-\infty \leqslant x \leqslant +\infty \quad (\forall x \in \mathbb{R}).$$

FIGURE 11 – La droite numérique achevée

Dans la théorie de l'intégrale de Lebesgue, il est commode de pouvoir considérer des fonctions à valeurs dans $\bar{\mathbb{R}}$. On qualifie de telles fonctions de *réelles étendues.*

Définition 5.4. La *tribu borélienne* de $\bar{\mathbb{R}}$ est l'ensemble $\mathcal{B}(\bar{\mathbb{R}})$ des parties de $\bar{\mathbb{R}}$ de la forme $B \cup C$ avec $B \in \mathcal{B}(\mathbb{R})$ et $C \subset \{-\infty, +\infty\}$ (donc C est l'un des quatre ensembles suivants : $\varnothing$, $\{-\infty\}$, $\{+\infty\}$ ou $\{-\infty, +\infty\}$).

Remarque 5.5. Cette définition a le mérite d'être suffisante pour la suite, et l'inconvénient de ne pas présenter $\mathcal{B}(\bar{\mathbb{R}})$ explicitement comme la tribu engendrée par la classe des ouverts de $\bar{\mathbb{R}}$. Nous n'avons d'ailleurs muni $\bar{\mathbb{R}}$ d'aucune métrique, ni d'aucune topologie qui permette de le faire. Deux exercices en fin de chapitre pallient cette carence.

Exercice 5.3 (Sur la tribu borélienne de $\bar{\mathbb{R}}$)**.**
1. Montrer que la "tribu borélienne de $\bar{\mathbb{R}}$" (définition 5.4) est une tribu.

2. Montrer que la tribu borélienne de $\bar{\mathbb{R}}$ est engendrée par la classe des intervalles de la forme $]a, b]$ avec $a, b \in \bar{\mathbb{R}}$.

3. La tribu borélienne de $\bar{\mathbb{R}}$ est-elle engendrée par la classe des intervalles de la forme $]a, b[\ a, b \in \bar{\mathbb{R}}$?

Définition 5.6. Un *pavé ouvert* (respectivement *fermé*) de $\mathbb{R}^d$ est un sous-ensemble de la forme

$$\prod_{1 \leqslant i \leqslant d}]a_i, b_i[\quad (\text{resp. } \prod_{1 \leqslant i \leqslant d} [a_i, b_i]).$$

Exercice 5.4 (Tribu borélienne produit)**.**
1. Montrer qu'un ouvert de $\mathbb{R}^2$ est une union dénombrable de pavés ouverts. Quelle différence notable a-t-on avec la dimension un ?

2. En déduire que $\mathcal{B}(\mathbb{R}^2)$ est engendrée par l'ensemble des pavés ouverts de $\mathbb{R}^2$.

3. Montrer que $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$.

L'extension des questions précédentes à $\mathbb{R}^d$ ou à $\mathbb{C}^d$ est immédiate.

4. Montrer que, si E et F sont deux espaces métrique séparables,

$$\mathcal{B}(E \times F) = \mathcal{B}(E) \otimes \mathcal{B}(F).$$

Autres exercices

Exercice 5.5 (Une distance sur $\bar{\mathbb{R}}$)**.**
1. Montrer que l'application

$$h : \mathbb{R} \to]-1, 1[, \quad x \mapsto \operatorname{Argth} x$$

est un homéomorphisme (bijection bicontinue).
On la prolonge en une application $h : \bar{\mathbb{R}} \to [-1, 1]$ posant

$$h(-\infty) = -1 \quad \text{et} \quad h(+\infty) = 1.$$

2. Montrer qu'elle induit une distance sur $\bar{\mathbb{R}}$.
3. Décrire une base de voisinages de $+\infty$; on rappelle qu'une *base de voisinages* d'un point x est un ensemble B_x de voisinages de x tel que pour tout voisinage V de x il existe $U \in B_x$ contenu dans V.
4. Montrer que la tribu engendrée par l'ensemble des ouverts de $\bar{\mathbb{R}}$ est bien la tribu décrite dans la définition 5.4.

Un autre point de vue plus naturel est le suivant ; mais il requiert de considérer une topologie "abstraite", qu'on ne donne pas comme la topologie d'une distance.

Exercice ★ 5.6 (La topologie de l'ordre sur $\bar{\mathbb{R}}$)**.**
Cet exercice utilise les notions d'ordre et de topologie.
Soit $(X, \leqslant)$ un ensemble non vide totalement ordonné. On note $<$ la relation d'ordre strict associée à $\leqslant$ ($a < b$ si et seulement si $a \leqslant b$ et $a \neq b$). Un *intervalle ouvert* est une partie de l'une des quatre formes suivantes :

$$]a, b[:= \{x,\ a < x < b\}, \quad]\alpha, a[:= \{x,\ x < a\}, \quad]a, \omega[:= \{x,\ a < x\}, \quad]\alpha, \omega[:= X,$$

où $a, b \in X$ (et où α et ω ne sont que deux symboles qui n'ont aucun sens hérité d'ailleurs).
1. Montrer que l'ensemble des parties O de X telles que quel que soit $a \in O$ la partie O contient un intervalle ouvert contenant a est une topologie de X ; c'est la *topologie de l'ordre* de $(X, \leqslant)$.
2. Dans le cas de $(\mathbb{R}, \leqslant)$, montrer que la topologie de l'ordre coïncide avec la topologie usuelle.
Notons $-\infty$ et $+\infty$ deux éléments quelconques n'appartenant pas à $\mathbb{R}$, et $X = \mathbb{R} \cup \{-\infty, +\infty\}$. Soit $\leqslant$ l'unique relation d'ordre sur X qui prolonge celle de $\mathbb{R}$ et telle que

$$-\infty \leqslant x \leqslant +\infty \quad (\forall x \in \mathbb{R}).$$

On munit X de la topologie de la relation d'ordre.
3. Donner une base de voisinages de $\pm\infty$.
4. Montrer que $\mathbb{R}$ est dense dans X. (Pour cette raison, on note souvent $X = \bar{\mathbb{R}}$, ce qui présuppose de connaître la topologie de X.)
5. Montrer que la fonction $\varphi : \mathbb{R} \to]0, 1[$ définie par

$$\varphi(x) = \frac{x}{1 + |x|}$$

se prolonge par continuité en une application $\bar{\mathbb{R}} \to [0,1]$ et que ce prolongement est un homéomorphisme.

6. En déduire une distance d sur X dont la topologie métrique associée soit la topologie de l'ordre de X.

(Le lecteur rigoureux remarquera que $]\alpha, \omega[= [-\infty, +\infty]$, ce qui montre que l'on ne peut généralement pas identifier α et ω aux plus petit et plus grand éléments de X, même à supposer que ceux-ci existent !)

7. Montrer que la tribu borélienne de X (c'est-à-dire engendrée par les ouverts de X) est la "tribu borélienne" de $\bar{\mathbb{R}}$ au sens de la définition 5.4.

Exercice 5.7 (Nombres diophantiens, *partiel 2013*).
Soit $n \mapsto q_n$ une bijection $\mathbb{N} \mapsto \mathbb{Q}$. Définissons

$$\begin{cases} I_{n,k} = \left]q_n - 2^{-n-k}, q_n + 2^{-n-k}\right[\quad (n,k \in \mathbb{N}) \\ L = \bigcap_{k\in\mathbb{N}} \bigcup_{n\in\mathbb{N}} I_{n,k}. \end{cases}$$

On voit que $\mathbb{Q} \subset L$. Définissons $L_0 = L \backslash \mathbb{Q}$ et $D = \mathbb{R} \backslash L$, de sorte que $\mathbb{R}$ se décompose en la partition suivante :

$$\mathbb{R} = \underbrace{\mathbb{Q} \cup L_0}_{=L} \cup D;$$

les nombres de L_0 sont qualifiés de *liouvilliens*, et ceux de D de *diophantiens*.

L'objectif est de montrer que presque tous les nombres réels sont diophantiens. (On verra dans l'exercice 11.1 que L_0 est non dénombrable.)

1. Montrer que L et D sont boréliens.

2. Montrer que D est Lebesgue-négligeable ; on pourra chercher une majoration convenable de la mesure de Lebesgue de U_k.

Chapitre 6

Applications mesurables et variables aléatoires

Si $f : E \to F$ est une application, rappelons que l'on note f^{-1} l'application

$$\mathcal{P}(F) \to \mathcal{P}(E), \quad B \mapsto f^{-1}(B) := \{x \in E,\ f(x) \in B\}.$$ [1]

Soit $f : (E, \mathcal{A}) \to (F, \mathcal{B})$ une application entre deux espaces mesurables. [2]

Définition 6.1. L'application f est *mesurable* ou est une *variable aléatoire* si

$$f^{-1}(B) \in \mathcal{A} \quad \text{pour toute } B \in \mathcal{B}.$$

Quand E et F sont des espaces métriques, on les munit par défaut de leur tribu borélienne. Si f est mesurable dans ce cas, on dit aussi qu'elle est *borélienne.*

Une propriété de mesurabilité plus faible est que $f : (E, \bar{\mathcal{B}}(E)) \to (F, \mathcal{B}(F))$ soit mesurable (la tribu borélienne de l'ensemble de départ étant remplacée par sa complétée, la tribu de Lebesgue). Dans ce cas, on dit aussi que f est *Lebesgue-mesurable.* [3]

Exercice 6.1 (Mesurabilité d'une fonction indicatrice)**.**
La fonction indicatrice $\mathbb{1}_A : (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ de $A \subset E$ est mesurable si et seulement si A est mesurable.

Proposition 6.2. *La composée de deux applications mesurables est mesurable.*

Démonstration. Soient $f : (E, \mathcal{A}) \to (F, \mathcal{B})$ et $g : (F, \mathcal{B}) \to (G, \mathcal{C})$ deux applications mesurables et composables. L'application $g \circ f : (E, \mathcal{A}) \to (G, \mathcal{C})$ vérifie que

$$(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$$

1. Si f est une bijection, f^{-1} désigne aussi la bijection réciproque $F \to E$ de f. Mais il n'y a pas (encore) d'ambiguité puisque les arguments de ces deux applications ne sont pas de même nature : des parties de F dans le premier cas et des éléments de F dans le second.

Le problème est que, pour simplifier, on note souvent $f^{-1}(y)$ à la place de $f^{-1}(\{y\}) \subset E$, et il faut alors comprendre ce qu'on lit...

2. Cette notation désigne une application $f : E \to F$ et précise en même temps que E et F sont munis des tribus $\mathcal{A}$ et $\mathcal{B}$.

3. La propriété d'être mesurable de $(E, \bar{\mathcal{B}}(E))$ dans $(F, \bar{\mathcal{B}}(F))$ est souvent délicate à vérifier, et peu utilisée.

pour toute partie $C \in \mathcal{C}$. Comme g est mesurable, $g^{-1}(C) \in \mathcal{B}$ et, comme f est mesurable, $f^{-1}(g^{-1}(C)) \in \mathcal{A}$. Donc $g \circ f$ est mesurable. □

Remarque 6.3. Dans la proposition précédente, il est sous-entendu que l'ensemble F intermédiaire (d'arrivée de f et de départ de g) est muni d'une seule et même tribu $\mathcal{B}$. La conclusion n'a bien sûr aucune raison d'être vraie en général si ce n'est pas le cas. Par exemple, il existe des fonctions Lebesgue-mesurables dont la composition ne soit pas Lebesgue-mesurable.

Proposition 6.4. *Soient $f : (E, \mathcal{A}) \to (F, \mathcal{B})$ une application et $\mathcal{C}$ une classe de parties de F qui engendre $\mathcal{B}$. Si*

$$f^{-1}(B) \in \mathcal{A} \quad \textit{pour toute } B \in \mathcal{C},$$

f est mesurable.

Démonstration. L'ensemble

$$f_*(\mathcal{A}) := \left\{B \subset F,\ f^{-1}(B) \in \mathcal{A}\right\}$$

est une tribu (cf. l'exercice 6.2). Par hypothèse, $\mathcal{C} \subset f_*(\mathcal{A})$. Or, $\mathcal{B} = \sigma(\mathcal{C})$ est la plus petite tribu contenant $\mathcal{C}$. Donc $\mathcal{B} \subset f_*(\mathcal{A})$. C'est dire que f est mesurable. □

Exercice 6.2 (Tribu image)**.**
1. Si $\mathcal{A}$ est une tribu de E, montrer que la classe

$$\text{"}f(\mathcal{A})\text{"} := \{f(A),\ A \in \mathcal{A}\}$$

n'est généralement pas une tribu de F, mais que

$$f_*(\mathcal{A}) := \left\{B \subset F,\ f^{-1}(B) \in \mathcal{A}\right\}$$

en est une (*tribu image* de $\mathcal{A}$ par f).
2. Montrer que $f : (E, \mathcal{A}) \to (F, \mathcal{B})$ est mesurable si et seulement si $\mathcal{B} \subset f_*(\mathcal{A})$, c'est-à-dire que $f_*(\mathcal{A})$ est la plus grosse tribu de F qui rende f mesurable.

Exercice 6.3 (Mesurabilité et continuité)**.**
1. Soit $f : E \to F$ une application entre deux espaces métriques (ou topologiques). Montrer que, si f est continue, elle est borélienne. La réciproque est-elle vraie ? [4]
2. Soit $f : \mathbb{R}^n \to \mathbb{R}^d$ une application borélienne. Montrer que son graphe

$$\Gamma_f = \{(x, f(x)) \in \mathbb{R}^n \times \mathbb{R}^d,\ x \in \mathbb{R}^n\}$$

est un borélien de $\mathbb{R}^n \times \mathbb{R}^d$.

Proposition 6.5 (Opérations sur les applications mesurables)**.** *1. Soient*

$$f_1 : (E, \mathcal{A}) \to (F_1, \mathcal{B}_1) \quad \textit{et} \quad f_2 : (E, \mathcal{A}) \to (F_2, \mathcal{B}_2)$$

deux applications mesurables. L'application

$$f = (f_1, f_2) : (E, \mathcal{A}) \to (F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2), \quad x \mapsto (f_1(x), f_2(x))$$

est mesurable.

4. Hésiter sur cette question à un examen énervera instantanément le correcteur, tellement les contre-exemples sont nombreux.

2. *Si $f, g : (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ sont mesurables, les fonctions suivantes le sont aussi :*

$$f+g,\ fg,\ \inf(f,g), \sup(f,g), f^+ := \sup(f,0), f^- := \sup(-f,0)$$

(f^+ et f^- sont les parties positive *et* négative *de f).*

3. *Si les fonctions $f_n : (E, \mathcal{A}) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ ($n \in \mathbb{N}$) sont des fonctions mesurables de E dans $\bar{\mathbb{R}}$, les fonctions suivantes le sont aussi :*

$$\inf_n f_n,\ \sup_n f_n,\ \liminf_n f_n,\ \limsup_n f_n. \quad (*)$$

En particulier, si la suite (f_n) converge simplement, $\lim_n f$ est mesurable.

(Pour les définitions des limites inférieure et supérieure, on se reportera au glossaire.)

Donc, l'ensemble des fonctions mesurables est l'un des premiers ensembles de fonctions connu à être stable par limite simple !

Démonstration. 1. Posons

$$\mathcal{C} = \{B_1 \times B_2; B_1 \in \mathcal{B}_1,\ B_2 \in \mathcal{B}_2\}.$$

Puisque $f^{-1}(B_1 \times B_2) = f_1^{-1}(B_1) \cap f_2^{-1}(B_2) \in \mathcal{A}$ pour tout $B_1 \times B_2 \in \mathcal{C}$, la conclusion découle de la proposition 6.4.

2. Par exemple, la fonction $f+g$ est la composée de

$$(E, \mathcal{A}) \to (\mathbb{R}^2, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})), \quad x \mapsto (f(x), g(x)),$$

qui est mesurable d'après le point qui précède, et de

$$+ : (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})), \quad (a,b) \mapsto a+b,$$

qui est continue, donc mesurable d'après l'exercice 6.3. On sait que la composée de deux applications mesurables est mesurable, *pourvu bien sûr que l'espace intermédiaire soit muni d'une seule et même tribu.* Or, il y a une incohérence a priori ici, sur la tribu dont $\mathbb{R}^2$ est muni : $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ dans un cas, et $\mathcal{B}(\mathbb{R}^2)$ dans le second. Mais ces deux tribus sont bien égales, d'après l'exercice 5.4. Donc $f+g$ est mesurable. Il en va de même des autres cas.

3. Soit $f(x) = \inf f_n(x)$. Il suffit de montrer que, pour tout $a \in \mathbb{R}$, on a $f^{-1}([-\infty, a[) \in \mathcal{A}$. Or,

$$f^{-1}([-\infty, a[) = \left\{x;\ \inf_n f_n(x) < a\right\} = \bigcup_n \{x;\ f_n(x) < a\},$$

d'où le résultat. Le cas de $\sup_n f_n$ est similaire. On en déduit que

$$\liminf_n f_n = \sup_{n \geqslant 0} \inf_{k \geqslant n} f_k$$

est mesurable, et de même pour

$$\limsup_n f_n = \inf_{n \geqslant 0} \sup_{k \geqslant n} f_k.$$

□

Exercice 6.4 (Quotient de fonctions mesurables)**.**
Soient $f, g : (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ deux fonctions mesurables. Montrer que la fonction

$$\frac{f}{g} : E \to \mathbb{R}, \quad x \mapsto \begin{cases} \frac{f(x)}{g(x)} & \text{si } g(x) \neq 0 \\ 0 & \text{sinon} \end{cases}$$

est mesurable.

Exercice 6.5 (Points de convergence)**.**
Montrer que, si (f_n) est une suite de fonctions mesurables de E dans $\bar{\mathbb{R}}$, l'ensemble

$$L := \left\{ x \in E;\ \lim_n f_n(x) \text{ existe} \right\}$$

est mesurable ; on pourra utiliser le fait que la limite de $f_n(x)$ existe si et seulement si les limites inférieure et supérieure coïncident.

Définition 6.6. Soient $\varphi : (E, \mathcal{A}) \to (F, \mathcal{B})$ une application mesurable et μ une mesure sur $(E, \mathcal{A})$. La *mesure image* de μ par φ est la mesure $\varphi(\mu)$ sur $(F, \mathcal{B})$ définie par

$$\varphi(\mu)(B) = \mu(\varphi^{-1}(B)).$$

Si μ est une probabilité, $f(\mu)$ est une probabilité, s'appelle la *loi* de φ, et se note aussi μ_φ.

Exercice 6.6 (Mesure image)**.**
Montrer que $\mu(\varphi)$ est une mesure, de même masse que μ.

Les notions introduites dans l'exercice suivant, de tribu engendrée par une application et de tribu trace, sont importantes pour la suite.

Exercice 6.7 (Tribu réciproque)**.**
Soit $f : E \to F$ une application entre deux ensembles.
1. Si $\mathcal{B}$ est une tribu de F, montrer que la classe

$$f^{-1}(\mathcal{B}) := \left\{ f^{-1}(B),\ B \in \mathcal{B} \right\}$$

est une tribu de E (*tribu réciproque* de $\mathcal{B}$ par f).
Remarque : Si la tribu $\mathcal{B}$ est fixée une fois pour toute et si l'on s'intéresse à différentes applications mesurables, $f^{-1}(\mathcal{B})$ s'appelle aussi la *tribu engendrée par* f et se note $\sigma(f)$. Supposons qu'un expérimentateur observe un système physique, caractérisé par son état $x \in E$, mais que la seule information accessible à son sujet soit $f(x) \in F$. Intuitivement, $\sigma(f)$ est l'ensemble des événements $A \subset E$ pour lesquels, pour chaque x, on peut savoir si $x \in A$ ou pas, uniquement sur la base de l'information fournie par $f(x)$.
2. Montrer que $f : (E, \mathcal{A}) \to (F, \mathcal{B})$ est mesurable si et seulement si $\mathcal{A} \supset f^{-1}(\mathcal{B})$, c'est-à-dire que $f^{-1}(\mathcal{B})$ est la plus petite tribu de E qui rende f mesurable.
3. Dans le cas où E est une partie de F et où f est l'inclusion $x \in E \mapsto x \in F$, montrer que la tribu réciproque est

$$f^{-1}(\mathcal{B}) = \{B \cap E; B \in \mathcal{B}\};$$

cette tribu s'appelle la *tribu trace* de $\mathcal{B}$ sur E.

4. Montrer que, si I est un intervalle de $\mathbb{R}$, $\mathcal{B}(I)$ est la tribu trace de $\mathcal{B}(\mathbb{R})$ sur I.

Exercice 6.8 (Tribu engendrée par une application)**.**
Soient $X : E \to (F, \mathcal{B})$ une application et $f : E \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction. Montrer que les deux propriétés suivantes sont équivalentes :

1. f est $\sigma(X)$-mesurable,
2. f se factorise par X, i.e. il existe une fonction mesurable $\varphi : (F, \mathcal{B}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que $f = \varphi \circ X$:

$$\begin{tikzcd} E \arrow[r, "f"] \arrow[d, "X"'] & \mathbb{R}. \\ F \arrow[ur, dotted, "\varphi"'] & \end{tikzcd}$$

Autres exercices

Exercice 6.9 (Image d'une fonction mesurable)**.**
Construire une fonction $f : \mathbb{R} \to \mathbb{R}$ Lebesgue-mesurable, nulle presque partout, et telle que $f(\mathbb{R}) = \mathbb{R}$.

Exercice 6.10 (Mesurabilité et inversibilité)**.**
Donner un exemple de bijection mesurable dont la réciproque n'est pas mesurable. De même, donner un exemple de bijection continue dont la réciproque n'est pas continue.

Exercice 6.11 (Tribu invariante, *examen 2012*)**.**
Soient $(E, \mathcal{A}, \mu)$ un espace mesuré et $\varphi : E \to E$ une application mesurable.

1. Montrer que $\mathcal{I} = \{A \in \mathcal{A},\ \varphi^{-1}(A) = A\}$ est une tribu de E (on l'appelle la *tribu invariante* de φ).

2. Soit $f : (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ une fonction $\mathcal{A}$-mesurable. Montrer que f est $\mathcal{I}$-mesurable si et seulement si elle est φ-invariante ($f \circ \varphi = f$).

Problème 6.1 (Complétion universelle [BJ06])**.**
Soit $\mathcal{M}(\mathcal{A})$ l'ensemble des probabilités sur $(E, \mathcal{A})$. Pour $\mu \in \mathcal{M}(\mathcal{A})$, on note ici $\mathcal{A}^\mu$ la tribu complétée de $\mathcal{A}$ pour μ. On définit la *complétion universelle* de $\mathcal{A}$ comme la tribu

$$\mathcal{A}^* = \cap_{\mu \in \mathcal{M}(\mathcal{A})} \mathcal{A}^\mu.$$

Montrer que si $(E, \mathcal{A})$ et $(F, \mathcal{B})$ sont des espaces mesurables et si $f : (E, \mathcal{A}) \to (F, \mathcal{B})$ est mesurable, l'application

$$f : (E, \mathcal{A}^*) \to (F, \mathcal{B}^*)$$

est mesurable.

Problème ★ 6.2 (Théorème de récurrence de Poincaré, *examen 2012*)**.**
Soient $(\Omega, \mathcal{A}, \mu)$ un espace de probabilité et $T : \Omega \to \Omega$ une application mesurable.

On note μ_T la probabilité image de μ par T, et T^n la n-ième composée itérée de T, définie par récurrence par $T^0 = \mathrm{id}$, $T^{n+1} = T \circ T^n$ $(n \geqslant 0)$.

Soit $A \in \mathcal{A}$. Un point $x \in A$ est A-*récurrent* s'il existe une infinité de rangs n tels que $T^n(x) \in A$. Notons $\hat{A}$ l'ensemble des points A-récurrents.

Montrer que, si μ est T-invariante, on a $\mu(\hat{A}) = \mu(A)$ (théorème de récurrence de Poincaré[5]) ; on pourra montrer pour cela que $A \backslash \hat{A}$ est l'union des $B_n = \{x \in A,\ T^{n-1}(x) \in A,\ \forall k \geqslant n\ T^k(x) \notin A\}$ $(n \geqslant 1)$, que pour tous $n, k, l \geqslant 1$ les parties $T^{-nk}(B_n)$ et $T^{-nl}(B_n)$ sont disjointes si $k \neq l$, et par l'absurde que $\mu(B_n) = 0$ pour tout n.

Problème ★ 6.3 (Graphe d'une application)**.**

1. Montrer que si $\mathcal{A}$ est une tribu engendrée par un ensemble dénombrable $\mathcal{C}$ de parties de E, il existe une fonction $f : E \to \mathbb{R}$ qui engendre $\mathcal{A}$.

2. En déduire que, si $(E, \mathcal{A})$ et $(F, \mathcal{B})$ sont des espaces mesurables et si une application $f : E \to F$ est mesurable, le graphe

$$G_f = \{(x, y) \in E \times F,\ y = f(x)\}$$

de f appartient à $\mathcal{A} \otimes \mathcal{B}$ si et seulement si il existe une sous-tribu $\mathcal{C}$ de $\mathcal{B}$ engendrée par une classe dénombrable de parties et telle que $\{y\} \in \mathcal{C}$ pour tout $y \in f(E)$.

5. Henri Poincaré, mathématicien, physicien et philosophe des sciences français (1854–1912).

Chapitre 7

Intégrale et espérance

À la fin du XIXe siècle, il apparut que l'intégrale de Riemann (celle qui est enseignée dans les cours de calcul différentiel) devrait être remplacée par une intégrale plus flexible et plus générale. Après plusieurs tentatives, c'est celle de Henri Lebesgue qui se montra la plus féconde [Leb01].

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré. Qu'est-ce que l'intégrale d'une fonction mesurable $f : E \to \mathbb{R}$? La mesure d'une partie mesurable A de E peut être interprétée comme l'intégrale de la fonction indicatrice de A :

$$\int \mathbb{1}_A \, d\mu := \mu(A).$$

Par linéarité, on prolonge immédiatement la définition de l'intégrale pour toute fonction "étagée". Puisque toute fonction mesurable f est limite simple d'une suite de fonctions étagées f_n, il ne reste qu'à prendre, pour l'intégrale de f, la limite des intégrales des f_n. Sauf que... cette méthode conduit à une notion d'intégration incohérente (pourquoi ? [1]). Le miracle est qu'il suffit de définir l'intégrale des fonctions par limite *croissante*, du moins si l'on se restreint aux fonctions positives. Il ne reste alors qu'à étendre la définition aux fonctions de signe quelconque, par différence, quand cela est possible.

Fonctions étagées positives

Supposons que $f : E \to \mathbb{R}$ est *étagée*, c'est-à-dire que f est mesurable et son image $f(E)$ est finie. Alors f s'exprime tautologiquement comme une combinaison linéaire finie de fonctions indicatrices :

$$f = \sum_{y \in f(E)} y \mathbb{1}_{f^{-1}(y)}. \tag{7.1}$$

1. L'intégrale des f_n n'a généralement pas de limite et, même à supposer que cette limite existe, elle dépend de la suite d'approximations étagées choisie. Par exemple, 0 est la limite simple des fonctions $f_n = \mathbb{1}_{[n,+\infty[}$, mais on ne voudrait pas définir l'intégrale de la fonction nulle comme la limite des intégrales des f_n, qui valent toutes $+\infty$. On retrouve ici la pathologie qui mettait en défaut la propriété de continuité extérieure d'une mesure quand celle-ci n'est pas finie. Pire, 0 est aussi la limite des fonctions $g_n = (-1)^n \mathbb{1}_{[n,+\infty[}$, mais les intégrales des g_n valent alternativement $+\infty$ et $-\infty$, donc n'ont pas de limite !

Cette décomposition de f est dite *canonique*, parce qu'elle compte le moins de termes possibles, à savoir le cardinal de $f(E)$; toutes les autres combinaisons linéaires de fonctions indicatrices, égales à f, ont strictement plus de termes (par exemple, si $a, b \in E$, on peut toujours ajouter $\mathbb{1}_{\{a,b\}} - \mathbb{1}_{\{a\}} - \mathbb{1}_{\{b\}}$). De plus, chaque terme est mesurable, puisque $f^{-1}(y) \in \mathcal{A}$ pour tout $y \in f(E)$.

Par exemple, une fonction en escalier* sur un intervalle de $\mathbb{R}$ est étagée. Mais la fonction indicatrice de $\mathbb{Q}$ sur $\mathbb{R}$ est étagée sans être en escalier.

Supposons de plus que f est positive.

Définition 7.1. L'*intégrale* de f par rapport à μ est

$$\int f \, d\mu := \sum_{y \in f(E)} y \, \mu(f^{-1}(y)) \in [0, +\infty],$$

avec la convention $0.\infty = 0$ dans le cas où $y = 0$ et $\mu(f^{-1}(y)) = \infty$.

Chaque terme dans la somme du membre de droite appartient à $[0, +\infty]$, donc la somme est bien définie. (En revanche, si f n'était pas supposée positive, on aurait un problème pour donner un sens à des sommes $y + \infty - \infty$ tout en préservant la propriété de linéarité du lemme 7.2).

Beaucoup de notations plus ou moins explicites coexistent :

$$\int f \, d\mu = \int_E f = \int f(x) \, d\mu(x) = \int f(x) \mu(dx) = \langle \mu, f \rangle = \mu(f) = ...$$

où il ne faut pas chercher à donner un sens à "l'opérateur d", dans ce cours. Certains cas particuliers ont aussi des notations disparates mais commodes :

— Si $\mu = \lambda$ (mesure de Lebesgue), on note

$$d\lambda(x) = \lambda(dx) = dx = dt = ...$$

selon le nom de la variable muette utilisée,

— Si $(E, \mathcal{A})$ est un intervalle $]a, b[$ ou $[a, b]$ de $\mathbb{R}$ *et si les singletons sont μ-négligeables*, on note aussi

$$\int_{]a,b[} f \, d\mu \left(= \int_{[a,b]} f \, d\mu \right) = \int_a^b f \, d\mu = \int_a^b f = - \int_b^a f = ...$$

(la première égalité provenant de la décomposition $f = f\mathbb{1}_{[a,b]} + f\mathbb{1}_{\{a\}} + f_{\{b\}}$ et du lemme 7.2 ci-dessous).

— Si μ est une probabilité, l'intégrale de f s'appelle son *espérance* et on note encore plus simplement

$$\int f \, d\mu =: E(f).$$

Sur les parties de $\mathbb{R}^n$, par défaut on utilise la tribu borélienne et la mesure de Lebesgue.

Lemme 7.2. *Soient f et g deux fonctions étagées positives.*

1. Si $a, b \geqslant 0$,

$$\int (af + bg)\, d\mu = a \int f\, d\mu + b \int g\, d\mu. \tag{7.2}$$

2. Si $f \leqslant g$,

$$\int f\, d\mu \leqslant \int g\, d\mu.$$

La première partie du lemme (7.2) ne découle pas directement de la définition de l'intégrale pour les fonctions étagées, parce que l'ensemble des fonctions indicatrices, qui engendre l'espace des fonctions étagées, n'en est pas une base : il existe une infinité de façon de décomposer une fonction étagée en combinaison linéaire de fonctions indicatrices et il n'est pas trivial que toutes ces décompositions soient compatibles du point de vue de l'intégration. En particulier, par récurrence on déduit de (7.2) si $a_i \in \mathbb{R}_+$ et $A_i \in \mathcal{A}$ $(1 \leqslant i \leqslant n)$,

$$\int \left(\sum_{1 \leqslant i \leqslant n} a_i \mathbb{1}_{A_i} \right) = \sum_{1 \leqslant i \leqslant n} a_i \mu(A_i). \tag{7.3}$$

(bien que la fonction à intégrer ne soit pas dans sa décomposition canonique, si certaines parties A_i se chevauchent).

Démonstration. 1. Si $a = 0$, on a manifestement $\int (0f) = 0 \int f$. Supposons $a \neq 0$. Alors $f(x) = y \Leftrightarrow (af)(x) = ay$, donc

$$\begin{aligned} \int (af) &= \sum_{z \in (af)(E)} z\, \mu((af)^{-1}(z)) \\ &\overset{z=ay}{=} \sum_{y \in f(E)} ay\, \mu(f^{-1}(y)) \\ &= a \sum_{y \in f(E)} y\, \mu(f^{-1}(y)) = a \int f. \end{aligned}$$

Montrons maintenant que

$$\int (f + g) = \int f + \int g.$$

En général, f, g et $f + g$ ne prennent pas les mêmes valeurs. Nous allons cependant découper les sommes donnant leurs intégrales de façon à ce qu'elles soient indexées par les mêmes ensembles.

Considérons la partition de E par les ensembles de niveaux[2] de g :

$$E = \bigcup_{z \in g(E)} g^{-1}(z)$$

2. Un *ensemble de niveau* d'une application g est une partie de son ensemble de départ E d'équation $g = z$, où z est un élément de l'ensemble d'arrivée de g. La collection des ensembles de niveau de g est une partition de E parce que : elle recouvre bien E, parce que tout élément a une image par g, par définition même d'une application ; les ensembles de niveau sont disjoints deux à deux parce qu'un élément $x \in E$ n'a qu'une seule image par g,par définition encore d'une application ; enfin, les ensembles de niveau sont non vides, parce que z est choisi dans l'image de g.

et cette union est disjointe. Par additivité,

$$\begin{aligned}\int f &= \sum_{y \in f(E)} y\, \mu(f^{-1}(y)) \\ &= \sum_{y \in f(E),\, z \in g(E)} y\mu(f^{-1}(y) \cap g^{-1}(z))\end{aligned}$$

et, symétriquement,

$$\int g = \sum_{y \in f(E),\, z \in g(E)} z\mu(f^{-1}(y) \cap g^{-1}(z)),$$

de sorte que

$$\begin{aligned}\int f + \int g &= \sum_{y \in f(E),\, z \in g(E)} (y+z)\mu(f^{-1}(y) \cap g^{-1}(z)) \\ &= \sum_{u \in (f+g)(E)} u \sum_{y \in f(E),\, z \in g(E),\, y+z=u} \mu(f^{-1}(y) \cap g^{-1}(z)) \\ &= \sum_{u \in (f+g)(E)} u\mu((f+g)^{-1}(u)) \\ &= \int (f+g),\end{aligned}$$

l'avant-dernière égalité étant justifiée par le fait que $(f+g)^{-1}(u)$ est l'union disjointe des $f^{-1}(y) \cap g^{-1}(z)$ avec $y \in f(E)$ et $z \in g(E)$ tels que $y+z=u$. L'éaglité (7.3) se déduit de la linéarité ainsi démontrée, par une simple récurrence.

2. Supposons $f \leqslant g$. Pour comparer les intégrales de f et de g, utilisons la même formule que précédemment :

$$\begin{cases} \int f = \sum_{y \in f(E),\, z \in g(E)} y\mu(f^{-1}(y) \cap g^{-1}(z)) \\ \int g = \sum_{y \in f(E),\, z \in g(E)} z\mu(f^{-1}(y) \cap g^{-1}(z)). \end{cases}$$

Pour tout $x \in E$, $f(x) \leqslant g(x)$, soit, en notant $y = f(x)$ et $z = g(x)$, $y \leqslant z$. Donc $\int f \leqslant \int g$.

□

Fonctions mesurables positives

Notons $\mathcal{E}_+$ l'ensemble des fonctions mesurables étagées positives sur E. Soit f une fonction positive étendue, c'est-à-dire une fonction

$$f : (E, \mathcal{A}) \to ([0, +\infty], \mathcal{B}([0, +\infty]))$$

(il est commode pour la suite d'autoriser f à prendre la valeur $+\infty$), supposée mesurable ; $\mathcal{B}([0, +\infty])$ est la tribu trace de $\mathcal{B}(\bar{\mathbb{R}})$ sur $[0, +\infty]$:

$$\mathcal{B}([0, +\infty]) = \left\{A \cap [0, +\infty],\ A \in \mathcal{B}(\bar{\mathbb{R}})\right\}.$$

Définition 7.3. L'*intégrale* de f est

$$\int f\,d\mu := \sup_{h\in\mathcal{E}_+,\,h\leqslant f} \int h\,d\mu.$$

La seconde propriété du lemme 7.2 montre que cette définition est cohérente avec la précédente quand f est étagée.

Propriétés 7.4. *1. L'intégration est croissante : si $f \leqslant g$, $\int f\,d\mu \leqslant \int g d\mu$.*

2. L'intégration est "définie" : si $f = 0$ en dehors d'un ensemble de mesure nulle[3]*, f est intégrable et $\int f\,d\mu = 0$.*

Démonstration. 1. La croissance découle de la définition même et de la croissance pour les fonctions étagées.

2. Si $\mu(f^{-1}(]0,+\infty])) = 0$, toute fonction de $\mathcal{E}_+$ minorant f est elle-même majorée par la fonction étagée $+\infty\,\mathbb{1}_{f^{-1}(]0,+\infty])}$, dont l'intégrale est nulle d'après la convention dans la définition 7.1. Donc $\int f = 0$.

□

Théorème 7.5 (de convergence monotone, de B. Levi[4]). *Si (f_n) est une suite croissante de fonctions mesurables positives étendues sur E,*

$$\lim\uparrow \int f_n\,d\mu = \int \lim\uparrow f_n\,d\mu.$$

La place et l'accord des mots sont essentiels : l'énoncé devient faux et perd même son sens si l'on suppose avoir "une suite de fonctions mesurables positives étendues croissantes" (alors $\lim f_n$ n'existe pas en général) ; c'est bien la suite, et non chacune des fonctions, qui doit être croissante.

Le théorème de convergence monotone est capital dans la théorie. La version ci-dessus n'est pas la plus générale, mais permet déjà de calculer des intégrales, comme le montre la remarque suivante (voir une mise en application dans l'exercice 7.9).

Remarque 7.6. Si f est une fonction mesurable positive étendue, il existe une suite croissante (f_n) de fonctions étagées convergeant vers f. En effet, il suffit de poser

$$f_n(x) = \max\left\{\frac{j}{2^n};\ j\in\mathbb{N},\ \frac{j}{2^n}\leqslant \min(f(x), 2^n)\right\}$$

autrement dit, de définir $f_n(x)$ comme le plus grand multiple de 2^{-n} majoré à la fois par $f(x)$ et par 2^n (voir la figure 12).

Les fonctions ainsi définies sont bien étagées positives puisque, n étant fixé, f_n prend ses valeurs dans l'ensemble fini

$$\left\{\frac{0}{2^n}, \frac{1}{2^n}, \ldots, \frac{2^{2n}}{2^n} = 2^n\right\}.$$

3. On dira bientôt : f nulle presque partout.
4. Beppo Levi, mathématicien italien (1875–1961)

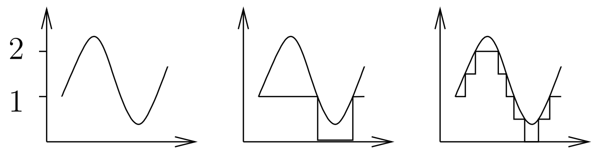


FIGURE 12 – Une fonction f et ses deux premières approximations étagées

De plus, la suite (f_n) est croissante. En effet, posons

$$A_n = \left\{ \frac{j}{2^n};\ j \in \mathbb{N},\ \frac{j}{2^n} \leqslant \min(f(x), 2^n) \right\}.$$

Comme tout nombre $j/2^n$ peut s'écrire $(2j)/2^{n+1}$,

$$\begin{aligned} A_n &= \left\{ \frac{j}{2^n};\ j \in \mathbb{N} \right\} \cap [0, f(x)] \cap [0, 2^n] \\ \subset A_{n+1} &= \left\{ \frac{j}{2^{n+1}};\ j \in \mathbb{N} \right\} \cap [0, f(x)] \cap [0, 2^{n+1}]. \end{aligned}$$

Comme de plus $f_n(x) = \max A_n$ et $f_{n+1}(x) = \max A_{n+1}$, on a bien $f_n(x) \leqslant f_{n+1}(x)$. Enfin, pour tout $x \in E$, $(f_n(x))$ converge vers $f(x)$. En effet, si $f(x) = +\infty$, $f_n(x) = 2^n \to +\infty$. Et sinon, à partir d'un certain rang, $2^n \geqslant f(x)$ donc

$$f_n(x) = \max \left\{ \frac{j}{2^n};\ j \in \mathbb{N},\ \frac{j}{2^n} \leqslant f(x) \right\}$$

donc il existe j tel que

$$f_n(x) = \frac{j}{2^n} \leqslant f(x) \leqslant \frac{j+1}{2^n},$$

donc

$$0 \leqslant f(x) - f_n(x) \leqslant \frac{j+1}{2^n} - \frac{j}{2^n} = \frac{1}{2^n} \to 0.$$

Démonstration du théorème. Remarquons d'abord que

— la limite simple $f := \lim \uparrow f_n : E \to [0, +\infty]$ existe parce que la suite (f_n) est croissante,

— comme l'intégration est croissante, la suite $(\int f_n d\mu)$ de $[0, +\infty]$ est croissante, donc possède une limite $\lim \uparrow \int f_n \, d\mu$.

Comme $f_n \leqslant f$ pour tout n et comme l'intégration est croissante, on a

$$\int f_n d\mu \leqslant \int f d\mu \quad (\forall n),$$

donc

$$\lim \uparrow \int f_n \, d\mu \leqslant \int f \, d\mu.$$

Il reste à montrer l'inégalité inverse. D'après la définition de l'intégrale de f, il s'agit de montrer que, pour toute fonction h étagée positive majorée par f,

$$\int h \leqslant \lim \int f_n.$$

C'est une situation typique, en analyse, où cela simplifie de se donner « un peu d'espace »[5] en démontrant le résultat équivalent que, pour tout $a \in [0,1[$,

$$a \int h \leqslant \lim \int f.$$

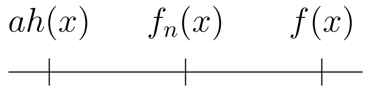
Soient donc h une fonction étagée positive :

$$h = \sum_{y \in h(E)} y \, \mathbb{1}_{\{h^{-1}(y)\}}$$

telle que $h \leqslant f$, et $a \in [0,1[$. L'ensemble

$$E_n = \{x \in E, \, ah(x) \leqslant f_n(x)\}$$

est mesurable et, puisque $a < 1$, $E = \bigcup \uparrow E_n$:



De plus,

$$ah\mathbb{1}_{E_n} \leqslant f_n,$$

et donc

$$\int ah\mathbb{1}_{E_n} \, d\mu = a \sum_{y \in h(E)} y \, \mu(h^{-1}(y) \cap E_n) \leqslant \int f_n \, d\mu.$$

Or, $E_n \uparrow E$, donc $h^{-1}(y) \cap E_n \uparrow h^{-1}(y)$ et, d'après la propriété de continuité intérieure d'une mesure,

$$\mu\left(h^{-1}(y) \cap E_n\right) \uparrow \mu\left(h^{-1}(y)\right).$$

Donc, en passant à la limite quand n tend vers $+\infty$ dans l'inégalité précédente il vient

$$a \sum_{y \in h(E)} y \, \mu(h^{-1}(y)) = a \int h \, d\mu \leqslant \lim \uparrow \int f_n \, d\mu,$$

d'où la seconde inégalité voulue. □

Proposition 7.7. *Si f et g sont mesurables positives et si $a, b \in \mathbb{R}_+$,*

$$\int (af + bg) = a \int f + b \int g.$$

Démonstration. D'après la remarque 7.6, il existe des suites croissantes (f_n) et (g_n) de fonctions étagées positives convergeant respectivement vers f et g. Alors

$$\begin{aligned}
\int (af + bg) &= \lim \uparrow \int (af_n + bg_n) && \text{(théorème de convergence monotone)} \\
&= \lim \uparrow \left(a \int f_n + b \int g_n\right) && \text{(linéarité pour les fonctions étagées)} \\
&= a \int f + b \int g.
\end{aligned}$$

□

5. « ϵ of a room », ce qui a donné son titre à deux livres de T. Tao [Tao10a, Tao10b].

Exercice 7.1 (Intégration d'une série de fonctions positives)**.**
Montrer que si (f_n) est une suite de fonctions mesurables positives étendues,

$$\int \sum_n f_n \, d\mu = \sum_n \int f_n \, d\mu.$$

Définition 7.8. Un propriété dépendant de $x \in E$ est *vraie μ presque partout* (souvent abrégé en *μ-p.p.* ou même *p.p.*) si elle est vraie en dehors d'un ensemble négligeable (définition 4.2 du chapitre 4). Si μ est une probabilité, on dit aussi μ *presque sûrement* (μ-p.s.).

Par exemple, dire que $f = g$ p.p. signifie

$$\mu(\{x \in E; \ f(x) \neq g(x)\}) = 0.$$

Exercice 7.2 (Fonction continue nulle presque partout)**.**
1. Montrer que la mesure de Lebesgue *charge les ouverts* : si O est un ouvert non vide de $\mathbb{R}^d$, $\lambda(O) > 0$.
2. Montrer que, si $f : U$ ouvert $\subset \mathbb{R}^d \to \mathbb{R}$ est nulle λ-presque partout et continue, f est nulle partout.
3. En déduire que, si $f, g : U$ ouvert $\subset \mathbb{R}^d \to \mathbb{R}$ sont égales λ-presque partout et continues, $f = g$ partout.

Proposition 7.9. *Soient f et g deux fonctions mesurables positives.*

1. Inégalité de Markov[6] : *pour tout $a > 0$,*

$$\mu(\{x \in E; \ f(x) \geqslant a\}) \leqslant \frac{1}{a} \int f \, d\mu.$$

2. *Par ailleurs, on a :*

$$\begin{cases} \int f \, d\mu < \infty \Rightarrow f < +\infty \quad p.p. \\ \int f \, d\mu = 0 \Leftrightarrow f = 0 \ p.p. \\ f = g \quad p.p. \Rightarrow \int f \, d\mu = \int g \, d\mu. \end{cases}$$

La réciproque des deux implications qui ne sont pas signalées comme des équivalences est fausse (pourquoi ?).

Démonstration. Notons $A_a = \{x \in E; \ f(x) \geqslant a\}$ pour $a \in [0, +\infty]$.

1. On a $a\mathbb{1}_{A_a} \leqslant f$ (sur A_a comme sur son complémentaire), donc

$$a\mu(A_a) = \int a\mathbb{1}_{A_a} \, d\mu \leqslant \int f \, d\mu.$$

6. Andreï Andreïevitch Markov, mathématicien russe (1856–1922), célèbre pour ses travaux en probabilités et notamment sa découverte des chaînes de Markov, qui sont à la source du calcul stochastique.

2. 1) Supposons que $\int f < \infty$. Il s'agit de montrer que $\mu(A_\infty) = 0$. Alors, d'après l'inégalité précédente,

$$\mu(A_1) \leqslant \int f \, d\mu < \infty.$$

Donc

$$\begin{aligned}
\mu(A_\infty) &= \mu\left(\bigcap_{n\geqslant 1} A_n\right) \\
&= \lim_n \downarrow \mu(A_n) && \text{(continuité exérieure)} \\
&= \lim_n \downarrow \frac{1}{n}\int f \, d\mu && \text{(inégalité de Markov)} \\
&= 0.
\end{aligned}$$

2) Par ailleurs, on a déjà vu que, si $f = 0$ p.p., $\int f = 0$. Réciproquement, supposons que $\int f = 0$. Alors, d'après l'inégalité de la première partie de la proposition,

$$\mu(\{x \in E;\ f(x) > 0\}) = \mu\left(\bigcup_{n\geqslant 1} A_{1/n}\right) \leqslant \sum_{n\geqslant 1} \mu\left(A_{1/n}\right),$$

où

$$\mu\left(A_{1/n}\right) \leqslant n \int f \, d\mu = 0,$$

ce qui démontre l'implication voulue.
3) Supposons enfin que $f = g$ p.p., donc que $\max(f, g) = \min(f, g)$ p.p. Donc,

$$\int \max(f,g) = \int \min(f,g) + \int (\max(f,g) - \min(f,g)) = \int \min(f,g).$$

Puisque

$$\min(f,g) \leqslant f \leqslant \max(f,g)$$

et de même pour g, il en découle que

$$\int f = \int \max(f,g) = \int g.$$

□

Exercice 7.3 (Intégrale d'une fonction strictement positive)**.**
Montrer que l'intégrale d'une fonction mesurable et $\geqslant 0$ et > 0 sur un ensemble de mesure non nulle est > 0.

L'exercice suivant est particulièrement important.

Exercice 7.4 (Mesure à densité)**.**
Soient f mesurable positive et $f\mu : \mathcal{A} \to [0, +\infty]$ définie par

$$(f\mu)(A) := \int_A f \, d\mu := \int \mathbb{1}_A \, f \, d\mu.$$

1. Montrer que $f\mu$ est une mesure sur $(E, \mathcal{A})$; on l'appelle la *mesure de densité f par rapport à μ*.

Attention, cette mesure n'est pas la même chose que l'image $\varphi(\mu)$ de μ par une application φ (définition 6.6), bien que les notations soient proches !

2. Montrer que, si une partie $A \in \mathcal{A}$ est μ-négligeable, elle est $(f\mu)$-négligeable (on dit que $f\mu$ est *absolument continue* par rapport à μ et l'on note $f\mu \ll \mu$; cf. la définition 31.1).

3. Montrer que, si g est une fonction mesurable positive,

$$\int g \, d(f\mu) = \int fg \, d\mu.$$

4. Montrer que, si $f\mu$ est σ-finie (définition 13.3), la densité f de $f\mu$ est unique à un ensemble μ-négligeable près ; on pourra d'abord le démontrer dans le cas où $f\mu$ est finie.

5. Trouver un contre-exemple si $f\mu$ n'est pas σ-finie.

Fonctions intégrables

Soit $f : E \to \mathbb{R}$ une fonction mesurable. Rappelons que les *parties positive* et *négative* de f sont respectivement les fonctions mesurables

$$f^+ := \sup(f, 0) \quad \text{et} \quad f^- := \sup(-f, 0).$$

Les fonctions f^+ et f^- sont positives, même f^- malgré son nom ! Notons que

$$f = f^+ - f^- \quad \text{et} \quad |f| = f^+ + f^-.$$

Définition 7.10. La fonction f est *μ-intégrable* si

$$\int |f| \, d\mu < \infty.$$

Dans ce cas, l'*intégrale de f* est le nombre réel

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.$$

On note $\mathcal{L}^1(E, \mathcal{A}, \mu)$ (ou $\mathcal{L}^1$, etc. selon le contexte) l'ensemble des fonctions *finies* μ-intégrables.

Comme $0 \leqslant f^\pm \leqslant |f|$, $\int f^\pm < \infty$ et la définition de l'intégrale de f a bien un sens, comme différence de deux nombres réels (finis).

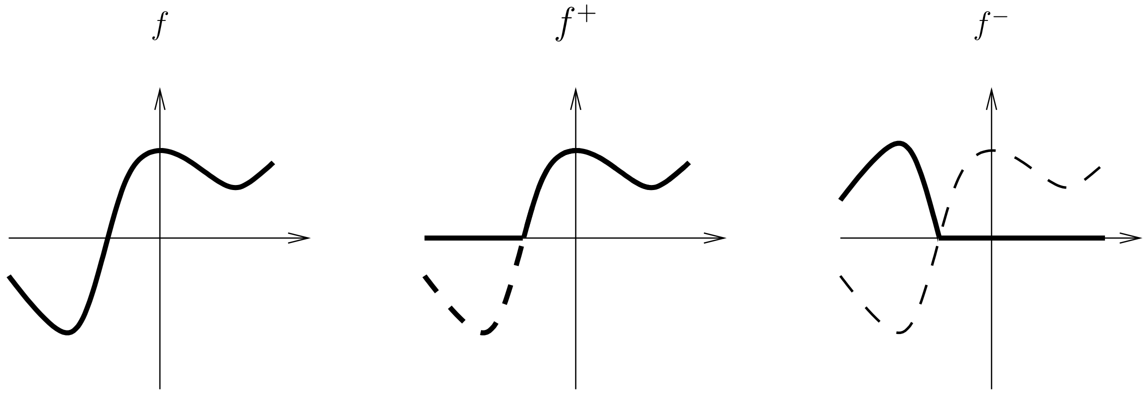


FIGURE 13 – Une fonction f et ses parties positive et négative.

Remarque 7.11. La terminologie est troublante : certaines fonctions ne sont pas qualifiées d'intégrables alors qu'elles ont une intégrale (alors infinie). On peut définir l'intégrale d'une fonctions f mesurable dans 3 cas sur 4 :

	$\int f^+ < \infty$	$\int f^+ = \infty$
$\int f^- < \infty$	$\int f \in \mathbb{R}$ f “intégrable”	$\int f = +\infty$
$\int f^- = \infty$	$\int f = -\infty$	$+\infty - \infty = ?$ $\int f$ non définie

Mais dans la suite, la théorie se concentre sur le cas où f est d'intégrale finie.

Propriétés 7.12. *1. Si $f \in \mathcal{L}^1$,*

$$\left|\int f\, d\mu\right| \leqslant \int |f|\, d\mu.$$

2. *$\mathcal{L}^1(E, \mathcal{A}, \mu)$ est une espace vectoriel réel, sur lequel l'application $f \mapsto \int f\, d\mu$ est une forme linéaire.*
3. *L'intégration est croissante : si $f, g \in \mathcal{L}^1$ et $f \leqslant g$, $\int f\, d\mu \leqslant \int g\, d\mu$.*
4. *Si $f, g \in \mathcal{L}^1$ et $f = g$ p.p., $\int f\, d\mu = \int g\, d\mu$.*

Démonstration. Les deux premiers points sont laissés en exercice : il suffit de revenir à la définition en terme d'intégrale des parties positive et négative.
Si f et g sont intégrables telles que $f \leqslant g$,

$$\int g = \int f + \int (g - f) \geqslant \int f.$$

Si f et g sont intégrables et égales p.p., f et g ont mêmes parties positive et négative, donc même intégrale. $\square$

Exercice 7.5 (Intégrale par rapport à la mesure de Dirac).
Montrer que, si $a \in E$ et si f est une fonction intégrable sur E,

$$\int f \, d\delta_a = f(a).$$

Exercice 7.6 (Intégration par rapport à la mesure de comptage).
Soit $(u_n)_{n\in\mathbb{N}}$ une suite réelle, vue comme une fonction mesurable sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}), c)$, où c est la mesure de comptage sur $\mathbb{N}$.
1. Montrer que, si (u_n) est positive,

$$\int u_n \, dc = \sum_{n\geqslant 0} u_n$$

(ce qui permet de transcrire beaucoup de résultat sur les intégrales, en termes de séries).
2. En déduire par exemple que si $(v_{n,m})_{(n,m)\in\mathbb{N}^2}$ est une suite double réelle positive,

$$\sum_n \sum_m v_{n,m} = \sum_m \sum_n v_{n,m}.$$

3. Montrer que (u_n) est intégrable par rapport à la mesure de comptage si et seulement si

$$\sum_n |u_n| < \infty,$$

auquel cas

$$\int u_n \, dc = \sum_n u_n = \sum_n u_n^+ - \sum_n u_n^-.$$

On dit alors que (u_n) est *sommable* au lieu de *intégrable.*

Exercice 7.7 (Règle de Chasles).
Si I est un intervalle de $\mathbb{R}$, $f \in \mathcal{L}^1(I)$ et $a, b, c \in I$, montrer cette version de la *règle de Chasles*[7] :

$$\int_a^c f = \int_a^b f + \int_b^c f.$$

Exercice 7.8 (Sur les fonctions intégrables sur $\mathbb{R}$).
1. Donner un exemple de fonction f positive intégrable sur $\mathbb{R}$, qui ne tende pas vers 0 en $+\infty$.

2. Montrer qu'une fonction f intégrable sur $\mathbb{R}$ et possédant une limite en $+\infty$ a forcément une limite nulle.

3. Montrer que, si f et g sont intégrables, $\max(f, g)$ est intégrable.

7. Michel Chasles, mathématicien français (1793–1880), connu pour ses travaux en géométrie projective et énumérative. Il a été victime d'une mystification célèbre, en acquérant une collection de lettres inédites prétendument de Pascal (établissant la découverte par Pascal de la loi de l'attraction universelle, antérieurement à Newton), d'Alexandre le Grand à Aristote, de Jules César à Vercingétorix, etc., en réalité toutes écrites dans un faux vieux français par le faussaire Vrain-Lucas.

Extension aux fonctions complexes

Soit $f : (E, \mathcal{A}) \to (\mathbb{C}, \mathcal{B}(\mathbb{C}))$ une fonction complexe mesurable (la mesurabilité équivaut au fait que les parties réelle et imaginaire de f sont mesurables).

Notons f_1 et f_2 les parties réelle et imaginaire de f. Remarquons que

$$\begin{aligned}\max(|f_1|, |f_2|) = \sqrt{\max(|f_1|^2, |f_2|^2)} &\leqslant |f| = \sqrt{|f_1|^2 + |f_2|^2} \\ &\leqslant |f_1| + |f_2|.\end{aligned}$$

Donc

$$\int |f| < \infty \Leftrightarrow \int |f_1| < \infty \quad \text{et} \quad \int |f_2| < \infty$$

soit,

$$|f| \in \mathcal{L}^1 \Leftrightarrow |f_1|, |f_2| \in \mathcal{L}^1. \tag{7.4}$$

Définition 7.13. La fonction f est *μ-intégrable* si

$$\int |f| \, d\mu < \infty.$$

Dans ce cas, l'*intégrale de* f est le nombre complexe

$$\int f \, d\mu = \int f_1 \, d\mu + i \int f_2 \, d\mu$$

(bien défini d'après (7.4)). On note $\mathcal{L}^1_{\mathbb{C}}(E, \mathcal{A}, \mu)$ l'ensemble des fonctions complexes *finies* intégrables.

Les propriétés 7.12 se généralisent directement au cas complexe.

Ce qui vient d'être fait pour les fonctions à valeurs dans $\mathbb{C} = \mathbb{R}^2$ se généralise *ispo facto* au cas de $\mathbb{R}^d$.

Formule d'intégration par rapport à une mesure image

Soient μ une mesure sur $(E, \mathcal{A})$ et $\varphi : (E, \mathcal{A}) \to (F, \mathcal{B})$ une application mesurable. Rappelons que la *mesure image* de μ par φ est la mesure notée $\varphi(\mu)$ (ou μ_φ, ou $\varphi_*(\mu)$, etc.) définie par

$$\varphi(\mu)(B) = \mu(\varphi^{-1}(B)), \quad B \in \mathcal{B}.$$

Soit aussi $f : (F, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ une fonction mesurable. Comme le diagramme suivant l'indique :

$$\begin{tikzcd} (E, \mu) \arrow[d, "\varphi"'] \arrow[dr, "f \circ \varphi"] & \\ (F, \varphi(\mu)) \arrow[r, "f"'] & \mathbb{R}, \end{tikzcd}$$

ces données permettent d'envisager d'intéger soit $f \circ \varphi$ par rapport à μ, soit f par rapport à $\varphi(\mu)$. Le théorème suivant dit que les deux opérations sont équivalentes.

Théorème 7.14 (Formule d'intégration par rapport à une mesure image). —
Si f est positive, on a l'égalité suivante, dans $\bar{\mathbb{R}}^+$:

$$\int_E f \circ \varphi \, d\mu = \int_F f \, d\varphi(\mu).$$

— *On a $f \circ \varphi \in L^1(\mu) \Leftrightarrow f \in L^1(f(\mu))$ et, si ces deux conditions équivalentes sont satisfaites, l'égalité ci-dessus est encore satisfaite, dans $\mathbb{R}$.*

Cette formule permet en théorie de se débarrasser de n'importe quelle fonction φ dans une intégrale $\int f \circ \varphi \, d\mu$ à calculer. La mesure image $\varphi(\mu)$ est parfois même mieux connue que μ elle-même, en probabilités. Un autre cas important où la mesure image est connue est celui où φ est un difféomorphisme entre deux ouverts de $\mathbb{R}^n$; on peut alors donner à la formule précédente une forme plus précise, appelée *formule du changement de variable* (chapitre 15).

Démonstration du théorème. Si $f = \mathbb{1}_B$ avec $B \in \mathcal{B}$,

$$\begin{aligned}
\int f \circ \varphi \, d\mu &= \int \mathbb{1}_B \circ \varphi \, d\mu, && (\text{où } \mathbb{1}_B \circ \varphi = \mathbb{1}_{\varphi^{-1}(B)}) \\
&= \mu(\varphi^{-1}(B)) && (\text{intégrale d'une fonction indicatrice}) \\
&= \varphi(\mu)(B) && (\text{définition de } \varphi(\mu)) \\
&= \int \mathbb{1}_B \, d\varphi(\mu) && (\text{intégrale d'une fonction indicatrice}).
\end{aligned}$$

La formule reste vraie, par linéarité, si f est une fonction mesurable étagée. Si f est mesurable positive, soit (f_n) une suite croissante de fonctions mesurables étagées positives qui tende vers f. Alors

$$\begin{aligned}
\int f \circ \varphi \, d\mu &= \int \lim \uparrow f_n \circ \varphi \, d\mu \\
&= \lim \uparrow \int f_n \circ \varphi \, d\mu && (\text{théorème de convergence monotone}) \\
&= \lim \uparrow \int f_n \, d\varphi(\mu) && (\text{d'après le cas précédent}) \\
&= \int f \, d\varphi(\mu) && (\text{théorème de convergence monotone}).
\end{aligned}$$

Considérons enfin une fonction f mesurable quelconque. La fonction $f \circ \varphi$ est intégrable par rapport à μ si et seulement si la quantité suivante est finie :

$$\begin{aligned}
\int |f \circ \varphi| \, d\mu &= \int |f| \circ \varphi \, d\mu \\
&= \int |f| \, d\varphi(\mu) && (\text{d'après le cas précédent}),
\end{aligned}$$

c'est-à-dire si f est intégrable par rapport à $\varphi(\mu)$. Dans ce cas, on obtient la formule en décomposant f en la différence de ses parties positive et négative, auxquelles on peut séparément appliquer la formule obtenue dans le cas précédent. □

Autres exercices

Exercice 7.9 (Exemple bête d'intégrale)**.**
Calculer l'intégrale de Lebesgue de la fonction

$$f : \begin{array}{ccc} [-2,2] & \to & \mathbb{R} \\ x & \mapsto & 1-|x| \end{array}$$

par rapport à la mesure de Lebesgue de $[-2,2]$, en n'utilisant que les définitions et le théorème de convergence monotone.

Exercice 7.10 (Inégalité de Jensen)**.**
Soient $(E, \mathcal{A}, \mu)$ un espace de probabilité, $f : E \to]a, b[$ $(-\infty \leqslant a < b \leqslant +\infty)$ une fonction intégrable relativement à μ et $\phi :]a, b[\to \mathbb{R}$ une fonction convexe.
1. Prouver l'*inégalité de Jensen* :[8]

$$\phi\left(\int f\, d\mu\right) \leqslant \int \phi(f)\, d\mu \ ;$$

on pourra utiliser le fait que le graphe d'une fonction convexe est l'enveloppe supérieure des fonctions affines qu'elle domine, et commencer par justifier l'existence des intérales apparaissant dans l'inégalité.
2. Écrire cette inégalité en termes de la probabilité image $f(\mu)$.
(C'est sous cette forme que l'inégalité de Jensen se comprend le mieux parce que μ et f n'y jouent séparément aucun rôle particulier : seule ν compte. En fait, la version véritablement géométrique, par opposition à fonctionnelle, de cette inégalité dit que le centre de masse d'une probabilité, quand il existe, est dans l'enveloppe convexe du support de la probabilité.)
3. En déduire l'*inégalité de Jensen finie*, obtenue dans le cas particulier où $E =]a, b[$, $\mathcal{A} = \mathcal{B}(]a, b[)$, $f(x) = x$ et $\mu = \sum_{i=1}^{n} \alpha_i \delta_{x_i}$, δ_x désignant la mesure de Dirac en x et les α_i étant des réels tels que $\sum_{i=1}^{n} \alpha_i = 1$.

Exercice 7.11 (Entropie d'une partition finie)**.**
Dans un article célèbre publié en 1948, Shannon[9] jeta les fondements d'une nouvelle discipline, la *théorie de l'information*, dans laquelle la notion d'*entropie* joue un rôle fondamental.

> My greatest concern was what to call it. I thought of calling it *information*. But the word was overly used, so I decided to call it *uncertainty*. When I discussed with John von Neumann, he had a better idea. He told me : « You should call it *entropy*, for two reasons. In the first place your uncertainty has been used in statistical mechanics under that name, so it already has a name. In second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage. » (Citation de C. Shannon, rapportée par M. Martin et J. England dans *Mathematical theory of entropy* (1981)

8. Johan Jensen, mathématicien danois (1859–1925)
9. Claude E. Shannon, mathématicien et ingénieur électricien américain (1916-2001)

L'entropie peut être définie à plusieurs niveaux. Le plus basique est celui des partitions finies. Soit donc une expérience aléatoire avec un nombre fini de résultats possibles, formant une partition mesurable finie $\mathcal{C}$ d'un espace de probabilité $(E, \mathcal{A}, \mu)$; mesurable signifie que $\mathcal{C} \subset \mathcal{A}$. On veut quantifier l'information $H(\mathcal{C})$ que cette expérience nous apporte : quasi-nulle pour une partition $\mathcal{C} = \{A, B\}$ avec

$$\mu(A) = 10^{-10} \quad \text{et} \quad \mu(B) = 1 - \mu(A)$$

(à la limite, si $\mu(A) = 0$, on sait juste que $\omega \in B$ p.s., ce qui est une information triviale puisqu'on sait a priori que $\omega \in \Omega$) et grande si

$$\mu(A) = \mu(B) = \frac{1}{2}.$$

Remarquons d'ailleurs que l'information apportée par le résultat de l'expérience correspond à l'incertitude avant que l'expérience n'ait eu lieu ; c'est la raison pour laquelle Shannon a hésité entre des termes aux sens aussi opposés que ceux d'*information* et d'*incertitude*.

Nous allons voir que les propriétés générales que l'on souhaiterait naturellement voir vérifiées par H déterminent quasiment H. Cherchons donc une expression convenable de H, comme une fonction des probabilités $\mu(A)$, $A \in \mathcal{C}$. Considérons d'abord le cas équiprobable où, pour tout $A \in \mathcal{C}$,

$$\mu(A) = \frac{1}{n}, \quad n = \sharp\mathcal{C},$$

et notons $f(n) = H(\mathcal{C})$ la fonction de n ainsi obtenue. Il est naturel d'imposer :

(E_1) La fonction f est croissante,

puisque par exemple la ville dans laquelle quelqu'un habite (sur Terre) contient plus d'information que son simple pays.

Considérons maintenant deux expériences indépendantes, la première avec n résultats possibles équiprobables, et la seconde avec m résultats possibles équiprobables. L'indépendance des deux expériences signifie que la connaissance de l'issue de la première expérience ne modifie pas l'incertitude (ou n'apporte aucune information) concernant l'issue de la seconde. Autrement dit,

(E_2) $f(nm) = f(n) + f(m)$.

(Le membre de droite est bien une somme, et non un produit : si par exemple la seconde expérience est triviale i.e. $f(m) = 0$, l'incertitude de la réunion des deux expériences combinées doit être celle de la première.)

Soit maintenant une expérience aboutissant à n issues possibles, que l'on regroupe en deux paquets $\mathcal{C}_1$ et $\mathcal{C}_2$ ($\{\mathcal{C}_1, \mathcal{C}_2\}$ est donc une partition de $\mathcal{C}$). Notons $A_1 = \cup_{A \in \mathcal{C}_\infty} A$ et $A_2 = \cup_{A \in \mathcal{C}_2} A$. Il est naturel d'estimer l'incertitude totale comme la somme de trois termes :

— l'incertitude de réaliser A_1 ou A_2
— l'incertitude de réaliser l'un des $A \in \mathcal{C}_1$, conditionnellement à la connaissance de A_1
— l'incertitude de réaliser l'un des $A \in \mathcal{C}_2$, conditionnellement à la connaissance de A_2.

Ceci se traduit par la formule :

$(E_3)\ H(\mathcal{C}) = H(\{A_1, A_2\}) + \mu(A_1)H(\mathcal{C}_1) + \mu(A_2)H(\mathcal{C}_2).$

1. ★ En supposant que les axiomes (E_1), (E_2) et (E_3) sont satisfaits, et que la fonction H dépend continûment des probabilités $\mu(A) \in]0,1[$, $A \in \mathcal{A}$, montrer qu'il exsiste $c > 0$ tel que, pour toute partition finie $\mathcal{C}$,

$$H(\mathcal{C}) = -c \sum_{A \in \mathcal{C}} \mu(A) \ln \mu(A) \in]0, +\infty[$$

avec la convention, pour chaque terme, $0 \ln 0 = 0$.

L'*entropie* de $\mathcal{C}$ est cette quantité $H(\mathcal{C})$ avec $c = 1$:

$$H(\mathcal{C}) = -\sum_{A \in \mathcal{C}} \mu(A) \ln \mu(A) \in]0, +\infty[\,.$$

2. Interpréter $H(\mathcal{C})$ comme l'intégrale sur E d'une certaine fonction $I_{\mathcal{C}}$ associée à $\mathcal{C}$, mesurable et à valeurs dans $[0, +\infty]$; $I_{\mathcal{C}}$ est la *fonction d'information* de $\mathcal{C}$. En déduire une définition de l'entropie d'une partition mesurable quelconque.

3. Le nombre fini $n \in \mathbb{N}_*$ de parties $A \in \mathcal{C}$ étant fixé, montrer que l'entropie de $\mathcal{C}$ est maximale si toutes les parties $A \in \mathcal{C}$ ont même mesure ; on pourra utiliser l'inégalité de Jensen finie (exercice 7.10).

Problème 7.1 (Fonctions de première et deuxième classes [Cho01]).
Soit E un espace métrique complet. Une fonction $f : E \to \mathbb{C}$ est *de première classe* si elle est la limite simple d'une suite de fonctions continues ; elle est *de deuxième classe* si elle est la limite simple d'une suite de fonctions de première classe.

1. Vérifier que l'ensemble $\mathcal{F}_1$ des fonctions de première classe est un espace vectoriel stable par les opérations $f \mapsto |f|$, $\bar{f}$, $1/f$ (pour les f qui ne s'annulent pas), ainsi que $(f, g) \mapsto fg$, $\min(f, g)$ et $\max(f, g)$ (pour les f réelles). Montrer qu'il en est de même pour l'ensemble $\mathcal{F}_2$ des fonctions de deuxième classe.

2. Montrer qu'une fonction de première classe possède des points de continuité ; on pourra utiliser la *propriété de Baire** de E, à savoir que l'intersection dénombrable d'ouverts denses de E est dense dans E.

3. Montrer que la fonction de Dirichlet $\mathbb{1}_{\mathbb{Q}}$ est de deuxième classe, mais pas de première classe.

En modifiant la fonction de Dirichlet sur un ensemble négligeable, on peut la rendre continue (en l'occurence, $\mathbb{1}_{\mathbb{Q}} = 0$ p.p.), donc de première classe. On peut cependant montrer qu'il existe des fonctions qu'on ne peut pas rendre de première classe par modification sur un ensemble négligeable. En revanche il n'en est pas de même avec la deuxième classe :

4. Montrer que, pour toute fonction $f \in \mathcal{L}^1(\mathbb{R})$, il existe une fonction g de deuxième classe telle que $f = g$ presque partout (Vitali).

De façon analogue (au prix d'une récurrence transfinie), on peut définir l'ensemble $\mathcal{F}_\alpha$ des fonctions de classe α pour tout nombre ordinal α, et montrer, à l'aide de l'exercice 9.1, que l'ensemble des fonctions borélienne n'est autre que la réunion des $\mathcal{F}_\alpha$.

Exercice 7.12 (Convergence en probabilité)**.**
Soient $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ et (A_n) une suite d'événements telle que $P(A_n)$ tend vers 0. Montrer que $E(X\mathbb{1}_{A_n})$ tend vers 0.

L'extension de l'intégrale aux fonctions à valeurs dans $\mathbb{C} = \mathbb{R}$ ou dans $\mathbb{R}^n$ se fait sans difficulté : il suffit de raisonner composante par composante. En revanche, l'extension aux fonctions à valeurs dans un espace vectoriel de dimension infinie est plus problématique. L'exercice suivant propose une telle extension.

Problème ★ 7.2 (Intégrale de Bochner [BJ06])**.**
Soient $(F, \|\cdot\|)$ un espace de Banach séparable * muni de sa tribu borélienne $\mathcal{B}(F)$, et F' le dual topologique de F.

1. Montrer qu'il existe une suite de fonctions boréliennes étagées $\phi_n : F \to F$ $(n \in \mathbb{N})$ telle qu'on ait la propriété suivante d'approximation de l'identité :

$$\|\phi_n(u) - u\| \to_n 0 \quad (\forall u \in F).$$

Montrer qu'on peut de plus construire ϕ_n de sorte que

$$\|\phi_n(u)\| \leqslant \|u\| \quad (\forall n \in \mathbb{N}, u \in F).$$

Soient $(E, \mathcal{A}, \mu)$ un espace mesuré et $\mathcal{L}^1_F = \mathcal{L}^1_F(E, \mathcal{A}, \mu)$ l'ensemble des fonctions mesurables $f : E \to F$ telles que

$$\int \|f\|\, d\mu < \infty.$$

2. Montrer que $\mathcal{L}^1_F$ est un espace vectoriel.

3. Montrer que, pour toute $f \in \mathcal{L}^1_F$, il existe un unique élément de F, appelé *intégrale de Bochner*[10] de f par rapport à μ et noté $\int f\, d\mu$, tel que, pour toute forme linéaire $\ell \in F'$,

$$\ell\left(\int f\, d\mu\right) = \int \ell \circ f\, d\mu;$$

on pourra commencer par définir cette intégrale quand f est étagée, et nulle en dehors d'une partie mesurable de E de mesure finie, puis prolonger cette définition en utilisant une approximation étagée de l'identité.[11]

4. Montrer que l'application d'intégration ainsi définie $\mathcal{L}^1_F \to F$ est linéaire et que

$$\left\|\int f\, d\mu\right\| \leqslant \int \|f\|\, d\mu.$$

10. Salamon Bochner, mathématicien américain d'origine austro-hongroise (1899–1982), connu pour ses travaux en analyse, théorie des probabilités et géométrie différentielle.
11. Cette construction est proche de celle de l'intégrale de Daniell.

Figure 14 – Henri Lebesgue (1875–1941)

Chapitre 8

Théorèmes de convergence

Soit $(E, \mathcal{A}, \mu)$ un espace mesuré. Dans ce chapitre, on s'intéresse à la question de savoir, si (f_n) est une suite de fonctions convergeant simplement, si l'on peut intervertir les opérations d'intégration et de limite :

$$\lim \int f_n \, d\mu = \int \lim f_n \, d\mu \ ? \tag{8.1}$$

Cette égalité n'est certainement pas vraie en général

Exemple 8.1 (Évasion à l'infini). Si $f_n = \mathbb{1}_{[n,n+1]}$ sur $\mathbb{R}$,

$$\lim \int f_n \, dx = \lim 1 = 1 \neq \int \lim f_n \, dx = 0.$$

En quelque sorte, la masse des f_n part à l'infini, n'en laissant aucune à la limite simple des f_n.

Exemple 8.2 (Élargissement infini). Si $f_n = \frac{1}{n}\mathbb{1}_{[0,n]}$, l'égalité (8.1) n'est pas satisfaite non plus bien que les f_n convergent uniformément vers 0, la masse des f_n se dispersant à l'infini sur un ensemble de mesure infinie.

Exemple 8.3 (Ascension infinie). Si $f_n = n\mathbb{1}_{[1/n,2/n]}$, l'égalité (8.1) n'est pas satisfaite non plus, la masse des f_n s'échappant verticalement.

Les hypothèses du théorème de convergence dominée ci-dessous permettront précisément d'éviter ces phénomènes. Avant cela, voyons deux théorèmes extrêmement utiles, plus généraux.

Rappelons d'abord le théorème de convergence monotone, démontré dans le chapitre précédent.

Théorème 8.4 (de convergence monotone)**.** *Si (f_n) est une suite croissante de fonctions mesurables positives sur E,*

$$\lim \uparrow \int f_n \, d\mu = \int \lim \uparrow f_n \, d\mu.$$

Exercice 8.1 (Variantes du théorème de convergence monotone)**.**
Les deux premières questions ont pour objectif d'étendre la notion de linéarité (de

la limite et de l'intégrale) aux sommes de deux termes dont l'un (mais seulement l'un) peut être infini.

1. Soient (u_n) et (v_n) deux suites de $\bar{\mathbb{R}}$ telles que

— $\lim u_n$ existe dans $[-\infty, +\infty]$

— $\lim v_n$ existe dans $\mathbb{R}$.

Montrer que
$$\lim(u_n + v_n) = \lim u_n + \lim v_n$$
(avec la convention naturelle que $\pm\infty + \ell = \pm\infty$ pour tout réel ℓ).

2. Soient $f, g : E \to \bar{\mathbb{R}}$ mesurables telles que

— f possède une intégrale $\in [-\infty, +\infty]$ (au sens que l'un au moins des deux nombres $\int f^+$ ou $\int f^-$ est fini)

— g est intégrable (au sens habituel que $\int |g| < \infty$, ce qui implique $\int g \in \mathbb{R}$).

Montrer que
$$\int (f + g) = \int f + \int g$$
(avec la même convention arithmétique que précédemment).

3. Montrer que, si (f_n) est une suite croissante de fonctions mesurables minorées par une fonction intégrable g, la conclusion du théorème de convergence monotone est vraie.

4. Que dire d'une suite décroissante de fonctions mesurables négatives ? D'une suite décroissante de fonctions mesurables majorée par une fonction g donnée qui soit intégrable ?

Le théorème de convergence monotone possède une version très utile où l'on ne suppose plus que la suite de fonctions f_n est croissante. (Attention, on ne peut plus invoquer la limite simple des f_n ou des $\int f_n$ parce que ces limites n'existent pas en général. Voir les notions de limite inférieure ou supérieure dans le glossaire.)

Théorème 8.5 (Lemme de Fatou[1])**.** *Si (f_n) est une suite de fonctions mesurables positives sur E,*
$$\int (\liminf f_n)\, d\mu \leqslant \liminf \int f_n\, d\mu.$$

Informellement, le lemme de Fatou dit que quand on prend la limite inférieure d'une suite de fonctions, la masse $\int f_n\, d\mu$ peut être détruite (comme dans les trois exemples du début du chapitre), mais pas créée.

Démonstration. On a
$$\liminf_n f_n = \lim_n \uparrow \left(\inf_{k \geqslant n} f_k \right),$$
donc, d'après le théorème de convergence monotone,
$$\int (\liminf f_n)\, d\mu = \lim_n \uparrow \int \left(\inf_{k \geqslant n} f_k \right)\, d\mu.$$

1. Pierre Joseph Louis Fatou, mathématicien et astronome français (1878–1929), célèbre pour ses nombreuses contributions en analyse, ainsi qu'en dynamique complexe.

Par ailleurs, pour tout $l \geqslant n$,
$$\inf_{k\geqslant n} f_k \leqslant f_l,$$
donc
$$\int \left(\inf_{k\geqslant n} f_k\right) d\mu \leqslant \inf_{l\geqslant n} \int f_l \, d\mu.$$
Les deux membres de cette inégalité sont croissants par rapport à n, et possède donc une limite quand n tend vers l'infini, de sorte que
$$\lim \uparrow \int \left(\inf_{k\geqslant n} f_k\right) d\mu \leqslant \lim_n \uparrow \inf_{l\geqslant n} \int f_l \, d\mu = \liminf \int f_n \, d\mu.$$
□

Exercice 8.2 (Variantes du lemme de Fatou)**.**
1. Montrer que la conclusion du lemme de Fatou reste vraie si l'on suppose seulement que (f_n) est une suite de fonctions mesurables minorées par une fonction intégrable g (le lemme de Fatou correspondant donc au cas particulier où $g = 0$).
2. Montrer que, si (f_n) est une suite de fonctions mesurables dominée par une fonction intégrable g, i.e. $|f_n| \leqslant g$ pour tout n,
$$\int \liminf f_n \leqslant \liminf \int f_n \leqslant \limsup \int f_n \leqslant \int \limsup f_n.$$

Le lemme de Fatou lui-même permet de démontrer le théorème suivant, dû à H. Lebesgue, où l'on ne suppose ni que la suite de fonctions est croissante, ni que les fonctions sont positives, mais où l'on fait deux hypothèses de remplacement : une hypothèse naturelle de convergence, et une hypothèse de domination permettant d'éviter le phénomène de fuite de masse à l'infini.

Théorème 8.6 (de convergence dominée)**.** *Si (f_n) une suite de fonctions, réelles ou complexes, mesurables, telle que*

1. *il existe une fonction mesurable f telle que*
$$f_n(x) \to f(x) \quad \mu\text{-}p.p.$$
2. *et il existe une fonction g positive intégrable qui* domine *les f_n p.p. : pour tout n,*
$$|f_n(x)| \leqslant g(x) \quad \mu\text{-}p.p.,$$

la fonction f est intégrable ($f \in \mathcal{L}^1$) et
$$\lim \int f_n \, d\mu = \int f \, d\mu \quad et \quad \lim \int |f - f_n| \, d\mu = 0.$$

Remarque 8.7. Dans l'hypothèse de domination, l'ensemble négligeable N_n en dehors duquel on a $|f_n(x)| \leqslant g(x)$ peut a priori dépendre de n, mais, l'union dénombrable d'ensembles négligeables étant elle-même négligeable, il revient au même de supposer que N_n ne dépend pas de n.

Démonstration. Remarquons d'abord que, d'après l'hypothèse de domination, pour tout n, f_n est intégrable et, par passage à la limite, $|f| \leqslant g$ donc f elle-même est intégrable, donc $|f - f_n| \leqslant |f| + |f_n|$ aussi. Donc les intégrales apparaissant dans l'énoncé sont définies et finies.

Par ailleurs, la seconde égalité à démontrer implique la première :

$$\left|\int f\,d\mu - \int f_n\,d\mu\right| = \left|\int (f - f_n)\,d\mu\right| \leqslant \int |f - f_n|\,d\mu \to 0.$$

Il s'agit donc de montrer que

$$\lim \int |f - f_n|\,d\mu = 0.$$

Commençons par supposer les hypothèses plus fortes suivantes :

1. il existe une fonction mesurable f telle que $f_n \to f$ simplement sur E
2. et il existe une fonction g positive intégrable qui domine (f_n) partout sur E : $|f_n| \leqslant g$.

On a

$$0 \leqslant |f - f_n| \leqslant |f| + |f_n| \leqslant 2g.$$

Donc, d'après la variante du lemme de Fatou pour les suites de fonctions dominées par une fonction intégrable (exercice 8.2),

$$0 \leqslant \liminf \int |f - f_n| \leqslant \limsup \int |f - f_n| \leqslant \int \limsup |f - f_n| = 0.$$

Donc

$$\liminf \int |f - f_n| = \limsup \int |f - f_n| = 0,$$

donc $\int |f - f_n| \to 0$.

Revenons aux hypothèses plus faibles de l'énoncé. Notons

$$A_1 = \{x \in E;\ f_n(x) \to f(x)\}$$

et,

$$A_2 = \{x \in E;\ \forall n,\ |f_n(x)| \leqslant g(x)\}.$$

Par hypothèse, le complémentaire de ces ensembles est négligeable. Sur l'ensemble des « bons $x \in E$ »

$$B = A_1 \bigcap A_2,$$

on a simultanément

$$\begin{cases} f_n(x) \to f(x) \\ |f_n(x)| \leqslant g(x) \quad (\forall n), \end{cases}$$

et, puisque $B^c = A_1^c \cup A_2^c$, le complémentaire de B aussi est négligeable :

$$\mu(B^c) \leqslant \mu(A_1^c) + \mu(A_2^c) = 0.$$

On peut appliquer la première partie de la démonstration aux fonctions $\mathbb{1}_B f_n$ et $\mathbb{1}_B f$, qui sont presque partout égales respectivement aux f_n et à f. Les intégrales ne sont pas modifiées par la multiplication des fonctions par $\mathbb{1}_B$, qui vaut 1 p.p., donc le résultat en découle. $\square$

Autres exercices

Exercice 8.3 (Une application du théorème de convergence monotone).
Soit $f : [0,1] \to \mathbb{R}$ une fonction mesurable. Déterminer la limite de

$$\int_0^1 \frac{dt}{\sqrt{f(t)^2 + 1/n}}.$$

Exercice 8.4 (Inégalité de Fatou stricte).
Notons λ la mesure de Lebesgue de l'intervalle $[-1,1]$. Soit $(f_n)_{n\in\mathbb{N}}$ la suite de fonctions définie par

$$f_n = \begin{cases} \mathbb{1}_{[0,1]} & \text{si } n \text{ pair} \\ \mathbb{1}_{[-1,0]} & \text{si } n \text{ impair} \end{cases}$$

Montrer que

$$\int \left(\liminf_{n\to+\infty} f_n \right) d\lambda < \liminf_{n\to+\infty} \int f_n \, d\lambda.$$

Exercice 8.5 (Une application du théorème de convergence dominée).
Soient a et b deux réels tels que $a < b$, et $f :]a,b[\to \mathbb{R}$ une fonction mesurable bornée possédant une limite $\ell \in \mathbb{R}$ en a^+. Montrer que, pour tout $t \in]a,b[$, la fonction $x \mapsto f(x)/\sqrt{(x-a)(t-x)}$ est intégrable sur $]a,t[$, puis calculer

$$\lim_{t\to a^+} \int_{]a,t[} \frac{f(x)}{\sqrt{(x-a)(t-x)}} \, dx;$$

on pourra admettre que la fonction $\frac{1}{\sqrt{(x-a)(t-x)}}$ est intégrable sur l'intervalle $[a,t]$.

Exercice 8.6 (Constante γ d'Euler, *rattrapage 2013*).
1. Montrer que la suite

$$c_n = \left(1 + \frac{1}{2} + \cdots + \frac{1}{n} \right) - \ln n$$

converge vers un nombre γ, appelé *constante d'Euler*[2].

À cause de la lenteur de la convergence de la suite (c_n), on connaît très mal la constante d'Euler. Par exemple, on ne sait même pas si elle est irrationnelle.

2. Montrer que

$$\int_1^n \left(\frac{1}{[x]} - \frac{1}{x} \right) dx \to \gamma,$$

où $[x]$ désigne la partie entière de x, et en déduire que $\gamma > 0$.

3. Montrer que

$$\gamma = -\int_{\mathbb{R}^+} e^{-x} \ln x \, dx;$$

on pourra calculer de deux façons différentes la limite de $\int_0^n \left(1 - \frac{x}{n}\right)^n \ln x \, dx$ quand n tend vers l'infini.

---

2. Leonhard Euler, éminent mathématicien et physicien suisse (1707–1783), qui fit d'importantes découvertes en calcul infinitésimal, analyse, mécanique, astronomie.

Exercice 8.7 (Interversions d'une somme et d'une intégrale)**.**
1. Soit a et b deux réels strictement positifs. Pour $x \in \mathbb{R}_*^+$, soit

$$f(x) = \frac{xe^{-ax}}{1 - e^{-bx}}.$$

Montrer que

$$\int_{\mathbb{R}^+} f\,dx = \sum_{n=0}^{\infty} \frac{1}{(a+nb)^2}.$$

2. Soit μ une mesure sur $\mathbb{R}$ telle que la fonction $x \mapsto e^{x^2}$ soit μ-intégrable. Donner un exemple d'une telle mesure μ, puis montrer que pour tout nombre complexe z on a

$$\sum_{n=0}^{\infty} \frac{z^n}{n!} \int_{\mathbb{R}} x^n\,d\mu(x) = \int_{\mathbb{R}} e^{zx}\,d\mu(x).$$

3. Montrer que la fonction $f : [1, +\infty[\to \mathbb{R}$ définie par

$$f(x) = \sum_{n=1}^{\infty} ne^{-nx}$$

est intégrable sur $[1, +\infty[$ relativement à la mesure de Lebesgue, et calculer son intégrale.

Exercice 8.8 (Calcul d'une intégrale gaussienne, *partiel 2013*)**.**
On veut retrouver l'égalité

$$\int_0^{\infty} e^{-u^2}\,du = \frac{\sqrt{\pi}}{2}.$$

Soit f la fonction définie sur $\mathbb{R}_+$ par

$$f(x) = \left(\int_0^x e^{-u^2}\,du\right)^2 + \int_0^1 \frac{e^{-x^2(1+t^2)}}{1+t^2}\,dt.$$

1. Montrer que $f(0) = \pi/4$ et que $\lim_{+\infty} f = \left(\int_0^{\infty} e^{-u^2}\,du\right)^2$.

2. Montrer que, si $g : \mathbb{R} \to \mathbb{R}$, $x \mapsto \int_0^x e^{-u^2}\,du$, le taux d'accroissement $(g(x+\xi) - g(x))/\xi$ a une limite quand ξ tend vers 0. En déduire que g est dérivable sur $\mathbb{R}$ et calculer sa dérivée.

3. Montrer que $h : \mathbb{R} \to \mathbb{R}$, $x \mapsto \int_0^1 \frac{e^{-x^2(1+t^2)}}{1+t^2}\,dt$ est dérivable sur $\mathbb{R}$ et calculer sa dérivée ; on pourra admettre (ce qui sera démontré ultérieurement) que h est dérivable et que sa dérivée s'obtient en dérivant sous l'intégrale. Conclure.

Problème ★ 8.1 (Formules de la moyenne, *examen 2012*)**.**
Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, f une fonction réelle mesurable sur Ω à valeurs dans un intervalle $[u, v]$ et g une fonction réelle intégrable sur Ω.

1. Montrer que fg est intégrable et que, si g est positive, il existe $k \in [u, v]$ tel que

$$\int_{\Omega} f\,g\,d\mu = k \int_{\Omega} g\,d\mu \quad (\textit{première formule de la moyenne}).$$

Supposons de plus que Ω est un intervalle $[a, b]$ de $\mathbb{R}$, que μ est la mesure de Lebesgue, et que f est positive décroissante. On veut montrer qu'il existe $\xi \in [a, b]$ tel que

$$\int_{[a,b]} f(x)\, g(x)\, dx = f(a) \int_a^\xi g(x)\, dx \quad (\textit{seconde formule de la moyenne}).$$

2. Montrer que $G(\xi) := \int_a^\xi g(x)\, dx$ est continue et que $G([a, b])$ est un intervalle $[m, M]$.

3. En supposant f étagée, montrer qu'il existe une subdivision $a = a_0 < a_1 < \cdots < b = a_n$ de l'intervalle $[a, b]$ telle que, sur chaque intervalle $]a_i, a_{i+1}[$, la fonction f est constante égale à un réel α_i ; en déduire que

$$\int_{[a,b]} f(x)\, g(x)\, dx = \sum_{1 \leqslant i \leqslant n-1} (\alpha_{i-1} - \alpha_i) G(a_i) + \alpha_{n-1} G(a_n);$$

puis que la seconde formule de la moyenne est satisfaite pour f.

4. Montrer la seconde formule de la moyenne dans le cas général.

Soient $f : [0, +\infty[\to \mathbb{R}$ positive, décroissante et de limite nulle, et $g(x) = \sin x$.

5. Déduire de la seconde formule de la moyenne que $F(x) := \int_0^x f(t) \sin t\, dt$ possède une limite quand x tend vers $+\infty$ (on pourra montrer que, si (x_n) tend vers l'infini, $(F(x_n))$ est de Cauchy).

6. ★ Peut-on en déduire que la fonction $f(t) \sin t$ est intégrable sur $[0, +\infty[$?


FIGURE 15 – Leonhard EULER (1707–1783) et Pierre Joseph Louis FATOU (1878–1929). Euler est l'un des mathématiciens les plus éminents et prolifiques de tous les temps.

Chapitre 9

Classes monotones ★

Le théorème de classe monotone, démontré par W. Sierpiński[1] en 1928, est un outil commode pour construire des mesures.

Définition 9.1. Une *classe monotone*[2] d'un ensemble E est une classe $\mathcal{M}$ de parties de E telle que

— $E \in \mathcal{M}$

— $A, B \in \mathcal{M} \Rightarrow B \backslash A \in \mathcal{M}$ (stabilité par différence)

— $A_1, A_2, ... \in \mathcal{M}$ et $A_1 \subset A_2 \subset ... \Rightarrow \cup_n A_n \in \mathcal{M}$ (stabilité par union croissante dénombrable).

Une tribu est, en particulier, une classe monotone. Comme on le verra, si μ et ν sont deux mesures sur $(E, \mathcal{A})$, les parties mesurables de E telles que $\mu(A) = \nu(A)$ forment un autre exemple de classe monotone.

Comme pour les tribus, on voit que l'intersection d'une famille non vide quelconque de classes monotones est une classe monotone. Si donc $\mathcal{C}$ est une classe quelconques de parties de E, la *classe monotone engendrée* par $\mathcal{C}$ est

$$\mathcal{M}(\mathcal{C}) := \bigcap_{\mathcal{M} \text{ classe monotone} \supset \mathcal{C}} \mathcal{M}.$$

On a donc $\mathcal{M}(\mathcal{C}) \subset \sigma(\mathcal{C})$.

Théorème 9.2 (de classe monotone)**.** *Si $\mathcal{C} \subset \mathcal{P}(E)$ est stable par intersection finie,*[3] *$\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$.*

L'idée est que les classes de parties stables par intersection finie sont souvent plus pratiques à décrire que les tribus. Si une propriété est facile à vérifier sur une telle classe, le théorème de classe monotone permet, dans les bons cas, d'en déduire que cette propriété est vraie plus généralement sur la tribu engendrée.

Des exemples de classes de parties stables par intersection finie sont :

— la classe des intervalles de $\mathbb{R}$

1. Wacław Sierpiński, mathématicien polonais (1882–1969), connu pour ses travaux en théorie des ensembles, topologie et théorie des fonctions.

2. Le terme de *λ-système* est synonyme.

3. Une telle classe s'appelle un *π-système*.

— la classe des pavés $A \times B$ d'un espace produit $E \times F$ ($A \subset E$, $B \subset F$).

Démonstration du théorème. Puisque toute tribu est une classe monotone, tautologiquement on a

$$\mathcal{M}(\mathcal{C}) \subset \sigma(\mathcal{C}).$$

Montrons l'inclusion inverse, c'est-à-dire que $\mathcal{M}(\mathcal{C})$ est une tribu. Pour qu'une classe monotone soit une tribu, il faut et il suffit qu'elle soit en plus stable par intersection finie, parce qu'alors, par passage au complémentaire elle sera stable par union finie, puis par passage à la limite croissante elle sera stable par union dénombrable. Montrons donc que $\mathcal{M}(\mathcal{C})$ est stable par intersection finie.

Soit $A \in \mathcal{M}(\mathcal{C})$ fixé. Posons

$$\mathcal{M}_1 = \{B \in \mathcal{M}(\mathcal{C});\ A \cap B \in \mathcal{M}(\mathcal{C})\}.$$

Il s'agit de montrer que $\mathcal{M}(\mathcal{C}) \subset \mathcal{M}_1$.

Supposons pour commencer que $A \in \mathcal{C}$. Puisque $\mathcal{C}$ est stable par intersection finie, $\mathcal{C} \subset \mathcal{M}_1$. Il suffit ensuite de vérifier que $\mathcal{M}_1$ est une classe monotone. Or :

— $E \in \mathcal{M}_1$

— Stabilité par différence : Soient $B, B' \in \mathcal{M}_1$ tels que $B \subset B'$. On a

$$A \cap (B \backslash B') = (A \cap B') \backslash (A \cap B) \in \mathcal{M}(\mathcal{C}),$$

donc $B' \backslash B \in \mathcal{M}_1$.

— Stabilité par union croissante dénombrable : Si $B_n \in \mathcal{M}_1$ croît par rapport à n,

$$A \cap (\cup B_n) = \cup (A \cap B_n) \in \mathcal{M}(\mathcal{C})$$

donc $\cup B_n \in \mathcal{M}_1$.

Donc $\mathcal{M}_1$ est bien une classe monotone.

Supposons maintenant $A \in \mathcal{M}(\mathcal{C})$. D'après le cas particulier précédent, l'intersection d'un élément de $\mathcal{C}$ et d'un élément de $\mathcal{M}(\mathcal{C})$ est dans $\mathcal{M}_1$, donc $\mathcal{C} \subset \mathcal{M}_1$. Les mêmes arguments que dans le cas particulier précédent s'appliquent pour montrer que $\mathcal{M}_1$ est une classe monotone. Donc $\mathcal{M}(\mathcal{C}) \subset \mathcal{M}_1$, donc $\mathcal{M}(\mathcal{C})$ est stable par intersection finie. □

On en déduit facilement le résultat d'unicité de mesure suivant, qui sera appliqué pour démontrer l'unicité de la mesure de Lebesgue (sous deux caractérisations différentes : proposition 11.4 et théorème 11.9), des mesures produit (proposition 13.4), des lois conjointes (définition 13.6 et remarque suivante), des mesures de Lebesgue-Stieltjes (exercice 9.1), etc.

Corollaire 9.3. *Soient μ et ν deux mesures sur $(E, \mathcal{A})$. Supposons qu'il existe une classe $\mathcal{C} \subset \mathcal{A}$ vérifiant :*

— *$\mathcal{C}$ est stable par intersection finie et $\sigma(\mathcal{C}) = \mathcal{A}$*

— *$\mu(A) = \nu(A)$ pour tout $A \in \mathcal{C}$*

— *$\mu(E) = \nu(E) < \infty$ ou il existe une suite croissante de parties $E_n \in \mathcal{C}$ telle que $E = \cup E_n$ et $\mu(E_n) = \nu(E_n) < \infty$ pour tout n.*

Alors $\mu = \nu$.

Démonstration. Traitons d'abord le cas où $\mu(E) = \nu(E) < \infty$. Posons

$$\mathcal{G} = \{A \in \mathcal{A};\ \mu(A) = \nu(A)\}.$$

Par hypothèse, $\mathcal{C} \subset \mathcal{G}$ et on voudrait montrer que $\mathcal{G} = \mathcal{A}$. L'ensemble $\mathcal{G}$ est une classe monotone (exercice). Donc $\mathcal{G}$ contient $\mathcal{M}(\mathcal{C})$, dont le théorème de classe monotone affirme que c'est $\sigma(\mathcal{C})$, soit $\mathcal{A}$ par hypothèse. Donc $\mathcal{G} = \mathcal{A}$. C'est dire que $\mu = \nu$.

Dans le second cas, pour tout entier n notons μ_n et ν_n les restrictions respectives de μ et de ν à E_n. D'après ce qu'on vient de démontrer, $\mu_n = \nu_n$. D'après la propriété de continuité extérieure des mesures (chapitre 4), pour tout $A \in \mathcal{A}$,

$$\mu(A) = \lim \uparrow \mu(A \cap E_n) = \lim \uparrow \nu(A \cap E_n) = \nu(A).$$

$\square$

Exercice ★ 9.1 (Mesures de Lebesgue-Stieltjes)**.**
1. Montrer que la relation

$$\mu(]a,b]) = F(b) - F(a) \quad (-\infty < a < b < \infty) \tag{9.1}$$

définit une équivalence entre

— les mesures μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *localement finies*, c'est-à-dire telles que $\mu(I) < \infty$ pour tout intervalle borné I ;

— et les fonctions $F : \mathbb{R} \to [-\infty, +\infty]$ croissantes, continues à droites et telles que $F(0) = 0$.

Pour définir μ à partir de F, on pourra introduire l'inverse à droite de F ($F \circ G = \mathrm{id}$) défini comme la fonction $G : \mathbb{R} \to \bar{\mathbb{R}}$ telle que

$$G(t) = \inf\{s \in \mathbb{R};\ F(s) \geqslant t\}, \quad t \in \mathbb{R},$$

puis poser $\mu = G(\lambda)$. (Une autre solution consiste à utiliser le théorème de Hahn-Kolmogorov.)

Une fonction F comme ci-dessus étant donnée, on note μ_F la mesure localement finie qui lui est associée, dite *de Lebesgue-Stieltjes*, et

$$\int f\, dF := \int f\, d\mu_F.$$

(Attention à ne pas utiliser la notation

$$\int_a^b f\, dF \quad \text{pour} \quad \int_{[a,b]} f\, dF \quad \text{ou} \int_{]a,b[} f\, dF$$

si l'on n'est pas certain que F soit continue, sans quoi les deux dernières intégrales peuvent ne pas être égales.)

2. Montrer que, si $F : \mathbb{R} \to]-\infty, +\infty[$ est croissante, dérivable et telle que $F(0) = 0$,

$$\mu_F = F'(x)\, dx.$$

3. Montrer la formule d'intégration par parties : si F et G sont deux fonctions $\mathbb{R} \to]-\infty, +\infty[$ croissantes, continues à droite et nulles en 0 :

$$\int_{[a,b]} F\, dG + \int_{[a,b]} G\, dF = F(b^+)G(b^+) - F(a^-)G(a^-).$$

L'exercice suivant donne un analogue fonctionnel du théorème de classe monotone, dont on notera la ressemblance formelle avec le théorème de Stone-Weierstrass*.

Problème ★ 9.1 (Théorème de classe monotone fonctionnel [Bog07])**.**
Soit $\mathcal{F}$ une classe de fonctions sur un ensemble E, contenant la fonction constante égale à 1. Soient encore $\mathcal{F}_0$ une partie de $\mathcal{F}$ et $\mathcal{E}$ la tribu engendrée par $\mathcal{F}_0$.

1. Supposons que $\mathcal{F}$ est un sous-espace fermé de l'ensemble des fonctions bornéees sur E muni de la norme uniforme, que $\mathcal{F}$ soit stable par passage à la limite croissante de fonctions positives uniformément bornées, et que $\mathcal{F}_0$ soit stable par multiplication. Montrer que $\mathcal{F}$ contient toutes les fonctions bornées $\mathcal{E}$-mesurables.

2. Application : Soient μ et ν deux probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ telle que $\int f\, d\mu = \int f\, d\nu$ pour toute fonction $f : \mathbb{R}^d \to \mathbb{R}$ bornée de classe C^∞. Montrer que $\mu = \nu$.

Chapitre 10

Le problème de la mesure. Mesures extérieures ★

Le concept de mesure $\mu(S)$ d'un solide S est fondamental en géométrie euclidienne. En petite dimension, $\mu(S)$ s'appelle la longueur (dimension 1), l'aire (dimension 2) ou le volume (dimension 3) de S.

Par exemple, pour calculer l'aire d'un triangle T, on se ramène à l'aire d'un rectangle par la construction de la figure 16, qui permet de retrouver que l'aire est la moitié du produit de la base par la hauteur.

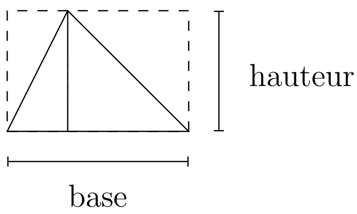


FIGURE 16 – L'aire du triangle est la moitié du produit de la base par la hauteur

De tels calculs peuvent être justifiés en faisant appel à l'intuition géométrique, ou bien en postulant l'existence d'une mesure $\mu(S)$ associée à tous les solides S, qui satisfasse certains axiomes géométriques raisonnables. L'avènement de la géométrie analytique, à l'époque de Descartes,[1] fit considérablement reculer les limites de l'horizon mathématique, puisqu'on commença à étudier des objets géométriques définis pas des équations. Et il ne fut plus évident de définir la mesure d'un sous-ensemble général de l'espace euclidien $\mathbb{R}^d$. C'est tout le problème de la mesure (ici décrit de façon assez vague) que de donner une telle définition.

Dans un chapitre antérieur, nous avons défini le concept de mesure mathématique de façon axiomatique, auxquels plusieurs siècles de recherches ont conduit. Nous allons maintenant voir de plus près comment le concept d'ensemble mesurable s'impose quand on essaye de construire une telle mesure.

Soit E un ensemble. Pour construire une certaine mesure sur E, nous allons partir

1. René DESCARTES, philosophe et mathématicien français (1596-1650). Son influence fut considérable. En particulier, il est considéré comme le père de la géométrie analytique.

d'un objet auxiliaire qui ressemble à une mesure, sans en vérifier tous les axiomes, et qui est défini sur $\mathcal{P}(E)$ entier.

Définition 10.1. Une *mesure extérieure* sur E est une fonction $\mu^* : \mathcal{P}(E) \to [0, +\infty]$ telle que

1. $\mu^*(\varnothing) = 0$
2. μ^* est croissante : $A \subset B \Rightarrow \mu^*(A) \leqslant \mu^*(B)$
3. μ^* est σ-sous-additive :
$$\mu^*\left(\bigcup A_k\right) \leqslant \sum \mu^*(A_k).$$

Nous verrons comment construire des mesures extérieures, notamment sur $\mathbb{R}$. Pour le moment, voyons comment construire à partir de μ^* une tribu $\mathcal{M}(\mu^*)$ (dépendant de μ^*), en restriction de qui μ^* est une mesure.
Si A et B sont deux parties quelconques de E, par sous-additivité on a toujours
$$\mu^*(A) \leqslant \mu^*(A \cap B) + \mu^*(A \cap B^c).$$
Pour une mesure, il faudrait avoir l'égalité puisque $A \cap B$ et $A \cap B^c$ sont des parties disjointes. Il revient à Carathéorody[2] d'avoir introduit la définition astucieuse suivante d'ensemble mesurable.

Définition 10.2. Une partie B de E est *μ^*-mesurable* si, pour toute partie A de E, on a
$$\mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c).$$
Notons $\mathcal{M}(\mu^*)$ l'ensemble des telles parties.

Autrement dit, que B soit μ^*-mesurable signifie que si l'on utilise B pour diviser une partie quelconque A en deux, on garde la propriété d'additivité. Si λ^* était finiment additive, tous les ensembles B seraient μ^*-mesurables. Mais on verra que ce n'est pas le cas par exemple avec la mesure extérieure de Lebesgue.

Théorème 10.3 (d'extension de Carathéodory)**.** *1. $\mathcal{M}(\mu^*)$ contient toute les parties B telles que $\mu^*(B) = 0$.*

2. $\mathcal{M}(\mu^*)$ est une tribu.

3. La restriction de μ^* à $\mathcal{M}(\mu^*)$ est une mesure complète de E.

Démonstration. Notons $\mathcal{M} = \mathcal{M}(\mu^*)$ et μ la restriction de μ^* à $\mathcal{M}$.

1. Si $\mu^*(B) = 0$, pour toute partie A de E on a
$$\begin{aligned} \mu^*(A) &\geqslant \mu^*(A \cap B^c) && \text{(croissance)} \\ &= \mu^*(A \cap B^c) + \mu^*(A \cap B) && (\text{car } \mu^*(A \cap B) \leqslant \mu^*(B) = 0), \end{aligned}$$
donc $B \in \mathcal{M}$.

2. Constantin Carathéordory, mathématicien grec (1873-1950), spécialiste d'analyse réelle, de calcul variationnel et de théorie de la mesure. Il a aussi donné une présentation axiomatique de la thermodynamique.

2. Montrons que $\mathcal{M}$ est une tribu. Le seul axiome qui ne soit pas immédiat est la stabilité par union dénombrable.

   Montrons d'abord la stabilité par union finie. Soient $B_1, B_2 \in \mathcal{M}$. Pour toute partie A de E,

$$\begin{aligned}\mu^*(A \cap (B_1 \cup B_2)) &= \mu^*(A \cap (B_1 \cup B_2) \cap B_1) + \mu^*(A \cap (B_1 \cup B_2) \cap B_1^c)\\ &\text{(car } B_1 \in \mathcal{M})\\ &= \mu^*(A \cap B_1) + \mu^*(A \cap B_2 \cap B_1^c)\\ &\text{(par simplification),}\end{aligned}$$

   donc

$$\begin{aligned}&\mu^*(A \cap (B_1 \cup B_2)) + \mu^*(A \cap (B_1 \cup B_2)^c)\\ &= \mu^*(A \cap B_1) + \mu^*(A \cap B_1^c \cap B_2) + \mu^*(A \cap B_1^c \cap B_2^c)\\ &= \mu^*(A \cap B_1) + \mu^*(A \cap B_1^c) \qquad \text{(car } B_2 \in \mathcal{M})\\ &= \mu^*(A) \qquad \text{(car } B_1 \in \mathcal{M}).\end{aligned}$$

   Donc $B_1 \cup B_2 \in \mathcal{M}$. Par récurrence, $\mathcal{M}$ est stable par union finie et, par stabilité par passage au complémentaire, $\mathcal{M}$ est stable par intersection finie.

   En particulier, si $B, B' \in \mathcal{M}$, $B' \backslash B = B' \cap B^c \in \mathcal{M}$. Pour démontrer la stabilité par union dénombrable, il suffit donc de montrer la stabilité par union dénombrable de parties disjointes deux à deux. Soient donc $B_k, \in \mathcal{M}$ $(k \in \mathbb{N})$ deux à deux disjoints. Montrons par récurrence que, pour tout $m \in \mathbb{N}$ et toute partie A de E,

$$\mu^*(A) = \sum_{0 \leqslant k \leqslant m} \mu^*(A \cap B_k) + \mu^*\left(A \cap \left(\bigcap_{0 \leqslant k \leqslant m} B_k^c\right)\right).$$

   L'égalité est vérifiée pour $m = 0$ parce que $B_0 \in \mathcal{M}$. L'égalité au rang $m+1$ se déduit de celle au rang m, puisque

$$\begin{aligned}&\mu^*\left(A \cap \left(\bigcap_{0 \leqslant k \leqslant m} B_k^c\right)\right)\\ &= \mu^*\left(A \cap \left(\bigcap_{0 \leqslant k \leqslant m} B_k^c\right) \cap B_{m+1}\right) + \mu^*\left(A \cap \left(\bigcap_{0 \leqslant k \leqslant m+1} B_k^c\right)\right)\\ &\text{(car } B_{m+1} \in \mathcal{M})\\ &= \mu^*(A \cap B_{m+1}) + \mu^*\left(A \cap \left(\bigcap_{0 \leqslant k \leqslant m+1} B_k^c\right)\right)\\ &\text{(car } B_{m+1} \text{ est disjoint de } B_0, ..., B_m).\end{aligned}$$

D'après l'égalité ainsi vérifiée par récurrence et d'après la croissance de μ^*,

$$\begin{aligned}\mu^*(A) &\geqslant \sum_{0\leqslant k\leqslant m} \mu^*(A\cap B_k) + \mu^*\left(A\cap\left(\bigcap_{k\geqslant 0} B_k^c\right)\right)\\ &\text{(croissance de } \mu^*\text{)}\\ &\geqslant \sum_{k\geqslant 0} \mu^*(A\cap B_k) + \mu^*\left(A\cap\left(\bigcap_{k\geqslant 0} B_k^c\right)\right)\\ &\text{(à la limite quand } m\to\infty\text{)}\\ &\geqslant \mu^*\left(A\cap\left(\bigcup_{k\geqslant 0} B_k\right)\right) + \mu^*\left(A\cap\left(\bigcap_{k\geqslant 0} B_k^c\right)\right)\\ &\text{(par } \sigma\text{-sous-additivité)),}\end{aligned}$$

ce qui montre que $\bigcup_{k\geqslant 0} B_k \in \mathcal{M}$.

3. Montrons enfin que μ est une mesure ; elle est automatiquement complète d'après la première partie du théorème. Le seul axiome à vérifier est la σ-additivité de μ. Soient $B_1, B_2, \ldots$ une famille dénombrable d'éléments de $\mathcal{M}$. On a vu que, pour toute partie A de E,

$$\mu^*(A) \geqslant \sum_{k\geqslant 0} \mu^*(A\cap B_k) + \mu^*\left(A\cap\left(\bigcap_{k\geqslant 0} B_k^c\right)\right).$$

Dans le cas où $A = \bigcup_{k\geqslant 0} B_k$, on obtient

$$\mu\left(\bigcup_{k\geqslant 0} B_k\right) \geqslant \sum_{k\geqslant 0} \mu(B_k),$$

qui est le sens manquant a priori pour la σ-additivité.

□

Remarque 10.4. La mesure extérieure n'est généralement pas σ-additive (l'argument est le même que pour construire une partie non borélienne de $\mathbb{R}$) ; cf. l'exercice 11.5. Un exemple élémentaire est donné par la mesure extérieure μ^* sur $\mathbb{N}$ définie par

$$\mu^*(A) = \begin{cases} 0 & \text{si } A \text{ est fini}\\ 1 & \text{sinon,}\end{cases}$$

pour laquelle on a $\sum_n \mu^*(\{n\}) = 0 < \mu^*(\cup\{n\}) = \mu^*(\mathbb{N}) = 1$.

Chapitre 11

La mesure de Lebesgue ⋆

Pour toute partie A de $\mathbb{R}$, on définit

$$\lambda^*(A) = \inf \sum (b_i - a_i),$$

où la borne inférieure porte sur l'ensemble des recouvrements dénombrables $\{]a_1, b_1[, \ldots\}$ de A par des intervalles ouverts.

Théorème 11.1. *1. λ^* est une mesure extérieure sur $\mathbb{R}$.*

2. La tribu $\mathcal{M}(\lambda^*)$ contient $\mathcal{B}(\mathbb{R})$.

3. Pour tous $a \leqslant b$,

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

Le qualificatif "extérieur" est donc justifié par le fait que l'on approche les parties quelconques de $\mathbb{R}$ par des recouvrements, qui les débordent.

Définition 11.2. La *tribu de Lebesgue* est $\mathcal{M}(\lambda^*)$. La *mesure de Lebesgue* est la restriction, notée λ, de λ^* à $\mathcal{B}(\mathbb{R})$.

Exemple 11.3. Du théorème découle immédiatement que $\lambda(\mathbb{R}) = +\infty$ et $\lambda(\mathbb{Q}) = 0$.

Comme le montre l'exercice 4.4, ce serait une grossière erreur de penser que toute partie Lebesgue-négligeable est dénombrable. En voici un autre exemple.

Exercice ⋆ 11.1 (Nombres liouvilliens)**.**
D'après l'exercice 5.7, l'ensemble L des nombres réels liouvilliens est Lebesgue-négligeable. Montrer que L est non dénombrable ; on pourra utiliser la propriété de Baire* de $\mathbb{R}$. (En 1844, Liouville en a déduit l'existence de nombres *transcendants*, c'est-à-dire de nombres réels qui ne sont la racine d'aucun polynôme à coefficients entiers.)

Démonstration du théorème. 1. Il est immédiat que $\lambda^*(\varnothing) = 0$ et que λ^* est croissante. Montrons que λ^* est σ-sous-additive, et choisissons pour cela une suite $(A_n)_{n\in\mathbb{N}}$ de parties de $\mathbb{R}$. S'il existe un entier n pour lequel $\lambda^*(A_n) = +\infty$, il n'y a rien à démontrer. Supposons donc que $\lambda^*(A_n) < +\infty$ pour tout

$n \in \mathbb{N}$. Soit $\epsilon > 0$. Par définition de λ^*, il existe des intervalles $]a_{n,i}, b_{n,i}[$ $(i, n \in \mathbb{N})$ tels que, pour tout n,

$$\begin{cases} A_n \subset \bigcup_i]a_{n,i}, b_{n,i}[\\ \lambda^*(A_n) \leqslant \sum_i (b_{n,i} - a_{n,i}) \leqslant \lambda^*(A_n) + \frac{\epsilon}{2^n}. \end{cases}$$

La famille dénombrable $(]a_{n,i}, b_{n,i}[)_{i,n\in\mathbb{N}}$ recouvre la réunion des A_n, et donc

$$\lambda^* \left(\bigcup_n A_n \right) \leqslant \sum_n \sum_i (b_{n,i} - a_{n,i}) \leqslant \sum_n \lambda^*(A_n) + 2\epsilon.$$

En faisant tendre ϵ vers 0, on obtient que λ^* est σ-sous-additive.

2. Comme $\mathcal{M}(\lambda^*)$ est une tribu, il suffit de montrer qu'elle contient une classe de parties de $\mathbb{R}$ qui engendre $\mathcal{B}(\mathbb{R})$, par exemple la classe des intervalles de la forme $]-\infty, \alpha]$ avec $\alpha \in \mathbb{R}$. Soient donc $\alpha \in \mathbb{R}$ et $B =]-\infty, \alpha]$. Il s'agit de vérifier que B est λ^*-mesurable, c'est-à-dire que, pour toute partie A de $\mathbb{R}$,
$$\lambda^*(A) \geqslant \lambda^*(A \cap B) + \lambda^*(A \cap B^c).$$
Soient $\epsilon > 0$ et $(]a_i, b_i[)_{i\in\mathbb{N}}$ un recouvrement de A. Les intervalles
$$]\min(a_i, \alpha), \min(b_i, \alpha) + \epsilon 2^{-i}[$$
recouvrent $A \cap B$, et les
$$]\max(a_i, \alpha), \max(b_i, \alpha)[$$
recouvrent $A \cap B^c$. Donc
$$\begin{cases} \lambda^*(A \cap B) \leqslant \sum_i (\min(b_i, \alpha) - \min(a_i, \alpha)) + 2\epsilon \\ \lambda^*(A \cap B^c) \leqslant \sum_i (\max(b_i, \alpha) - \max(a_i, \alpha)). \end{cases}$$
En ajoutant ces inégalités, le membre de droite se simplifie et donc
$$\lambda^*(A \cap B) + \lambda^*(A \cap B^c) \leqslant \sum_i (b_i - a_i) + 2\epsilon;$$
en faisant tendre ϵ vers 0,
$$\lambda^*(A \cap B) + \lambda^*(A \cap B^c) \leqslant \sum_i (b_i - a_i);$$
en prenant enfin l'infimum sur tous les recouvrements de A, on obtient l'inégalité voulue.

3. Par définition,
$$\lambda([a, b]) \leqslant b - a.$$
Il reste à montrer l'inégalité inverse. Supposons que
$$[a, b] \bigcup_{i\in\mathbb{N}}]a_i, b_i[.$$

Comme $[a, b]$ est compact, il existe $N \in \mathbb{N}$ tel que

$$[a, b] \bigcup_{0 \leqslant i \leqslant N}]a_i, b_i[.$$

Par un raisonnement élémentaire, on voit que

$$b - a \leqslant \sum_{0 \leqslant i \leqslant N} (b_i - a_i).$$

En faisant tendre N vers l'infini, on obtient

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

et cette inégalité est donc une égalité. Mais d'après la définition de λ^* on voit que $\lambda(\{a\}) = \lambda(\{b\}) = 0$ et, comme λ est additive, $\lambda(]a, b[) = \lambda([a, b])$. $\square$

Proposition 11.4. *La mesure sur* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *telle que, pour tout intervalle non vide* $]a, b[$, $\lambda(]a, b[) = b - a$ *est unique.*

Démonstration. Soit. λ' une deuxième mesure satisfaisant la même propriété. La proposition découle du corollaire 9.3 appliqué avec prenant pour $\mathcal{C}$ la classe des intervalles ouverts (dont on sait qu'elle engendre $\mathcal{B}(\mathbb{R})$) et pour suite de parties E_n les intervalles $] - n, n[$ $(n \in \mathbb{N})$. $\square$

Rappelons qu'un pavé ouvert de $\mathbb{R}^d$ est un ensemble de la forme

$$P = \prod_{1 \leqslant i \leqslant d}]a_i, b_i[.$$

Le *volume* de P est, par définition

$$\mathrm{vol}(P) = \prod_{1 \leqslant i \leqslant d} (b_i - a_i).$$

Pour toute partie A de $\mathbb{R}^d$, on définit

$$\lambda^*(A) = \inf \sum_{i \in \mathbb{N}} \mathrm{vol}(P_i),$$

où l'infimum porte sur l'ensemble des recouvrements dénombrables $(P_i)_{i \in \mathbb{N}}$ de A par des pavés ouverts.

Le théorème précédent se généralise directement en dimension d.

Théorème 11.5.
1. *λ^* est une mesure extérieure sur $\mathbb{R}^d$.*
2. *La tribu $\mathcal{M}(\lambda^*)$ contient $\mathcal{B}(\mathbb{R}^d)$.*
3. *Pour tout pavé ouvert ou fermé P, $\lambda^*(P) = \mathrm{vol}(P)$.*

La démonstration, peu différente du cas unidimensionnel, est laissée à la sagacité du lecteur.

Définition 11.6. La *mesure de Lebesgue d-dimensionnelle* est la restriction, notée λ (ou parfois λ_d), de λ^* à $\mathcal{B}(\mathbb{R}^d)$.

On peut se demander si la tribu $\mathcal{M}(\lambda^*)$ est beaucoup plus grande que $\mathcal{B}(\mathbb{R})$. Nous allons voir que non, dans un certain sens.

Proposition 11.7. *La tribu $\mathcal{M}(\lambda^*)$ est la tribu complétée de $\mathcal{B}(\mathbb{R}^d)$ (cf. chapitre 4).*

Démonstration. On a vu que $\mathcal{M}(\lambda^*)$ contient $\mathcal{B}(\mathbb{R}^d)$ (theoreme 11.5) et l'ensemble $\mathcal{N}$ des parties N de $\mathbb{R}^d$ telles que $\lambda^*(N) = 0$ (théorème 10.3). Comme $\mathcal{M}(\lambda^*)$ est une tribu, $\mathcal{M}(\lambda^*)$ contient donc la tribu borélienne complétée

$$\bar{\mathcal{B}}(\mathbb{R}^d) = \sigma\left(\mathcal{B}(\mathbb{R}^d) \cup \mathcal{N}\right).$$

Inversement, montrons que $\mathcal{M}(\lambda^*) \subset \bar{\mathcal{B}}(\mathbb{R}^d)$. Soit $A \in \mathcal{M}(\lambda^*)$. Comme A est la réunion dénombrables des $A \cap]-n, n[^d$, il suffit de démontrer que $A \in \bar{\mathcal{B}}(\mathbb{R}^d)$ dans le cas où $A \subset]-N, N[^d$, le cas général s'en déduisant. On a alors $\lambda^*(A) < \infty$, et on peut trouver des pavés ouverts $P_{n,i} \subset]-N, N[^d$ $(i, n \in \mathbb{N}_*)$ tels que

$$\begin{cases} A \subset \bigcup_i P_{n,i} \\ \lambda^*(A) \leqslant \sum_i \mathrm{vol}(P_{n,i}) \leqslant \lambda^*(A) + \frac{1}{n}. \end{cases}$$

Posons

$$B_n = \bigcup_i P_{n,i} \quad \text{et} \quad B = \bigcap_n B_n.$$

Alors $A \subset B \in \mathcal{B}(\mathbb{R}^d)$ et, pour tout n,

$$\lambda^*(B) \leqslant \sum_i \mathrm{vol}(P_{n,i}) \leqslant \lambda^*(A) + \frac{1}{n},$$

ce qui montre que $\lambda^*(A) = \lambda^*(B)$. En remplaçant A par $]-N, N[^d \backslash A$, on construit de même une partie borélienne $\tilde{B}$ de $]-N, N[^d$ contenant $]-N, N[^d \backslash A$ et telle que $\lambda^*(]-N, N[^d \backslash A) = \lambda^*(\tilde{B})$. Posons enfin $B' =]-N, N[^d \backslash \tilde{B}$; on a $B' \subset A$ et $\lambda^*(B') = \lambda^*(A)$. Donc B et B' sont deux boréliens tels que

$$B' \subset A \subset B \quad \text{et} \quad \lambda(B \backslash B') = 0.$$

Donc, d'après l'exercice 4.3, $A \in \bar{\mathcal{B}}(\mathbb{R}^d)$. □

Remarque ★ 11.8. On peut cependant montrer que, du point de vue des cardinaux, les tribus $\mathcal{M}(\lambda^*)$ et $\mathcal{B}(\mathbb{R}^d)$ sont très différentes : la première a le cardinal de $\mathcal{P}(\mathbb{R})$, tandis que la seconde a seulement le cardinal de $\mathbb{R}$.

Théorème 11.9 (Caractérisation de la mesure de Lebesgue)**.** *La mesure de Lebesgue sur $\mathbb{R}^d$ est invariante par translation :*

$$\lambda(A + x) = \lambda(A) \quad (\forall A \in \mathcal{B}(\mathbb{R}^d),\ x \in \mathbb{R}^n).$$

Inversement, si μ est une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ finie sur les parties bornées et invariante par translation, il existe un réel $c \geqslant 0$ tel que $\mu = c\lambda$.

Démonstration. Soit τ_x la translation $y \mapsto y - x$ dans $\mathbb{R}^d$. La mesure image (cf. la définition 6.6) de λ par σ_x est définie par

$$\sigma_x(\lambda)(A) = \lambda(\sigma_x^{-1}(A)) = \lambda(x + A) \quad (\forall A \in \mathcal{B}(\mathbb{R}^d)).$$

Il s'agit donc de montrer que $\sigma_x(\lambda)(A) = \lambda(A)$ pour tout $A \in \mathcal{B}(\mathbb{R}^d)$. Si A est un pavé, A et $x+A$ ayant même volume, l'égalité est satisfaite. Le cas général découle donc du corollaire 9.3.

Réciproquement, soit μ une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ invariante par translation. Soit

$$c = \mu\left([0,1[^d\right).$$

Pour tout entier $n \geqslant 1$, $[0,1[^d$ est l'union disjointe de n^d pavés obtenus par translation du pavé $[0,1/n[^d$, donc

$$\mu\left(\left[0, \frac{1}{n}\right[^d\right) = \frac{c}{n^d}.$$

On généraliser cette formule à un pavé de coordonnées réelles quelconques, en l'encadrant avec deux pavés du type précédent. Soient $a_1, ..., a_d \geqslant 0$. Notons $[x]$ la partie entière de x. On a

$$\prod_i \left[0, \frac{[na]}{n}\right[\subset \prod_i [0, a_i[\subset \prod_i \left[0, \frac{[na_i]+1}{n}\right[,$$

donc

$$\begin{cases} \mu\left(\prod_i [0,a_i[\right) \geqslant \mu\left(\prod_i \left[0, \frac{[na]}{n}\right[\right) = \left(\prod_i [na_i]\right) \frac{c}{n^d} \\ \mu\left(\prod_i [0,a_i[\right) \leqslant \mu\left(\prod_i \left[0, \frac{[na]+1}{n}\right[\right) = \left(\prod_i ([na_i]+1)\right) \frac{c}{n^d}. \end{cases}$$

D'après les propriétés de continuité intérieure et extérieure de μ (exercice 4.2), quand n tend vers l'infini on obtient

$$\mu\left(\prod_i [0, a_j[\right) = c \prod_i a_i = c\lambda\left(\prod_i [0, a_j[\right).$$

Donc les mesures μ et $c\lambda$ coïncident sur tous les pavés de la forme $\prod_i [a_i, b_i[$. Ici encore, le corollaire 9.3 permet de conclure : $\mu = c\lambda$. □

Autres exercices

Exercice 11.2 (Recouvrement de $\mathbb{Q}$)**.**
Montrer que, pour tout $\epsilon > 0$, on peut recouvrir $\mathbb{Q}$ par une famille dénombrable d'intervalles ouverts, dont la longueur totale est $\leqslant \epsilon$.

Exercice 11.3 (Mesures de Lebesgue 1 et 2-dimensionnelles)**.**
Montrer que la mesure de Lebesgue bidimensionnelle de l'intervalle

$$I = \{(x,0),\ 0 \leqslant x \leqslant 1\} \subset \mathbb{R}^2$$

est nulle (et donc que les mesures de Lebesgue 1-dimensionnelle et 2-dimensionnelle sont distinctes).

Exercice 11.4 (Régularité de la mesure de Lebesgue).
Montrer que la mesure de Lebesgue est *régulière* : pour toute partie $A \in \bar{\mathcal{B}}(\mathbb{R}^d)$,

$$\lambda(A) = \inf\{\lambda(U),\, U \text{ ouvert } \supset A\} = \sup\{\lambda(K),\, K \text{ compact} \subset A\}.$$

Donner un exemple pour lequel $\lambda(A) \neq \sup\{\lambda(U),\, U \text{ ouvert } \subset A\}$.

Exercice 11.5 (Un ensemble non Lebesgue-mesurable).
Soit A un *ensemble de Vitali*,[1] construit de la façon suivante. Soit R la relation d'équivalence sur $\mathbb{R}$ qui identifie deux nombres réels dont la différence est un nombre rationnel. Chaque classe d'équivalence de R possédant un représentant dans l'intervalle $[0,1[$, on peut définir $A \subset [0,1[$ comme un système de représentants de R, c'est-à-dire une partie de $[0,1[$ qui contienne un et un seul point de chaque classe d'équivalence (l'existence d'une telle partie repose sur l'axiome du choix* non dénombrable).

1. Montrer, en utilisant le fait que $\mathbb{R} = \cup_{r\in\mathbb{Q}}(A + r)$, que $\lambda^*(A) > 0$.

Posons $X = \cup_{q\in\mathbb{Q}\cap[-1,1]}(A + q)$.

2. Montrer, en utilisant le fait que $[0,1] \subset X \subset [-1,2]$, que

$$1 \leqslant \lambda^*(X) \leqslant 3 < +\infty = \sum_{q\in\mathbb{Q}\cap[-1,1]} \lambda^*(A + q)$$

(ainsi, la famille $(A + q)_{q\in\mathbb{Q}\cap[0,1]}$ est un contre-exemple à la σ-additivité de λ^*) et que, donc, la mesure de Lebesgue ne se prolonge pas en une mesure sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ invariante par translation.

On vient de voir que λ^* n'est pas σ-additive. Pire :

3. Déduire de ce qui précède que λ^* n'est pas même finiment additive : il existe des parties bornées disjointes E et F de $\mathbb{R}$ telles que $\lambda^*(E \cup F) \neq \lambda^*(E) + \lambda^*(F)$.

(En revanche, à l'aide du théorème de Hahn-Banach, on peut construire un prolongement finiment additif de la mesure de Lebesgue à $\mathcal{P}(\mathbb{R})$, qui soit invariant par translation !)

Remarque ★ 11.10. Ce type d'exemple est relié au *paradoxe de Banach-Tarski*,[2,3] qui démontre (avec l'axiome du choix non-dénombrable) qu'il existe une partition finie de la boule unité de $\mathbb{R}^3$ telle que l'union de ses éléments déplacés par rotation et translation forme *deux* boules unité ! Une telle partition est bien sûr faite de parties non Lebesgue-mesurables. La construction de telles parties requiert des arguments de théorie des groupes moyennables. Ce paradoxe aide à comprendre pourquoi la mesure de Lebesgue (qui est σ-additive et invariante par rotations et translations, et qui charge les ouverts) ne peut pas mesurer toutes les parties de $\mathbb{R}^n$.

Par ailleurs, un théorème de Solovay[4] affirme que la question de l'existence d'ensembles non Lebesgue-mesurables est indécidable sans l'axiome du choix non-

---

1. Giuseppe Vitali, mathématicien italien (1875–1932)
2. Stefan Banach, mathématicien polonais (1892–1945), fondateur de l'analyse fonctionnelle
3. Alfred Tarski, mathématicien et logicien polonais (1801–1945), principalement connu pour ses travaux en théorie des modèles et en logique.
4. Robert Solovay, mathématicien américain (né en 1938), spécialiste de théorie des ensembles.

dénombrable, c'est-à-dire que l'on peut construire un modèle de théorie des ensembles dans lequel toute partie de $\mathbb{R}$ est Lebesgue-mesurable [Sol70].

Curieusement, l'existence d'ensembles non Lebesgue-mesurables implique aussi l'existence d'ensembles non boréliens, comme le montre l'exercice suivant.

Problème 11.1 (Un ensemble Lebesgue-mesurable non-borélien)**.**
Notons, pour tout $x \in [0,1]$,

$$x = \sum_{n \geqslant 1} \frac{\epsilon_n}{2^n}, \quad \epsilon_n \in \{0,1\}$$

le développement dyadique infini de x (tout nombre non nul possède en effet un unique développement dyadique qui ne se termine pas pas une suite infinie de zéros) ; cf. [Tao09].

1. Montrer que, pour tout entier $n \geqslant 1$, la fonction $\epsilon_n : [0,1] \to \{0,1\}$ ainsi définie est borélienne.

Soit $f : [0,1] \to [0,1]$ la fonction

$$x = \sum_{n \geqslant 1} \frac{\epsilon_n}{2^n} \longmapsto \sum_{n \geqslant 1} \frac{2\epsilon_n}{3^n}.$$

Comme l'exercice 4.4 le montre, f réalise une bijection de $[0,1]$ sur l'ensemble de Cantor triadique C. Nous utiliserons uniquement le fait que C est de mesure nulle.

2. Montrer que f est borélienne.

Soient A un ensemble de Vitali $\in \mathcal{P}(\mathbb{R}) \backslash \bar{\mathcal{B}}(\mathbb{R})$, comme celui construit dans l'exercice 11.5, et $B = f(A)$.

3. Montrer que B est Lebesgue-mesurable.

4. Montrer que f n'est pas borélienne.

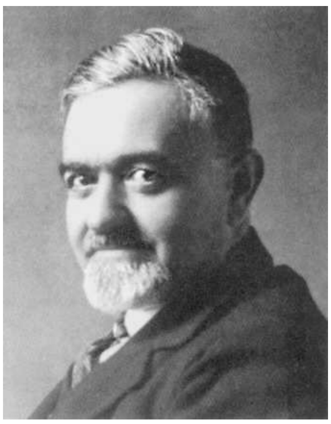
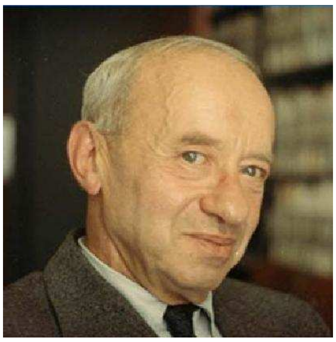




FIGURE 17 – Giuseppe VITALI (1875–1932), Alfred TARSKI (1901–1983) et Robert Martin SOLOVAY (né en 1938) ont eu tous trois des contributions marquantes en théorie de la mesure

Problème 11.2 (Mesure de Hausdorff [Tao10a])**.**
La mesure de Hausdorff d-dimensionnelle[5] est une mesure adaptée pour les objets d-dimensionnels dans $\mathbb{R}^n$. Nous allons faire jouer le rôle des les pavés ouverts dans la construction de la mesure de Lebesgue, par par les boules ouvertes

$$B(x,r) = \{y \in \mathbb{R}^n;\ \|y-x\| < r\},$$

auxquelles on assigne la mesure r^d (la constante multiplicative éventuelle qu'on pourrait ajouter n'a aucune importance qualitative). Pour toute partie $A \subset \mathbb{R}^n$, notons donc

$$h^*_{d,\infty}(A) = \inf \sum_{k\geqslant 1} r_k^d,$$

où la borne inférieure porte sur les familles dénombrables de boules $B(x_k, r_k)$ qui recouvrent A.

1. Montrer que $h^*_{d,\infty}$ est une mesure extérieure.

Comme dans la construction de la mesure de Lebesgue, on peut donc considérer la restriction de $h^*_{d,\infty}$ à la tribu $\mathcal{M}(h^*_{d,\infty})$ des parties $h^*_{d,\infty}$-mesurables de $\mathbb{R}^n$. Le problème est que, si $d < n$, la plupart des parties de $\mathbb{R}^n$ ne sont pas $h^*_{d,\infty}$-mesurables !

2. Dans le cas $n = 1$ et $d = 1/2$, montrer que

$$h^*_{1/2,\infty}([a,b]) = \left(\frac{b-a}{2}\right)^{1/2}.$$

3. En déduire que

$$h^*_{1/2,\infty}([0,1]) = h^*_{1/2,\infty}(]1,2]) = \frac{1}{\sqrt{2}} \quad et \quad h^*_{1/2,\infty}([0,2]) = 1,$$

puis que $[0,1]$ n'est pas $h^*_{1/2,\infty}$-mesurable.

On voit d'où vient le problème : la façon la plus efficace de recouvrir $[0,2]$ n'est pas de recouvrir séparément $[0,1]$ et $]1,2]$, mais de recouvrir $[0,2]$ directement par une grosse boule. Pour pallier cette difficulté, on va se limiter aux boules de petit rayon. Posons

$$h^*_{d,r}(A) = \inf \sum_{k\geqslant 1} r_k^d,$$

en faisant maintenant porter la borne inférieure sur toutes les suites de boules $B(x_k, r_k)$ recouvrant A avec $r_k \leqslant r$; puis, $h^*_{d,r}$ étant croissante par rapport à r, définissons la *mesure extérieure de Hausdorff d-dimensionnelle*

$$h^*_d(A) = \lim_{r\geqslant 0} h_{d,r}(A).$$

4. Montrer que, si $d > n$, $h^*_d \equiv 0$.

5. Felix Hausdorff, mathématicien allemand (1868–1942). Il fut l'un des fondateurs de la topologie générale, et fit de nombreuses découvertes en théorie des ensembles, théorie de la mesure et analyse fonctionnelle.

Dans toute la suite, on suppose donc que $0 \leqslant d \leqslant n$.

5. Montrer que les $h_{d,r}$, $r > 0$, et donc h_d^* sont des mesures extérieures sur $\mathbb{R}^n$.

6. Soient A et B deux parties de $\mathbb{R}^n$ dont la distance soit > 0 :

$$\operatorname{dist}(A, B) = \inf\{\|x - y\|;\ x \in A,\ y \in B\} > 0.$$

Montrer que

$$h_d^*(A \cup B) = h_d^*(A) + h_d^*(B).$$

7. En déduire que tout borélien, et donc toute partie Lebesgue-mesurable, est h_d^*-mesurable ; on pourra remarquer qu'il suffit de considérer un fermé A de $\mathbb{R}^n$, puis montrer que, pour toute partie E,

$$h_d^*(E \cap A) + h_d^*(E \backslash A_{1/m}) \leqslant h_d^*(E) \leqslant h_d^*(E \cap A) + h_d^*(E \backslash A),$$

où $A_{1/m}$ est le $1/m$-voisinage de A.

On note h_d la mesure, restriction de h_d^* à $\mathcal{M}(h_d^*)$.

8. Montrer que h_0 est la mesure de comptage sur $(\mathbb{R}^n, \mathcal{P}(\mathbb{R}^n))$.

9. Considérons à l'opposé la mesure n-dimensionnelle.

9.1. Montrer que, si $\lambda(A) = 0$, $h_n(A) = 0$.

On admettra qu'alors il existe une fonction c telle que $h_n = c\lambda$ (théorème de Radon-Nikodym 31.2) et que, comme h_n et λ sont invariantes par translation, c l'est aussi, i.e. c est constante.

9.2. Montrer que, si l'on note ω_n la mesure de Lebesgue de la boule unité en dimension n,

$$h_n = \frac{1}{\omega_n}\lambda.$$

10. Montrer que la restriction de h_d à un sous-espace affine V de $\mathbb{R}^n$ est égal à $1/\omega_d$ fois la mesure de Lebesgue d-dimensionnelle sur V.

11. En déduire que h_d n'est pas σ-finie (définition 13.3) si $d < n$.

12. Soient U un ouvert de $\mathbb{R}^d$ et $\phi : U \to \mathbb{R}^n$ une immersion (c'est-à-dire une application dont la différentielle soit de rang d en tout point). Montrer que, pour tout compact K de U,

$$h_d(\phi(K)) = \int_K J\, dh_d = \frac{1}{\omega_d}\int_K J\, d\lambda_d,$$

où le *jacobien* J est

$$J = \left(\sum_M (\det M)^2\right)^{1/2},$$

la somme étant étendue à tous les mineurs $d \times d$ de $d\phi$.

13. Que dit cette formule dans le cas d'une courbe $\gamma : [a, b] \to \mathbb{R}^n$ de classe C^1 telle que γ' ne s'annule pas ? Pourquoi la formule reste-t-ell vraie dans le cas d'une courbe C^1 par morceaux et non-dégénérée par morceaux ?

14. Soient $d < d'$ et A un borélien de $\mathbb{R}^n$. Montrer que :

14.1. si $h_d(A) < \infty$, $h_{d'}(A) = 0$;

14.2. si $h_{d'}(A) > 0$, $h_d(A) = \infty$.

15. En déduire que

$$h_{d'}(B^d(0,1)) = \begin{cases} \infty & \text{si} \quad d' < d \\ 1 & \text{si} \quad d' = d \\ 0 & \text{si} \quad d' > d. \end{cases}$$

16. Montrer plus généralement que, pour tout borélien A de $\mathbb{R}^n$, il existe un unique réel dans $[0,n]$ (appelé la *dimension de Hausdorff* de A et noté $\dim_H(A)$), tel que

$$h_d(A) = \begin{cases} \infty & \text{si} \quad d < \dim_H(A) \\ 0 & \text{si} \quad d > \dim_H(A). \end{cases}$$

17. Montrer que la dimension de Haussdorff de l'ensemble triadique de Cantor vaut $1/2$.

La construction faite en début de chapitre de la mesure extérieure de Lebesgue à partir de la longueur des intervalles peut être axiomatisée, pour ensuite servir dans d'autres constructions. Rappelons la notions d'algèbre, déjà rencontrée dans le problème 3.1.

Définition 11.11. Une *algèbre (de Boole)*[6] sur un ensemble E est une classe de parties $\mathcal{A}_0$ vérifiant les axiomes suivants :
— Ensemble vide : $\varnothing \in \mathcal{A}_0$
— Stabilité par complément : si $A \in \mathcal{A}_0$, $A^c \in \mathcal{A}_0$
— Stabilité par union finie : si $A, B \in \mathcal{A}_0$, $A \cup B \in \mathcal{A}_0$.
Autrement dit, ce qui manque à une algèbre pour être une tribu est la stabilitié par union dénombrable.

Une *prémesure* sur $(E, \mathcal{A}_0)$ est une fonction $\mu_0 : \mathcal{A}_0 \to [0, +\infty]$ σ-additive : pour toute suite $A_1, A_2, ... \in \mathcal{A}_0$ de parties disjointes deux à deux telle que $\bigcup A_j \in \mathcal{A}_0$,

$$\mu\left(\bigcup A_i\right) = \sum \mu(A_i).$$

Problème 11.3 (Théorème d'extension de Hahn-Kolmogorov [Tao11])**.**
Appelons *partie élémentaire* de $\mathbb{R}^d$ une partie qui soit la réunion finie de produits d'intervalles bornés de $\mathbb{R}$.

1. Montrer que les unions finies d'intervallles et, plus généralement, les parties élémentaires forment une algèbre respectivement sur $\mathbb{R}$ et sur $\mathbb{R}^d$.

2. Sans utiliser la mesure de Lebesgue, montrer que le volume des parties élémentaires de $\mathbb{R}^d$ est une prémesure.

3. Construire un exemple de mesure finiment additive (soit une fonction μ_0 : $\mathcal{A}_0 \to [0, +\infty]$ finiment additive avec $\mu_0(\varnothing) = 0$) qui ne soit pas une prémesure.

6. Il existe une autre notion d'algèbre de Boole, qui est abstraction de celle-ci et que nous ne considèrerons pas ici.

Indication. Prendre $E = \mathbb{N}$ et $\mathcal{A}_0 = \mathcal{P}(\mathbb{N})$, puis définir μ_0 séparément sur les ensembles finis et infinis.

L'objectif est de montrer le *théorème de Hahn-Kolmogorov*[7] : *Toute prémesure μ_0 sur une algèbre $\mathcal{A}_0$ de E se prolonge en une mesure (σ-additive) $\mu : \mathcal{A} \to [0, +\infty]$.*

4. Montrer que la formule

$$\mu^*(A) = \inf \mu_0(E_n),$$

où la borne inférieure porte sur l'ensemble des recouvrements dénombrables $\{E_0, E_1, ...\}$, $E_i \in \mathcal{A}_0$ ($\forall i$), de A, définit une mesure extérieure sur E.

On définit $\mathcal{A}$ comme l'ensemble des parties μ^*-mesurables et μ comme la restriction de μ^* à $\mathcal{A}$. D'après le théorème de Carathéodory, $\mathcal{A}$ est une tribu et μ une tribu sur $(E, \mathcal{A})$.

5. Montrer que toute partie $A \in \mathcal{A}_0$ est μ^*-mesurable, c'est-à-dire que $\mathcal{A}$ contient $\mathcal{A}_0$.

6. Montrer que, pour toute partie $A \in \mathcal{A}_0$, $\mu(A) = \mu_0(A)$, c'est-à-dire que μ prolonge μ_0.

(La construction de la mesure de Lebesgue est donc un cas particulier de ce qui précède.)

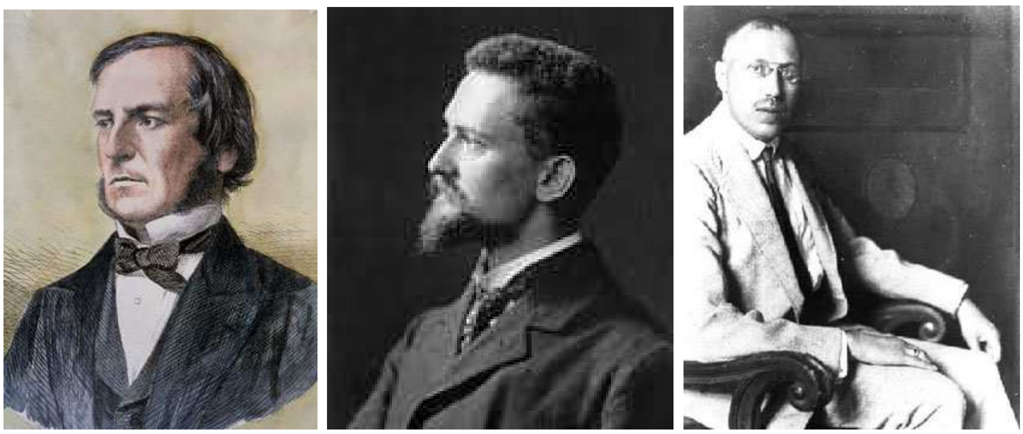


FIGURE 18 – George Boole (1815–1864), Felix HAUSDORFF (1868–1942) et Hans HAHN (1879–1934). Felix Hausdorff se suicida pour éviter la déportation.

7. Hans HAHN, mathématicien autrichien (1879–1934), qui s'est illustré en analyse fonctionnelle, topologie, analyse réelle et théorie des ensembles.

Chapitre 12

Lien avec l'intégrale de Riemann

Rappelons qu'une fonction $g : [a, b] \to \mathbb{R}$ est *en escalier* * s'il existe une subdivision finie $a_0 = a < a_1 < ... < a_{n-1} < a_n = b$ telle que g soit constante sur chaque intervalle $]a_{i-1}, a_i[$ (sa valeur aux points a_i pouvant être encore différente des deux valeurs prises à gauche et à droite) :

$$g = \sum_{1 \leqslant i \leqslant n} y_i \mathbb{1}_{]a_{i-1}, a_i[} + \sum_{0 \leqslant i \leqslant n} z_i \mathbb{1}_{\{a_i\}}.$$

En particulier, g est étagée (puisqu'elle prend au plus $2n + 1$ valeurs $y_1, ..., y_n$, $z_0, z_1, ..., z_n$), borélienne (c'est une combinaison linéaire de fonctions indicatrices d'intervalles, donc de boréliens), et intégrable par rapport à la mesure de Lebesgue ($g \in \mathcal{L}^1$) puisque

$$|g| = \sum_{1 \leqslant i \leqslant n} |y_i| \mathbb{1}_{]a_{i-1}, a_i[} + \sum_{0 \leqslant i \leqslant n} |z_i| \mathbb{1}_{\{a_i\}}$$

et donc

$$\int_{[a,b]} |g| \, dx = \sum_{1 \leqslant i \leqslant n} |y_i|(a_i - a_{i-1}) < \infty.$$

De plus, son intégrale de Lebesgue

$$\int_{[a,b]} g \, dx = \sum_{1 \leqslant i \leqslant n} y_i(a_i - a_{i-1})$$

coïncide avec son intégrale de Riemann.

(En revanche, une fonction étagée *n'est pas*, en général, en escalier, comme le montre l'exemple de la fonction de Dirichlet $\mathbb{1}_{\mathbb{Q}}$.)

La définition générale de l'intégrale de Riemann, ainsi que quelques unes de ses première propriétés, ont été rappelées dans le chapitre 1 ; nous noterons ici $I(f)$ l'intégrale de Riemann d'une fonction f quand elle existe. Une fonction $f : [a, b] \to \mathbb{R}$ est *Riemann-intégrable* s'il existe deux suites (g_n) et (h_n) de fonctions en escalier vérifiant

$$g_n \leqslant f \leqslant h_n \quad \text{et} \quad \lim I(g_n) = \lim I(h_n).$$

On définit alors l'*intégrale de Riemann* de f par l'égalité

$$I(f) = \lim I(g_n) = \lim I(h_n).$$

Proposition 12.1. *Si f est Riemann-intégrable, f est Lebesgue-mesurable et ses intégrales de Riemann et de Lebesgue coïncident.*

Démonstration. Choisissons deux suites (g_n) et (h_n) comme avant la proposition. Quitte à remplacer g_n par $\max(g_1, ..., g_n)$, on peut supposer que la suite (g_n) est croissante, et de même que la suite (h_n) est décroissante. On peut alors définir

$$g_\infty := \lim \uparrow g_n \leqslant f$$

et

$$h_\infty := \lim \downarrow h_n \geqslant f.$$

Les fonctions g_∞ et h_∞ sont mesurables et bornées. D'après la variante du théorème de convergence dominée de l'exercice 8.1, puisque $g_n \geqslant g_1 \in \mathcal{L}^1$ et $h_n \leqslant h_1 \in \mathcal{L}^1$,

$$\begin{cases} \int_{[a,b]} g_\infty \, d\lambda = \lim \uparrow \int_{[a,b]} g_n \, d\lambda = \lim \uparrow I(g_n) = I(f) \\ \int_{[a,b]} h_\infty \, d\lambda = \lim \downarrow \int_{[a,b]} h_n \, d\lambda = \lim \downarrow I(h_n) = I(f). \end{cases}$$

Donc

$$\int_{[a,b]} (h_\infty - g_\infty) \, d\lambda = 0.$$

Puisque $g_\infty \leqslant h_\infty$, l'égalité précédente entraîne

$$g_\infty = h_\infty \quad \text{p.p.}$$

Comme $g_\infty \leqslant f \leqslant h_\infty$, f coïncide p.p. avec une fonction borélienne. Donc f est Lebesgue-mesurable, i.e. mesurable de $([a,b], \bar{\mathcal{B}}([a,b]))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$). De plus,

$$\int_{[a,b]} f \, d\lambda = \int_{[a,b]} g \, d\lambda = I(f).$$

□

Remarque 12.2. On voit ici l'intérêt d'avoir complété la tribu borélienne de $[a,b]$: une fonction Riemann-intégrable est Lebesgue mesurable, mais généralement pas borélienne.

Remarque 12.3. Dans le chapitre 18, on verra que l'intégrale de Lebesgue peut être vue, en un certain sens, comme la complétion métrique de l'intégrale de Riemann. L'avantage de la théorie de Lebesgue est de réaliser, par construction, les éléments de l'espace complet (que nous appellerons l'espace L^1) comme des (classes de) fonctions, ce que le procédé abstrait de complétion d'un espace normé ne ferait pas.

Exercice ★ 12.1 (Fonctions bornées Riemann-intégrables [Rud76]).
Soit $f : [a,b] \to \mathbb{R}$ une fonction bornée. Montrer que f est Riemann-intégrable si et seulement si l'ensemble de ses points de discontinuités est négligeable.

Chapitre 13

Mesures produits

Soient $(E, \mathcal{A})$ et $(F, \mathcal{B})$ deux espaces mesurables. Rappelons la définition suivante.

Définition 13.1. La *tribu produit* de $E \times F$ est la tribu engendrée par l'ensemble des *pavés mesurables*, c'est-à-dire des parties de la forme $A \times B$ avec $A \in \mathcal{A}$ et $B \in \mathcal{B}$.

Exercice 13.1 (Caractérisation de la tribu produit)**.**
Montrer que la tribu produit est la plus petite tribu qui rende mesurables les deux projections canoniques $\pi_E : E \times F \to E$ et $\pi_F : E \times F \to F$.

Lemme 13.2. *1. Soit $C \in \mathcal{A} \otimes \mathcal{B}$. Pour tous $\xi \in E$ et $\eta \in F$, les sections*

$$\begin{cases} C_\xi := \{y \in F;\ (\xi, y) \in C\} \in \mathcal{B} \\ C^\eta := \{x \in E;\ (x, \eta) \in C\} \in \mathcal{A}. \end{cases}$$

sont mesurables.

2. Soit $f : E \times F \to G$ une application mesurable. Pour tous $\xi \in E$ et $\eta \in F$, les apllications partielles

$$f(\xi, \cdot) : y \mapsto f(\xi, y) \quad et \quad f(\cdot, \eta) : x \mapsto f(x, \eta)$$

sont respectivement $\mathcal{B}$-mesurable et $\mathcal{A}$-mesurable.

Démonstration. 1. Un $\xi \in E$ étant fixé, soit

$$\mathcal{C} = \{C \in \mathcal{A} \otimes \mathcal{B};\ C_\xi \in \mathcal{B}\}.$$

On a $\mathcal{C} \subset \mathcal{A} \otimes \mathcal{B}$ et l'on veut montrer l'inclusion inverse. Si $C = A \times B$,

$$C_\xi = \begin{cases} B & \text{si } \xi \in A \\ \varnothing & \text{si } \xi \notin A. \end{cases}$$

Donc $\mathcal{C}$ contient les pavés mesurables. De plus, $\mathcal{C}$ est une tribu (exercice). Puisque $\mathcal{A} \otimes \mathcal{B}$ est la plus petite tribu contenant les pavés mesurables, $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$.

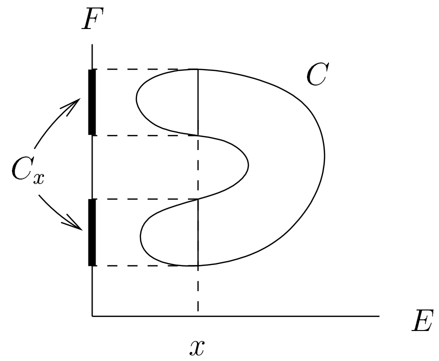


FIGURE 19 – Section d'une partie mesurable

2. Pour toute partie mesurable D de G,

$$f(\xi,\cdot)^{-1}(D) = \{y \in F;\ (\xi,y) \in f^{-1}(D)\} = \big(f^{-1}(D)\big)_\xi \in \mathcal{B}.$$

□

Exercice 13.2 (Mesurabilité sur les espaces produits)**.**
1. Montrer que $f = (f_1, f_2) : (G, \mathcal{C}) \to (E \times F, \mathcal{A} \otimes \mathcal{B})$ est mesurable si et seulement si ses composantes $f_1 : (G, \mathcal{C}) \to (E, \mathcal{A})$ et de $f_2 : (G, \mathcal{C}) \to (F, \mathcal{B})$ le sont.

L'objectif de la suite est de trouver un exemple d'ensemble non mesurable dont les sections le soient, et d'en déduire un exemple de fonction définie sur un espace produit, qui ne soit pas mesurable mais dont les applications partielles le soient.

Considérons le cas $E = F = \mathbb{R}$, $\mathcal{A} = \mathcal{B} = \sigma(\{x\}, x \in \mathbb{R}\})$ (tribu des parties dénombrables ou de complémentaire dénombrable), $\Delta = \{(x,x)\}_{x \in \mathbb{R}}$ (diagonale de $\mathbb{R}^2$), $f = \mathbb{1}_\Delta$. On veut montrer en particulier que Δ n'est pas dans la tribu produit $\mathcal{A}^{\otimes 2}$, mais cette tribu produit n'est pas facile à décrire. Nous allons construire une sur-tribu qui ne contient pas Δ. [1]

Soient $\mathcal{L}$ la classe des unions dénombrables de droites affines horizontales ou verticales dans $\mathbb{R}^2$; et $\mathcal{M}$ la classe des parties A de $\mathbb{R}^2$ telles que A ou A^c soit incluse dans une partie $L \in \mathcal{L}$.

2. Montrer que $\mathcal{M}$ est une tribu contenant $\mathcal{A}^{\otimes 2}$.

3. Montrer que $\Delta \notin \mathcal{M}$ et conclure.

(On peut rapprocher ce phénomène du fait qu'il existe des fonctions de deux variables continues par rapport à chacune des deux variables mais pas conjointement, comme par exemple

$$f : \mathbb{R}^2 \to \mathbb{R}, \quad (x,y) \mapsto \begin{cases} \frac{xy}{x^2+y^2} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{sinon.)} \end{cases}$$

Exercice ★ 13.3 (Produit des tribus de Lebesgue)**.**
1. Montrer que le produit de deux tribus de Lebesgues (sur $\mathbb{R}^d$ et sur $\mathbb{R}^\delta$) n'est

1. Merci à J. Lehec et É. Séré d'avoir rectifié ma construction erronée.

pas la tribu de Lebesgue de l'espace produit ($\mathbb{R}^{d+\delta}$).
2. Montrer que la tribu de Lebesgue de $\mathbb{R}^{d+\delta}$ est la complétion du produit des tribu de Lebesgue de $\mathbb{R}^d$ et de $\mathbb{R}^\delta$.

Supposons maintenant que $(E, \mathcal{E}, \mu)$ et $(F, \mathcal{F}, \nu)$ sont deux espaces mesurés.

Définition 13.3. La mesure μ est *σ-finie* si il existe une suite de parties mesurables E_n telles que $E = \cup_n E_n$ et $\mu(E_n) < +\infty$ pour tout n.

Pour construire une mesure sur le produit $(E \times F, \mathcal{A} \otimes \mathcal{B})$, il est commode de supposer que μ et ν sont σ-finies, ce que nous ferons.
Par exemple, $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda_n)$ est σ-fini (c'est la réunion des boules centrées en l'origine et de rayon $n \in \mathbb{N}$). En revanche, $\mathbb{R}^n$ muni de la mesure de comptage ne l'est pas. Cependant, la plupart des espaces mesurés que l'on rencontre en analyse (dont, trivialement, les espaces probabilisés) sont σ-finis.

Proposition 13.4 (Mesure produit). *Si $(E, \mathcal{A}, \mu)$ et $(F, \mathcal{B}, \nu)$ sont σ-finis, il existe une unique mesure $\mu \otimes \nu$ sur $(E \times F, \mathcal{A} \otimes \mathcal{B})$ telle que*

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B) \tag{13.1}$$

pour tout $A \in \mathcal{A}$, $B \in \mathcal{B}$. Cette mesure est σ-finie.

Démonstration. L'unicité d'écoule du corollaire 9.3 appliqué avec, pour ensemble $\mathcal{C}$ d'unicité, la classe des pavés mesurables.
Montrons l'existence de $m := \mu \otimes \nu$. Pour tout $C \in \mathcal{A} \otimes \mathcal{B}$, on voudrait poser

$$m(C) = \int_E \nu(C_x)\, \mu(dx). \tag{13.2}$$

D'après le lemme 13.2, les sections C_x sont dans $\mathcal{B}$ donc $\nu(C_x)$ est bien défini. Mais pour que la formule (13.2) ait un sens, il faut démontrer que l'application $x \mapsto \nu(C_x)$ est $\mathcal{A}$-mesurable.
Supposons d'abord ν finie. La classe

$$\mathcal{G} = \{C \in \mathcal{A} \otimes \mathcal{B};\ x \mapsto \nu(C_x) \text{ est } \mathcal{A}\text{-mesurable}\}$$

contient les pavés mesurables (parce que $\nu((A \times B)_x) = \mathbb{1}_A(x)\nu(B)$), dont l'ensemble est stable par intersection finie. Donc la classe monotone engendrée par $\mathcal{G}$ est la tribu engendrée par $\mathcal{G}$. Or, $\mathcal{G}$ est une classe monotone (si $C \subset C'$, $\nu(C' \backslash C) = \nu(C') - \nu(C)$ parce que ν est finie, et si (C_n) est une suite croissante, $\nu((\cup C_n)_x) = \lim \uparrow \nu((C_n)_x)$ par convergence monotone). Donc $\mathcal{G}$ est une tribu, contenant les pavés mesurables. Donc $\mathcal{G}$ est contenue dans $\mathcal{A} \otimes \mathcal{B}$, et lui est donc égale.
Dans le cas général où ν est seulement σ-finie, il existe une suite croissante de parties $B_n \in \mathcal{B}$ de mesure finie, telle que $F = \cup B_n$. Le même raisonnement que dans le cas fini montre que $x \mapsto \nu(B_n \cap C_x)$ est mesurable. Par suite, $x \mapsto \nu(C_x) = \lim \uparrow \nu(B_n \cap C_x)$ l'est aussi.

La fonction m ainsi définie est bien une mesure sur $(E \times E, \mathcal{A} \otimes \mathcal{B})$. En effet, si (C_n) est une suite de parties deux à deux disjointes dans $\mathcal{A} \otimes \mathcal{B}$,

$$\begin{aligned} m\left(\bigcup C_n\right) &= \int_E \nu\left(\bigcup (C_n)_x\right) \mu(dx) \\ &= \int_E \sum \nu((C_n)_x)\, \mu(dx) \\ &= \sum \int_E \nu((C_n)_x)\, \mu(dx) \qquad\qquad \text{(convergence monotone)} \\ &= \sum m(C_n). \end{aligned}$$

La formule (13.2) dit immédiatement que

$$m(A \times B) = \mu(A) \times \nu(B).$$

Et si l'on définit enfin m' par la formule symétrique

$$m'(C) = \int_F \mu(C^y)\, \nu(dy),$$

le corollaire 9.3 entraîne $m = m'$, d'où la dernière égalité voulue. □

Remarque 13.5. Le produit ainsi défini sur les mesures σ-finies est associatif :

$$(\mu_1 \otimes \mu_2) \otimes \mu_3 = \mu_1 \otimes (\mu_2 \otimes \mu_3)$$

parce que cette mesure est caractérisée par ses valeurs $\mu_1(A_1)\mu_2(A_2)\mu_3(A_3)$ sur les pavés mesurables $A_1 \times A_2 \times A_3$. Sans ambiguité on peut donc noter

$$\mu_1 \otimes \mu_2 \otimes \mu_3,$$

sans parenthèses, le produit de trois mesures σ-finies. Par récurrence, il en va de même pour le produit de n mesures σ-finies.

Loi conjointe et lois marginales

Soient X, Y deux variables aléatoires sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. On a déjà défini la *loi* de X, par exemple, comme la probabilité image de P par X :

$$P_X(A) := P(X^{-1}(A)) = P(X \in A) \quad (\forall A \in \mathcal{B}(\mathbb{R})).$$

Définition 13.6. La *loi conjointe* de X et de Y est la loi du vecteur aléatoire $(X, Y) : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$, caractérisée par :

$$P_{(X,Y)}(A \times B) := P(\{X \in A\} \cap \{Y \in B\}) \quad (\forall A, B \in \mathcal{B}(\mathbb{R})).$$

Remarques 13.7. — Le fait que la probabilité $P_{(X,Y)}$ soit caractérisée par la valeur qu'elle prend sur les pavés boréliens découle du corollaire 9.3 du théorème de classe monotone.

— X et Y seront qualifiées d'*indépendantes* si

$$P_{(X,Y)}(A \times B) := P(\{X \in A\})\, P(\{Y \in B\}) \quad (\forall A, B \in \mathcal{B}(\mathbb{R})).$$

Définition 13.8. Inversement, les *lois marginales* d'un vecteur aléatoire (X, Y) à valeurs dans $\mathbb{R}^2$ sont les lois de ses composantes.

Les lois marginales sont bien sûr déterminées par la loi conjointe :

$$P_X(A) = P_{(X,Y)}(A \times \mathbb{R}) \quad \text{et} \quad P_Y(A) = P_{(X,Y)}(\mathbb{R} \times A).$$

Autres exercices

Exercice 13.4 (Inégalité de Bell [Can13])**.**
Soient X, X', Y, Y' quatre variables aléatoires à valeurs dans $\{-1, 1\}$. Supposons les quatre variables définies sur un espace de probabilité commun $(\Omega, \mathcal{A}, P)$ (ce qu'on peut ne pas savoir a priori, dans une expérience aléatoire donnée). Alors le quadruplet possède une loi conjointe. Montrer l'inégalité de Bell [2] suivante est satisfaite :

$$|E(XY) + E(X'Y) + E(XY') - E(X'Y')| \leqslant 2;$$

on pourra pour cela calculer l'espérance de la variable $Z = (X+X')Y+(X-X')Y'$ et montrer que $|Z| \leqslant 2$.

(Cette inégalité très simple a eu un retentissement considérable parce que, en mécanique quantique, on peut construire des variables qui violent cette inégalité, en conformité d'ailleurs avec l'expérience. Ceci a conduit à abandonner certaines interprétations philosophiques classiques de la mécanique quantique.)

Problème 13.1 (Mesure de Bernoulli [3])**.**
Soient S un ensemble fini de cardinal $\geqslant 2$ et $\Omega = S^{\mathbb{N}}$ l'ensemble de tous les résultats possibles quand on tire au hasard un élément de S une infinité dénombrable de fois. En théorie de l'information, S joue le rôle d'un alphabet (fini) et Ω de l'ensemble des mots (de longueur infinie)

$$\omega = (\omega_0, \omega_1, \omega_2, \ldots);$$

les fonctions $\pi_i : \Omega \to S$, $\omega \mapsto \omega_i$ sont les projections canoniques, encore appelées *fonctions coordonnées.*

Un *cylindre de rang* n est une partie de Ω de la forme

$$C_H = \{\omega,\ (\omega_0, \ldots, \omega_n) \in H\} = \{(\pi_0, \ldots, \pi_n) \in H\}$$

pour une certaine partie $H \subset S^n$ (autrement dit, A est déterminée par une contrainte quelconque donnée, sur les n-premières composantes de ω). Soit $\mathcal{C}$ l'ensemble des cylindres de tous les rangs possibles.

1. Montrer que $\mathcal{C}$ est une algèbre (définition 11.11), mais pas une σ-algèbre.

2. John Stewart Bell, physicien britannique (1928–1990), célèbre pour ses travaux en mécanique quantique.

3. Jacob Bernoulli, mathématicien suisse (1654/6–1705), ayant travaillé en calcul différentiel, calcul des variations et probabilités. Il est notamment l'auteur de la première version de la loi des grands nombres.

Soient p_s, $s \in S$, des probabilités sur S. Définissons une fonction P sur $\mathcal{C}$

$$P(C_H) = \sum_{(\omega_0,\dots,\omega_n)\in H} p_{\omega_0}\cdots p_{\omega_n}.$$

Par exemple, si $H = \{(\omega_0^o, \dots, \omega_n^o)\}$, $P(C_H) = p_{\omega_0^o}\cdots p_{\omega_n^o}$.

2. Montrer que cette définition est consistante, c'est-à-dire que, si $C_H = C_K$ pour deux parties a priori distinctes $H \subset S^n$ et $K \subset S^m$,

$$\sum_{(\omega_0,\dots,\omega_n)\in H} p_{\omega_0}\cdots p_{\omega_n} = \sum_{(\omega_0,\dots,\omega_m)\in K} p_{\omega_0}\cdots p_{\omega_m}$$

3. Montrer que P est additive au sein de $\mathcal{C}$.

4. On souhaite montrer que P est σ-additive au sein de $\mathcal{C}$.

4.1. Montrer que, si (C_k) est une suite décroissante de cylindres non vides, la limite $\cap_k C_k$ est non vide ; on pourra utiliser un argument diagonal.

(Ce résultat est un cas particulier du théorème de Tychonov : si S est muni de la topologie discrète, l'espace topologique produit S^∞ est compact non vide.)

4.2. Montrer que P est extérieurement continue : si (C_k) est une suite décroissante de cylindres qui converge vers $\varnothing$, $P(C_k) \to 0$.

4.3. Conclure.

Donc P est une prémesure sur l'algèbre $\mathcal{C}$, qui, d'après le théorème de Hahn-Kolmogorov (problème 11.3), se prolonge en une mesure, encore notée P et appelée *mesure produit*, sur $(\Omega, \sigma(\mathcal{C}))$ (*théorème de Bernoulli*).

Problème 13.2 (Théorème de Kolmogorov [Bil95])**.**
Soient $(\Omega, \mathcal{A}, P)$ un espace de probabilité et T un ensemble ; typiquement, T est un intervalle réel, ou $\mathbb{N}$, ou $\mathbb{Z}$, et représente le temps. Une famille $(X_t)_{t\in T}$ de variables aléatoires s'appelle un *processus stochastique*. Si $t_1, \dots, t_k$ sont des éléments distincts de T, notons $\mu_{t_1\dots t_k}$ la loi du vecteur aléatoire

$$(X_{t_1}, \dots, X_{t_k}) : \Omega \to \mathbb{R}^k.$$

Ces *marginales de dimension finie* du processus satisfont les deux conditions de compatibilité suivantes :

— symétrie : si π est une permutation de $\{1, \dots, k\}$,

$$\mu_{t_1,\dots,t_k}(B_1 \times \cdots \times B_k) = \mu_{t_{\pi 1}\dots t_{\pi k}}(B_{\pi 1} \times \cdots \times B_{\pi k})$$

(par exemple, si $\mu_{s,t} = \nu \otimes \nu'$, $\mu_{t,s} = \nu' \otimes \nu$) ;

— compatibilité relativement aux restrictions :

$$\mu_{t_1\dots t_{k-1}}(B_1 \times \cdots \times B_{k-1}) = \mu_{t_1\dots t_{k-1}t_k}(B_1 \times \cdots \times B_{k-1} \times \mathbb{R}).$$

On se donne dorénavant une probabilité $\mu_{t_1\dots t_k}$ sur $\mathbb{R}^k$ pour tout $k \in \mathbb{N}_*$ et toute famille finie $(t_1, \dots, t_k)$ d'éléments distincts de T.

1. Montrer que ces probabilités satisfont les deux conditions de compatibilité précédentes si et seulement si

$$\mu_{t_1\dots t_k} = \mu_{u_1\dots u_m} \circ \psi^{-1},$$

pout toute application $\psi : \mathbb{R}^m \to \mathbb{R}^k$ de la forme

$$\psi(x_1, ..., x_m) = (x_{\pi^{-1}1}, ..., x_{\pi^{-1}k}),$$

où π est une permutation de $\{1, ..., k\}$ et

$$(u_{\pi^{-1}1}, ..., u_{\pi^{-1}m}) = (t_1, ..., t_k, t_{k+1}, ..., t_m).$$

Pour tout $t \in \mathbb{R}$, notons $Z_t : \mathbb{R}^T \to \mathbb{R}$ la t-ième projection canonique (ou t-ième fonction coordonnée) $x \mapsto x(t) = x_t$. Soit

$$\mathcal{B}^T = \sigma(\{Z_t,\ t \in T\});$$

c'est la généralisation de la tribu borélienne $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R})^{\otimes n}$ de $\mathbb{R}^n$, qui peut être vue comme la tribu engendrée par les fonctions coordonnées (cf. exercice exo :borel-produit). Par ailleurs, un *cylindre* de $\mathbb{R}^T$ est une partie de la forme et

$$C_{t_1 \dots t_k, B} = \{x \in \mathbb{R}^T,\ (x_{t_1}, ..., x_{t_k}) \in B\},$$

où $t_1, ..., t_k \in T$ et $B \in \mathcal{B}(\mathbb{R})$ (c'est la généralisation d'un pavé mesurable). On note enfin $\mathcal{B}_0^T$ l'ensemble des cylindres.

2. Montrer que $\mathcal{B}_0^T$ est une algèbre (définition 11.11) mais généralement pas une tribu.

L'objectif est de montrer le théorème de Kolmogorov suivant : si les probabilités $\mu_{t_1 \dots t_k}$ sont compatibles au sens de la première question, il existe une probabilité P sur $(\mathbb{R}^T, \mathcal{B}^T)$ telle que les marginales de dimension finie du processus $(Z_t)_{t \in T}$ soient les $\mu_{t_1 \dots t_k}$.

3. Montrer que, pour tout cylindre $C_{t_1 \dots t_k, B}$, on peut définir sans ambiguité

$$P(C_{t_1 \dots t_k, B}) = \mu_{t_1 \dots t_k}(B).$$

4. Montrer que P est finiment additive.

5. Montrer que P est σ-additive au sein de $\mathcal{B}_0^T$; on pourra commencer par montrer que, si $A_1, A_2, ... \in \mathcal{B}_0^T$ est une suite décroissante tendant vers l'ensemble vide, $P(A_n)$ tend vers 0.[4]

6. Conclure.

4. Une partie de l'argument à utiliser revient à démontrer que le produit dénombrable d'ensembles compacts est compact.

Chapitre 14

Intégration sur un espace produit

Soient $(E, \mathcal{A}, \mu)$ et $(F, \mathcal{B}, \nu)$ deux espaces mesurés σ-finis.

La formule dans le théorème suivant permet de ramener le calcul d'une intégrale par rapport à une mesure produit, au calcul de deux intégrales simples imbriquées.

C'est cette formule qui constitue le coeur du théorème. Les deux premières parties de l'énoncé ne sont que les propriétés qui permettent de donner un sens aux différentes intégrales de la formule.

Théorème 14.1 (Tonelli[1]). *Soit $f : E \times F \to [0, +\infty]$ une fonction mesurable.*

1. *Pour tout $x \in E$, $f(x, \cdot)$ est $\mathcal{B}$-mesurable et pour tout $y \in F$, $f(\cdot, y)$ est $\mathcal{A}$-mesurable.*
2. *La fonction $x \mapsto \int f(x, y)\, d\nu(y)$ est $\mathcal{A}$-mesurable et la fonction $y \mapsto \int f(x, y)\, d\mu(x)$ est $\mathcal{B}$-mesurable.*
3. *Enfin,*

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

Démonstration. 1. La mesurabilité des fonctions partielles a déjà été démontrée (lemme 13.2).

2. Si $f = \mathbb{1}_C$ avec $C \in \mathcal{A} \otimes \mathcal{B}$, on a déjà vu que les fonctions

$$x \mapsto \int f(x, y)\, d\nu(y) = \nu(C_x) \quad \text{et} \quad y \mapsto \int f(x, y)\, d\mu(x) = \mu(C^y)$$

sont respectivement $\mathcal{A}$-mesurable et $\mathcal{B}$-mesurable. Par linéarité, les fonctions

$$x \mapsto \int f(x, y)\, d\nu(y) \quad \text{et} \quad y \mapsto \int f(x, y)\, d\mu(x)$$

sont encore mesurables quand f est une fonction étagée positive. Par passage à la limite simple, elles le sont aussi quand f est réelle étendue positive,

1. Leonida TONELLI, mathématicien italien (1885–1946), connu pour sa première version du théorème ci-dessus, ainsi que pour ses travaux fondamentaux en calcul des variations, sur la semi-continuité de la fonctionnelle d'action.

puisque, si l'on choisit des fonctions f_n étagées positives croissant vers f quand n tend vers l'infini, d'après le théorème de convergence monotone,

$$\begin{cases} \int f(x,y)\,d\nu(y) = \lim\uparrow \int f_n(x,y)\,d\nu(y) \\ \int f(x,y)\,d\nu(y) = \lim\uparrow \int f_n(x,y)\,d\mu(x). \end{cases}$$

3. *Formule d'intégration.* Concentrons-nous sur la première égalité, la seconde se montrant de façon symétrique. Cette égalité est vraie si $f = \mathbb{1}_C$ avec $C \in \mathcal{A} \otimes \mathcal{B}$, puisqu'elle dit que

$$\mu \otimes \nu(C) = \int_E \nu(C_x)\,d\mu(x),$$

ce qui est la définition même de $\mu \otimes \nu$ dans la démonstration du théorème précédent. On en déduit le résultat voulu quand f est étagée positive (par linéarité des deux membres de l'égalité), puis quand f est mesurable positive étendue (par passage à la limite croissante en approchant f par une suite croissante de fonctions étagées positives).

□

Exercice 14.1 (Théorème de Tonelli pour les séries)**.**
Soit $(x_{n,m})_{n,m\in\mathbb{N}}$ une suite double réelle positive étendue (i.e. $x_{n,m} \in [0,+\infty]$ pour tous $n, m \in \mathbb{N}$). Soit c_d la mesure de comptage sur $\mathbb{N}^d$ (cf. la première partie de l'exercice 7.6).
1. Montrer que $c_1{}^{\otimes 2}(= c_1 \otimes c_1) = c_2$ et que

$$\int_{\mathbb{N}^2} x_{n,m}\,dc_2(n,m) = \sum_{(n,m)\in\mathbb{N}^2} x_{n,m}.$$

2. Montrer avec le théorème de Tonelli que

$$\sum_{(n,m)\in\mathbb{N}^2} x_{n,m} = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} x_{n,m} = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} x_{n,m}.$$

Cette démonstration donne un rôle plus symétrique aux deux sommations que celle décrite dans l'exercice 7.6 via le théorème de convergence monotone, et permet la généralisation (modeste) suivante.
Soient E et F deux ensembles quelconques, dénombrables ou pas, et $(x_{n,m})_{n\in E,m\in F}$ une suite de réels étendus indexée par $E \times F$.
3. Montrer que

$$\sum_{(n,m)\in E\times F} x_{n,m} = \sum_{n\in E}\sum_{m\in F} x_{n,m} = \sum_{m\in F}\sum_{n\in E} x_{n,m}.$$

(L'exercice 3.2 montre que, si ces sommes sont finies, l'ensemble des termes non nuls est dénombrable.)

Théorème 14.2 (Fubini[2])**.** *Soit* $f \in \mathcal{L}^1(E \times F, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$.

1. $\mu(dx)$*-p.p., la fonction* $f(x, \cdot)$ *est dans* $\mathcal{L}^1(F)$ *et*
   $\nu(dy)$*-p.p., la fonction* $f(\cdot, y)$ *est dans* $\mathcal{L}^1(E)$
2. $\int f(\cdot, y)\, d\nu(y)$ *est dans* $\mathcal{L}^1(E)$ *et*
   $\int f(x, \cdot)\, d\mu(x)$ *est dans* $\mathcal{L}^1(F)$.
3. *Enfin,*

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

Remarques 14.3. — Dans le deuxième point de l'énoncé, par exemple la fonction $F : x \mapsto \int f(\cdot, y)\, d\nu(y)$ n'est a priori définie qu'en dehors d'une partie négligeable N de E, d'après le premier point. Il est implicite qu'on prolonge F sur N en posant par exemple $F|_N = 0$, sachant de toute façon que l'intégrabilité de F ainsi que la valeur de l'intégrale de F ne dépendent pas de ce prolongement (proposition 7.4 et 7.9).
— D'autre part, les mêmes conclusions restent valables si f est une fonction complexe $\in \mathcal{L}^1_{\mathbb{C}}(E \times F, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$.

Démonstration. 1. On sait déjà que l'application partielle $f(x, \cdot)$ est mesurable. D'après le théorème de Tonelli appliqué à $|f|$,

$$\int_E \left(\int_F |f(x,y)|\, d\nu(y) \right) d\mu(x) = \int |f|\, d\mu \otimes \nu < \infty,$$

ce qui entraîne que, $\mu(dx)$-p.p., $\int_F |f(x,y)|\, d\nu(y) < \infty$, i.e. $f(x, \cdot) \in \mathcal{L}^1(F)$.

2. En utilisant la décomposition $f = f^+ - f^-$ de f en ses parties positive et négative, et en utilisant le théorème de Tonelli, on voit que la fonction de x

$$\int f(\cdot, y)\, d\nu(y) = \int f^+(\cdot, y)\, d\nu(y) - \int f^-(\cdot, y)\, d\nu(y)$$

est $\mathcal{A}$-mesurable et que l'intégrale de sa valeur absolue vaut

$$\begin{aligned} \int_E \left| \int_F f(x,y)\, d\nu(y) \right| d\mu(x) &\leqslant \int_E \left(\int_F |f(x,y)|\, d\nu(y) \right) d\mu(x) \\ &= \int |f|\, d\mu \otimes d\nu < \infty, \end{aligned}$$

donc enfin que $\int f(\cdot, y)\, d\nu(y) \in \mathcal{L}^1(E)$.

3. La première égalité de la dernière formule voulue s'obtient par différence des égalités analogues pour les parties positive et négative de f,

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

□

2. Guido Fubini, mathématicien italien (1879–1943), principalement connu pour ce théorème d'intégration sur les espaces produits, ainsi que pour sa découverte de la métrique de Fubini-Study en géométrie kählérienne.

Exercice 14.2 (Théorème de Fubini pour les séries).
Si $(x_{n,m})_{n,m\in\mathbb{N}}$ est une suite double réelle (ou complexe), sommable ($\sum_{(n,m)\in\mathbb{N}^2}|x_{n,m}|<\infty$),

$$\sum_{(n,m)\in\mathbb{N}^2} x_{n,m} = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} x_{n,m} = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} x_{n,m}.$$

Exercice 14.3 (Contre-exemple à la sommabilité).
La conclusion du théorème de Fubini est-elle valide dans le cas de la suite double réelle $(x_{n,m})_{,n,m\in\mathbb{N}}$ définie par

$$x_{n,m} = \begin{cases} +1 & \text{si } n=m \\ -1 & \text{si } n=m+1 \\ 0 & \text{sinon ?} \end{cases}$$

Exercice 14.4 (Interprétation de l'intégrale comme aire).
Soient $(E,\mathcal{A},\mu)$ un espace mesuré σ-fini et $f : E \to [0,+\infty]$ une fonction positive étendue. Montrer que f est mesurable si et seulement si l'hypographe de f

$$H_f := \{(x,t)\in E\times\mathbb{R},\ 0\leqslant t\leqslant f(x)\}$$

est mesurable dans $(E\times\mathbb{R},\mathcal{A}\otimes\mathcal{B}(\mathbb{R}))$, auquel cas

$$\mu\otimes\lambda(H_f) = \int_E f(x)\,d\mu(x).$$

Autres exercices

Exercice 14.5 (Applications élémentaires du théorème de Fubini).
Dans cet exercice, pour les questions d'intégrabilité en dimension un, on peut : soit utiliser ce qui est connue de l'intégrale de Riemann et de l'égalité entre les intégrales de Lebesgue et de Riemann ; soit retrouver directement ces résultats directement avec l'intégrale de Lebesgue, en admettant la formule du changement de variable et la formule fondamentdale du calcul différentiel.

1. Étudier l'intégrabilité de

$$f(x,y) = \frac{1}{(1+x+y)^\alpha}$$

sur $[0,+\infty[^2$ en fonction du paramètre $\alpha\in\mathbb{R}$, et calculer l'intégrale de f dans les cas où cette intégrale est finie.

2. Utiliser le fait que $1/x = \int_0^\infty e^{-xt}\,dt$ pour montrer que

$$\lim_{a\to\infty}\int_0^a \frac{\sin x}{x}\,dx = \frac{\pi}{2}.$$

Puis montrer que pourtant la fonction $\sin x/x$ n'est pas intégrable sur $[0,+\infty[$; on pourra, en raisonnant par l'absurde, en déduire que $\sin^2 x/x$ serait elle-même intégrable, et montrer que ceci conduit à une contradiction.

3. Soient $0<a<b$ deux réels et soit f la fonction réelle sur $[0,1]\times[a,b]$ définie par $f(x,y)=x^y$. On muni $[0,1]\times[a,b]$ de la tribu borélienne. Montrer que f est

intégrable par rapport à la mesure de Lebesgue et en déduire la valeur de

$$\int_0^1 \frac{x^b - x^a}{\ln x}\, dx.$$

4. Pour tout $y \in]0, +\infty[$, notons $[y]$ la partie entière de y, $q(y) = [y/\pi] + 1$ et $r(y) = y - [y/\pi]\pi$. Soit $f : \mathbb{R}_*^{+2} \to \mathbb{R}$ la fonction définie par

$$f(x, y) = e^{-xq(y)^2\sqrt{r(y)}}.$$

La fonction f est-elle intégrable par rapport à la mesure de Lebesgue sur $]0, +\infty[^2$? (On pourra même calculer son intégrale en utilisant le fait que $\sum_{n\geqslant 1} n^{-2} = \pi^2/6$.)

5. Pour $\alpha \in \mathbb{R}$, soit $f_\alpha : \mathbb{R}^2 \to \mathbb{R}$ la fonction définie par $f_\alpha(x, y) = e^{-x^2-\alpha xy-y^2}$. Déterminer les valeurs de α pour lesquelles f_α est intégrable par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$ et, quand elle est intégrable, calculer son intégrale I_α en utilisant le fait que $\int_{\mathbb{R}} e^{-x^2}\, dx = \sqrt{\pi}$.

6. Montrer que

$$\int_{\mathbb{R}} e^{-x^2}\, dx = \sqrt{\pi}\ ;$$

on pourra utiliser l'astuce qui consiste à élever cette intégrale au carré, à convertir par application du théorème de Fubini le résultat en une intégrale sur $\mathbb{R}^2$, puis à passer en coordonnées polaires.

Exercice 14.6 (Racines de l'équation de degré deux)**.**
Soient a, b, c trois variables aléatoires réelles indépendantes de loi uniforme sur $[0, 1]$. On considère l'équation

$$ax^2 + 2bx + c = 0.$$

Calculer la probabilité des événements suivants :
$A = \{$ les racines sont réelles distinctes $\}$
$B = \{$ les racines sont complexes non réelles $\}$
$C = \{$ il y a une racine double $\}$.

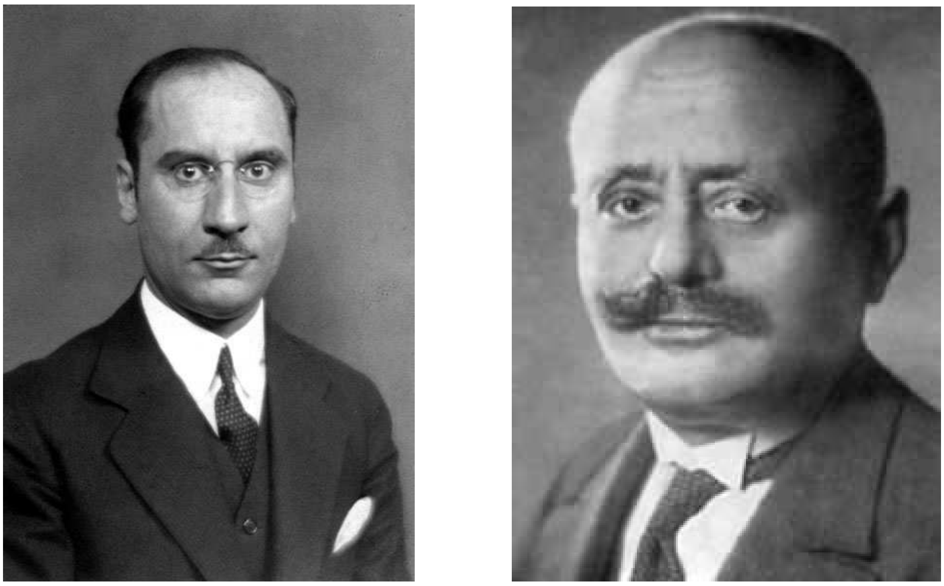


FIGURE 20 – Leonida TONELLI (1885–1946) et Guido FUBINI (1879–1943), deux illustres représentants de l'école italienne d'analyse

Chapitre 15

Changements de variable dans $\mathbb{R}^n$

Dans la formule de transfert (théorème 7.14), on ne sait pas a priori quelle est la loi du changement de variable. En revanche, dans le cas où cette application est un difféomorphisme entre deux ouverts de $\mathbb{R}^n$, et où la mesure initiale est la mesure de Lebesgue, on sait dire précisément ce qu'est la mesure image : c'est une mesure à densité par rapport à la mesure de Lebesgue, dont la densité est le déterminant jacobien. Ceci précise considérablement la formule de transfert.

Il faut déjà comprendre le cas des changements de variables affines.

Proposition 15.1. *Soient M une matrice réelle $d \times d$ inversible, $a \in \mathbb{R}^d$ et φ : $\mathbb{R}^d \to \mathbb{R}^d$, $x \mapsto Mx + a$. La mesure image de λ par φ^{-1} vérifie*

$$\lambda(\varphi(A)) = |\det M|\,\lambda(A) \quad (\forall A \in \mathcal{B}(\mathbb{R}^d)).$$

Cette formule remarquable dit en particulier que le facteur par lequel l'aire d'une figure est multiplié quand on transforme la figure par φ ne dépend pas de la figure !

Démonstration. Comme la bijection réciproque φ^{-1} est continue (de même que φ), pour tout borélien A, $\varphi(A) = (\varphi^{-1})^{-1}(A)$ est borélien, donc le membre de gauche de l'égalité annoncée a bien un sens. On cherche donc à calculer $\lambda(\varphi(A))$ en fonction de $\lambda(A)$.

Comme la mesure de Lebesgue est invariante par translation (première partie du théorème 11.9),

$$\lambda(\varphi(A)) = \lambda(MA + a) = \lambda(MA).$$

De plus, pour tous $b \in \mathbb{R}^d$ et $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\lambda(M(b + A)) = \lambda(Mb + MA) = \lambda(MA).$$

Donc la mesure $A \mapsto \lambda(MA)$ est invariante par translation. Enfin, cette mesure est finie sur les parties compactes de $\mathbb{R}^d$, puisque M, en tant qu'application linéaire entre des espaces de dimension finie, est bornée, et que λ possède elle-même cette propriété. Donc, d'après la seconde partie du théorème 11.9, il existe un réel $c \geqslant 0$ tel que, pour tout $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\lambda(MA) = c\lambda(A). \tag{15.1}$$

Le plus gros est fait, il ne reste qu'à montrer que $c = |\det M|$, en testant l'égalité (15.1) sur un borélien bien choisi. Nous allons en fait procéder en trois étapes, selon la nature de M :

— Si M est orthogonale et B^d désigne une boule euclidienne de $\mathbb{R}^d$, M est isométrique donc $MB^d = B^d$, ce qui montre que $c = 1 = |\det M|$.

— Si M est symétrique positive, donc définie (puisque inversible), il existe une matrice orthogonale P telle que

$$P^{-1}MP = \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{pmatrix}, \quad \alpha_1, ..., \alpha_d > 0.$$

Pour trouver c dans ce cas, nous allons tester l'égalité (15.1) sur l'image $A = P([0,1]^d)$ par P de la "boule ℓ^∞" : comme le diagramme

$$\begin{tikzcd} A = P([0,1]^d) \arrow[r, "M"] & MA = P(\prod_i [0,\alpha_i]) \\ {[0,1]^d} \arrow[u, "P"] \arrow[r, "P^{-1}MP"] & \prod_i [0,\alpha_i] \arrow[u, "P"'] \end{tikzcd}$$

commute,

$$\begin{aligned} \lambda(MP([0,1]^d))) &= \lambda(P(\prod_i [0,\alpha_i])) && \text{(diagramme commutatif)} \\ &= \lambda(\prod_i [0,\alpha_i]) && (P \text{ orthgonale}) \\ &= \prod_i \alpha_i = \det M && (\text{définition de } \lambda) \\ &= (\det M)\, \lambda([0,1]^d) \end{aligned}$$

et $c = \det M$.

— Dans le cas général, on peut toujours décomposer M en le produit PS d'une matrice P orthogonale et d'une matrice S symétrique définie positive (prendre $S = ({}^tMM)^{1/2}$ et $P = MS^{-1}$). D'après les deux cas qui précèdent,

$$c = |\det P|\,|\det S| = |\det M|.$$

□

Remarque 15.2. Si M n'est pas inversible, son image est contenue dans un hyperplan, et l'on se convaincra facilement que la mesure image de λ par φ ne prend que les deux valeurs 0 ou ∞.

Une application du rapport entre aire et déterminant est la suivante.

Exercice 15.1 (Deuxième loi de Kepler)**.**
Soit $x : \mathbb{R} \to \mathbb{R}^2$ une courbe paramétrée vérifiant l'équation différentielle

$$x''(t) = f(x(t))\, x(t),$$

où $f : \mathbb{R}^2 \to \mathbb{R}$ est une fonction de classe C^1 ; autrement dit, on suppose que l'accélération est parallèle à la position. Montrer que l'aire $A(t)$ balayée par le vecteur $x(t)$ croît linéairement avec le temps (deuxième loi de Kepler).

Soient U et V deux ouverts de $\mathbb{R}^d$, et $\varphi : U \to V$ un difféomorphisme.* Si l'on note $x = (x_1, ..., x_d)$ les coordonnées des points de U et $\varphi = (\varphi_1, ..., \varphi_d)$ celles de φ, la *matrice jacobienne* de φ en x est la matrice $d \times d$

$$\varphi'(x) = \left(\frac{\partial\varphi}{\partial x_1}(x), \quad \cdots, \quad \frac{\partial\varphi}{\partial x_d}(x)\right) = \begin{pmatrix} \frac{\partial\varphi_1}{\partial x_1}(x) & \cdots & \frac{\partial\varphi_1}{\partial x_d}(x) \\ \vdots & & \vdots \\ \frac{\partial\varphi_d}{\partial x_1}(x) & \cdots & \frac{\partial\varphi_d}{\partial x_d}(x) \end{pmatrix}$$

et le *(déterminant) jacobien* de φ en x est

$$J\varphi(x) = \det \varphi'(x).$$

On notera dx et dy les mesures de Lebesgue respectivement sur U et V ; cette notation n'utilise pas la lettre "d" de façon cohérente, mais elle pratique, et sans risque de confusion.

Dans le cas de la proposition 15.1, où $\varphi(x) = Mx + a$, la formule démontrée devient, avec les nouvelles notations,

$$\varphi^{-1}(dy) = |\det M|\, dx. \tag{15.2}$$

Le théorème suivant donne la formule analogue dans le cas général, pour laquelle il suffit de remplacer $\det M$ par le jacobien (maintenant variable) de φ.

Théorème 15.3 (Formule du changement de variable)**.** *1. La mesure image de dy par φ^{-1} est la mesure de densité $|\det \varphi'|$ par rapport à dx :*

$$\varphi^{-1}(dy) = |J\varphi(x)|\, dx. \tag{15.3}$$

2. Pour toute fonction borélienne $f : V \to \mathbb{R}_+$,

$$\int_V f(y)\, dy = \int_U f(\varphi(x))\, |J\varphi(x)|\, dx. \tag{15.4}$$

3. Pour toute fonction borélienne $f : V \to \mathbb{R}$, $f \in \mathcal{L}^1(dy)$ si et seulement si $f \circ \varphi\, |\det \varphi'| \in \mathcal{L}^1(dx)$, auquel cas la formule (15.4) *est encore vraie.*

Remarque 15.4. La valeur absolue, autour du déterminant jacobien, n'a rien de surprenant, puisque ce facteur est la densité (forcément positive) de la nouvelle mesure par rapport à la mesure de Lebesgue.

En dimension $d = 1$ quand $U = [a, b]$, on écrit certes souvent cette formule

$$\int_{\varphi(a)}^{\varphi(b)} f(y)\, dy = \int_a^b f(\varphi(x))\, \varphi'(x)\, dx,$$

sans valeur absolue. Mais, dans le cas où $\varphi' < 0$, c'est-à-dire où φ renverse l'orientation, cela revient bien à écrire

$$\begin{aligned} \int_{\varphi(b)}^{\varphi(a)} f(y)\, dy &= \int_a^b f(\varphi(x))\, (-\varphi'(x))\, dx \\ &= \int_a^b f(\varphi(x))\, |\varphi'(x)|\, dx \end{aligned}$$

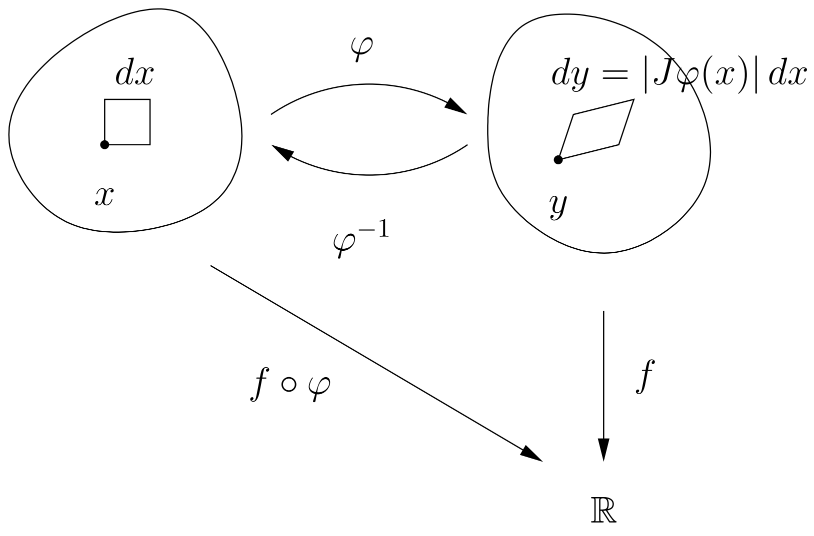


FIGURE 21 – Transformation infinitésimale des volumes

(autrement dit, il n'y a pas de signe négatif dans la version “Lebesgue” de la formule du changement de variable parce qu'il n'y a pas non plus d'orientation dans l'intervalle d'intégration).

Démonstration du théorème. Montrons d'abord que la première affirmation implique les deux suivantes. Si f est borélienne positive sur V,

$$\begin{aligned}
\int_V f(y)\,dy &= \int_V f \circ \varphi \circ \varphi^{-1}(y)\,dy && (\varphi \circ \varphi^{-1} = \mathrm{id}\,) \\
&= \int_U f \circ \varphi\, d(\varphi^{-1}(dy)) && \text{(formule de transfert appliquée à } \varphi^{-1}) \\
&= \int_U f \circ \varphi\, |J\varphi(x)|\,dx && \text{(première affirmation du théorème).}
\end{aligned}$$

Si f est borélienne de signe quelconque sur V, la caractérisation de son intégrabilité se montre en appliquant la formule précédente à $|f|$, après quoi la formule du changement de variable se déduit pour f de celle appliquée aux parties positive et négative de f.

On en est réduit à démontrer la première partie du théorème, c'est-à-dire que, pour toute partie borélienne A de U,

$$\lambda(\varphi(A)) = (|J\varphi|\,dx)(A) = \int_A |J\varphi(x)|\,dx. \tag{15.5}$$

Le lemme suivant permet de se ramener, à une petite erreur près, à une transformation linéaire décrite dans la proposition 15.1.

Lemme 15.5. *Soit K un compact de U. Quel que soit $\epsilon > 0$, il existe $\delta > 0$ tel que, pour tout cube C d'arêtes parallèles aux axes, de centre $x_0 \in K$ et de côté de longueur $< \delta$,*

$$(1-\epsilon)\,|J\varphi(x_0)|\,\lambda(C) \leqslant \lambda(\varphi(C)) \leqslant (1+\epsilon)\,|J\varphi(x_0)|\,\lambda(C).$$

Démonstration du lemme. Si $\delta > 0$ est assez petit, d'une part $\delta < \frac{1}{d}\mathrm{dist}(K, U^c)$; d'autre part, comme φ' est continue et comme K est compact, pour tout $x_0 \in K$ et $x \in \mathbb{R}^d$ tel que $\|x - x_0\| < d\delta$,

$$\|\varphi(x) - \varphi(x_0) - \varphi'(x_0) \cdot (x - x_0)\| \leqslant \epsilon\|x - x_0\|.$$

Si l'on note $f(v) = \varphi(x_0) + \varphi'(x_0) \cdot v$ ($v \in \mathbb{R}^d$), et si $\|x - x_0\| < d\delta$,

$$\varphi(x) = f(x - x_0) + h(x, x_0), \quad \text{avec} \quad \|h(x, x_0)\| \leqslant \epsilon\|x - x_0\|.$$

En posant $g(x, x_0) = \varphi'(x_0)^{-1} \cdot h(x, x_0)$, on obtient

$$\varphi(x) = f(x - x_0 + g(x, x_0)), \quad \text{avec} \quad \|g(x, x_0)\| \leqslant a\epsilon\|x - x_0\|,$$

avec $a := \max_{v \in K} \|\varphi'(z)\| < \infty$.
Soient C un cube centré en x_0 et de côté $r < \delta$, et $\tilde{C}$ le cube centré en 0 translaté de C. On a

$$\varphi(C) \subset f((1 + da\epsilon)\tilde{C}),$$

donc

$$\begin{aligned}
\lambda(\varphi(C)) &\leqslant \lambda(f((1 + da\epsilon)\tilde{C})) && \text{(croissance)} \\
&= |J\varphi(x_0)|\lambda((1 + da\epsilon)\tilde{C}) && \text{(proposition précédente)} \\
&= (1 + da\epsilon)^d |J\varphi(x_0)|\lambda(C) && \text{(homogénéité),}
\end{aligned}$$

ce qui donne la majoration souhaitée. La minoration apparaît de façon analogue. □

Fin de la démonstration du théorème. Soit $\mathcal{C}_n$ l'ensemble des *cubes élémentaires d'ordre* n, c'est-à-dire des cubes de la forme

$$C = \prod_{1 \leqslant j \leqslant d} \left] \frac{k_j}{2^n}, \frac{k_j + 1}{2^n} \right[, \quad k_j \in \mathbb{Z}.$$

Soient C_0 un cube élémentaire d'ordre n_0 fixé, tel que $\tilde{C_0} \subset U$, et $\epsilon > 0$. Fixons $n \geqslant n_0$ assez grand pour que

— la conclusion du lemme soit vraie pour $K = \tilde{C_0}$ et $\delta = 2^{-n}$,

— et pour tous $x, x' \in K$ tels que $\|x - x'\| \leqslant d\delta$,

$$(1 - \epsilon)|J\varphi(x)| \leqslant |J\varphi(x')| \leqslant (1 + \epsilon)|J\varphi(x)|. \tag{15.6}$$

Alors, en notant x_C le centre d'un cube C,

$$\begin{aligned}
\lambda(\varphi(C_0)) &= \sum_{C \in \mathcal{C}_n,\, C \subset C_0} \lambda(\varphi(C)) && \text{(additivité)} \\
&\leqslant (1 + \epsilon) \sum_{C \in \mathcal{C}_n,\, C \subset C_0} |J\varphi(x_C)|\, \lambda(C) && \text{(lemme)} \\
&\leqslant (1 + \epsilon)^2 \sum_{C \in \mathcal{C}_n,\, C \subset C_0} \int_C |J\varphi(x)|\, dx && \text{(équation (15.6))} \\
&= (1 + \epsilon)^2 \int_{C_0} |J\varphi(x)|\, dx && \text{(additivité).}
\end{aligned}$$

De façon analogue, on obtient la minoration

$$\lambda(\varphi(C_0)) \geqslant (1-\epsilon)^2 \int_{C_0} |J\varphi(x)|\, dx.$$

En passant à la limite quand ϵ tend vers 0, on obtient l'égalité (15.5) dans le cas d'un cube élémentaire C_0 quelconque pourvu que son adhérence soit contenue dans U. Il s'agit de comparer

— la mesure μ image de λ par φ^{-1} : $\mu(A) = \lambda(\varphi(A))$ pour tout $A \in \mathcal{B}(U)$

— et la mesure ν de densité $|J\varphi|$ par rapport à λ sur A : $\nu(A) = \int_A |J\varphi(x)|\, d\lambda$ pour tout $A \in \mathcal{B}(U)$.

Ces deux mesures coïncident sur tout cube élémentaire C d'adhérence contenue dans U. La classe de ces cubes est stable par intersection finie, et engendre la tribu borélienne de U. D'après le théorème de classe monotone, on bien $\mu = \nu$. □

Exercice 15.2 (Intégrale curviligne)**.**
Soient I et J deux intervalles compacts de $\mathbb{R}$, $\alpha : I \to \mathbb{R}^3$ et $\beta : J \to \mathbb{R}^3$ deux applications injectives de classe C^1 et $\phi : I \to J$, $s \mapsto t = \phi(s)$ un difféomorphisme* de classe C^1 tels que $\alpha = \beta \circ \phi$. Les deux chemins α et β sont donc deux paramétrages de la courbe $C = \alpha(I) = \beta(J)$. Notons λ_I et λ_J les restrictions de la mesure de Lebesgue respectivement à I et à J.

1. Montrer que si $f : \alpha(I) \to \mathbb{R}^+$ est une fonction mesurable positive on a

$$\int_I f \circ \alpha \ \|\alpha'\| \ d\lambda_I = \int_J f \circ \beta \ \|\beta'\| \ d\lambda_J.$$

2. En déduire que les mesures $\alpha(\|\alpha'\|\, \lambda_I)$ (image par α de la mesure de densité $\|\alpha'\|$ par rapport à λ_I) et $\beta(\|\beta'\|\, \lambda_J)$ coïncident. Notons l_c cette mesure de $\mathbb{R}^3$.

3. Interpréter l'égalité de la question précédente et donner une formule pour la longueur $L(C)$ de la courbe C faisant intervenir l_C.

4. Calculer la longueur d'un cercle de rayon $r > 0$.

Soit $\hat{\rho}$ une mesure sur I et soit ρ la mesure image de $\|\alpha'\|\, \hat{\rho}$ par α. La mesure ρ peut s'interpréter comme une répartition de masse, de charge électrique, etc. le long de C.

5. Donner une expression de la longueur de la cardioïde dont l'équation en coordonnées polaires est

$$r = 1 + \cos\theta, \quad \theta \in [-\pi, \pi],$$

et calculer la masse de cette cardioïde en supposant que sa répartition de masse est donnée par la mesure

$$\hat{\rho} = \delta_0 + \frac{1}{\sqrt{1+\sin\theta}}\, \lambda_{[-\pi,\pi]},$$

où δ_0 est la mesure de Dirac en $\theta = 0$.

Remarque 15.6. L'exercice qui précède montre que l'intégrale d'une fonction scalaire le long d'une courbe dépend en général du paramétrage choisi. En revanche, si l'on définit l'intégrale de la 1-forme différentielle (champ de formes linéaires) $\alpha = g(s)\,ds$ par la formule

$$\int_C \alpha = \int g(s)\,ds,$$

on obtient un résultat indépendant des coordonnées choisies pour peu que celles-ci préservent l'orientation. En effet, si $s = \phi(t)$ avec $\phi'(t) > 0$, $\alpha = g(\phi(t))\,\phi'(t)\,dt$ (c'est ici que l'on voit qu'une forme différentielle ne se transforme pas par changement de coordonnées de la même façon qu'une fonction, c'est-à-dire par simple composition) et donc,

$$\int g(s)\,ds = \int g(\phi(t))\,\phi'(t)\,dt.$$

La généralisation de ceci aux dimension supérieures conduit au concept de forme différentielle, pour lequel nous renvoyons par exemple au livre [Arn76]. L'exercice suivant peut être interprété en terme de formes différentielles de degré deux.

Problème 15.1 (Intégrale de surface)**.**
Soient K et L deux compacts de $\mathbb{R}^2$, $\alpha : K \to \mathbb{R}^3$ et $\beta : L \to \mathbb{R}^3$ deux surfaces paramétrées injectives de classe C^1 et $\phi : K \to L$, $(s,t) \mapsto (u,v) = \phi(s,t)$ un difféomorphisme de classe C^1 tels que $\alpha = \beta \circ \phi$. Les deux applications α et β sont deux paramétrages de la surface $S = \alpha(K) = \beta(L)$.

1. Montrer que si $f : \mathbb{R}^3 \to \mathbb{R}^+$ une fonction mesurable positive on a

$$\int_K f \circ \alpha \left\| \frac{\partial \alpha}{\partial s} \wedge \frac{\partial \alpha}{\partial t} \right\| ds \otimes dt = \int_L f \circ \beta \left\| \frac{\partial \beta}{\partial u} \wedge \frac{\partial \beta}{\partial v} \right\| du \otimes dv.$$

2. En déduire une mesure σ_S sur $\mathbb{R}^3$ qui dépend de S mais pas son paramétrage. En déduire une formule pour l'aire $A(S)$ de la surface S, que l'on justifiera rapidement.

3. Calculer l'aire de la sphère et du tore, respectivement d'équations

$$\mathbb{S}^2 : x^2 + y^2 + z^2 = 1 \quad \text{et} \quad \mathbb{T}^2 : (\sqrt{x^2 + y^2} - a)^2 + z^2 = r^2$$

$(0 \leqslant r \leqslant a)$.

4. Montrer que pour la sphère la mesure $\sigma_{\mathbb{S}^2}$ est invariante par rotation.

Considérons maintenant un compact K de $\mathbb{R}^2$ et une fonction positive $f : K \to \mathbb{R}^+$ de classe C^1. Notons

$$S = \{(x,y,z) \in \mathbb{R}^3,\ z = f(x,y)\}$$

le graphe de f.

5. Trouver un paramétrage $\alpha : K \to \mathbb{R}^3$ de S et en déduire une expression de l'aire de S comme une intégrale sur K.

6. Calculer cette aire dans le cas où $K = \{(x,y) \in \mathbb{R}^2,\ x^2 + y^2 \leqslant 1\}$ et où $f(x,y) = \frac{1}{2}(x^2 + y^2)$.

7. En revenant au cas général du début de l'exercice, montrer que

$$\left\| \frac{\partial \alpha}{\partial s} \wedge \frac{\partial \alpha}{\partial t} \right\|^2 = \left\| \frac{\partial \alpha}{\partial s} \right\|^2 \left\| \frac{\partial \alpha}{\partial t} \right\|^2 - \left(\frac{\partial \alpha}{\partial s} \cdot \frac{\partial \alpha}{\partial t} \right)^2$$

et en déduire une formule qui donne l'aire d'une surface dans $\mathbb{R}^n$ en fonction de son paramétrage $\alpha : K \subset \mathbb{R}^2 \to \mathbb{R}^n$.

La moindre des choses pour notre théorie de l'intégration est de permettre de retrouver les formules donnant le volume de certains solides, telle qu'Archimède de Syracuse[1] savait les calculer il y a plus de 20 siècles, par la célèbre méthode d'exhaustion.

Exercice 15.3 (Calculs de volumes de solides).
1. Calculer le volume de l'ellipsoïde solide de demi grands axes a, b et c ; on rappelle que ce solide a pour équation $x^2/a^2 + y^2/b^2 + z^2/c^2 \leqslant 1$.
2. Pour $a > r > 0$, calculer le volume du tore solide $\hat{A}$ obtenu par révolution autour de l'axe des z de

$$A = \{(y, z) \ : \ (y - a)^2 + z^2 \leqslant r^2\}.$$

3. Soit A un borélien du demi-plan (y, z), $y \geqslant 0$. Montrer que l'ensemble $\hat{A}$ de $\mathbb{R}^3$ obtenu en le faisant tourner autour de l'axe Oz est borélien et que son volume vaut

$$V = 2\pi \int_A y \, dy \, dz.$$

4. Montrer que le domaine sous le *simplexe standard de dimension* n, soit l'ouvert de $\mathbb{R}^n_+$ défini par l'inéquation

$$x_1 + \cdots + x_n \leqslant 1,$$

a pour volume $V_n = \frac{1}{n!}$.
5. Calculer le volume de la boule $\mathbb{B}^n$ de rayon 1 en dimension n ; on pourra faire une récurrence sur la dimension.

Autres exercices

Exercice 15.4 (Calculs de densités).
Soit X une variable aléatoire de densité f par rapport à la mesure de Lebesgue. Montrer que les deux variables aléatoires suivantes admettent une densité, que l'on calculera :

$$|X| \quad \text{et} \quad a\sin(X + \alpha) \quad (a > 0, \alpha \in \mathbb{R}).$$

Problème ★ 15.2 (Mesurabilité et applications linéaires).
1. Soit $\varphi : \mathbb{R}^d \to \mathbb{R}^n$ une application linéaire injective. Montrer que, si A est un borélien de $\mathbb{R}^d$, $\varphi(A)$ est un borélien de $\mathbb{R}^n$.

L'affirmation analogue est fausse en général si l'on ne suppose pas φ injective. En fait, il existe un borélien de $\mathbb{R}^2$ dont la projection canonique sur l'axe des abscisses n'est pas borélienne. Ce fait surprenant fut démontré par Souslin[2] en 1917. Auparavant, on pensait que ce n'était pas possible, et Lebesgue, commettant un erreur

1. Archimède de Syracuse, savant grec (287 à 212 avant notre ère), l'un des plus grands mathématiciens de tous les temps
2. Mikhail Yakovlevich Souslin, mathématicien russe

Figure 22 – Archimède de Syracuse. Il porta la méthode d'exhaustion à un tel degré de virtuosité que les tentatives de le dépasser restèrent vaines pendant mille ans au Moyen-Âge.

célèbre,[3] pensait avoir démontré le contraire ! Ce fut le départ d'une branche des mathématiques maintenant appelée la *théorie descriptive des ensembles.*

2. Soit $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ une application linéaire inversible. Montrer que, si A appartient à la tribu de Lebesgue de $\mathbb{R}^d$, il en est de même de $\varphi(A)$.

3. $\star$ Peut-on dire la même chose pour une application linéaire quelconque $\varphi : \mathbb{R}^d \to \mathbb{R}^n$? On précisera la réponse en fonction de n et d.

Exercice 15.5 (Exemple de coordonnées polaires)**.**
Soient R et Θ deux variables aléatoires réelles indépendantes de lois respectives exponentielle de paramètre $1/2$ et uniforme sur $[0, 2\pi]$. Montrer que le vecteur aléatoire

$$Z = (\sqrt{R}\cos\Theta, \sqrt{R}\sin\Theta)$$

de $\mathbb{R}^2$ suit une loi $N(0, I)$.

Exercice 15.6 (Mesure invariante par une transformation, *examen 2012*)**.**
Soient $(\Omega, \mathcal{A}, \mu)$ un espace de probabilité et $T : \Omega \circlearrowleft$ une transformation mesurable (cette notation signifie $T : \Omega \to \Omega$). On note μ_T la probabilité image de μ par T , et T^n la n-ième composée itérée de T, définie par récurrence par $T^0 = \mathrm{id}$, $T^{n+1} = T \circ T^n$ $(n \geqslant 0)$.

1. Montrer en général que μ est invariante ($\mu_T = \mu$) si et seulement si

$$\int (\varphi \circ T - \varphi)\, d\mu = 0$$

3. Lebesgue savait que les opération de projection et d'intersection ne commutent pas en général, mais pensait qu'elles commutent dans le cas d'une intersection décroissante. C'est faux, comme le montre l'exemple des $\mathbb{R} \times [n, +\infty[$.

pour toute fonction $\varphi \in L^1(\mu)$ telle que $\varphi \circ T \in L^1(\mu)$; on pourra commencer par montrer que

$$(\mu_T - \mu)(A) = \int (\mathbb{1}_A \circ T - \mathbb{1}_A)\, d\mu \quad (\forall A \in \mathcal{A}).$$

2. On suppose dans cette question que $\Omega =]0,1[$, $\mathcal{A}$ est la tribu borélienne, μ est la mesure de densité $((\ln 2)(1+x))^{-1}$ par rapport à la mesure de Lebesgue, et $T :]0,1[\to]0,1[$ est l'application de Gauss définie par

$$T(x) = \begin{cases} \frac{1}{x} - k & \text{si } \frac{1}{k+1} < x < \frac{1}{k},\ k \in \mathbb{N}_* \\ 1/2 & \text{si } x = 1/k,\ k \in \mathbb{N}_*. \end{cases}$$

Montrer que μ est invariante ; on pourra, dans le calcul de $\int_{]0,1[} \varphi \circ T\, d\mu$, découper $]0,1[$ en $\cup_{k \geqslant 1} \left[\frac{1}{k+1}, \frac{1}{k}\right[$, puis utiliser le fait que $[(y+k+1)(y+k)]^{-1} = (y+k)^{-1} - (y+k+1)^{-1}$.

Supposons dorénavant que la mesure μ est invariante. On dira qu'elle est *ergodique* si, pour toute partie mesurable $A \in \mathcal{A}$ invariante ($T^{-1}(A) = A$) [4], on a $\mu(A) = 0$ ou $\mu(A) = 1$.

3. Montrer que, si T est bijective,

$$T^{-1}(A) = A \Leftrightarrow A = T(A)$$

pour toute partie A de Ω, et donner un contre-exemple avec une transformation T non bijective.

4. Montrer que μ est ergodique si et seulement si, pour toute fonction $\varphi \in L^1(\mu)$ telle que $\varphi \circ T = \varphi$, φ est constante μ-p.p.

$\star$ Considérons dorénavant le cas où $\Omega = \mathbb{T}^2 := \mathbb{R}^2/\mathbb{Z}^2$, $\mathcal{A}$ est la tribu borélienne, $T(x_1, x_2) = (x_1 + \alpha_1, x_2 + \alpha_2)$ pour un certain $(\alpha_1, \alpha_2) \in \mathbb{R}^2$.

5. Montrer que, si $\frac{\alpha_1}{\alpha_2}$ est irrationnel, la mesure de Lebesgue est l'unique mesure invariante, et qu'elle est ergodique.

6. Montrer que, si $\frac{\alpha_1}{\alpha_2}$ est rationnel, il existe une infinité de mesure invariantes, une infinité de mesure ergodiques, et une infinité de mesure invariantes non ergodiques.

Problème 15.3 (Potentiel newtonien)**.**

Si A est un borélien de $\mathbb{R}^3$, le *potentiel newtonien* de A dans $\mathbb{R}^3$ est la fonction réelle négative $\phi_A : \mathbb{R}^3 \backslash A \to \mathbb{R}$ définie par

$$\phi_A(x) = -\frac{1}{4\pi} \int_A \frac{1}{\|x-y\|}\, dy.$$

En Physique, le potentiel détermine l'attraction newtonienne d'une particule en x créée par la distribution de masse (ici choisie uniforme) $\mathbb{1}_A\, dy$.

4. Une bonne définition de l'invariance ne devrait pas dépendre de parties négligeables. Ici, pour simplifier on impose $T^{-1}(A) \Delta A = \varnothing$ (où Δ désigne la différence symétrique de deux parties : $A \Delta B = (A \backslash B) \cup (B \backslash A)$). En général, on impose seulement que cette différence symétrique soit négligeable : $\mu(T^{-1}(A) \Delta A) = 0$.

On suppose dorénavant que A est la boule

$$A = \{x \in \mathbb{R}^3,\ \|x\| \leqslant R\},$$

et, étant donné l'invariance par rotation de ϕ_A, on suppose aussi que $x = (0, 0, a)$, $a > R$, de sorte que

$$\phi_A(x) = -\frac{1}{4\pi} \int_{\|y\| \leqslant R} \frac{1}{\sqrt{y_1^2 + y_2^2 + (y_3 - a)^2}}\, dy.$$

Montrer que

$$\phi_A(x) = -\frac{\mathrm{vol}(A)}{4\pi} \frac{1}{\|x\|} = \frac{R^3}{3} \frac{1}{\|x\|},$$

c'est-à-dire que le potentiel newtonien est le même que si toute la "masse" vol(A) était concentrée en l'origine. On pourra utiliser des coordonnées cylindriques (r, θ, y_3) telles que $y = (r \cos\theta, r \sin\theta, y_3)$.

Chapitre 16

Intégrales à paramètres

Quand on intègre par rapport à une variable une fonction de deux variables, on obtient une fonction de la variable restante (le paramètre). Les exemples sont nombreux et importants : transformation de Fourier, convolution, intégrales de Fresnel et de Laplace, etc. Dans ce chapitre, nous étudions deux questions naturelles : quelle est la régularité de la fonction obtenue ? et quel est le comportement asymptotique de cette fonction au voisinage de certaines valeurs importantes de son argument ?

Soient (U, d) un espace métrique (typiquement, un ouvert de $\mathbb{R}^d$), $(E, \mathcal{A}, \mu)$ un espace mesuré et $f : U \times E \to \mathbb{R}$ (ou $\mathbb{C}$) une application.

Théorème 16.1 (Continuité sous l'intégrale)**.** *Supposons :*

1. *$\mu(dx)$-p.p., l'application $f(\cdot, x)$ est continue en un point $u_0 \in U$ fixé ;*
2. *pour tout $u \in U$, l'application $f(u, \cdot)$ est mesurable ;*
3. *il existe une fonction $g \in \mathcal{L}^1(E)$ telle que, pour tout $u \in U$,*

$$|f(u, x)| \leqslant g(x) \quad \mu(dx)-\text{p.p.}$$

Alors la fonction $F : u \mapsto \int f(u, x)\, d\mu(x)$ est définie sur U et continue en u_0.

Démonstration. D'après les deuxième et troisième hypothèses, pour tout u la fonction $f(u, \cdot)$ est intégrable, donc $F(u)$ est bien définie. Ensuite, soit (u_n) une suite de U convergeant vers u_0. D'après la première hypothèse, quand n tend vers l'infini,

$$f(u_n, x) \to f(u_0, x) \quad \mu(dx)-\text{p.p.}$$

D'après la troisième hypothèse, cette convergence est dominée. Donc

$$\lim_n \int f(u_n, x)\, d\mu(x) = \int f(u_0, x)\, d\mu(x).$$

Donc F est continue en u_0. □

Exercice 16.1 (Continuité de certaines fonctions intégrales)**.**
Montrer que les fonctions suivantes de u sont continues sur $\mathbb{R}$:
1. La *transformée de Fourier* de φ,

$$\hat{\varphi}(u) := \int e^{iux}\varphi(x)\, d\lambda(x),$$

où $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$.
2. La *convolée*

$$h * \varphi\,(u) := \int h(u-x)\,\varphi(x)\, d\lambda(x),$$

où $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ et où $h : \mathbb{R} \to \mathbb{R}$ est une fonction continue bornée.
3. $F(u) := \int_{]-\infty,u]} \varphi(x)\, d\mu(x)$, où μ estune mesure diffuse sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$.

Restreignons-nous maintenant au cas où U est un intervalle ouvert de $\mathbb{R}$. On considère toujours une fonction $f : U \times E \to \mathbb{R}$ (ou $\mathbb{C}$).

Théorème 16.2 (Dérivation sous l'intégrale)**.** *Supposons :*

1. *$\mu(dx)$-p.p., $f(\cdot, x)$ est dérivable en un point $u_0 \in U$ donné ;*
2. *pour tout $u \in U$, $f(u, \cdot) \in \mathcal{L}^1(E)$;*
3. *il existe une fonction $g \in \mathcal{L}^1(E)$ telle que, pour tout $u \in U$,*

$$|f(u,x) - f(u_0,x)| \leqslant g(x)|u-u_0| \quad \mu(dx)\text{−p.p.}$$

Alors $F : U \to \mathbb{R}$, $u \mapsto \int f(x,u)\, d\mu(x)$ est dérivable en u_0 et de dérivée

$$F'(u_0) = \int \frac{\partial f}{\partial u}(u_0, x)\, d\mu(x)$$

(où, dans le membre de droite, la notation $\frac{\partial f}{\partial u}(u_0, x)$ est à prendre au sens de "0" pour les x éventuels pour lesquels $f(\cdot, x)$ n'est pas dérivable en u_0, les tels x formant de toute façon un ensemble négligeable).

Comme la dérivabilité est une propriété locale, il est souvent préférable, pour montrer la dérivabilité en point u_0, de choisir un intervalle U suffisamment petit.

Démonstration. Soient (u_n) une suite de $U \backslash \{u_0\}$ convergeant vers u_0 et

$$g_n(x) = \frac{f(u_n, x) - f(u_0, x)}{u_n - u_0}.$$

D'après la première hypothèse, $g_n(x) \to \frac{\partial f}{\partial u}(u_0, x)$, $\mu(dx)$-p.p. D'après la deuxième hypothèse, la fonction F est bien définie. D'après la troisième hypothèse et le théorème de convergence dominé,

$$\lim_n \frac{F(u_n) - F(u_0)}{u_n - u_0} = \lim_n \int g_n(x)\, d\mu(x) = \int \frac{\partial f}{\partial u}(u_0, x)\, d\mu(x).$$

□

Pour vérifier les hypothèses du théorème précédent, on est souvent amené à

— montrer que μ-p.p. la fonction $f(\cdot, x)$ est de classe C^1 sur un intervalle compact, donc que sa dérivée est bornée,

— puis appliquer le théorème des accroissements finis.

Exercice 16.2 (Dérivée de certaines fonctions intégrales)**.**
Montrer que les fonctions suivantes de u sont dérivables sur $\mathbb{R}$, et calculer leur dérivée :

1. La transformée de Fourier de φ, où $\varphi, x\varphi(x) \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$.

2. La convolée $h * \varphi$, où $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ et où $h : \mathbb{R} \to \mathbb{R}$ est une fonction de classe C^1, bornée et de dérivée bornée.

3. $F(u) := \int_{]-\infty,u]} (u-x)^+ \, \varphi(x) \, d\mu(x)$, où μ est une mesure diffuse sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et $\varphi(x), x\varphi(x) \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$.

Exercice 16.3 (Intégrale holomorphe)**.**
Cet exercice s'adresse au lecteur familier avec la $\mathbb{C}$-dérivabilité.
Soient D un ouvert de $\mathbb{C}$ et $f : D \times E \to \mathbb{C}$ une fonction complexe. Supposons :

1. pour tout $u \in D$, $f(u, \cdot) \in \mathcal{L}^1(E)$;
2. $\mu(dx)$-p.p., $f(\cdot, x)$ est $\mathbb{C}$-dérivable en un point $u_0 \in U$ donné ;
3. il existe une fonction $g \in \mathcal{L}^1(E)$ telle que, pour tout $u \in D$,

$$|f(u,x) - f(u_0,x)| \leqslant g(x)|u - u_0| \quad \mu(dx)\text{–p.p.}$$

Alors $F : \mathbb{R} \to \mathbb{R}$, $u \mapsto \int f(x,u) \, d\mu(x)$ est $\mathbb{C}$-dérivable en u_0 et de dérivée

$$F'(u_0) = \int \frac{\partial f}{\partial u}(u_0, x) \, d\mu(x)$$

Exercice 16.4 (Méthode de Laplace, *examen 2011*)**.**
Soit $f :]0,1[\to \mathbb{R}$ une fonction borélienne intégrable par rapport à la mesure de Lebesgue. On suppose que f possède une limite $f(1^-) \in \mathbb{R}_*$ à gauche en 1. On veut démontrer l'équivalent suivant quand n tend vers $+\infty$:

$$I_n = \int_0^1 x^n \, f(x) \, dx \sim \frac{f(1^-)}{n}.$$

1. Montrer que I_n tend vers 0 quand n tend vers $+\infty$.

2. Montrer l'équivalence voulue dans le cas où f est de classe C^1 sur $[0,1]$, en faisant une intégration par partie.

3. Pourquoi ne peut-on pas généralement appliquer le théorème de convergence dominée directement à nI_n ?

4. Démontrer l'équivalent de I_n en coupant l'intervalle $[0,1]$ en deux : un voisinage de 1 sur lequel f est bornée et où l'on pourra faire le changement de variables $y = x^n$, et son complémentaire dont la contribution est négligeable quand n tend vers l'infini.

Exercice 16.5 (Formule de Stirling, *examen 2012*).
Soit
$$f(n) = \int_0^\infty t^n\, e^{-t}\, dt, \quad n \in \mathbb{N}.$$

1. Montrer par récurrence que $f(n) = n!$ si $n \in \mathbb{N}$.
2. Montrer que
$$f(n) = \sqrt{n}\, n^n\, e^{-n} \int_{-\sqrt{n}}^{+\infty} \left(1 + \frac{y}{\sqrt{n}}\right)^n e^{-y\sqrt{n}}\, dy.$$
Indication : on pourra partir du membre de droite.
3. En déduire la formule de Stirling :
$$f(n) \sim_{n\to+\infty} \sqrt{2\pi n}\, n^n\, e^{-n};$$
on rappelle que $\int_{\mathbb{R}} e^{-y^2/2}\, dy = \sqrt{2\pi}$.

Problème 16.1 (Formule de Watson).
Soit F une fonction réelle définie sur un voisinage de 0 dans $]0, +\infty[$. Un *développement asymptotique* en 0 est une fonction de la forme $\sum_{n\leqslant k\leqslant N} a_k h_k$, où $a_n, ..., a_N$ sont des nombres réels et $h_n, ..., h_N$ sont des fonctions définies sur un voisinage de 0 dans $]0, +\infty[$, tels que, pour tout $k \in \{n, ..., N\}$,
$$F(x) = \sum_{n\leqslant j\leqslant k} a_j h_j(x) + o(h_k(x))$$
quand x tend vers 0. Les développements asymptotiques furent introduits par Poincaré à la fin du XIXe siècle pour étudier le mouvement des planètes sur de longs intervalles de temps, le rôle du temps étant alors joué par $1/x$. On se propose ici d'étudier quelques développements asymptotiques particuliers.
On utilisera à plusieurs reprises la *fonction* $\Gamma :]0, +\infty[\to \mathbb{R}$ *d'Euler* définie par
$$\Gamma(y) = y^{-1} \int_0^\infty e^{-t}\, t^y\, dt. \tag{16.1}$$

1. Montrer que, pour tout $y > 0$, $y\Gamma(y) = \Gamma(y+1)$ et en déduire que, pour tout $n \in \mathbb{N}$, $\Gamma(n+1) = n!$.
2. On considère dans cette question une fonction F de la forme
$$F(x) = \int_0^\infty e^{-t/x}\, f(t)\, dt$$
où $f :]0, +\infty[\to \mathbb{R}$ est une fonction continue sur $]0, +\infty[$, telle que :
— il existe un entier $K \geqslant 0$ et un réel $C > 0$ tels que $f(t) \leqslant Ct^K$ sur $]1, +\infty[$,
— et il existe un entier $N \geqslant 0$, deux réels $\lambda > 0$ et $\mu > 0$ et des réels $a_0, ..., a_N$ tels que
$$f(t) = \sum_{0\leqslant k\leqslant N} a_k\, t^{(k+\lambda-\mu)/\mu} + o\left(t^{(k+\lambda-\mu)/\mu}\right)$$
quand t tend vers 0.

Le principal objectif est de montrer la formule suivante (Watson, 1952) :

$$F(x) = \sum_{0\leqslant k\leqslant N} a_k\,\Gamma\left(\frac{k+\lambda}{\mu}\right)\, x^{(k+\lambda)/\mu} + o\left(x^{(N+\lambda)/\mu}\right) \tag{16.2}$$

quand x tend vers 0.

3. Pourquoi F est-elle bien définie ?

On note

$$\rho_N(t) = f(t) - \sum_{0\leqslant k\leqslant N} a_k\, t^{(k+\lambda-\mu)/\mu}$$

le reste du développement asymtotique de f.

4. Montrer qu'il existe $T > 0$ tel que

$$\int_0^T e^{-t/x}\,\rho_N(t)\,dt = o\left(x^{(N+\lambda)/\mu}\right).$$

5. Soit $X > 0$. Montrer que la fonction

$$\Phi_N :]T, +\infty[\to \mathbb{R}, \quad t \mapsto \int_T^t e^{-s/X}\,\rho_N(s)\,ds$$

est continue bornée, et que, si X est assez petit et $0 < x < X$,

$$\int_T^\infty e^{-t/x}\rho_N(t)\,dt = \left(\frac{1}{x} - \frac{1}{X}\right)\int_T^\infty e^{-(1/x-1/X)t}\Phi_N(t)\,dt.$$

En déduire que, pour tout $n \geqslant 0$,

$$\int_T^\infty e^{-t/x}\,\rho_N(t)\,dt = o\left(x^n\right).$$

6. Conclure.

Nous allons déduire de la formule de Watson un développement asymptotique de la fonction Γ en $+\infty$; on se ramène à un développement asymptotique en 0 en notant $x = 1/y$.

7. Montrer par un changement de variable que

$$\Gamma(y) = e^{-y}\,y^y \int_{-1}^\infty e^{-y\phi(s)}\,ds, \quad \phi(s) = s - \ln(1+s),$$

et tracer le graphe de ϕ sur $]-1, +\infty[$.

8. Montrer que, sur $]-1, 1[$,

$$\phi(s) = s^2 \sum_{k\geqslant 0} \frac{(-1)^k}{k+2}\, s^k.$$

Soient ϕ_- et ϕ_+ les restrictions de ϕ aux intervalles $]-1,0[$ et $]0,+\infty[$.

9. Montrer que la fonction $\sqrt{\phi_+}$ se prolonge sur $]-1,+\infty[$ en une fonction V de classe C^∞. Calculer le développement limité de V en 0 à l'ordre 3.

10. Montrer que, si $\epsilon > 0$ est assez petit, la fonction V est une bijection de $]-\epsilon,+\infty[$ dans $]V(-\epsilon),+\infty[$.

11. Montrer qu'il existe des réels b_1, b_2 et b_3 tels que la bijection réciproque de V satisfasse, quand v tend vers 0,

$$V^{-1}(v) = b_1 v + b_2 v^2 + b_3 v^3 + o\left(v^3\right)$$

et en déduire que, quand q tend vers 0,

$$\phi_+^{-1}(q) = b_1\sqrt{q} + b_2 q + b_3 q^{3/2} + o\left(q^{3/2}\right).$$

12. En déduire le développement asympotique

$$(\phi_+^{-1})'(q) = \frac{1}{\sqrt{2q}} + \frac{2}{3} + \frac{\sqrt{q}}{6\sqrt{2}} + o(\sqrt{q}),$$

en admettant pouvoir dériver terme à terme le développement de ϕ_+^{-1}.

On admettra que, de même, quand q tend vers 0,

$$(\phi_-^{-1})'(q) = -\frac{1}{\sqrt{2q}} + \frac{2}{3} - \frac{\sqrt{q}}{6\sqrt{2}} + o(\sqrt{q}).$$

13. Montrer que

$$\Gamma(y) = e^{-y}\, y^y \int_0^\infty e^{-yq}\left((\phi_+^{-1})'(q) - (\phi_-^{-1})'(q)\right)\, dq.$$

14. En déduire le raffinement suivant de la formule de Stirling (exercice 16.5) :

$$\Gamma(y) = e^{-y}\, y^y\, \left(\frac{2\pi}{y}\right)^{1/2}\left(1 + \frac{1}{12\, y} + o\left(\frac{1}{y}\right)\right)$$

quand y tend vers l'infini ; on pourra utiliser le fait que $\Gamma(1/2) = \sqrt{\pi}$.

Problème 16.2 (Sommation au plus petit terme)**.**
Soit la fonction $F :]0,+\infty[\to \mathbb{R}$ définie par

$$F(x) = \int_1^{+\infty} e^{-t/x} t^{-1}\, dt.$$

On va voir qu'une série divergente peut être utile pour calculer une valeur approchée de F en un point particulier.

1. Montrer que F est bien définie et de classe C^∞ sur $]0,+\infty[$.

2. Montrer que, pour tout $N \geqslant 0$ et tout $x > 0$

$$F(x) = S_N(x) + R_N(x)$$

avec

$$\begin{cases} S_N(x) = \sum_{1\leqslant k\leqslant N}(-1)^{k-1}\,(k-1)!\,x^k\,e^{-1/x} \\ R_N(x) = (-1)^N\,N!\,x^N\,\int_1^{+\infty} e^{-t/x}t^{-(N+1)}\,dt. \end{cases}$$

3. Pour tout $x > 0$ fixé, montrer que la série $\sum(-1)^{k+1}\,(k-1)!\,x^k$ diverge et que la suite $(R_N(x))_{N\geqslant 1}$ n'est pas bornée.

4. Montrer que, si $N \in \mathbb{N}$ et $x > 0$,

$$|R_N(x)| \leqslant |r_N(x)|;$$

on pourra commencer par faire le changement de variable $t = xs$.

5. Montrer le reste est de l'ordre du premier terme négligé : pour tout $N \geqslant 0$,

$$R_N(x) \sim_{x\to 0} r_N(x) = (-1)^N\,N!\,x^{N+1}\,e^{-1/x}.$$

6. Montrer que, pour tout $x > 0$, la suite $(|r_N(x)|)_{N\geqslant 0}$ est décroissante jusqu'à un certain rang, puis croissante. (On ne demande pas de montrer cela pour la suite $(|R_N(x)|)_{N\geqslant 0}$.)

Quand on utilise $S_N(x)$ comme valeur approchée de $F(x)$, on dit que l'erreur relative est

$$E_N(x) = \left|\frac{R_N(x)}{F(x)}\right|.$$

7. Montrer que, si N est pair : $N = 2M$ avec $M \geqslant 0$, et si $0 < x \leqslant 1/N$,

$$E_N(x) \leqslant \frac{N!\,x^{N+1}}{\displaystyle\sum_{0\leqslant \ell\leqslant M-1}(1-(2\ell+1)x)\,(2\ell)!\,x^{2\ell+1}}.$$

8. Vérifier que

$$E_4\left(\frac{1}{10}\right) \leqslant 3\,.\,10^{-3}.$$

Problème ★ 16.3 (Inégalité de Schauder [Tao10a])**.**

Soient $\alpha \in]0,1[$ et f une fonction α-höldérienne sur $\mathbb{R}^3$, nulle en dehors de la boule unité. Soit u la fonction définie sur $\mathbb{R}^3$ par

$$u(x) = \frac{1}{4\pi}\int_{\mathbb{R}^3}\frac{f(y)}{\|x-y\|}\,dy.$$

1. Montrer que u est continue sur $\mathbb{R}^3$.

2. Montrer que u est de classe C^1 et que

$$\frac{\partial u}{\partial x_j}(x) = -\frac{1}{4\pi}\int_{\mathbb{R}^3}(x_j-y_j)\frac{f(y)}{\|x-y\|^3}\,dy \quad (j = 1,2,3).$$

3. Montrer que u est de classe C^2 et que

$$\frac{\partial^2 u}{\partial x_j\partial x_k}(x) = \frac{1}{4\pi}\lim_{\epsilon\to 0}\int_{\|x-y\|\geqslant\epsilon}\left[\frac{3(x_j-y_j)(x_k-y_k)}{\|x-y\|^5} - \frac{\delta_{jk}}{\|x-y\|^3}\right]f(y)\,dy;$$

on pourra établir cette formule d'abord dans le cas particulier où $f(x) = 0$, puis dans celui où f est constante au voisinage de x.

4. Montrer que $\Delta u = f$ sur $\mathbb{R}^3$.

5. Montrer que $u \in C^{2,\alpha}(\mathbb{R}^3)$, et établir l'*inégalité de Schauder* suivante :

$$\|u\|_{C^{2,\alpha}(\mathbb{R}^3)} \leqslant C_\alpha \|f\|_{C^{0,\alpha}(\mathbb{R}^3)},$$

où C_α ne dépend que de α.

Cette inégalité remarquable dit que, si Δu est α-höldérienne, il en est de même de toutes les autres dérivées secondes de u (y compris celles qui n'apparaissent pas dans le laplacien !). Ce phénomène est connu sous le nom de *régularité elliptique*, et repose de façon cruciale sur le fait que le laplacien est un opérateur elliptique. On peut montrer que l'inégalité est fausse en général si $\alpha = 0$, ce qui explique la nécessité d'introduire les espaces de Hölder.

Chapitre 17

Lien avec le calcul différentiel

L'objet de ce chapitre est de retrouver les deux "théorèmes fondamentaux de l'analyse réelle", qui donnent des conditions suffisantes pour que

— $\left(\int_a^x f\,dt\right)' = f(x)$, i.e. la fonction $F(x) = \int_a^x f\,dt$ est une solution de l'équation différentielle $F'(x) = f(x)$ de donnée f,

— $\int_a^x F'(t)\,dt = F(x) - F(a)$, i.e. l'équation différentielle $F'(x) = f(x)$ possède au plus une solution une fois que $F(a)$ est donnée, à savoir $F(a) + \int_a^x f(t)\,dt$, ce qui montre en particulier que la variation $F(x) - F(a)$ ne dépend que de F' (par exemple, si $F' \equiv 0$, $F(x) \equiv F(a)$) ;

nous choisissons des hypothèses de régularité non optimales pour que ces théorèmes soient des conséquences simples du théorème de convergence dominée.

Théorème 17.1 (Premier théorème fondamental de l'analyse, de dérivation de l'intégrale d'une fonction continue)**.** *Si $f : [a,b] \to \mathbb{R}$ est continue, la fonction*

$$F : [a,b] \to \mathbb{R}, \quad x \mapsto \int_a^x f(t)\,dt$$

est dérivable et $F' = f$.

Remarque 17.2. Le théorème de dérivation sous l'intégrale tel que nous l'avons énoncé ne s'applique pas directement à la fonction

$$x \mapsto \int_0^x f(t)\,dt = \int_a^b \mathbb{1}_{[a,x]}(t)\,f(t)\,dt$$

parce que, à cause de la discontinuité en $x = t$, l'hypothèse de domination du taux d'accroissement n'est pas vérifiée. D'ailleurs, si le théorème s'appliquait, la conclusion serait que $F' = 0$, puisque $x \mapsto \mathbb{1}_{[a,x]}(t)\,f(t)$ est presque partout dérivable de dérivée nulle. Or, comme on le verra, ce n'est pas le cas ! Le plus simple est de repartir du théorème de convergence dominée.

Démonstration. Soit $x \in [a,b]$. Pour tout $h \neq 0$ tel que $x + h \in [a,b]$, on pose

$$\tau_x(h) = \frac{F(x+h) - F(x)}{h}.$$

Un changement de variable ($t = \varphi(s) = h + s$) dans le premier terme et la règle de Chasles (exercice 7.7) montrent que

$$\tau_x(h) = \frac{1}{h}\int_x^{x+h} f(t)\,dt.$$

Le changement de variable $t = \varphi(s) = x + hs$, $0 \leqslant s \leqslant 1$, de jacobien h, montre que [1]

$$\tau_x(h) = \int_0^1 f(x + hs)\,ds.$$

Comme f est continue, $f(x+hs) \to f(x)$ quand h tend vers 0, et f est bornée, donc dominée par une constante (forcément intégrable sur l'intervalle compact $[a, b]$). Par convergence dominée, on a donc

$$\lim_{h\to 0} \tau_x(h) = \int_0^1 f(x)\,ds = f(x).$$

Donc F est dérivable en x (plus précisément, F est dérivable sur $]a, b[$, à droite en a et à gauche en b) et $F'(x) = f(x)$. □

Mentionnons sans démonstration le célèbre théorème suivant.

Théorème ★ 17.3 (de dérivation de Lebesgue)**.** *Si $f \in \mathcal{L}^1([a, b])$, la fonction*

$$F : [a, b] \to \mathbb{R}, \quad x \mapsto \int_a^x f(t)\,dt$$

est continue, dérivable p.p. et $F' = f$ p.p.

L'hypothèse est plus faible puisque la fonction f n'est plus supposée continue (et donc intégrable sur $[a, b]$), mais seulement intégrable. La conclusion aussi est plus faible puisque F est dérivable non plus partout mais presque partout. Il reste que la démonstration est considérablement plus délicate que celle du “premier théorème fondamental” tel que nous l'avons énoncé. Une superbe référence sur ces questions est le livre de théorie de la mesure de T. Tao [Tao11]. [2]

Théorème 17.4 (Second théorème fondamental de l'analyse réelle, d'intégration d'une dérivée bornée)**.** *Si $F : [a, b] \to \mathbb{R}$ est dérivable (p.p. n'est pas suffisant ici !) de dérivée bornée, F' est borélienne et*

$$\int_a^b F'(t)\,dt = F(b) - F(a).$$

(Un exemple de fonction F dérivable sur $[0, 1]$ mais dont la dérivée n'est pas bornée est $F : x \mapsto x^2 \sin \frac{1}{x^2}$. Dans cet exemple, la conclusion du théorème reste vraie, mais pour d'autres exemples plus compliqués ce n'est pas le cas...)

1. Si $h < 0$, $|J\varphi(s)| = |h| = -h$, donc $\int_x^{x+h} f = -\int_{[x+h,x]} f(t)\,dt = h\int_{[0,1]} f(x + hs)\,ds$ (cf. la remarque 15.4).

2. Terry TAO, mathématicien australien vivant aux USA (né en 1975), travaillant entre autres en analyse harmonique, équations aux dérivées partielles, analyse combinatoire, matrices aléatoires, théorie analytique des nombres.

Démonstration. Soient α et β tels que $a < \alpha \leqslant \beta < b$. On va montrer que

$$\int_\alpha^\beta F'(t)\,dt = F(\beta) - F(\alpha).$$

La formule voulue en découle par passage à la limite quand α tend vers a^+ et β vers b^-. En effet, dans cette limite, le membre de gauche de l'égalité tend vers $\int_a^b F'(t)\,dt$ par convergence dominée ($\mathbb{1}_{[\alpha,\beta]}F'(t) \to \mathbb{1}_{]a,b[}F'(t) = F'(t)$ p.p., et $|\mathbb{1}_{[\alpha,\beta]}F'(t)|$ est bornée). Quant au membre de droite, il tend vers $F(b) - F(a)$ parce que F est continue.

Soit $x \in [a,b]$. Pour $h \neq 0$ tel que $x + h \in [a,b]$, soit

$$\tau_x(h) = \frac{F(x+h) - F(x)}{h}$$

le taux d'accroissement de F entre x et $x + h$.

Nous allons calculer de deux façons différentes la limite de $\int_\alpha^\beta \tau_x(h)\,dx$ quand h tend vers 0.

1. Comme F est dérivable sur $[\alpha,\beta]$, quand h tend vers 0, $\tau_x(h)$ tend vers $F'(x)$ sur $[\alpha,\beta]$. De plus, d'après l'inégalité des accroissements finis, si l'on note M une borne de $|F'|$ sur $[\alpha,\beta]$,

$$|\tau_x(h)| \leqslant M \in \mathcal{L}^1([\alpha,\beta]).$$

   Par convergence dominée, quand h tend vers 0,

$$\int_\alpha^\beta \tau_x(h)\,dx \to \int_\alpha^\beta F'(x)\,dx.$$

2. Si $x + h \in [a,b]$, une translation dans le premier terme ($t = h + s$) et la règle de Chasles (exercice 7.7) montrent que

$$\int_\alpha^\beta \tau_x(h)\,dx = \frac{1}{h}\left(\int_\beta^{\beta+h} f - \int_\alpha^{\alpha+h} f\right).$$

   Maintenant, les changements de variables $t = \beta + hs$ et $t = \alpha + hs$ pour chacun des deux termes montrent que

$$\int_\alpha^\beta \tau_x(h)\,dx = \int_0^1 \left(F(\beta+hs) - F(\alpha+hs)\right)\,ds.$$

   Par continuité de F, quand h tend vers 0 on a

$$F(\alpha+hs) \to F(\alpha) \quad \text{et} \quad F(\beta+hs) \to F(\beta).$$

   De plus, F est bornée. Donc, par convergence dominée,

$$\int_\alpha^\beta \tau_x(h)\,dx \to F(\beta) - F(\alpha).$$

Par unicité de la limite, on a bien

$$\int_\alpha^\beta F'(t)\,dt = F(\beta) - F(\alpha),$$

puis en passant à la limite comme indiqué en début de démonstration, la formule annoncée. □

Remarques ★ 17.5. (1) L'hypothèse "dérivée bornée" n'est pas nécessaire, mais simplifie grandement la démonstration. En toute généralité, les fonctions vérifiant l'égalité

$$\int_a^x F'(t)\,dt = F(x) - F(a) \quad (\forall x \in [a,b])$$

sont les fonctions *absolument continues*. Elles sont caractérisées par le fait que

$$\sum_i |f(t_i) - f(s_i)|$$

tend vers 0 quand $([s_i, t_i])_i$ est une famille finie quelconque d'intervalles disjoints dont la longueur totale $\sum_i (t_i - s_i)$ tend vers 0. On peut montrer en particulier que, pour une telle fonction, l'image d'un ensemble négligeable est négligeable, ce qui n'est pas le cas de "l'escalier du diable" décrit dans l'exemple 17.7. Voir [Tao11].

(2) Le théorème 17.4 donne lieu à une généralisation appelée *formule de Stokes*, la plus importante formule de la physique mathématique. Cette formule de Stokes fait le lien entre les opérations de dérivation d de l'objet à intégrer et de "bord géométrique" ∂ du domaine d'intégration, et affirme que les deux sont adjointes l'une de l'autre relativement à la forme bilinéaire d'intégration : $\int_{\partial D} f = \int_D df$. Pour lui donner son sens, il faut introduire les *variétés différentielles à bord* (D) et les *formes différentielles* (f), qui sont une sorte de généralisation géométrique de la notion de mesure. La *cohomologie de de Rham* est alors l'obstruction à la formule du second théorème fondamental sur une variété différentielle.

Exercice 17.1 (Critère de Riemann)**.**
Soit $\alpha \in \mathbb{R}$. Montrer que la fonction

$$f :]1, +\infty[\to \mathbb{R}, \quad x \mapsto x^\alpha$$

est intégrable si et seulement si $\alpha < -1$.

Exercice 17.2 (Dérivée par rapport aux bornes)**.**
Reprenons les hypothèse du second théorème fondamental, et supposons de plus que α et β sont deux fonctions dérivables $]c, d[\to [a, b]$. Quelle est la dérivée de

$$\phi : u \mapsto \int_{\alpha(u)}^{\beta(u)} F'(t)\,dt \ ?$$

Si l'on affaiblit l'hypothèse de dérivabilité, l'une des deux inégalités du second théorème fondamental subsiste.

Proposition ★ 17.6 (Intégration d'une dérivée positive). *Si $F : [a, b] \to \mathbb{R}$ est croissante, continue en a et b, et presque partout dérivable,*

$$\int_a^b F'(x)\,dx \leqslant F(b) - F(a),$$

où $F'(x)$ désigne conventionnellement la dérivée de F quand elle existe et la fonction nulle ailleurs.

D'après un théorème de dérivabilité de Lebesgue, la propriété que F soit presque partout dérivable est en fait une conséquence de ce qu'elle est monotone.

Démonstration. Comme toute fonction croissante continue, F est borélienne : en effet, l'image réciproque d'un intervalle est un intervalle, et les intervalles engendrent la tribu borélienne. De plus, comme f est croissante, elle est bornée par $\max\{|F(a)|, |F(b)|\}$, donc intégrable.
Définissons sur $[a, b]$ la suite des fonctions intégrables

$$g_n(x) = \begin{cases} \text{le taux d'accroissement de } F \text{ entre } x \text{ et } x + 1/n \text{ si } x + 1/n \leqslant b \\ 0 \text{ sinon,} \end{cases}$$

soit

$$g_n(x) = \begin{cases} n(F(x + 1/n) - F(x)) & \text{si } a \leqslant x \leqslant b - 1/n \\ 0 & \text{si } b - 1/n < x \leqslant b. \end{cases}$$

Ces fonctions sont positives et, pour presque tout x la suite $(g_n(x))$ converge vers $F'(x)$. La fonction F' est donc positive là où elle existe et l'on pose conventionnellement $F'(x) = 0$ ailleurs. Comme $\liminf g_n = F'$ presque partout, F est Lebesgue-mesurable et le lemme de Fatou implique

$$\int_a^b F'(x)\,dx \leqslant \liminf \int_a^b g_n(x)\,dx.$$

D'autre part, posons $\phi(x) = \int_a^x F(t)\,dt$; alors

$$\int_a^b g_n(x)\,dx = n(\phi(b) - \phi(b - 1/n)) - n(\phi(a + 1/n) - \phi(a)).$$

Comme F est continue en a et b, le membre de droite tend vers $F(b) - F(a)$. $\square$

Exemple ★ 17.7 (Escalier du diable de Lebesgue). Il existe une fonction $F : [0, 1] \to [0, 1]$ continue, croissante, telle que $F(0) = 0$ et $F(1) = 1$ et telle que presque partout F soit dérivable de dérivée nulle, de sorte que

$$\int_0^1 F'(t)\,dt = 0 < F(1) - F(0) = 1.$$

Nous allons construire une telle fonction, en plusieurs étapes.

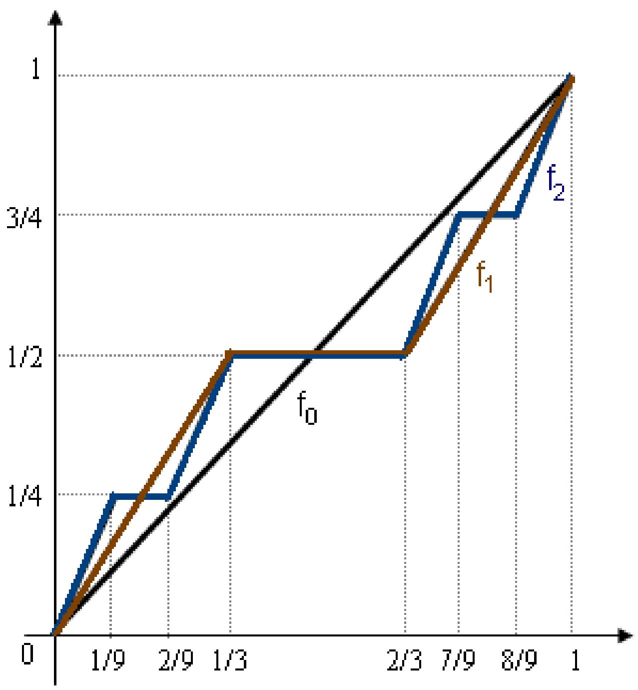


FIGURE 23 – Fonctions F_0, F_1, F_2 de l'exemple 17.7

1. Soit (A_n) la suite de parties de $[0,1]$ définie par
$$A_0 = [0,1], \quad A_{n+1} = \frac{A_n}{3} \cup \frac{2+A_n}{3}.$$
A_n est un fermé, réunion disjointe de 2^n intervalles fermés de longueur 3^{-n}. Soit maintenant (F_n) la suite de fonctions sur $[0,1]$ définie par
$$F_n(x) = (3/2)^n \int_0^x \mathbb{1}_{A_n}(t)\, dt.$$
On a $F_n(0) = 0$ et, comme $\lambda(A_n) = (2/3)^n$, $F_n(1) = 1$. D'autre part, F_n est continue et croissante sur $[0,1]$.
2. Montrons que (F_n) converge uniformément. Si I est l'un des 2^n intervalles compacts dont A_n est la réunion, $\lambda(I) = 3^{-n}$ et $\lambda(I \cap A_{n+1}) = 2/3\lambda(I)$. Donc
$$\int_I \mathbb{1}_{A_n}(t)\, dt = \frac{3}{2} \int_I \mathbb{1}_{A_{n+1}}(t)\, dt.$$
En outre, F_n est constante sur chacun des $2^n - 1$ intervalles ouverts composant A_n^c ; et il en est de même de la fonction F_{n+1} puisque $A_n^c \subset A_{n+1}^c$. Donc l'égalité précédente entraîne que, quel que soit $x \in A_n^c$,
$$F_n(x) = \sum_{I \subset A_n \cap [0,x]} (3/2)^n \int_I \mathbb{1}_{A_n}(t)\, dt = \sum_{I \subset A_n \cap [0,x]} (3/2)^{n+1} \int_I \mathbb{1}_{A_{n+1}}(t)\, dt = F_{n+1}(x),$$
où I est toujours l'un des 2^n intervalles composants A_n. En particulier, pour

un tel I, $F_{n+1}(\min I) = F_n(\min I)$. Donc, pour tout $x \in I$,

$$\begin{aligned} |F_{n+1}(x) - F_n(x)| &\leqslant |F_{n+1}(x) - F_{n+1}(\min I)| + |F_n(x) - F_n(\min I)| \\ &\leqslant (3/2)^{n+1} \int_I \mathbb{1}_{A_{n+1}}(t)\, dt + (3/2)^n \int_I \mathbb{1}_{A_n}(t)\, dt = 2^{-n+1}. \end{aligned}$$

La suite (F_n) est donc uniformément de Cauchy, donc converge uniformément vers une fonction F, nécessairement continue et croissante.

3. De plus, pour tout $m \geqslant n$, $F_m = F_n$ sur chacun des intervalles ouverts dont A_n^c est la réunion. Donc $F = F_n$ sur A_n^c. Or, F_n est constante sur A_n^c. En particulier F y est dérivable de dérivée nulle. Finalement, F est dérivable de dérivée nulle sur $\cup_{n \geqslant 0} A_n^c$, qui est de mesure pleine.

Remarque : Dans cette construction, l'intersection décroissante des A_n est un compact K d'intérieur vide et de mesure nulle, mais non dénombrable : c'est l'*ensemble triadique de Cantor* déjà rencontré,

$$C = \left\{ \sum_1^\infty \frac{x_n}{3^n},\ x_n \in \{0, 2\} \right\}.$$

Pour aller plus loin dans cette direction, il convient d'introduire les notions de *dérivée faible* et de *distribution*, ce que nous ne ferons pas ici. En revanche, voici une dernière formule très utile à la fois en calcul différentiel, en théorie des distributions et en calcul stochastique.

Intégration par parties

Soient $f, g \in \mathcal{L}^1([a, b])$ $(a < b)$. Posons, pour tout $x \in [a, b]$,

$$\begin{cases} F(x) = \int_a^x f\, dt \\ G(x) = \int_a^x g\, dt. \end{cases}$$

Théorème 17.8 (Formule d'intégration par parties)**.**

$$\int_a^b fG\, dt + \int_a^b Fg\, dt = \Big[FG \Big]_a^b := F(b)G(b) - F(a)G(a).$$

Démonstration. La formule équivaut à la formule plus symétrique suivante :

$$\int_a^b f(t)(G(t) - G(a))\, dt = \int_a^b (F(b) - F(t))g(t)\, dt.$$

Or,

$$\begin{aligned}
\int_a^b f(t)(G(t)-G(a))\,dt &= \int_a^b f(t)\left(\int_a^t g(s)\,ds\right)dt && \text{(définition de } G\text{)}\\
&= \int_a^b \left(\int_a^b \mathbb{1}_{s\leqslant t} f(t)g(s)\,ds\right)dt && \text{(linéarité)}\\
&= \int_a^b \left(\int_a^b \mathbb{1}_{s\leqslant t} f(t)g(s)\,dt\right)ds && \text{(théorème de Fubini)}\\
&= \int_a^b \left(\int_s^b f(t)\,dt\right)g(s)\,ds && \text{(linéarité)}\\
&= \int_a^b g(s)(F(b)-F(s))\,ds,
\end{aligned}$$

l'application du théorème de Fubini étant justifiée par le fait que la fonction $\varphi(s,t) := \mathbb{1}_{s\leqslant t} f(t)g(s)$ est dans $\mathcal{L}^1([a,b]^2)$, puisque, d'après le théorème de Tonelli,

$$\int_{[a,b]^2} |\varphi(s,t)|\,ds\,dt \leqslant \int_{[a,b]^2} |f(t)|\,|g(s)|\,ds\,dt = \left(\int_a^b |f|\,dt\right)\left(\int_a^b |g|\,ds\right) < \infty.$$

□

Exercice 17.3 (Intégration par parties bis)**.**
1. Montrer que la formule du théorème 17.8 reste vraie si on ajoute une constante à F ou à G.
2. Si $f : [a,b] \to \mathbb{R}$ est une fonction intégrable et si $G : [a,b] \to \mathbb{R}$ est de classe C^1, montrer que

$$\int_a^b fG = [FG]_a^b - \int_a^b f(t)G'(t)\,dt, \quad F(x) = \int_a^x f(t)\,dt;$$

ce sont les hypothèses et la formules souvent les plus pertinentes pour « intégrer par parties ».
3. En supposant que f et g sont continues, démontrer plus simplement la formule d'intégration par parties du théorème 17.8 ; on pourra utiliser les deux théorèmes fondamentaux.

Exercice 17.4 (Partie finie de Hadamard, *examen 2011*)**.**
Soit $\phi : [0,1] \to \mathbb{R}$ une fonction de classe C^1.
1. Montrer qu'il existe une fonction θ continue telle que

$$\phi(x) = \phi(0) + x\,\theta(x).$$

2. Dans quels cas la fonction $x \mapsto \phi(x)/x$ est-elle intégrable sur $[0,1]$? En déduire que, dans tous les cas, quand ϵ tend vers 0, la limite, notée P.f. $\int_0^1 \phi(t)/t\,dt$, de

$$\int_\epsilon^1 \frac{\phi(t)}{t}\,dt + \phi(0)\,\ln\epsilon,$$

existe (*partie finie de Hadamard* de $\int_0^1 \phi(t)/t\,dt$).

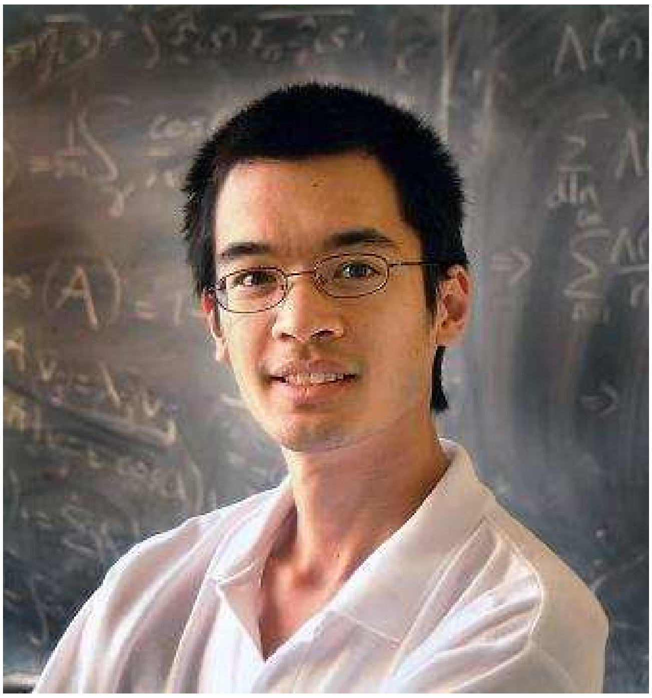


FIGURE 24 – Terry TAO (né en 1975)

Chapitre 18

Les espaces $L^1(\mu)$ et $L^\infty(\mu)$

Dans ce chapitre, $(E, \mathcal{A}, \mu)$ est un espace mesuré donné. Toutes les parties de E considérées sont $\mathcal{A}$-mesurables. Pour être concret, on va s'intéresser aux fonctions boréliennes $f : E \to \mathbb{R}$, mais la théorie des fonctions à valeurs complexes ou dans $\mathbb{R}^n$ est complètement parallèle.

On a défini la notion d'*intégrabilité* pour une telle fonction f, que l'on note $f \in \mathcal{L}^1$ et qui signifie que

$$\int |f|\, d\mu < \infty.$$

Remarque 18.1. On pourrait aussi considérer des fonctions à valeurs dans $\bar{\mathbb{R}}$, mais on a vu que, si une telle fonction est intégrable, elle est finie en dehors d'un ensemble négligeable et que donc on peut modifier cette fonction en lui donnant la valeur nulle là où elle était infinie, sans modifier son intégrale. Donc, étendre l'ensemble d'arrivée à $\bar{\mathbb{R}}$ ne conduirait à aucune généralisation significative.

La fonction $f \in \mathcal{L}^1 \mapsto \int |f|\, d\mu$ vérifie les propriétés

— d'homogénéité :

$$\int |af|\, d\mu = |a| \int |f| \quad (\forall a \in \mathbb{R})$$

— et d'inégalité triangulaire :

$$\int |f+g|\, d\mu \leqslant \int |f|\, d\mu + \int |g|\, d\mu,$$

mais pas

— de séparation :

$$\int |f|\, d\mu \nRightarrow f = 0$$

puisque l'on a seulement

$$\int |f|\, d\mu \Rightarrow f = 0 \text{ p.p.}$$

(si $a \in E$ et $\mu(\{a\}) = 0$, la fonction $\mathbb{1}_{\{a\}}$ a une intégrale nulle mais n'est pas identiquement nulle), ce qui fait de $\int |\cdot|\, d\mu$ une *semi-norme*.

Dans cette théorie, on aurait donc envie d'identifier les fonctions nulles presque partout à 0, et, plus généralement, d'identifier deux fonctions qui sont égales presque partout.

Définition 18.2. Deux fonctions f et g sont *équivalentes* si elles sont égales μ-presque partout.

La *classe d'équivalence* d'une fonction f est l'ensemble $[f]$ des fonctions g qui lui sont équivalentes.
On vérifie que la relation définie est bien d'une relation d'équivalence (sinon elle porterait mal son nom), c'est-à-dire que les classes d'équivalences forment une partition de l'ensemble des fonctions considérées.

Définition 18.3. Le *premier espace de Lebesgue* $L^1(E, \mathcal{A}, \mu)$ (aussi noté L^1, etc. en l'absence d'ambiguité) est l'ensemble des classes d'équivalence de fonctions dans $\mathcal{L}^1(E, \mathcal{A}, \mu)$.

À proprement parler, un élément $[f]$ de L^1 n'est donc pas exactement une fonction $E \to \mathbb{R}$. Il est cependant typique de faire l'abus de langage de parler d'une fonction f appartenant à L^1, et il est entendu qu'il s'agit en fait d'une fonction définie à équivalence près. Cela n'a pas de sens de s'intéresser à la valeur d'une classe de fonctions sur un ensemble négligeable, puisque la classe d'équivalence contient des fonctions prenant des valeurs distinctes en ces points. Mais, du point de vue des propriétés intégrales de f, seule sa classe compte et l'identification de f à sa classe simplifie les énoncés puisqu'elle permet de supprimer les “presque partout” qui traînent dans beaucoup d'affirmations.

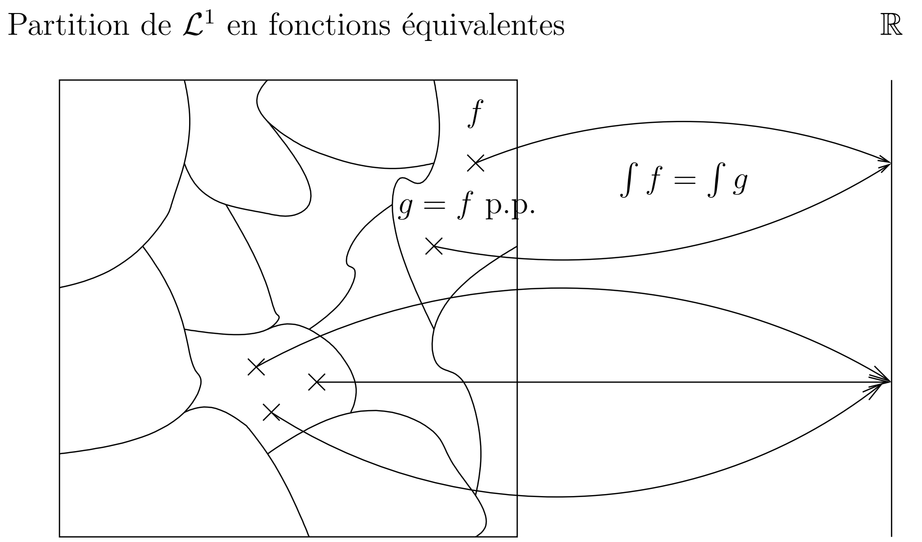


FIGURE 25 – L'intégrale définie sur L^1

L'intégrale

$$\int |f| \, d\mu$$

est inchangée si l'on remplace f par n'importe quelle fonction équivalente. Ceci permet de voir la fonction

$$f \mapsto \|f\|_{L^1} := \int |f| \, d\mu$$

comme une fonction sur l'espace L^1, à valeurs dans $]-\infty, +\infty[$. Cette fonction est bien une norme, puisque l'axiome de séparation est maintenant vérifié : si $f \in L^1$,

$$\|f\|_{L^1} = 0 \Rightarrow f = 0,$$

où la dernière égalité est à prendre au sens de L^1 : pour toute fonction $f_0 \in f$, $f_0 = 0$ p.p.

De même, les opérations arithmétiques habituelles gardent un sens dans L^1, par exemple la somme de deux (classes de) fonctions : si $f \sim f'$ et $g \sim g'$, alors $f + g \sim f' + g'$. On peut donc définir la somme de deux classes d'équivalence comme la classe d'équivalence de la somme de n'importe quels représentants.

Théorème 18.4. $(L^1, \|\cdot\|_{L^1})$ *est un espace de Banach.*[1]

Rappel 18.5. Soit $(E, \|\cdot\|)$ un espace vectoriel normé. Les deux propriétés suivantes sont équivalentes :

— E est de Banach (i.e. complet) ;

— toute série $\sum u_n$ de E absolument convergente (i.e. $\sum_{n\geqslant 0} \|u_n\| < \infty$) est convergente (i.e. $\sum_{n=0}^{N} u_n$ converge vers une limite dans E quand N tend vers l'infini).

Démonstration du théorème. Soit $(f_n)_{n\in\mathbb{N}}$ une suite de L^1 telle que $\sum_n f_n$ soit absolument convergente. On veut montrer que $\sum f_n$ est convergente. Par convergence monotone,

$$\int \left| \sum_{n\geqslant 0} |f_n| \right| d\mu = \int \sum_{n\geqslant 0} |f_n|\, d\mu = \sum_{n\geqslant 0} \int |f_n|\, d\mu,$$

c'est-à-dire

$$\left\| \sum_{n\geqslant 0} |f_n| \right\|_{L^1} = \sum_{n\geqslant 0} \|f_n\|_{L^1},$$

le membre de droite étant fini par hypothèse. En particulier, la fonction positive $\sum_{n\geqslant 0} |f_n|$ est finie presque partout, i.e. $\sum_{n\geqslant 0} f_n(x)$ est une série réelle absolument convergente pour μ-presque tout x. Comme $\mathbb{R}$ lui-même est complet, $\sum_{n\geqslant 0} f_n(x)$ converge pour μ-presque tout x ; notons

$$F(x) = \sum_{n\geqslant 0} f_n(x).$$

D'après le théorème de convergence dominée, la somme partielle $\sum_{0\leqslant n\leqslant N} f_n$ converge vers F dans L^1, ce qui conclut la démonstration. □

Exemple 18.6 (Espace ℓ^1). Dans le cas où $E = \mathbb{N}$ et où μ est la mesure de comptage (qui charge tous les points de $\mathbb{N}$, par définition), on a a déjà vu que $\mathcal{L}^1$ est l'ensemble des suites réelles (u_n) sommables, c'est-à-dire telles que $\sum_{n\geqslant 0} |u_n| < \infty$. Comme chaque suite $(u_n)_{n\in\mathbb{N}}$ n'est équivalente qu'à elle-même, $L^1 \equiv \mathcal{L}^1$. On note $\ell^1 = \ell^1(\mathbb{N})$ ce dernier espace.

1. Stefan Banach, mathématicien polonais (1892–1945), l'un des fondateurs de l'analyse fonctionnelle

Plus généralement, on définit l'espace $\mathcal{L}^p$ $(0 < p < \infty)$ comme l'espace vectoriel des fonctions mesurables $f : E \to \mathbb{R}$ dont la puissance p-ième soit intégrable :

$$\int |f|^p \, d\mu < \infty,$$

puis l'*espace de Lebesgue* L^p comme l'ensemble des classes d'équivalence de telles fonctions f. On munit ce dernier de la norme

$$\|f\|_{L^p} := \left(\int |f|^p \, d\mu \right)^{1/p}.$$

Il est non trivial, mais élémentaire, de montrer que L^p est un espace vectoriel. (Cf. le problème 18.1, question 2.) La même démonstration que pour L^1 montre que $(L^p, \| \cdot \|_{L^p})$ est un espace de Banach (théorème de Riesz [2]).
On définit aussi l'espace L^∞, de la façon suivante.

Définition 18.7. $f : E \to \mathbb{R}$ est *essentiellement bornée* si il existe $M \geqslant 0$ tel que $f \leqslant M$ μ-p.p. La *norme* L^∞ de f est la borne inférieure des tels M :

$$\|f\|_{L^\infty} := \inf\{M;\ \mu(\{x;\ f(x) > M\}) = 0\}.$$

Enfin, $\mathcal{L}^\infty(E, \mathcal{A}, \mu)$ est l'espace des fonctions de norme L^∞ finie, et l'*espace de Lebesgue* L^∞ est l'ensemble de ses classes d'équivalence.

(Le soin de vérifier que $\| \cdot \|_{L^\infty}$ est une norme est laissé au lecteur.)

Remarque 18.8. Soit $A \in \mathcal{A}$ une partie telle que $0 \leqslant \mu(A) < \infty$. Soient de plus $a > 0$ (amplitude) et $f = a\mathbb{1}_A$. Pour tout $p \in]0, +\infty[$,

$$\|f\|_{L^p} = a(\mu(A))^{1/p} \to_{p \to +\infty} a = \|f\|_{L^\infty},$$

donc dans ce cas au moins la norme L^∞ est la limite des normes L^p quand p tend vers l'infini. Par ailleurs, dans ce cas on voit aussi que les normes L^p sont une combinaison de la "hauteur" a et de la "largeur" $\mu(A)$, bien que ces deux quantités ne soient pas définies pour des fonctions plus générales.

Tous ces espaces de Lebesgue sont des exemples importants d'espaces de Banach. Mais dans ce cours nous n'utiliserons que L^1 (dont la définition même est appelée par la théorie), L^2 (un espace de Hilbert, que nous étudierons dans le chapitre suivant) et L^∞ (le dual topologique de L^1).

Proposition 18.9. $(L^\infty, \| \cdot \|_{L^\infty})$ *est un espace de Banach.*

La démonstration est encore plus simple que pour les autres espaces L^p.

2. Frigyes Riesz, mathématicien hongrois (1880–1956), aux nombreuses contributions en analyse fonctionnelle

Démonstration. Si $\sum f_n$ une série absolument convergente de L^∞, pour tout entier N, d'après l'inégalité triangulaire,

$$\left\| \sum_{0\leqslant n\leqslant N} |f_n| \right\|_{L^\infty} \leqslant \sum_{0\leqslant n\leqslant N} \|f\|_{L^\infty} \leqslant \sum_{n\geqslant 0} \|f\|_{L^\infty} < \infty,$$

donc μ-p.p. la série numérique $\sum f_n(x)$ converge absolument, et, puisque $\mathbb{R}$ est complet, cette série converge vers une limite $F(x)$ dont on voit qu'elle appartient à L^∞. □

Exemple 18.10 (Espace ℓ^∞). Dans le cas où $E = \mathbb{N}$ et où μ est la mesure de comptage, on note ℓ^∞ l'espace $\mathcal{L}^\infty = L^\infty$. Pour une suite de E, “essentiellement borné” signifie “borné”.

Terminons ce chapitre en indiquant le lien qui unit L^1 et L^∞.[3] Rappelons que le *dual (topologique)* d'un espace vectoriel normé F est l'espace $F' := L(F, \mathbb{R})$ des formes linéaires continues sur F.

Exercice 18.1 (Dual en dimension finie)**.**
Montrer que le dual de $\mathbb{R}^n$ est isomorphe à $\mathbb{R}^n$.

Exercice 18.2 (Dual d'un espace normé)**.**
Montrer que, si F est non trivial, F' est non trivial ; on pourra (et devra !) utiliser le théorème de Hahn-Banach.

Théorème 18.11. *Si μ est σ-finie, le dual de L^1 s'identifie à L^∞.*

Début de la démonstration du théorème. Si $\varphi \in L^\infty$, la forme linéaire

$$T_\varphi : L^1 \to \mathbb{R}, \quad f \mapsto \int \varphi f \, d\mu$$

est bien définie sur L^1 et continue parce que

$$\int |\varphi f| \, d\mu \leqslant \|\varphi\|_{L^\infty} \int |f| \, d\mu = \|\varphi\|_{L^\infty} \|f\|_{L^1}. \tag{18.1}$$

Réciproquement, nous montrerons dans le chapitre sur L^2 que l'application ainsi définie

$$L^\infty \to L^{1\prime}, \quad \varphi \mapsto T_\varphi,$$

est surjective. □

Remarque 18.12. Étant donné la symétrie de l'estimation (18.1), on pourrait être tenté de penser que, de même que L^∞ est le dual de L^1, L^1 est le dual de L^∞. Or, en général ce n'est pas le cas ; on dit que L^1 n'est pas *réflexif* parce qu'il n'est pas isomorphe à son *bidual* : $L^1 \subsetneq L^{1\prime\prime}$. L'estimation en question montre bien que L^1 s'injecte dans $L^{\infty\prime}$. Mais cette injection n'est pas surjective : il existe des formes linéaires continues sur L^∞ qui ne soient pas de la forme $\varphi \mapsto \int \varphi f \, d\mu$ avec $f \in L^1$.

3. Le théorème ci-dessous est de même nature que le théorème de Radon-Nikodym, et lui est “quasiment” équivalent.

Voyons-en un exemple dans le cas de $\ell^\infty(\mathbb{N})$. Soit $c(\mathbb{N})$ le sous-espace de $\ell^\infty(\mathbb{N})$ constitué par les suites convergentes. La fonctionnelle “limite”

$$\lambda : c(\mathbb{N}) \to \mathbb{R}, \quad u = (u_n) \mapsto \lim u_n$$

est bornée de norme 1. D'après le théorème de Hahn-Banach, λ possède un prolongement

$$\lambda : \ell^\infty(\mathbb{N}) \to \mathbb{R}$$

de norme 1. Supposons par l'absurde que ce prolongement soit de la forme

$$\lambda(u) = \sum_{n \geqslant 0} a_n u_n \quad (\forall u \in \ell^\infty(\mathbb{N}))$$

pour une certaine suite $(a_n) \in \ell^1(\mathbb{N})$. En appliquant λ successivement à toutes les suites $(u_n) = (\mathbb{1}_{\{k\}}(n))$ avec $k \in \mathbb{N}$, on trouve

$$\lambda(\mathbb{1}_{\{k\}}) = a_k;$$

mais, puisque la suite $(\mathbb{1}_k)$ est dans $c(\mathbb{N})$ et que sa limite est nulle, ceci implique $a = 0$, ce qui est absurde puisque λ est non triviale.

Autres exercices

L'exercice suivant montre qu'on aurait pu définir L^1 comme le complété de l'espace des fonctions étagées intégrables (quotienté par la relation d'égalité presque partout), ou comme le complété de l'espace des fonctions continues à support compact, relativement à la norme $\|\cdot\|_{L^1}$. L'énorme inconvénient d'une telle définition, si courte soit-elle, est que les points ajoutés à ces ensembles pour former L^1 ne seraient pas explicitement « réalisés » comme des fonctions.

Exercice 18.3 (Densité de $Lip_c(\mathbb{R}^d)$ dans $L^1(\mathbb{R}^d)$)**.**
Dans cet exercice, on établit un résultat de densité dans l'espace $L^1 = L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$. Dans chaque question, ϵ est un réel > 0 arbitraire.

1. Montrer que, si O est un ouvert borné de $\mathbb{R}^d$, il existe une fonction lischitzienne g à support (fermé) compact telle que $\|\mathbb{1}_O - g\|_{L^1} \leqslant \epsilon$.

2. En déduire que, si A est est un borélien de $\mathbb{R}^d$ de mesure finie, il existe une fonction lischitzienne g à support compact telle que $\|\mathbb{1}_A - g\|_{L^1} \leqslant \epsilon$. On pourra utiliser la régularité extérieure de la mesure de Lebesgue, à savoir que, pour tout $\epsilon > 0$, il existe un ouvert O contenant A et tel que $\lambda(O \backslash A) < \epsilon$ (exercice 11.4).

3. En déduire que, si f est une fonction étagée intégrable, il existe une fonction lischitzienne g à support compact telle que $\|f - g\|_{L^1} \leqslant \epsilon$.

4. En déduire que, si $f \in L^1$, il existe une fonction lipschitzienne g à support compact telle que $\|f - g\|_{L^1} \leqslant \epsilon$.

Remarques : (1) La densité des fonctions étagées intégrables dans L^1 est vraie pour un espace mesuré quelconque — la démonstration est la même.

(2) Dans $\mathbb{R}^d$, avec les opérateurs de convolution, on verra mieux : l'espace des fonctions C^∞ à support compact est dense dans L^1 (exercice 19.3).

Problème 18.1 (Inégalités de Hölder et de Minkowski).
Soient $(E, \mathcal{A}, \mu)$ un espace mesuré, $p, q \in]0, +\infty[$ deux réels *conjugués*, au sens que

$$\frac{1}{p} + \frac{1}{q} = 1,$$

et f et g deux fonctions réelles mesurables sur E.

1. Montrer l'*inégalité de Hölder* :

$$\|fh\|_{L^1} \leqslant \|f\|_{L^p} \|h\|_{L^q}.$$

on pourra remarquer que l'inégalité est triviale lorsque le membre de droite est infini, puis, dans le cas contraire, utiliser l'inégalité de Jensen pour la probabilité de densité $|f|^p / \int |f|^p$ par rapport à μ, la fonction $u = |g|/|f|^{p-1} \mathbb{1}_{f \neq 0}$ et la fonction convexe $q \mapsto u^q$.

(En particulier, cette inégalité montre que, si $f \in L^p$ et $g \in L^q$, $fg \in L^1$.)

2. Montrer que, si $f, g \in L^p$, $f + g \in L^p$.

3. Montrer l'*inégalité de Minkowski* :

$$\|f + g\|_{L^p} \leqslant \|f\|_{L^p} + \|g\|_{L^p}.$$

Problème 18.2 (Borne supérieure dans $\mathcal{L}^1$ et L^1).
On munit les deux espaces $\mathcal{L}^1$ et L^1 de la relation d'ordre définie par

$$f \leqslant g \Leftrightarrow g - f \geqslant 0.$$

Soit (f_i) une famille d'éléments positifs de $\mathcal{L}^1$ ou de L^1, majorés par un autre élément $g \geqslant 0$, stable par l'opération "sup fini".

1. Montrer que, dans $\mathcal{L}^1$, la famille (f_i) n'a pas nécessairement une borne supérieure, et que, si elle en a une, on n'a pas toujours $\int \sup f_i = \sup \int f_i$.

2. Montrer que, dans L^1, la famille (f_i) a nécessairement une borne supérieure, et que $\int \sup f_i = \sup \int f_i$.

Problème 18.3 ($\ell^1(\mathbb{N})$ comme dual).
On considère les deux sous-espaces suivants de $\ell^\infty(\mathbb{N})$:

— $c_c(\mathbb{N})$, constitué des suites à support compact (i.e. nulles à partir d'un certain rang dépendant de la suite),
— $c_0(\mathbb{N})$, constitué des suites de limite nulle.

1. Montrer que le dual $c_c(\mathbb{N})'$ est isomorphe à $\ell^1(\mathbb{N})$.

2. Montrer que le complété de $c_c(\mathbb{N})$ est isomorphe à $c_0(\mathbb{N})$.

3. En déduire que le dual $c_0(\mathbb{N})'$ est isomorphe à $\ell^1(\mathbb{N})$.

4. Montrer que le dual de $\ell^1(\mathbb{N})$ est $\ell^\infty(\mathbb{N})$.

Problème 18.4 (L^1 non réflexif).

1. Montrer qu'un espace de Banach est réflexif si et seulement si son dual est réflexif.

2. En déduire que si un espace mesuré $(E, \mathcal{A}, \mu)$ contient une suite infinie de parties disjointes de mesure > 0, $L^1(E, \mathcal{A}, \mu)$ et $L^\infty(E, \mathcal{A}, \mu)$ ne sont pas réflexifs.

FIGURE 26 – Stefan BANACH (1892–1945)

Chapitre 19

Convolution

Les raisons pour introduire le produit de convolution des fonctions sont nombreuses. Par exemple, la convolution permet de construire des “approximations de l’identité” et elle est transformée en le produit habituel par la transformation de Fourier (voir le chapitre sur la transformation de Fourier). Plus physiquement, considérons un système mécanique, financier, biologique, ou autre, que l’on excite avec un signal d’entrée $f(t)$ et qui réagit avec un signal de sortie $s(t)$. Il est courant que l’opérateur $A : f \mapsto s$ possède les propriétés suivantes :

- Linéarité : $A(\alpha f + \beta g) = \alpha A(f) + \beta A(g)$.
- Homogénéité dans le temps : $A(f(\tau - \cdot)) = A(f)(\tau - \cdot)$ (i.e. si on applique en entré le même signal décalé dans le temps, on obtient en sortie le même signal de sortie décalé du même intervalle de temps)
- Causalité : si $f(t) = 0$ pour $t \leqslant t_0$, $A(f)(t) = 0$ pour $t \leqslant t_0$.

L’opération de convolution telle que nous allons la définir modélise, dans une large mesure, ce type de situation.

Théorème et Définition 19.1. *Si* $f, g \in L^1 = L^1_{\mathbb{C}}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$, *le* produit de convolution *de* f *et* g *est la fonction*

$$f * g : \mathbb{R}^d \to \mathbb{C}, \quad x \mapsto \int f(x-y)g(y)\,dy = \int g(x-y)f(y)\,dy. \tag{19.1}$$

Il appartient à L^1 *et*

$$\|f * g\|_{L^1} \leqslant \|f\|_{L^1}\|g\|_{L^1}. \tag{19.2}$$

Démonstration. Le produit de deux fonctions intégrables n’est généralement pas intégrable. Pourtant, ici c'est bien le cas. En effet,

$$\begin{aligned}
\iint |f(x-y)g(y)|\,dx\,dy &= \int |g(y)| \left(\int |f(x-y)|\,dx \right) dy \quad \text{(théorème de Tonelli)} \\
&= \int |g(y)| \left(\int |f(x)|\,dx \right) dy \\
&\text{(changement de variable } x' = x-y, \text{ à } y \text{ fixé)} \\
&= \left(\int |f(x)|\,dx \right) \left(\int |g(y)|\,dy \right) \quad \text{(linéarité)} \\
&= \|f\|_{L^1} \|g\|_{L^1} < \infty,
\end{aligned}$$

donc $(x, y) \mapsto f(x-y)g(y)$ est intégrable sur $\mathbb{R}^d \times \mathbb{R}^d$. Donc, d'après le théorème de Fubini,

— dx-p.p., la fonction $y \mapsto f(x-y)g(y)$ est intégrable (donc le produit de convolution est bien défini p.p. par la première formule intégrale donnée)

— et $f * g$ est intégrable.

L'estimation annoncée découle du calcul précédent :

$$\|f * g\|_{L^1} = \int |f * g(x)|\,dx \leqslant \iint |f(x-y)g(y)|\,dx\,dy \leqslant \|f\|_{L^1} \|g\|_{L^1}.$$

Il reste à démontrer que

$$\int f(x-y)g(y)\,dy = \int g(x-y)f(y)\,dy$$

(commutativité du produit de convolution). Mais cette égalité provient de la formule du changement de variable $y' = x - y$ (de jacobien $(-1)^d$). □

Par exemple, si $f = \frac{1}{2}\mathbb{1}_{[-1,1]}$ sur $\mathbb{R}$, la convolution par f, en un point x, revient à moyenner sur l'intervalle $[x-1, x+1]$, comme le changement de variable $y' = x-y$ le montre :

$$f * g(x) = \int g(x-y)f(y)\,dy = \frac{1}{2}\int_{-1}^{1} g(x-y)\,dy = \frac{1}{2}\int_{x-1}^{x+1} g(y')\,dy'.$$

Exercice 19.1 (Dérivée d'une convolée)**.**
Soient $f \in L^1$ et g continue et bornée.

1. Montrer que l'équation (19.1) définit une fonction $f * g$ continue et bornée.

2. Soit $r \in \mathbb{N}$. Montrer que si les dérivées partielles d'ordre r de g sont continues et bornées, les dérivées partielles d'ordre r de $f * g$ le sont aussi, et

$$\frac{\partial(f * g)}{\partial x_j} = f * \frac{\partial g}{\partial x_j}, \quad \frac{\partial^2(f * g)}{\partial x_j \partial x_k} = f * \frac{\partial^2 g}{\partial x_j \partial x_k}, \quad \ldots$$

Exercice 19.2 (Convolution d'une fonction dans L^1_{loc})**.**
Notons $L^1_{loc} = L^1_{loc}(\mathbb{R}^n)$ l'espace des fonctions f *localement intégrables*, c'est-à-dire telles que, pour tout compact $K \subset \mathbb{R}^d$, la fonction $f\mathbb{1}_K$ soit intégrable. Soient

$f \in L^1_{loc}$ et $g \in L^1$ nulle en dehors d'un certain compact K de $\mathbb{R}^d$. Montrer que l'intégrale

$$f * g(x) = \int f(x-y)g(y)\,dy$$

est définie dx-p.p. et que $f * g \in L^1_{loc}$.

Exercice 19.3 (Approximations de l'identité)**.**
Soient $f \in L^1 = L^1(\mathbb{R}^d)$ et

$$h(x) = \frac{1}{\pi^{d/2}} e^{-\|x\|^2} \quad \text{et} \quad h_n(x) = n^d h(nx) \quad (x \in \mathbb{R}^d,\ n \in \mathbb{N}).$$

1. Montrer que $h, h_n \in C^\infty \cap L^1$ et que $\int h = \int h_n = 1$.

2. Montrer que, pour tout n, $f * h_n$ est de classe C^∞ et donc que C^∞ est dense dans L^1.

Pour cette raison, la suite des opérateurs de convolution $f \mapsto f * h_n$ s'appelle aussi une *suite régularisante*.

3. Montrer que $(f*h_n)$ converge vers f dans L^1 ; on pourra commencer par supposer que f est continue à support compact, puis utiliser le résultat de densité démontré dans l'exercice 18.3.

La suite des opérateurs de convolution $f \mapsto f * h_n$ et, par abus de langage, la suite des fonctions h_n elle-même, s'appellent une *approximation de l'identité*.

L'objectif final étant de démontrer que l'espace C_0^∞ des fonctions C^∞ à support compact est dense dans L^1, on désire d'abord définir un analogue de la fonction h à support compact.

4. Montrer que la fonction

$$\varphi : x \mapsto \begin{cases} \exp\left(-\frac{1}{1-\|x\|^2}\right) & \text{si } \|x\| < 1 \\ 0 & \text{si } \|x\| \geqslant 1 \end{cases}$$

est dans C_0^∞ ; on pourra considérer la fonction intermédiaire sur $\mathbb{R}$

$$u \mapsto \begin{cases} \exp\left(-\frac{1}{u}\right) & \text{si } u > 0 \\ 0 & \text{si } u \leqslant 0. \end{cases}$$

5. Montrer que les fonctions $\eta(x) := \frac{\varphi}{\int \varphi}$ et $\eta_n(x) := n^d \eta(nx)$ vérifient les propriétés qui nous ont été utiles pour h et h_n, en plus d'être à support compact.

6. Montrer que, si $f \in C_0^0$, $f * \eta_n \in C_0^\infty$.

7. En déduire que C_0^∞ est dense dans L^1.

Autres exercices

Problème ★ 19.1 (Autour du théorème de Steinhaus [Tao10a])**.**
1. Soit $E \subset \mathbb{R}^d$ une partie Lebesgue-mesurable de mesure > 0. Montrer que l'ensemble

$$E - E := \{x - y;\ x, y \in E\}$$

contient un voisinage ouvert de l'origine (théorème de Steinhaus).

Indication : Se ramener au cas où E est borné, puis considérer la convolution $\mathbb{1}_E * \mathbb{1}_{-E}$, où $-E := \{-y;\ y \in E\}$.

Un *homomorphisme* $f : \mathbb{R}^d \to \mathbb{C}$ est une application vérifiant $f(x+y) = f(x)+f(y)$ quels que soient $x, y \in \mathbb{R}^d$.

2. Montrer que les homomorphismes mesurables sont continus.

Indication : Pour tout disque D centré à l'origine dans $\mathbb{C}$, montrer que $f^{-1}(z + D)$ est de mesure non nulle pour au moins un $z \in \mathbb{C}$, puis utiliser la question précédente.

3. Montrer que f est un homomorphisme mesurable si et seulement si f est de la forme

$$f(x_1, ..., x_d) = x_1 z_1 + \cdots + x_d z_d$$

pour certains coefficients complexes $z_1, ..., z_d$, quels que soient $x_1, ..., x_d \in \mathbb{R}$.

Indication : L'établir d'abord quand $x_1, ..., x_d$ sont rationnels, puis utiliser la question précédente.

4. (Pour les lecteurs familiers avec le lemme de Zorn, cf. [Tao10a, § 2.4]) Montrer qu'il existe des homomorphismes $f : \mathbb{R}^d \to \mathbb{C}$ qui ne sont pas de la forme précédente.

Indication : Utiliser la structure de $\mathbb{Q}$-espace vectoriel de $\mathbb{R}^d$ et de $\mathbb{C}$, et le fait (conséquence du lemme de Zorn) que ces espaces vectoriels possèdent une base algébrique. (Ceci fournit une construction alternative d'ensembles non Lebesgue-mesurables.)

Chapitre 20

Transformée de Fourier

Dans ce chapitre, nous notons $L^1 = L^1_{\mathbb{C}}(\mathbb{R}^d, \mathcal{B}(\mathbb{R}), \lambda)$.

Définition 20.1. La *transformée de Fourier* [1] d'une fonction $f \in L^1$ est la fonction

$$\mathcal{F}(f) = \hat{f} : \mathbb{R}^d \to \mathbb{C}, \quad \xi \mapsto \int_{\mathbb{R}^d} f(x) e^{-ix\cdot\xi}\, dx,$$

où $x \cdot \xi = x_1\xi_1 + \cdots + x_d\xi_d$.
La *transformée de Fourier* d'une mesure μ finie sur $\mathbb{R}^d$ est la fonction

$$\mathcal{F}(\mu) = \hat{\mu} : \mathbb{R}^d \mapsto \mathbb{C}, \quad \xi \mapsto \int_{\mathbb{R}^d} e^{-ix\cdot\xi}\, d\mu(x).$$

En probabilités, la *fonction caractéristique* d'une variable aléatoire $X : (\Omega, \mathcal{A}, P) \to \mathbb{R}^d$ est la fonction $\Phi_X := \widehat{P_X}$, c'est-à-dire telle que

$$\Phi_X(\xi) = \int e^{-ix\cdot\xi}\, dP_X(x) = E\left(e^{-iX\cdot\xi}\right).$$

La cohérence de ces définitions est que, si μ a une densité f, $\hat{\mu} = \hat{f}$.
Dans ce chapitre, nous nous focalisons sur la première des trois définitions. Les fonctions caractéristiques seront utilisée ultérieurement en probabilités.

Théorème 20.2. *La fonction $\hat{f}$ est continue, tend vers 0 à l'infini et vérifie*

$$\|\hat{f}\|_{L^\infty} \leqslant \|f\|_{L^1}. \tag{20.1}$$

Démonstration. Le fait que $\hat{f}$ soit continue a déjà été démontré dans l'exercice 16.1. La majoration (20.1) découle de la croissance de l'intégrale. Il reste à démontrer que la limite de $\hat{f}$ à l'infini est nulle ("lemme de Riemann-Lebesgue").

1. Joseph Fourier, mathématicien français (1768–1830), célèbre pour son étude des équations de la chaleur et des ondes.

Supposons d'abord que f est continue à support compact. Pour $\xi \neq 0$,

$$\begin{aligned}\hat{f}(\xi) &= \int e^{-ix\cdot\xi} f(x)\,dx \\ &= -\int e^{-i\left(x-\pi\frac{\xi}{\|\xi\|^2}\right)\cdot\xi} f(x)\,dx \\ &= -\int e^{-ix\cdot\xi} f\left(x+\pi\frac{\xi}{\|\xi\|^2}\right)\,dx,\end{aligned}$$

donc

$$\hat{f}(\xi) = \frac{1}{2}\int e^{-ix\cdot\xi}\left(f(x) - f\left(x+\pi\frac{\xi}{\|\xi\|^2}\right)\right)\,dx.$$

Soit B une boule contenant le support de f :

$$\hat{f}(\xi) = \frac{1}{2}\int_{B\cup(B-\pi\xi/\|\xi\|^2)} e^{-ix\cdot\xi}\left(f(x) - f\left(x+\pi\frac{\xi}{\|\xi\|^2}\right)\right)\,dx.$$

De l'expression précédente, il vient

$$\left|\hat{f}(\xi)\right| \leqslant \mathrm{Vol}\,(B) \sup_{|y-x|\leqslant\pi\xi/\|\xi\|^2} |f(y)-f(x)|.$$

La fonction f étant continue à support compact, elle est uniformément continue. Donc le membre de droite de la dernière inégalité tend vers 0 quand ξ tend vers l'infini.

Soit maintenant $f \in L^1$. D'après l'exercice 19.3, il existe des fonctions f_n continues à support compact telles que

$$\int |f-f_n|\,dx \to 0.$$

Il découle de l'inégalité (20.1) que $\|\hat{f}-\hat{f}_n\|_{L^\infty}$ tend vers 0. Or, pour tout ξ,

$$|\hat{f}(\xi)| \leqslant |\hat{f}(\xi)-\hat{f}_n(\xi)| + |\hat{f}_n(\xi)|;$$

dans le membre de droite, le premier terme est majoré par $\|f-f_n\|_{L^1}$, qui est arbitrairement petit si n est assez grand, uniformément par rapport à ξ ; et le second terme est alors arbitrairement petit si $\|\xi\|$ est assez grand, d'après le cas particulier des fonctions continues à support compact. Donc $\hat{f}$ tend vers 0 à l'infini. □

Si l'on utilise un théorème d'approximation plus fort, on obtient une démonstration très simple, comme le montre l'exercice suivant en dimension $d = 1$; dans le cas général (d quelconque), la *formule de Stokes* doit se subtituer à la formule d'intégration par parties, ce que nous ne ferons pas ici.

Exercice 20.1 (Sur le lemme de Riemann-Lebesgue)**.**
Montrer que le lemme de Riemann-Lebesgue en dimension $d = 1$ pour les fonctions f de classe C^∞ à support compact ; on pourra faire une intégration par parties.

Exemple 20.3 (Indicatrice d'un intervalle).

$$\widehat{\mathbb{1}_{[a,b]}}(\xi) = \begin{cases} \frac{2\sin\left(\frac{b-a}{2}\xi\right)}{\xi} e^{-i\frac{a+b}{2}\xi} & \text{si } \xi \neq 0 \\ 0 & \text{si } \xi = 0. \end{cases}$$

En revanche, d'après le théorème, $\mathbb{1}_{[a,b]}$ (étant discontinue) n'est la transformée de Fourier d'aucune fonction L^1.

Exemple 20.4 (Gaussienne). Pour $a > 0$, on a

$$\mathcal{F}\left(e^{-ax^2/2}\right)(\xi) = \sqrt{\frac{2\pi}{a}}\, e^{-\xi^2/(2a)}.$$

Première démonstration. On a

$$\mathcal{F}\left(e^{-ax^2/2}\right)(\xi) = e^{-\xi^2/(2a)} \int e^{-a\left(x+\frac{i\xi}{2a}\right)^2/2}\, dx. \tag{20.2}$$

En écrivant que l'intégrale de e^{-az^2} sur le contour orienté dessiné sur la figure 27 est nul, puis en montrant que les contributions des côtés verticaux tend vers 0 quand on étire à l'infini le rectangle horizontalement, on voit que l'intégrale du membre de droite de (20.2) est égale à $\int e^{-ax^2/2}\, dx = \sqrt{\frac{2\pi}{a}}$ (exercice 14.5).

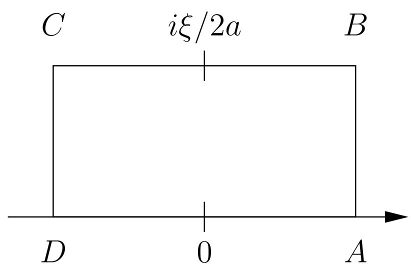


FIGURE 27 – Le contour d'intégration dans l'exemple 20.4

Deuxième démonstration. Par application du théorème de convergence dominée (ou, plus directement, du théorème 16.2 ou de l'exercice 16.2), la transformée de Fourier de $f(x) = e^{-ax^2/2}$ est dérivable de dérivée

$$\hat{f}'(\xi) = -i \int e^{-ix\xi} \left(xe^{-ax^2/2}\right) dx.$$

Une intégration par partie montre que

$$\hat{f}(\xi) = -\frac{\xi}{a} \int e^{-ix\xi} e^{-ax^2/2}\, dx = -\frac{\xi}{a} \hat{f}(\xi).$$

Donc il existe un réel C tel que

$$\hat{f}(\xi) = Ce^{-\xi^2/(2a)}.$$

Mais, on sait que

$$\hat{f}(0) = \int_{\mathbb{R}} e^{-ax^2/2}\,dx = \sqrt{\frac{2\pi}{a}}$$

(exercice 14.5). D'où la formule annoncée.
Par le changement de variable élémentaire $x' = (x-\mu)/\sigma$ et en prenant $a = 1$, on en déduit que la fonction caractéristique d'un variable aléatoire X de loi gaussienne $N(\mu, \sigma^2)$

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

est

$$\Phi_X(\xi) = e^{-i\mu\xi - \frac{1}{2}\sigma^2\xi^2}.$$

Théorème 20.5 (Formule d'inversion)**.** *Si $f \in L^1$ et $\hat{f} \in L^1$,*

$$f = \frac{1}{(2\pi)^d}\bar{\mathcal{F}}(\hat{f}) \quad p.p., \tag{20.3}$$

où l'on note classiquement

$$\bar{\mathcal{F}}(g)(x) = \check{g}(x) = \int_{\mathbb{R}^d} e^{ix\cdot\xi} g(\xi)\,d\xi$$

(par rapport à $\mathcal{F}$, on a juste remplacé i par $-i$).

Remarque 20.6. Plus explicitement, (20.3) s'écrit

$$f(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^n} e^{i(x-y)\cdot\xi} f(y)\,dy\right) d\xi.$$

Formellement, on aimerait intervertir l'ordre des deux intégrales (que faire d'autre ?) et écrire le membre de droite comme

$$\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} e^{i(x-y)\cdot\xi}\,d\xi\right) f(y)\,dy,$$

puis essayer de se convaincre que cette expression vaut $f(x)$. Mais l'utilisation ici du théorème de Fubini est illicite parce que la fonction $(y, \xi) \mapsto e^{i(x-y)\cdot\xi)} f(y)$ n'est pas intégrable sur $\mathbb{R}^{2d}$. Nous allons adapter cette idée en multipliant cette fonction par $e^{-\epsilon^2\|\xi\|^2/4}$ (qui, en particulier, tend vers 1 quand ϵ tend vers 0), puis faire tendre ϵ vers 0 dans les deux expressions obtenues en intervertissant l'ordre d'intégration.

Démonstration. Soit, pour $n \geqslant 0$,

$$I_\epsilon(x) = \frac{1}{(2\pi)^d}\iint_{\mathbb{R}^{2d}} e^{i(x-y)\cdot\xi} e^{-\epsilon^2\|\xi\|^2/4} f(y)\,dy\,d\xi.$$

Cette fois-ci, le théorème de Fubini s'applique. En intégrant par rapport à ξ et à y dans les deux ordres possibles, on va obtenir deux expressions de la limite de I_n quand n tend l'infini, et, par unicité de la limite, obtenir l'égalité voulue.

En intégrant d'abord par rapport à y, on obtient

$$I_\epsilon(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{ix\cdot\xi} e^{-\epsilon^2\|\xi\|^2/4} \hat{f}(\xi) d\xi.$$

La fonction qui reste à intégrer par rapport à ξ est dominée par $|\hat{f}| \in L^1$ et tend vers $e^{ix\cdot\xi}\hat{f}(\xi)$ quand ϵ tend vers 0. D'après le théorème de convergence dominée,

$$\lim_{\epsilon\to 0} I_\epsilon(x) = \frac{1}{(2\pi)^d} \bar{\mathcal{F}}(\hat{f})(x).$$

En revanche, en intégrant d'abord par rapport à ξ, on obtient

$$I_\epsilon(x) = \int h_\epsilon(x-y) f(y)\, dy = h_\epsilon * f(x),$$

avec

$$h_\epsilon(x) = \frac{1}{(2\pi)^d} \int e^{ix\cdot\xi} e^{-\epsilon^2\|\xi\|^2/4}\, d\xi.$$

Rappelons l'égalité montrée dans l'exemple 20.4 (où l'on a interverti les rôles de x et de ξ, et pris la conjugaison complexe des deux membres), en dimension 1 :

$$\int_{\mathbb{R}} e^{ix_1\xi_1}\, e^{-a\xi_1^2/2}\, d\xi_1 = \sqrt{\frac{2\pi}{a}}\, e^{-x_1^2/(2a)}.$$

En dimension d, d'après le théorème de Fubini on obtient

$$\int_{\mathbb{R}^d} e^{ix\cdot\xi}\, e^{-a\|\xi\|^2/2}\, d\xi = \left(\frac{2\pi}{a}\right)^{d/2} e^{-\|x\|^2/(2a)}$$

et, en prenant $a = \epsilon^2/2$,

$$\int_{\mathbb{R}^d} e^{ix\cdot\xi}\, e^{-\epsilon^2\|\xi\|^2/4}\, d\xi = \left(\frac{4\pi}{\epsilon^2}\right)^{d/2} e^{-\|x\|^2/\epsilon^2}.$$

Donc

$$h_\epsilon(x) = \frac{1}{\pi^{d/2}\epsilon^d} e^{-\|\xi\|^2/\epsilon^2} = \frac{1}{\epsilon^d} h_1\left(\frac{\xi}{\epsilon}\right).$$

La convolution par ces fonctions forme une approximation de l'identité et, de la même façon que dans l'exercice 19.3 (où n jouait le rôle joué ici par $1/\epsilon$), on voit que

$$\|f - h_\epsilon * f\|_{L^1} = \|f - I_\epsilon\|_{L^1} \to_{\epsilon\to 0} 0.$$

On admet provisoirement que ceci implique qu'il existe une suite ϵ_n tendant vers 0 telle que $I_{\epsilon_n} \to f$ p.p. (cette réciproque partielle du théorème de convergence dominée sera démontrée dans le chapitre sur les modes de convergence). Donc $f = (2\pi)^{-d}\bar{\mathcal{F}}(\hat{f})$ p.p. □

Corollaire 20.7 (Injectivité)**.** *Si $f, g \in L^1$ et $\hat{f} = \hat{g}$, $f = g$.*

Autrement dit, la transformée de Fourier d'une fonction L^1 la caractérise.

Démonstration. Sous ces hypothèses, $\widehat{f-g} = 0 \in L^1$. Donc, d'après la formule d'inversion, $f - g = (2\pi)^{-d}\bar{\mathcal{F}}(\widehat{f-g}) = 0$ p.p. □

Remarque 20.8. Dans le corollaire, il n'est pas nécessaire que $\hat{f}$ et $\hat{g}$ soient sommables, parce que la formule d'inversion est appliquée à la fonction $\widehat{f-g}$ qui, par hypothèse ici, est nulle.

La formule d'inversion peut-être adaptée aux probabilités (voir l'exercice 20.7). On en déduit alors de la même façon l'analogue du corollaire précédent.

Théorème 20.9 (admis)**.** *La transformation de Fourier des mesures de probabilités est injective.*

Remarque 20.10 (Normalisation)*.* Si l'on note

$$\hat{f}(\xi) = C^n \int e^{-i\alpha x\cdot\xi} f(x)\,dx \quad (\alpha, C \in \mathbb{R}_*)$$

la formule d'inversion devient

$$f(x) = \left(\frac{|\alpha|}{2\pi C}\right)^n \int e^{i\alpha x\cdot\xi}\hat{f}(\xi)\,d\xi.$$

Théorème 20.11. *Soit $f \in L^1 \cap C^1$ dont les dérivées partielles aussi sont intégrables. Alors*

$$\mathcal{F}\left(\frac{\partial f}{\partial x_j}\right) = i\xi_j\hat{f}(\xi).$$

Soit g une fonction intégrable telle que les fonctions $x \mapsto x_j g(x)$ soient intégrables. Alors $\hat{g}$ est de classe C^1 et

$$\mathcal{F}\left(x_j g(x)\right) = i\frac{\partial \hat{g}}{\partial \xi_j}(\xi).$$

C'est une propriété remarquable de la transformation de Fourier de convertir une dérivation en une opération algébrique (la multiplication par ξ_j). Ceci explique l'importance de la transformation de Fourier dans l'étude des équations aux dérivées partielles. L'exercice 20.3 donne un exemple de résolution d'équation aux dérivées partielles par transformation de Fourier.

Plus quantitativement, on peut aussi remarquer que la transformation de Fourier échange les propriétés de régularité et de décroissance à l'infini : par récurrence, on voit qu'une fonction très régulière (ayant beaucoup de dérivées successives dans L^1) aura une transformée de Fourier petite à l'infini (telle qu'elle reste intégrable après multiplication par un monôme de degré élevé), et vice-versa.

Démonstration. Pour démontrer la seconde partie du théorème, il suffit d'appliquer le théorème de dérivation sous le signe somme. Regardons donc la première partie.

Traitons le cas de la dimension $d = 1$. Il est faux en général qu'une fonction intégrable est de limite nulle à l'infini. C'est pourtant le cas ici, parce que f et f'

sont supposées intégrables. En effet, d'après le deuxième théorème fondamental de l'analyse réelle,

$$f(x) = f(0) + \int_0^x f'(t)\,dt \to_{x\to\pm\infty} f(0) + \int_0^{\pm\infty} f'(t)\,dt.$$

Donc $\lim_{\pm\infty} f$ existe. De plus, cette limite est forcément nulle, sans quoi f ne pourrait pas être intégrable.

Une intégration par parties montre que

$$\int_{-R}^{R} e^{-ix\xi} f'(x)\,dx = \left[e^{-ix\xi} f(x)\right]_{-R}^{R} + i\xi \int_{-R}^{R} e^{-x\xi} f(x)\,dx.$$

Grâce au fait que $\lim_{\pm\infty} f = 0$, par convergence dominée on obtient, quand R tend vers l'infini :

$$\int_{\mathbb{R}} e^{-ix\xi} f'(x)\,dx = i\xi \int_{\mathbb{R}} e^{-ix\xi} f(x)\,dx.$$

Nous laisserons au lecteur le soin d'adapter cet argument en dimension d. □

Corollaire 20.12. *Soit $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ une variable aléatoire de carré intégrable. Alors sa fonction caractéristique Φ_X est de classe C^2 et*

$$\Phi_X(\xi) = 1 - iE(X)\cdot\xi - \frac{1}{2}{}^t\xi E(X^tX)\xi + o(\|\xi\|^2).$$

Dans cette formule, ${}^tX$ désigne la matrice transposée de X. En particulier, X^tX est la matrice carré $d\times d$ de coefficients X_jX_k. On peut écrire, plus explicitement,

$$\Phi_X(\xi) = 1 - i\sum_{1\leqslant j\leqslant d} \xi_j E(X_j) - \frac{1}{2}\sum_{1\leqslant j,k\leqslant d} \xi_j\xi_k E[X_jX_k] + o(\|\xi\|^2).$$

Exercice 20.2 (Transformation de Fourier et convolution)**.**
Montrer que, si $f, g \in L^1$,

$$\widehat{f * g}(\xi) = \hat{f}(\xi)\hat{g}(\xi).$$

Remarque ★ 20.13. Nous avons mentionné dans la première définition de ce chapitre comment la transformation de Fourier se généralise aux mesures finies, ce qui revêt une importance considérable en théorie des probabilités. L'analyse harmonique moderne s'intéresse à une autre direction de généralisation, gigantesque, à savoir l'analyse harmonique sur les groupes. Cette voie a atteint un certain stade d'aboutissement uniquement pour les groupes abéliens localements compacts. On munit un tel groupe G de sa mesure de Haar, son unique mesure invariante par l'action à gauche, à homothétie près. Un *caractère* est un morphisme de G dans le cercle unité de $\mathbb{C}$. Les caractères de G forment eux-mêmes un groupe, appelé *groupe dual* $\hat{G}$ de G. Par exemple, le groupe dual de $\mathbb{R}$ est $\hat{\mathbb{R}} = \mathbb{R}$ (les caractères étant les fonctions $x \mapsto e^{itx}$, $t \in \mathbb{R}$), et celui de $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ est $\hat{\mathbb{T}} = \mathbb{Z}$ (les caractères sont les fonctions $x \mapsto e^{itx}$, $t \in \mathbb{Z}$). On définit alors la transformée de Fourier d'une fonction $f \in L^1(G)$ par

$$\hat{f}(t) = \int_G \bar{t}(g)\, f(g)\,dg.$$

On peut alors montrer que, sous certaines conditions, la formule d'inversion suivante est vraie :

$$f(g) = \int_{\hat{G}} \hat{f}(t) t(g) \, dg.$$

Ces outils opèrent un rapprochement remarquable de l'analyse et de la théorie des groupes et des représentations, et la théorie des nombres [Dei05, Gur98].

Dans l'exercice suivant, on résoud "l'équation de la chaleur"

$$\frac{\partial u}{\partial t}(t, x) = \Delta u(t, x), \tag{20.4}$$

où u est une fonction sur $\mathbb{R}^+ \times \mathbb{R}^d$, $t \in \mathbb{R}^+$ et $x \in \mathbb{R}^d$, et où $\Delta u(t, x) = u_{x_1 x_1} + \ldots + u_{x_d x_d}$ est le *laplacien* de u. Les méthodes utilisées fonctionnent remarquablement bien, en général, pour les équations aux dérivées partielles à coefficients constants sur $\mathbb{R}^d$. L'équation de la chaleur est importante à plusieurs titres :

— En physique, on montre qu'elle régit l'évolution de la température dans certains milieux. En effet, dans une petite boule B de l'espace, l'énergie thermique est

$$\int_B c \, u \, dx,$$

où u est la température et c la capacité thermique, ici supposée constante. De plus, la loi de Fourier affirme que le flux de chaleur à travers le bord ∂B de B est

$$\oint_{\partial B} \kappa \, \mathtt{grad} \, u \cdot \nu \, dS,$$

où la conductivité thermique κ est une constante, où ν est le vecteur normal sortant de ∂B et où cette intégrale est une intégrale de surface. Comme $\mathtt{div} \, \mathtt{grad} = \Delta$, la formule de Stokes affirmant que

$$\oint_{\partial B} \kappa \, \mathtt{grad} \, u \cdot \nu \, dS = \int_B \kappa \, \Delta \, u \, dx,$$

la conservation de l'énergie impose

$$\int_B (c \, \partial_t u - \kappa \, \Delta \, u) \; dx = 0.$$

Comme cette équation est vérifiée pour toute boule B, on en déduit que la fonction sous l'intégrale est nulle :

$$\partial_t u = D \, \Delta u,$$

avec $D = \frac{\kappa}{c}$ (coefficient de diffusion). Le facteur D est inessentiel est peut être supposé égal à 1 : il suffit de considérer plutôt la fonction $\tilde{u}(s, x) := u(Ds, x)$ (ce qui revient à changer d'unité de temps).

— L'équation de la chaleur peut aussi être vue, en complexifiant le temps (c'est la *rotation de Fick* des physiciens), comme l'équation de Schrödinger de la mécanique quantique, pour une particule dans le vide.

— En probabilités, l'équation de la chaleur est intimement liée au mouvement brownien, ce que nous allons voir ici de façon informelle. Considérons une particule se mouvant le long de l'axe réel de façon aléatoire, suivant les règles suivantes, où Δx et Δt sont deux paramètres > 0 donnés :
  - — pendant chaque intervalle de temps $[n\Delta t, (n+1)\Delta t]$ (avec $n \in \mathbb{N}$), la particule se déplace aléatoirement vers la gauche ou vers la droite ;
  - — les accroissements successifs $x_{n+1} - x_n$ entre les instants $n\Delta t$ et $(n+1)\Delta t$ sont des variables aléatoires $\Omega \to \{\pm\Delta x\}$ indépendantes (voir la définition de l'indépendance dans le chapitre qui lui sera bientôt consacré) et identiquement distribuées, de loi de Bernoulli $\frac{1}{2}(\delta_{-\Delta x} + \delta_{\Delta x})$.

  Alors la probabilité $p(t,x)$ de trouver à l'instant t la particule à l'abscisse x vérifie, d'après la formule des probabilités totales,

$$p(t+\Delta t, x) = \frac{1}{2}\left(p(t, x-\Delta x) + p(t, x+\Delta x)\right). \tag{20.5}$$

  Un tel système dynamique aléatoire s'appelle une *marche aléatoire* (symétrique, ici, puisqu'à chaque étape la probabilité d'aller à gauche ou à droite est la même).

  On s'intéresse à la limite d'une telle marche aléatoire aléatoire quand Δt et Δx tendent chacun vers 0. Il est alors commode de supposer que $p(t,x)$ est une fonction définie sur $\mathbb{R}^+ \times \mathbb{R}$ (et non plus seulement en des valeurs discrètes de t et de x) et que la formule (20.5) est encore valable. Alors, en écrivant

$$\begin{cases} p(t+\Delta t, x) = p(t,x) + \frac{\partial p}{\partial t}(t,x)\Delta t + o(\Delta t) \\ p(t, x \pm \Delta x) = p(t,x) \pm \frac{\partial p}{\partial x}(t,x)\Delta x + \frac{1}{2}\frac{\partial^2 p}{\partial x^2}(t,x)\Delta x^2 + o(\Delta x^2), \end{cases}$$

  après division par Δt l'équation devient

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}\frac{\Delta x^2}{\Delta t} + o\left(\frac{\Delta x^2}{\Delta t}\right) + o(1).$$

  L'équation a une limite non triviale si Δt et Δx tendent vers 0 de façon que le rapport

$$\frac{\Delta x^2}{\Delta t} = 2D$$

  soit constant (le facteur 2 étant là pour une pure raison esthétique) ; D s'appelle le *coefficient de diffusion.* Alors l'équation limite est l'équation de diffusion de la chaleur

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}.$$

  Le comportement qualitatif de ses solutions ne dépend pas de la valeur précise de D : on peut toujours se ramener à $D = 1$ en faisant le changement d'inconnue consistant à poser $u(t,x) = p(D.t, x)$ (ce qui revient à changer d'unité de temps).

  Mais, dans cette limite, qu'est devenu le mouvement de la particule initialement considérée ? Les accroissements normalisés $\xi_j = \frac{x_{j+1}-x_j}{\Delta x}$ sont indépendants, identiquement distribués, d'espérance nulle et de variance 1. Le théorème central limite (voir plus loin le chapitre qui lui est consacré) affirme que la position

$$x_n = \Delta x.(\xi_0 + \xi_1 + \cdots + \xi_{n-1})$$

de la particule à l'instant $n\Delta t$, à supposer que l'on choisisse $\Delta x = \sqrt{2tDn}$, converge quand $n \to +\infty$ en loi vers une variable aléatoire réelle $B(t)$ de loi normale $N(0, 2Dt)$. La position limite se note $B(t, \omega)$, de sorte que
— pour tout t, la fonction $\omega \mapsto B(t, \omega)$ est une variable aléatoire,
— pour tout ω, la fonction $t \mapsto B(t, \omega)$ est une courbe dans l'espace des positions (ici $\mathbb{R}$), définissant la trajectoire de la particule.

On montre que les trajectoires $t \mapsto B(t, \omega)$ sont des courbes continues nulle part dérivables. B s'appelle un *mouvement brownien*[2] ou un *processus de Wiener.*[3]

Voyons comment trouver heuristiquement (c'est-à-dire en raisonnant formellement, en faisant toutes les hypothèses supplémentaires de régularité et d'intégrabilité nécessaires pour intégrer, dériver sous le signe d'intégration, etc.) une solution de l'équation (20.4) avec la transformation de Fourier. Supposons qu'il existe une solution de cette équation, vérifiant la condition unitiale $u(0, x) = \varphi(x)$ (une certaine fonction de x donnée). Prenons la transformée de Fourier de l'équation de la chaleur par rapport à la variable spatiale x, le paramètre t étant fixé :

$$\int_{\mathbb{R}^d} \partial_t u e^{-ix\cdot\xi}\, dx = \int_{\mathbb{R}^d} \Delta u e^{-ix\cdot\xi}\, dx.$$

En supposant que le théorème 20.11 s'applique, on obtient

$$\partial_t \hat{u}(t, \xi) = -\|\xi\|^2 \hat{u}(t, \xi),$$

où $\hat{u}$ est la transformée de Fourier de u par rapport à la variable spatiale x. On a ainsi converti l'équation de la chaleur en une équation différentielle ordinaire (puisque seule la dérivée de $\hat{u}$ par rapport au temps intervient), la variable spatiale duale ξ pouvant être prise comme paramètre. Comme de plus

$$\hat{u}(0, \xi) = \hat{\varphi}(\xi),$$

on obtient

$$\hat{u}(t, \xi) = \varphi(\xi)\, e^{-\|\xi\|^2 t}$$

soit, d'après la formule démontrée dans l'exercice 20.2 (à supposer toujours que celle-ci s'applique),

$$u(t, x) = \varphi *_x \frac{1}{(2\pi)^{d/2}} \bar{\mathcal{F}}\left(e^{-\|\xi\|^2 t}\right),$$

où "$*_x$" désigne le produit de convolution par rapport à x, soit

$$u(t, x) = \varphi(x) *_x \mathcal{N}(t, x)$$

2. D'après Robert Brown, paléobotaniste écossais (1773–1858). Il étudia au microsope le mouvement de graines de pollen de *Clarkia pulchella* immergés dans l'eau, ce que l'on modélise aujourd'hui par un mouvement brownien.

3. Norbert Wiener, mathématicien américain (1894–1964).

avec

$$\mathcal{N}(t,x) = \frac{1}{(4\pi t)^{d/2}} e^{-\|x\|^2/(4t)}. \tag{20.6}$$

L'exercice suivant justifiera le degré de validité de cette formule.

Exercice 20.3 (Équation de la chaleur)**.**
Considérons l'*équation de la chaleur*

$$\frac{\partial u}{\partial t}(t,x) = \Delta u(t,x),$$

où u est une fonction sur $\mathbb{R}^+ \times \mathbb{R}^d$, $t \in \mathbb{R}^+$ et $x \in \mathbb{R}^d$. Imposons de plus la distribution initiale $u(0,\cdot) = \varphi \in L^1(\mathbb{R}^d)$. On montre en Physique que la température d'un espace homogène isotrope sans source de chaleur définit une solution de ce problème de Cauchy.

Soient $\mathcal{N} : \mathbb{R}^+_* \times \mathbb{R}^d \to \mathbb{R}$ le *noyau de la chaleur* défini par l'équation (20.6) et $u : \mathbb{R}^+_* \times \mathbb{R}^d \to \mathbb{R}$ la fonction définie par

$$u(t,x) = \left(\mathcal{N}(t,\cdot) * \varphi\right)(x).$$

1. u est une fonction de classe C^∞ sur $\mathbb{R}^+_* \times \mathbb{R}^d$.

2. u est une solution de l'équation de la chaleur sur $\mathbb{R}^+_* \times \mathbb{R}^d$.

3. $u(t,\cdot)$ converge vers φ dans $L^1(\mathbb{R}^d)$ quand t tend vers 0.

Pour tout $t > 0$ on note symboliquement $e^{t\Delta}$ le *propagateur de la chaleur*

$$e^{t\Delta} : L^1(\mathbb{R}^d) \to \mathrm{C}^\infty(\mathbb{R}^d), \quad \varphi \mapsto u(t,\cdot) = \mathcal{N}(t,\cdot) * \varphi.$$

4. Montrer que $e^{t\Delta}$ est un opérateur continu positif $L^1(\mathbb{R}^d) \to L^1(\mathbb{R}^d)$.

5. Montrer que $e^{t\Delta}$ se prolonge de façon unique en un endomorphisme continu $L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$; on pourra utiliser sans le démontrer que la transformation de Fourier se prolonge en une isométrie de $L^2(\mathbb{R}^d)$.

Montrer que quelle que soit la fonction $\varphi \in L^2$, que la fonction $e^{t\Delta}\varphi$ satisfait l'équation de la chaleur sur $\mathbb{R}^+_* \times \mathbb{R}^d$ et $u(t,\cdot)$ converge vers φ dans L^2 quand t tend vers 0.

Remarque 20.14. La solution de l'équation de la chaleur sur $\mathbb{R}^+_* \times \mathbb{R}^d$ construite dans l'exercice précédent grâce à la transformation de Fourier peut être elle-même utilisée pour résoudre l'équation sur un domaine spatial borné de la forme

$$\Omega = \{x \in \mathbb{R}^d,\ 0 \leqslant x_j \leqslant a_j,\ 0 \leqslant j \leqslant n\}.$$

La méthode consiste à se ramener l'espace $\mathbb{R}^d$ entier par réflexions dénombrables du domaine.

Exercice 20.4 (Non surjectivité de la transformation de Fourier)**.**
Soit

$$\theta(\xi) = \int_1^\xi \frac{e^{i\eta}}{\eta}\, d\eta, \quad \xi > 0.$$

1. Montrer que la fonction θ est continue sur $]0,+\infty[$ et qu'elle possède une limite finie en $+\infty$.

2. Montrer que, si f est une fonction intégrable sur $\mathbb{R}$ et si $\widehat{f}$ est sa transformée de Fourier, l'expression

$$\phi(\xi) = \int_1^\xi \frac{\widehat{f}(\eta)}{\eta}\, d\eta$$

a une limite finie quand ξ tend vers $+\infty$.

Soit g la fonction paire définie sur $\mathbb{R}$ par

$$g(x) = \begin{cases} 1 & \text{si } |x| \leqslant 1 \\ 1/(1+\ln(1+|x|)) & \text{si } |x| \geqslant 1. \end{cases}$$

3. Montrer que g est continue et tend vers 0 en $\pm\infty$, mais que g n'est la transformée de Fourier d'aucune fonction intégrable.

Exercice 20.5 (Rotations minimales)**.**
Soient $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ et

$$R_\alpha : \mathbb{T}^d \to \mathbb{T}^d, \quad x \mapsto x + \alpha.$$

On identifie $\mathbb{T}^d$ à $[0,1[^d$, et on le munit de la restriction de la mesure de Lebesgue.

Montrer que les deux propriétés suivantes sont équivalentes :

— les nombres $(1, \alpha_1, ..., \alpha_d)$ sont libres dans le $\mathbb{Q}$-espace vectoriel $\mathbb{R}$ (on dit que R_α est minimale)

— l'unique mesure de probabilité invariante par R_α est la mesure de Lebesgue.

Problème 20.1 (Une inégalité de concentration [BJ06])**.**
Soient X une variable aléatoire réelle, P_X sa loi, $\Phi_X : t \mapsto E(e^{-itX})$ sa fonction caractéristique et

$$C_X : [0,+\infty[\to [0,1], \quad y \mapsto \sup_{x\in\mathbb{R}} P_X([x, x+y])$$

sa *fonction de concentration*. On suppose que P_X est diffuse, c'est-à-dire que $P_X(a) = 0$ pour tout $a \in \mathbb{R}$.

1. Tracer le graphe de la fonction C_X dans le cas d'une variable de loi uniforme sur un intervalle $[a-h, a+h]$ ($a \in \mathbb{R}$, $h \in \mathbb{R}_*^+$).

On revient au cas général.

2. Montrer que C_X est croissante.

3. Montrer que C_X est continue à droite.

4. Soient $T > 0$ et $a \in \mathbb{R}$. Montrer :

$$E\left[2\frac{1-\cos T(X+a)}{T^2(X+a)^2}\right] = \frac{1}{T^2}\int_{-T}^{T} (T-|t|)\, e^{-ita}\phi_X(t)\, dt.$$

5. ? ? ? En déduire qu'il existe un réel A universel (indépendant de X) tel que, si $\lambda > 0$ et $0 \leqslant x\lambda \leqslant \pi$,

$$C_X(x) \leqslant \frac{A}{\lambda}\int_0^\lambda |\Phi_X(t)|\, dt.$$

Soient $\tilde{X}$ une copie indépendante de X, $b > 0$ et Z indépendante de $(X, \tilde{X})$ et de densité

6. Montrer que

7. En déduire :

$$C_X(y) \geqslant \frac{1}{2\pi^2} \frac{y}{1+by} \int_{-b/2}^{b/2} |\Phi_X(t)|^2 \, dt.$$

Exercice 20.6 (Inégalité d'Heisenberg (R. Krikorian))**.**
Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction de classe C^2 telle que $xf(x)$ et $f'(x)$ soient dans L^2. Montrer que

$$\left(\int_{\mathbb{R}} x^2 f(x)^2 \, dx \right) \left(\int_{\mathbb{R}} f'(x)^2 \, dx \right) \geqslant \frac{1}{4};$$

quels sont les cas d'égalité ?

Indication : On pourra considérer les opérateurs A_t et A_t^* tels que $A_t f(x) = f'(x) + txf(x)$ et $A_t^* f(x) = -f'(x) + txf(x)$, et montrer que $\int_{\mathbb{R}} (A_t^* A_t f)(x) \, dx \geqslant 0$.

(L'exercice 22.4 donne une version duale de cette inégalité, au sens de la transformation de Fourier.)

Exercice 20.7 (Inversion de la transformation de Fourier)**.**
Soit μ une probabilité sur $\mathbb{R}^d$ telle que $\hat{\mu} \in L^1(\mathbb{R}^d)$. Montrer que μ admet une densité par rapport à la mesure de Lebesgue et que cette densité vaut

$$f_\mu(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{\mu}(\xi) \, e^{i x \cdot \xi} \, d\xi.$$

Figure 28 – Joseph Fourier (1768–1830)

Chapitre 21

L'espace de Hilbert $L^2(\mu)$

On a démontré dans un chapitre précédent que $L^1(E, \mathcal{A}, \mu)$ est un espace de Banach (théorème 18.4) et que la démonstration de la complétude de L^1 se transpose *mutatis mutandis* à tous les espaces L^p. L'espace L^2 jouit de la propriété supplémentaire qu'il est hilbertien, puisque sa norme est manifestement issue du produit scalaire hilbertien (disons complexe, si l'on considère des fonctions complexes, mais on pourrait tout aussi bien se limiter aux fonctions réelles)

$$\langle f|g\rangle = \int \overline{f(x)} g(x)\, d\mu(x).$$

Rappelons les propriétés qui font de cette application $\langle\cdot|\cdot\rangle$ un *produit scalaire préhilbertien* :[1]

— sesquilinéarité :

$$\langle af|g\rangle = \bar{a}\langle f|g\rangle \quad \text{et} \quad \langle f|ag\rangle = a\langle f, g\rangle$$

— symétrie hermitienne :

$$\langle g|f\rangle = \overline{\langle f|g\rangle}$$

— positivité :

$$\langle f|f\rangle \geqslant 0$$

— caractère défini :

$$\|f\|_{L^2} := \langle f|f\rangle = 0 \Rightarrow f = 0.$$

On a donc démontré le résultat suivant.

Théorème 21.1. $L^2(E, \mathcal{A}, \mu)$ *est un espace de Hilbert.*

Ainsi, L^2 bénéficie du lot de propriétés très fortes qui accompagne tous les espaces de Hilbert. Par exemple, l'inégalité de Cauchy-Schwarz dit que

$$|\langle f|g\rangle| \leqslant \|f\|_{L^2}\|g\|_{L^2},$$

1. David Hilbert, mathématicien allemand (1862–1943), dont l'influence fut considérable. Il définit une liste de 23 problèmes fondamentaux, qui a orienté une grande partie des recherches mathématiques au XXe siècle.

soit
$$\left|\int \bar{f} g \, d\mu\right| \leqslant \left(\int |f|^2 \, d\mu\right)^{1/2} \left(\int |g|^2 \, d\mu\right)^{1/2}.$$
Voyons trois applications simples de l'inégalité de Cauchy-Schwarz.

Exemple 21.2. Le produit scalaire $\langle\cdot|\cdot\rangle : L^2 \times L^2 \to \mathbb{C}$ est continu.

Exemple 21.3. Si μ est finie, $L^2 \subset L^1$. En effet, si $f \in L^2$,
$$\|f\|_{L^1} = \int |f| \, d\mu \leqslant \|f\|_{L^2} \|1\|_{L^2} = \|f\|_{L^2} \sqrt{\mu(E)} < \infty.$$

Exemple 21.4. La formule du produit de convolution définit un produit des fonctions dans $L^2(\mathbb{R}^d)$, à valeurs dans l'espace des fonctions continues bornées sur $\mathbb{R}^d$, puisque
$$|f * g(x)| \leqslant \int |f(y)g(x-y)| \, dy \leqslant \|f\|_{L^2} \|g(x-\cdot)\|_{L^2} = \|f\|_{L^2} \|g\|_{L^2}.$$

Nous rappellerons les propriétés de réflexivité, existence et unicité de la projection sur un convexe fermé non vide, existence de base hilbertiennes etc., au fur et à mesure des besoins.

Fin de la démonstration du théorème 18.11. L'argument suivant est dû à von Neumann.[2] Soit $\lambda \in L^{1\prime}$. On veut montrer que, sous l'hypothèse que μ soit σ-finie, il existe une fonction $g \in L^\infty$ telle que, pour toute fonction $f \in L^1$, on ait
$$\lambda(f) = \int fg \, d\mu.$$
Notons b la norme de λ :
$$|\lambda(f)| \leqslant b\|f\|_{L^1} \quad (\forall f \in L^1).$$
Supposons d'abord que μ soit finie. Alors l'inégalité de Cauchy-Schwarz :
$$|\langle f|g\rangle| \leqslant \|f\|_{L^2} \|g\|_{L^2}$$
appliquée avec $g = 1$:
$$\int |f| \leqslant \left(\int |f|^2\right)^{1/2} \mu(E)^{1/2}$$
implique que $L^2 \subset L^1$, donc $L^{1\prime} \subset L^{2\prime}$. Donc λ peut être vue comme une forme linéaire sur L^2, par simple restriction. Cette restriction est continue, comme le montre encore l'inégalité de Cauchy-Schwarz :
$$|\lambda(f) \leqslant b\|f\|_{L^1} = b \int |f| \cdot \mathbb{1}_E d\mu \leqslant \|f\|_{L^2} \|\mathbb{1}_E\|_{L^2}.$$

2. John von Neumann, mathématicien américano-hongrois (1903–1957), qui s'est illustré dans des domaines aussi variés que l'analyse fonctionnelle, la théorie des ensembles, la mécanique quantique, l'informatique et l'économie. En informatique, il a imaginé l'architecture (qui porte aujourd'hui son nom) utilisée dans tous les ordinateurs modernes.

Dans le cas de L^2, la surjectivité nous est donnée gratuitement par la théorie hilbertienne : d'après le théorème de Riesz,* il existe une unique fonction $g \in L^2$ telle que

$$\lambda(f) = \int fg\,d\mu$$

pour toute $f \in L^2$. Remarquons que, pour tout ensemble A de mesure finie, $\mathbb{1}_A \in L^1 \cap L^2$ et

$$\left|\int \mathbb{1}_A\, g\, d\mu\right| = |\lambda(\mathbb{1}_A)| \leqslant b\|\mathbb{1}_A\|_{L^1} \leqslant b\mu(A).$$

Donc

$$|g(x)| \leqslant b\,\text{p.p.}$$

(exercice : justifier cette affirmation), donc $g \in L^\infty$ et $\|g\|_{L^\infty} \leqslant b$.

Comme l'espace des fonctions étagées intégrables est dense dans L^1 (exercice 18.3), par continuité

$$\lambda(f) = \int fg\,d\mu$$

pour toute $f \in L^1$. Ceci démontre la surjectivité dans le cas où $\mu(E) < \infty$.

Le cas général où μ est seulement σ-finie s'en déduit. □

Autres exercices

Exercice 21.1 (Séries de Fourier L^2)**.**
Les séries de Fourier sont un cas particulier de la transformée de Fourier, mais on peut les étudier indpendemment. Ici, nous nous attachons à la théorie L^2.

Considérons l'intervalle $C = [0, 1[$ muni de la tribu borélienne et de la mesure de Lebesgue. On considère l'espace de Hilbert $L^2(C)$ des fonctions complexes, muni du produit scalaire complexe (produit hermitien) défini par

$$\langle\phi|\psi\rangle = \int_0^1 \phi(t)\bar{\psi}(t)\,dt.$$

Un tel produit hermitien définit naturellement une norme par la formule :

$$\|\phi\|_{L^2} = \sqrt{\langle\phi|\phi\rangle}.$$

On vérifiera ou on admettra que les résultats démontrés dans le Cours pour les espaces de Hilbert réels se transposent aux espaces de Hilbert complexes.

Soit $(e_n)_{n\in\mathbb{Z}}$ la famille de fonctions définie par $e_n(t) = e^{in2\pi t}$.

1. Montrer que la famille $(e_n)_{n\in\mathbb{Z}}$ est orthonormée.

Notons $C_1^\infty(\mathbb{R})$ l'ensemble des fonctions 1-périodiques et de classe C^∞ sur $\mathbb{R}$. Soit $\phi \in C_1^\infty(\mathbb{R})$; par restriction à C, on peut voir ϕ comme une fonction sur C. Pour tout $n \in \mathbb{Z}$ on pose

$$c_n(\phi) = \langle\phi|e_n\rangle = \int_C \phi(t)\bar{e}_n(t)\,dt,$$

puis

$$S_N(\phi)(t) = \sum_{n=-N}^{N} c_n(\phi)e_n(t).$$

2. Montrer en faisant deux intégrations par parties que la série de somme partielle $S_N(\phi)(t)$ converge.

3. Montrer que

$$S_N(\phi)(t) = \int_0^1 \phi(\theta)\frac{\sin\left(\pi(2N+1)(t-\theta)\right)}{\sin \pi(t-\theta)}\, d\theta.$$

4. En déduire que S_N converge vers ϕ uniformément sur C.

5. En déduire que si ϕ appartient à l'orthogonal de $(e_n)_{n\in\mathbb{Z}}$ alors $\phi = 0$. En admettant la densité de $C_1^\infty(\mathbb{R})$ dans $L^2(C)$, montrer que $(e_n)_{n\in\mathbb{Z}}$ est une base hilbertienne de $L^2(C)$.

Soient $\alpha \in \mathbb{R}$ et $g \in L^2(C)$. Considérons l'équation

$$f(t+\alpha \pmod 1) - f(t) = g(t)$$

d'inconnue $f \in L^2(C)$.

6. Montrer que, si g n'est pas de moyenne nulle, l'équation n'a pas de solution.

7. Montrer que, si α est rationnel, en général l'équation n'a pas de solution.

8. Donner un exemple avec $\alpha \notin \mathbb{Q}$ où l'équation admet une solution non constante.

9. ★ Donner un exemple où l'équation admet une solution avec $\alpha \notin \mathbb{Q}$ et $g(t) = \sum_{n\in\mathbb{Z}} e_n(t)/2^n$.

Comme seconde application des séries de Fourier, on se propose de retrouver la formule classique suivante :

$$\sum_{n\geqslant 1} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

10. Calculer les coefficients de Fourier de la fonction ϕ qui est impaire et 1-périodique, qui vaut 1 sur $[0, 1/2[$.

11. Exprimer $\sum_{n\geqslant 1} \frac{1}{n^2}$ en fonction de $\sum_{n\geqslant 0} \frac{1}{2n+1}$.

12. En déduire le résultat cherché, en utilisant l'égalité de Parseval.

Problème 21.1 (Séries de Fourier bi-dimensionnelles)**.**
Soit $f : \mathbb{R}^2 \to \mathbb{R}$ une fonction de classe C^3, qui soit 1-périodique par rapport à chacun de ses deux arguments. On définit, pour tout $k = (k_1, k_2) \in \mathbb{Z}^2$, le k-ième *coefficient de Fourier* de f,

$$\hat{f}_k = \int_0^1 \int_0^1 f(\theta)\, e^{-i\,2\pi\, k\cdot\theta}\, d\theta_1\, d\theta_2 \in \mathbb{C}$$

et, pour tout $N \in \mathbb{N}$ et tout $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$, la *série de Fourier*

$$S_N[f](\theta) = \sum_{k\in K_N} \hat{f}_k\, e^{i\,2\pi\, k\cdot\theta},$$

où $K_N = \{k \in \mathbb{Z}^2,\ |k_1| \leqslant N,\ |k_2| \leqslant N\}$ et $k \cdot \theta = k_1\theta_1 + k_2\theta_2$.
L'objectif de cette partie est de montrer que, pour tout $\epsilon > 0$, il existe $N \in \mathbb{N}$ tel que

$$\max_{\theta \in \mathbb{R}^2} |f(\theta) - S_N[f](\theta)| \leqslant \epsilon,$$

ce que nous utiliserons dans la partie **??**.

1. La fonction

$$D_N : \mathbb{R}^2 \to \mathbb{C}, \quad \theta \mapsto \sum_{k \in K_N} e^{i\,2\pi\,k\cdot\theta}$$

est-elle réelle ?

2. Montrer que

$$S_N[f](\theta) = \int_0^1 \int_0^1 f(t)\, D_N(\theta - t)\, dt_1\, dt_2.$$

3. En déduire que, pour tout $\theta \in \mathbb{R}^2$,

$$f(\theta) - S_N[f](\theta) = \int_0^1 \int_0^1 \varphi[f](\theta, t) \left(\prod_{j=1,2} \sin\left[(2N+1)\pi(\theta_j - t_j)\right] \right) dt_1\, dt_2, \quad (21.1)$$

avec

$$\varphi[f](\theta, t) = \frac{f(\theta) - f(t)}{\sin\left[\pi(\theta_1 - t_1)\right] \sin\left[\pi(\theta_2 - t_2)\right]};$$

on pourra commencer par écrire que $f(\theta) = \int_0^1 \int_0^1 f(\theta)\, D_N(\theta - t)\, dt_1\, dt_2$.

4. Dans le cas où il existe une fonction $g : \mathbb{R} \to \mathbb{R}$ telle que $f(\theta) = g(\theta_1)$, déduire de la formule (21.1) que

$$f(\theta_1, 0) - S_N[f](\theta_1, 0) = \int_0^1 \psi[g](\theta_1, t_1)\ \sin\left[(2N+1)\pi(\theta_1 - t_1)\right]\, dt_1$$

avec

$$\psi[g] : (\theta_1, t_1) \mapsto \frac{g(\theta_1) - g(t_1)}{\sin\left[\pi(\theta_1 - t_1)\right]},$$

montrer que $\psi[g]$ est de classe C^1 et en déduire que la série $(S_N[f])_{N \in \mathbb{N}}$ converge vers f uniformément.

5. Montrer que, dans le cas général, si l'on pose, pour $\theta^o = (\theta_1^o, \theta_2^o) \in \mathbb{R}$,

$$h_{\theta^o}(\theta) = f(\theta^o + \theta) - f(\theta_1^o + \theta_1, \theta_2^o) - f(\theta_1^o, \theta_2^o + \theta_2) + f(\theta^o),$$

la fonction $\varphi[h_{\theta^o}](0, \cdot)$ est de classe C^1, et que $S_N[h_{\theta^o}](0)$ converge vers 0 uniformément par rapport à θ^o. Conclure.

Exercice 21.2 (Convolution dans L^2)**.**
Soient $f, g \in L^2(\mathbb{R})$.

1. Montrer, en utilisant l'inégalité de Cauchy-Schwarz, que la fonction $f * g$ est partout définie et bornée, en valeur absolue, par $\|f\|_{L^2}\, \|g\|_{L^2}$.

2. Montrer que $f * g$ est continue et tend vers 0 à l'infini.

3. $f * g$ est-elle toujours dans L^2 ?

Problème ★ 21.2 (Universalité de la convolution)**.**
1. Montrer que la formule du produit de convolution définit un produit de $L^2(\mathbb{R}^d) \times L^2(\mathbb{R}^d)$ dans l'espace $C_b(\mathbb{R}^d)$ des fonctions continues bornées.

2. Soit A un opérateur continu de $L^2(\mathbb{R}^d)$ dans $C_b(\mathbb{R}^d)$, continu (c'est-à-dire qu'il existe $C > 0$ tel que, pour toute $f \in L^2(\mathbb{R}^d)$, $\|Af\|_{L^\infty} \leqslant C\|f\|_{L^2}$) et commutant aux translations (c'est-à-dire tel que $A(f(\cdot - \tau)) = (Af)(\cdot - \tau)$ pour toute $f \in L^2(\mathbb{R}^d)$ et tout $\tau \in \mathbb{R}^d$). Montrer qu'il existe $a \in L^2(\mathbb{R}^d)$ tel que, pour toute $f \in L^2(\mathbb{R}^d)$, $Af = a * f$; on pourra utiliser le théorème de Riesz.

Dans le résultat précédent, on aimerait pouvoir supprimer les hypothèses restrictives sur A. Mais l'exemple suivant montre qu'on ne peut pas le faire, du moins sans utiliser la théorie des distributions.

3. Montrer qu'il n'existe pas de fonction $a \in L^1(\mathbb{R})$ telle que, pour toute fonction $f \in L^1(\mathbb{R})$, $f = a * f$ (ce qui correspond au cas où A est l'identité de $L^1(\mathbb{R})$).

Problème ★ 21.3 (De la formule de Taylor aux fonctions d'Hermite, *examen 2013*)**.**
Soient $(\Omega, \mathcal{A}, P)$ un espace de probabilité et $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle. Pour tout $\xi \in \mathbb{R}$, on note $\epsilon_\xi : \mathbb{R} \to \mathbb{R}$ la fonction $x \mapsto e^{x\xi}$ et $\epsilon_\xi^{(n)}$ sa dérivée n-ième.

1. Supposons qu'il existe une suite $(p_n)_{n\in\mathbb{N}}$ de polynômes d'une variable telle que, pour tout $\xi \in \mathbb{R}$, on a

$$\epsilon_\xi(x) = \sum_{n\in\mathbb{N}} E(\epsilon_\xi^{(n)}(X))\,\frac{p_n(x)}{n!}.$$

On cherche ici à montrer que, forcément, les polynômes p_n satisfont une certaine relation de récurrence et certaines propriétés qui les déterminent. Dans cette question, on raisonnera donc par condition suffisante, en supposant notamment que l'on peut, dans les calculs, intervertir le signe de sommation avec les opérations de dérivation et d'intégration.

1.1. Calculer les dérivées $\epsilon_\xi^{(n)}$ pour $n \in \mathbb{N}$, montrer que

$$\frac{e^{x\xi}}{E(e^{X\xi})} = \sum_{n\geqslant 0} \frac{p_n(x)}{n!}\xi^n, \tag{21.2}$$

puis en déduire que les polynômes p_n sont totalement déterminés par les relations suivantes :

$$p_0(x) = 1, \quad p_n' = np_{n-1}, \quad \int p_n \, dP_X = 0 \quad (\forall n \geqslant 1). \tag{21.3}$$

1.2. Calculer les p_n dans le cas où $P_X = \delta_a$ (mesure de Dirac) avec $a \in \mathbb{R}$, ainsi que p_n pour $n = 0, 1, 2$ dans le cas où P_X est la loi de Bernoulli $\frac{1}{2}(\delta_0 + \delta_1)$.

2. On veut montrer, la "formule de Taylor" suivante, pour toute fonction f réelle de classe C^∞ sur $\mathbb{R}$ telle que, pour tout $n \in \mathbb{N}$, $f^{(n)}(X)$ soit intégrable, et pour tout $n \in \mathbb{N}$:

$$f(x) = \sum_{0\leqslant k\leqslant n} E(f^{(k)}(X))\frac{p_n(x)}{k!} + E\left(\int_0^{X-x} \frac{p_n(x+t)}{n!} f^{(n+1)}(X-t)\,dt\right).$$

2.1. Montrer que, pour toute fonction f réelle de classe C^∞ et pour tout $n \in \mathbb{N}$, on a

$$\sum_{0\leqslant k\leqslant n}\left(\frac{p_k(x)}{k!}f^{(k)}(y)-\frac{p_k(y)}{k!}f^{(k)}(x)\right)=\int_0^{y-x}\frac{p_n(x+t)}{n!}f^{(n+1)}(y-t)\,dt.$$

2.2. Conclure ; on pourra intégrer la formule précédente par rapport à y.

2.3. Dans le cas où P_X est la loi de Bernoulli $\frac{1}{2}(\delta_0+\delta_1)$, exprimer la formule obtenue en termes des *polynômes d'Euler* E_n définis par l'égalité

$$\frac{2e^{x\xi}}{e^\xi+1}=\sum_{n\geqslant 1}\frac{E_n(x)}{n!}\xi^n.$$

Calculer E_n pour $n=0,1,2$.

On se spécialise dorénavant au cas où X est de loi gaussienne $N(0,1/2)$, c'est-à-dire de densité $e^{-x^2}/\sqrt{\pi}$. Les polynômes p_n (définis par (21.2) ou (21.3)) s'appellent alors les *polynômes d'Hermite*[3] et se notent H_n.

3.

3.1. Montrer que, pour tout $\xi \in \mathbb{R}$, la variable aléatoire $e^{X\xi}$ est intégrable, et que la fonction $h : \xi \mapsto E(e^{X\xi})$ est dérivable sur $\mathbb{R}$, calculer sa dérivée, et en déduire que $h(\xi)=e^{\xi^2/4}$ (on pourra trouver une équation différentielle vérifiée par h, et la résoudre).

3.2. En déduire que

$$H_n(x)=(-1)^n2^{-n}e^{x^2}\frac{d^n}{dx^n}\left(e^{-x^2}\right),$$

et calculer H_n pour $n=0$, 1 et 2.

Pour $n \geqslant 0$, notons ψ_n la *n-ième fonction d'Hermite* définies sur $\mathbb{R}$ par

$$\psi_n(x)=e^{-x^2/2}H_n(x).$$

4.

4.1. Montrer que $(\psi_n|\psi_m)_{L^2}=\int_{\mathbb{R}}\psi_n(x)\,\psi_m(x)\,dx=0$ si $n\neq m$ et que $(\psi_n|\psi_n)_{L^2}\neq 0$; on pourra faire des intégrations par parties et utiliser la récurrence de (21.3).

4.2. ★ Soit $\varphi \in L^2$ une fonction L^2-orthogonale à chaque ψ_n et telle que $e^{x^2/2}\varphi(x) \in L^1(\mathbb{R})$. Montrer que $\varphi=0$; on pourra remarquer que, pour tout $\xi \in \mathbb{R}$,

$$0=\sum_{n\geqslant 0}\frac{\xi^n}{n!}\int e^{-x^2/2}H_n(x)\varphi(x)\,dx,$$

montrer que le membre de droite de cette égalité s'exprime comme la convolée de e^{-x^2} et de $e^{x^2/2}\varphi(x)$, puis prendre la transformée de Fourier.

3. Charles Hermite, mathématicien français (1822–1901), analyste qui avait en horreur les fonctions continues nulle part différentiables !

On admettra que (ψ_n) est une base hilbertienne de L^2 (ce que l'on peut voir en utilisant la question précédente).

5. On veut maintenant prolonger la transformation de Fourier à $L^2(\mathbb{R})$, par un argument du à Wiener [Wie66].

5.1. Montrer que $\psi_{n+1}(x) = \frac{1}{2}(x\psi_n(x) - \psi_n'(x))$ pour tout $n \in \mathbb{N}$.

5.2. En déduire que $\hat{\psi}_{n+1}(\xi) = -\frac{i}{2}\left(\xi\hat{\psi}_n(\xi) - \hat{\psi}_n'(\xi)\right)$, puis que $\hat{\psi}_n = \sqrt{2\pi}(-i)^n\psi_n$.

5.3. Montrer qu'il existe une unique application linéaire continue $\mathcal{F}_2 : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ telle que $\mathcal{F}_2(\psi_n) = \sqrt{2\pi}(-i)^n\psi_n$; cette application est la *transformation de Fourier–Plancherel.*[4]

5.4. Montrer que $\mathcal{F}_2$ est une isométrie de $L^2(\mathbb{R})$ et que $\mathcal{F}_2 \circ \mathcal{F}_2 = s$, où $(s(\varphi))(x) = \varphi(-x)$ (en particulier, $\mathcal{F}_2^4$ est l'identité de $L^2(\mathbb{R})$).

5.5. En déduire que L^2 est la somme directe de quatre sous-espaces invariants par $\mathcal{F}_2$.

Problème 21.4 (Transformations linéaires de $L^2(\mathbb{R})$)**.**
Cet exercice utilise la transformation de Fourier-Plancherel, définie dans l'exercice 21.3.

Soit $T : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ une application linéaire continue telle que, pour tout $\alpha \in \mathbb{R}$ et toute $f \in L^2(\mathbb{R})$, $T(\tau_\alpha f) = \tau_\alpha T(f)$, où $(\tau_\alpha f)(x) = f(x - \alpha)$ (translation de α du graphe de f). Montrer qu'il existe une fonction $\varphi \in L^\infty(\mathbb{R})$ telle que, pour toute $f \in L^2(\mathbb{R})$,

$$\widehat{Tf} = \varphi.\widehat{f}.$$

4. Michel Plancherel, mathématicien suisse (1885–1967)

FIGURE 29 – David HILBERT (1862–1943) et John von NEUMANN (1903–1957), deux prodiges dont le génie s'est exprimé dans tous les domaines des mathématiques pour le premier, et pour le second des mathématiques les plus théoriques à l'informatique, la mécanique quantique et l'économie.

Chapitre 22

Covariance

Soient $(\Omega, \mathcal{A}, P)$ un espace de probabilité et $X \in L^2(\Omega, \mathcal{A}, P)$.
L'inégalité de Cauchy-Schwarz ($E(|XY|) \leqslant E(X^2)^{1/2}E(Y^2)^{1/2}$) montre, dans le cas particulier où $Y = 1$ que

$$E(|X|) \leqslant E(X^2)^{1/2},$$

donc que $L^1 \subset L^2$ (ce qui n'est pas le cas en général dans un espace mesuré dont la mesure est infinie). Donc X et X^2 sont intégrables :

$$E(X^2) \leqslant \infty \quad \text{et} \quad E(|X|) \leqslant \infty.$$

Définition 22.1. La *variance* de X est

$$\mathrm{var}(X) := E\left((X - E(X))^2\right) = E(X^2) - E(X)^2.$$

L'*écart-type* de X est

$$\sigma_X = \sigma(X) := \sqrt{\mathrm{var}(X)}.$$

La variance de X mesure la dispersion de X autour de sa moyenne. En particulier, $\mathrm{var}(X) = 0$ si et seulement si X est constante p.p.

Exercice 22.1 (Calcul élémentaire de variance)**.**
Calculer la variance d'une variable aléatoire uniforme sur $[a, b]$.

Proposition 22.2. *Inégalité de Markov : Si $X \geqslant 0$ et $a > 0$,*

$$P(X \geqslant a) \leqslant \frac{1}{a}E(X).$$

Inégalité de Bienaymé-Tchebicheff :[1, 2] *Si $X \in L^2$ et $a > 0$,*

$$P(|X - E(X)| \geqslant a) \leqslant \frac{1}{a^2}\mathrm{var}(X).$$

1. Irénée Jules Bienaymé, probabiliste et statisticien français (1796–1878).
2. Pafnouti Lvovitch Tchebychev, mathématicien russe (1821–1894), qui s'est illustré en probabilités et en théorie des nombres.

Démonstration. L'inégalité de Markov a déjà été démontrée, dans la proposition 7.9. L'inégalité de Bienaymé-Tchebychev s'en déduit en l'appliquant à la variable positive $(X - E(X))^2$. □

Définition 22.3. Si $X, Y \in L^2$, la *covariance* de X et de Y est

$$\mathrm{cov}(X,Y) = E((X-E(X))(Y-E(Y))) = E(XY) - E(X)E(Y).$$

Si $X \in L^2_{\mathbb{R}^d}$ (c'est-à-dire $X = (X_1, ..., X_d) : \Omega \to \mathbb{R}^d$ et $E(X_i^2) \leqslant \infty$ pour tout $i \in \{1, ..., d\}$), la *matrice de covariance* de X est la matrice $d \times d$ notée K_X et définie par

$$K_X = \left((X - E(X))^t(X - E(X))\right).$$

Remarques 22.4. — La covariance de X et de Y mesure la corrélation entre X et Y, puisque par exemple les événements pour lesquels X et Y sont simultanément supérieures (ou inférieures) à leur espérance ont une contribution positive.
Cette notion de corrélation ne doit pas être confondue avec celle de dépendance, que l'on décrira ultérieurement. (Grossièrement, X et Y sont indépendantes si *n'importe quelle fonction* de X est décorrélée de *n'importe quelle fonction* de Y.)

— Si l'on pose $Z = (Z_1, ..., Z_d) := X - E(X)$, la matrice Z^tZ est la matrice carrée symétrique

$$Z^tZ = \begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix} \begin{pmatrix} Z_1, \cdots, Z_d \end{pmatrix} = \begin{pmatrix} Z_1^2 & \cdots & Z_1Z_d \\ \vdots & & \vdots \\ Z_1Z_d & \cdots & Z_d^2 \end{pmatrix},$$

de sorte que le coefficient (i,j) de K_X est $\mathrm{cov}(X_i, X_j)$.

— Avec ces notations, l'inégalité de Cauchy-Schwarz dit :

$$|\mathrm{cov}(X,Y)| \leqslant \sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)} = \sigma_X\sigma_Y.$$

Exercice 22.2 (Matrice de covariance)**.**
Soit $X \in L^2_{\mathbb{R}^d}$.

1. Montrer que, si $a \in \mathbb{R}^d$ et $M \in M_d(\mathbb{R})$,

$$K_{MX+a} = MK_X{}^tM.$$

2. Montrer que K_X est positive, au sens que ${}^t\xi K_X \xi \geqslant 0$ pour tout $\xi \in \mathbb{R}^d$.

Exercice 22.3 (Développement d'une fonction caractéristique)**.**
Soit $X \in L^2(\Omega)$ une variable aléatoire centrée ($E(X) = 0$). Montrer que

$$\Phi_X(\xi) = 1 - \frac{1}{2}{}^t\xi K_X\xi + o(\|\xi\|^2).$$

Si $f \in L^1(\mathbb{R})$ et $\alpha \in \mathbb{R}_*$, un calcul immédiat montre que

$$\mathcal{F}(f(x/\alpha))(\xi) = \alpha\mathcal{F}(f(x))(\alpha\xi).$$

En particulier, si $f(x)$ et $\mathcal{F}(f(x))(\xi)$ sont de carré intégrable et si $\alpha > 1$, la variance de $f(x/\alpha)$ est multipliée à partir de celle de $f(x)$ par facteur α^2, tandis que la variance de $\mathcal{F}(f(x/\alpha))(\xi)$ est divisée par rapport à celle de $\mathcal{F}(f(x))(\xi)$ par un facteur α : en faisant des homothéties, on voit que le produit des variances d'une fonction et de sa transformée de Fourier est constant. L'exercice suivant montre que ce produit possède en fait une minoration universelle (indépendante de f) !

Exercice 22.4 (Inégalité d'Heisenberg)**.**
Cet exercice utilise la transformation de Fourier-Plancherel, introduite dans l'exercice 21.3.
On munit l'espace

$$H = \{\psi \in L^2_{\mathbb{C}}(\mathbb{R}),\ x\psi(x) \in L^2 \text{ et } \xi\hat{\psi}(\xi) \in L^2\}$$

de la forme bilinéaire

$$(\varphi|\psi) = \int \left(\varphi\bar{\psi} + |x|\varphi\bar{\psi}\right)\, dx + \int |\xi|\, \hat{\varphi}(\xi)\, \bar{\hat{\psi}}(\xi)\, d\xi.$$

1. Montrer que H un espace de Hilbert.

Soit $\psi \in H$. Notons x_ψ et ξ_ψ les centres de masse des mesures $\frac{|\psi(x)|^2}{\|\psi\|_2^2}\, dx$ et $\frac{|\hat{\psi}(\xi)|^2}{\|\psi\|_2^2}\, d\xi$:

$$x_\psi = \int_{\mathbb{R}} x|\psi(x)|^2 \frac{dx}{\|\psi\|_2^2}, \quad \xi_\psi = \int_{\mathbb{R}} \xi|\hat{\psi}(\xi)|^2 \frac{d\xi}{\|\psi\|_2^2},$$

et Δx_ψ et $\Delta\xi_\psi$ les résolutions temporelle et fréquentielle de ψ :

$$\Delta x_\psi = \left(\int_{\mathbb{R}} |x - x_\psi|^2 |\psi(x)|^2 \frac{dx}{\|\psi\|_2^2}\right)^{1/2}, \quad \Delta\xi_\psi = \left(\int_{\mathbb{R}} |\xi - \xi_\psi|^2 |\hat{\psi}(x)|^2 \frac{d\xi}{\|\psi\|_2^2}\right)^{1/2}.$$

La *résolution conjointe* de ψ est le produit $\Delta x_\psi \cdot \Delta\xi_\psi$.

2. Montrer l'inégalité d'Heisenberg [3] : quelle que soit $\psi \in L^2(\mathbb{R})$,

$$\Delta x_\psi \cdot \Delta\xi_\psi \geqslant \frac{1}{2};$$

on pourra d'abord remarquer qu'on peut supposer que $x_\psi = \xi_\psi = 0$, puis passer par le cas des fonctions de classe C^∞ à support compact, pour lesquelles l'inégalité découle d'une intégration par parties, de l'inégalité de Cauchy-Schwarz et du fait que la transformation de Fourier-Plancherel est une isométrie de $L^2(\mathbb{R})$.

3. Montrer que l'égalité est atteinte uniquement pour les gaussiennes ; on pourra utiliser le fait que l'inégalité de Cauchy-Schwarz est atteinte uniquement pour les vecteurs colinéaires.

L'inégalité ci-dessus possède des implications profondes en Mécanique quantique (principe d'incertitude d'Heisenberg), ainsi qu'en théorie des ondelettes [KLR98, p. 311].

3. Werner Heisenberg, physicien allemand (1901–1976), l'un des découvreurs de la Mécanique quantique.

Chapitre 23

Indépendance

> Measure theory ends and probability theory begins with the definition of independence.
>
> R. Durrett

Le concept d'indépendance peut être défini au niveau des événements, des variables aléatoires ou des tribus. Ce dernier niveau est le plus difficile à appréhender, mais aussi le plus fondamental.

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

Définition 23.1. Deux événements $A, B \in \mathcal{A}$ sont *indépendants* si

$$P(A \cap B) = P(A) \cap P(B). \tag{23.1}$$

Cette définition a le mérite d'être symétrique par rapport à A et B (elle se généralise donc facilement à plusieurs événements) et de garder un sens même si A ou B sont négligeables. Mais, dans le cas où $P(A) > 0$, l'égalité (23.1) s'interprète en disant que la *probabilité conditionnelle de B sachant A*, définie comme

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

coïncide avec $P(B)$; autrement dit, le fait de savoir que A est réalisé ne modifie pas la probabilité de B.

Exemple 23.2.
- *Lancé d'un dé :* L'univers est $\Omega = \{1, ..., 6\}$ et $P(\{\omega\}) = 1/6$ pour tout $\omega \in \Omega$. Les événements $A = \{1, 2\}$ et $B = \{1, 3, 5\}$ sont indépendants, mais A et $C = \{3, 5\}$ ne le sont pas.
- *Lancé de deux dés :* L'univers est $\Omega = \{1, ..., 6\}^2$ et la probabilité est uniforme ($P(\{\omega\}) = 1/36$ pour tout $\omega \in \Omega$). Les événements $A = \{1\} \times \{1, ..., 6\}$ et $B = \{1, ..., 6\} \times \{1\}$ sont indépendants.

Définition 23.3. Des événements A_i indexés par un ensemble I quelconque sont *indépendants* si, pour toute partie finie non vide $J \subset I$ on a

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j). \tag{23.2}$$

Remarque 23.4. Pour que les A_i soient indépendants, il ne suffit pas que l'égalité (23.2) soit satisfaite pour toutes les paires $J = \{i, j\} \subset I$.

En revanche, si I est fini, il suffit que l'égalité (23.2) soit satisfaite pour $J = I$, puisque, si J est strictement inclus dans I, on peut toujours poser $A_i = \Omega$ pour tout $i \in I \backslash J$, de sorte que

$$\bigcap_{j \in J} A_j = \bigcap_{i \in I} A_i \quad \text{et} \quad \prod_{j \in J} P(A_j) = \prod_{i \in i} P(A_i).$$

Définition 23.5. Soit $(\mathcal{B}_i)_{i \in I}$ une famille de tribus de Ω incluses dans $\mathcal{A}$ (on dit que les $\mathcal{B}_i$ sont des *sous-tribus* de $\mathcal{A}$). Les sous-tribus $\mathcal{B}_i$ sont *indépendantes* si, pour toute partie finie non vide $J \subset I$,

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

pour toute famille $(A_j)_{j \in J} \in \prod_{j \in J} \mathcal{B}_j$.

Soit $(X_i : \Omega \to (E_i, \mathcal{E}_i))_{i \in I}$ une famille de variables aléatoires. Les X_i sont *indépendantes* si les tribus $\sigma(X_i) = \{X_i^{-1}(B_i),\ B_i \in \mathcal{E}_i\}$ le sont.

Exemple 23.6. Le principal modèle de construction de tribus indépendantes est celui des espaces produits. Considérons en effet deux espaces de probabilité $(\Omega_i, \mathcal{A}_i, P_i)$, $i = 1, 2$, et le produit $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P = P_1 \otimes P_2)$. Soient $\mathcal{B}_i$, $i = 1, 2$ les deux sous-tribus de $\mathcal{A}$ “horizontale” et “verticale” :

$$\begin{cases} \mathcal{B}_1 = \sigma\left(\{A_1 \times \Omega_2,\ A_1 \in \mathcal{A}_1\}\right) \\ \mathcal{B}_2 = \sigma\left(\{\Omega_1 \times A_2,\ A_2 \in \mathcal{A}_2\}\right). \end{cases}$$

Pour toutes parties $A_1 \times \Omega_2 \in \mathcal{A}_1$ et $\Omega_1 \times A_2 \in \mathcal{A}_2$, comme

$$(A_1 \times \Omega_2) \cap (\Omega_1 \times A_2) = A_1 \times A_2,$$

on a bien

$$\begin{aligned} P\left(A_1 \times \Omega_2 \ \cap \ \Omega_2 \times A_2\right) &= P(A_1 \times A_2) \\ &= P_1(A_1)\, P_2(A_2) \qquad\qquad (\text{car } P = P_1 \otimes P_2) \\ &= P(A_1 \times \Omega_1)\, P(\Omega_2 \times A_2). \end{aligned}$$

En particulier, n'importe quelles variables aléatoires $X_i : \Omega_i \to E_i$, $i = 1, 2$, se relèvent canoniquement en des variables aléatoires $\tilde{X}_i : \Omega \to E_i$ (définies par $\tilde{X}_i(\omega_1, \omega_2) = X_i(\omega_i)$), qui sont chacune $\mathcal{B}_i$-mesurable, et qui sont donc indépendantes l'une de l'autre.

Exemple 23.7. Étant donné une variable aléatoire $X : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E})$, on a couramment besoin d'en construire une ou plusieurs « copies indépendantes », i.e. des variables aléatoires X_1 et X_2 ayant la même loi que X, mais indépendantes l'une de l'autre (ce qui interdit de choisir simplement $X_1 = X_2 = X$) ; voir par exemple l'exercice 26.4. Les deux variables

$$X_i : \Omega^2 \to E, \quad \omega = (\omega_1, \omega_2) \mapsto X(\omega_i)$$

ont même loi et l'exemple 23.6 montre qu'elles sont indépendantes. Cette construction se généralise à un nombre fini quelconque de copies (sur l'espace produit Ω^n, muni des tribu et probabilité produits), ou même à une infinité dénombrable de copies (sur l'espace Ω^∞, muni de la probabilité de Bernoulli construite dans le problème 13.1).

Lemme 23.8. *Des événements A_i $(i \in I)$ sont indépendants si et seulement si les tribus $\sigma(\{A_i\})$ $(i \in I)$ le sont.*

Démonstration. Rappelons que $\sigma(\{A_i\}) = \{\varnothing, A_i, A_i^c, \Omega\}$. Donc, si ces tribus sont indépendantes, les A_i le sont trivialement.

Réciproquement, supposons les A_i indépendants. Soient J une partie finie de I et, pour tout $j \in J$, $B_j \in \sigma(\{A_j\})$. On veut montrer que

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

Si l'un des B_j est l'ensemble vide, la conclusion est vraie puisque les deux membres de l'égalité sont nuls. Si $B_k = \Omega$ pour un certain $k \in J$,

$$\begin{cases} P\left(\bigcap_j B_j\right) = P\left(\bigcap_{j\neq k} B_j\right) \\ \prod_j P(B_j) = \prod_{j\neq k} P(B_j), \end{cases}$$

donc, par une récurrence finie, on se ramène au cas où les B_j sont différents de Ω. On peut donc supposer que, pour tout j, $B_j = A_j$ ou A_j^c. Puisque en général on a

$$\text{"}P(A^c \cap B) = P(B) - P(A \cap B),\text{"}$$

si il existe un seul $k \in J$ tel que $B_k = A_k^c$, on voit que

$$\begin{aligned} P\left(\bigcap_j B_j\right) &= P\left(A_k^c \bigcap \bigcap_{j\neq k} A_j\right) \\ &= P\left(\bigcap_{j\neq k} A_j\right) - P\left(A_k \bigcap \bigcap_{j\neq k} A_j\right) \\ &= \prod_{j\neq k} P(A_j) - P(A_k) \prod_{j\neq k} P(A_j) \\ &= P(A_k^c) \prod_{j\neq k} P(A_j) \\ &= \prod_j P(A_j). \end{aligned}$$

Par une nouvelle récurrence finie, on voit que l'égalité voulue est satisfaite dans tous les cas. □

Théorème 23.9. *Quel que soit $i \in \{1, ..., n\}$, soit $X_i : (\Omega, \mathcal{A}, P) \to (E_i, \mathcal{E}_i)$ une variable aléatoire. Les propriétés suivantes sont équivalentes :*

(1) $X_1, ..., X_n$ sont indépendantes.

(2) Quels que soient $B_1 \in \mathcal{E}_1,, \ ..., \ B_n \in \mathcal{E}_n$,

$$P\left(\bigcap_{1\leqslant i\leqslant n}\{X_i \in B_i\}\right) = \prod_{1\leqslant i\leqslant n} P(X_i \in B_i).$$

(3) La loi conjointe des X_i est le produit des lois marginales :

$$P_X = \bigotimes_{1\leqslant i\leqslant n} P_{X_i}.$$

(4) Pour toute fonction mesurable $h : \prod_{1\leqslant i\leqslant n} E_i \to \mathbb{R}^+$,

$$\int_{\Omega^n} h\left(X_1(\omega_1), ..., X_n(\omega_n)\right)\, dP^{\otimes n}(\omega_1, ..., \omega_n) = \int_{\Omega} h(X_1(\omega), ..., X_n(\omega))\, dP(\omega).$$

Si les X_i sont réelles, l'item supplémentaire suivant est équivalent aux précédents :

(5) La fonction caractéristique de X est le produit tensoriel des fonctions caractéristiques des X_i :

$$\Phi_X(\xi) = \prod_{1\leqslant i\leqslant n} \Phi_{X_i}(\xi_i)$$

("tensoriel" signifie ici qu'on ne donne pas le même argument aux Φ_{X_i}).

Démonstration. Pour simplifier les notations, nous allons démontrer le théorème pour $n = 2$ et laisserons le cas général (qui ne présente pas de difficulté autre qu'au niveau des notations) au lecteur. Notons $X : (\Omega, \mathcal{A}, P) \to (E, \mathcal{E})$ et $Y = (\Omega, \mathcal{A}, P) \to (F, \mathcal{F})$ les deux variables aléatoires.

— Les propriétés (1) et (2) sont équivalentes par définition.

— *(2) ⇔ (3) :* Si $A \in \mathcal{A}$ et $B \in \mathcal{B}$, par défintion des probabilités image $P_{(X,Y)}$, P_X et P_Y,

$$P(\{X \in A\} \cap \{Y \in B\}) = P((X, Y) \in A \times B) = P_{(X,Y)}(A \times B)$$

et

$$P(\{X \in A\})P(\{Y \in B\}) = P_X(A)\, P_Y(B).$$

Par ailleurs, d'après la proposition 13.4, le membre de droite de la dernière égalité vaut $P_X \otimes P_Y(A \times B)$, et ceci caractérise $P_X \otimes P_Y$. Donc, la propriété (2) équivaut bien au fait que $P_{(X,Y)} = P_X \otimes P_Y$ (3).

— *(3) ⇒ (4) :* Si $h : E \times F \to \mathbb{R}^+$ est mesurable,

$$\int_{\Omega^2} h(X(\omega), Y(\omega'))\, dP \otimes P(\omega, \omega')$$

$$= \int_\Omega \left(\int_\Omega h(X(\omega), Y(\omega'))\, dP(\omega) \right) dP(\omega') \quad \text{(théorème de Tonelli)}$$

$$= \int_F \left(\int_E h(x,y)\, dP_X(x) \right) dP_Y(y) \quad \text{(formule de transfert)}$$

$$= \int_{E\times F} h(x,y) d(P_X \otimes P_Y)(x,y) \quad \text{(théorème de Tonelli)}$$

$$= \int_{E\times F} h(x,y)\, dP_{(X,Y)}(x,y) \quad \text{(indépendance)}$$

$$= \int_\Omega h(X,Y)\, dP \quad \text{(formule de transfert).}$$

— *(4) ⇒ (2) :* Si $A \in \mathcal{A}$ et $B \in \mathcal{B}$,

$$\begin{aligned}
P(\{X \in A\} \cap \{Y \in B\}) = P_{(X,Y)}(A \times B) &= \int_{E\times F} \mathbb{1}_{A\times B}\, dP_{(X,Y)} \\
&= \int_\Omega \mathbb{1}_{A\times B}(X(\omega), Y(\omega))\, dP(\omega) \quad \text{(formule de transfert)} \\
&= \int_{\Omega^2} \mathbb{1}_{A\times B}(X(\omega), Y(\omega'))\, dP^{\otimes 2}(\omega, \omega') \\
&\qquad \text{(par l'hypothèse (4))} \\
&= P(\{X \in A\}) P(\{Y \in B\}) \quad \text{(théorème de Tonelli).}
\end{aligned}$$

— *(2) ⇒ (5) :* Si les X_i sont indépendantes,

$$\begin{aligned}
\Phi_X(\xi) &= E\left(e^{-iX\cdot\xi}\right) \\
&= E\left(\prod_j e^{-iX_j\xi_j}\right) \\
&= \prod_j E\left(e^{iX_j\xi_j}\right) \\
&= \prod_j \Phi_{X_j}(\xi_j) = \left(\bigotimes_j \Phi_{X_j}\right)(\xi).
\end{aligned}$$

— *(5) ⇒ (2) :* Sous l'hypothèse (5),

$$\widehat{P_X} = \bigotimes_j \widehat{P_{X_j}} = \widehat{\bigotimes_j P_{X_j}}.$$

D'après le théorème d'injectivité 20.9 (admis), on a donc

$$P_X = \bigotimes_j P_{X_j}.$$

□

Exercice 23.1 (Sur l'indépendance)**.**
1. Si X et Y sont indépendantes, n'importe quelle fonction de X est indépendante de n'importe quelle fonction de Y.

2. Si $X, Y \in L^2$ sont indépendantes, $\mathrm{cov}(X, Y) = 0$.

3. Si X et Y sont dans L^1 et indépendantes, $XY \in L^1$. Donner un exemple montrant que XY n'est pas forcément dans L^1 si X et Y sont dans L^1 mais pas indépendantes.

4. (H. Doss) Soit X de loi uniforme sur $[0, 1]$. Montrer qu'il existe une infinité de fonctions de X indépendantes. Faire de même en supposant seulement que X a une densité par rapport à la mesure de Lebesgue.

Terminons ce chapitre par la définition de la convolée de deux probabilités, concept utile dans l'étude de la somme de deux variables aléatoires.

Définition 23.10. Si μ et ν sont deux probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, la *convolée* de μ et ν est la probabilité, notée $\mu * \nu$, image de $\mu \otimes \nu$ par l'addition $(x, y) \mapsto x + y$: pour toute fonction mesurable $\varphi : \mathbb{R}^d \to [0, +\infty[$,

$$\int_{\mathbb{R}^d} \varphi(z)\, d(\mu * \nu)(z) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x + y)\, d(\mu \otimes \nu)(x, y).$$

Exercice 23.2 (Convolée de probabilités à densité).
Soient μ et ν sont deux probabilités sur $\mathbb{R}^d$ possédant une densité par rapport à la mesure de Lebesgue :

$$\mu = f\lambda \quad \text{et} \quad \nu = g\lambda.$$

Montrer que la probabilité $\mu * \nu$ possède une densité et que cette densité est $f * g$:

$$(f\lambda) * (g\lambda) = (f * g)\lambda.$$

Exercice 23.3 (Somme de deux variables indépendantes).
Soient $X, Y : \Omega \to \mathbb{R}^d$ deux vecteurs aléatoires indépendants. Montrer les propriétés suivantes de $X + Y$:

1. La loi de $X + Y$ est $P_X * P_Y$. Si X et Y ont des densités f_X et f_Y, $X + Y$ a une densité, qui est $f_X * f_Y$.
2. La fonction caractéristique de $X + Y$ est
$$\Phi_{X+Y}(\xi) = \Phi_X(\xi)\Phi_Y(\xi).$$
3. Si de plus $X, Y \in L^2_{\mathbb{R}^d}$, les covariances s'ajoutent :
$$K_{X+Y} = K_X + K_Y.$$

Exercice 23.4 (Convolée de deux variables de Poisson indépendantes).
Quelle est la loi de la somme de deux variables aléatoires de Poisson indépendantes, de paramètres respectifs λ et μ ?

Exercice 23.5 (Indépendance de tribus).
Montrer que des tribus $\mathcal{B}_1, ..., \mathcal{B}_n$ sont indépendantes si et seulement si, pour toutes variables aléatoires réelles positives $X_1, ..., X_n$ telles que X_i soit $\mathcal{B}_i$-mesurable pour tout $i \in \{1, ..., n\}$,

$$E[X_1 \cdots X_n] = E[X_1] \cdots E[X_n].$$

Exercice 23.6 (Exemples de variables indépendantes)**.**
1. Soient R une variable de loi exponentielle de paramètre 1 et Θ une variable uniforme sur $[0,1]$; on suppose R et Θ indépendantes. Monter que les variables

$$X = \sqrt{R}\cos(2\pi\Theta) \quad \text{et} \quad Y = \sqrt{R}\sin(2\pi\Theta)$$

sont indépendantes.
2. Soient X et Y des variables aléatoires indépendantes à valeurs dans $\mathbb{N}_*$, avec

$$P(\{X = n\}) = P(\{Y = n\}) = \frac{1}{2^n}.$$

Trouver les probabilités des événements suivants :
2.1. $\{\min(X,Y) \leqslant n\}$
2.2. $\{X = Y\}$
2.3. $\{Y > X\}$
2.4. $\{X \text{ divise } Y\}$ (dans ce cas on pourra laisser le résultat sous la forme de la somme d'une série).

Exercice 23.7 (Mesurabilité par rapport à deux tribus indépendantes)**.**
Soient $(\Omega, \mathcal{A}, P)$ un espace de probabilité, $\mathcal{F}$ et $\mathcal{G}$ deux sous-tribus indépendantes de $\mathcal{A}$, et X une variable aléatoire réelle mesurable à la fois par rapport à $\mathcal{F}$ et à $\mathcal{G}$. Montrer que X est constante presque partout.

Exercice 23.8 (Lemme de Borel-Cantelli)**.**
Soit (A_n) une suite d'événements.
1. Montrer que, si $\sum P(A_n) < \infty$, $P(\limsup A_n) = 0$ *i.e.*, p.s. $x \in \Omega$ est contenu dans au plus un nombre fini de A_n.
2. Montrer réciproquement que si $P(\limsup A_n) = 0$ et si les A_n sont indépendants, $\sum P(A_n) < \infty$.
Application :
3. Soit (X_n) une suite de variables aléatoires réelles indépendantes, chacune de loi exponentielle telle que

$$P(X_n > x) = e^{-x} \quad (\forall x \geqslant 0).$$

3.1. Pour $\alpha > 0$, montrer que

$$P\left(\{X_n \geqslant \alpha \ln n \text{ pour infinité de valeurs de } n\}\right) = \begin{cases} 0 & \text{si } \alpha > 1 \\ 1 & \text{si } \alpha \leqslant 1. \end{cases}$$

3.2. Montrer que $\limsup \frac{X_n}{\ln n} = 1$ presque sûrement.

Problème ★ 23.1 (Loi du $0-1$ de Kolmogorov)**.**
Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité.
1. La tribu $\mathcal{A}$ est *triviale*[1] si tout événement $A \in \mathcal{A}$ est de probabilité 0 ou 1.

1. Cette définition généralise la définition de l'exercice 3.1.

Montrer que $\mathcal{A}$ est triviale si et seulement si l'une des deux propriétés (équivalentes) suivantes sont satisfaites :

1.1. $\mathcal{A}$ est indépendante d'elle-même.

1.2. Toute variable aléatoire $X : (\Omega, \mathcal{A}) \to (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$ est constante presque sûrement.

2. On démontre ici un critère d'indépendance de tribus qui sera utile ensuite.

2.1. Soient $\mathcal{I}$ et $\mathcal{J}$ deux classes d'événements stables par intersection finie. Montrer que, si $\mathcal{I}$ et $\mathcal{J}$ sont indépendantes (au sens que $P(I \cap J) = P(I)P(J)$ pour tous $I \in \mathcal{I}$ et $J \in \mathcal{J}$), $\sigma(\mathcal{I})$ et $\sigma(\mathcal{J})$ sont indépendantes ; on pourra utiliser le corollaire 9.3 du théorème 9.2 de classe monotone.

2.2. En guise d'exemple immédiat, en déduire que deux variables aléatoires réelles X et Y sont indépendantes si et seulement si

$$P(X \leqslant x; Y \leqslant y) = P(X \leqslant x)P(Y \leqslant y) \quad (\forall x, y \in \mathbb{R}).$$

3. Soient $(\mathcal{E}_n)$ une suite croissante de sous-tribus de $(\Omega, \mathcal{A})$ et $(\mathcal{F}_n)$ une suite décroissante de sous-tribus de $(\Omega, \mathcal{A})$. On suppose que

— $\mathcal{F}_0 \subset \sigma(\cup \mathcal{G}_n)$

— pour tout n, $\mathcal{F}_n$ et $\mathcal{G}_n$ sont indépendantes.

Montrer que la tribu $\cap \mathcal{G}_n$ est triviale (*loi du* $0-1$ *de Kolmogorov*).

4. Soit (X_n) une suite de variables aléatoires réelles indépendantes. On note

$$\mathcal{E}_n = \sigma(\{X_1, ..., X_n\}), \quad \mathcal{F}_n = \sigma(\{X_{n+1}, X_{n+2}, ...\}) \quad (n \in \mathbb{N})$$

et

$$\mathcal{F}_\infty = \cap \mathcal{F}_n;$$

$\mathcal{F}_\infty$ est la *tribu de queue* (ou la *tribu asymptotique*) de (X_n).

4.1. Montrer que, pour tout n, $\mathcal{E}_n$ et $\mathcal{F}_n$ sont indépendantes.

4.2. Montrer que la tribu $\mathcal{F}_\infty$ est triviale.

4.3. Montrer les deux propriétés suivantes :

— Soit (X_n) converge p.s., soit elle diverge p.s.

— Les variables suivantes sont p.s. constantes :

$$\liminf X_n, \quad \limsup X_n, \quad \liminf \frac{1}{n}(X_1 + \cdots X_n), \quad \limsup \frac{1}{n}(X_1 + \cdots X_n).$$

Problème 23.2 (Espérance et indépendance)**.**
Trouver deux variables aléatoires réelles dépendantes X et Y telles que $E(XY) = E(X)E(Y)$.

Problème 23.3 (Nombres diophantiens encore)**.**
Soient $\tau > 2$ et $\gamma > 0$.

1. Utiliser le lemme de Borel-Cantelli pour montrer que

$$A := \left\{ x \in [0,1];\ |x - \frac{p}{q}| \leqslant \frac{\gamma}{q^\tau} \text{ pour une infinités d'entiers } p, q \right\}$$

est de mesure nulle. *Indication :* Il suffit de considérer les entiers p tels que $0 \leqslant p \leqslant q$ (pourquoi ?), et l'on pourra utiliser le fait que la série $\sum \frac{1}{q^{\tau-1}}$ converge.
Un nombre $x \in \mathbb{R}$ est *diophantien* s'il existe des nombres réels $\gamma, \tau > 0$ tels que

$$\left| x - \frac{p}{q} \right| > \frac{\gamma}{q^\tau} \tag{23.3}$$

pour touts entiers relatifs $q \neq 0$ et p (cf. l'exercice 5.7).

2. Montrer que presque tout nombre réel est diophantien. *Indication :* Montrer qu'il suffit de se restreindre à l'intervalle $[0, 1]$, et qu'on peut toujours choisir γ et τ rationnels avec $\tau > 2$. Utiliser ensuite la première question.

Chapitre 24

Vecteurs gaussiens

Définition 24.1. Si $\sigma > 0$ et $\mu \in \mathbb{R}$, la *loi gaussienne (unidimensionnelle)*[1] $N_1(\mu, \sigma^2)$ est la probabilité sur $\mathbb{R}$ de densité

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

(par rapport à la mesure de Lebesgue). La *loi gaussienne* $N_1(\mu, 0)$ est la mesure de Dirac δ_μ.

On dit que la loi est

— *réduite* si $\mu = 0$ et $\sigma = 1$

— *centrée* si $\mu = 0$

— *dégénérée* si $\sigma = 0$ (elle n'a alors pas de densité).

Si $\sigma > 0$, la transformée de Fourier de $N_1(\mu, \sigma^2)$ est (en se ramenant facilement à l'exemple 20.4),

$$\widehat{N_1(\mu, \sigma^2)}(\xi) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{\mathbb{R}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} - ix\xi\right) dx = \exp\left(-i\mu\xi - \frac{\sigma^2}{2}\xi^2\right).$$

Dans le cas dégénéré, on obtient

$$\widehat{N_1(\mu, 0)}(\xi) = \widehat{\delta_\mu}(\xi) = \int_{\mathbb{R}} e^{-ix\xi} d\delta_\mu(x) = e^{-i\mu\xi}.$$

Exercice 24.1 (Simulation de Box-Muller)**.**
Soient U_1 et U_2 deux variables aléatoires indépendantes uniformes sur $]0,1[$. Montrer que

$$X = \sqrt{-2\ln U_2} \cos 2\pi U_1 \quad \text{et} \quad Y = \sqrt{-2\ln U_2} \cos 2\pi U_2$$

sont indépendantes normales centrées.

Définir la loi gaussienne d-dimensionnelle requiert quelques préliminaires. Nous allons en fait commencer par définir ce qu'est un vecteur gaussien sans caractériser sa loi.[2] Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité.

1. Johann Carl Friedrich Gauss, mathématicien allemand (1777-1855), dont l'ampleur des contributions le firent surnommer le *prince des mathématiciens*.

2. La difficulté vient du fait que nous n'allons pas décrire l'analogue en dimension d de la mesure de Dirac.

Définition 24.2. Un vecteur aléatoire $X : (\Omega, \mathcal{A}, P) \to \mathbb{R}^d$ est *gaussien* si, pour tout $\eta \in \mathbb{R}^d$, la variable réelle $\eta \cdot X = \sum_i \eta_i X_i$ suit une loi gaussienne.

Remarques 24.3. – En prenant comme vecteurs η les d vecteurs de la base canonique de $\mathbb{R}^d$, on voit que les composantes d'un vecteur gaussien sont gaussiennes. Mais, comme on le verra, la réciproque est fausse !
– Par ailleurs, si X est gaussien, tout vecteur de la forme $MX + a$ où $M \in M_{d,k}(\mathbb{R})$ et $a \in \mathbb{R}^k$ est un vecteur gaussien.

Théorème 24.4. *Un vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ est gaussien si et seulement si il existe un vecteur $\mu \in \mathbb{R}^d$ et une matrice symétrique $K \in M_d(\mathbb{R})$ tels que*

$$\Phi_X(\xi) = \exp\left(-i\mu \cdot \xi - \frac{1}{2}{}^t\xi K \xi\right). \tag{24.1}$$

De plus, μ et K sont l'espérance et la matrice de covariance de X.

Démonstration. Remarquons d'abord que, si $a \in \mathbb{R}$ et $M \in M_d(\mathbb{R})$,

$$\Phi_{MX+a}(\xi) = e^{-ia\cdot\xi}\Phi_X({}^tM\xi). \tag{24.2}$$

($\Leftarrow$) Soient $\eta \in \mathbb{R}^d$ et $Y = \eta \cdot X$. D'après la formule (24.2) (appliquée avec $M = {}^t\eta$ et $a = 0$), pour tout $\xi \in \mathbb{R}$,

$$\begin{aligned}\Phi_Y(\xi) &= \Phi_X(\xi\eta) \\ &= \exp\left(-i(\eta\cdot\mu)\,\xi - \frac{{}^t\eta K\eta}{2}\xi^2\right).\end{aligned}$$

On reconnaît la transformée de Fourier de la loi gaussienne $N_1(\eta\cdot\mu, {}^t\eta K\eta)$. Donc, par injectivité de la transformation de Fourier des probabilités, Y suit la loi $N_1(\eta\cdot\mu, {}^t\eta K\eta)$. Donc X est gaussien.

($\Rightarrow$) Supposons X gaussien. On veut calculer $\Phi_X(\eta)$ mais on ne connait a priori que la loi des $\eta \cdot X$. Fixons $\eta \in \mathbb{R}^d$ et notons μ et K l'espérance et la matrice de covariance de X. Alors la variable $Y = \eta \cdot X$ est gaussienne de loi $N_1(\eta\cdot\mu, {}^t\eta K\eta)$. Donc, d'après l'exemple 20.4,

$$\Phi_Y(\xi) = \exp\left(-i(\eta\cdot\mu)\xi - \frac{1}{2}({}^t\eta K\eta)\xi^2\right).$$

Donc

$$\Phi_X(\eta) = E\left(e^{iX\cdot\eta}\right) = \Phi_{\eta\cdot X}(1) = \Phi_Y(1)$$

a bien l'expression voulue.

Enfin, en général on a, d'après le corollaire 20.12,

$$\Phi_X(\xi) = 1 - i{}^tE(X)\xi - \frac{1}{2}{}^t\xi E(X{}^tX)\xi + o(\|\xi\|^2),$$

donc[3]

$$\begin{aligned}\ln\Phi_X(\xi) &= -i{}^tE(X)\xi - \frac{1}{2}{}^t\xi E(X{}^tX)\xi + \frac{1}{2}\left({}^tE(X)\xi\right)^2 + o(\|\xi\|^2) \\ &= -i{}^tE(X)\xi - \frac{1}{2}{}^t\xi\left(E(X{}^tX) - E(X){}^tE(X)\right)\xi + o(\|\xi\|^2).\end{aligned}$$

3. $\ln(1+y) = y - \frac{y^2}{2} + o(y^2)$ quand y tend vers 0, n'est-ce pas ?

En comparant avec l'expression 24.1, on identifie les coefficients μ et K comme l'espérance et la matrice de covariance. $\square$

Définition 24.5. Soient $\mu \in \mathbb{R}^d$ et $K \in M_d(\mathbb{R})$ symétrique. La *loi gaussienne* $N_d(\mu, K)$ est l'unique loi sur $\mathbb{R}^d$ dont la transformée de Fourier est la fonction

$$\exp\left(-i\mu \cdot \xi - \frac{1}{2}{}^t\xi K\xi\right),$$

et un vecteur aléatoire est *gaussien* si sa loi est gaussienne.

Proposition 24.6. *Si X est un vecteur gaussien, ses composantes sont indépendantes si et seulement si sa matrice de covariance K_X est diagonale.*

Démonstration. Si $X_1, ..., X_d$ sont indépendantes, c'est un fait général que les covariances

$$\operatorname{cov}(X_j, X_k) = E(X_jX_k) - E(X_j)E(X_k) = 0$$

sont nulles si $j \neq k$ (exercice 23.1).
Réciproquement, si K est diagonale :

$$K = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_d^2 \end{pmatrix},$$

la fonction caractéristique de X vérifie

$$\Phi_X(\xi) = \prod_j \Phi_{X_j}(\xi_j).$$

Donc les X_j sont indépendantes. $\square$

Exercice 24.2 (Vecteur non gaussien)**.**
Soient X une variable aléatoire de loi $N(0,1)$ et Z indépendante de X telle que $P(\{Z = 1\}) = P(\{Z = -1\}) = 1/2$. Montrer que $Y = ZX$ est gaussien mais que (X, Y) n'est pas gaussien.

Théorème 24.7. *Soit X un vecteur gaussien de loi $N_d(\mu, K)$. Il existe des réels $\lambda_1, ..., \lambda_d \geqslant 0$, des variables gaussiennes $Y_1, ..., Y_d$ indépendantes de lois respectives $N_1(0, \lambda_1)$, ..., $N_1(0, \lambda_d)$ et une matrice orthogonale $A \in O_d(\mathbb{R})$ tels que*

$$X = AY + \mu.$$

Démonstration. La matrice K étant une matrice de covariance, elle est symétrique positive (rappelons que ceci se voit par exemple en remarquant que ${}^t\xi K_X \xi = K_{{}^t\xi X} = \operatorname{var}({}^t\xi \cdot X) \geqslant 0$). Donc, il existe une matrice orthogonale A et une matrice diagonale Λ telles que

$$K = A\Lambda{}^tA.$$

Le vecteur

$$Y = {}^tA(X - \mu)$$

est gaussien, centré, et de covariance

$$K_Y = {}^tAKA = A^{-1}KA = \Lambda$$

(rappelons en effet que, avec les notations de l'exercice 22.2, on a $K_{MX+a} = MK_X{}^tM$). Comme Λ est diagonale, les Y_j sont indépendantes. De plus, $X = AY + \mu$. □

Corollaire 24.8. *Un vecteur gaussien de covariance K_X admet une densité si et seulement si* $\det K_X \neq 0$. *Sa densité est alors*

$$f_X(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\det K_X}} \exp\left(-\frac{1}{2}{}^t(x-\mu)K_X^{-1}(x-\mu_X)\right).$$

Démonstration. Supposons d'abord $\det K_X \neq 0$. Reprenons les notations du théorème précédent : il existe un vecteur aléatoire gaussien Y de composantes indépendantes $Y_j \sim N_1(0, \lambda_j)$ et une matrice orthogonale A tels que $X = AY + \mu$.
Comme $Y = A^{-1}(X - \mu)$, $K_Y = A^{-1}K_XA$ donc $\det K_Y = \det K_X \neq 0$ et, puisque $K_Y = \mathrm{Diag}\,(\lambda_1, ..., \lambda_d)$, $\lambda_j > 0$ pour tout j. Donc, pour tout j, la loi de Y_j est

$$P_{Y_j} = \frac{1}{\sqrt{2\pi\lambda_j}} \exp\left(-\frac{y_j^2}{2\lambda_j}\right) dy_j.$$

Comme les Y_j sont indépendantes, la loi conjointe de Y est le produit tensoriel de ces lois :

$$\begin{aligned} P_Y = \bigotimes_j P_{Y_j} &= \prod_j \frac{1}{\sqrt{2\pi\lambda_j}} \exp\left(-\frac{y_j^2}{2\lambda_j}\right) dy_j \\ &= \frac{1}{(2\pi)^{d/2}\sqrt{\det K_Y}} \exp\left(-{}^tyK_Y^{-1}y\right) dy = f_Y(y)\,dy, \end{aligned}$$

où l'on a utilisé $K_Y = \mathrm{Diag}\,(\lambda_1^{-1}, ..., \lambda_d^{-1})$ et $\det K_Y = \prod_j \lambda_j$. La loi de Y étant connue, celle de $X = AY + \mu$ s'en déduit par un simple changement de variable : pour toute fonction réelle borélienne positive h sur $\mathbb{R}^d$,

$$\begin{aligned} \int h(X)\,dP &= \int h(AY+\mu)\,dP \\ &= \int h(Ay+\mu)\,dP_Y(y) = \int h(Ay+\mu)\,f_Y(y)\,dy \\ &= \int h(x)\,f_Y(A^{-1}(x-\mu))\,dx \\ &\qquad (y = \varphi(x) = A^{-1}(x-\mu),\ |J\varphi(x)| = |\det A^{-1}| = 1), \end{aligned}$$

donc X possède une densité, qui vaut $f_X(x) = f_Y(A^{-1}(X - \mu))$, d'où la formule voulue.

Réciproquement, supposons que $\det K_X \neq 0$. Il existe $a \in \mathbb{R}^d$ tel que $K_Xa = 0$. La variable aléatoire $Z = a \cdot X$ a pour espérance $a \cdot \mu_X$ et variance $\sigma_Z^2 = {}^taK_Xa = 0$. Donc $Z = a \cdot \mu$ p.p. Donc $a \cdot (X - \mu) = 0$ p.p. Soit H l'hyperplan

$$H = \mu + a^\perp = \{x \in \mathbb{R}^d,\ (x - \mu) \cdot a = 0\}.$$

Presque sûrement, $X \in H$ alors que $\lambda_d(H) = 0$. Si X avait une densité, on aurait

$$1 = P(X \in H) = \int_H f_X(x)\,dx = 0,$$

ce qui est absurde. □

Dans le cas dégénéré, la démonstration précédente montre que, si l'on suppose par exemple que $\sigma_1, ..., \sigma_r > 0$ tandis que $\sigma_{r+1} = ... = \sigma_d = 0$, et si V est le sous-espace vectoriel de dimension r image par A des r premiers vecteurs de la base canonique de $\mathbb{R}^n$, la loi de X est portée par V et X possède une densité relativement à la mesure de Lebesgue associée aux coordonnées d'une base orthogonale quelconque de V.

Exercice 24.3 (Vecteur gaussien standard)**.**
Soit X un vecteur gaussien comme dans le théorème, avec $\det K > 0$. Montrer qu'il existe une matrice B telle que $X = \mu + BY$ où Y suit la loi $N_d(0, I)$.

Exercice 24.4 (Exemple de vecteur gaussien [BJ06])**.**
Soit $X = (X_1, X_2, X_3)$ un vecteur gaussien de densité

$$f(x) = K \exp -\frac{1}{2}\left(7\|x\|^2 + 6x_1x_2 + 8x_2x_3\right).$$

Soit de plus Y le vecteur défini par $X = AY$ avec

$$A = \frac{1}{5}\begin{pmatrix} 3 & 4\sqrt{2} & 3 \\ -5 & 0 & 5 \\ 4 & -3\sqrt{2} & 4 \end{pmatrix}.$$

Montrer que Y admet une densité de probabilité, et déterminer : cette densité, la valeur de K, la matrice de covariance de Y et celle de X.

FIGURE 30 – Johann Carl Friedrich GAUSS (1777–1855)

Chapitre 25

Modes de convergence ⋆

Soient $(E, \mathcal{A}, \mu)$ un espace mesuré et f_n $(n \in \mathbb{N})$ et f des fonctions réelles boréliennes sur E. Parmi les dizaines de modes de convergence possibles, mentionnons les suivants ; nous ne rappelons pas les convergences simple ou uniforme.

Définition 25.1. (f_n) converge vers f...

1. *(simplement) presque partout* ($f_n \xrightarrow{p.p.} f$) si
$$f_n(x) \to_n f(x) \quad \mu(dx) - p.p.;$$
2. *en norme* L^p ($f_n \xrightarrow{L^p} f$) si
$$\|f_n - f\|_{L^p} \to 0;$$
3. *presque uniformément* ($f_n \xrightarrow{p.u.} f$) si, pour tout $\epsilon > 0$, il existe un ensemble exceptionnel $A \in \mathcal{A}$ de mesure $\mu(A) \leqslant \epsilon$ en dehors duquel
$$f_n \to f \quad \text{uniformément};$$
4. *en mesure* ($f_n \xrightarrow{\mu} f$) si, pour tout $\epsilon > 0$,
$$\mu\left(\{x \in E;\ |f_n(x) - f(x)| \geqslant \epsilon\}\right) \to_n 0.$$

En Probabilités, la terminologie est la suivante :
— convergence presque partout (p.p.) = convergence *presque sûre (p.s.)*
— convergence L^1 = convergence *en moyenne*
— convergence L^2 = convergence *en moyenne quadratique*
— convergence en mesure = convergence *en probabilité.*

Rappelons la signification des deux cas les plus importants, dans ce cours, de convergence en norme L^p :
— $f_n \xrightarrow{L^1} f$ si $\int |f_n(x) - f(x)|\, d\mu \to 0$;
— $f_n \xrightarrow{L^\infty} f$ (on dit aussi *essentiellement uniformément*) si, pour tout $\epsilon > 0$, il existe un rang N tel que, pour tout $n \geqslant N$,
$$|f_n(x) - f(x)| \leqslant \epsilon \quad \mu(dx) - p.p.$$

Si l'on note Z_ϵ l'ensemble négligeable (a priori dépendant de ϵ) en dehors duquel $|f_n(x)-f(x)| \leqslant \epsilon$, on voit que, *a fortiori*, la même propriété est satisfaite en dehors de l'ensemble lui aussi négligeable mais de plus indépendant de ϵ, défini par $Z := \bigcup_{k\geqslant 1} Z_{1/k}$.

La notion suivante de convergence est plus faible que les précédentes (et ne requiert même pas que les fonctions f_n soient définie sur le même espace mesuré).

Définition 25.2. Une suite (ν_n) de mesures finies sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *converge étroitement* vers une mesure finie ν sur le même espace, ce qu'on note $\nu_n \xrightarrow{e} \nu$, si, pour toute fonction g continue bornée sur $\mathbb{R}$,

$$\int g\, d\nu_n \to \int g\, d\nu.$$

Une suite (f_n) de fonctions réelles boréliennes sur $(E, \mathcal{A}, \mu)$ *converge en loi (étroitement)* vers f, ce que l'on note $f_n \xrightarrow{\mathcal{L}} f$ (ou, abusivement, $f_n \xrightarrow{\mathcal{L}} f(\mu)$), si $f_n(\mu) \xrightarrow{e} f(\mu)$, i.e.

$$\int g \circ f_n\, d\mu \to \int g \circ f\, d\mu$$

pour toute fonction g continue bornée sur $\mathbb{R}$.

Théorème 25.3. *Les implications indiquées figure 31 sont toujours vraies.*

Ce labyrinthe d'implications est un enfer pour quiconque essayerait de l'apprendre par coeur sans avoir idée de la démonstration.

Sur cette figure, il est tacite que "$f_{n_k} \xrightarrow{p.p.} f$" signifie que *il existe* une suite extraite (f_{n_k}) de (f_n) qui converge vers f p.p.

Rappelons aussi une propriété déjà mentionée, qui n'est pas indiquée sur la figure : si $\mu(E) < \infty$, la convergence L^2 implique la convergence L^1, comme une simple application de l'inégalité de Cauchy-Schwarz le montre.

$$\begin{tikzcd}
 & f_n \xrightarrow{L^\infty} f \arrow[d, Rightarrow] & \\
f_{n_k} \xrightarrow{p.p.} f & f_n \xrightarrow{p.p.} f \arrow[l, Rightarrow] \arrow[r, Leftrightarrow, "\text{Egorov}", "\mu(E)<\infty"'] \arrow[d, Rightarrow] & f_n \xrightarrow{p.u.} f \\
 & f_n \xrightarrow{\mu} f \arrow[ul, Rightarrow] \arrow[r, Leftrightarrow, "|f_n|\leqslant g\in L^p"] \arrow[d, Rightarrow, "\mu(E)<\infty"] & f_n \xrightarrow{L^p} f \\
 & f_n \xrightarrow{\mathcal{L}} f \arrow[r, Leftrightarrow, "\mu(E)<\infty", "\text{Lévy}"'] & \Phi_{f_n} \xrightarrow{simpl.} \Phi_f .
\end{tikzcd}$$

FIGURE 31 – Le labyrinthe des convergences

Démonstration. Certaines de ces implications sont triviales :

— $f_n \xrightarrow{p.p.} f \Rightarrow$ il existe une suite extraite (f_{n_k}) telle que $f_{n_k} \xrightarrow{p.p.} f$:
N'importe quelle suite extraite convient, y compris la suite extraite triviale d'indices $n_k = k$.

— $f_n \xrightarrow{L^\infty} f \Rightarrow f_n \xrightarrow{p.u.} f$:
La suite (f_n) converge vers f uniformément, en dehors d'une partie négligeable Z de E. Donc, dans la définition de la convergence presque uniforme, il suffit de prendre trivialement $A_\epsilon := Z$ pour tout $\epsilon > 0$.

— $f_n \xrightarrow{p.u.} f \Rightarrow f_n \xrightarrow{p.p.} f$:
Pour tout $\epsilon > 0$, il existe une partie A_ϵ, de mesure $< \epsilon$, en dehors de laquelle f_n converge vers f uniformément, donc simplement. Donc f_n converge vers f simplement en dehors de $Z := \cap_{k\geqslant 1} A_{1/k}$, qui est négligeable puisque, par continuité extérieure de la mesure (ou par convergence monotone),

$$\mu(Z) = \mu(\cap_k A_{1/k}) = \lim \mu(A_{1/k}) \leqslant \lim \frac{1}{k} = 0.$$

— $f_n \xrightarrow{\mathcal{L}} f \Rightarrow \Phi_{f_n} \xrightarrow{simplement} \Phi_f$:
Pour tout $\xi \in \mathbb{R}$, la fonction $\mathbb{R} \to \mathbb{C}$, $y \mapsto e^{-iy\xi}$ est continue bornée, donc, par définition de la convergence en loi,

$$\Phi_{f_n}(\xi) = \int e^{-if_n\cdot\xi}\, d\mu \Longrightarrow \int e^{-if\xi}\, d\mu = \Phi_f(\xi).$$

D'autres implications ne sont pas triviales :

— $f_n \xrightarrow{p.p.} f \Rightarrow f_n \xrightarrow{p.u.} f$ (théorème de Severini-Egorov [1, 2]) :
Soit

$$C_{n,r} = \bigcup_{m\geqslant n} \left\{|f_m - f| > 2^{-r}\right\} \quad (n, r \in \mathbb{N}).$$

Fixons r. La suite $(C_{n,r})_n$ est décroissante et, comme (f_n) converge p.s. vers f, $P(\cap_n C_{n,r}) = 0$. Donc, d'après le théorème de convergence monotone,

$$\lim_n P(C_{n,r}) = 0.$$

Donc quel que soit $\epsilon > 0$ il existe $n = n(r,\epsilon)$ tel quel que

$$P(C_{n(r,\epsilon),r}) \leqslant \epsilon 2^{-r-1}.$$

L'ensemble

$$A_\epsilon := \bigcup_{r\in\mathbb{N}} C_{n(r,\epsilon),r}$$

est de mesure $< \epsilon$ et, en dehors de A_ϵ, (f_n) converge uniformément.

1. Carlo Severini, mathématicien italien (1872–1951), qui travailla en analyse réelle, théorie de la mesure et de l'intégration, et en théorie des équations aux dérivées partielles. Le théorème démontré dans cet exercice est souvent appelé *théorème d'Egorov*, mais il fut démontré par Severini en 1910, soit un an avant Egorov. C'est une illustration du "principe d'Arnold" que les théorèmes ne portent pas le nom de leur découvreur, ce principe s'appliquant d'ailleurs à lui-même.

2. Dmitri Fyodorovich Egorov, mathématicien russe et soviétique (1869–1931), célèbre pour ses travaux en analyse et géométrie différentielle.

— $f_n \xrightarrow{p.p.} f \Rightarrow f_n \xrightarrow{\mu} f$:
Soient $\epsilon > 0$ et, pour tout n, $A_n = \{x, |f_n(x) - f(x)| \geqslant \epsilon\}$. Il s'agit de montrer que $\mu(A_n) \to_n 0$. L'ensemble $\limsup A_n$ est l'ensemble des x tels que $|f_n(x) - f(x)| \geqslant \epsilon$ pour une infinité de valeurs de n ; pour de tels x, $f_n(x) \nrightarrow f(x)$. Donc, d'après l'hypothèse, $\mu(\limsup A_n) = 0$ et, d'après le lemme de Fatou,

$$\limsup \mu(A_n) \leqslant \mu\left(\limsup A_n\right) = 0.$$

Donc $\lim \mu(A_n) = 0$.

— $f_n \xrightarrow{L^p} \Rightarrow f_n \xrightarrow{\mu} f$:
Comme $\epsilon^p \mathbb{1}_{|f_n - f| \geqslant \epsilon} \leqslant |f_n - f|^p$,

$$\epsilon^p \mu(\{x, |f_n(x) - f(x)| \geqslant \epsilon\}) \leqslant \int |f_n - f|^p \, d\mu = \|f_n - f\|_{L^p}^p \to 0.$$

— $f_n \xrightarrow{\mu} f \Rightarrow f_{n_k} \xrightarrow{p.p.} f$:
Pour tout $\epsilon > 0$, $\mu(\{|f_n - f| > \epsilon\}) \to 0$. Par récurrence, on peut donc construire une suite strictement croissante (n_k) telle que

$$\mu\left(A_k\right) \leqslant 2^{-k}, \quad \text{avec } A_k = \{|f_{n_k} - f| > 2^{-k}\}.$$

On a

$$\sum_k \mu(A_k) = \sum 2^{-k} < \infty.$$

Donc

$$\begin{aligned} \mu(\limsup_k A_k) &= \mu(\cap_k \cup_{l \geqslant k} A_l) \\ &= \lim_k \mu(\cup_{l \geqslant k} A_l) \quad \text{(continuité extérieure)} \\ &\leqslant \lim_k \sum_{l \geqslant k} \mu(A_l) \quad (\sigma\text{-additivité}) \\ &= 0 \quad \text{(reste d'une série convergente)} \end{aligned}$$

(cette dernière partie du raisonnement est semblable à celle de la première partie de l'exercice 23.8 ; ici puisque la mesure est éventuellement infinie, l'égalité due à la continuité extérieure est justifiée par le fait que $\mu(\cup_l A_l) \leqslant \sum_l \mu(A_l) < \infty$). De plus, pour tout $x \notin \limsup A_k$ (donc, en dehors d'un ensemble négligeable), il existe $K \geqslant 1$ tel que, pour tout $k \geqslant K$, $x \notin A_k$, c'est-à-dire $|f_{n_k}(x) - f(x)| \leqslant 2^{-k}$. Donc (f_{n_k}) converge vers f p.p.

— $f_n \xrightarrow{L^p} f \Rightarrow f_n \xrightarrow{\mu} f$:
Pour tout $\epsilon > 0$,

$$\mu(\{|f_n - f| > \epsilon\}) \leqslant \frac{1}{\epsilon^p} \int |f_n - f|^p \, d\mu = \frac{1}{\epsilon^p} \|f_n - f\|_{L^p}^p \to 0.$$

— $(f_n \xrightarrow{\mu} f$ et $|f_n| \leqslant g \in L^p) \Rightarrow f_n \xrightarrow{L^p} f$:
D'après l'hypothèse de domination, $f_n \in L^p$ pour tout n. Nous allons montrer que $|f| \leqslant g$ p.p. (donc que $f \in L^p$), puis que $\|f_n - f\|_{L^p}$ tend vers 0.

Remarquons d'abord que, pour tout $\epsilon > 0$ et pour tout n,

$$\begin{aligned}\{|f| > g + \epsilon\} \subset \{|f| > |f_n| + \epsilon\} &= \{|f| - |f_n| > \epsilon\} \\ &\subset \{|f - f_n| > \epsilon\},\end{aligned}$$

donc

$$\mu(\{|f| > g + \epsilon\}) \leqslant \mu(\{|f - f_n| > \epsilon\}).$$

À la limite quand n tend vers l'infini,

$$\mu(\{|f| > g + \epsilon\}) \leqslant \lim \mu(\{|f - f_n| > \epsilon\}) = 0.$$

À la limite quand ϵ tend vers 0, d'après le théorème de convergence monotone,

$$\mu(\{|f| > g\}) = 0.$$

Donc $|f| \leqslant g$ p.p.
Maintenant, supposons par l'absurde que f_n ne tende pas vers f dans L^p. Il existe $\epsilon > 0$ et une suite (n_k) tendant vers l'infini telle que, pour tout k,

$$\int |f_{n_k} - f|^p \, d\mu \geqslant \epsilon.$$

Or, (f_{n_k}) converge vers f en probabilité, donc possède une suite extraite $(f_{n_{k_l}})$ qui converge vers f p.p. D'après le théorème de convergence dominée,

$$\int |f_{n_{k_l}} - f|^p \, d\mu \to 0,$$

ce qui est incompatible avec ce qui précède.

— $(f_n \xrightarrow{\mu} f$ et $\mu(E) < \infty) \Rightarrow f_n \xrightarrow{\mathcal{L}} f$:
Commençons par le cas facile où $f_n \xrightarrow{p.p.} f$. Pour toute fonction g continue bornée sur $\mathbb{R}$, $g \circ f_n$ converge vers $g \circ f$ p.p. et, g étant bornée (disons par un réel M), $|g \circ f_n| \leqslant M \in L^1(\mu)$ (nous avons supposé μ finie). D'après le théorème de convergence dominée, $\int g \circ f_n \, d\mu \to \int g \circ f \, d\mu$.
Dans le cas général, supposons par l'absurde qu'il existe une fonction g continue bornée sur $\mathbb{R}$ telle que $\int g \circ f_n \, d\mu \nrightarrow \int g \circ f \, d\mu$. Alors il existe $\epsilon > 0$ et une suite n_k strictement croissante telle que

$$\left| \int g \circ f_n \, d\mu - \int g \circ f \, d\mu \right| > \epsilon \quad (\forall k).$$

Mais d'après l'une des implications précédemment démontrée, comme $f_{n_k} \xrightarrow{\mu} f$, il existe une suite extraite $(f_{n_{k_l}})$ convergeant simplement p.p. vers f, ce qui est absurde d'après le "cas facile" ci-dessus.

— $\Phi_{f_n} \xrightarrow{simplement} f \Rightarrow f_n \xrightarrow{\mathcal{L}} f$ (théorème de Lévy) : admis.

□

Énonçons plus précisément le théorème de Lévy pour les probabilités, ici admis.

Théorème 25.4 (Lévy)**.** *Une suite (ν_n) de probabilités sur $\mathbb{R}^d$ converge étroitement vers une probabilité ν sur $\mathbb{R}^d$ si et seulement si $\widehat{\nu_n}$ converge simplement vers $\widehat{\nu}$.*
De même, une suite (X_n) de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$ converge en loi vers un vecteur aléatoire X si et seulement si Φ_{X_n} converge simplement vers Φ_X.

Si une formule d'inversion de la transformation de Fourier des probabilités existait, le théorème de Lévy (dans le sens difficile $\Leftarrow$) serait une sorte de théorème de continuité de la transformation de Fourier inverse (voir l'exercice 20.7 dans le cas où $\hat{\nu}, \hat{\nu}_n \in L^1(\mathbb{R}^d)$).

Bien que les modes de convergence diffèrent notablement les uns des autres, ceux de la définition 25.1 sont tous compatibles dans le sens qu'ils ne sont jamais en désaccord sur la fonction vers laquelle une suite (f_n) converge, en dehors d'un ensemble négligeable. Plus précisément :

Exercice ★ 25.1 (Unicité de la limite [Tao11])**.**
Soient f_n $(n \in \mathbb{N})$, f et g des fonctions mesurables $E \to \mathbb{R}$. Si (f_n) converge vers f dans l'un quelconque des sens de la définition 25.1 et vers g dans un autre de ces sens, $f = g$ p.p. Pourquoi cela n'est-il pas vrai avec la convergence en loi ?

Exercice 25.2 (Convergence de $\alpha_n \mathbb{1}_{A_n}$ [Tao11])**.**
Soient (α_n) une suite de nombres réels > 0 et (A_n) une suite de parties $\in \mathcal{A}$ de mesure > 0. Pour simplifier, supposons vérifiée l'alternative suivante :
— soit (α_n) tend vers 0
— soit il existe $c > 0$ tel que $\alpha_n \geqslant c$ pour tout n
(si (α_n) ne tend pas vers 0, la seconde propriété est vérifiée pour une sous-suite ; donc on ne perd pas beaucoup en généralité à faire cette hypothèse).

Nous allons déterminer si les fonctions $f_n := \alpha_n \mathbb{1}_{A_n}$ convergent vers 0, dans différents modes. Il s'avère que la réponse dépend plus ou moins des trois quantités suivantes :
— la *hauteur* α_n de la n-ième fonction f_n,
— la *largeur* $\mu(A_n)$ de la n-ième fonction f_n,
— le n-ième *support de queue* $A_n^* = \cup_{k \geqslant n} A_k$ de la suite (f_n) (de sorte que $\cap_n A_n^* = \limsup_n A_n$).
Établir les affirmations suivantes :

1. f_n converge uniformément vers 0 si et seulement si $\alpha_n \to 0$.

2. f_n converge en norme L^∞ vers 0 si et seulement si $\alpha_n \to 0$.

3. f_n converge presque uniformément vers 0 si et seulement si $\alpha_n \to 0$ ou $\mu(A_n^*) \to 0$.

4. f_n converge (simplement) vers 0 si et seulement si $\alpha_n \to 0$ ou $\cap A_n^* = \varnothing$.

5. f_n converge (simplement) vers 0 p.p. si et seulement si $\alpha_n \to 0$ ou $\mu(\cap A_n^*) = 0$.

6. f_n converge en mesure vers 0 si et seulement si $\alpha_n \to 0$ ou $\mu(A_n) \to 0$.

7. f_n converge en norme L^p vers 0 si et seulement si $\alpha_n^p \mu(A_n) \to 0$ $(p > 0)$.

Exercice 25.3 (Convergence L^1 rapide)**.**
Supposons que $f_n, f : E \to \mathbb{R}$ sont des fonctions mesurables telles que

$$\sum_n \|f_n - f\|_{L^1} < \infty$$

(non seulement la suite $\|f_n - f\|_{L^1} \to 0$, ce qui serait la convergence L^1, mais elle converge de sorte que la série dont cette suite est le terme général converge).

1. Montrer que (f_n) converge p.p. vers f.

Indication : Commencer avec une fonction indicatrice.

2. Montrer que (f_n) converge presque uniformément vers f (ce qui implique la propriété demandée dans la première question, mais qui est légèrement plus difficile à démontrer).

3. En déduire que la convergence en norme L^1 implique la convergence p.p. ou presque uniforme d'une sous-suite.

Exercice 25.4 (Sur la composition des limites)**.**
Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction continue.

1. Montrer que, si f_n converge vers f p.p., $g \circ f_n$ converge vers $g \circ f$ p.p.

2. Monter que, si μ est finie et si f_n converge vers f en mesure, $g \circ f_n$ converge vers $g \circ f$ en mesure.

Autres exercices

Exercice 25.5 (Théorème de McMillan, *examen 2011*)**.**
Soit $X = \{1, ..., r\}$ $(r > 0)$ muni d'une probabilité P. On notera $p_x = P(\{x\})$ si $x \in X$. On appelle *entropie* de X le réel positif

$$h = h(p_1, ..., p_r) = -\sum_{x \in X} p_x \ln p_x,$$

avec la convention $0 \ln 0 = 0$ dans les cas où $p_x = 0$ (voir l'exercice 7.11).

1. Montrer que h atteint son maximum quand $p_x = 1/r$ pour tout $x = 1, ..., r$, d'abord en supposant que $p_x > 0$ pour tout x (on pourra maximiser h sous la contrainte $p_1 + \cdots + p_r = 1$), puis dans le cas général.

Soit $\Omega = X^n$ l'espace des suites de n tirages indépendants dans X, muni de la probabilité produit. Soit $\nu_x : \Omega \to \{0, ..., n\}$ la variable aléatoire définie par $\nu_x(\omega) =$ le nombre de composantes de $\omega = (\omega_1, ..., \omega_n)$ égales à x.

2. Calculer la loi de ν_x en fonction des p_x, $x \in X$.

3. Un nombre $\delta > 0$ étant fixé, posons

$$C_n = \{\omega,\ |\nu_x(\omega)/n - p_x| \leqslant \delta\ (\forall x \in X)\}.$$

Quelle est la limite de $P(C_n)$ quand n tend vers l'infini ? On pourra utiliser la loi faible des grands nombres.

4. En déduire que, si $\alpha > 0$ et si δ est assez petit en fonction de α, pour tout entier n et pour tout $\omega \in C_n$,

$$e^{-n(h+\alpha)} \leqslant P(\{\omega\}) \leqslant e^{-n(h-\alpha)}.$$

5. En déduire enfin le théorème de McMillan : il existe $\beta > 0$ tel que, si n est assez grand, le cardinal de C_n satisfait

$$e^{n(h-\beta)} \leqslant |C_n| \leqslant e^{n(h+\beta)}.$$

Au regard des deux dernières questions, interpréter la place de C_n dans Ω.

Exercice 25.6 (Théorème de Weierstrass)**.**
Soit $f \in C([0,1])$. On définit le n-ième polynôme de Bernstein associé à f :

$$B_n(x) = \sum_{0 \leqslant k \leqslant n} C_n^k x^k (1-x)^{n-k} f\left(\frac{k}{n}\right).$$

On veut montrer que B_n converge vers f uniformément sur $[0,1]$. Soit $\epsilon > 0$.
1. Montrer que

$$B_n(x) - f(x) = \sum_{0 \leqslant k \leqslant n} C_n^k x^k (1-x)^{n-k} \left(f\left(\frac{k}{n}\right) - f(x)\right).$$

Soit $\delta > 0$. Notons

$$K_1 = \{k,\ |k/n - x| \leqslant \delta\} \quad \text{et} \quad K_2 = \{0, ..., n\} \backslash K_1,$$

et coupons en conséquence la somme obtenue dans la première question en

$$B_n(x) - f(x) = \Sigma_1 + \Sigma_2, \quad \Sigma_i = \sum_{k \in K_i} C_n^k x^k (1-x)^{n-k} \left(f\left(\frac{k}{n}\right) - f(x)\right)$$

2. Montrer que, si δ est assez petit, $|\Sigma_1| \leqslant \epsilon$.
On veut maintenant majorer Σ_2 convenablement. Notons $M = \sup_{[0,1]} |f|$. On a

$$\Sigma_2 \leqslant 2M \sum_{k : |k/n - x| \geqslant \delta} C_n^k x^k (1-x)^{n-k}.$$

Pour le problème déterministe consistant à majorer la somme du membre de droite, nous allons utiliser un argument probabiliste dû à Bernstein.
Soit $S_n = X_1 + \cdots + X_n$ la somme de n variables aléatoires indépendantes identiquement distribuées, de loi de Bernoulli de paramètre $x \in [0,1]$.
3. Montrer que

$$|\Sigma_2| \leqslant 2M \frac{x(1-x)}{\delta^2 n},$$

et en déduire que $|\Sigma_2| \leqslant \epsilon$ si n est assez grand ; pour montrer l'inégalité, on pourra appliquer l'inégalité de Bienaymé-Tchebychev à la variable aléatoire S_n/n.
Conclure.

Exercice 25.7 (Développements décimaux, *examen 2011*)**.**
Dans cet exercice, on définit la fonction caractéristique d'une variable aléatoire réelle X comme $\Phi_X(\xi) = E(e^{+iX\xi})$; ce changement de convention ne modifie pas les propriétés essentielles des fonctions caractéristiques. Soit $(X_n)_{n \geqslant 1}$ une suite de variables aléatoires réelles indépendantes définies sur le même espace de probabilité $(\Omega, \mathcal{A}, P)$, de même loi uniforme sur l'ensemble d'entiers $\{0, 1, 2, \cdots, 9\}$.
1. Calculer la loi et la fonction caractéristique communes des X_n.

Pour tout $n \geqslant 1$, on définit Y_n comme la somme partielle de rang n de la série $\sum_k \frac{X_k}{10^k}$ (une bonne référence sur les développements décimaux est l'appendice de [Tao09]) :

$$Y_n = \frac{X_1}{10} + \frac{X_2}{10^2} + \cdots + \frac{X_n}{10^n}.$$

2. Montrer que la suite (Y_n) converge simplement, vers une variable aléatoire que l'on notera Y. Converge-t-elle en probabilité ? En loi ?

3. Montrer que la fonction caractéristique de Y_n est

$$\Phi_{Y_n}(t) = \frac{1}{10^n}\frac{1-\exp(it)}{1-\exp(10^{-n}it)} \quad (\forall t \in \mathbb{R}) \ ;$$

on pourra commencer par supposer $t \notin 2\pi\mathbb{Z}$.

4. En déduire que la loi de Y est uniforme sur $[0,1]$; on pourra commencer par calculer la fonction caractéristique de Y.

Exercice ★ 25.8 (Mesure invariante par une application, *examen 2012*)**.**
Soient $(\Omega, \mathcal{A}, \mu)$ un espace de probabilité et T une transformation mesurable de Ω (c'est-à-dire une application mesurable $\Omega \to \Omega$, ce qu'on note encore $T : \Omega \circlearrowleft$) .

On note $T(\mu)$ la probabilité image de μ par T, et T^n la n-ième composée itérée de T, définie par récurrence par $T^0 = \mathrm{id}$, $T^{n+1} = T \circ T^n$ $(n \geqslant 0)$.

1. On suppose dans cette question qu'il existe un entier $k \geqslant 1$ et un point $x \in \Omega$ *k-périodique*, i.e. tel que $T^k(x) = x$. Montrer que, si δ_a est la mesure de Dirac en $a \in \Omega$, la mesure $\mu = \frac{1}{k}(\delta_x + \delta_{T(x)} + ... + \delta_{T^{k-1}(x)})$ est T-invariante (*i.e.* $T(\mu) = \mu$ sur $\mathcal{A}$).

2. On suppose de plus, dans cette question, que Ω est une partie compacte non vide de $\mathbb{R}^n$ (ou un espace métrique compact non vide) et T un homéomorphisme. On cherche à montrer qu'il existe une probabilité T-invariante sur Ω.

2.1. Donner un exemple de probabilité μ sur Ω.

Pour cette probabilité, on pose

$$\bar{\mu}_n = \frac{1}{n}(\mu + T(\mu) + \cdots + T^{n-1}(\mu)).$$

On admet que $(\bar{\mu}_n)$ possède au moins une valeur d'adhérence étroite $\bar{\mu}$ (c'est une conséquence du théorème de Banach-Alaoglu).

2.2. Conclure.

On dit que : une partie $A \in \mathcal{A}$ est *(presque) invariante* si $\mu((T^{-1}(A))\Delta A) = 0$ [3] ; la mesure μ est *ergodique* si toute partie invariante a pour mesure 0 ou 1.

3. Montrer que μ est ergodique si et seulement si toute fonction mesurable presque invariante ($f \circ T = f$ p.s.) est constante p.p.

4. Nous allons caractériser les mesures ergodiques en termes du spectre de la transformation unitaire U induite sur $L^2_{\mathbb{C}}(\Omega, \mathcal{A}, \mu)$ par T :

$$U : L^2_{\mathbb{C}}(\Omega, \mathcal{A}, \mu) \circlearrowleft, \quad f \mapsto f \circ T.$$

3. La *différence symétrique* de A et B est définie par $A\Delta B := (A\backslash B) \cup (B\backslash A)$

4.1. Supposons T ergodique. Montrer qu' une fonction propre de U, presque sûrement, est non nulle. En déduire que le spectre de U est un sous-groupe du cercle unité, et que toute valeur propre est simple.

4.2. En déduire que μ est ergodique si et seulement si 1 est valeur propre simple de U.

Exercice ★ 25.9 (Exemple de convergence en mesure)**.**
Soit (X_n) une suite de v.a.r.i.i.d. de $(\Omega, \mathcal{A}, P)$ dans $[0,1]$. Soit

$$Y_n = e^{-nX_1} + \cdots + e^{-nX_n}.$$

L'objectif est de calculer la limite de $P(Y_n < 1)$ quand n tend vers l'infini.

On définit la *transformée de Laplace* d'une v.a.r. positive Z ou d'une fonction $f : \mathbb{R}^+ \to \mathbb{R}$ comme les fonctions respectives

$$L(Z)(u) = \int_\Omega e^{-uZ}\, dP \quad \text{et} \quad L(f)(u) = \int_{\mathbb{R}^+} e^{-ux} f(x)\, dx,$$

là où elles existent ($u \in \mathbb{R}$).

1. Deux entiers $n, k \geqslant 1$ étant fixés, exprimer la transformée de Laplace de $Z = e^{-nX_k}$ comme une intégrale sur un intervalle de $\mathbb{R}$; en déduire une expression de $L(Y_n)(u)$.

Soit une fonction $f : \mathbb{R}^+ \to \mathbb{R}$ qui vaut 1 sur $[0,1]$, et qui vérifie

$$xf'(x) = -f(x-1)$$

presque partout sur $]1, +\infty[$. On admettra que sa transformée de Laplace existe sur $\mathbb{R}$.

2. Montrer que

$$L(f)(u) = e^\gamma \exp\left(\int_0^u \frac{e^{-s} - 1}{s}\, ds\right),$$

où γ est la constante d'Euler (définie dans l'exercice 8.6) ; pour calculer le facteur e^γ, on pourra calculer la limite de $uL(f)(u)$ de deux façons différentes.

3. Montrer que $L(Y_n)$ converge uniformément vers $e^{-\gamma} L(f)$.

On admettra ici pouvoir en déduire que Y_n converge en loi vers $e^{-\gamma} f$. On en déduit alors directement que la limite de $P(Y_n < 1)$ vaut $e^{-\gamma}$.

Chapitre 26

Sommes de variables indépendantes

Les sommes de variables aléatoires jouent un rôle important en probabilités, ne serait-ce que de part les probabilités empiriques (voir l'exercice 26.3). Dans ce chapitre, on étudie un résultat classique, dans le cas le plus simple, où les variables sont indépendantes.

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles $\in L^1$ indépendantes de même loi, d'espérance μ et d'écart-type σ ; les X_n sont dans L^2 si et seulement si $\sigma < \infty$. Posons

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}. \tag{26.1}$$

Par linéarité on a

$$E(\bar{X}_n) = \mu \quad \text{et} \quad \operatorname{var}(\bar{X}_n) = \frac{\sigma^2}{n};$$

pour la variance, on a utilisé l'indépendance de $X_1, ..., X_n$.

Exercice 26.1 (Loi faible des grands nombres)**.**
Supposons que $X_n \in L^2$.

1. Montrer avec une inégalité de Markov que

$$\bar{X}_n \xrightarrow[n\to\infty]{P} \mu.$$

2. Montrer que

$$\bar{X}_n \xrightarrow[n\to\infty]{L^2} \mu$$

(ce qui d'ailleurs implique le résultat de convergence de la question précédente).

Il est remarquable que la limite μ obtenue est déterministe (variable aléatoire constante). Ce sera encore le cas pour la loi forte des grands nombres (ce qui découle de la loi du $0-1$ de Kolmogorov). La loi faible peut se déduire de la loi forte (voir ci-dessous), mais cette déduction n'est pas évidente, et s'avère plus compliquée que la démonstration directe de la loi faible.

Théorème 26.1 (Loi forte des grands nombres, Kolmogorov)**.** *Si* $X_n \in L^1$,

$$\bar{X}_n \xrightarrow{p.s.} \mu.$$

La démonstration suivante utilisera par deux fois le critère suivant de convergence presque sûre vers 0 :

Si (Z_n) est une suite de variables aléatoires réelles positives telle que la série $\sum E(Z_n)$ converge, $Z_n \to 0$ p.s. En effet, d'après le théorème de convergence monotone,

$$E\left(\sum Z_n\right) = \sum E(Z_n) < \infty,$$

donc $\sum Z_n < \infty$ p.s., donc $Z_n \to 0$ p.s.

Démonstration dans le cas où $X_n \in L^2$. Si l'on pose $Y_n = X_n - \mu$, il s'agit de montrer que

$$\bar{Y}_n := \frac{Y_1 + \cdots + Y_n}{n} \xrightarrow{\text{p.s.}} 0.$$

D'après la loi faible des grands nombres, $\bar{Y}_n$ tend vers 0 en norme L^2. On sait que donc il existe une suite extraite (Y_{n_k}) qui converge presque sûrement vers 0 (théorème 25.3), mais il est difficile d'en déduire quelque chose sur la suite (Y_n) entière. Ici, on peut être plus précis, et trouver une suite extraite explicite.

1. *Une suite extraite convergeant p.s.* On a

$$E(\bar{Y}_n^2) = \operatorname{var}(\bar{X}_n) = \frac{\sigma^2}{n},$$

donc, si les termes de la suite dont les indices sont de la forme $n = k^2$, avec $k \geqslant 1$, forment une série convergente :

$$\sum E(\bar{Y}_{k^2}^2) < \infty,$$

donc, d'après la remarque précédant la démonstration appliquée aux $Z_k = Y_{k^2}^2$,

$$\bar{Y}_{k^2} \xrightarrow{\text{p.s.}} 0.$$

Il reste maintenant à contrôler les termes intermédiaires de la suite $\bar{Y}_n$, précisément ceux qui ne sont pas de la forme $n = k^2$, en le comparant au terme d'indice k^2 le plus proche.

2. *Contrôle des termes intermédiaires.* Pour tout $n \in \mathbb{N}$, notons

$$q(n) = \max\{k^2 \leqslant n,\ k \in \mathbb{N}\}$$

le plus grand carré d'entier, $\leqslant n$; on utilisera le fait que

$$q(n) \leqslant n \leqslant (\sqrt{q(n)} + 1)^2 = q(n) + 2\sqrt{q(n)} + 1.$$

Montrons que

$$\Delta_n := \bar{Y}_n - \frac{q(n)}{n}\bar{Y}_{q(n)} = \frac{1}{n} \sum_{q(n)+1 \leqslant j \leqslant n} Y_j$$

tend vers 0 p.s. On a

$$E\left(\Delta_n^2\right) = \frac{n - q(n)}{n^2}\sigma^2 \leqslant \frac{2\sqrt{q(n)} + 1}{n^2}\sigma^2 \leqslant \frac{3\sigma^2}{n^{3/2}},$$

donc, d'après la remarque précédant la démonstration appliquée aux $Z_n = \Delta_n^2$, $\Delta_n \to 0$ p.s., c'est-à-dire

$$\bar{Y}_n - \frac{q(n)}{n} Y_{q(n)} \to 0 \quad \text{p.s.},$$

où $q(n)/n \leqslant 1$ et $Y_{q(n)} \to 0$ p.s. La conclusion en découle. □

Exemple 26.2 (Méthode de Monte-Carlo[1]). Soient f une fonction réelle mesurable bornée sur $[0,1]$ et (U_n) une suite de variables aléatoires réelles indépendantes de loi uniforme dans $[0,1]$. Alors

$$\frac{1}{n}(f(U_1) + \cdots + f(U_n)) \xrightarrow{p.s.} E(f(U_1)) = \int_0^1 f(x)\,dx.$$

De plus, d'après l'inégalité de Bienaymé-Tchebytchev,

$$P\left(\left|\frac{f(U_1) + \cdots + f(U_n)}{n} - \int_0^1 f(x)\,dx\right| \geqslant \epsilon\right) \leqslant \frac{\operatorname{var}(X_1)}{n\epsilon^2} = \frac{1}{12n\epsilon^2}.$$

Cette méthode de calcul d'une intégrale est surtout intéressante si f est peu régulière, et encore plus pour des intégrales sur des espaces de grande dimension.

Exercice 26.2 (Grandes déviations).
Pour tout entier n, soit S_n une variable aléatoire de loi binomiale $B(n,p)$. On veut montrer que, pour tout $\epsilon > 0$ assez petit, il existe $H > 0$ tel que, pour tout n,

$$P\left(\left|\frac{S_n}{n} - p\right| \geqslant \epsilon\right) \leqslant e^{-nH}.$$

1. Montrer que, pour tout $a \geqslant 0$,

$$P\left(\frac{S_n}{n} \geqslant a\right) \leqslant \exp\left(-n \sup_{s \geqslant 0}(as - \ln(1 - p + pe^s))\right).$$

2. Montrer que, pour tout $\epsilon > 0$ assez petit, il existe $h > 0$ tel que, pour tout n,

$$P\left(\frac{S_n}{n} \geqslant p + \epsilon\right) \leqslant e^{-nh}.$$

Prouver l'estimation symétrique et conclure.

Exercice 26.3 (Théorème fondamental de la statistique).
Soit $(X_n)_{n \geqslant 1}$ une suite de vecteurs aléatoires sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^d$, deux à deux indépendants et de même loi m sur $\mathbb{R}^d$.

Pour tout $\omega \in \Omega$ et tout $n \geqslant 1$, la *probabilité empirique* $m_n(\omega, \cdot)$ est la probabilité définie sur $\mathbb{R}^d$ par

$$m_n(\omega, \cdot) = \frac{1}{n}\left(\delta_{X_1(\omega)} + \cdots + \delta_{X_n(\omega)}\right),$$

de sorte que, pour tout $A \in \mathcal{B}(\mathbb{R}^d)$, $m_n(A)$ est le nombre d'indices $k \in \{1, ..., n\}$ tels que $X_k(\omega) \in A$. Montrer que (m_n) converge étroitement vers m P-p.s. (définition 25.2).

1. Du nom de l'un des quartiers de la principauté de Monaco, célèbre pour son futile casino.

Exercice 26.4 (Inégalité de Jensen par la loi des grands nombres).
Soit $\varphi :]a,b[\to \mathbb{R}$ une fonction convexe (a et b pouvant éventuellement être infinis) ; on rappelle que, pour tous $x_1, ..., x_n \in \mathbb{R}$,

$$\varphi\Big(\frac{x_1 + \cdots + x_n}{n}\Big) \leqslant \frac{\varphi(x_1) + \cdots + \varphi(x_n)}{n}$$

(ceci est même une définition possible de la convexité). Soit de plus $X : \Omega \to]a,b[$ une variable aléatoire réelle telle que X et $\varphi(X)$ soient intégrables. Montrer que

$$\varphi(E(X)) \leqslant E(\varphi(X))$$

en utilisant la loi forte des grands nombres ; on admettra qu'il existe une suite (X_n) de copies de X indépendantes (voir l'exemple 23.7).

L'exercice suivant montre que la loi (faible ou forte) des grands nombres peut être violée par des variables aléatoires non intégrables (« à queue épaisse »).

Exercice 26.5 (Somme de variables de Cauchy, *examen de juillet 2014*).
Soient U une variable aléatoire uniforme sur $]-\pi/2, \pi/2[$ et $\alpha \in \mathbb{R}\backslash\{0\}$.

1. Trouver la loi de $C = \alpha \tan U$ et, si elles existent, son espérance et sa variance ; C s'appelle une *variable de Cauchy* de paramètre α.

2. Calculer la transformée de Fourier inverse de la fonction $\varphi : \mathbb{R} \to \mathbb{R}$, $t \mapsto e^{-|t|}$ et en déduire que la fonction caractéristique de C est $\Phi_C(t) = e^{-\alpha|t|}$.

Soit $(X_n)_{n\geqslant 1}$ une suite de variables de Cauchy indépendantes de paramètre α.

3. Que la loi des grands nombres nous apprend-elle sur la limite, quand n tend vers l'infini, de $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$?

4. Déterminer, pour $n \geqslant 1$, les lois de $\bar{X}_n$ et de $\bar{X}_{2n} - \bar{X}_n$.

5. La suite $(\bar{X}_n)$ converge-t-elle dans L^1 ? En probabilités ? Presque sûrement ?

Problème 26.1 (Une loi forte facile).

1. Soient $X_1, X_2, ...$ des variables aléatoires indépendantes centrées telles qu'il existe $C > 0$ telle que

$$E(X_n^4) \leqslant C < \infty \quad (\forall n);$$

en particulier, $X_n \in L^4$, mais on ne suppose pas que les X_n sont identiquement distribuées.

Montrer que

$$\bar{X}_n := \frac{X_1 + \cdots + X_n}{n} \to 0 \text{ p.s.};$$

on pourra montrer que $\sum \bar{X}_n^4$ converge p.s.

2. Soient $X_1, X_2, ...$ des variables aléatoires indépendantes de même espérance μ et telles qu'il existe $C > 0$ telle que

$$E(X_n^4) \leqslant C < \infty \quad (\forall n).$$

Montrer que

$$\frac{X_1 + \cdots + X_n}{n} \to \mu \text{ p.s.};$$

pour vérifier que $(E((X_n - \mu)^4))$ est bornée, on pourra utiliser l'inégalité de Minkowski, ou une inégalité élémentaire.

Problème ★ 26.2 (Le théorème ergodique de Birkhoff (1931))**.**
Soient $(E, \mathcal{E})$ un espace mesurable, $f : E \to E$ une application mesurable et μ une mesure de probabilité de E.

On suppose que μ est invariante : pour toute partie $A \in \mathcal{E}$, on a $f_*\mu(A) := \mu(f^{-1}(A)) = \mu(A)$. On suppose aussi que μ est ergodique : pour toute partie $A \in \mathcal{E}$ telle que $f^{-1}(A) = A$ on a $\mu(A) = 0$ ou 1.

Soit enfin $\psi : E \to \mathbb{R}$ une fonction μ-intégrable. Le but du problème est de montrer le théorème de Birkhoff :[2] pour μ-presque tout point $x \in E$ la moyenne des valeurs de ψ le long de l'orbite positive $x, f(x), f(f(x)), \ldots$ de x converge vers la moyenne de ψ :

$$\frac{1}{n}\sum_{k=0}^{n-1} \psi(f^k(x)) \to \int_E \psi \, d\mu \quad \text{quand } n \to +\infty.$$

1. Interpréter ce résultat lorsque ψ est la fonction indicatrice d'une partie $B \in \mathcal{E}$.

2. Dans le cas où E est l'ensemble fini $\{1, \ldots, p\}$, où $\mathcal{E} = \mathcal{P}(E)$, et où μ est la probabilité uniforme ($\mu(\{k\}) = 1/p$ quel que soit $k \in \{1, \ldots, p\}$), caractériser les permutations f pour lesquelles μ est ergodique.

3. En quoi le théorème de Birkhoff renforce-t-il le théorème de récurrence de Poincaré ?

Commençons par considérer une fonction auxiliaire $\varphi : E \to \mathbb{R}$ qui soit μ-intégrable. Notons

$$S_n(x) = \sum_{k=0}^{n-1} \varphi \circ f^k(x) \quad \text{et} \quad F_n(x) = \max\left(S_1(x), \ldots, S_n(x)\right).$$

4. Montrer que pour tout x on a

$$F_{n+1}(x) = \varphi(x) + \max\left(0, F_n \circ f(x)\right). \tag{26.2}$$

5. Montrer que la partie $A = \{x, F_n(x) \to +\infty\}$ est mesurable.

6. Montrer que A est invariante ($f^{-1}(A) = A$) en utilisant l'égalité (26.2). En déduire que A est négligeable ou de mesure pleine.

7. En utilisant encore l'identité (26.2), montrer que sur A la suite des fonctions $F_{n+1} - F_n \circ f$ tend simplement vers φ ; puis justifier soigneusement que

$$0 \leqslant \int_A (F_{n+1} - F_n)\, d\mu = \int_A (F_{n+1} - F_n \circ f)\, d\mu \to_{n\to+\infty} \int_A \varphi \, d\mu;$$

8. En déduire que si $\int_E \varphi \, d\mu < 0$, $\mu(A) = 0$.

9. En déduire que si $\int_E \varphi \, d\mu < 0$,

$$\limsup_n \frac{S_n}{n} \leqslant 0 \quad \mu\text{-presque partout.}$$

2. George David Birkhoff, mathématicien américain (1884–1944), célèbre pour ses travaux en théorie des systèmes dynamiques.

10. Montrer que quel que soit $\varepsilon > 0$ on a

$$\limsup_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \psi \circ f^k \leqslant \int_E \psi \, d\mu + \varepsilon \quad \mu\text{-presque partout ;}$$

on pourra appliquer la question (9) à la fonction $\varphi = \psi - \int_E \psi d\mu - \varepsilon$.

Le même raisonnement appliqué à $-\psi$ montre que quel que soit $\varepsilon > 0$ on a

$$\liminf_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \psi \circ f^k \geqslant \int_E \psi \, d\mu - \varepsilon \quad \mu\text{-presque partout.}$$

En faisant tendre ε vers 0 on voit que les limites inférieure et supérieure coïncident μ-presque partout, ce qui prouve le théorème annoncé.

11. Montrer que, pour une transformation uniquement ergodique, la somme de Birkhoff converge *partout* vers la moyenne.

Voici deux applications du théorème ergodique de Birkhoff à la théorie des nombres.

Considérons le cas particulier où E est l'intervalle $[0, 1[$ muni de la tribu borélienne, $f : [0, 1[\to [0, 1[$ l'application

$$x = 0, x_1x_2x_3... \mapsto 0, x_2x_3x_4... = 10\, x \pmod 1 = \text{partie décimale de } 10\, x,$$

et ψ la fonction indicatrice de la partie $B = [0, 1/10[$.

12. Montrer que la mesure de Lebesgue λ est invariante.

13. Montrer qu'il existe une orbite dense, c'est-à-dire qu'il existe un $x \in [0, 1[$ tel que l'adhérence de l'ensemble $\{f^k(x)\}_{k\in\mathbb{N}}$ égale $[0, 1[$.

On admettra qu'alors λ est ergodique.

14. Caractériser les nombres $x = 0, x_1x_2x_3...$ tels que $f^k(x) \in [0, 1/10[$, puis interpréter en une phrase l'affirmation du théorème de Birkhoff dans cette situation.

15. ★ Montrer que le chiffre de gauche du nombre 2^n, $n \in \mathbb{N}$, en base 10 est plus souvent un 7 qu'un 8. On pourra admettre que, si $f : [0, 1[\to [0, 1[$ est la rotation $x \mapsto x + \alpha \pmod 1$, avec $\alpha \notin \mathbb{Q}$, alors f est ergodique ; ceci peut se démontrer en utilisant le théorème d'injectivité de la transformation de Fourier des mesures finies.

Problème 26.3 (Décalage de Bernoulli)**.**
Soient S un ensemble fini et $\Omega = S^{\mathbb{N}}$ l'ensemble des suites de S. Soient

$$\pi_k : \Omega \to S, \quad \omega = (\omega_0, \omega_1, ...) \mapsto \pi_k(\omega) = \omega_k$$

la k-ième projection canonique.

On rappelle (exercice 13.1) que Ω peut être muni d'une mesure produit, définie sur la tribu engendrée par l'ensemble $\mathcal{C}$ des cylindres de rang quelconque.

Soit

$$T : \Omega \circlearrowleft, \quad \omega = (\omega_0, \omega_1, ...) \mapsto T\omega = (\omega_1, \omega_2, ...)$$

le *décalage de Bernoulli* ; T "perd" le premier élément de ω et décale les autres d'un rang vers la gauche : $(T\omega)_k = \omega_{k+1}$, soit $\pi_k \circ T = \pi_{k+1}$ pour tout $k \in \mathbb{N}$.

1. Montrer que T est mesurable.

2. Montrer que P est invariante (par T : $T(P) = P$).

3. Montrer que T est ergodique, c'est-à-dire que les seules parties A telles que $T^{-1}(A) = A$ sont de probabilité 0 ou 1 ; on pourra commencer par considérer le cas d'un cylindre, puis en déduire ou admettre que la propriété est vérifiée plus généralement pour toute partie $A \in \sigma(\mathcal{C})$.

Soit $X : S \to \mathbb{R}$ une variable aléatoire discrète. Pour tout $k \in \mathbb{N}$, on définit

$$X_k = X \circ \pi_k : \Omega \to \mathbb{R}, \quad \omega \mapsto X(\omega_k).$$

4. Vérifier que $X_k \circ T = X_{k+1}$.

5. Montrer que les variables aléatoires X_k sont indépendantes et identiquement distribuées.

6. Montrer comment, dans ce cas, le théorème ergodique de Birkhoff implique la loi forte des grands nombres :

$$\frac{1}{n} \sum_{0 \leqslant k \leqslant n-1} X_k \to E(X_1) \quad \text{p.s.}$$

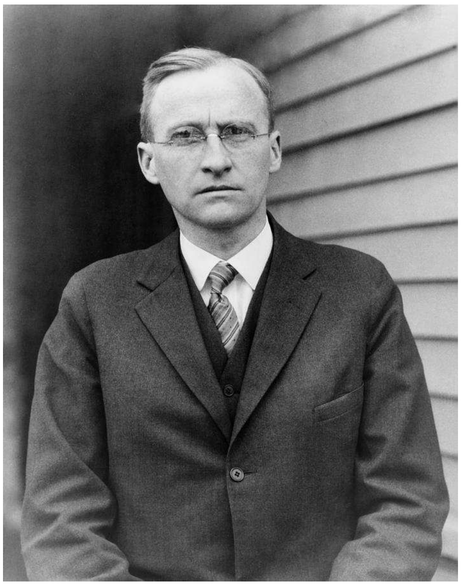


FIGURE 32 – George David BIRKHOFF (1884–1944)

Chapitre 27

Théorème central de la limite

La première version du théorème central de la limite fut postulée par DE MOIVRE en 1733, qui, dans un article remarquable, comprit que la distribution du résultat d'un jeu de pile ou face peut être approchée par la loi normale. En particulier, et contrairement à la situation décrite par la loi des grands nombres, la limite ici n'est pas déterministe. L'importance de cette découverte ne fut pas comprise de son temps. Le résultat fut même oublié, jusqu'à ce que LAPLACE le remît au goût du jour dans son traité monumental sur la *Théorie Analytique des Probabilités*, publié en 1812. Laplace élargit la portée du théorème en considérant plus généralement la loi binomiale. Ce travail reçut lui-même peu d'attention. C'est à la fin du XIXe siècle que LYAPUNOV en donna l'énoncé actuel et le démontra rigoureusement.

Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles $\in L^2$ indépendantes identiquement distribuées. Notons $\mu = E(X_1)$, $\sigma = \sigma_{X_1}$ et

$$S_n = X_1 + \cdots + X_n.$$

Avec ces données, la loi forte des grands nombres affirme que

$$\frac{S_n}{n} - \mu \xrightarrow{\text{p.s.}} 0.$$

Mais à quelle vitesse cette convergence a-t-elle lieu ? Par exemple, on pourrait se demander naïvement s'il existe un réel $\alpha > 0$ tel que

$$n^\alpha \left(\frac{S_n}{n} - \mu \right) \to C,$$

où C est une constante non nulle ; la multiplication par n^α fonctionne comme une loupe pour la quantité entre parenthèses.

Exercice ★ 27.1 (Théorème de Moivre)**.**
Supposons que les X_n sont des variables de Bernoulli de paramètre 1/2, indépendantes.

1. Montrer que, pour tous $a < b$,

$$2^{-n} \sum_{\frac{n}{2}+a\sqrt{n}\leqslant k\leqslant \frac{n}{2}+b\sqrt{n}} C_n^k \to_{n\to\infty} \sqrt{\frac{2}{\pi}} \int_a^b e^{-2x^2}\, dx;$$

on admettra que, si une suite (ν_n) de probabilités sur $\mathbb{R}$ converge étroitement vers une probabilité ν, pour tous $a < b$ on a

$$\int_a^b d\nu_n \to \int_a^b d\nu$$

(ce qui n'est pas trivial parce que $\mathbb{1}_{[a,b]} \notin C_b^0(\mathbb{R})$).

2. Montrer, plus précisément, le théorème de Moivre : [1]

$$\sqrt{n}2^{-n}C_n^k = \sqrt{\frac{2}{\pi}}\exp\left(-\frac{2}{n}\left(k-\frac{n}{2}\right)^2\right) + o(1)$$

en utilisant la formule de Stirling.

D'après cet exemple, il s'avère que la réalité est plus compliquée, pour deux raisons :

- La limite n'est jamais déterministe (c'est-à-dire constante). Autrement dit, le miracle de la loi forte ne se reproduit à aucune échelle α.
- La limite doit être prise dans un sens considérablement plus faible que celui d'une limite simple, en l'occurence dans le sens de la convergence en loi.

Ceci étant, pour $\alpha = 1/2$ il se produit un fait absolument remarquable : $\sqrt{n}\left(\frac{S_n}{n} - \mu\right)$ converge en loi vers une loi normale ! En ce sens, la vitesse de convergence est donc $\sqrt{n}$.

Théorème 27.1.

$$\frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - \mu\right) \xrightarrow{\mathcal{L}} Y \sim N(0,1).$$

En particulier, la loi de $\frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - \mu\right)$ converge étroitement vers $N(0,1)$.

Démonstration. La fonction caractéristique de $Y_n := \frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - \mu\right)$ vaut

$$\begin{aligned}
\Phi_{Y_n}(\xi) &= E\left(\exp i\frac{S_n - n\mu}{\sigma\sqrt{n}}\xi\right) \\
&= E\left(\exp i(S_n - n\mu)\frac{\xi}{\sigma\sqrt{n}}\right) \\
&= \Phi_{S_n - n\mu}\left(\frac{\xi}{\sigma\sqrt{n}}\right) \\
&= \prod_{1\leqslant j\leqslant n}\Phi_{X_j-\mu}\left(\frac{\xi}{\sigma\sqrt{n}}\right) = \left(\Phi_{X_1-\mu}\left(\frac{\xi}{\sigma\sqrt{n}}\right)\right)^n,
\end{aligned}$$

où

$$\phi(u) := \Phi_{X_1-\mu}(u) = 1 - \sigma^2\frac{u^2}{2} + o(u^2)$$

1. Abraham de Moivre, mathématicien français (1667–1754), connu pour ses contributions à la théorie des probabilités et pour la formule de Moivre en analyse complexe.

(puisque $E(X_1 - \mu) = 0$ et $\mathrm{var}(X_1 - \mu) = \sigma^2$). Donc

$$\Phi_{Y_n}(\xi) = \exp\left[n \ln\left(1 - \frac{\xi^2}{2n} + o\left(\frac{1}{n}\right)\right)\right] \to e^{-\xi^2/2} = \Phi_Y(\xi)$$

si Y est une variable quelconque de loi $N(0,1)$. D'après le théorème de Lévy (théorème 25.4), ceci implique que Y_n converge en loi vers Y. □

Exercice 27.2 (Démonstration probabiliste de la formule de Stirling [LM06])**.** Supposons que les X_n soient indépendantes, et de loi de Poisson de paramètre $\lambda = 1$:

$$P(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!} = \frac{1}{ek!} \quad (k \geqslant 0).$$

1. Vérifier que

$$E\left(\left(\frac{S_n - n}{\sqrt{n}}\right)^-\right) = \frac{n^{n+\frac{1}{2}}}{e^n n!},$$

où le signe négatif en exposant désigne la partie négative de la variable aléatoire entre parenthèses.

2. Montrer que

$$\frac{S_n - n}{\sqrt{n}} \xrightarrow{\mathcal{L}} Y,$$

où Y est de loi $N(0,1)$.

3. En déduire que

$$E\left(\left(\frac{S_n - n}{\sqrt{n}}\right)^-\right) \to E(Y^-) = \frac{1}{\sqrt{2\pi}};$$

on admettra que $\left(\frac{S_n-n}{\sqrt{n}}\right) \xrightarrow{\mathcal{L}} Y^-$.

4. En déduire que, quand n tend vers l'infini,

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Théorème 27.2 (Théorème central de la limite vectoriel)**.** *Si $(X_n)_{n\geqslant 1}$ est une suite de variables aléatoires $\in L^2_{\mathbb{R}^d}(\Omega, \mathcal{A}, P)$ indépendantes de même loi,*

$$\sqrt{n}\left(\frac{X_1 + \cdots + X_n}{n} - E(X_1)\right) \xrightarrow[n\to\infty]{\mathcal{L}} N\left(0, K_{X_1}\right).$$

Remarque 27.3. On ne peut pas trivialement déduire cette version vectorielle du théorème en appliquant la version réelle à chaque composante des X_n parce que la loi conjointe de la limite, à supposer qu'elle existe, n'est pas déterminée par les lois marginales.

Exercice : se convaincre que pourtant la démonstration est la même, *mutatis mutandis*, que dans le cas scalaire.

Chapitre 28

Le théorème ergodique de von Neumann ★

Ce chapitre est une incartade en *théorie ergodique*, la branche de la théorie des systèmes dynamiques qui a pour objet de justifier l'utilisation de concepts probabilistes dans l'étude des systèmes dynamiques déterministes. Précisément, on démontre ici le théorème ergodique de von Neumann, qui généralise la loi (faible) des grands nombres à certaines suites de variables aléatoires éventuellement dépendantes, possédant une interprétation dynamique. La conclusion du théorème fait surgir l'*espérance conditionnelle*, qui sera étudiée plus en détail dans les chapitres ultérieurs. [1]

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $T : \Omega \circlearrowleft$ une *transformation mesurable* de Ω, c'est-à-dire une application mesurable $(\Omega, \mathcal{A}) \to (\Omega, \mathcal{A})$.

Définition 28.1. Un événement $A \in \mathcal{A}$ est *presque invariant* si les fonctions indicatrices $\mathbb{1}_A$ et $\mathbb{1}_{T^{-1}(A)}$ sont égales presque sûrement. [2]

Lemme 28.2. *L'ensemble $\mathcal{I}$ des événements presque invariants est une tribu ; $\mathcal{I}$ s'appelle la* tribu presque invariante *de T.*

Démonstration. Soit A presque invariante. On a

$$\begin{aligned}
\mathbb{1}_{A^c} &= 1 - \mathbb{1}_A \\
&= 1 - \mathbb{1}_{T^{-1}(A)} \quad \text{p.s.} && \text{(car } \mathbb{1}_A = \mathbb{1}_{T^{-1}(A)} \quad \text{p.s.)} \\
&= \mathbb{1}_{T^{-1}(A)^c} \\
&= \mathbb{1}_{T^{-1}(A^c)} && \text{(car } T^{-1}(A^c) = T^{-1}(A)^c\text{),}
\end{aligned}$$

donc A^c est presque invariante.

Soient $A_1, A_2, \ldots$ une famille dénombrable de parties presque invariantes. Pour tout i, il existe une partie N_i de probabilité nulle telle que

$$A_i \cap N_i^c = T^{-1}(A_i) \cap N_i^c.$$

1. La généralisation analogue de la loi forte des grands nombres est le théorème ergodique de Birkhoff, démontré dans les exercices 26.2 et 30.2.

2. De façon équivalente, A est presque invariant si la différence symétrique $A \Delta T^{-1}(A)$ est de mesure nulle.

La partie $N = \cup N_i$ est aussi de probabilité nulle, et

$$(\cup_i A_i) \cap N^c = \cup_i (A_i \cap N^c) = \cup_i (T^{-1}(A_i) \cap N^c) = T^{-1}(\cup_i A_i) \cap N^c,$$

donc

$$\mathbb{1}_{\cup_i A_i} = \mathbb{1}_{T^{-1}(\cup_i A_i)} \quad \text{p.s.},$$

donc $\cup_i A_i$ est presque invariante. Donc $\mathcal{I}$ est une tribu. □

Exemple 28.3. Supposons Ω fini : $\Omega = \{x_1, ..., x_n\}$, muni de la probabilité uniforme. Si T est une permutation circulaire, par exemple[3]

$$\begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ 2 & 3 & \cdots & n & 1 \end{pmatrix},$$

ses seules parties invariantes sont $\varnothing$ et Ω. Comme P est uniforme, les parties presque invariantes de T sont en fait invariantes. Alors $\mathcal{I} = \{\varnothing, \Omega\}$.

Exercice 28.1 (Complétion de la tribu invariante)**.**
Montrer que la complétion de la tribu invariante est la tribu presque invariante.

Lemme 28.4. *Une variable aléatoire réelle f sur $(\Omega, \mathcal{A})$ est presque invariante au sens que $f \circ T = f$ p.s., si et seulement si f est $\mathcal{I}$-mesurable.*

Démonstration. Supposons que $f \circ T = f$ p.s. et soit $B \in \mathcal{B}(\mathbb{R})$. Il s'agit de montrer que $f^{-1}(B)$ est presque invariante. Soit N une partie de probabilité nulle en dehors de laquelle $f \circ T = f$. En restriction à N^c,

$$\mathbb{1}_{f^{-1}(B)} = \mathbb{1}_{(f \circ T)^{-1}(B)} = \mathbb{1}_{T^{-1}(f^{-1}(B))},$$

donc $f^{-1}(B)$ est presque invariante. Donc f est $\mathcal{I}$-mesurable.

Réciproquement, supposons que f soit $\mathcal{I}$-invariante. Dans le cas où f est une fonction étagée positive, il existe des parties $A_1, ..., A_n \in \mathcal{I}$ et des réels $a_1, ..., a_n > 0$ tels que

$$f = \sum_i a_i \mathbb{1}_{A_i}.$$

Alors

$$f \circ T = \sum_i a_i \mathbb{1}_{T^{-1}(A_i)} = \sum_i a_i \mathbb{1}_{A_i} = f \quad \text{p.s..}$$

Si f est positive, il existe une suite croissante (f_n) de fonctions étagés positives $\mathcal{I}$-mesurables, qui converge vers f. D'après le cas précédent, pour tout i,

$$f_i \circ T = f_i \quad \text{p.s..}$$

et l'égalité est réalisée en dehors d'une partie négligeable N_i. En dehors de la partie elle-aussi négligeable $N = \cup_i N_i$, $f_i = f_i \circ T$ pour tout i, donc, en passant à la limite quand i tend vers l'infini, $f = f \circ T$ en dehors de N. Donc $f = f \circ T$ p.s.

Si enfin f est de signe quelconque, ses parties positives et négatives sont presque invariantes d'après le cas précédent. Donc f aussi. □

3. Cette notation est classique pour une permutation. Elle signifie que $T(1) = 2$, $T(2) = 3$, etc. L'adjectif "circulaire" signifie que les orbites $\{T^k(x_i)\}_{k \in \mathbb{N}}$ sont Ω entier, i.e. il n'y a qu'une seule orbite.

On suppose dorénavant que $f \in L^2(\mathcal{A}) = L^2(\Omega, \mathcal{A}, P)$. On rappelle que $L^2(\mathcal{A})$ possède un produit scalaire hilbertien réel $\langle g|h \rangle = \int g\, h\, dP$.

Lemme 28.5. *$L^2(\mathcal{I}) = L^2(\Omega, \mathcal{I}, P)$ s'identifie à un sous-espace fermé de $L^2(\mathcal{A})$.*

Démonstration. $L^2(\mathcal{I})$ est un espace vectoriel au même titre que $L^2(\mathcal{A})$. Comme $\mathcal{I}$ est une sous-tribu de $\mathcal{A}$, une fonction $\mathcal{I}$-mesurable est automatiquement $\mathcal{A}$-mesurable. Donc toute classe de fonctions $\mathcal{I}$-mesurables (pour la relation d'égalité presque sûre) est incluse dans une unique classe de fonctions $\mathcal{A}$-mesurables. Donc il existe une application $i : L^2(\mathcal{I}) \to L^2(\mathcal{A})$, qui est injective. De plus, le produit scalaire sur $L^2(\mathcal{I})$ n'est autre que la restriction du produit scalaire sur $L^2(\mathcal{A})$. Donc i est une isométrie. Enfin, si une suite (f_n) de $L^2(\mathcal{I})$ converge vers f dans $L^2(\mathcal{A})$, comme la norme $L^2(\mathcal{A})$ d'un élément de $L^2(\mathcal{I})$ est aussi sa norme $L^2(\mathcal{I})$, (f_n) converge dans $L^2(\mathcal{I})$, et sa limite doit être la même, f. Donc $f \in L^2(\mathcal{I})$. Donc $L^2(\mathcal{I})$ est fermé. □

Définition 28.6. L'espérance conditionnelle $E(f|\mathcal{I})$ de f sachant $\mathcal{I}$ est la projection orthogonale de f sur $L^2(\mathcal{I})$.

Dans le chapitre 29, on définira plus généralement l'espérance conditionnelle d'une variable aléatoire sachant une sous-tribu (le fait que la sous-tribu soit la tribu presque invariante d'une transformation mesurable ne jouant ici aucun rôle).

En conséquence, $E(f|\mathcal{I})$ est caractérisée, parmi les éléments de $L^2(\mathcal{I})$, par le fait que

$$\langle E(f|\mathcal{I})|g \rangle = \langle f|g \rangle \tag{28.1}$$

pour tout $g \in L^2(\mathcal{I})$.

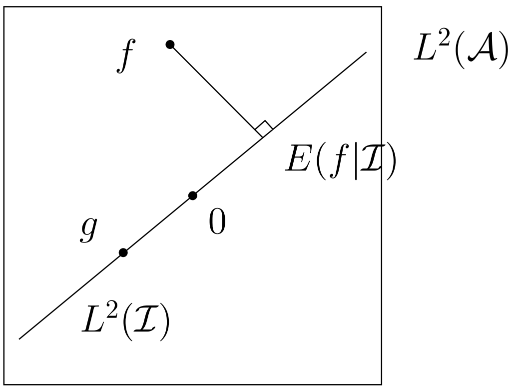


FIGURE 33 – La projection orthogonale de f sur $L^2(\mathcal{I})$

Lemme 28.7. $E(E(f|\mathcal{I})) = E(f)$.

Démonstration. Il suffit de prendre $g = 1 \in L^2(\mathcal{I})$ dans l'équation (28.1). □

On suppose dorénavant que P est invariante par $T : T(P) = P$.

Lemme 28.8. *$E(f \circ T) = E(f)$ et $E(f \circ T|\mathcal{I}) = E(f|\mathcal{I})$.*

Démonstration. La première égalité provient de ce que

$$\begin{aligned} E(f \circ T) &= \int f \circ T \, dP \\ &= \int f \, dT(P) && \text{(formule de transfert)} \\ &= \int f \, dP = E(f) && (T(P) = P). \end{aligned}$$

Par ailleurs, en prenant $g = 1$ dans l'équation (28.1) et en remplaçant f par $f \circ T - f$, on obtient

$$\langle f \circ T - f|g\rangle = \langle E(f \circ T - f|\mathcal{I})|g\rangle.$$

Or, g est invariante presque sûrement, donc le membre de gauche vaut

$$\begin{aligned} \langle f \circ T - f|g\rangle &= \int f \circ T \, g - \int fg \\ &= \int f \circ T \, g \circ T - \int fg && (g = g \circ T \quad \text{p.s.}) \\ &= \int fg \, d(T(P) - \int fg && \text{(formule de transfert)} \\ &= 0 && (T(P) = P). \end{aligned}$$

Donc

$$\langle E(f \circ T - f|\mathcal{I})|g\rangle = 0 \quad (\forall g \in L^2(\mathcal{I})),$$

donc

$$E(f \circ T|\mathcal{I}) = E(f|\mathcal{I}).$$

□

Le résultat principal de ce chapitre est le suivant.

Théorème 28.9 (von Neumann)**.**

$$S_N(f) = \frac{1}{N} \sum_{0 \leqslant n \leqslant N-1} f \circ T^n \xrightarrow{L^2}_{N \to \infty} E(f|\mathcal{I})$$

Remarque 28.10. Si $f = \mathbb{1}_A$, $A \in \mathcal{A}$, $S_N(f)(x)$ est le nombre d'entiers n compris entre 0 et N tels que $T^n(x)$ soit dans A. Autrement dit, c'est la fréquence de passage dans A de la suite des composées itérées de x par T.

Si $\mathcal{I}$ ne contient que des parties négligeables ou certaines (on dit que T est *ergodique*), $E(f|\mathcal{I})$ est constante presque sûrement, donc, d'après la question précédente, $E(f|\mathcal{I}) = E(f)$. Comme de plus $f = \mathbb{1}_A$, $E(f|\mathcal{I}) = P(A)$.

Sous ces hypothèses, le théorème ergodique affirme donc que la fréquence limite de passage dans A de l'orbite de x est simplement la probabilité de A, presque sûrement.

Démonstration. Par définition, $E(f|\mathcal{I})$ possède un représentant $\mathcal{I}$-mesurable. Donc, d'après le lemme 28.4, $E(f|\mathcal{I})$ est preque invariante : $E(f|\mathcal{I}) \circ T = E(f|\mathcal{I})$ p.s. Donc, par récurrence, $E(f|\mathcal{I}) \circ T^n = E(f|\mathcal{I})$ p.s. pour tout $n \in \mathbb{N}$. Donc, pour tout $N \in \mathbb{N}$,

$$S_N(E(f|\mathcal{I})) = \frac{1}{N} \sum_{0 \leqslant n \leqslant N-1} E(f|\mathcal{I}) \circ T^n = \frac{1}{N} \sum_{0 \leqslant n \leqslant N-1} E(f|\mathcal{I}) = E(f|\mathcal{I}).$$

On s'intéresse dorénavant à $f_0 = f - E(f|\mathcal{I})$, qui peut s'interpréter comme la "partie aléatoire" de f. Il s'agit de montrer que la suite $(S_N(f_0))$ converge vers 0 dans L^2.

Supposons provisoirement que T est bijective d'inverse mesurable (c'est une pure commodité de calcul). Si l'on pose

$$F_N = \frac{1}{N^2} \sum_{0 \leqslant n,m \leqslant N-1} f_0 \circ T^{n-m} \quad (N \geqslant 1),$$

alors

$$\begin{aligned}
\|S_N(f_0)\|_{L^2}^2 &= \frac{1}{N^2} \int \left(\sum_{0 \leqslant n \leqslant N-1} f_0 \circ T^n \right)^2 dP \\
&= \frac{1}{N^2} \int \sum_{0 \leqslant n,m \leqslant N-1} (f_0 \circ T^n \, f_0 \circ T^m) \, dP \\
&= \frac{1}{N^2} \int \sum_{0 \leqslant n,m \leqslant N-1} \left(f_0 \circ T^{n-1} \, f_0 \circ T^{m-1}\right) \, dP \\
&\text{(formule de transfert, avec } T(P) = P) \\
&= \frac{1}{N^2} \int \sum_{0 \leqslant n,m \leqslant N-1} \left(f_0 \circ T^{n-m} \, f_0\right) \, dP \qquad \text{(récurrence)} \\
&= (F_N | f_0)_{L^2} .
\end{aligned}$$

De plus, d'après l'inégalité triangulaire,

$$\|F_N\|_{L^2} \leqslant \frac{1}{N^2} \sum_{0 \leqslant n,m \leqslant N-1} \|f_0 \circ T^{n-m}\|_{L^2}.$$

Or, la formule de transfert et le fait que $T(P) = P$ montrent que, si $k \geqslant 1$, $\|f_0 \circ T^k\|_{L^2}^2 = \int (f \circ T^k)^2 = \int (f \circ T^{k-1})^2 = \cdots = \int f^2 = \|f\|_{L^2}^2$. Si $k \leqslant -1$, la propriété précédente appliquée avec T^{-1} au lieu de T montre l'identité analogue. Donc $\|f \circ T^k\|_{L^2} = \|f_0\|_{L^2}$ pour tout $k \in \mathbb{Z}$. Donc

$$\|F_N\|_{L^2} \leqslant \|f_0\|_{L^2},$$

et en particulier la suite (F_N) est bornée dans L^2.

Si (F_N) tend vers 0, d'après l'inégalité de Cauchy-Schwarz,

$$\|S_N(f_0)\|_{L^2}^2 = (F_N|f_0)_{L^2} \leqslant \|F_N\|_{L^2} \, \|f_0\|_{L^2} \to 0,$$

de quoi la conclusion du théorème ergodique découle.

Supposons par l'absurde que (F_N) ne tende pas vers 0. On a

$$\begin{aligned} F_N \circ T - F_N &= \frac{1}{N^2} \sum_{0 \leqslant n,m \leqslant N-1} \left(f_0 \circ T^{n-m+1} - f_0 \circ T^{n-m}\right) \\ &= \frac{1}{N^2} \sum_{0 \leqslant m \leqslant N-1} \left(\sum_{1 \leqslant n \leqslant N} f_0 \circ T^{n-m} - \sum_{0 \leqslant n \leqslant N-1} f_0 \circ T^{n-m} \right) \\ &= \frac{1}{N^2} \sum_{0 \leqslant m \leqslant N-1} \left(f_0 \circ T^{N-m} - f_0 \circ T^{-m}\right) \\ &\quad \text{(série télescopique)} \end{aligned}$$

donc, quand N tend vers l'infini,

$$\|F_N \circ T - F_N\|_{L^2} \leqslant \frac{2\|f_0\|_{L^2}}{N} \to 0.$$

Donc

$$F_\infty \circ T = F_\infty,$$

donc F_∞ est $\mathcal{I}$-mesurable (question 28.4), donc $F_\infty \in L^2(\mathcal{I})$, donc $E(F_\infty|\mathcal{I}) = F_\infty$. D'après le théorème de Banach-Alaoglu, il existe un suite extraite (F_{N_j}) et une fonction $F_\infty \in L^2(\mathcal{A})$ non nulle telles que, pour toute $g \in L^2(\mathcal{A})$,

$$\langle F_{N_j}|g\rangle \longrightarrow_{j\to\infty} \langle F_\infty|g\rangle.$$

Par ailleurs, on a $E(f_0|\mathcal{I}) = 0$, donc

$$E(F_N|\mathcal{I}) = 0.$$

Donc, pour toute $g \in L^2(\mathcal{A})$,

$$0 \equiv E(E(F_{N_j}|\mathcal{I})\, g) = E(E(F_{N_j}|\mathcal{I})\, E(g|\mathcal{I})) = E(F_{N_j}\, E(g|\mathcal{I}))$$

et, d'après ce qui précède, cette dernière quantité (nulle) tend, quand j tend vers l'infini, vers

$$E(F_\infty\, E(g|\mathcal{I})) = E(E(F_\infty|\mathcal{I})\, g).$$

Donc, d'après le théorème de représentation de Riesz, $E(F_\infty|\mathcal{I}) = 0$. D'après ce qui précède, ceci montre que $F_\infty = 0$, ce qui est contraire à l'hypothèse. Donc (F_N) tend vers 0, et le théorème est vérifié, dans le cas où T est une transformation bimesurable.

Lorsque T n'est pas inversible, il suffit de remplacer F_N par

$$F_N = \frac{1}{N^2} \sum_{0 \leqslant n,m \leqslant N-1} f_0 \circ T^{|n-m|}.$$

Le reste des arguments sont les mêmes. $\square$

Exercice 28.2 (Exemple de variables aléatoires dépendantes).
Trouver un exemple dans lequel les variables $f \circ T^n$ $(n \geqslant 0)$ ne sont pas indépendantes ; on pourra utiliser la loi des grands nombres.

Problème 28.1 (Décalage de Bernoulli et loi faible des grands nombres).
Cet exercice reprend partiellement le problème 26.3, dans un cadre plus simple.
Soient S un ensemble fini et $\Omega = S^{\mathbb{N}}$ l'ensemble des suites de S. Soient

$$\pi_k : \Omega \to S, \quad \omega = (\omega_0, \omega_1, ...) \mapsto \pi_k(\omega) = \omega_k$$

la k-ième projection canonique.
On rappelle (exercice 13.1) que Ω peut être muni d'une mesure produit, définie sur la tribu engendrée par l'ensemble $\mathcal{C}$ des cylindres de rang quelconque.
Soit

$$T : \Omega \circlearrowleft, \quad \omega = (\omega_0, \omega_1, ...) \mapsto T\omega = (\omega_1, \omega_2, ...)$$

le *décalage de Bernoulli* ; T "perd" le premier élément de ω et décale les autres d'un rang vers la gauche : $(T\omega)_k = \omega_{k+1}$, soit $\pi_k \circ T = \pi_{k+1}$ pour tout $k \in \mathbb{N}$.
On rappelle que T est mesurable et ergodique, et que P est invariante (par T : $T(P) = P$).
Soit $X : S \to \mathbb{R}$ une variable aléatoire discrète. Pour tout $k \in \mathbb{N}$, on définit

$$X_k = X \circ \pi_k : \Omega \to \mathbb{R}, \quad \omega \mapsto X(\omega_k),$$

de sorte que

$$X_k \circ T = X \circ \pi_k \circ T = X \circ \pi_{k+1} = X_{k+1}.$$

1. Montrer que les variables aléatoires X_k sont indépendantes et identiquement distribuées.

2. Montrer comment, dans ce cas, le théorème ergodique de von Neumann implique la loi faible des grands nombres :

$$\frac{1}{n} \sum_{0 \leqslant k \leqslant n-1} X_k \xrightarrow{L^2} E(X).$$

Exercice 28.3 (Généralisation aux mesures infinies).
On considère la situation analogue, où l'espace probabilisé est remplacé par un espace mesuré $(\Omega, \mathcal{A}, \mu)$ tel que $\mu(\Omega) = \infty$. On se restreint au cas ergodique, où la tribu presque invariante de T est triviale : $\mathcal{I} = \{\varnothing, \Omega\}$. Montrer que $S_N(f)$ converge vers 0 dans L^2.

Chapitre 29

Espérance conditionnelle dans L^2

Intuitivement, l'espérance conditionnelle d'une variable aléatoire réelle par rapport à une sous-tribu est la variable aléatoire mesurable pour cette sous-tribu "la plus proche" de la variable initiale. Son calcul explicite est rarement possible, mais c'est un outil fondamental pour estimer une variable aléatoire quand on ne dispose que d'informations incomplètes.

Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité. Considérons pour commencer un événement $B \in \mathcal{A}$ de probabilité non nulle. On définit classiquement la *probabilité conditionnelle sachant* B, notée $P(\cdot|B)$, par la formule :

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (\forall A \in \mathcal{A}).$$

Cette probabilité induit l'*espérance conditionnelle sachant* B, c'est-à-dire l'espérance relativement à la probabilité $P(\cdot|B)$:

$$E(X|B) = \frac{E(X\mathbb{1}_B)}{P(B)}$$

pour toute variable aléatoire réelle intégrable X.

Nous allons généraliser cette notion d'espérance conditionnelle, en définissant l'espérance conditionnelle

- sachant une partition dénombrable $\mathcal{B}$, par exemple la partition par les niveaux d'une variable aléatoire discrète
- ou sachant une sous-tribu $\mathcal{B}$, par exemple la tribu engendrée par une variable aléatoire mesurable.

Dans ces deux cas, l'espérance conditionnelle n'est plus un nombre, mais une variable aléatoire $\mathcal{B}$-mesurable qui n'est a priori constante que dans le cas d'une partition ou tribu triviales (l'espérance conditionnelle sachant un événement B, comme définie ci-dessus, correspondant à l'espérance conditionnelle sachant la partition $\{B, B^c\}$ ou sachant la tribu $\sigma(\{B\}) = \{\varnothing, B, B^c, \Omega\}$). L'espérance conditionnelle sachant une tribu est le cas plus difficile à appréhender, mais aussi le plus fondamental.

Conditionnement par une partition dénombrable

Soit $\mathcal{B} \subset \mathcal{A}$ une partition dénombrable de Ω. Par exemple, si Y est une variable aléatoire à valeurs dans un espace F dénombrable, les niveaux de Y [1] forment une partition dénombrable de Ω.

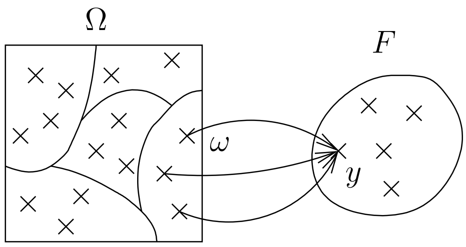


FIGURE 34 – La partition de Ω par les niveaux de Y

Les événements $A \in \mathcal{B}$ ont une probabilité nulle ou pas ; notons

$$\mathcal{B}' = \{B \in \mathcal{B},\ P(B) > 0\}.$$

Comme $\sum_{B\in\mathcal{B}\setminus\mathcal{B}'} P(B) = 0$, pour définir une variable aléatoire presque sûrement il suffit de la définir sur $\Omega' = \cup_{B\in\mathcal{B}'} B$.

Comme prolongement de l'espérance conditionnelle sachant un événement B, en faisant varier la partie B dans $\mathcal{B}'$ on obtient la définition suivante.

Définition 29.1. L'*espérance conditionnelle* de $X \in L^1(\Omega, \mathcal{A}, P)$ sachant $\mathcal{B}$ est la fonction $E(X|\mathcal{B}) \in L^1(\Omega, \sigma(\mathcal{B}), P)$ définie par

$$E(X|\mathcal{B})(\omega) = E(X|B) \quad \text{si } \omega \in B \quad (\forall B \in \mathcal{B}'),$$

soit

$$E(X|\mathcal{B}) = \sum_{B\in\mathcal{B}'} E(X|B)\mathbb{1}_B.$$

Comme $E(X|\mathcal{B})$ est constante sur chaque $B \in \mathcal{B}$, $E(X|\mathcal{B})$ est bien $\sigma(\mathcal{B})$-mesurable. Le fait que $E(X|\mathcal{B})$ soit intégrable découle du calcul suivant :

$$\begin{aligned}
E(|E(X,\mathcal{B})|) &= \sum_{B\in\mathcal{B}'} \frac{|E(X\mathbb{1}_B)|}{P(B)} P(B) && \text{(convergence monotone)}\\
&\leqslant \sum_{B\in\mathcal{B}'} E(|X|\mathbb{1}_B) && \text{(inégalité triangulaire)}\\
&= E(|X|\mathbb{1}_{\cup_{B\in\mathcal{B}'}B}) && \text{(convergence monotone)}\\
&= E(|X|) < \infty && (\mathbb{1}_{\cup_{B\in\mathcal{B}'}B} = 1 \text{ p.s.}).
\end{aligned}$$

Exemple 29.2. Soient $\Omega = \{1, 2, 3\}$ muni de la probabilité uniforme et $\mathcal{B} = \{\{1, 2\}, \{3\}\}$. L'espérance conditionnelle de $X \in L^1$ sachant $\mathcal{B}$ est

$$E(X|\mathcal{B}) = \frac{f(1) + f(2)}{2}\mathbb{1}_{\{1,2\}} + f(3)\mathbb{1}_{\{3\}}.$$

1. Rappelons qu'un *niveau* de $Y : \Omega \to F$ est une partie de Ω d'équation $Y(\omega) = y$ pour un certain $y \in F$ donnée.

Exercice 29.1 (Exemple élémentaire d'espérance conditionnelle)**.**
Dans l'expérience d'un lancé de dé à six faces équiprobables, soit Y la variable aléatoire définie par

$$Y(\omega) = \begin{cases} 0 & \text{si } \omega \text{ est pair} \\ 1 & \text{si } \omega \text{ est impair} \end{cases}.$$

Calculer l'espérance conditionnelle de $X(\omega) := \omega$ sachant Y (c'est-à-dire sachant la partition par les niveaux de Y).

Proposition 29.3. *Si $X \in L^2(\mathcal{A})$, $E(X|\mathcal{B}) \in L^2(\sigma(\mathcal{B}))$. Si de plus $Z \in L^2(\sigma(\mathcal{B}))$,*

$$E(XZ) = E(E(X|\mathcal{B})Z).$$

Démonstration. Supposons donc $X \in L^2(\mathcal{A})$. Alors,

$$\begin{aligned}
E(E(X|B)^2) &= E\left(\sum_{B\in\mathcal{B}'} E(X|B)^2 \mathbb{1}_B\right) && \\
&= \sum E(X\mathbb{1}_B)^2 && \text{(convergence monotone)} \\
&\leqslant \sum E(X^2\mathbb{1}_B) && \text{(inégalité de Jensen)} \\
&= E\left(X^2 \sum \mathbb{1}_B\right) && \text{(convergence monotone)} \\
&= E(X^2) < \infty && \Big(\sum_{B\in\mathcal{B}'} \mathbb{1}_B = 1 \quad \text{p.s.}\Big),
\end{aligned}$$

donc $E(X|\mathcal{B}) \in L^2(\sigma(\mathcal{B}))$.
Soit de plus $Z \in L^2(\sigma(\mathcal{B}))$. Z est constante sur chaque événement $B \in \mathcal{B}$, donc de la forme

$$Z = \sum_{B\in\mathcal{B}} z_B \, \mathbb{1}_B,$$

où l'on a noté $z_B \in \mathbb{R}$ la valeur prise par Z sur B. De plus, d'après l'inégalité de Cauchy-Schwarz,

$$XZ, \; E(X|\mathcal{B})Z \in L^2(\mathcal{A}).$$

Donc

$$\begin{aligned}
E(E(X|\mathcal{B})Z) &= E\left(\sum_{B\in\mathcal{B}} E(X|B)\mathbb{1}_B z_B\right) && \text{(car } 1 = \textstyle\sum_{B\in\mathcal{B}} \mathbb{1}_B) \\
&= \sum_{B\in\mathcal{B}} E(E(X|B)\mathbb{1}_B z_B) && \text{(convergence dominée par } |E(X|\mathcal{B})\, Z| \in L^1) \\
&= \sum_{B\in\mathcal{B}} E(X\mathbb{1}_B \, z_B) && \\
&= E(XZ) && \text{(convergence dominée par } |XZ| \in L^1).
\end{aligned}$$

□

La définition donnée de l'espérance conditionnelle ne se généralise pas de façon évidente au cas d'une tribu qui ne soit pas engendrée par une partition. Mais la dernière proposition pallie cette difficulté. En effet, elle s'interprète géométriquement de la façon suivante : pour toute variable aléatoire $Z \in L^2(\sigma(\mathcal{B}))$,

$$\langle X - E(X|\mathcal{B})|Z\rangle = 0,$$

c'est-à-dire que $X - E(X|\mathcal{B})$ est orthogonale à $L^2(\mathcal{B})$, soit encore que $E(X|\mathcal{B})$ est la projection orthogonale de X sur $L^2(\sigma(\mathcal{B}))$. On justifiera ci-dessous que cette propriété suffit à définir uniquement $E(X|\mathcal{B})$.

Conditionnement par une tribu

Soit maintenant $\mathcal{B}$ est une sous-tribu quelconque de $\mathcal{A}$. (L'exercice 3.5 avait montré que beaucoup de tribus ne sont pas engendrées par une partition.)

Comme Ω et P sont fixés dans cette discussion, on note $L^2(\mathcal{A}) = L^2(\Omega, \mathcal{A}, P)$ et $L^2(\mathcal{B}) = L^2(\Omega, \mathcal{B}, P)$.

Lemme 29.4. *Il existe une injection canonique isométrique* $L^2(\mathcal{B}) \hookrightarrow L^2(\mathcal{A})$.

Démonstration. $L^2(\mathcal{B})$ est un espace vectoriel hilbertien réel, au même titre que $L^2(\mathcal{A})$. Comme $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$, une fonction $\mathcal{B}$-mesurable est automatiquement $\mathcal{A}$-mesurable. Toute classe de fonctions $\mathcal{B}$-mesurables (pour la relation d'égalité presque sûre) est donc incluse dans une unique classe de fonctions $\mathcal{A}$-mesurables (mais ce n'est généralement pas une égalité, parce que un classe dans l'espace des fonctions $\mathcal{A}$-mesurable contient aussi des représentants $\mathcal{A}$-mesurables qui, a priori, ne sont pas $\mathcal{B}$-mesurables).

Donc il existe une application $j : L^2(\mathcal{B}) \to L^2(\mathcal{A})$, et cette application est linéaire, injective (si $f \in L^2(\mathcal{B})$ et $j(f) = 0$ dans $L^2(\mathcal{A})$, les représentants de f sont nuls presque partout, donc $f = 0$ dans $L^2(\mathcal{B})$) et isométrique (si $f, g \in L^2(\mathcal{B})$, $\langle j(f)|j(g)\rangle_{L^2(\mathcal{A})} = E(j(f)j(g)) = E(fg) = \langle f|g\rangle$).

De plus, supposons qu'une suite (f_n) de $L^2(\mathcal{B})$ est telle que $(j(f_n))$ converge vers g dans $L^2(\mathcal{A})$. La suite $(j(f_n))$ est de Cauchy. Comme j est isométrique, la suite (f_n) l'est aussi. Comme $L^2(\mathcal{B})$ est complet, (f_n) converge donc vers une classe $f \in L^2(\mathcal{B})$ et, par unicité de la limite et continuité de j, $j(f) = g$. Donc $j(L^2(\mathcal{B}))$ est un sous-espace fermé de $L^2(\mathcal{A})$. □

Dans la suite, on identifiera $L^2(\mathcal{B})$ au sous-espace fermé $j(L^2(\mathcal{B}))$ des classes de fonctions $f \in L^2(\mathcal{A})$ qui possèdent un représentant $\mathcal{B}$-mesurable.

Définition 29.5. Soient $\mathcal{B}$ une sous-tribu de $\mathcal{A}$ et $X \in L^2(\mathcal{A})$. L'espérance conditionnelle de X sachant $\mathcal{B}$ est la projection orthogonale, notée $E(X|\mathcal{B})$, de X sur le sous-espace fermé $L^2(\mathcal{B})$. Autrement dit, $E(X|\mathcal{B})$ est l'unique variable aléatoire $\hat{X} \in L^2(\mathcal{B})$ telle que, pour toute variable aléatoire $Z \in L^2(\mathcal{B})$,

$$E(XZ) = E(E(X|\mathcal{B})Z). \tag{29.1}$$

Il en découle que $E(X|\mathcal{B})$ est la meilleure approximation de X (au sens de la norme L^2) par une variable aléatoire $\mathcal{B}$-mesurable :

$$\|X - E(X|\mathcal{B})\|_{L^2} = \min_{\hat{X} \in L^2(\mathcal{B})} \|X - \hat{X}\|_{L^2},$$

puisque, en décomposant X comme $X = (X - E(X|\mathcal{B})) + E(X|\mathcal{B})$, on voit (théorème de Pythagore) que

$$\|X - \hat{X}\|_{L^2}^2 = \|X - E(X|\mathcal{B})\|_{L^2}^2 + \|E(X|\mathcal{B}) - \hat{X}\|_{L^2}^2 \geqslant \|X - E(X|\mathcal{B})\|_{L^2}^2.$$

Dans le cas où $\mathcal{B}$ est la tribu engendrée par une variable aléatoire Y : $\mathcal{B} = \sigma(Y)$, on note

$$E(X|Y) = E(X|\sigma(Y)).$$

Définition 29.6. La probabilité de A sachant la sous-tribu $\mathcal{B}$ est

$$P(A|\mathcal{B}) = E(\mathbb{1}_A|\mathcal{B}) \in L^1(\mathcal{B}).$$

Propriété 29.7. *Si $\mathcal{C}$ est une sous-tribu de $\mathcal{B}$ ($\mathcal{B}$ étant elle-même une sous-tribu de $\mathcal{A}$),*

$$E(E(X|\mathcal{B})|\mathcal{C}) = E(X|\mathcal{C}).$$

Les deux exemples extrêmes d'espérance conditionnelle sachant $\mathcal{B}$, selon que $\mathcal{B}$ est fine ou grossière, sont les suivants.

Exemple 29.8. Si $\mathcal{B} \supset \sigma(X)$ (c'est-à-dire si $X \in L^2(\mathcal{B})$), $E(X|\mathcal{B}) = X$.

Exemple 29.9. Si X et $\mathcal{B}$ sont indépendantes (c'est-à-dire si $\sigma(X)$ et $\mathcal{B}$ sont indépendantes), pour toute $Z \in L^2(\mathcal{B})$, X et Z sont indépendantes, donc

$$\begin{aligned} E(XZ) &= E(X)E(Z) && \text{(indépendance)} \\ &= E(E(X)Z) && \text{(linéarité).} \end{aligned}$$

Comme $E(X)$ est constante, elle est mesurable par rapport à n'importe quelle tribu, en particulier $\mathcal{B}$. Donc $E(X)$ satisfait la propriété caractéristique de $E(X|\mathcal{B})$, donc

$$E(X|\mathcal{B}) = E(X).$$

Donc, en prenant pour $\mathcal{B}$ une sous-tribu quelconque de $\mathcal{A}$ et en prenant $\mathcal{C} = \{\varnothing, \Omega\}$ (tribu indépendante de toute variable aléatoire), la propriété 29.7 montre que

$$E(E(X|\mathcal{B})) = E(E(X|\mathcal{B})|\mathcal{C}) = E(X|\mathcal{C}) = E(X)$$

(ce qui se voit aussi en prenant $Z = 1$ dans l'identité (29.1)).

Exercice 29.2 (Exemple élémentaire d'espérance conditionnelle)**.**
Calculer $E(X|\mathcal{B})$ quand $\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}([0,1])$, quand P est la mesure de Lebesgue et quand $\mathcal{B}$ est la tribu engendrée par l'ensemble des singletons de $[0,1]$.

Exercice 29.3 (Propriétés élémentaires de l'espérance conditionnelle)**.**
Montrer les propriétés suivantes :

1. L'application $L^2(\mathcal{A}) \to L^2(\mathcal{B})$, $X \mapsto E(X|B)$ est linéaire.

2. $E(E(X|B)) = E(X)$

3. Si $X \leqslant X'$ p.s., $E(X|\mathcal{B}) \leqslant E(X'|\mathcal{B})$.

Exercice 29.4 (Conditionnement gaussien).
Soit $(Y_1, ..., Y_d, X)$ un vecteur gaussien centré. On note $Y = (Y_1, ..., Y_d)$.
1. Montrer que $E(X|Y)$ est la projection orthgonale de X sur l'espace vectoriel (de dimension finie !) engendré par $Y_1, ..., Y_d$, donc qu'en particulier il existe un vecteur $\lambda \in \mathbb{R}^d$ tel que

$$E(X|Y) = \lambda \cdot Y.$$

Posons

$$\sigma^2 = E\left((X - \lambda \cdot Y)^2\right).$$

2. Montrer que, si $\sigma = 0$, pour toute fonction borélienne $h : \mathbb{R} \to \mathbb{R}_+$, $E(h(X)|Y) = h(X)$.
3. Montrer que, si $\sigma \neq 0$, pour toute fonction borélienne $h : \mathbb{R} \to \mathbb{R}_+$,

$$E(h(X)|Y) = \int_{\mathbb{R}} h(x) q_{m,\sigma^2}(x)\, dx$$

où $m = \lambda \cdot Y$ et q_{m,σ^2} est la densité de $N(m, \sigma^2)$:

$$q_{m,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Chapitre 30

Espérance conditionnelle dans L^1

Soient $(\Omega, \mathcal{A}, P)$ un espace de probabilité et $\mathcal{B}$ une sous-tribu.

Théorème et Définition 30.1. *Si $X : \Omega \to [0, +\infty]$ est mesurable, la formule*

$$E(X|\mathcal{B}) := \lim \uparrow E(\min(X, n)|\mathcal{B}) \quad \text{p.s.}$$

définit une variable aléatoire $\Omega \to [0, +\infty]$ (à un ensemble négligeable près) $\mathcal{B}$-mesurable, caractérisée par la propriété suivante : pour toute variable aléatoire $\mathcal{B}$-mesurable positive Z,

$$E(XZ) = E(E(X|\mathcal{B})Z).$$

Démonstration. La variable aléatoire $X_n = \min(X, n)$ est bornée, donc dans $L^2(\mathcal{A})$. D'après l'exercice 29.3, la suite $(E(X_n|\mathcal{B}))$ est croissante p.s., donc possède une limite p.s. parmi les variables aléatoires $\Omega \to [0, +\infty]$, que l'on note $E(X|\mathcal{B})$.

Soit Z est une variable aléatoire $\mathcal{B}$-mesurable positive. Pour tout n,

$$E(X_n Z) = E(E(X_n|\mathcal{B})Z).$$

Donc, par convergence monotone,

$$E(XZ) = E(E(X|\mathcal{B})Z).$$

Réciproquement, supposons que $\hat{X}$ soit une variable aléatoire $\mathcal{B}$-mesurable positive telle que, pour toute variable aléatoire Z $\mathcal{B}$-mesurable positive on ait

$$E(XZ) = E(\hat{X}Z).$$

*** (fin à rédiger) □

Théorème et Définition 30.2. *Soit $X \in L^1(\mathcal{A})$. Il existe une unique variable aléatoire $E(X|\mathcal{B}) \in L^1(\mathcal{B})$ telle que, pour tout événement $B \in \mathcal{B}$,*

$$E(X\mathbb{1}_B) = E(E(X|\mathcal{B})\mathbb{1}_B). \tag{30.1}$$

Plus généralement, pour toute variable aléatoire Z réelle $\mathcal{B}$-mesurable bornée,

$$E(XZ) = E(E(X|\mathcal{B})Z). \tag{30.2}$$

Démonstration du théorème 30.2. Soient $X^{\pm}$ les parties positive et négative de X. D'après le théorème-définition 30.1, *** (fin à rédiger). □

Remarque 30.3. On peut aussi montrer le théorème 30.2 en utilisant le théorème général suivant, de prolongement des applications linéaires continues :

Soient E et F deux espaces de Banach, E' un sous-espace dense de E, et $u : E' \to F$ une application linéaire continue. Il existe une unique application linéaire continue $E \to F$ qui prolonge u.

Exercice 30.1 (Propriétés élémentaires de l'espérance conditionnelle dans L^1)**.**
Montrer les propriétés suivantes :

1. L'espérance conditionnelle ainsi définie est linéaire pour les combinaisons linéaires à coefficients positifs.

2. Si X est $\mathcal{B}$-mesurable, $E(X|\mathcal{B}) = X$.

3. Convergence monotone : si (X_n) est une suite croissante positive,

$$\lim \uparrow E(X_n|\mathcal{B}) = E(\lim \uparrow |\mathcal{B}) \quad \text{p.s.}$$

4. Lemme de Fatou : Si (X_n) est positive,

$$E(\liminf |\mathcal{B}) \leqslant \liminf E(X_n|\mathcal{B}).$$

5. Convergence dominée : Si $X_n \to X$ p.s. et $|X_n| \leqslant Z \in L^1$,

$$E(X|\mathcal{B}) = \lim E(X_n|\mathcal{B}).$$

6. Inégalité de Jensen conditionnelle : Si $\phi\ :]a, b[\to \mathbb{R}$ est convexe positive et $X \in L^1$ est à valeurs dans $]a, b[$,

$$\phi(E(X|\mathcal{B})) \leqslant E(\phi(X)|\mathcal{B})$$

(on pourra s'inspirer de la démonstration de l'inégalité de Jensen non conditionnelle vue dans l'exercice 7.10).

Exercice 30.2 (Conditionnement des variables à densité)**.**
Soient X et Y deux vecteurs aléatoires à valeurs respectivement dans $\mathbb{R}^m$ et dans $\mathbb{R}^n$, dont la loi conjointe possède une densité f : pour toute fonction borélienne $h : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^+$,

$$E[h(X,Y)] = \int_{\mathbb{R}^m\times\mathbb{R}^n} h(x,y)f(x,y)\,dx\,dy.$$

1. Quelle est la densité f_Y de Y ?

2. Calculer $E[h(X)|Y]$, en utilisant le fait que, pour toute fonction borélienne $g : \mathbb{R}^n \to \mathbb{R}^+$,

$$E[h(X)g(Y)] = E[E[X|Y]g(Y)].$$

Définition 30.4. Soient $(E, \mathcal{E})$ et $(F, \mathcal{F})$ deux espaces probabilisables. Une *probabilité de transition* ou un *noyau de transition* de E dans F est une application

$$\nu : E \times \mathcal{F} \to [0, 1]$$

telle que
— pour tout $x \in E$, $\nu(x, \cdot)$ est une probabilité sur $(F, \mathcal{F})$
— et, pour tout $B \in \mathcal{F}$, $\nu(\cdot, B)$ est $\mathcal{E}$-mesurable.
De plus, si X et Y sont deux variables aléatoires à valeurs respectivement dans $(E, \mathcal{E})$ et $(F, \mathcal{F})$, une *loi conditionnelle* de Y sachant X est une probabilité de transition ν de E dans F telle que, pour toute fonction h mesurable positive sur $(F, \mathcal{F})$,

$$E[h(Y)|X] = \int \nu(X, dy)h(y).$$

Exercice 30.3 (Densité conditionnelle)**.**
Soient X et Y deux vecteurs aléatoires à valeurs respectivement dans $\mathbb{R}^m$ et dans $\mathbb{R}^n$, dont la loi conjointe possède une densité f. Calculer la loi conditionnelle et X sachant Y, ainsi que la densité conditionnelle correspondante.

Autres exercices

Problème 30.1 (Exemple de variable aléatoire gaussienne, *examen 2012*)**.**
Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires réelles sur $(\Omega, \mathcal{A}, P)$.

On note, si $n \geqslant 1$, $\vec{X}_n = (X_1, ..., X_n)$ et, si $(x_n)_{n\geqslant 1}$ est une suite réelle, $\vec{x}_n = (x_1, ..., x_n)$.

On suppose que X_1 est de loi gaussienne $N(0, 1)$ et que, pour tout $n \geqslant 1$ et tout $\vec{x}_n \in \mathbb{R}^n$, une loi conditionnelle de X_{n+1} sachant $\vec{X}_n$ est la loi gaussienne $N(x_n, 1)$; cela signifie que

$$f_{X_{n+1}}^{\vec{X}_n=\vec{x}_n}(x_{n+1}) := \frac{f_{\vec{X}_{n+1}}(\vec{x}_{n+1})}{f_{\vec{X}_n}(\vec{x}_n)} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_{n+1} - x_n)^2}{2}\right).$$

1. Quelle est la densité de $\vec{X}_2 = (X_1, X_2)$?

2. Montrer que $E(X_{n+1}|\vec{X}_n) = X_n$ et $E({X_{n+1}}^2|\vec{X}_n) = 1 + {X_n}^2$; en déduire les moyenne et variance de X_n, puis que la suite (X_n) ne converge pas dans L^2.

3. Montrer que la densité de $\vec{X}_n$ est de la forme $f_{\vec{X}_n}(\vec{x}_n) = (2\pi)^{-n/2} e^{-{}^t\vec{x}_n \cdot A_n \cdot \vec{x}_n/2}$ pour une matrice A_n qu'on déterminera, puis calculer la fonction caractéristique $\varphi_{\vec{X}_n}$ de $\vec{X}_n$.

4. Quelle est la loi du vecteur bi-dimensionnel (X_j, X_k), $j < k$? Calculer la limite en loi de $(X_j, X_k/\sqrt{k})$ quand k tend vers l'infini, j étant fixé.

5. $(*)$ Calculer la limite en loi de la suite des variables aléatoires

$$Z_n = \frac{1}{n\sqrt{n}} \sum_{1\leqslant j\leqslant n} X_j.$$

Problème ⋆ 30.2 (Théorème ergodique de Birkhoff)**.**
Cet exercice fait suite à l'exercice 26.2.

1. Montrer la version générale du *théorème ergodique de Birkhoff* :

Si $f : E \to E$ est une application $\mathcal{E}$-mesurable préservant la probabilité μ (c'est-à-dire $f(\mu) = \mu$) et si $\psi \in L^1(\mathcal{E})$, quand n tend vers $+\infty$ on a

$$\frac{1}{n}\sum_{k=0}^{n-1}\psi(f^k(x)) \to \psi_{\mathcal{I}} \quad \mu\text{-presque partout,}$$

où $\mathcal{I} = \{A \in \mathcal{E},\ f^{-1}(A) = A\}$ est la *tribu invariante* de f.

On adaptera l'exercice 26.2 (où l'on supposait f ergodique, et où la tribu invariante $\mathcal{I}$ était donc la tribu grossière), en posant notamment, à la fin, $\varphi = \psi - \psi_{\mathcal{I}} - \varepsilon$.

2. Vérifier directement la validité du théorème de Birkhoff dans le cas où $E = \{1, ..., n\}$, $n \in \mathbb{N}_*$, $\mathcal{E} = \mathcal{P}(E)$, et f est une permutation de E se décomposant en plusieurs cycles.

Chapitre 31

Théorème de Radon-Nikodym ★

Définition 31.1. Soient μ et ν deux mesures sur un espace mesurable $(E, \mathcal{A})$.

— ν est *absolument continue* par rapport à μ, ce que l'on note $\nu \ll \mu$, si

$$\mu(A) = 0 \Rightarrow \nu(A) = 0 \quad (\forall A \in \mathcal{A}).$$

— ν est *étrangère à* μ, ce que l'on note $\nu \perp \mu$, s'il existe $N \in \mathcal{A}$ tel que $\mu(N) = 0$ et $\nu(N^c) = 0$.

Par exemple, la mesure de Dirac en un réel est étrangère à la mesure de Lebesgue. Dans le problème 31.1, un exemple est donné d'une mesure étrangère à la mesure de Lebesgue et qui ne charge pas les singletons.

Inversement, si f est une fonction mesurable positive, la mesure $f\mu$ est absolument continue par rapport à μ. Le théorème suivant montre que l'on construit ainsi toutes les mesures absolument continues par rapport à μ.

Théorème 31.2 (Radon-Nikodym [1, 2]). *Soient μ et ν deux mesures σ-finies sur $(E, \mathcal{A})$. Il existe un unique couple (ν_a, ν_e) de mesures σ-finies sur $(E, \mathcal{A})$ telles que*

$$\nu = \nu_a + \nu_e, \quad \text{avec} \quad \begin{cases} \nu_a \ll \nu \\ \nu_e \perp \mu. \end{cases}$$

De plus il existe une fonction mesurable $g : E \to \mathbb{R}_+$ tell que, pour tout $A \in \mathcal{A}$,

$$\nu_a(A) = \int_A g\, f\mu$$

et la fonction g est unique à un ensemble de μ-mesure nulle près.

Démonstration. Supposons d'abord que μ et ν sont finies et que $\mu \geqslant \nu$ (c'est-à-dire que $\mu(A) \geqslant \nu(A)$ pour tout $A \in \mathcal{A}$). Comme $L^2(\mu) \subset L^1(\mu) \subset L^1(\nu)$, on peut considérer l'application

$$\Phi : L^2(\mu) \to \mathbb{R}, \quad f \mapsto \int f\, d\nu.$$

1. Johann Karl August Radon, mathématicien hongrois (1887–1956), spécialiste de calcul variationnel et d'analyse fonctionnelle.

2. Otto Marcin Nikodým, mathématicien polonais (1887–1974), dont les contributions concernent des domaines allant de la théorie de la mesure à la mécanique quantique.

L'inégalité de Cauchy-Schwarz montre que

$$|\Phi(f)| \leqslant \left(\int f^2\,d\nu\right)^{1/2} \nu(E)^{1/2} \leqslant \left(\int f^2\,d\mu\right)^{1/2} \nu(E)^{1/2} = \nu(E)^{1/2}\|f\|_{L^2(\mu)}.$$

Donc Φ est une forme linéaire continue sur l'espace de Hilbert $L^2(\mu)$. D'après le théorème de Riesz,* il existe une fonction $h \in L^2(\mu)$ telle que, pour toute $f \in L^2(\mu)$,

$$\Phi(f) = \int fh\,d\mu.$$

En particulier, en prenant $f = \mathbb{1}_A$, on obtient que, pour tout $A \in \mathcal{A}$,

$$\nu(A) = \int_A h\,d\mu.$$

Dans la suite, on aura besoin de savoir que $0 \leqslant h \leqslant 1$ μ-p.p. Pour le montrer, remarquons que, pour tout $\epsilon > 0$,

$$0 \leqslant \nu(\{h(x) \leqslant -\epsilon\}) = \int_{\{h(x)\leqslant -\epsilon\}} h\,d\mu \leqslant -\epsilon\mu(\{h(x) \leqslant -\epsilon\}),$$

donc que $\mu(\{h(x) \leqslant -\epsilon\}) = 0$; et que, pour tout $\epsilon > 0$,

$$\begin{aligned}\mu(\{h(x) \geqslant 1+\epsilon\}) &\geqslant \nu(\{h(x) \geqslant 1+\epsilon\})\\ &= \int_{\{h(x)\geqslant 1+\epsilon\}} h\,d\mu\\ &\geqslant (1+\epsilon)\mu(\{h(x) \geqslant 1+\epsilon\}),\end{aligned}$$

donc que $\mu(\{h(x) \geqslant 1+\epsilon\}) = 0$.

Si μ et ν sont toujours finies mais pas forcément ordonnées, on peut appliquer la conclusion du cas précédent) ν et $\mu+\nu$: il existe une fonction mesurable h telle que, pour toute fonction $f \in L^2(\mu+\nu)$,

$$\int f\,d\nu = \int fh\,d(\mu+\nu).$$

En particulier, pour toute fonction f mesurable bornée,

$$\int f\,d\nu = \int fh\,d\mu + \int fh\,d\nu,$$

donc

$$\int f(1-h)\,d\nu = \int fh\,d\mu. \tag{31.1}$$

Puisque $1-h \geqslant 0$ μ-p.p. donc ν-p.p., en utilisant le théorème de convergence monotone, on voit que cette dernière égalité est vraie pour toute fonction f mesurable positive.

D'autre part, en posant $N = \{h = 1\}$ et en prenant $f = \mathbb{1}_N$, on voit que $\mu(N) = 0$. La mesure $\nu_e = \mathbb{1}_N \nu$ est donc étrangère à μ. Et, en remplaçant f par $\mathbb{1}_{N^c}(1-h)^{-1}f$ dans l'égalité (31.1), on obtient que, pour toute fonction f mesurable positive,

$$\int_{N^c} f\,d\nu = \int_{N^c} f\frac{h}{1-h}\,d\mu = \int fg\,d\mu, \quad g = \mathbb{1}_{N^c}\frac{h}{1-h}.$$

En posant $\nu_a = \mathbb{1}_{N^c}\nu = g\mu$, on obtient bien une mesure absolument continue par rapport à μ, vérifiant la propriété indiquée dans l'énoncé.

Il reste à généraliser ce qui précède au cas où μ et ν sont non plus finies. Supposons donc μ et ν seulement σ-finies. Alors il existe une partition mesurable dénombrable $(E_n)_{n\in\mathbb{N}}$ de E telle que, pour tout n, $\mu(E_n) < \infty$ et $\nu(E_n) < \infty$. Soient μ_n et ν_n les restrictions respectives de μ et de ν à E_n. D'après la démonstration dans le cas fini, il existe une décomposition

$$\nu_n = \nu_n^a + \nu_n^e$$

de ν_n et une fonction mesurable positive g_n telles que $\nu_n^a = g\mu$ et $\nu_n^e \perp \mu$. On peut prolonger g_n par 0 en dehors de E_n. Il suffit alors de poser

$$\nu_a = \sum_{n\in\mathbb{N}} \nu_n^a, \quad \nu_e = \sum_{n\in\mathbb{N}} \nu_n^e \quad \text{et} g = \sum_{n\in\mathbb{N}} g_n.$$

Montrons l'unicité du couple (ν_a, ν_e). Soit $(\tilde{\nu}_a, \tilde{\nu}_e)$ un autre couple tel que $\nu = \tilde{\nu}_a + \tilde{\nu}_e$, $\nu_a \ll \mu$ et $\nu_e \perp \mu$. Pour tout $A \in \mathcal{A}$,

$$\nu_a(A) - \tilde{\nu}_a(A) = \nu_e(A) - \tilde{\nu}_e(A).$$

Soient N et $\tilde{N}$ les ensembles μ-négligeables portant respectivement ν_e et $\tilde{\nu}_e$. Si l'on note $M = N \cup \tilde{N}$,

$$\nu_e(A) - \tilde{\nu}_e(A) = \nu_e(A \cap M) - \tilde{\nu}_e(A \cap M) = \nu_a(A \cap M) - \tilde{\nu}_a(A \cap M) = 0$$

puisque $\nu_a \ll \mu$ et $\tilde{\nu}_a \ll \mu$. L'unicité de g à un ensemble μ-négligeable près a été démontrée dans l'exercice 7.4. $\square$

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé et $\mathcal{B}$ une sous-tribu. Le théorème de Radon-Nikodym peut être utiliser pour définir directement l'espérance conditionnelle des variables aléatoires L^1 sans passer par L^2 (le détour ayant déjà été fait dans la démonstration du théorème). On note $L^1(\mathcal{A}) = L^1(\Omega, \mathcal{A}, P)$ et $L^1(\mathcal{B}) = L^1(\Omega, \mathcal{B}, P)$.

Démonstration du théorème 30.2 par le théorème de Radon-Nikodym. Supposons d'abord $X \geqslant 0$. Considérons la mesure de densité X par rapport à P :

$$(XP)(A) := E(X\mathbb{1}_A) \quad (\forall A \in \mathcal{A}).$$

Elle est bien sûr absolument continue par rapport à P, et la mesure XP restreinte à la tribu $\mathcal{B}$ est elle-même absolument continue par rapport à $P|_{\mathcal{B}}$. D'après le théorème de Radon-Nikodým, $(XP)|_{\mathcal{B}}$ possède une densité par rapport à $P_{\mathcal{B}}$, qui est une fonction $\mathcal{B}$-mesurable que nous noterons $E(X|\mathcal{B})$:

$$\forall B \in B, \quad E(X\mathbb{1}_B) = E(E(X|\mathcal{B})\mathbb{1}_B).$$

En prenant $B = \Omega$, on voit que $E(E(X|\mathcal{B})) = E(X) < \infty$, donc que $E(X|\mathcal{B}) \in \mathcal{L}^1(\Omega, \mathcal{B}, P)$. La propriété (30.2) s'en déduit immédiatement, d'après l'exercice 7.4. Si maintenant X est de signe quelconque, en notant $X = X^+ - X^-$ sa décomposition en la différence de ses parties positive et négative, on voit qu'il suffit de poser

$$E(X|\mathcal{B}) = E(X^+|\mathcal{B}) - E(X^-|\mathcal{B}).$$

Montrons que la variable aléatoire $E(X|\mathcal{B})$ trouvée est uniquement définie par la propriété (30.1) (pour cette raison, la propriété (30.1) est souvent appelée la *propriété caractéristique* de $E(X|\mathcal{B})$). Soient X' et X'' deux variables aléatoires telles que

$$\forall B \in \mathcal{B}, \quad E(X'\mathbb{1}_B) = E(X''\mathbb{1}_B) = E(X\mathbb{1}_B).$$

En particulier, avec $B = \{X' > X''\} \in \mathcal{B}$, on obtient

$$E((X' - X'')\mathbb{1}_{X'-X''>0}) = E((X - X)\mathbb{1}_{X'-X''>0}) = 0,$$

donc $X' \leqslant X''$ p.s. et, par symétrie, $X' = X''$ dans $L^1(\mathcal{B})$. □

Problème 31.1 (Une mesure étrangère à la mesure de Lebesgue)**.**
Montrer que la mesure de Lebesgue-Stieltjes μ_F (exercice 9.1) associée à l'escalier du diable de Lebesgue F (exemple 17.7) est étrangère à la mesure de Lebesgue, et ne charge pas les singletons.

Chapitre 32

Chaînes de Markov

Dans ce chapitre nous examinons un exemple important de suite de variables aléatoires dépendantes.

Soient S un ensemble fini et, pour toute paire $(x, y) \in S$, $p_{xy} \in [0, 1]$, tels que

$$\sum_{y \in S} p_{xy} = 1 \quad (x \in S).$$

Définition 32.1. Une *chaîne de Markov* (on dit aussi *processus de Markov*) associée aux *probabilités de transition* p_{xy} est une suite $(X_n)_{n \geqslant 0}$ de variables aléatoires telle que

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, ..., X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) = p_{x_n x_{n+1}}$$

pour tout n et tout $(x_0, ..., x_{n+1})$ tels que

$$P(X_0 = x_0, ..., X_n = x_n) > 0.$$

On peut penser aux éléments de S comme les états possibles d'un certain système (physique, économique, biologique, etc.), et à X_n comme l'état de ce système au temps n. Une chaîne de Markov représente alors l'évolution du système au cours du temps. La contrainte principale est que la loi conditionnelle de l'état au temps $n + 1$ sachant l'état au temps n ne doit pas dépendre des états passés antérieurs, $X_0, ..., X_{n-1}$. Les exemples d'application sont innombrables, et cette définition conduit à une riche théorie, qu'ici nous n'effleurons qu'à peine.

Exercice 32.1 (Marche aléatoire symétrique)**.**
Soit un mobile se mouvant sur les points entiers de la droite. On suppose qu'il est en 0 à l'instant $t = 0$, et qu'à chaque instant entier il se déplace de ± 1 avec la probabilité $1/2$, de façon indépendante des déplacements précédents. On note X_n la position à l'instant n (donc $X_0 = 0$), N_n le nombre de déplacements vers la droite à l'instant n, et M_n le nombre de déplacements vers la gauche.

1. Montrer que X_n a la parité de n (on pourra utiliser le fait que $N_n + M_n = n$ et $N_n - M_n = X_n$) et en déduire que X_n est à valeurs dans $\{-n, -n + 2, ..., n - 2, n\}$.

2. Remarquer que N_n suit une loi binomiale, et en déduire la loi de X_n.

Soit $Z_n = \pm 1$ la variable aléatoire $X_n - X_{n-1}$ $(n \geqslant 1)$, de sorte que

$$X_n = Z_1 + \cdots + Z_n.$$

Par hypothèse, les Z_n sont indépendantes les unes des autres et ont même loi

$$\frac{1}{2}(\delta_{-1} + \delta_{+1}).$$

La suite $(X_n)_{n\geqslant 0}$ s'appelle une *marche aléatoire symétrique partant de* 0. La *trajectoire* de $\omega \in \Omega$ pour ce processus est la suite de points $(n, X_n(\omega))_{n\in\mathbb{N}}$.

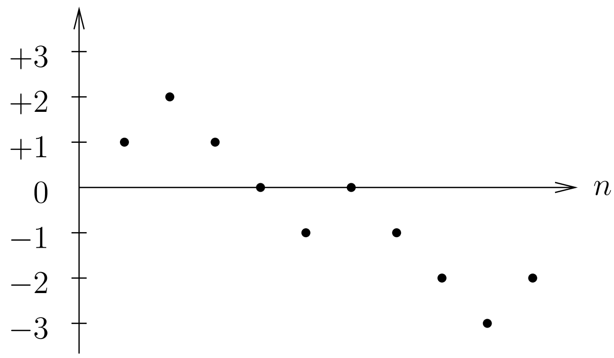


FIGURE 35 – Une trajectoire d'une marche aléatoire

3. Montrer que la probabilité de ne pas repasser par 0 jusqu'à l'instant n (i.e. $X_1, ..., X_n \neq 0$) est

$$P(X_{n-1} = 0) + P(X_{n-1} = 1).$$

Soit T_a le temps mis pour aller de 0 à a :

$$T_a(\omega) := \min\{n \geqslant 0;\ X_n(\omega) = a\}.$$

4. ★ Montrer que, quel que soit $a \in \mathbb{Z}$, T_a est finie p.s.

5. Montrer l'*inégalité des grandes déviations* : pour tous $a > 0$ et $n \geqslant 0$,

$$P(|X_n| > a) \leqslant 2e^{-a^2/2n};$$

on pourra utiliser l'inégalité de Markov à la variable aléatoire e^{uX_n} avec $u > 0$.

6. En déduire que, pour tout $m \geqslant 1$,

$$\sum_{n\geqslant 1} P(|X_n| > \frac{n}{m}) \leqslant \sum_{n\geqslant 1} 2e^{-n/2m^2},$$

puis que

$$\frac{X_n}{n} \to 0 \text{ p.s.}$$

Exercice ★ 32.2 (Marche aléatoire asymétrique [LG06], *partiel 2012*)**.**
Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes, de même loi

$$P_{X_1}(1) = p, \quad P_{X_1}(-1) = 1 - p, \quad 0 < p < 1.$$

Soit $(Z_n)_{n\geqslant 1}$ la suite de changements de signes de (X_n) : $Z_n = \mathbb{1}_{\{X_n \neq X_{n+1}\}}$.

1. Montrer que si les Z_n sont indépendantes, $p = 1/2$ (la réciproque est vraie, mais on ne demande pas de la démontrer) ; on pourra utiliser en particulier l'indépendance de Z_1 et Z_2.

On pose $S_0 = 0$ et, pour $n \geqslant 1$, $S_n = \sum_{1\leqslant k\leqslant n} X_k$. Soient

$$\Sigma = \sum_{n\geqslant 0} \mathbb{1}_{\{S_n=0\}} \quad \text{et} \quad A = \{\Sigma < \infty\};$$

donc, la variable aléatoire Σ compte le nombre d'annulation des S_n, et A est l'événement "les S_n s'annulent seulement un nombre fini de fois".

2. Montrer qu'il existe une constante C, qu'on déterminera, telle que

$$P(S_{2n} = 0) \sim_{n\to+\infty} \frac{C}{\sqrt{n}}(4p(1-p))^n$$

(on pourra utiliser la formule de Stirling), tandis que $P(S_{2n+1} = 0) = 0$.

3. (*) En déduire que $P(A) = 0$ si $p = 1/2$ et $P(A) = 1$ si $p \neq 1/2$; on pourra calculer l'espérance de Σ.

Figure 36 – Andreï Markov (1856–1922)

Annexe A

Borel par lui-même

Le texte qui suit est extrait des *Leçons sur la théorie des fonctions* (1898) [Bor98]. Il donne une définition ce ce que nous appelons maintenant les parties boréliennes de $\mathbb{R}$. L'objectif est de donner une mesure à des parties de l'intervalle $[0, 1]$ engendrées par union dénombrable et passage au complémentaire. Mais la démonstration d'existence de la mesure manque. Elle aurait été "longue et fastidieuse", estima Borel, et résulte des travaux ultérieurs de Lebesgue.

Les ensembles mesurables.

Tous les ensembles que nous considérerons seront formés de points compris entre 0 et 1. Lorsqu'un ensemble sera formé de tous les points compris dans une infinité dénombrable d'intervalles n'empiétant pas les uns sur les autres et ayant une longueur totale s, nous dirons que l'ensemble *a pour mesure* s. Lorsque deux ensembles n'ont pas de points communs, et que leurs mesures sont s et s', l'ensemble obtenu en les réunissant, c'est-à-dire leur somme, a pour mesure $s + s'$. D'ailleurs, il importe peu dans la définition de la mesure d'un ensemble, ou dans celle de la somme de deux ensembles, que l'on néglige, ou que l'on tienne tel compte que l'on veut des extrémités des intervalles, en infinité dénombrable.

Plus généralement, si l'on a une infinité dénombrable d'ensembles n'ayant deux à deux aucun point commun et ayant respectivement pour mesures $s_1, s_2, \ldots, s_n, \ldots$, leur somme (ou

---

(1) On comparera avec fruit les définitions que nous allons donner avec les définitions plus générales que donne M. Jordan dans son *Cours d'Analyse*. Le problème que nous étudions ici est d'ailleurs tout différent de celui qu'a résolu

ensemble formé par leur réunion) a pour mesure

$$s_1 + s_2 + \ldots + s_n + \ldots.$$

Tout cela est une conséquence de la définition de la mesure. Voici maintenant des définitions nouvelles : si un ensemble E a pour mesure s, et contient tous les points d'un ensemble E′ dont la mesure est s', *l'ensemble* $E - E'$, *formé des points de* E *qui n'appartiennent pas à* E′, *sera dit avoir pour mesure* $s - s'$; de plus, si un ensemble est la somme d'une infinité dénombrable d'ensembles sans partie commune, sa mesure sera la somme des mesures de ses parties et enfin les ensembles E et E′ ayant, en vertu de ces définitions, s et s' comme mesures, et E renfermant tous les points de E′, l'ensemble $E - E'$ aura pour mesure $s - s'$.

Le théorème fondamental démontré pages 41-43 nous assure que ces définitions ne seront jamais contradictoires entre elles (¹); nous sommes donc libres de les adopter; nous sommes d'ailleurs assurés aussi que la mesure d'un ensemble ne sera jamais une quantité négative; mais un ensemble peut avoir pour mesure *zéro* et avoir la puissance du continu. Tel est l'ensemble E considéré tantôt. Si nous reprenons les notations de la page 45 et si nous désignons par α_n la mesure de E_n $\left(\alpha_n < \frac{M}{n}\right)$, l'ensemble $E_n - E_{n+1}$ aura pour mesure $\alpha_n - \alpha_{n+1}$ (nous savons que E_n renferme tous les points de E_{n+1}). L'ensemble A des points qui n'appartiennent pas à E_n a pour mesure $1 - \alpha_n$ (c'est la différence de l'ensemble de tous les points du segment 0-1 et de E_n). L'ensemble des points qui n'appartiennent pas à E peut être regardé comme formé en ajoutant à A les ensembles $E_n - E_{n+1}$, $E_{n+1} - E_{n+2}$, $\ldots$; sa mesure est donc

$$1 - \alpha_n + (\alpha_n - \alpha_{n+1}) + (\alpha_{n+1} - \alpha_{n+2}) + \ldots = 1$$

puisque α_m tend vers zéro pour m infini. Donc, l'ensemble E obtenu en retranchant cet ensemble de l'ensemble de tous les points 0-1, a pour mesure *zéro*.

Ainsi un ensemble qui a pour mesure zéro peut être non dénom-

(¹) Il est du moins aisé d'obtenir ce résultat par des procédés tout à fait analogues à ceux que l'on a employés pour établir ce théorème.

brable; mais *tout ensemble dénombrable a pour mesure zéro;* c'est une conséquence aisée de ce qui précède.

Les ensembles dont on peut définir la mesure en vertu des définitions précédentes seront dits par nous ensembles mesurables, sans que nous entendions impliquer par là qu'il n'est pas possible de donner une définition de la mesure d'autres ensembles; mais une telle définition nous serait inutile; elle pourrait même nous gêner, si elle ne laissait pas à la *mesure* les propriétés fondamentales que nous lui avons attribuées dans les définitions que nous avons données ([1]).

Ces propriétés essentielles, que nous résumons ici parce qu'elles nous seront utiles, sont les suivantes : La mesure de la somme d'une infinité dénombrable d'ensembles est égale à la somme de leurs mesures; la mesure de la différence de deux ensembles est égale à la différence de leurs mesures ([2]); *la mesure n'est jamais négative; tout ensemble dont la mesure n'est pas nulle n'est pas dénombrable.* C'est surtout de cette dernière propriété que nous ferons usage. Il est d'ailleurs expressément entendu que nous ne parlerons de mesure qu'à propos des ensembles que nous avons appelés *mesurables.*

Cependant, si un ensemble E contient tous les éléments d'un ensemble mesurable E_1, de mesure α, nous pourrons dire que la mesure de E est supérieure à α, sans nous inquiéter si E est mesurable ou non. Inversement, si E_1 contient tous les éléments de E, nous dirons que la mesure de E est inférieure à α. Les mots *supérieure* et *inférieure* n'excluent d'ailleurs pas l'égalité.

([1]) Le procédé que nous avons employé revient en réalité à ceci : nous avons reconnu qu'une définition de la mesure ne pouvait être utile que si elle avait certaines propriétés fondamentales : nous avons posé *a priori* ces propriétés et ce sont elles qui nous ont servi à définir la classe d'ensembles que nous regardons comme mesurables. Cette manière de procéder présente de grandes analogies avec les méthodes introduites par M. J. Drach, en Algèbre et dans la théorie des équations différentielles (*voir*, par exemple, l'Ouvrage cité (p.23) et *Comptes rendus*, janvier 1895). Dans tous les cas, elle procède de la même idée fondamentale : définir les éléments nouveaux que l'on introduit, à l'aide de leurs propriétés *essentielles,* c'est-à-dire de celles qui sont strictement indispensables pour les raisonnements qui doivent suivre.

([2]) Bien entendu quand on parle de la somme de plusieurs ensembles, on suppose qu'ils n'ont, deux à deux, aucun élément commun et, quand on parle de leur différence, on suppose que l'un d'eux renferme tous les éléments de l'autre.

Il est aisé de voir que les propriétés essentielles s'étendent, avec des modifications convenables, à ces nouvelles définitions : en quelque sorte, un calcul d'égalités se trouve remplacé par un calcul d'inégalités qui peut parfois rendre les mêmes services.

Nous allons démontrer, en terminant, une proposition importante, qui montrera la liaison intime entre les notions diverses introduites dans ce Chapitre : *tout ensemble parfait* (1) *limité* (2) *est mesurable*.

Soit en effet A un ensemble parfait dont tous les points sont dans l'intervalle o-1 et soit α un point de cet intervalle, n'appartenant pas à A; je dis qu'il existe un intervalle ab comprenant α et ne renfermant pas de point de A, a et b étant d'ailleurs des points de A. (L'un des points a ou b pourrait ne pas appartenir à A; il coïnciderait alors avec le point o ou le point 1). En effet, A étant parfait, α n'appartient pas à A'; il existe donc un nombre ε tel qu'il n'y ait pas de point de A dans l'intervalle α, $\alpha+\varepsilon$, ε étant positif. D'autre part, en excluant le cas où il n'y aurait pas de point de A dans l'intervalle α-1, il existe un nombre ε' tel qu'il y ait au moins un point de A dans l'intervalle α, $\alpha+\varepsilon'$. Il est manifeste que tout nombre positif η, ou bien a la même propriété que ε, ou bien a la même propriété que ε' et l'on voit aisément qu'il existe un nombre b tel que, si $\eta < b$, il n'y a pas de point de A dans l'intervalle α, $\alpha+\eta$, tandis qu'il y en a si $\eta > b$. Or, il est aisé de voir que le point b, ou bien appartient à A, ou bien (3) appartient à A'; or, A est parfait; donc, dans tous les cas, b appartient à A. On démontrerait de même l'existence d'un point a en prenant ε et η négatifs; l'intervalle ab est l'intervalle cherché.

Ainsi, A étant un ensemble parfait, tout point α qui n'appartient pas à A se trouve dans un intervalle ab ne renfermant d'autre point de A que ses extrémités. Si nous considérons main-

(1) Le mot *parfait* peut être pris ici dans le sens large de M. Jordan; le résultat serait vrai, *a fortiori*, s'il avait le sens que lui donne M. Cantor.

(2) On dit qu'un ensemble est limité lorsque la distance de deux quelconques de ses points est inférieure à un nombre fixe.

(3) Car, si le point b n'appartient pas à A, il y a des points de A distincts de b, dans l'intervalle b, $b+h$, quelque petit que soit le nombre positif h; donc b appartient à A'.

tenant un autre point α', ne faisant pas partie de A et non situé dans ab, nous aurons un intervalle analogue $a'b'$, etc. D'ailleurs, l'ensemble des intervalles que l'on peut ainsi définir est certainement dénombrable, puisque la somme de leurs longueurs est au plus égale à l'unité. L'ensemble A s'obtient donc en retranchant de l'ensemble o-1 une infinité dénombrable d'ensembles mesurables; il est donc mesurable [1].

Cette démonstration nous fait en même temps connaître le moyen le plus général de construction d'un ensemble parfait situé sur le segment o, 1. Pour se donner un tel ensemble, il suffit de se donner une infinité dénombrable d'intervalles $a_1 b_1$, $a_2 b_2$, ..., $a_n b_n$, ..., assujettis d'ailleurs à certaines restrictions. On en conclut aisément que l'ensemble de tous ces ensembles parfaits à la puissance du continu. Nous verrons que, d'autre part, l'ensemble de tous les ensembles de points situés sur le segment o, 1, a une puissance plus élevée que celle du continu [2]. Cette remarque suffit à montrer quelle restriction considérable on apporte à la notion d'ensemble, lorsqu'on assujettit un ensemble à être *parfait* et, par suite, combien il est naturel que l'on puisse être conduit ainsi à des résultats plus simples.

---

[1] Nous avons eu d'ailleurs un exemple d'un ensemble mesurable non parfait; car il était dense dans tout intervalle.

[2] *Voir* Note I, pages 109-110.

Annexe B

Lebesgue par lui-même

Ce texte de trois pages de Lebesgue [Leb01] définit l'intégrale et la mesure de Lebesgue, qui sont d'une portée immense.

ANALYSE MATHÉMATIQUE. — *Sur une généralisation de l'intégrale définie.* Note de M. **H. Lebesgue**, présentée par M. Picard.

« Dans le cas des fonctions continues, il y a identité entre les notions d'intégrale et de fonction primitive. Riemann a défini l'intégrale de certaines fonctions discontinues, mais toutes les fonctions dérivées ne sont pas intégrables, au sens de Riemann. Le problème de la recherche des fonctions primitives n'est donc pas résolu par l'intégration, et l'on peut désirer une définition de l'intégrale comprenant comme cas particulier celle de Riemann et permettant de résoudre le problème des fonctions primitives (1).

» Pour définir l'intégrale d'une fonction continue croissante

$$\gamma(x)\,(a \leqq x \leqq b),$$

on divise l'intervalle (a, b) en intervalles partiels et l'on fait la somme des quantités obtenues en multipliant la longueur de chaque intervalle partiel

(1) Ces deux conditions imposées *a priori* à toute généralisation de l'intégrale sont évidemment compatibles, car toute fonction dérivée intégrable, au sens de Riemann, a pour intégrale une de ses fonctions primitives.

par l'une des valeurs de y quand x est dans cet intervalle. Si x est dans l'intervalle (a_i, a_{i+1}), y varie entre certaines limites m_i, m_{i+1}, et réciproquement si y est entre m_i et m_{i+1}, x est entre a_i et a_{i+1}. De sorte qu'au lieu de se donner la division de la variation de x, c'est-à-dire de se donner les nombres a_i, on aurait pu se donner la division de la variation de y, c'est-à-dire les nombres m_i. De là deux manières de généraliser la notion d'intégrale. On sait que la première (se donner les a_i) conduit à la définition donnée par Riemann et aux définitions des intégrales par excès et par défaut données par M. Darboux. Voyons la seconde.

» Soit la fonction y comprise entre m et M. Donnons-nous

$$m = m_0 < m_1 < m_2 < \ldots < m_{p-1} < \mathrm{M} = m_p;$$

$y = m$, quand x fait partie d'un ensemble E_0; $m_{i-1} < y \leqq m_i$ quand x fait partie d'un ensemble E_i.

» Nous définirons plus loin les mesures λ_0, λ_i de ces ensembles. Considérons l'une ou l'autre des deux sommes

$$m_0\lambda_0 + \Sigma m_i\lambda_i; \quad m_0\lambda_0 + \Sigma m_{i-1}\lambda_i;$$

si, quand l'écart maximum entre deux m_i consécutifs tend vers zéro, ces sommes tendent vers une même limite indépendante des m_i choisis, cette limite sera par définition l'intégrale des y qui sera dite *intégrable*.

» Considérons un ensemble de points de (a, b); on peut d'une infinité de manières enfermer ces points dans une infinité dénombrable d'intervalles; la limite inférieure de la somme des longueurs de ces intervalles est la mesure de l'ensemble. Un ensemble E est dit *mesurable* si sa mesure augmentée de celle de l'ensemble des points ne faisant pas partie de E donne la mesure de (a, b) (1). Voici deux propriétés de ces ensembles : une infinité d'ensembles mesurables E_i étant donnée, l'ensemble des points qui font partie de l'un au moins d'entre eux est mesurable; si les E_i n'ont deux à deux aucun point commun, la mesure de l'ensemble obtenu est la somme des mesures E_i. L'ensemble des points communs à tous les E_i est mesurable.

» Il est naturel de considérer d'abord les fonctions telles que les ensembles qui figurent dans la définition de l'intégrale soient mesurables. On trouve que : *si une fonction limitée supérieurement en valeur absolue est*

---

(1) Si l'on ajoute à ces ensembles des ensembles de mesures nulles convenablement choisis, on a des ensembles mesurables au sens de M. Borel (*Leçons sur la théorie des fonctions*).

lle que, quels que soient A *et* B, *l'ensemble des valeurs de* x *pour lesquelles* *a a* $A < y \leqq B$ *est mesurable, elle est intégrable* par le procédé indiqué. Une elle fonction sera dite *sommable*. L'intégrale d'une fonction sommable st comprise entre l'intégrale par défaut et l'intégrale par excès. De sorte ue, *si une fonction intégrable au sens de Riemann est sommable, l'intégrale t la même avec les deux définitions*. Or, *toute fonction intégrable au sens e Riemann est sommable,* car l'ensemble de ses points de discontinuité est e mesure nulle, et l'on peut démontrer que si, en faisant abstraction d'un nsemble de valeurs de x de mesure nulle, il reste un ensemble en chaque oint duquel une fonction est continue, cette fonction est sommable. Cette propriété permet de former immédiatement des fonctions non intégrables au sens de Riemann et cependant sommables. Soient $f(x)$ et $\varphi(x)$ eux fonctions continues, $\varphi(x)$ n'étant pas toujours nulle; une fonction ui ne diffère de $f(x)$ qu'aux points d'un ensemble de mesure nulle partout ense et qui en ces points est égale à $f(x) + \varphi(x)$ est sommable sans être ntégrable au sens de Riemann. *Exemple :* La fonction égale à o si x irrationnel, égale à 1 si x rationnel. Le procédé de formation qui précède iontre que l'ensemble des fonctions sommables a une puissance supérieure au continu. Voici deux propriétés des fonctions de cet ensemble.

» 1° *Si* f *et* φ *sont sommables,* $f + \varphi$ *et* $f\varphi$ *le sont* et l'intégrale de $f + \varphi$ st la somme des intégrales de f et de φ.

» 2° *Si une suite de fonctions sommables a une limite, c'est une fonction ommable.*

» L'ensemble des fonctions sommables contient évidemment $y = k$ et $y = x$; donc, d'après 1°, il contient tous les polynomes et comme, d'après 2°, il contient toutes ses limites, il contient donc toutes les fonctions coninues, toutes les limites de fonctions continues, c'est-à-dire les fonctions le première classe (voir Baire, *Annali di Matematica*, 1899), il contient outes celles de seconde classe, etc.

» En particulier, *toute fonction dérivée, limitee supérieurement en valeur ibsolue,* étant de première classe, *est sommable* et l'on peut démontrer que *on intégrale, considérée comme fonction de sa limite supérieure, est une de ses onctions primitives.*

» Voici maintenant une application géométrique : si $|f'|$, $|\varphi'|$, $|\psi'|$ sont imitées supérieurement, la courbe

$$x = f(t), \qquad y = \varphi(t), \qquad z = \psi(t)$$

a pour longueur l'intégrale de $\sqrt{f'^2 + \varphi'^2 + \psi'^2}$. Si $\varphi = \psi = 0$, on a la variation totale de la fonction f à variation limitée. Dans le cas où f', φ', ψ' n'existent pas, on peut obtenir un théorème presque identique en remplaçant les dérivées par les nombres dérivés de Dini. »

ANALYSE MATHÉMATIQUE. — *Sur les intégrales analytiques des équations différentielles du premier ordre dans le voisinage de conditions initiales singulières.* Note de M. **Henri Dulac**, présentée par M. Painlevé.

« X et Y étant des fonctions de x et de y holomorphes et nulles pour $x = y = 0$, considérons, dans le champ complexe, aux environs des valeurs singulières $x = y = 0$, les intégrales de l'équation

$$X\,dy + Y\,dx = 0. \tag{1}$$

On connaît divers résultats relatifs au cas où le point singulier est un point d'intersection simple des courbes $X = 0$, $Y = 0$. Je me propose de rechercher dans quelle mesure on peut étendre ces résultats aux autres cas.

» Soient : n l'ordre minimum des termes de X et Y, P et Q l'ensemble de ces termes d'ordre n de X et de Y (P ou Q peut être nul). Le point singulier sera dit d'ordre n.

» I. *Recherche des intégrales algébroïdes passant par l'origine.* — A la méthode de Briot et Bouquet je substitue la méthode suivante, qui permet aussi d'obtenir toutes ces intégrales. Cette méthode, ainsi que je l'ai constaté récemment, ne diffère guère que par des détails d'exposition d'une méthode employée par M. I. Bendixson.

» L'équation homogène $yP + xQ = 0$ fournit, pour $\frac{y}{x}$, $(n+1)$ valeurs égales ou inégales. Soit α une de ces valeurs; deux cas peuvent se présenter :

» 1° La valeur $\frac{y}{x} = \alpha$ n'annule pas à la fois P et Q. En posant $y = x(t+\alpha)$ on met en évidence une seule intégrale holomorphe, quel que soit l'ordre de multiplicité de α. On permute les rôles de x et de y pour étudier les intégrales tangentes à $x = 0$;

» 2° La racine α, d'ordre de multiplicité r, annule à la fois P et Q. Le même changement de variable nous ramène à l'étude d'une équation pour laquelle $x = t = 0$ est un point singulier au plus d'ordre r. Cette équation, outre l'intégrale $x = 0$, qui ne fournit pas d'intégrale pour (1),

Annexe C

Quelques lois classiques

Lois discrètes

Loi uniforme sur un ensemble fini E

$$\frac{1}{\sharp E}\sum_{x\in E}\delta_x$$

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

$$p\delta_1 + (1-p)\delta_0$$

Loi d'une pièce qui tombe sur pile (1) avec la probabilité p et sur face (0) avec la probabilité $1-p$.

Loi binomiale $B(n,p)$ ($n \in \mathbb{N}^*$, $p \in [0,1]$)

$$\sum_{0\leqslant k\leqslant n} C_n^k p^k (1-p)^{n-k}\delta_k$$

Loi du nombre de piles obtenus en n lancers avec la pièce précédente.

Loi géométrique de paramètre $p \in]0,1[$

$$(1-p)\sum_{n\in\mathbb{N}} p^n\delta_n.$$

Loi du nombre de piles obtenues avant le premier face.

Loi de Poisson de paramètre $\lambda > 0$

$$e^{-\lambda}\sum_{n\in\mathbb{N}}\frac{\lambda^n}{n!}\delta_n$$

Si $X_k \sim B(k,p_k)$ et $kp_k \to \lambda$, $P(X_k = n) \to_{k\to\infty} e^{-\lambda}\frac{\lambda^n}{n!}$.

Lois continues

Loi uniforme sur $[a, b]$ $(a < b)$

$$\frac{1}{b-a}\mathbb{1}_{[}a, b]\, dx.$$

Loi de Cauchy

$$\frac{1}{\pi\alpha\left(1+\left(\frac{x}{\alpha}\right)^2\right)}\, dx$$

Loi de $\alpha \tan Y$ si Y est de loi uniforme sur $]-\pi/2, \pi/2[$.

Loi du χ^2 à n degrés de liberté

$$\frac{1}{\Gamma(n/2)2^{n/2}}\,\frac{x^{n/2-1}}{e^{x/2}}\,\mathbb{1}_{x>0}\, dx$$

Loi de $Y_1^2+\cdots+Y_n^2$ si $Y_1, ..., Y_n$ sont gaussiennes centrées réduites et indépendantes,

Loi exponentielle de paramètre λ

$$\lambda e^{-\lambda x}\mathbb{1}_{\mathbb{R}^+}(x)\, dx.$$

Loi gaussienne unidimensionnelle $N(\mu, \sigma^2)$

$$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right)\, dx$$

Loi log-normale

$$\mathbb{1}_{x>0}\,\frac{1}{x\sqrt{2\pi}}\,\exp\left(-\frac{(\ln x)^2}{2}\right)\, dx.$$

Loi de e^Y si Y est gausienne centrée réduite.

Annexe D

La double terminologie

> ... An attentive reader will observe, and perhaps resent, that in other chapters a function is f or g, and so on, whereas in this chapter a function is more likely to be x or y, and so on, at the other end of the alphabet. This difference is traditional, and is one of the principal features that distinguishes probability from the rest of measure theory.
>
> J. L. Doob [Doo94]

Analyse	**Probabilités**
Espace mesuré $(E, \mathcal{A}, \mu)$	Espace probabilisé $(\Omega, \mathcal{A}, P)$
Partie mesurable : $A \in \mathcal{A}$	Événement : $A \in \mathcal{A}$
Point : $x \in E$	Réalisation du hasard : $\omega \in \Omega$
Application mesurable f	Variable aléatoire X
Mesure μ	Probabilité P $(P(\Omega) = 1)$
Presque partout (p.p.)	Presque sûrement (p.s.)
Mesure image par f : $f(\mu) = \mu_f$	Loi de X : P_X
Fonction réelle mesurable $f : E \to \mathbb{R}$	Variable aléatoire réelle $X : \Omega \to \mathbb{R}$
$f^{-1}(B)$	$\{X \in B\}$
Intégrale $\int f \, d\mu$	Espérance $E(X) = \int X \, dP$
Transformée de Fourier $\hat{\mu}$ $(E = \mathbb{R}^d)$	Fonction caractéristique $\Phi_X = \widehat{X(P)}$
Convergence simple presque partout (p.p.)	Convergence presque sûre (p.s.)
Convergence L^1	Convergence en moyenne
Convergence L^2	Convergence en moyenne quadratique
Convergence en mesure	Convergence en probabilité.

Glossaire

La plupart de ces définitions sont expliquées dans n'importe quel livre d'analyse.

Axiome du choix Soient I un ensemble et $(E_i)_{i\in I}$ une famille d'ensembles non vides indexée par I. Alors il existe une famillle $(x_i)_{i\in I}$ d'éléments telle que $x_i \in E_i$ pour tout $i \in I$; ou, de façon équivalente, le produit cartésien $\prod_{i\in I} E_i$ est non vide. Cet axiome est trivial quand A est un singleton*, et se démontre par récurrence si A est fini. Mais, si A est infini, on ne peut pas déduire cet axiome des autres axiomes de la théorie des ensembles.

Base d'ouverts d'un espace métrique (ou topologique) Classe **B** d'ouverts telle que tout ouvert soit exactement une union d'éléments de **B**. Par exemple, les intervalles ouverts de $\mathbb{R}$ forment une base d'ouverts (exercice 5.1).

Classe En théorie des ensemble, la notion de classe généralise celle d'ensemble. Dans ce cours, nous n'utilisons pas le mot "classe" dans un sens précis, mais simplement, pour éviter les répétitions au sein d'une phrase, dans le sens d'un ensemble dont les éléments sont eux-mêmes des ensembles. Par exemple, une tribu de E est une classe de parties de E.

Difféomorphisme Application de classe C^1 entre deux ouverts de $\mathbb{R}^n$, dont l'inverse soit aussi de classe C^1.

Ensemble dénombrable Ensemble E tel qu'il existe une injection* de E dans l'ensemble $\mathbb{N}$ des nombres entiers naturels. [1]

Par convention, si $(x_e)_{e\in\varnothing}$ est une famille vide de nombres réels (indexée par l'ensemble vide),

$$\sum_{e\in\varnothing} x_e = 0 \quad \text{et} \quad \prod_{e\in\varnothing} x_e = 1;$$

cela respecte la croissance de la somme de nombres positifs, et du produit de nombre $\geqslant 1$. De même, si $(A_e)_{e\in\varnothing}$ est une famille vide de parties d'un ensemble Ω,

$$\bigcup_{e\in\varnothing} A_e = \varnothing \quad \text{et} \quad \bigcap_{e\in\varnothing} A_e = \Omega.$$

1. L'autre convention possible, en usage en particulier chez les mathématiciens anglo-saxons, est de dire que E est dénombrable s'il est en bijection avec $\mathbb{N}$.

En conséquence, une fonction σ-additive d'ensembles (mesure) est nulle en l'ensemble vide.

Espace métrique séparable Espace métrique contenant une partie dénombrable dense.

Fonction Application à valeurs dans $\mathbb{R}$ (on la qualifie alors de *réelle*), $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ (on la dit alors *réelle étendue*) ou $\mathbb{C}$ (on la dit alors *complexe*).

Fonction en escalier Fonction définie sur un intervalle borné $[a, b]$ de $\mathbb{R}$, telle qu'il existe une subdivision* $a_0 = a < ... < a_n = b$ et des réels $y_1, ..., y_n$ tels que

$$f|_{]a_{i-1},a_i[} = y_i \quad (\forall i = 1, ..., n).$$

Injection Une injection entre deux ensembles E et F est une application $f : E \to F$ telle que deux éléments distincts de E ne sont peuvent pas être envoyés sur la même image : $f(x) = f(x') \Rightarrow x = x'$ (pour tous $x, x' \in E$).

Limites inférieure et supérieure Ces notions existent dans plusieurs contextes.
1) Si $(u_n)_{n\in\mathbb{N}}$ est une suite réelle, ses limites inférieure et supérieure sont

$$\begin{cases} \liminf u_n = \sup_{n\geqslant 0} \inf_{k\geqslant n} u_k \in \bar{\mathbb{R}} \\ \limsup u_n = \inf_{n\geqslant 0} \sup_{k\geqslant n} u_k \in \bar{\mathbb{R}}. \end{cases}$$

Ces nombres (éventuellement infinis) existent toujours, contrairement à la limite de (u_n). En effet, comme la suite $(\inf_{k\geqslant n} u_k)_{n\geqslant 0}$ est croissante, sa limite existe et donc

$$\liminf u_n = \lim_n \inf_{k\geqslant n} u_k;$$

de même,

$$\limsup u_n = \lim_n \sup_{k\geqslant n} u_k.$$

(Ces formules peuvent éviter d'hésiter entre les définitions des limites inférieure et supérieure !)

Exercice : Montrer que $\liminf u_n \leqslant \limsup u_n$ (on pourra montrer que lim inf est la plus petite valeur d'adhérence de (u_n) et que lim sup est la plus grande).

La suite (u_n) a une limite si et seulement si ses limites inférieure et supérieure coïncident, auquel cas ces trois notions de limite coïncident.

1') Si (f_n) est une suite de fonctions réelles, donc d'un ensemble E dans $\mathbb{R}$, en utilisant ce qui précède, on peut définir les fonctions $\liminf f_n$ et $\limsup f_n$ par les formules

$$(\liminf f_n)(x) = \liminf(f_n(x)) \quad \text{et} \quad (\limsup f_n)(x) = \limsup(f_n(x)) \quad (\forall x \in E).$$

2) Si $(A_n)_{n\in\mathbb{N}}$ est une suite de parties d'un ensemble E, ses limites inférieure et supérieure sont

$$\begin{cases} \liminf A_n = \cup_{n\geqslant 0} \cap_{k\geqslant n} A_k \\ \limsup A_n = \cap_{n\geqslant 0} \cup_{k\geqslant n} A_k. \end{cases}$$

La limite inférieure de (A_n) est l'ensemble des éléments de E qui appartiennent à une infinité de A_n ; la limite supérieure est l'ensemble des éléments de E qui évitent au plus un nombre fini de A_n.

Le lien avec les suites de fonctions numériques est le suivant :

$$\liminf \mathbb{1}_{A_n} = \mathbb{1}_{\liminf A_n} \quad \text{et} \quad \limsup \mathbb{1}_{A_n} = \mathbb{1}_{\limsup A_n}.$$

Ouvert d'un espace métrique Partie O contenant une (petite) boule ouverte centrée en chacun de ses points.

Propriété de Baire Dans un espace métrique complet, une partie l'intersection dénombrable d'ouverts denses est dense ; en passant au complémentaire, on obtient la propriété équivalente que l'union dénombrable de fermés d'intérieur vide est d'intérieur vide.

Une intersection dénombrable d'ouverts s'appelle un G_δ, et une union dénombrable de fermés, un F_σ [2] La propriété ci-dessus se paraphrase en disant qu'un G_δ est dense, et qu'un F_σ est d'intérieur vide.

Remarquons qu'une intersection dénombrable d'ouverts denses est plus qu'une partie dense ; notamment, elle est non dénombrable (bien qu'il puisse exister des parties dénombrables denses), puisqu'un ouvert dense privé d'un point reste un ouvert dense. Par exemple, $\mathbb{Q}$ est dense mais n'est pas un G_δ de $\mathbb{R}$, tandis que l'ensemble des nombres liouvilliens (pourtant de mesure nulle), est un G_δ dense de $\mathbb{R}$ (exercice 11.1).

Singleton Ensemble à un élément. Par exemple, $\{0\}$, ou $\{36\}$ sont des singletons de $\mathbb{N}$. Formellement, il faut distinguer l'entier $0 \in \mathbb{N}$ du singleton $\{0\} \subset \mathbb{N}$.

Subdivision d'un invervalle $[a, b]$ Partie finie $\{a_0 = a < ... < a_n = b\}$ de $[a, b]$, contenant les extrémités.

Théorème de Hahn-Banach Soient X un espace vectoriel normé et Y un sous-espace de X. Toute forme linéaire continue $\eta \in Y'$ sur Y possède un prolongement $\tilde{\eta} \in X'$ sur X ayant la même norme d'opérateur : $\tilde{\eta}|_Y = \eta$ et $\|\tilde{\eta}\| = \|\eta\|$.

Théorème de représentation de Riesz Si H est un espace de Hilbert, pour toute forme linéaire continue $\lambda : H \to \mathbb{C}$ il existe un unique vecteur $a \in H$ tel que, pour tout $x \in H$, $\lambda(x) = (a|x)$. L'application $a \mapsto (a|\cdot)$ est une isométrie (anti-isométrie sur le corps des nombres complexes) isométrique.

2. L'origine de ces symboles est la suivante : G provient de l'allemand *Gebiet* (voisinage), δ de l'allemand *Durchschnit* (intersection), F du français *fermé* et σ du français *somme* (union).

Théorème de Stone-Weierstrass Soient X un espace topologique compact et $C(X)$ l'espace des fonctions réelles continues sur X, muni de la norme sup. Si R est une sous-algèbre de $C(X)$ possédant la propriété des deux points (à savoir que, pour tous points distincts x et y de X et pour tous réels a et b, il existe $f \in R$ telle que $f(x) = a$ et $f(y) = b$), R est dense dans $C(X)$.

Bibliographie

[Arn76] V. I. Arnold. *Les méthodes mathématiques de la mécanique classique.* Éditions Mir, Moscow, 1976. Traduit du russe par Djilali Embarek.

[Arn99] V. I. Arnold. Kepler's second law and the topology of Abelian integrals (according to Newton) [Kvant **1987**, no. 12, 17–21]. In *Kvant selecta : algebra and analysis, II*, volume 15 of *Math. World*, pages 131–140. Amer. Math. Soc., Providence, RI, 1999.

[Bil95] P. Billingsley. *Probability and measure.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[BJ06] M. Brancovan and T. Jeulin. *Probabilités : niveau M1.* Mathématiques à l'Université. Ellipses, Paris, 2006. La couv. porte en plus : Cours et exercices corrigés.

[Bog07] V. I. Bogachev. *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007.

[Bor98] É. Borel. *Leçons sur la théorie des fonctions : éléments et principes de la théorie des ensembles, applications à la théorie des fonctions.* Collection de monographies sur la théorie des fonctions. Gauthier-Villars et Fils, 1898.

[BP06] M. Briane and G. Pagès. *Théorie de l'intégration : cours & exercices : licence & master de mathématiques.* Vuibert, 2006.

[Can13] B. Candelpergher. *Théorie des probabilités.* Calvage et Mounet, 2013.

[Cho01] G. Choquet. *Les Cours à la Sorbonne. Les Cours à l'École polytechnique.* Mathématiques. Ellipses, Paris, 2001.

[Dei05] A. Deitmar. *A first course in harmonic analysis.* Universitext. Springer-Verlag, New York, second edition, 2005.

[Doo94] J. L. Doob. *Measure theory*, volume 143 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, 1994.

[Dos11] H. Doss. Cours d'intégration. Cours polycopié, 2011. Licence de Mathématiques, Université Paris-Dauphine.

[Gur98] V. P. Gurarii. *Group methods in commutative harmonic analysis*, volume 25 of *Encyclopaedia of Mathematical Sciences.* Springer-Verlag, Berlin, 1998. Translated from the Russian original by D. and S. Dynin.

[JP03] J. Jacod and P. Protter. *Probability essentials.* Universitext. Springer-Verlag, Berlin, second edition, 2003.

[KLR98] J.-P. Kahane and P.-G. Lemarié-Rieusset. *Séries de Fourier et ondelettes.* Nouvelle Bibliothèque mathématique. Cassini, 1998.

[KS92] A.N. Kolmogorov and A.N. Shiryayev. *Selected Works of A. N. Kolmogorov : Probability theory and mathematical statistics.* Number vol. 2 in Mathematics and its applications (Kluwer Academic Publishers). : Soviet series. Kluwer Academic Publishers, 1992.

[KS07] L. B. Koralov and Y. G. Sinai. *Theory of probability and random processes.* Universitext. Springer, Berlin, second edition, 2007.

[Leb01] H. Lebesgue. Sur une généralisation de l'intégrale définie. *Comptes-rendus de l'Académie des sciences*, 1901.

[LG06] J.-F. Le Gall. Intégration, probabilités et processus aléatoires. Cours polycopié, 2006. École normale supérieure, Paris.

[LM06] Y. Lacroix and L. Mazliak. *Probabilités. Variables aléatoires, convergences, conditionnement.* Ellipses, 2006.

[Rie54] B. Riemann. Ueber die darstellbarkeit einer function durch eine trigonometrische reihe [sur la représentation d'une fonction comme série trigonométrique]. In *Proceedings of the Royal Philosophical Society at Göttingen*, volume 13, pages 87–132, 1854.

[Rud76] W. Rudin. *Principles of mathematical analysis.* McGraw-Hill Book Co., New York, third edition, 1976. International Series in Pure and Applied Mathematics.

[Rud80] W. Rudin. *Analyse réelle et complexe.* Masson, Paris, 1980. Traduit de la première édition anglaise par N. Dhombres et F. Hoffman, troisième édition.

[Rud87] W. Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York, third edition, 1987.

[Sko05] A. V. Skorokhod. *Basic principles and applications of probability theory.* Springer-Verlag, Berlin, 2005. Edited by Yu. V. Prokhorov, Translated from the 1989 Russian original by B. D. Seckler.

[Sol70] R. M. Solovay. A model of set-theory in which every set of reals is Lebesgue measurable. *Ann. of Math. (2)*, 92 :1–56, 1970.

[Tao09] T. Tao. *Analysis. I*, volume 37 of *Texts and Readings in Mathematics.* Hindustan Book Agency, New Delhi, second edition, 2009.

[Tao10a] T. Tao. *An epsilon of room, I : real analysis*, volume 117 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2010. Pages from year three of a mathematical blog.

[Tao10b] T. Tao. *An epsilon of room, II.* American Mathematical Society, Providence, RI, 2010. Pages from year three of a mathematical blog.

[Tao11] T. Tao. *An introduction to measure theory*, volume 126 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2011.

[Wie66] N. Wiener. *Generalized harmonic analysis and Tauberian theorems.* The M.I.T. Press, Cambridge, Mass.-London, 1966.

Index