Théorie de la Mesure [1]

S. Friedli

Departamento de Matemática, UFMG
E-mail address: sacha@mat.ufmg.br

1. Cours donné à l'Institut de Mathématiques (UNIGE), automne 2010 et 2013.

Table des matières

1. Conventions

Les entiers naturels : $\mathbb{N} := \{1, 2, \dots\}$. Les indices i, j, k, l, m, n seront toujours utilisés pour symboliser ses éléments.

La droite réelle : $\mathbb{R}$.

La droite achevée : $\mathbf{R} := \mathbb{R} \cup \{+\infty, -\infty\}$. Pour tout $x \in \mathbb{R}$ on a $-\infty < x < +\infty$. Les opérations arithmétiques standards ($+$, $-$ et $\cdot$) s'étendent naturellement à $\mathbf{R}$. Si $x \in \mathbb{R}$, alors (les lignes qui suivent doivent être interprétées en prenant soit toujours le signe "du haut", soit celui "du bas")

$$(\pm\infty) + (\pm\infty) = \pm\infty\,, \quad x \pm (\pm\infty) = (\pm\infty) \pm x = \pm\infty\,,$$
$$(\pm\infty) \cdot (\pm\infty) = +\infty\,, \quad (\pm\infty) \cdot (\mp\infty) = -\infty\,,$$
$$x \cdot (\pm\infty) = (\pm\infty) \cdot x = \begin{cases} \pm\infty & \text{si } x > 0, \\ \mp\infty & \text{si } x < 0, \\ 0 & \text{si } x = 0\,. \end{cases}$$

Obs : On ne définit jamais "$\infty - \infty$", et la convention $0 \cdot \infty = 0$ jouera un rôle important dans la suite.

Les rationnels : $\mathbb{Q}$. Si $I \subset \mathbb{R}$, $\mathbb{Q}_I := \mathbb{Q} \cap I$.

Limites : Si $x_n \in \mathbb{R}$ est une suite monotone non-décroissante convergeant vers x (évent. ∞), on notera $x_n \nearrow x$, ou encore $\lim_n \uparrow x_n = x$. De même, si $x_n \in \mathbb{R}$ est une suite monotone non-croissante convergeant vers x (évent. $-\infty$), on notera $x_n \searrow x$, ou encore $\lim_n \downarrow x_n = x$.

Ensembles. Si X est un ensemble, alors $\mathcal{P}(X)$ dénote la famille de tous les sous-ensembles de X ; $\mathcal{P}(X)$ contient aussi $\varnothing$. Le complémentaire de $A \in \mathcal{P}(X)$ s'écrit $A^c := X \backslash A$. Pour des ensembles A_α quelconques, on a les Relations de de Morgan :

$$\Big(\bigcup_\alpha A_\alpha\Big)^c = \bigcap_\alpha A_\alpha^c\,, \quad \Big(\bigcap_\alpha A_\alpha\Big)^c = \bigcup_\alpha A_\alpha^c\,.$$

La fonction indicatrice de A est définie par

$$1_A(x) := \begin{cases} 1 & \text{si } A \ni x\,, \\ 0 & \text{sinon.} \end{cases}$$

Si X, Y sont deux ensembles, et $f : X \to Y$, alors pour $B \subset Y$, $f^{-1}(B) := \{x \in X : f(x) \in B\}$. Pour toute collection d'ensembles $B_\alpha \subset Y$, on a $f^{-1}(B_\alpha^c) = (f^{-1}(B_\alpha))^c$, et

$$f^{-1}\Big(\bigcup_\alpha B_\alpha\Big) = \bigcup_\alpha f^{-1}(B_\alpha)\,, \quad f^{-1}\Big(\bigcap_\alpha B_\alpha\Big) = \bigcap_\alpha f^{-1}(B_\alpha)\,.$$

CHAPITRE 1

Introduction

Dans le but de prouver une affirmation concernant la convergence de la série de Fourier, Henri Léon Lebesgue (1875-1941) fut amené à développer une nouvelle théorie de l'intégration, qu'il présenta dans sa thèse **[11]**, intitulée *Intégrale, longueur, aire.* Cette théorie s'avéra plus tard avoir un champ d'application bien plus vaste que celle inventée par Newton et Leibniz au 17ème siècle, et développée par Riemann et ses contemporains au 19ème.

L'intégration au sens de Lebesgue est aujourd'hui utilisée dans pratiquement tous les domaines des mathématiques. Elle joue un rôle particulièrement important dans la formulation rigoureuse de la Théorie des Probabilités (Kolmogorov, 1933) et dans la Théorie des Systèmes Dynamiques.

Ce cours expose les grands chapitres de cette théorie, comme décrite dans la plupart des livres classiques sur le sujet. Une brève description bibliographique se trouve à la fin de cette introduction.

1. L'Intégrale selon Riemann

Rappelons brièvement la définition de l'intégrale définie de Riemann pour une fonction $f : [a, b] \to \mathbb{R}$, que l'on supposera positive et bornée pour simplifier.

Considérons une partition π de l'intervalle $[a, b]$: $x_0 \equiv a < x_1 < x_2 < \cdots < x_{n-1} < x_n \equiv b$, et définissons pour chaque intervalle $I_i := [x_i, x_{i+1})$ les deux nombres $m_i := \inf\{f(x) : x \in I_i\}$, $M_i := \sup\{f(x) : x \in I_i\}$. On définit alors les fonctions en escalier

$$f_\pi^- := \sum_{i=0}^{n-1} m_i 1_{I_i}\,, \quad f_\pi^+ := \sum_{i=0}^{n-1} M_i 1_{I_i}\,,$$

qui sont deux approximations de f satisfaisant $f_\pi^- \leq f \leq f_\pi^+$. On définit alors les sommes de Riemann

$$I(f_\pi^-) := \sum_{i=0}^{n-1} m_i \ell(I_i)\,, \quad I(f_\pi^+) := \sum_{i=0}^{n-1} M_i \ell(I_i)\,,$$

où $\ell(I)$ représente la longueur de l'intervalle I, c'est-à-dire

$$\ell([a,b)) := b - a\,. \tag{1.1}$$

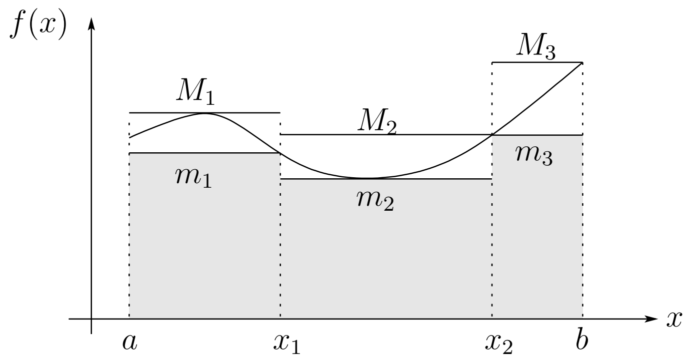


Figure 1. Une fonction $f : [a,b] \to \mathbb{R}$ et les deux approximations $f_\pi^\pm$ (ici, $\pi = (x_0 \equiv a, x_1, x_2, x_3 \equiv b)$), correspondant aux sommes de Riemann $I(f_\pi^\pm)$.

La fonction f est intégrable (au sens de Riemann) si

$$\inf_\pi I(f_\pi^+) = \sup_\pi I(f_\pi^-) < \infty\,.$$

Si f est intégrable, alors la valeur commune des deux nombres ci-dessus est notée

$$\int_a^b f(x)\,dx\,.$$

Le procédé qui mène à l'intégrale de Riemann d'une fonction f consiste donc à *discrétiser son domaine.*

Il est bien connu que les fonctions *continues* sont intégrables au sens de Riemann. Dans ce cas particulier, l'intégrale de Riemann garantit en plus l'existence d'une *primitive* de la fonction f :

THÉORÈME 1.1 (Théorème Fondamental de l'Analyse). *Si f est continue, alors la fonction*

$$x \mapsto \int_a^x f(s)ds\,,$$

est dérivable, et sa dérivée est égale à $f(x)$. En particulier, si F est une primitive quelconque de f, alors

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

Même si une fonction n'a pas besoin d'être continue pour être intégrable, l'ensemble de ses points de discontinuité ne peut pas être trop important : si f possède trop de discontinuités sur un petit intervalle I, alors les termes des sommes de Riemann pour lesquels $I_i \subset I$ se comportent mal dans la limite des partitions fines. Donc la continuité apparaît de façon essentielle dans la définition de l'intégrale de Riemann. En fait, Lebesgue a montré (on le verra dans le Théorème 3.28) qu'une fonction est intégrable au sens de Riemann *si et seulement si* l'ensemble de ses points de discontinuités est négligeable [1].

Malgré son grand succès rencontré en mathématique (calcul différentiel, géométrie analytique, ...), en physique (mécanique, thermodynamique, ...) et en ingénierie, l'intégrale de Riemann souffre de quelques défauts importants, dont nous allons énoncer les principaux.

Passages à la limite. La notion d'intégrabilité au sens de Riemann manque de robustesse, comme le montre l'exemple suivant. Soit $x_1, x_2, \dots$ une énumération des rationnels de l'intervalle $[0,1]$. Définissons $f_n : [0,1] \to \mathbb{R}$ par $f_n(x) := 1_{Q_n}(x)$, où $Q_n := \{x_1, \dots, x_n\}$. Par le théorème mentionné plus haut, chaque fonction f_n est intégrable, puisqu'elle possède un nombre fini de discontinuités. Par contre, la fonction limite $f(x) := \lim_{n\to\infty} f_n(x) \equiv 1_{\mathbb{Q}\cap[0,1]}(x)$ existe, est bornée, mais discontinue en *tout* point $x \in [0,1]$. Elle n'est donc pas intégrable.

Échange limite-intégrale. En analyse, on a souvent besoin de pouvoir échanger une limite avec une intégrale. Le résultat suivant est un des seuls qui donne une condition sous laquelle cet échange peut être fait pour l'intégrale de Riemann :

Soit $f_n : [0,1] \to \mathbb{R}$ une suite de fonctions intégrables convergeant uniformément vers une fonction f. Alors f est intégrable, et $\int_0^1 f(x)dx =$

1. La définition précise d'ensemble négligeable sera donnée plus bas.

$\lim_{n\to\infty}\int_0^1 f_n(x)dx$.
La condition de convergence uniforme est très forte, trop pour que ce théorème soit vraiment utile. Pour le voir, il suffit de considérer la suite de fonctions $f_n(x) := \frac{e^{-nx}}{\sqrt{x}}$ (avec $f(0)=0$). Il est facile de voir que f_n est intégrable (l'intégrale impropre $\int_0^1 f_n(x)dx$ existe pour tout n). De plus, comme $f_n(x)\to 0$ pour tout $x\in[0,1]$, il semble raisonnable d'affirmer que

$$\lim_{n\to\infty}\int_0^\infty f_n(x)dx = 0\,.$$

Pourtant la convergence $f_n\to 0$ n'est pas uniforme, et il faut donc prouver cette convergence "à la main".

Un autre exemple de résultat dont la preuve n'utilise que des outils classiques d'analyse mais requiert une certaine ingéniosité :
Soit f_n une suite de fonctions continues sur $[0,1]$, telle que $0\leq f_n\leq 1$, et telle que $f_n(x)\to 0$ pour tout $x\in[0,1]$. Alors

$$\int_0^1 f_n(x)dx \to 0\,.$$

Les deux exemples précédents sont donc des cas où l'échange de limites

$$\lim_{n\to\infty}\int_0^1 f_n(x)dx = \int_0^1 \lim_{n\to\infty} f_n(x)dx$$

peut être justifié, sans que la convergence soit uniforme.

Donc l'intégrale au sens de Riemann n'a pas de théorème général sur l'échange intégrale-limite, et chaque cas doit être considéré séparément.

Non-complétude de $C[a,b]$. Soit $C[a,b]$ l'espace vectoriel des fonctions continues sur $[a,b]$. Dans diverses situations, on s'intéresse à définir une norme sur $C[a,b]$ à partir de l'intégrale

$$\|f\| := \Big(\int_a^b |f(x)|^2 dx\Big)^{1/2}\,.$$

Par exemple, en analyse fonctionnelle, on s'intéresse à ce que muni de cette norme, l'espace $C[a,b]$ soit complet (i.e. que toute suite de Cauchy converge). Or il est facile de voir que la norme définie ci-dessus ne satisfait justement pas cette propriété.

L'intégrale de Lebesgue permet d'écarter chacune de ces difficultés.

2. L'idée originale de Lebesgue

Comme mentionné plus haut, la définition de l'intégrale de Riemann d'une fonction $f : [a,b] \to \mathbb{R}$ repose sur une approximation obtenue à partir de partitions de son *domaine* $[a,b]$. L'idée fondamentale à l'origine de la théorie de Lebesgue est *d'approximer la fonction par des partitions de son ensemble image*, c'est-à-dire $f([a,b])$.

Considérons une fonction $f : [a,b] \to \mathbb{R}$, que l'on supposera à nouveau positive et bornée pour simplifier. Il existe alors deux nombres $m < M$ tels que $m \leq f(x) \leq M$ pour tout $x \in [a,b]$. Considérons alors une partition Π de l'intervalle $[m,M]$: $y_0 \equiv m < y_1 \cdots < y_n \equiv M$.

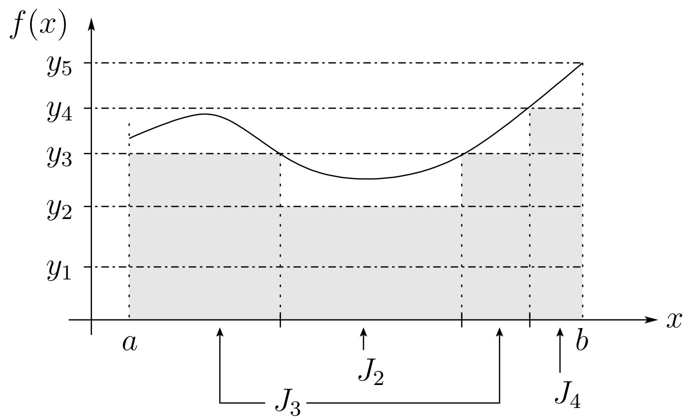


FIGURE 2. La même fonction que sur la Figure 1, approximée via un partitionnement de l'ensemble image.

À chaque intervalle $[y_i, y_{i+1})$ correspond un sous-ensemble du domaine de f donné par

$$J_i := f^{-1}([y_i, y_{i+1}))\,.$$

On définit alors une fonction étagée associée à f et à la partition Π :

$$\varphi^{\Pi} := \sum_{i=0}^{n-1} y_i 1_{J_i}\,, \tag{1.2}$$

et son intégrale :

$$\int \varphi^{\Pi} := \sum_{i=0}^{n-1} y_i \ell(J_i)\,. \tag{1.3}$$

On dit que f est intégrable au sens de Lebesgue si

$$\int f := \sup_{\Pi} \int \varphi^{\Pi} < \infty\,.$$

Sans que ça soit évident à première vue, cette nouvelle définition de l'intégrale s'avère être beaucoup plus générale que celle de Riemann, et la contient comme cas particulier (dans le sens où toute fonction intégrable au sens de Riemann est intégrable au sens de Lebesgue). Voyons quelle est la principale difficulté apparaissant avec cette nouvelle définition.

Suivant le comportement de f, les ensembles $J_i = f^{-1}([y_i, y_{i+1}))$ peuvent ne plus être des intervalles, et avoir une structure compliquée. Par exemple, si

$$f(x) := \begin{cases} \sin \frac{1}{x} & \text{si } x \in (0,1]\,, \\ 0 & \text{si } x = 0\,, \end{cases} \tag{1.4}$$

alors pour tout $[c,d) \subset [-1,1]$, $f^{-1}([c,d))$ est une union disjointe, infinie dénombrable, d'intervalles contenus dans $[0,1]$. Ou encore, si $f(x) := 1_{\mathfrak{K}}(x)$, où $\mathfrak{K}$ est l'ensemble de Cantor sur $[0,1]$, alors $f^{-1}([1-\delta, 1+\delta)) = \mathfrak{K}$, que l'on sait être un ensemble non-dénombrable, nulle part dense.

Donc pour pouvoir définir $\sum_{i=0}^{n-1} y_i \ell(J_i)$, on doit avant tout savoir définir précisément les nombres, "$\ell(J_i)$". On doit donc être en possession d'une *fonction d'ensemble* $\ell(\cdot)$, qui généralise la notion de *longueur* pour les intervalles, et qui coïncide avec celle-ci sur les intervalles. Cette fonction d'ensemble s'appelle *mesure*.

La mesure devra alors satisfaire à quelques propriétés fondamentales. Par exemple, pour la fonction (1.4), on doit s'assurer que *la mesure de* $f^{-1}([c,d))$ *est égale à la somme (infinie) des longueurs des intervalles.*

Donc, la construction de l'intégrale de Lebesgue se fera en deux étapes principales. En premier lieu, on commencera par élaborer une *Théorie de la Mesure* permettant de distinguer les ensembles de la droite auxquels on sait associer une notion de *mesure*, qui généralise la notion de longueur pour les intervalles. Ces ensembles seront appelés *mesurables.* Ensuite, on se restreindra aux fonctions f telles que tous les ensembles de la forme $f^{-1}(I)$ sont mesurables, et l'intégration de ces fonctions par rapport à la mesure de Lebesgue se fera en suivant les lignes décrites plus haut.

Nous l'avons vu, pour l'intégrale de Riemann, la continuité est un ingrédient central, et la convergence des intégrales vers l'intégrale de la limite n'est garantie qu'au prix d'une convergence uniforme. Or la continuité ne disparaît pas complètement dans la théorie de Lebesgue : même si la continuité de la fonction n'est plus essentielle, c'est la continuité de la *mesure* qui jouera un rôle fondamental dans tous les théorèmes de la théorie de l'intégration : des ensembles proches (dans un sens qui sera rendu précis dans le chapitre suivant) ont des mesures proches.

Dans un but pédagogique, nous construirons la mesure de Lebesgue sur l'intervalle $[0,1]$, dans la section suivante. Les méthodes introduites dans la construction de la mesure de Lebesgue n'étant pas restreintes aux sous-ensembles de la droite, nous les développerons en toute généralité dans le Chapitre 2. De même, la théorie de l'intégration par rapport à une mesure peut se faire en toute généralité ; nous y consacrerons le Chapitre 3.

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

Cette section est consacrée à la construction de la mesure de Lebesgue sur l'intervalle $[0,1]$. Ce cas particulier est celui sur lequel Lebesgue a originalement élaboré sa théorie, et il montre déjà toutes les difficultés rencontrées dans le problème plus général de construire des mesures sur des espaces abstraits. La construction de la mesure de Lebesgue en dimensions supérieures sera étudiée en détails dans la Section 6 du Chapitre 2.

Notre objectif est donc de pouvoir *mesurer* le plus possible de sous-ensembles de $[0,1]$. Avant de commencer, on peut se demander ce qu'on entend exactement par *mesure*, et quelles sont les propriétés qu'elle doit satisfaire. Citons de la Vallée Poussin **[6]** (Chapitre 2, page 1) :

> "La mesure des grandeurs repose sur un axiome fondamental : *le tout est la somme des parties.* Si une grandeur se partage en parties de même espèce, la mesure du tout sera la somme des mesures des parties. Euclide, qui a formulé cet axiome, a fondé sur

> lui la théorie de la mesure des figures de la géométrie élémentaire et tous ses successeurs ont fait de même [2].
>
> Toute généralisation de la théorie de la mesure, telle la mesure des ensembles, doit donc se proposer comme première condition de respecter le principe précédent.
>
> La mesure d'un ensemble E est un nombre, *positif ou nul*, attaché à cet ensemble et qui en dépend. On exprime cette correspondance en disant que la mesure est une *fonction de l'ensemble* E. D'après ce qui précède, cette fonction doit satisfaire à deux conditions essentielles :
>
> (1) Elle doit être *additive*, c'est-à-dire que la mesure d'une somme d'ensembles sans point commun deux à deux doit être la somme des mesures des parties.
>
> (2) Elle doit se réduire à la mesure au sens élémentaire dans le cas des figures élémentaires, segment, rectangle, etc.
>
> La première de ces deux conditions, *l'additivité*, peut être étendue de deux manières : une fonction est *additive au sens restreint* si elle n'est additive que pour un nombre fini de termes ; elle est *additive au sens complet* si elle est encore additive pour une infinité dénombrable de termes.
>
> Le progrès essentiel obtenu par MM. Borel et Lebesgue dans la théorie de la mesure est d'avoir réalisé l'additivité *au sens complet*. Toute la supériorité de leur théorie vient de là. Il importe toutefois de dire que la première idée de cette théorie revient à M. Borel. L'oeuvre propre de M. Lebesgue ne commence qu'avec les intégrales définies."

L'*additivité au sens complet* dont parle de la Vallée Poussin est communément appelée *additivité dénombrable* (ou σ-*additivité*) : si (A_n) est une famille dénombrable d'ensembles deux à deux disjoints, alors

$$\ell\Big(\bigcup_n A_n\Big) = \sum_n \ell(A_n)\,. \tag{1.5}$$

2. Pour définir la mesure des figures géométriques, il faut au principe précédent joindre celui-ci : des figures égales (superposables) ont même mesure.

Sur les réels, la mesure est donc une généralisation de la notion classique de longueur pour les intervalles. Or la longueur peut être vue comme une fonction d'ensemble

$$\ell : \{\text{intervalles de } [0,1]\} \to \mathbb{R}_+ ,$$

telle que pour un intervalle J de la forme $[a,b]$, $[a,b)$, $(a,b]$ ou (a,b),

$$\ell(J) = b - a . \tag{1.6}$$

Nous commencerons par définir une fonction d'ensemble sur $\mathcal{P}([0,1])$, c'est-à-dire sur toutes les parties de $[0,1]$, appelée *mesure extérieure* :

$$\ell^* : \mathcal{P}([0,1]) \to \mathbb{R}_+ ,$$

telle que

$$\ell^*(J) = \ell(J)$$

pour tout intervalle J. Pour obtenir l'additivité dénombrable, nous nous verrons obligés de *restreindre* ℓ^* à une sous-famille $\mathcal{M}^* \subset \mathcal{P}([0,1])$ sur laquelle (1.5) est satisfaite. Les ensembles de $\mathcal{M}^*$ seront appelés *mesurables*, et la restriction

$$\ell^* : \mathcal{M}^* \to \mathbb{R}_+$$

la *mesure de Lebesgue sur* $[0,1]$.

Les principales étapes qui mènent à la mesure de Lebesgue seront utilisées à nouveau dans le Chapitre 2, lors de la construction de mesures sur des espaces plus généraux.

3.1. Les ensembles négligeables. Pour commencer, donnons la définition d'ensemble négligeable, mentionnée plus haut.

Un recouvrement d'un ensemble quelconque $A \subset [0,1]$ est une famille dénombrable d'intervalles ouverts $\mathcal{I} = \{I\}$ telle que $A \subset \bigcup_{I\in\mathcal{I}} I$.

DÉFINITION 1.2. *A est* négligeable *si pour tout $\epsilon > 0$ il existe un recouvrement $\mathcal{I} = \{I\}$ de A tel que $\sum_{I\in\mathcal{I}} \ell(I) < \epsilon$.*

En d'autres termes, A est négligeable si

$$\inf \sum_{I\in\mathcal{I}} \ell(I) = 0 ,$$

où l'infimum est sur tous les recouvrements de A. Par exemple, un point $A = \{x\}$ est négligeable.

3.2. Mesure extérieure. Un procédé naturel pour mesurer un ensemble est de l'approximer par des ensembles plus simples que l'on sait déjà mesurer. Comme on l'a fait ci-dessus, on peut le recouvrir par des intervalles ouverts et chercher à minimiser la somme totale des intervalles.

DÉFINITION 1.3. *Soit $A \subset [0, 1]$. La* mesure extérieure de *A est définie par*

$$\ell^*(A) := \inf \sum_{I \in \mathcal{I}} \ell(I) \,, \tag{1.7}$$

où l'infimum est sur tous les recouvrements de A.

On peut vérifier (exercice) que les recouvrements utilisés pour définir ℓ^* peuvent être pris avec des intervalles quelconques, pas nécessairement ouverts ou fermés.

La mesure extérieure est donc une fonction d'ensemble sur $[0, 1]$, $\ell^* : \mathcal{P}([0, 1]) \to \mathbb{R}_+$. Pour tout $A \in \mathcal{P}([0, 1])$, le nombre $\ell^*(A)$ est tel que pour tout $\epsilon > 0$ on peut trouver un recouvrement $\mathcal{I}$ de A tel que

$$\sum_{I \in \mathcal{I}} \ell(I) - \epsilon \leq \ell^*(A) \leq \sum_{I \in \mathcal{I}} \ell(I) \,.$$

Les ensembles négligeables sont donc ceux pour lesquels $\ell^*(A) = 0$. Donnons quelques propriétés élémentaires de la mesure extérieure :

Normalisation : $\ell^*(\varnothing) = 0$.

Consistance : si $J \subset [0, 1]$ est un intervalle, alors $\ell^*(J) = \ell(J)$. Prenons par exemple $J = [a, b]$. Alors J peut être recouvert par $J' = (a-\delta, b+\delta)$, et donc $\ell^*(J) \leq \ell(J') = \ell(J)+2\delta$. Puisque δ est arbitraire, $\ell^*(J) \leq \ell(J)$. Ensuite, par définition, pour tout $\epsilon > 0$ il existe un recouvrement $\mathcal{I}$ de J tel que $\ell^*(J) \geq \sum_{I \in \mathcal{I}} \ell(I) - \epsilon$. Or clairement, $\sum_{I \in \mathcal{I}} \ell(I) \geq \ell(J)$ (exercice), ce qui montre, comme ϵ est arbitraire, que $\ell^*(J) \geq \ell(J)$.

Monotonicité : si $A \subset B$, alors $\ell^*(A) \leq \ell^*(B)$. En effet, tout recouvrement de B est aussi un recouvrement de A.

Par exemple, pour tout A, $\ell^*(A) \leq \ell^*([0, 1]) = \ell([0, 1]) = 1 - 0 = 1$.

Sous-additivité dénombrable : si $(A_n)_{n \geq 1}$ est une famille dénombrable, alors

$$\ell^*\Big(\bigcup_{n \geq 1} A_n\Big) \leq \sum_{n \geq 1} \ell^*(A_n) \,. \tag{1.8}$$

En particulier, si $(A_n)_{n=1,2,\dots,N}$ est finie, alors

$$\ell^*\Big(\bigcup_{n=1}^N A_n\Big) \leq \sum_{n=1}^N \ell^*(A_n)\,. \tag{1.9}$$

En effet, fixons $\epsilon > 0$. Pour chaque A_n il existe un recouvrement $\{I_{n,k}\}_{k\geq 1}$ tel que $\ell^*(A_n) \geq \sum_{k\geq 1} \ell(I_{n,k}) - \epsilon/2^n$. Il est clair que $\{I_{n,k}\}_{n,k}$ est un recouvrement de $\bigcup_{n\geq 1} A_n$. Donc

$$\begin{aligned}
\ell^*\Big(\bigcup_{n\geq 1} A_n\Big) \leq \sum_{n,k} \ell(I_{n,k}) &= \sum_{n\geq 1}\Big(\sum_{k\geq 1}\ell(I_{n,k})\Big) \\
&\leq \sum_{n\geq 1}\big(\ell^*(A_n) + \epsilon/2^n\big) \\
&\leq \sum_{n\geq 1}\ell^*(A_n) + \epsilon\,.
\end{aligned}$$

On obtient la deuxième affirmation en appliquant la première avec $A_k = \varnothing$ pour tout $k \geq N+1$.

La sous-additivité dénombrable implique en particulier que *les réunions dénombrables d'ensembles négligeables sont négligeables*, et que les ensembles dénombrables sont négligeables (le contraire n'est pas vrai, voir exercice). Aussi, $\ell^*(A) > 0$ seulement si A est non-dénombrable.

3.3. Les ensembles mesurables. La mesure extérieure généralise donc la notion de longueur pour les intervalles. Il nous reste à voir si elle satisfait à la condition (1.5) d'additivité dénombrable.

La propriété de sous-additivité finie (1.9) de ℓ^* implique que si A, B sont deux ensembles quelconques, alors

$$\ell^*(A\cup B) \leq \ell^*(A) + \ell^*(B)\,.$$

Si en plus ces ensembles sont *disjoints*, on s'attend à obtenir une égalité (voir la citation de de la Vallée Poussin) :

$$\ell^*(A\cup B) = \ell^*(A) + \ell^*(B)\,. \tag{1.10}$$

Cette identité peut être vérifiée dans certains cas simples, par exemple lorsque A et B sont à distance positive (exercice). Ou encore, on sait par consistance que $\ell^*([0,\frac12)) = \frac12$, $\ell^*([\frac12,1]) = \frac12$, et $\ell^*([0,1]) = 1$, donc

$$\ell^*([0,\tfrac12)) + \ell^*([\tfrac12,1]) = \ell^*([0,1])\,. \tag{1.11}$$

Mais de manière générale, on aimerait que la mesure extérieure satisfasse (1.10), au moins quand $B = A^c := [0,1]\backslash A$:

$$\ell^*(A) + \ell^*(A^c) = 1\,. \tag{1.12}$$

Cette propriété n'est malheureusement pas vraie pour *tout* ensemble $A \subset [0,1]$ (voir la Section 6.5 du Chapitre 2). Donc il n'y a aucun espoir de montrer que ℓ^* est σ-additive sur $\mathcal{P}([0,1])$.

Donc pour pouvoir utiliser effectivement ℓ^*, on doit la *restreindre* à une sous-classe de $\mathcal{P}([0,1])$. On pourrait par exemple se restreindre à la classe des ensembles A sur lesquels (1.12) est satisfaite. On verra (exercice) que la définition suivante, bien que plus abstraite, est équivalente et plus maniable :

DÉFINITION 1.4. *Un ensemble $A \subset [0,1]$ est* **mesurable (au sens de Lebesgue) par rapport à** ℓ^* *si*

$$\ell^*(E) = \ell^*(E \cap A) + \ell^*(E \cap A^c) \quad \forall E \subset [0,1]\,. \tag{1.13}$$

Il est clair que si A est mesurable, alors (1.12) est satisfaite. On peut en fait vérifier (exercice) que (1.13) et (1.12) sont deux définitions équivalentes de la mesurabilité.

Dans la suite, on dénotera par $\mathcal{M}^*$ la famille des ensembles mesurables par rapport à ℓ^*. Pour vérifier qu'un ensemble A est mesurable, il suffit de vérifier

$$\ell^*(E) \geq \ell^*(E \cap A) + \ell^*(E \cap A^c)\,. \tag{1.14}$$

En effet, l'inégalité contraire est satisfaite par sous-additivité. Commençons à étudier les propriétés de $\mathcal{M}^*$.

LEMME 1.5. $\mathcal{M}^*$ *est une* **algèbre** *:*

(1) $\varnothing \in \mathcal{M}^*$,

(2) Si $A \in \mathcal{M}^*$, alors $A^c \in \mathcal{M}^*$,

(3) Si $A_n \in \mathcal{M}^*$ pour tout $n = 1, \dots, N$, alors $\bigcup_{n=1}^N A_n \in \mathcal{M}^*$.

DÉMONSTRATION. Le premier point est évident. Le deuxième suit de la symétrie de (1.13) par rapport à A et A^c. Soient ensuite $A, B \in \mathcal{M}^*$. Pour montrer que $A \cup B \in \mathcal{M}^*$, on utilise la sous-additivité de ℓ^* : comme $E \cap (A \cup B) = \{E \cap A\} \cup \{E \cap (A^c \cap B)\}$, on a

$$\ell^*(E \cap (A \cup B)) \leq \ell^*(E \cap A) + \ell^*(E \cap A^c \cap B)\,.$$

Donc, par la mesurabilité de B, puis par celle de A,

$$\begin{aligned}
\ell^*(E\cap(A\cup B)) &+ \ell^*(E\cap(A\cup B)^c) \\
&\leq \ell^*(E\cap A)+\ell^*(E\cap A^c\cap B)+\ell^*(E\cap(A\cup B)^c) \\
&= \ell^*(E\cap A)+\underbrace{\ell^*(E\cap A^c\cap B)+\ell^*(E\cap A^c\cap B^c)}_{=\ell^*(E\cap A^c)} \\
&= \ell^*(E)\,,
\end{aligned}$$

donc $A\cup B\in\mathcal{M}^*$. La troisième affirmation s'obtient par induction. □

On montre ensuite que

Lemme 1.6. *Si $A_n\in\mathcal{M}^*$ ($n=1,\dots,N$) sont disjoints, alors pour tout $E\in\mathcal{P}([0,1])$:*

$$\ell^*\big(E\cap\bigcup_{n=1}^N A_n\big)=\sum_{n=1}^N\ell^*(E\cap A_n)\,. \tag{1.15}$$

*En particulier, ℓ^* est **additive** sur $\mathcal{M}^*$: si $A_n\in\mathcal{M}^*$ ($n=1,\dots,N$) sont disjoints, alors*

$$\ell^*\big(\bigcup_{n=1}^N A_n\big)=\sum_{n=1}^N\ell^*(A_n)\,. \tag{1.16}$$

Démonstration. Pour $N=2$, soient $A_1,A_2\in\mathcal{M}^*$ disjoints, et $B:=A_1\cup A_2$. Alors par la mesurabilité de A_2,

$$\begin{aligned}
\ell^*(E\cap B) &= \ell^*(E\cap\underbrace{B\cap A_2}_{=A_2})+\ell^*(E\cap\underbrace{B\cap A_2^c}_{=A_1}) \\
&= \ell^*(E\cap A_1)+\ell^*(E\cap A_2)\,.
\end{aligned}$$

On peut ensuite itérer ce procédé et obtenir (1.15), qui implique (1.16) en prenant $E=[0,1]$. □

Il nous reste à vérifier que ces propriétés s'étendent à des unions infinies dénombrables.

Lemme 1.7. $\mathcal{M}^*$ *est une **tribu** :*

(1) $\varnothing\in\mathcal{M}^*$,

(2) Si $A\in\mathcal{M}^*$, alors $A^c\in\mathcal{M}^*$,

(3) Si $A_n\in\mathcal{M}^*$ pour tout $n\geq 1$, alors $\bigcup_{n\geq1}A_n\in\mathcal{M}^*$.

Démonstration. Les deux premières propriétés ont été prouvées plus haut. Pour la troisième, soient $A_n\in\mathcal{M}^*$, que l'on commence par supposer disjoints. Soit $A := \bigcup_n A_n$ et $B_N := \bigcup_{n\leq N} A_n \in \mathcal{M}^*$. Pour tout E,

$$\begin{aligned}\ell^*(E) &= \ell^*(E\cap B_N^c) + \ell^*(E\cap B_N)\\ &= \ell^*(E\cap B_N^c) + \sum_{n\leq N} \ell^*(E\cap A_n)\\ &\geq \ell^*(E\cap A^c) + \sum_{n\leq N} \ell^*(E\cap A_n)\,.\end{aligned}$$

En prenant $N\to\infty$, et par sous-σ-additivité,

$$\begin{aligned}\ell^*(E) &\geq \ell^*(E\cap A^c) + \sum_{n} \ell^*(E\cap A_n)\\ &\geq \ell^*(E\cap A^c) + \ell^*(E\cap A)\,,\end{aligned}$$

ce qui montre que $A\in\mathcal{M}^*$. Si les ensembles A_n ne sont pas disjoints, on définit $B_1 := A_1$, $B_n := A_n\cap A_1^c\cap\cdots\cap A_{n-1}^c$ ($n\geq 2$), qui sont disjoints, et se réduit au cas précédent puisque $\bigcup_n A_n = \bigcup_n B_n \in \mathcal{M}^*$. □

Pour finir,

LEMME 1.8. *La fonction d'ensemble* $\ell^* : \mathcal{M}^* \to [0,1]$ *est une* **mesure** *:*

(1) $\ell^*(\varnothing) = 0$,

(2) ℓ^* *est* σ-**additive** *: si* $(A_n)_{n\geq 1}$ *est une famille disjointe d'ensembles mesurables, alors* $\ell^*(\bigcup_n A_n) = \sum_n \ell^*(A_n)$.

DÉMONSTRATION. Soient $A_n\in\mathcal{M}^*$, disjoints deux-à-deux, $A := \bigcup_n A_n$. Par sous-additivité, $\ell^*(A)\leq\sum_n \ell^*(A_n)$. Mais pour tout N, par additivité et monotonicité, $\sum_{n\leq N}\ell^*(A_n) = \ell^*(\bigcup_{n\leq N} A_n)\leq \ell^*(A)$. On conclut en prenant $N\to\infty$. □

Même si $\mathcal{M}^*$ est une classe d'ensembles sur laquelle la mesure extérieure est σ-additive, elle pourrait être vide ! Il est donc important de vérifier qu'elle est suffisamment riche en ensemble. En particulier, on aimerait qu'elle contienne au moins les intervalles :

LEMME 1.9. $\mathcal{M}^*$ *contient les intervalles.*

DÉMONSTRATION. Soit par exemple $A = (a,b)\subset[0,1]$ (les autres cas se traitent de façon similaire), et $E\subset[0,1]$ un ensemble quelconque. Soit $\epsilon>0$. Alors il existe un recouvrement $\mathcal{I} = \{I\}$ de E tel que $\ell^*(E)\geq\sum_I \ell(I)-\epsilon$. Définissons alors les intervalles ouverts $I_- := I\cap A$, $I_+ := I\cap A^c$. Par la définition même de longueur pour les intervalles,

$$\ell(I) = \ell(I_-) + \ell(I_+)\,.$$

Or il est clair que $\{I_-\}$ est un recouvrement de $E \cap A$, et que $\{I_+\}$ est un recouvrement de $E \cap A^c$. Donc

$$\sum_I \ell(I) = \sum_{I_-} \ell(I_-) + \sum_{I_+} \ell(I_+) \geq \ell^*(E \cap A) + \ell^*(E \cap A^c)\,.$$

Donc $\ell^*(E) \geq \ell^*(E \cap A) + \ell^*(E \cap A^c) - \epsilon$. □

Puisque c'est une tribu, et comme elle contient les intervalles, la classe $\mathcal{M}^*$ contient également tous les ouverts (qui sont des unions dénombrables d'intervalles ouverts), tous les fermés (dont les complémentaires sont des ouverts), et a fortiori tous les compacts. Donc $\mathcal{M}^*$ contient, de fait, *beaucoup* d'ensembles. On peut donc se demander si *tous* les sous-ensembles de $[0,1]$ sont mesurables. Ce n'est pas le cas ; nous en donnerons un exemple d'ensemble non-mesurable dans la Section 6.5.

Même s'il n'existe pas de caractérisation explicite de tous les ensembles mesurables, nous voyons que tout ensemble $A \subset [0,1]$ qui peut être *construit* est mesurable. Ici, par *construit* on entend *construit à partir d'intervalles, par une suite d'opérations élémentaires, à savoir : unions dénombrables, intersections dénombrables, complémentaires.*

DÉFINITION 1.10. *La mesure extérieure ℓ^*, restreinte aux ensembles mesurables $\mathcal{M}^*$, est appelée la* **mesure de Lebesgue sur** $[0,1]$. *On l'écrira λ^1.*

Par construction, sur les intervalles, λ^1 coïncide avec ℓ. On verra plus tard que λ^1 est l'unique mesure ayant cette propriété.

Remarquons qu'à part dans la preuve de la consistance et du Lemme 1.9, on a assez peu utilisé les particularités de $[0,1]$: la restriction de la mesure extérieure à la classe des ensembles mesurables est un procédé tout à fait général et sera réutilisé dans le Chapitre 2.

4. À propos de la bibliographie

La littérature sur la Théorie de la Mesure et de l'Intégration est vaste.

Les ouvrages les plus accessibles sont certainement le petit livre de Bartle [**1**], ainsi que le livre de Kolmogorov et Fomine [**10**]. Le livre de Tao [**17**], plus récent, est également recommandé.

Certaines références, plus complètes, sont par exemple Folland [**7**], Rana [**14**], ou le livre classique de Rudin [**15**]. Très précis, l'ouvrage de Gapaillard [**8**], ou encore celui de Bauer [**2**].

Comme la Théorie de la Mesure est intimement liée à la Théorie des Probabilités, ces deux théories sont souvent développées en parallèle. Les ouvrages classiques de Billingsley [**3**], Durrett [**13**] ou de Shiryaev [**16**] en sont un bon exemple. Une autre bonne référence est le livre de Neveu [**12**].

À titre de complément, on pourra également consulter l'ouvrage plus ancien de Halmos [**9**], le double pavé de Bogachev [**4**], ainsi que le livre original de Lebesgue [**11**], ou celui de de la Vallée Poussin [**6**].

Pour un texte profond, avec une approche historique particulièrement détaillée, on pourra consulter l'ouvrage de Bressoud [**5**].

CHAPITRE 2

Tribus et Mesures

Dans l'introduction, la recherche d'une fonction d'ensemble σ-additive sur $[0,1]$ a conduit à distinguer une classe $\mathcal{M}^*$ de sous-ensembles appelés mesurables, formant une tribu. C'est en restreignant la mesure extérieure ℓ^* aux mesurables que nous avons obtenu la mesure de Lebesgue. Le triplet $([0,1], \mathcal{M}^*, \ell^*)$ est appelé *espace mesuré*.

On pourrait passer maintenant à l'intégration des fonctions $f : [0,1] \to \mathbb{R}$. Or la structure "(ensemble, tribu, mesure)" peut être définie sur des ensembles plus abstraits que $[0,1]$. En fait, la construction de la plus grand partie de la théorie de l'intégration selon Lebesgue est basée sur les propriétés élémentaires de la mesure, essentiellement l'additivité dénombrable, et non sur la structure détaillée de l'ensemble sous-jacent. On a donc avantage, avant de passer à l'intégration des fonctions sur $[0,1]$, à développer la théorie de la mesure sur des espaces plus généraux.

Ceci sera fait dans le présent chapitre : nous définirons les espaces mesurés $(X, \mathcal{F}, \mu)$ abstraits, donnerons leurs propriétés et le moyen de les construire. Dans le chapitre suivant, la théorie de l'intégration par rapport à μ sera développée en toute généralité, pour des fonctions à valeurs réelles $f : X \to \mathbb{R}$. Cette théorie contiendra l'intégration des fonctions $f : [0,1] \to \mathbb{R}$ par rapport à la mesure de Lebesgue comme cas particulier.

1. Tribus

Soit X un ensemble. Comme précédemment, nous noterons $\mathcal{P}(X)$ la famille des parties $A \subset X$. On suppose toujours que $\varnothing \in \mathcal{P}(X)$. Le complémentaire d'un ensemble $A \in \mathcal{P}(X)$ s'écrira $A^c := X \backslash A$.

DÉFINITION 2.1. *Une famille de sous ensembles de X, $\mathcal{F} \subset \mathcal{P}(X)$, est une* **tribu**[a] *si elle satisfait aux conditions suivantes.*

(1) $\varnothing \in \mathcal{F}$.

(2) Si $A \in \mathcal{F}$, *alors* $A^c \in \mathcal{F}$.

(3) Si $A_n \in \mathcal{F}$ *pour tout* n, *alors* $\bigcup_n A_n \in \mathcal{F}$.

Les éléments de $\mathcal{F}$ *sont appelés* **ensembles mesurables**, *et la paire* $(X, \mathcal{F})$ *est appelée* **espace mesurable**[b].

a. Dans la littérature, on trouve aussi σ-algèbre, ou σ-algèbre de Boole. En anglais : σ-field, σ-algebra.

b. On peut interpréter cette terminologie en sachant que les ensembles mesurables sont ceux auxquels on pourra associer une mesure. Une tribu peut donc être considérée comme le *domaine* d'une mesure.

Observons que (1) et (2) impliquent que $X \in \mathcal{F}$. De plus, par (2) et (3), $\bigcap_n A_n = (\bigcup_n A_n^c)^c \in \mathcal{F}$. Donc une tribu est une famille d'ensembles qui contient X, et qui est stable sous complémentaires, sous unions dénombrables et sous intersections dénombrables.

La tribu la plus simple sur un ensemble X est la famille $\mathcal{F} := \{\varnothing, X\}$, appelée tribu triviale. Ensuite, si $E \subset X$ est non-vide et différent de X, alors $\mathcal{F} := \{\varnothing, X, E, E^c\}$ est aussi une tribu, contenant quatre ensembles mesurables.

En général, une tribu contient "beaucoup" d'ensembles. Il est clair que la tribu la plus riche en sous-ensembles de X est celle qui contient *tous* les sous-ensembles : $\mathcal{F} := \mathcal{P}(X)$, appelée tribu discrète. Cette tribu est utilisée en général quand X est fini ou dénombrable, par exemple quand $X = \mathbb{N}$. Mais quand X n'est pas dénombrable, la tribu discrète contient trop d'ensembles et n'est en général pas utilisable[1].

EXEMPLE 2.2. La famille $\mathcal{F}$ contenant les sous-ensembles $E \subset X$ tels que ou E est dénombrable, ou E^c est dénombrable, est une tribu (exercice).

Dans la pratique, on a souvent besoin qu'une tribu contienne certains ensembles particuliers. Sur les réels par exemple, il est naturel de vouloir une tribu qui contienne les intervalles, et on peut se poser la question : sur $X = \mathbb{R}$, quelle est la plus petite tribu qui contient les intervalles ? Il existe un procédé simple permettant de répondre à cette question :

1. Nous verrons plus loin qu'il n'est en général pas possible de définir une mesure sur tous les sous-ensembles de X.

LEMME 2.3. *Soit* $\mathcal{C} \subset \mathcal{P}(X)$ *une famille de sous-ensembles. Il existe une tribu minimale contenant* $\mathcal{C}$*, que l'on notera* $\sigma(\mathcal{C})$*, appelée la tribu* engendrée par $\mathcal{C}$. *Elle est donnée par* $\sigma(\mathcal{C}) := \bigcap\{\mathcal{A} : \mathcal{A} \supset \mathcal{C}, \mathcal{A} \text{ tribu}\}$.

Ce procédé sera utilisé surtout pour définir des tribus sur des ensembles de nature continue, comme dans l'exemple suivant :

EXEMPLE 2.4. Considérons $X = \mathbb{R}^d$. Soit $\mathcal{O}$ la famille des ouverts de $\mathbb{R}^d$ (par rapport à la métrique de la distance Euclidienne). La tribu $\mathcal{B}(\mathbb{R}^d) := \sigma(\mathcal{O})$ s'appelle la tribu des ensembles Boréliens, ou simplement la tribu des Boréliens. On peut aussi définir les Boréliens de n'importe quel sous-ensemble de $\mathbb{R}^d$. Par exemple, $\mathcal{B}([0, 1])$.

En général, si X est un espace topologique (par exemple un espace métrique) dont les ouverts sont notés $\mathcal{O}$, on appellera également $\mathcal{B}(\mathcal{O}) := \sigma(\mathcal{O})$ la tribu des Boréliens.

EXEMPLE 2.5. Considérons ensuite le cas $X = \mathbb{R}^\infty$, c'est-à-dire l'espace des suites de nombres réels, $x = (x_1, x_2, \dots)$. Soit $\pi_n : \mathbb{R}^\infty \to \mathbb{R}^n$ la projection canonique définie par $\pi_n(x_1, x_2, \dots) := (x_1, x_2, \dots, x_n)$. Pour tout $n \geq 1$, on pose $\mathcal{C}_n := \pi_n^{-1}(\mathcal{B}(\mathbb{R}^n))$. Alors $\mathcal{C} := \bigcup_{n\geq 1} \mathcal{C}_n$ est appelée l'algèbre des cylindres. On définit alors une tribu sur X par $\mathcal{F} := \sigma(\mathcal{C})$.

EXEMPLE 2.6. Soient $(X, \mathcal{F})$ et $(Y, \mathcal{G})$ deux espaces mesurables, et

$$X \times Y := \{(x, y) : x \in X, y \in Y\}$$

leur produit cartésien. On peut définir une tribu naturelle sur $X \times Y$ en considérant la famille des rectangles

$$\mathcal{R} := \big\{A \times B : A \in \mathcal{F}, B \in \mathcal{G}\big\} .$$

Alors, $\mathcal{F} \otimes \mathcal{G} := \sigma(\mathcal{R})$ est la tribu produit. L'espace mesurable $(X \times Y, \mathcal{F} \otimes \mathcal{G})$ est l'espace produit. Les produits seront étudiés dans le Chapitre 6.

REMARQUE 2.1. Il est important de souligner que le procédé décrit dans le Lemme 2.3 est utile pour introduire des tribus jouissant de certaines propriétés, mais *non-constructif* : en général, on ne peut pas caractériser exactement les ensembles contenus dans $\sigma(\mathcal{C})$.

EXEMPLE 2.7. Une application permet de lever une structure mesurable d'un ensemble sur un autre ensemble. Soit X un ensemble, $(Y, \mathcal{G})$ un espace mesurable, et $f : X \to Y$ une application. Alors $f^{-1}(\mathcal{G}) := \{f^{-1}(G) : G \in \mathcal{G}\}$ est une tribu sur X, appelée la tribu engendrée par f.

1.1. Applications Mesurables.

DÉFINITION 2.8. *Soient* $(X, \mathcal{F})$, $(Y, \mathcal{G})$ *deux espaces mesurables. Une application* $f : X \to Y$ *est dite* $(\mathcal{F}, \mathcal{G})$-**mesurable** *(ou plus simplement* **mesurable***) si la préimage d'un mesurable est un mesurable, c'est-à-dire si* $B \in \mathcal{G}$ *implique* $f^{-1}(B) \in \mathcal{F}$.

La façon dont on définit la mesurabilité d'une application rappelle la *continuité* en topologie : si $(X, \mathcal{X})$, $(Y, \mathcal{Y})$ sont deux espaces topologiques ($\mathcal{X}$ et $\mathcal{Y}$ sont les ouverts de X, resp. Y), une fonction $f : X \to Y$ est dite continue si la préimage d'un ouvert de Y est un ouvert de X.

Il est facile de vérifier que la composition de deux fonctions mesurables est mesurable. Pour vérifier qu'une fonction est mesurable on utilisera souvent le résultat suivant :

LEMME 2.9. *Soient* $(X, \mathcal{F})$, $(Y, \mathcal{G})$ *deux espaces mesurables, et soit* $f : X \to Y$. *Si une famille* $\mathcal{C} \subset \mathcal{P}(Y)$ *est telle que* $\sigma(\mathcal{C}) = \mathcal{G}$, *et telle que* $f^{-1}(C) \in \mathcal{F}$ *pour tout* $C \in \mathcal{C}$, *alors* f *est* $(\mathcal{F}, \mathcal{G})$*-mesurable.*

DÉMONSTRATION. $\mathcal{G}_0 := \{B \in \mathcal{G} : f^{-1}(B) \in \mathcal{F}\} \subset \mathcal{G}$ est une tribu. Comme elle contient $\mathcal{C}$, elle contient aussi $\sigma(\mathcal{C})$. Donc $\mathcal{G}_0 = \mathcal{G}$. □

En particulier, une application continue entre deux espaces topologiques $(X, \mathcal{O}_X)$, $(Y, \mathcal{O}_Y)$, est $(\mathcal{B}(\mathcal{O}_X), \mathcal{B}(\mathcal{O}_Y))$-mesurable.

2. Mesures

DÉFINITION 2.10. *Soit* $(X, \mathcal{F})$ *un espace mesurable. Une fonction d'ensemble* $\mu : \mathcal{F} \to [0, +\infty]$ *est une* **mesure** *si elle satisfait aux conditions suivantes :*

(1) $\mu(\varnothing) = 0$,

(2) μ *est* σ-**additive** *: si* A_n *est une suite d'ensembles mesurables deux-à-deux disjoints alors*

$$\mu\Big(\bigcup_{n\geq 1} A_n\Big) = \sum_{n\geq 1} \mu(A_n)\,. \tag{2.1}$$

Le triplet $(X, \mathcal{F}, \mu)$ *est appelé* **espace mesuré***.*

En particulier, (2.1) implique que μ est **additive** : si $A, B \in \mathcal{F}$ sont disjoints, alors $\mu(A \cup B) = \mu(A) + \mu(B)$.

En général, la construction d'un espace mesuré non-trivial est une tâche délicate. En particulier, notre premier exemple d'espace mesuré,

$([0,1], \mathcal{M}^*, \ell^*)$, a utilisé plusieurs propriétés fondamentales de la droite.

Parmi les façons triviales de définir une mesure sur un ensemble X, mentionnons les exemples suivants. Si $\mu(A) := 0$ pour tout $A \in \mathcal{F}$, alors μ est une mesure. Ou encore, si $\mu(A) := 0$ si $A = \varnothing$, $+\infty$ sinon, alors μ est aussi une mesure. Passons à des exemples plus intéressants.

Exemple 2.11. Sur X muni d'une tribu quelconque, fixons $x \in X$ et définissons, pour tout $A \in \mathcal{F}$,

$$\delta_x(A) = 1_A(x) .$$

Alors δ_x est une mesure appelée masse de Dirac en x.

Exemple 2.12. Si $A \in \mathcal{F}$ est un ensemble fini, posons $\mu_c(A) := \mathrm{card}(A)$. Si A est infini, $\mu_c(A) := +\infty$. Alors μ_c est appelée mesure de comptage. Sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, $\mu_c(A)$ compte simplement le nombre de points contenus dans A.

Exemple 2.13. On peut généraliser la mesure de comptage en considérant une famille de nombres non-négatifs $\rho = (\rho(k))_{k\in\mathbb{N}}$, et définir $\mu_\rho(A) := \sum_{k\in A} \rho(k)$. La mesure de comptage correspond à $\rho(k) := 1$.

Exemple 2.14. Soit $(X, \mathcal{F}, \mu)$ un espace mesuré, $(Y, \mathcal{G})$ un espace mesurable. Une application mesurable $f : X \to Y$ permet de *transporter* μ de X sur Y. En effet,

$$\nu := \mu \circ f^{-1}$$

définit une mesure sur $(Y, \mathcal{G})$, appelée transport de μ par f.

On verra que certaines propriétés d'un espace mesurable $(X, \mathcal{F}, \mu)$ peuvent dépendre fortement de la masse que la mesure associe à l'espace tout entier. La mesure est finie si $\mu(X) < \infty$, et σ-finie si il existe une suite $X_n \in \mathcal{F}$, $\mu(X_n) < \infty$, telle que $X = \bigcup_n X_n$. La mesure de comptage sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ est σ-finie. Dans l'Exemple 2.13, en prenant $\rho(k) = k^{-2}$ on obtient une mesure μ_ρ finie. Par contre si $\rho(k) \equiv \epsilon > 0$, alors μ_ρ est seulement σ-finie. La mesure de comptage sur $([0,1], \mathcal{P}([0,1]))$ n'est pas σ-finie [2].

Une mesure est une probabilité si $\mu(X) = 1$; dans ce cas, $(X, \mathcal{F}, \mu)$ est appelé espace de probabilité. Par exemple, $([0,1], \mathcal{M}^*, \ell^*)$ est un espace de probabilité, qui modélise l'expérience qui consiste à tirer un nombre $x \in [0,1]$ au hasard (uniformément). Dans l'Exemple 2.13, on obtient

2. En effet, si elle l'était, cela signifie que $[0,1]$ pourrait s'écrire comme une union dénombrable d'ensembles finis, ce qui est impossible puisque $[0,1]$ n'est pas dénombrable.

un espace de probabilité si ρ satisfait à la condition $\sum_k \rho(k) = 1$.

Nous commencerons par donner quelques propriétés générales de la mesure, puis passerons ensuite à la construction de mesures.

2.1. Continuité. Dans cette section et la suivante, on suppose donné un espace mesuré $(X, \mathcal{F}, \mu)$.

Si $A_n \in \mathcal{F}$ est une suite croissante ($A_n \subset A_{n+1}$), alors $A := \bigcup_n A_n \in \mathcal{F}$, et on écrit $A_n \nearrow A$. De même, si $A_n \in \mathcal{F}$ est une suite décroissante ($A_{n+1} \subset A_n$), alors $A_n \searrow A$, où $A := \bigcap_n A_n \in \mathcal{F}$.

Proposition 2.15 (Sous-additivité, continuité).

1. *Si $A, B \in \mathcal{F}$, avec $A \subset B$, alors $\mu(A) \leq \mu(B)$.*
2. *Si $A_n \in \mathcal{F}$ pour tout n, alors $\mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$.*
3. *Si $A_n \in \mathcal{F}$ est croissante, $A_n \nearrow A$, alors $\mu(A_n) \nearrow \mu(A)$.*
4. *Si $A_n \in \mathcal{F}$ est décroissante, $A_n \searrow A$, et si $\mu(A_1) < \infty$, alors $\mu(A_n) \searrow \mu(A)$.*

Démonstration. (1) Si $C := B \backslash A$, alors $\mu(B) = \mu(A) + \mu(C) \geq \mu(A)$. (2) On peut écrire $\bigcup_n A_n = \bigcup_n B_n$, où $B_1 := A_1$, et $B_n := A_n \cap (A_1 \cup \cdots \cup A_{n-1})^c$ pour $n \geq 2$. Comme les B_n sont mesurables et disjoints, on a $\mu(\bigcup_n A_n) = \sum_n \mu(B_n)$. Mais $B_n \subset A_n$, et donc $\mu(B_n) \leq \mu(A_n)$, ce qui prouve la deuxième affirmation. (3) Observons que si $B_1 := A_1$ et (pour $n \geq 2$) $B_n := A_n \backslash A_{n-1}$, alors les B_n sont disjoints, $A_n = \bigcup_{k=1}^n B_k$, et $\bigcup_n B_n = A$. Donc

$$\mu(A) = \mu\Big(\bigcup_k B_k\Big) = \sum_k \mu(B_k) = \lim_{n\to\infty} \sum_{k=1}^n \mu(B_k) = \lim_{n\to\infty} \mu(A_n)\,.$$

(4) La suite $B_n := A_1 \backslash A_n$ est croissante et $B_n \nearrow A_1 \backslash A =: B$. Par (3) on a donc $\lim_n \mu(A_n) = \mu(A_1) - \lim_n \mu(B_n) = \mu(A_1) - \mu(B) = \mu(A)$. $\square$

Observons que la condition $\mu(A_1) < \infty$ dans (4) est nécessaire. En effet, en prenant $X = \mathbb{N}$ avec μ la mesure de comptage et $A_n = \{n, n+1, \dots\}$, alors $A_n \searrow \varnothing$, mais $\mu_c(A_n) = \infty$ pour tout n.

2.2. Les ensembles de mesure nulle ; complétude. Un fait inévitable, en Théorie de la Mesure, est la présence d'ensembles non-vides mais invisibles du point de vue de la mesure :
On définit les ensembles de mesure nulle :

$$\mathcal{N}_\mu := \{N \in \mathcal{F} : \mu(N) = 0\}\,, \tag{2.2}$$

Ainsi que les **ensembles négligeables (par rapport à μ)**, définis par

$$\mathcal{N}eg_\mu := \{B \in \mathcal{P}(X) : \exists N \in \mathcal{N}_\mu, N \supset B\}. \tag{2.3}$$

Les ensembles négligeables ne peuvent être ignorés ; ils joueront un rôle central dans toute la théorie de l'intégration. Leur propriété fondamentale, facile à vérifier, est la suivante.

LEMME 2.16. *Les unions dénombrables d'ensembles négligeables (ou de mesure nulle) sont négligeables.*

Dans la définition donnée ci-dessus, les ensembles négligeables ne sont pas nécessairement mesurables. On le verra plus tard, il est souvent nécessaire de travailler avec un espace mesuré dans lequel tous les ensembles négligeables sont mesurables. Dans ce cas, on dira que $(X, \mathcal{F}, \mu)$ est **complet**. Or tout espace mesurable peut être complété, comme le montre la proposition suivante.

PROPOSITION 2.17. *Soit $(X, \mathcal{F}, \mu)$ un espace mesurable. La tribu*

$$\overline{\mathcal{F}}_\mu := \sigma(\mathcal{F} \cup \mathcal{N}eg_\mu) \tag{2.4}$$

est la **tribu complétée par rapport à μ**. *On a*

$$\overline{\mathcal{F}}_\mu = \{A \in \mathcal{P}(X) : \exists B, B' \in \mathcal{F} : B \subset A \subset B',\, \mu(B' \backslash B) = 0\}. \tag{2.5}$$

Soit $A \in \overline{\mathcal{F}}_\mu$. Si B, B' sont comme dans (2.5), on définit

$$\overline{\mu}(A) := \mu(B) = \mu(B'). \tag{2.6}$$

Alors $\overline{\mu}$ définit une mesure sur $(X, \overline{\mathcal{F}}_\mu)$. L'espace mesuré

$$\overline{(X, \mathcal{F}, \mu)} := (X, \overline{\mathcal{F}}_\mu, \overline{\mu}) \tag{2.7}$$

est complet, appelé le **complété** *de $(X, \mathcal{F}, \mu)$. De plus, $\overline{\mu}$ est l'unique mesure sur $(X, \overline{\mathcal{F}}_\mu)$ qui coïncide avec μ sur $\mathcal{F}$.*

DÉMONSTRATION. Soit $\mathcal{F}_1$ la famille définie par le membre de droite de (2.5). Il est facile de voir que $\mathcal{F}_1$ est une tribu. Puisqu'elle contient $\mathcal{F}$ et $\mathcal{N}eg_\mu$, on a $\overline{\mathcal{F}}_\mu \subset \mathcal{F}_1$. Soit alors $A \in \mathcal{F}_1$ et $B, B' \in \mathcal{F}$ tels que $B \subset A \subset B'$, $\mu(B' \backslash B) = 0$. En écrivant $A = B \cup (A \cap B^c)$, puisque $B \in \mathcal{F}$ et $A \cap B^c \subset B' \backslash B$, on a $A \cap B^c \in \mathcal{N}eg_\mu$, et donc $A \in \overline{\mathcal{F}}_\mu$. Donc $\mathcal{F}_1 \subset \overline{\mathcal{F}}_\mu$.

On vérifie que le nombre $\overline{\mu}(A)$ en (2.6) est bien défini, c'est-à-dire ne dépend pas du choix de B, B'. En effet, soient C, C' deux autres ensembles mesurables tels que $C \subset A \subset C'$, $\mu(C' \backslash C) = 0$. Par monotonicité, $\mu(C) \leq \mu(B') = \mu(B)$. De même, $\mu(B) \leq \mu(C)$. Donc $\mu(C) = \mu(B)$, et aussi $\mu(C') = \mu(B')$.

Finalement, on montre facilement que $\overline{\mu}$ est une mesure sur $(X, \overline{\mathcal{F}}_\mu)$. $\square$

Au vu du résultat précédent, on pourra toujours supposer que les espaces mesurables considérés sont complets. On vérifie facilement (exercice) que $([0,1], \mathcal{M}^*, \ell^*)$ est complet.

On dit qu'une propriété dépendant de $x \in X$ est vraie μ-presque partout si l'ensemble des points $x \in X$ pour lesquels elle n'est pas vérifiée est négligeable par rapport à μ. Par exemple, si f, g sont deux fonctions sur X, alors "$f = g$ presque partout" signifie que $\{x \in X : f(x) \neq g(x)\}$ est négligeable.

3. Un Critère d'unicité

Une mesure peut être vue comme une collection de nombres $\{\mu(A) : A \in \mathcal{F}\}$, et on peut se demander s'il est nécessaire de connaître *tous* ces nombres pour connaître μ. En d'autres termes : est-ce que la connaissance de $\mu(C)$ pour certains ensembles mesurables C appartenant à une sous-classe $\mathcal{C} \subset \mathcal{F}$ permet de reconstruire μ sur toute la tribu $\mathcal{F}$?

Théorème 2.18. *Soient μ et ν deux mesures sur $(X, \mathcal{F})$, et soit $\mathcal{C} \subset \mathcal{F}$ une famille d'ensembles stable par intersection (i.e. $A, B \in \mathcal{C} \Rightarrow A \cap B \in \mathcal{C}$) telle que $\sigma(\mathcal{C}) = \mathcal{F}$ et telle que $\mu(C) = \nu(C)$ pour tout $C \in \mathcal{C}$.*

(1) *Si $\mu(X) = \nu(X) < \infty$, alors $\mu = \nu$.*
(2) *Si il existe une suite $X_n \in \mathcal{C}$, $X_n \nearrow X$ telle que $\mu(X_n) = \nu(X_n) < \infty$, alors $\mu = \nu$.*

La preuve sera basée sur la notion de classe de Dynkin [3].

Définition 2.19. *Une famille $\mathcal{D} \subset \mathcal{P}(X)$ est une* classe de Dynkin *si*

(1) *$X \in \mathcal{D}$.*
(2) *Si $A, B \in \mathcal{D}$, $A \subset B$, alors $B \backslash A \in \mathcal{D}$*
(3) *$A_n \in \mathcal{D}$, $A_n \nearrow A$ implique $A \in \mathcal{D}$.*

Observons qu'une classe de Dynkin est aussi stable sous complémentation (donc en particulier $\varnothing \in \mathcal{D}$) et sous intersections dénombrables décroissantes, et que toute tribu est une classe de Dynkin. En outre,

Lemme 2.20. *Si $\mathcal{D}$ est une classe de Dynkin stable par intersections (finies), alors c'est une tribu.*

Démonstration. Soient $A_n \in \mathcal{D}$. Si $\mathcal{D}$ est stable par intersections, elle est aussi stable par réunions finies, ce qui implique que

3. Dans la littérature, on trouvera aussi l'appellation *classe monotone*.

$B_N := \bigcup_{n\leq N} A_n \in \mathcal{D}$. On a $B_N \nearrow \bigcup_n A_n$, et puisque $\mathcal{D}$ est une classe de Dynkin, on a $\bigcup_n A_n \in \mathcal{D}$, et donc $\mathcal{D}$ est une tribu. □

Comme dans le Lemme 2.3 pour les tribus, on peut considérer la plus petite classe de Dynkin $\delta(\mathcal{C})$ contenant une famille donnée d'ensembles $\mathcal{C} \subset \mathcal{P}(X)$. $\delta(\mathcal{C})$ est appelée la classe de Dynkin engendrée par $\mathcal{C}$. On a alors :

Théorème 2.21. *Si $\mathcal{C} \subset \mathcal{P}(X)$ est stable par intersections (finies), alors*

$$\delta(\mathcal{C}) = \sigma(\mathcal{C}) .$$

Démonstration. Clairement, $\delta(\mathcal{C}) \subset \sigma(\mathcal{C})$. Pour voir si $\delta(\mathcal{C}) \supset \sigma(\mathcal{C})$, il faut vérifier que $\delta(\mathcal{C})$ est une tribu. Pour montrer que $\delta(\mathcal{C})$ est une tribu, par le Lemme 2.20 il suffit de montrer qu'elle est stable par intersections finies.
Fixons $C \in \mathcal{C}$ et posons

$$\mathcal{D}_1 := \{B \in \delta(\mathcal{C}) : B \cap C \in \delta(\mathcal{C})\} .$$

Il est facile de voir que $\mathcal{D}_1$ est une classe de Dynkin. Puisque $\mathcal{C}$ est stable par intersections, on a $\mathcal{C} \subset \mathcal{D}_1$, et donc $\mathcal{D}_1 \supset \delta(\mathcal{C})$. Ceci implique que pour tout $C \in \mathcal{C}$ et tout $B \in \delta(\mathcal{C})$, on a $C \cap B \in \delta(\mathcal{C})$. Fixons alors $B \in \delta(\mathcal{C})$ et posons encore

$$\mathcal{D}_2 := \{A \in \delta(\mathcal{C}) : A \cap B \in \delta(\mathcal{C})\} .$$

Comme avant, on montre aussi facilement que $\mathcal{D}_2$ est une classe de Dynkin contenant $\mathcal{C}$, et donc que $\mathcal{D}_2 \supset \delta(\mathcal{C})$. Ceci implique que pour tout $A \in \delta(\mathcal{C})$ et pour tout $B \in \delta(\mathcal{C})$, $A \cap B \in \delta(\mathcal{C})$. Donc $\delta(\mathcal{C})$ est stable par intersections. □

Preuve du Théorème 2.18 : (1) : Soit $\mathcal{D} := \{A \in \mathcal{F} : \mu(A) = \nu(A)\} \subset \mathcal{F}$. On a $X \in \mathcal{D}$, et si $A, B \in \mathcal{D}$, $A \subset B$, alors comme ces ensembles ont des mesures finies,

$$\mu(B\backslash A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B\backslash A) ,$$

et donc $B\backslash A \in \mathcal{D}$. Puis, si $A_n \in \mathcal{D}$, $A_n \nearrow A$, alors par continuité,

$$\mu(A) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \nu(A_n) = \nu(A) ,$$

et donc $A \in \mathcal{D}$. Donc $\mathcal{D}$ est une classe de Dynkin. Comme elle contient $\mathcal{C}$ et que $\mathcal{C}$ est stable par intersections, on a $\mathcal{D} \supset \delta(\mathcal{C}) = \sigma(\mathcal{C}) = \mathcal{F}$ par le Théorème 2.21. Donc $\mathcal{D} = \mathcal{F}$.
(2) : Soient $\mu_n(\cdot) := \mu(\cdot \cap X_n)$ et $\nu_n(\cdot) := \nu(\cdot \cap X_n)$. Comme $\mu_n(C) = \nu_n(C)$ pour tout $C \in \mathcal{C}$ et que $\mu_n(X) = \nu_n(X) < \infty$, on peut appliquer

(1), et conclure que $\mu_n = \nu_n$. Ensuite, pour tout $A \in \mathcal{F}$, on a $\mu(A) = \lim_n \mu_n(A) = \lim_n \nu_n(A) = \nu(A)$, ce qui prouve que $\mu = \nu$. □

4. Le Théorème d'Extension de Carathéodory

Cette section décrit le moyen généralement utilisé pour construire une mesure sur un espace mesurable. Il suit essentiellement la méthode présentée dans l'introduction pour construire la mesure de Lebesgue sur $[0,1]$, qui s'applique dans le cadre plus général d'un espace mesurable abstrait $(X, \mathcal{F})$.

4.1. Mesures extérieures.

DÉFINITION 2.22. *Soit X un ensemble. Une fonction d'ensemble $\mu^* : \mathcal{P}(X) \to [0, +\infty]$ est une* **mesure extérieure** *si elle satisfait aux conditions suivantes :*

(1) $\mu^*(\varnothing) = 0$,

(2) Si $A \subset B$, alors $\mu^*(A) \leq \mu^*(B)$.

(3) μ^* est **sous-additive** *: si $(A_n)_{n\geq 1}$ est une suite quelconque, alors*

$$\mu^*\Big(\bigcup_{n\geq 1} A_n\Big) \leq \sum_{n\geq 1} \mu^*(A_n)\,. \tag{2.8}$$

Nous l'avons vu, la fonction d'ensemble

$$\ell^*(A) := \inf \sum_I \ell(I)\,, \tag{2.9}$$

où l'infimum porte sur tous les recouvrements dénombrables de A par des intervalles, et où $\ell(I)$ est la longueur de l'intervalle I, définit une mesure extérieure sur $X = [0,1]$. Ce procédé peut être généralisé :

EXEMPLE 2.23. Soit $\mathcal{C} \subset \mathcal{P}(X)$ une famille contenant $\varnothing$ et recouvrant X, dans le sens où $X = \bigcup_n C'_n$ pour une suite $C'_n \in \mathcal{C}$. Soit $\rho : \mathcal{C} \to [0, +\infty]$ une fonction d'ensemble telle que $\rho(\varnothing) = 0$. Soit, pour $A \in \mathcal{P}(X)$,

$$\rho^*(A) := \inf \sum_j \rho(C_j)\,, \tag{2.10}$$

où l'infimum porte sur tous les recouvrements finis ou dénombrables de A par des éléments $C_j \in \mathcal{C}$. Alors ρ^* est une mesure extérieure (exercice).

Une mesure extérieure est en général utilisée dans le but de pouvoir en extraire une mesure. On procède donc comme dans l'Introduction, en la restreignant à une famille plus petite que $\mathcal{P}(X)$.

Définition 2.24. *Soit μ^* est une mesure extérieure sur X. Un ensemble $A \in \mathcal{P}(X)$ est dit* **mesurable (par rapport à μ^*)** *si*

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c), \quad \forall E \in \mathcal{P}(X). \tag{2.11}$$

Théorème 2.25. *Soit $\mathcal{M}^*$ la famille des ensembles mesurables par rapport à μ^*. Alors*

(1) $\mathcal{M}^*$ est une tribu.

(2) La restriction de μ^* à $\mathcal{M}^*$ est une mesure.

De plus, l'espace mesuré $(X, \mathcal{M}^*, \mu^*)$ est complet.

Démonstration. Les deux premières affirmations se prouvent comme les Lemmes 1.7 et 1.8. Il reste à montrer que $(X, \mathcal{M}^*, \mu^*)$ est complet. Soit donc $A \in \mathcal{P}(X)$ un ensemble négligeable. Il existe $N \in \mathcal{M}^*$ tel que $A \subset N$ et $\mu^*(N) = 0$. Donc aussi $\mu^*(A) = 0$, et pour tout E on peut écrire

$$\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A^c) \leq \mu^*(E),$$

donc $A \in \mathcal{M}^*$. □

4.2. Algèbres et pré-mesures. Dans les cas concrets, une mesure extérieure est définie à partir d'une fonction d'ensemble ρ déjà connue sur une classe d'ensembles $\mathcal{C}$, comme dans l'Exemple 2.23. Sur $[0, 1]$ par exemple, on a défini ℓ^* en supposant que l'on savait calculer la longueur des intervalles, $\ell[a, b] := b - a$.

Or en général, sans autres conditions sur la paire $(\rho, \mathcal{C})$, on ne sait pas si $\mathcal{M}^*$ contient la classe originale $\mathcal{C}$, et encore moins si la mesure extérieure ρ^* coïncide avec ρ sur $\mathcal{C}$. Nous allons donc donner plus de structure à la fonction d'ensemble initiale.

Définition 2.26. *Une famille $\mathcal{A} \subset \mathcal{P}(X)$, est une* **algèbre** *(ou* **algèbre de Boole***) si elle satisfait aux conditions suivantes.*

(1) $\varnothing \in \mathcal{A}$.

(2) Si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$.

(3) Si $A, B \in \mathcal{A}$, alors $A \cup B \in \mathcal{A}$ (et donc $A \cap B \in \mathcal{A}$).

La différence avec une tribu est qu'une algèbre est stable sous intersections et réunions *finies*.

DÉFINITION 2.27. *Soit $\mathcal{A}$ une algèbre sur X. Une fonction d'ensemble $\rho : \mathcal{A} \to [0, +\infty]$ est une* **pré-mesure** *si elle satisfait aux conditions suivantes :*

(1) $\rho(\varnothing) = 0$,

(2) Si A_n est suite d'éléments de $\mathcal{A}$ disjoints deux-à-deux, et *si $\bigcup_n A_n \in \mathcal{A}$, alors*

$$\rho\Big(\bigcup_{n\geq 1} A_n\Big) = \sum_{n\geq 1} \rho(A_n)\,. \tag{2.12}$$

Comme pour les mesures, (2.12) implique que ρ est **additive** : si $A, B \in \mathcal{A}$ sont disjoints, alors $\rho(A \cup B) = \rho(A) + \rho(B)$. Mais dans le cas de réunions *dénombrables*, comme $\mathcal{A}$ n'est pas nécessairement stable par réunions dénombrables, il est important de supposer que $\bigcup_n A_n \in \mathcal{A}$.

PROPOSITION 2.28. *Soit ρ une pré-mesure sur une algèbre $\mathcal{A}$. Soient ρ^* la mesure extérieure obtenue comme en* (2.10), *et $\mathcal{M}^*$ la tribu des ensembles mesurables par rapport à ρ^*. Alors*

(1) $\mathcal{M}^* \supset \mathcal{A}$.

(2) La restriction de ρ^* à $\mathcal{A}$ coïncide avec ρ.

DÉMONSTRATION. (1) Soit $A \in \mathcal{A}$, et $E \in \mathcal{P}(X)$. Par la définition de $\rho^*(E)$, pour tout $\epsilon > 0$ il existe un recouvrement de E par des ensembles $B_n \in \mathcal{A}$ tel que $\rho^*(E) \geq \sum_n \rho(B_n) - \epsilon$. Mais, puisque ρ est additive sur $\mathcal{A}$, on peut écrire $\rho(B_n) = \rho(B_n \cap A) + \rho(B_n \cap A^c)$. Puis, comme les ensembles $B_n \cap A \in \mathcal{A}$ recouvrent $E \cap A$, on a $\sum_n \rho(B_n \cap A) \geq \rho^*(E \cap A)$. De même, $\sum_n \rho(B_n \cap A^c) \geq \rho^*(E \cap A^c)$. Donc $\rho^*(E) \geq \rho^*(E \cap A) + \rho^*(E \cap A^c) - \epsilon$, ce qui montre que $A \in \mathcal{M}^*$ puisque ϵ est arbitraire. (2) Soit encore $A \in \mathcal{A}$. Comme $\{A\}$ recouvre A, on a $\rho^*(A) \leq \rho(A)$. Pour montrer l'inégalité inverse, considérons un recouvrement quelconque de A par une suite $B_n \in \mathcal{A}$. Soient $B_n' := B_n \cap A \in \mathcal{A}$. Alors les ensembles $D_1 := B_1'$ et $D_n := B_n' \cap (D_1 \cup \dots D_{n-1})^c \in \mathcal{A}$ sont disjoints, et $\bigcup_n D_n = A$. Donc, comme ρ est une pré-mesure,

$$\rho(A) = \rho\Big(\bigcup_n D_n\Big) = \sum_n \rho(D_n) \leq \sum_n \rho(B_n') \leq \sum_n \rho(B_n)\,.$$

Par conséquent, $\rho(A) \leq \rho^*(A)$, ce qui implique $\rho^*(A) = \rho(A)$. □

Comme pour les mesures, une pré-mesure sur une algèbre $\mathcal{A}$ est dite **σ-finie** s'il existe une suite $X_n \in \mathcal{A}$ telle que $X = \bigcup_n X_n$, $\rho(X_n) < \infty$.

THÉORÈME 2.29 (Théorème d'Extension de Carathéodory). *Soit $\mathcal{A}$ une algèbre sur X, et ρ une pré-mesure sur $(X, \mathcal{A})$. Soit $\mathcal{F} := \sigma(\mathcal{A})$ la tribu engendrée par $\mathcal{A}$. Alors il existe une mesure μ sur $(X, \mathcal{F})$, appelée* **extension de** *ρ, qui coïncide avec ρ sur $\mathcal{A}$. Si ρ est σ-finie, cette extension est unique.*

DÉMONSTRATION. Soit ρ^* la mesure extérieure associée à ρ, et $\mathcal{M}^*$ la tribu des ensembles mesurables par rapport à ρ^*. Comme $\mathcal{M}^* \supset \mathcal{A}$, on a aussi $\mathcal{M}^* \supset \sigma(\mathcal{A}) \equiv \mathcal{F}$. Soit alors μ la restriction de ρ^* à $\mathcal{F}$. Par la Proposition 2.28, $\mu(A) = \rho^*(A) = \rho(A)$ pour tout $A \in \mathcal{A}$.

Supposons que ρ soit σ-finie. Soit ν une mesure sur $(X, \mathcal{F})$ qui coïncide avec μ sur $\mathcal{A}$. Comme $\mathcal{A}$ est stable sous intersections, $\mu = \nu$ par le par le point (2) du Théorème 2.18. □

Donc par construction on sait que la mesure μ, qui étend ρ, est en fait donnée par ρ^*.

5. Mesures de Lebesgue-Stieltjes sur $\mathbb{R}$

La mesure de Lebesgue sur $[0, 1]$ est construite à partir de la notion classique de longueur pour les intervalles :

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

Plus généralement, sur la droite réelle, on peut considérer une fonction $g : \mathbb{R} \to \mathbb{R}$ non-décroissante, et pondérer la mesure des intervalles à l'aide de g :

$$\ell_g([a, b]) := g(b) - g(a)\,. \tag{2.13}$$

Supposons que l'on veuille étendre ℓ_g à une mesure sur la droite, à l'aide du Théorème de Carathéodory. Puisque ℓ_g est définie sur les intervalles, on souhaite l'étendre à la plus petite tribu contenant les intervalles, c'est-à-dire la tribu des Boréliens, $\mathcal{B}(\mathbb{R})$. Pour appliquer le Théorème 2.29, on donne aux intervalles une structure d'algèbre et à ℓ_g celle de pré-mesure.

Soit $\mathcal{I}$ la famille des intervalles de la droite ouverts à gauche et fermés à droite, c'est-à-dire de la forme $(a, b]$, (a, ∞) ou $(-\infty, b]$, avec $a < b$ finis. Il est clair que l'intersection de deux éléments de $\mathcal{I}$ est à nouveau un élément de $\mathcal{I}$, et que le complément d'un élément de $\mathcal{I}$ peut s'écrire comme une union finie disjointe d'éléments de $\mathcal{I}$. Donc $\mathcal{A}$, la famille des unions finies disjointes d'éléments de $\mathcal{I}$, est une algèbre. Il est facile de montrer que cette algèbre génère les Boréliens de la droite, $\mathcal{B}(\mathbb{R})$ (exercice).

Remarque 2.2. $\mathcal{A}$ peut également être vue comme la plus petite algèbre sur $\mathbb{R}$ contenant les intervalles bornés $(a,b]$.

Soit ensuite $g:\mathbb{R}\to\mathbb{R}$ une fonction non-décroissante, et $g(a^+):=\lim_{x\to a^+}\downarrow g(x)$, $g(a^-):=\lim_{x\to a^-}\uparrow g(x)$. Définissons aussi $g(\infty):=\sup_x g(x)$ et $g(-\infty):=\inf_x g(x)$, qui peuvent a priori être $+\infty$, ou $-\infty$. On utilise g pour définir, pour tout $(a,b]\in\mathcal{I}$ (sans exclure les cas $a=-\infty$ et $b=+\infty$),

$$\ell_g\big((a,b]\big):=g(b)-g(a)\,. \tag{2.14}$$

Proposition 2.30. *Soit $g:\mathbb{R}\to\mathbb{R}$ comme ci-dessus et continue à droite : $g(a^+)=g(a)$ pour tout $a\in\mathbb{R}$. Pour tout $A=\bigcup_{j=1}^n(a_n,b_n]\in\mathcal{A}$, posons*

$$\rho(A):=\sum_{j=1}^n\ell_g((a_j,b_j])\,. \tag{2.15}$$

Alors ρ définit une prémesure sur $\mathcal{A}$.

Démonstration. Un ensemble $A\in\mathcal{A}$ peut toujours s'écrire de plusieurs manières comme une union disjointe d'intervalles [4]. On doit donc vérifier que $\rho(A)$ ne dépend pas de la représentation choisie pour A. Prenons pour commencer un intervalle $A=(a,b]$, et écrivons-le sous la forme d'une union finie disjointe d'intervalles adjacents : $(a,b]=\bigcup_{j=1}^n(a_j,b_j]$, où $a=a_1<b_1=a_2<b_2\dots a_n<b_n=b$. En téléscopant les termes de la somme,

$$\begin{aligned}\sum_{j=1}^n\ell_g((a_j,b_j])&=\sum_{j=1}^n g(b_j)-g(a_j)\\&=\dots\\&=g(b_n)-g(a_1)=g(b)-g(a)=\ell_g((a,b])\end{aligned} \tag{2.16}$$

Soient maintenant $\bigcup_i I_i=\bigcup_j J_j=A$ deux façons d'écrire A comme union finie disjointe d'intervalles (pas nécessairement adjacents). Alors pour chaque i, $I_i=\bigcup_j I_i\cap J_j$, et pour chaque j, $J_j=\bigcup_i J_j\cap I_i$. Donc,

4. Par exemple, $(0,1]=(0,\frac12]\cup(\frac12,1]$, ou $(0,1]=(0,\frac13]\cup(\frac13,\frac45]\cup(\frac45,1]$, etc.

par (2.16),

$$\begin{aligned}\sum_i \ell_g(I_i) &= \sum_i \Big(\sum_j \ell_g(I_i \cap J_j) \Big) \\ &= \sum_j \Big(\sum_i \ell_g(I_i \cap J_j) \Big) \\ &= \sum_j \ell_g(I_j)\,.\end{aligned}$$

Ceci montre que ρ est bien définie.

Par définition, ρ est additive sur $\mathcal{A}$, et il reste à vérifier l'additivité dénombrable. On peut se restreindre au cas d'un intervalle $A = (a, b] \in \mathcal{I}$, mis sous la forme d'une union dénombrable disjointe $\bigcup_{n\geq 1} I_n$, où $I_n = (a_n, b_n]$. Puisque A peut s'écrire $A = \bigcup_{n=1}^N I_n \cup \bigcup_{n>N} I_n$, et que chacune de ces unions appartient à $\mathcal{A}$, on peut borner, pour tout N,

$$\rho(A) \geq \rho\Big(\bigcup_{n=1}^N I_n \Big) = \sum_{n=1}^N \rho(I_n)\,.$$

En prenant $N \to \infty$,

$$\rho(A) \geq \sum_{n\geq 1} \rho(I_n)\,.$$

Pour montrer l'inégalité inverse, commençons par le cas où a et b sont tous les deux finis. Fixons $\epsilon > 0$. Comme g est continue à droite, il existe $a' > a$ tel que $g(a') \leq g(a)+\epsilon$, et $b'_n > b_n$ tel que $g(b'_n) \leq g(b_n)+\epsilon/2^n$. Il est clair que les intervalles $I'_n = (a_n, b'_n)$ recouvrent l'intervalle compact $[a', b]$. On peut donc en extraire un sous-recouvrement fini I'_{n_k}. Pour simplifier l'écriture, supposons que ce sous-recouvrement est $I'_1, \dots, I'_K$, que $a_1 < a_2 < \dots < a_K$, et que $a_{j+1} < b'_j$, comme sur la Figure 1.

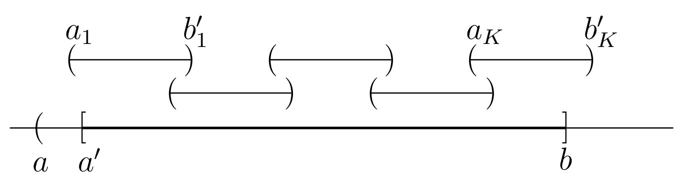


FIGURE 1. Représentation schématique du recouvrement de $[a', b]$ par les intervalles $I'_j = (a_j, b'_j)$, $j = 1, \dots, K$.

On a $\rho(A) = g(b) - g(a) \leq g(b) - g(a') + \epsilon \leq g(b'_K) - g(a_1) + \epsilon$. Or

$$\begin{aligned} g(b'_K) - g(a_1) &= g(b'_K) - g(a_K) + \sum_{j=1}^{K-1} (g(a_{j+1}) - g(a_j)) \\ &\leq g(b'_K) - g(a_K) + \sum_{j=1}^{K-1} (g(b'_j) - g(a_j)) \\ &\leq g(b_K) - g(a_K) + \sum_{j=1}^{K-1} (g(b_j) - g(a_j)) + \epsilon \\ &= \sum_{j=1}^{K} \rho(I_j) + \epsilon \\ &\leq \sum_{n\geq 1} \rho(I_n) + \epsilon\,. \end{aligned}$$

Comme $\epsilon > 0$ est arbitraire, ceci montre que $\rho(A) \leq \sum_{n\geq 1} \rho(I_n)$.

Dans le cas où une des bornes de l'intervalle A est infinie, on se ramène au cas précédent. Par exemple, si $A = (-\infty, b]$ et I_n un recouvrement, pour tout $L > 0$ on obtient, par le même procédé que celui utilisé ci-dessus, $\rho((-L, b]) \leq \sum_{n\geq 1} \rho(I_n)$. En prenant $L \to \infty$, $\rho((-L, b]) \to \rho((-\infty, b])$. □

On peut maintenant considérer l'extension de ρ à la tribu engendrée par $\mathcal{A}$:

THÉORÈME 2.31. *Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction non-décroissante et continue à droite. Alors il existe une unique mesure μ_g sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, appelée* mesure de Lebesgue-Stieltjes associée à *g, telle que*

$$\mu_g((a, b]) = g(b) - g(a)\,. \tag{2.17}$$

Inversément, si μ est une mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, finie sur les ensembles bornés, alors en définissant $g(x) := \mu((0, x])$ pour $x > 0$, $g(0) := 0$, et $g(x) := -\mu((x, 0])$ pour $x < 0$, alors g est non-décroissante, continue à droite, et $\mu_g = \mu$.

DÉMONSTRATION. Par la Proposition 2.30, ρ définie en (2.15) est une pré-mesure sur l'algèbre des unions finies disjointes d'intervalles de $\mathcal{I}$. Comme $\mathbb{R} = \bigcup_{n\in\mathbb{Z}} (n, n + 1]$, ρ est σ-finie. Par le Théorème 2.29, ρ possède une unique extension à la tribu engendrée par $\mathcal{A}$, qui dans ce cas coïncide avec celle des Boréliens. On note cette extension μ_g.

Inversément, si μ est une mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, finie sur les ensembles bornés, alors la fonction $g(x)$ définie dans l'énoncé est non-décroissante et continue à droite (par la continuité de μ). Comme μ_g coïncide avec μ sur les intervalles de $\mathcal{I}$, le Théorème 2.29 implique $\mu_g = \mu$. $\square$

Puisque la mesure μ_g est obtenue en *restreignant* la mesure extérieure de $\mathcal{M}_g^*$ (les ensembles mesurables par rapport à la mesure extérieure associéé à ρ) aux Boréliens, on voit que l'on peut en fait considérer μ_g comme une mesure sur $\mathcal{B}(\mathbb{R})$ aussi bien que sur $\mathcal{M}_g^*$. Or $\mathcal{M}_g^*$ peut être a priori plus grande que $\mathcal{B}(\mathbb{R})$.

Lorsque $g = \mathrm{id}$, c'est-à-dire $g(x) = x$ pour tout $x \in \mathbb{R}$, la mesure construite dans le Théorème 2.31 est la mesure de Lebesgue sur $\mathbb{R}$; on la notera

$$\lambda^1 := \mu_{\mathrm{id}} .$$

On reviendra à la mesure de Lebesgue dans la section suivante.

Pour les mesures *finies* sur la droite, le théorème ci-dessus est en général énoncé de la façon suivante, plus commune en théorie des probabilité :

THÉORÈME 2.32.

(1) *Soit μ une mesure finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, et soit $F_\mu(x) := \mu((-\infty, x])$. Alors $F_\mu : \mathbb{R} \to \mathbb{R}_+$ est non-décroissante, bornée, continue à droite et $F_\mu(-\infty) := \lim_{x \searrow -\infty} F_\mu(x) = 0$.*

(2) *Inversément, si $F : \mathbb{R} \to \mathbb{R}_+$ est une fonction non-décroissante, bornée, continue à droite et $F(-\infty) = 0$, alors il existe une unique mesure μ_F sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, telle que $F_{\mu_F} = F$.*

Quand μ est une probabilité, F_μ est appelée la fonction de répartition de μ. Ce résultat montre que l'étude d'une mesure de probabilité sur la droite peut se réduire à celle de sa fonction répartition F_μ, pour laquelle d'autres outils d'analyse peuvent être utilisés. Par exemple, la convergence d'une suite de mesures de probabilité sur la droite peut être caractérisée par la convergence de la suite de ses fonctions de répartition. Ceci est très utilisé dans les théorèmes limites en probabilité, notamment dans la preuve du Théorème Central Limite.

6. La mesure de Lebesgue sur $\mathbb{R}^d$

La construction de la mesure de Lebesgue sur la droite, dans la section précédente, s'étend sans encombres au cas de dimensions supérieures :

$\mathbb{R}^d$ avec $d \geq 2$. La mesure de Lebesgue sur $\mathbb{R}^d$, avec $d \geq 2$, sera notée λ^d.

Dans cette section nous donnerons quelques-unes de ses propriétés, comme l'invariance sous translations, son unicité et sa régularité. Dans le Chapitre 6, on donnera une autre construction de λ^d, en définissant le produit $\lambda^d := \lambda^1 \otimes \cdots \otimes \lambda^1$ (d fois).

6.1. Les Boréliens. La tribu naturelle sur $\mathbb{R}^d$ est la plus petite qui contient tous les ouverts, appelée la tribu des Boréliens (voir Exemple 2.4), $\mathcal{B}(\mathbb{R}^d) := \sigma(\mathcal{O})$.

Il est utile de savoir que les Boréliens peuvent être générés par des familles d'ensembles dont la structure est plus simple que celle des ouverts.

Un pavé est un ensemble de la forme

$$P = J_1 \times J_2 \times \cdots \times J_d\,,$$

où chaque J_i est un intervalle de $\mathbb{R}$. Si tous les J_i sont ouverts, le pavé est ouvert ; s'ils sont tous fermés, le pavé est fermé. S'ils sont ouverts à gauche et fermés à droite, le pavé est semi-ouvert. Une autre famille d'ensembles est donnée par les quadrants :

$$Q = (-\infty, b_1] \times (-\infty, b_2] \times \cdots \times (-\infty, b_d]\,.$$

Finalement, les semi-espaces sont les ensembles de la forme $H^i_\lambda = \{x : x \cdot e_i \leq \lambda\}$, où $\lambda \in \mathbb{R}$ et e_i est un des vecteurs unitaires de la base canonique de $\mathbb{R}^d$.

LEMME 2.33. *Considérons les familles de $\mathcal{P}(\mathbb{R}^d)$ définies par*

$$\mathcal{P} := \{\textit{pavés}\}\,,\ \mathcal{Q} := \{\textit{quadrants}\}\,,\ \mathcal{H} := \{\textit{semi-espaces}\}\,,$$

où il est entendu que les pavés de $\mathcal{P}$ peuvent être ouverts, fermés ou semi-ouverts. Alors $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{P}) = \sigma(\mathcal{Q}) = \sigma(\mathcal{H})$.

La preuve est laissée en exercice.

6.2. La mesure. Sur $[0, 1]$, la mesure extérieure des parties était définie par des recouvrements dénombrables par des intervalles. En dimensions supérieures, on utilise les pavés (voir Figure 2).

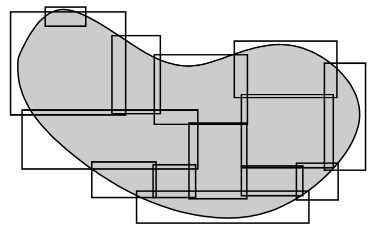


FIGURE 2. Recouvrement par des pavés en dimension 2.

Le volume d'un pavé $P = J_1 \times \cdots \times J_d$, dont l'intervalle J_i a des extrémités $a_i \leq b_i$ (ouvertes ou fermées), est donné par

$$\mathrm{vol}(P) := \prod_{j=1}^{d} (b_j - a_j)\,.$$

Soit $\mathcal{A}$ la plus petite algèbre sur $\mathbb{R}^d$ engendrée par les pavés (voir Remarque 2.2). Comme $\mathcal{P} \subset \mathcal{A} \subset \mathcal{B}(\mathbb{R}^d)$, $\mathcal{A}$ génère aussi les Boréliens de $\mathbb{R}^d$:

$$\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R}^d)\,.$$

On peut, comme pour ℓ_g dans la Proposition 2.28, étendre naturellement $\mathrm{vol}(\cdot)$ sur $\mathcal{A}$ et montrer qu'elle satisfait aux propriétés d'une prémesure. L'extension de $\mathrm{vol}(\cdot)$ obtenue s'appelle la mesure de Lebesgue sur $\mathbb{R}^d$. Puisqu'elle est obtenue en restreignant la mesure extérieure associée à $\mathrm{vol}(\cdot)$, on la note

$$\lambda^d := \mathrm{vol}^*\,.$$

Plus explicitement, si B est un Borélien, ou un ensemble mesurable (par rapport à vol^*),

$$\lambda^d(B) = \mathrm{vol}^*(B) := \inf \sum_P \mathrm{vol}(P)\,, \qquad (2.18)$$

où l'infimum porte sur les recouvrements de B par des familles dénombrables de pavés. En dimension $d = 1$, $\mathrm{vol}(\cdot)$ coïncide avec $\ell(\cdot)$ et $\mathrm{vol}^*(\cdot)$ avec $\ell^*(\cdot)$.

Notons $\mathcal{M}^* \subset \mathcal{P}(\mathbb{R}^d)$ la tribu des ensembles mesurables par rapport à vol^*, au sens de la Définition 2.24 appelée tribu des ensembles mesurables au sens de Lebesgue. $\mathcal{M}^*$ contient les pavés, et $\mathrm{vol}^*(P) = \mathrm{vol}(P)$ pour tout pavé P.

La mesure de Lebesgue est σ-finie en toute dimension, puisque on peut écrire $\mathbb{R}^d$ comme d'union des pavés $P_n = [-n,n]^d$, et que le volume de chacun de ces pavés est fini.

On sait par le Théorème 2.25 que $(\mathbb{R}^d, \mathcal{M}^*, \lambda^d)$ est complet. En fait,

LEMME 2.34. $\overline{(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)} = (\mathbb{R}^d, \mathcal{M}^*, \lambda^d)$.

DÉMONSTRATION. On utilise la charactérisation de la Proposition 2.17. Comme $\mathcal{M}^* \supset \mathcal{B}(\mathbb{R}^d)$, pour montrer que $\mathcal{M}^* \supset \overline{\mathcal{B}(\mathbb{R}^d)}_{\lambda^d}$, il suffit de voir que les ensembles négligeables de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$ sont tous dans $\mathcal{M}^*$. Soit donc N un ensemble négligeable. Donc il existe un Borélien B, $N \subset B$, tel que $\lambda^d(B) = 0$. Mais alors $\mathrm{vol}^*(N) \leq \mathrm{vol}^*(B) = \lambda^d(B) = 0$. Comme $\mathcal{M}^*$ est complète, $N \in \mathcal{M}^*$.

Pour montrer que $\mathcal{M}^* \subset \overline{\mathcal{B}(\mathbb{R}^d)}_{\lambda^d}$, prenons $A \in \mathcal{M}^*$. On peut supposer que A est borné : il existe $L > 0$ tel que $A \subset [-L, L]^d$. En particulier, $\mathrm{vol}^*(A) < \infty$. Pour tout $k \geq 1$, il existe un recouvrement de A par des pavés $P_n^{(k)}$, tel que $\mathrm{vol}^*(A) \geq \sum_n \mathrm{vol}(P_n^{(k)}) - 1/k$. Soit $B' = \bigcap_{k\geq 1} \bigcup_n P_n^{(k)}$. Alors bien-sûr $B' \in \mathcal{B}(\mathbb{R}^d)$, $B' \supset A$, et $\lambda^d(B') = \lambda^d(A)$. On peut faire de même pour l'ensemble $A^C := [-L, L]^d \backslash A$: il existe $\widetilde{B} \in \mathcal{B}(\mathbb{R}^d)$, $\widetilde{B} \supset A^C$, tel que $\lambda^d(\widetilde{B}) = \lambda^d(A^C)$. En posant $B := [-L, L]^d \backslash \widetilde{B}^c$, $B \subset A$ et $\lambda^d(B) = \lambda^d(A)$. On a donc bien trouvé deux Boréliens B, B' tels que $B \subset A \subset B'$, $\lambda^d(B' \backslash B) = 0$. Donc $A \in \overline{\mathcal{B}(\mathbb{R}^d)}_{\lambda^d}$. □

6.3. Invariance et Unicité. Pour tout $x \in \mathbb{R}^d$, définissons la translation $\theta_x y := y + x$. Pour un Borélien B, $\theta_x^{-1} B := \{y : \theta_x y \in B\} \in \mathcal{B}(\mathbb{R}^d)$ [5]. Si μ est une mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, on peut donc définir la mesure translatée $\theta_x \mu$ par

$$\theta_x \mu(B) := \mu(\theta_x^{-1} B) \quad \forall B \in \mathcal{B}(\mathbb{R}^d)\,.$$

THÉORÈME 2.35. *La mesure de Lebesgue est invariante sous translations :*

$$\lambda^d \circ \theta_x^{-1} = \lambda^d\,, \quad \forall x \in \mathbb{R}^d\,.$$

DÉMONSTRATION. Pour tout pavé P, on a

$$\theta_x \lambda^d(P) = \lambda^d(\theta_{-x} P) = \mathrm{vol}(\theta_{-x} P) = \mathrm{vol}(P) = \lambda^d(P)\,.$$

Mais les pavés génèrent les Boréliens et sont stables par intersections. Aussi, la suite de pavés $P_n := [-n, n]^d$ est telle que $\lambda^d(P_n) < \infty$, et

5. Clairement, $\theta_x : \mathbb{R}^d \to \mathbb{R}^d$ est continue, donc si O est un ouvert de $\mathbb{R}^d$, alors $\theta_x^{-1}(O)$ est un ouvert de $\mathbb{R}^d$. Comme les ouverts génèrent $\mathcal{B}(\mathbb{R}^d)$, le Lemme 2.9 implique que θ_x est mesurable, donc que $\theta_x^{-1} B \in \mathcal{B}(\mathbb{R}^d)$.

$P_n \nearrow \mathbb{R}^d$. Par le Théorème 2.18, on conclut que $\theta_x\lambda^d$ et λ^d coïncident sur les Boréliens. Par la Proposition 2.17, elles coïncident aussi sur les ensembles mesurables. □

Or la mesure de Lebesgue est essentiellement la seule mesure sur $\mathcal{B}(\mathbb{R}^d)$ qui soit invariante par translations :

Théorème 2.36. *À une constante multiplicative près, λ^d est l'unique mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ qui soit invariante sous translations et finie sur les Boréliens bornés.*

Démonstration. Supposons que μ est une mesure invariante sous translations et finie sur les parties bornées de $\mathbb{R}^d$. Soit $c := \mu([0,1]^d)$. Nous allons montrer que $\mu = c\lambda^d$. Considérons un pavé semi-ouvert $P = (0, b_1] \times \cdots \times (0, b_d] \equiv \prod_{i=1}^d (0, b_i]$, dont le volume vaut $\lambda^d(P) = \prod_{i=1}^d b_i$. Soit n un entier. On décompose P en une union disjointe de petits pavés, obtenus par translations de $P_0 = (0, 1/n]^d$ dont le volume est $1/n^d$. Soit $k_i(n)$ le plus grand nombre d'intervalles disjoints, de taille $1/n$, contenus dans $(0, b_i]$. Clairement, $k_i(n)/n \to b_i$, et on a

$$\prod_{i=1}^d (0, k_i(n)/n] \subset P \subset \prod_{i=1}^d (0, (k_i(n)+1)/n]\,,$$

donc

$$\mu\Big(\prod_{i=1}^d (0, k_i(n)/n]\Big) \leq \mu(P) \leq \mu\Big(\prod_{i=1}^d (0, (k_i(n)+1)/n]\Big)\,.$$

Mais $\prod_{i=1}^d (0, k_i(n)/n]$ est une union disjointe de translatés de P_0. Comme μ est invariante par translations,

$$\begin{aligned}\mu\Big(\prod_{i=1}^d (0, k_i(n)/n]\Big) &= \Big(\prod_{i=1}^d k_i(n)\Big)\mu((0, 1/n]^d)\\ &= \Big(\prod_{i=1}^d k_i(n)\Big)\frac{\mu((0,1]^d)}{n^d} = c\prod_{i=1}^d \frac{k_i(n)}{n}\,.\end{aligned}$$

Donc, lorsque $n \to \infty$,

$$\mu\Big(\prod_{i=1}^d (0, k_i(n)/n]\Big) \to c\prod_{i=1}^d b_i \equiv c\lambda^d(P)\,.$$

La borne supérieure converge vers la même limite, donc $\mu(P) = c\lambda^d(P)$. Comme précédemment, ceci implique $\mu = c\lambda^d$. □

Plus généralement, considérons les transformations linéaires inversibles de $\mathbb{R}^d$ dans $\mathbb{R}^d$, notées $GL(d,\mathbb{R})$. Pour toute transformation $T \in GL(d,\mathbb{R})$, $\det T \neq 0$, on peut montrer (voir par exemple [**7**]) que $\lambda^d(TB) = |\det T|\lambda^d(B)$ pour tout borélien $B \in \mathcal{B}(\mathbb{R}^d)$. Donc :

THÉORÈME 2.37. *Pour toute transformation $T \in GL(d,\mathbb{R})$,*

$$\lambda^d \circ T = |\det T|\lambda^d\,.$$

En particulier, la mesure de Lebesgue est **invariante sous transformations orthogonales** *(telles que $TT^t = 1$), par exemple sous les rotations.*

Ces résultats seront utilisés dans la Section 6.2 lorsque nous étudierons les changements de variables pour l'intégrale par rapport à la mesure de Lebesgue.

6.4. Régularité.

THÉORÈME 2.38. *La mesure de Lebesgue est* **régulière** *: pour tout ensemble mesurable $A \in \mathcal{M}^*$,*

$$\lambda^d(A) = \inf\{\lambda^d(G) : G \supset A,\, G \text{ ouvert}\}\,, \tag{2.19}$$

$$= \sup\{\lambda^d(K) : K \subset A,\, K \text{ compact}\}\,. \tag{2.20}$$

Ce résultat implique que l'on peut *approximer* la mesure de n'importe quel mesurable par des ouverts le contenant, ou par des compacts qu'il contient.

DÉMONSTRATION. Soit $\lambda_0(A) := \inf\{\lambda^d(G) : G \supset A,\, G \text{ ouvert}\}$. Par monotonicité, on a $\lambda^d(A) \leq \lambda^d(G)$ pour tout ouvert $G \supset A$. En particulier, $\lambda^d(A) \leq \lambda_0(A)$. Mais puisque λ^d coïncide avec la mesure extérieure vol* sur les ensembles mesurables, on a pour tout $\epsilon > 0$ un recouvrement de A par des pavés ouverts P_n tel que $\lambda^d(A) \geq \sum_n \text{vol}(P_n) - \epsilon$. En prenant $G := \bigcup_n P_n$, qui est ouvert, on a $\lambda^d(A) \geq \lambda^d(G) - \epsilon$. Ceci implique $\lambda^d(A) \geq \lambda_0(A)$.
Pour montrer (2.20), posons $\lambda_1(A) := \sup\{\lambda^d(K) : K \subset A,\, K \text{ compact}\}$. Par monotonicité, on a $\lambda^d(A) \geq \lambda^d(K)$ pour tout compact $K \subset A$. En particulier, $\lambda^d(A) \geq \lambda_1(A)$. Pour l'inégalité contraire, on peut supposer que A est contenu dans un pavé fermé $C \subset \mathbb{R}^d$ (le cas échéant, on peut toujours écrire $\lambda^d(A) = \lim_n \lambda^d(A \cap [-n,n]^d)$). On peut alors appliquer le raisonnement précédent à l'ensemble $\widetilde{A} := C\backslash A$: pour tout $\epsilon > 0$ il existe un ouvert $G \supset \widetilde{A}$ tel que $\lambda^d(\widetilde{A}) \geq \lambda^d(G) - \epsilon$. Mais alors $K := C \cap G^c$ est un compact, et

$$\lambda^d(K) \geq \lambda^d(C) - \lambda^d(G) \geq \lambda^d(C) - \lambda^d(\widetilde{A}) - \epsilon = \lambda^d(A) - \epsilon\,.$$

Ceci implique $\lambda^d(A) \leq \lambda_1(A)$. $\square$

La régularité est en fait une propriété générale des mesures finies sur $\mathbb{R}^d$.

Proposition 2.39. *Toute mesure μ finie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est régulière : pour tout $A \in \mathcal{B}(\mathbb{R}^d)$,*

$$\mu(A) = \inf\{\mu(G) : G \supset A, G \text{ ouvert}\} \tag{2.21}$$
$$= \sup\{\mu(K) : K \subset A, K \text{ compact}\}\,. \tag{2.22}$$

Démonstration. Soit $\mathcal{A}$ la famille des Boréliens A tels que pour tout $\epsilon > 0$, il existe un fermé F et un ouvert G tels que $F \subset A \subset G$ tels que $\mu(G \backslash F) \leq \epsilon$.

On montre d'abord que $\mathcal{A}$ contient les fermés. Si A est fermé, alors pour tout $n \geq 1$, considérons l'ouvert $G_n := \bigcup_{x \in A} B_{1/n}(x)$. Alors $G_n \searrow A$, et donc $\mu(G_n) \searrow \mu(A)$ puisque μ est finie. Pour tout $\epsilon > 0$, on peut donc prendre $F := A$ et n suffisamment grand, tel que $\mu(G_n \backslash F) = \mu(G_n) - \mu(F) \leq \epsilon$.

Montrons ensuite que $\mathcal{A}$ est une tribu. Clairement, $\mathbb{R}^d \in \mathcal{A}$ puisque $\mathbb{R}^d$ est à la fois ouvert et fermé. Si $A \in \mathcal{A}$ alors pour tout $\epsilon > 0$ il existe $F \subset A \subset G$ tels que $\mu(G \setminus F) \leq \epsilon$. Donc $F^c \supset A^c \supset G^c$, et comme $\mu(F^c \setminus G^c) = \mu(G \setminus F) \leq \epsilon$, ceci montre que $A^c \in \mathcal{A}$. Soit $A_n \in \mathcal{A}$, et $A := \bigcup_n A_n$. Fixons $\epsilon > 0$. Alors pour tout $n \geq 1$, il existe un fermé F_n et un ouvert G_n tels que $\mu(G_n \backslash F_n) \leq \epsilon/2^{n+1}$. Soit $G \supset A$ l'ouvert $G := \bigcup_{n \geq 1} G_n$, et la suite de fermés $F_N := \bigcup_{n=1}^N F_n \subset A$. On a $F_N \nearrow F$, et

$$\mu(G \backslash F) \leq \sum_{n \geq 1} \mu(G_n \backslash F_n) \leq \epsilon/2.$$

Mais comme $\mu(G \backslash F_N) \searrow \mu(G \backslash F)$, on peut prendre N suffisamment grande de façon à ce que $\mu(G \backslash F_N) \leq \epsilon$. Ceci montre que $A \in \mathcal{A}$.

Soit E un Borélien et $F \subset E$ un fermé tel que $\mu(F) \geq \mu(E) - \epsilon/2$. Soit C_n la suite de cubes fermés (et bornés) $C_n := [-n, n]^d$. Comme $\mu(F \cap C_n) \nearrow \mu(F)$, on peut prendre n suffisamment grand, tel que $\mu(F \cap C_n) \geq \mu(F) - \epsilon$. Comme $K := F \cap C_n$ est compact, ceci montre que $\mu(K) \geq \mu(E) - \epsilon$. Par conséquent, $\mu(E) = \sup\{\mu(K) : K \subset E, K \text{ compact}\}$. $\square$

6.5. Ensembles non-mesurables. On l'a vu, les Boréliens contiennent les pavés, les ouverts, les fermés, les intersections dénombrables, les unions dénombrables d'intersections dénombrables, etc. Donc il est naturel de se poser la question : existe-t-il des sous-ensembles de $\mathbb{R}^d$ qui ne sont pas Boréliens ?

On considère le cas de la droite pour simplifier. Nous allons présenter un argument qui montre qu'il peut exister des parties de $\mathbb{R}$ qui ne sont pas des Boréliens, et qui ne sont même pas mesurables au sens de Lebesgue. Soulignons le fait que cet argument est basé sur l'Axiome du Choix, et que sans lui, il n'est pas possible de prouver l'existence d'ensembles non-mesurables (Solovay, 1970).

Considérons sur $\mathbb{R}$ la relation d'équivalence

$$x \sim y \Leftrightarrow x - y \in \mathbb{Q}\,.$$

Considérons alors l'ensemble des classes d'équivalences $\mathsf{C} := \mathbb{R}/\sim$. Pour chaque classe $c \in \mathsf{C}$, *choisissons* [6] un représentant x_c. On peut toujours supposer que $x_c \in [0,1]$. Définissons

$$H := \{x_c : c \in \mathsf{C}\} \subset [0,1].$$

Supposons que H appartient à la classe des ensembles mesurables au sens de Lebesgue, en particulier que $\lambda^1(H) = \lambda^1(\theta_x H)$ pour toute amplitude x, et voyons pourquoi cette hypothèse mène à une contradiction. D'abord, observons que $\bigcup_{q\in\mathbb{Q}} \theta_q H = \mathbb{R}$, et que cette union est disjointe. En effet, si $x \in \theta_q H \cap \theta_{q'} H$, alors $x = x_c + q = x_{c'} + q'$, et donc $x_c - x_{c'} = q' - q \in \mathbb{Q}$, et donc x_c et $x_{c'}$ sont dans la même classe d'équivalence, ce qui implique $c = c'$, donc $x_c = x_{c'}$ et donc $q = q'$.

Ensuite, si la mesure de H était nulle, alors celle de $\mathbb{R}$ le serait aussi, puisque par σ-additivité et invariance translationnelle,

$$\lambda^1(\mathbb{R}) = \sum_{q\in\mathbb{Q}} \lambda^1(\theta_q H) = \sum_{q\in\mathbb{Q}} \lambda^1(H)\,.$$

On a donc $\lambda^1(H) > 0$. On peut ensuite observer que puisque $H \subset [0,1]$, on a $\bigcup_{q\in\mathbb{Q}\cap[0,1]} \theta_q H \subset [0,2]$. Mais comme $\lambda^1(\theta_q H) = \lambda^1(H) > 0$ pour tout q,

$$\infty = \sum_{q\in\mathbb{Q}\cap[0,1]} \lambda^1(\theta_q H) = \lambda^1\Big(\bigcup_{q\in\mathbb{Q}\cap[0,1]} \theta_q H\Big) \leq \lambda^1([0,2]) = 2\,,$$

une contradiction.

On peut maintenant répondre à une question posée dans l'introduction : comme H n'est pas mesurable, il existe une partie $E \subset [0,1]$ telle que

$$\ell^*(E) < \ell^*(E\cap H) + \ell^*(E \cap H^c)\,.$$

6. Axiome du Choix : si $(\mathcal{E}_\alpha)_{\alpha\in T}$ est une famille quelconque d'ensembles non-vides, alors il existe une fonction de choix sur $(\mathcal{E}_\alpha)_{\alpha\in T}$, c'est-à-dire une fonction qui à chaque ensemble E_α fait correspondre exactement un élément $x_\alpha \in E_\alpha$.

Puisqu'il existe des ensembles qui ne sont pas mesurables, on peut ensuite se demander s'il existe des ensembles qui sont mesurables mais qui ne sont pas des Boréliens. Il y a en fait beaucoup plus de mesurables que de Boréliens ; voir [**7**] page 38.

CHAPITRE 3

Intégrer par rapport à une mesure

Sur un espace mesuré $(X, \mathcal{F}, \mu)$ donné, on aimerait ensuite *intégrer* une fonction $f : X \to \mathbb{R}$ par rapport à la mesure μ.

On l'a vu dans l'introduction, la fonction f doit respecter une condition vis-à-vis de la structure mesurable de $(X, \mathcal{F})$: puisqu'on voudra pouvoir calculer leur mesure, les ensembles de la forme $f^{-1}(I)$ $(I \subset \mathbb{R})$ doivent appartenir à la tribu $\mathcal{F}$. L'intégration sera donc restreinte à ce type de fonctions, appelées *mesurables*.

On commencera par définir l'intégrale des fonctions dites *étagées*. Celles-ci prennent un nombre fini de valeurs et sont l'analogue (bien que de nature très différente), dans l'intégration de Riemann, d'une approximation d'une fonction continue par une fonction en escalier. Leur intégrale sera définie très naturellement en termes de la mesure μ. L'extension de l'intégrale aux fonctions mesurables sera immédiate puisque celles-ci sont approximables par des suites de fonctions étagées. L'intégrale (au sens de Lebesgue) d'une fonction mesurable f par rapport à la mesure μ sera notée

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

On décrira ensuite les théorèmes fondamentaux pour l'intégrale au sens de Lebesgue. En particulier, on donnera des conditions suffisantes pour que la suite des intégrales d'une suite de fonctions mesurables f_n convergeant vers f, converge vers l'intégrale de f :

$$\lim_{n\to\infty} \int f_n d\mu = \int \lim_{n\to\infty} f_n d\mu \,. \tag{3.1}$$

1. Mesurabilité

Comme dans le membre de droite de (3.1), on voudra intégrer des fonctions qui sont limites de suites de fonctions. Comme ces limites peuvent parfois être infinies, on admettra que f prenne les valeurs $\pm\infty$. Donc on considère des fonctions $f : X \to \mathbf{R}$, où $\mathbf{R}$ est la droite achevée (voir page 1).

La tribu sur $\mathbf{R}$ est naturellement définie par

$$\mathcal{B}(\mathbf{R}) := \sigma(\mathcal{B}(\mathbb{R}) \cup \{\pm\infty\}) .$$

Donc les sous-ensembles mesurables de $\mathbf{R}$ sont ceux de la forme A, $A \cup \{\infty\}$, $A \cup \{-\infty\}$ ou $A \cup \{\pm\infty\}$, où A est un Borélien de $\mathbb{R}$. On est donc dans un cas particulier d'application mesurable (voir Définition 2.8) :

DÉFINITION 3.1. *Une fonction $f : X \to \mathbf{R}$ est* **mesurable** *si $f^{-1}(B) \in \mathcal{F}$ pour tout $B \in \mathcal{B}(\mathbf{R})$. On notera l'ensemble des fonctions mesurables $\mathcal{M} = \mathcal{M}(X, \mathcal{F})$, et $\mathcal{M}^+ = \mathcal{M}^+(X, \mathcal{F})$ l'ensemble des fonctions mesurables non-négatives* [a].

[a] $\mathcal{M}$ est à ne pas confondre avec l'ensemble des ensembles mesurables associés à une mesure extérieure, noté $\mathcal{M}^*$.

Par exemple, les fonctions constantes ($f \equiv c$) sont mesurables puisque $f^{-1}(B) = X$ si $B \ni c$, $\varnothing$ sinon. Pour montrer qu'une fonction est mesurable, il suffit de vérifier la condition ci-dessus pour les ensembles B qui sont des demi-droites. Pour simplifier les notations, écrivons $\{f \leq a\}$ au lieu de $\{x \in X : f(x) \leq a\}$.

PROPOSITION 3.2. *Soit $f : X \to \mathbf{R}$. Les affirmations suivantes sont équivalentes.*

(1) f est mesurable.

(2) $\{f \leq a\} \in \mathcal{F}$ pour tout $a \in \mathbb{R}$

(3) $\{f < a\} \in \mathcal{F}$ pour tout $a \in \mathbb{R}$

(4) $\{f \geq a\} \in \mathcal{F}$ pour tout $a \in \mathbb{R}$

(5) $\{f > a\} \in \mathcal{F}$ pour tout $a \in \mathbb{R}$

DÉMONSTRATION. Comme $\{f < a\} = \bigcup_{n \geq 1}\{f \leq a - 1/n\}$ et $\{f \leq a\} = \bigcap_{n \geq 1}\{f < a + 1/n\}$, on a (2)⇔(3). De la même manière, (4)⇔(5). Or (2)⇔(5) puisque $\{f \leq a\} = \{f > a\}^c$ et (3)⇔(4) puisque $\{f < a\} = \{f \geq a\}^c$. Donc il reste à vérifier (1)⇔(2). Or (1)⇒(2) suit de la définition de mesurabilité (les ensembles $(-\infty, a]$ sont des Boréliens). Supposons alors que (2) est satisfaite. Observons que les ensembles $\{f = \pm\infty\}$ sont mesurables puisque, par exemple, $\{f = +\infty\} = \bigcap_{n \geq 1}\{f > n\}$. Par le Lemme 2.9, il suffit de vérifier que la tribu des Boréliens $\mathcal{B}(\mathbb{R})$ peut être générée par n'importe quelle famille d'intervalles du type $(-\infty, a]$, $(-\infty, a)$, $[a, \infty)$ ou (a, ∞). Ceci a déjà été prouvé dans le Lemme 2.33. □

La mesurabilité d'une fonction est une propriété stable, comme le montre le résultat suivant, dont la preuve est basée sur l'utilisation répétée de la Proposition 3.2.

PROPOSITION 3.3. *(1) Soit $f_n : X \to \mathbf{R}$ une suite de fonctions mesurables. Alors $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, $\liminf_n f_n$ sont mesurables. En particulier, si $f = \lim_n f_n$ existe, alors elle est mesurable.*

(2) Soient $f, g : X \to \mathbf{R}$ deux fonctions mesurables. Alors $a \cdot f$ ($a \in \mathbb{R}$), $f \cdot g$ et $f + g$, f/g (lorsqu'elles sont définies partout) sont mesurables, ainsi que $|f|$, f^2, $f^+ := \sup\{f, 0\}$, $f^- := \sup\{-f, 0\} = (-f)^+$.

DÉMONSTRATION. (1) On a $\{\sup_n f_n \leq a\} = \bigcap_n \{f_n \leq a\}$, qui appartient à $\mathcal{F}$. Comme $\inf_n f_n = -\sup_n(-f_n)$, et comme

$$\limsup_n f_n = \inf_n \sup_{m \geq n} f_m\,, \quad \liminf_n f_n = \sup_n \inf_{m \geq n} f_m\,,$$

ceci prouve les autres affirmations. (2) $a \cdot f$ est clairement mesurable. Ensuite, $\{f < g\}$ est mesurable, puisque $\{f < g\} = \bigcup_{q \in \mathbb{Q}} \{f < q\} \cap \{g > q\} \in \mathcal{F}$. Donc $\{f + g < a\} = \{f < a - g\} \in \mathcal{F}$, et $f + g$ (puis $f - g$) est mesurable. Montrons ensuite que f^2 est mesurable. Ceci suit de la remarque suivante : $\{f^2 \geq a\} = X$ si $a \leq 0$, et $\{f^2 \geq a\} = \{f \geq \sqrt{a}\} \cup \{f \leq -\sqrt{a}\}$ si $a > 0$. Ensuite, comme $f \cdot g = \frac{1}{4}((f+g)^2 - (f-g)^2)$, ceci montre que $f \cdot g$ est mesurable. En particulier, f^+ est mesurable puisque $f^+ = f \cdot 1_{\{f \geq 0\}}$. Finalement, $|f|$ est aussi mesurable puisque $\{|f| \leq a\} = \{f \leq a\} \cap \{f \geq -a\}$. □

1.1. Fonctions égales presque partout. Considérons maintenant la comparaison des fonctions, en supposant donnée une mesure μ sur $(X, \mathcal{F})$.

Deux fonctions mesurables f, g sont égales μ-presque partout (ou simplement presque partout) si $\{f \neq g\}$ est négligeable par rapport à μ.

THÉORÈME 3.4. *Soit $f : X \to \mathbf{R}$ mesurable et $g : X \to \mathbf{R}$ une fonction quelconque. Si $f = g$ presque partout, et si $(X, \mathcal{F}, \mu)$ est complet, alors g est aussi mesurable.*

DÉMONSTRATION. Par la complétude de l'espace, $\{f = g\}$ et $\{f \neq g\}$ sont mesurables. Pour tout $a \in \mathbb{R}$, on a

$$\{g \leq a\} = \big(\{g \leq a\} \cap \{f = g\}\big) \cup \big(\{g \leq a\} \cap \{f \neq g\}\big)\,.$$

Mais $\{g \leq a\} \cap \{f = g\} = \{f \leq a\} \cap \{f = g\}$, qui est mesurable. Aussi, $\{g \leq a\} \cap \{f \neq g\}$ est mesurable puisqu'il est contenu dans $\{f \neq g\}$, qui est négligeable. □

2. Intégrer les fonctions étagées

Dans le reste de la section, on suppose donné un espace mesuré $(X, \mathcal{F}, \mu)$.

DÉFINITION 3.5. *Une fonction mesurable* $\varphi : X \to \mathbf{R}$ *est* **étagée** *si elle ne prend qu'un nombre fini de valeurs. On notera l'ensemble des fonctions étagées* $\mathcal{E} = \mathcal{E}(X, \mathcal{F})$, *et* $\mathcal{E}^+ = \mathcal{E}^+(X, \mathcal{F})$ *l'ensemble des fonctions étagées non-négatives.*

Soit φ une fonction étagée dont les valeurs possibles sont $\{a_1, \dots, a_n\}$. On peut considérer les ensembles disjoints $E_k := \{x \in X : \varphi(x) = a_k\}$. Puisque φ est mesurable, on a $E_k \in \mathcal{F}$. De plus, les ensembles E_k forment toujours une partition de X : $\bigcup_k E_k = X$. On peut donc écrire φ sous la forme

$$\varphi = \sum_{k=1}^{n} a_k 1_{E_k} . \tag{3.2}$$

La définition suivante est alors naturelle :

DÉFINITION 3.6. *Soit* $\varphi \in \mathcal{E}^+(X, \mathcal{F})$ *une fonction étagée, représentée comme en* (3.2). **L'intégrale de Lebesgue de** φ **par rapport à** μ *est définie par*

$$\int \varphi d\mu := \sum_{k=1}^{n} a_k \mu(E_k) . \tag{3.3}$$

Rappelons que l'on adopte toujours la convention suivante :

$$0 \cdot \infty = \infty \cdot 0 = 0 .$$

Le fait que φ soit non-négative garantit qu'il n'apparaît aucune indétermination de la forme $\infty - \infty$ dans le terme de droite de (3.3) (même si l'intégrale peut être infinie, par exemple si l'un des a_k est $+\infty$ et si $\mu(E_k) > 0$). Par contre, comme la représentation (3.2) de φ en une combinaison de fonctions indicatrices n'est pas unique, on doit vérifier que la définition de l'intégrale ne dépend pas de la représentation choisie.

LEMME 3.7. *Si* $\sum_k a_k 1_{E_k}$ *et* $\sum_j b_j 1_{E'_j}$ *sont deux représentations de la même fonction étagée* φ, *alors*

$$\sum_k a_k \mu(E_k) = \sum_j b_j \mu(E'_j) .$$

DÉMONSTRATION. Il suffit de construire la partition plus fine $E_k \cap E'_j$. Or, puisque $a_k = b_j$ sur $E_k \cap E'_j$, on peut écrire

$$\begin{aligned}\sum_k a_k \mu(E_k) &= \sum_k a_k \mu(E_k \cap \bigcup_j E'_j) \\ &= \sum_k \sum_j a_k \mu(E_k \cap E'_j) \\ &= \sum_j \sum_k b_j \mu(E_k \cap E'_j) \\ &= \sum_j b_j \mu(E'_j \cap \bigcup_k E_k) = \sum_j b_j \mu(E'_j)\,.\end{aligned}$$

□

Donc $\int \varphi d\mu$ est bien définie, même si la représentation de φ n'est pas unique.

Si $A \in \mathcal{F}$, alors $\varphi = 1_A$ est mesurable, étagée, et

$$\int 1_A d\mu \equiv \mu(A)\,.$$

Considérons par exemple $([0,1], \mathcal{B}([0,1]), \lambda^1)$. On a vu dans l'introduction que la fonction $1_{\mathbb{Q}_{[0,1]}}$ n'est pas intégrable au sens de Riemann. Mais comme $\mathbb{Q}_{[0,1]}$ est un Borélien, $1_{\mathbb{Q}_{[0,1]}}$ est étagée et son intégrale vaut

$$\int 1_{\mathbb{Q}_{[0,1]}} d\lambda^1 = \lambda^1(\mathbb{Q}_{[0,1]}) = \sum_{q \in \mathbb{Q}_{[0,1]}} \lambda^1(\{q\}) = 0\,.$$

Donc à l'inverse de celle de Riemann, l'intégrale de Lebesgue permet d'intégrer $1_{\mathbb{Q}}$, qui est discontinue en tout $x \in \mathbb{R}$.
Observons aussi que sur $X = \mathbb{R}^d$, à la différence de l'intégrale de Riemann, la définition (3.3) fait sens même si φ n'est pas à support borné.

Donnons les propriétés élémentaires de l'intégrale pour les fonctions étagées :

LEMME 3.8. *Soient $\varphi, \psi \in \mathcal{E}^+$. Alors $\varphi + \psi \in \mathcal{E}^+$, et*

(1) Si $c \geq 0$, alors $\int c\varphi d\mu = c \int \varphi d\mu$.
(2) $\int (\varphi + \psi) d\mu = \int \varphi d\mu + \int \psi d\mu$.
(3) Si $\varphi \leq \psi$, alors $\int \varphi d\mu \leq \int \psi d\mu$.

DÉMONSTRATION. La première affirmation est triviale. Pour la seconde, supposons que $\varphi = \sum_k a_k 1_{E_k}$, $\psi = \sum_j b_j 1_{F_j}$. Soient c_l les valeurs possibles distinctes de $\varphi+\psi$, c'est-à-dire des sommes a_k+b_j. À chaque c_l on associe l'ensemble G_l, union des ensembles $E_k \cap F_j$ pour lesquels $a_k + b_j = c_l$. Alors $\varphi + \psi = \sum_l c_l 1_{G_l}$, et par définition

$$\begin{aligned}
\int (\varphi+\psi) d\mu &= \sum_l c_l \mu(G_l) \\
&= \sum_{k,j} (a_k + b_j) \mu(E_k \cap F_j) \\
&= \sum_k a_k \mu\big(E_k \cap \bigcup_j F_j\big) + \sum_j b_j \mu\big(F_j \cap \bigcup_k E_k\big) \\
&= \sum_k a_k \mu(E_k) + \sum_j b_j \mu(F_j) \\
&= \int \varphi d\mu + \int \psi d\mu \,.
\end{aligned}$$

où on a utilisé l'additivité de la mesure μ et le fait que les ensembles E_k, resp. F_j, forment une partition de X. Pour la dernière affirmation, il suffit d'utiliser des représentations pour φ et ψ utilisant la même partition $E_k \cap F_j$. □

Le lemme suivant sera le point de départ des théorèmes de convergence.

LEMME 3.9. *Soit $\varphi \in \mathcal{E}^+(X, \mathcal{F})$. Alors la fonction d'ensemble $\nu : \mathcal{F} \to [0, +\infty]$ définie par*

$$\nu(A) := \int_A \varphi d\mu \equiv \int 1_A \varphi d\mu \quad \forall A \in \mathcal{F} \tag{3.4}$$

est une mesure sur $(X, \mathcal{F})$.

DÉMONSTRATION. Clairement, $\nu(\varnothing) = 0$. Soit $\varphi = \sum_k a_k 1_{E_k}$ une représentation de φ. Si A_n est une suite disjointe d'ensembles mesurables, alors $1_{\bigcup_n A_n} = \sum_n 1_{A_n}$, et

$$\nu\Big(\bigcup_n A_n\Big) = \int 1_{\bigcup_n A_n} \varphi d\mu = \int \sum_k a_k 1_{E_k \cap \bigcup_n A_n} d\mu \,.$$

Or la fonction $\sum_k a_k 1_{E_k \cap \bigcup_n A_n}$ est étagée. Donc, par la linéarité de l'intégrale et par les propriétés fondamentales de la mesure,

$$\begin{aligned}\int \sum_k a_k 1_{E_k \cap \bigcup_n A_n} d\mu &= \sum_k a_k \mu\Big(E_k \cap \bigcup_n A_n\Big)\\ &= \sum_k a_k \sum_n \mu(E_k \cap A_n)\\ &= \sum_n \sum_k a_k \mu(E_k \cap A_n)\\ &= \sum_n \int 1_{A_n} \varphi d\mu\\ &\equiv \sum_n \nu(A_n)\,.\end{aligned}$$

L'argument ci-dessus a utilisé deux fois l'échange $\sum_k \sum_n = \sum_n \sum_k$, justifié par le fait que les termes sommés sont non-négatifs. $\square$

3. Intégrer les fonctions mesurables positives

On continue avec un espace mesuré $(X, \mathcal{F}, \mu)$. Une propriété essentielle des fonctions mesurables positives est qu'elles sont approximables par des fonctions étagées :

PROPOSITION 3.10 (Approximation par des fonctions étagées). *Soit $f \in \mathcal{M}^+$. Alors il existe une suite de fonctions étagées $\varphi_n \in \mathcal{E}^+$ telle que $\varphi_n(x) \leq \varphi_{n+1}(x) \leq f(x)$ et $\varphi_n(x) \nearrow f(x)$ pour tout $x \in X$. De plus, si f est bornée, alors $\varphi_n \nearrow f$ uniformément.*

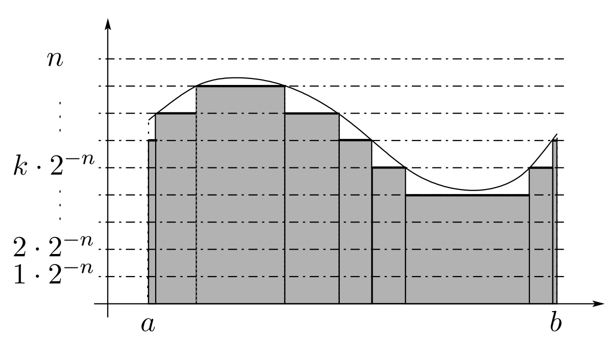


FIGURE 1. Approximation d'une fonction mesurable positive f par une fonction étagée φ, et l'intégrale de φ au sens de Lebesgue.

Démonstration. Fixons $n \geq 1$ et définissons, pour $k = 0, \dots, n2^n - 1$, $E_k := f^{-1}[k2^{-n}, (k+1)2^{-n}) \in \mathcal{F}$, et $E_{n2^n} := f^{-1}([n, +\infty])$. Posons alors $\varphi_n := \sum_{k=0}^{n2^n} k2^{-n} 1_{E_k}$. Si $x \in E_k$, $k < n2^n$, alors $f(x) - 2^n \leq \varphi_n(x) \leq f(x)$. D'autre part, si f est bornée, alors $E_{n2^n} = \varnothing$ lorsque n est suffisamment grand. □

Définition 3.11. *Soit $f \in \mathcal{M}^+(X, \mathcal{F})$. L'intégrale de Lebesgue de f par rapport à μ est définie par*

$$\int f d\mu := \sup \Big\{ \int \varphi d\mu : \varphi \in \mathcal{E}^+(X, \mathcal{F}),\, \varphi \leq f \Big\}. \tag{3.5}$$

Exemple 3.12. Considérons un ensemble dénombrable $X = \{x_n\}_{n \in \mathbb{N}}$, avec $\mathcal{F}$ la tribu discrète. Alors une mesure μ s'obtient toujours à partir d'une fonction $\rho : X \to [0, +\infty]$, $\mu(A) = \sum_{x \in A} \rho(x)$, et l'intégration d'une fonction $f : X \to \mathbb{R}$ par rapport à μ s'écrit

$$\int f d\mu = \sum_{x \in X} f(x)\rho(x)$$

Lorsque μ est une probabilité, cette expression est l'expression bien connue de espérance mathématique de f par rapport à μ.

Exemple 3.13. Si X est muni de la tribu discrète, et si δ_x est la masse de Dirac au point x, alors

$$\int f d\delta_x = f(x)\,.$$

Les deux premières affirmations du prochain résultat suivent directement de la définition. La troisième sera prouvée plus bas.

Lemme 3.14. *Soient $f, g \in \mathcal{M}^+$.*

(1) Si $c \geq 0$, alors $\int c f d\mu = c \int f d\mu$.
(2) Si $f \leq g$, alors $\int f d\mu \leq \int g d\mu$.
(3) $\int (f+g) d\mu = \int f d\mu + \int g d\mu$.

Passons maintenant au premier résultat fondamental de la théorie de l'intégration :

Théorème 3.15 (Théorème de la Convergence Monotone). *Soit $f_n \in \mathcal{M}^+$ une suite monotone non-décroissante : $f_n(x) \leq f_{n+1}(x)$ pour tout x. Soit $f(x) := \lim_{n\to\infty} f_n(x)$ (évent. infini). Alors $f \in \mathcal{M}^+$, et*

$$\lim_{n\to\infty} \uparrow \int f_n d\mu = \int f d\mu\,. \tag{3.6}$$

DÉMONSTRATION. Comme $f_n \leq f$, on a $\limsup_n \int f_n d\mu \leq \int f d\mu$. Soit ensuite $a > 1$. Si $X_n := \{x : f(x) \leq a f_n(x)\}$, alors $X_n \nearrow X$. Si $\varphi \leq f$ est une fonction étagée, on peut écrire

$$\int f_n d\mu \geq \int_{X_n} f_n d\mu \geq \frac{1}{a} \int_{X_n} f d\mu \geq \frac{1}{a} \int_{X_n} \varphi d\mu \,.$$

Si $\nu(\cdot) := \int_\cdot \varphi d\mu$, on a $\nu(X_n) \nearrow \nu(X) = \int \varphi d\mu$. Donc $\liminf_n \int f_n d\mu \geq 1/a \int \varphi d\mu$. En prenant le supremum sur $\varphi \leq f$, suivi de $a \searrow 1$, on conclut que $\liminf_n \int f_n d\mu \geq \int f d\mu$. □

On peut donner une définition plus naturelle de l'intégrale pour les fonctions mesurables non-négatives. Soit $f \in \mathcal{M}^+$, et φ_n une suite croissante de fonctions étagées convergeant vers f. Alors

$$\int f d\mu := \lim_{n\to\infty} \uparrow \int \varphi_n d\mu \,. \tag{3.7}$$

En effet, le Théorème de la Convergence Monotone garantit que le nombre défini par la limite du membre de droite ne dépend pas de la suite φ_n choisie.

Avec cette définition, la preuve du point (3) du Lemme 3.14 est élémentaire : soit φ_n (resp. ψ_n) une approximation de f (resp. g) par une suite croissante de fonctions étagées. Alors $\varphi_n + \psi_n$ est une approximation de $f + g$, et

$$\begin{aligned}\int (f+g) d\mu &= \lim_{n\to\infty} \int (\varphi_n + \psi_n) d\mu \\ &= \lim_{n\to\infty} \int \varphi_n d\mu + \lim_{n\to\infty} \int \psi_n d\mu = \int f d\mu + \int g d\mu \,.\end{aligned}$$

Une autre conséquence de la convergence monotone :

LEMME 3.16. *Si $f_n \in \mathcal{M}^+$, alors $\sum_n f_n \in \mathcal{M}^+$, et $\int \sum_n f_n d\mu = \sum_n \int f_n d\mu$.*

DÉMONSTRATION. En effet, la suite $F_N := \sum_{k\leq N} f_k$ est monotone non-décroissante, et $\sum_n f_n = \lim_N F_N$. □

Pour les suites qui ne sont pas monotones, on a un résultat plus faible, mais très utile :

LEMME 3.17 (Lemme de Fatou). *Si $f_n \in \mathcal{M}^+$, alors*

$$\int \big\{ \liminf_{n\to\infty} f_n(x) \big\} d\mu \leq \liminf_{n\to\infty} \int f_n(x) d\mu \tag{3.8}$$

DÉMONSTRATION. Soit $g_m := \inf_{n \geq m} f_n$. Alors $g_m \leq f_n$ pour tout $n \geq m$, et donc $\int g_m d\mu \leq \inf_{n \geq m} \int f_n d\mu$, ce qui donne

$$\begin{aligned}\int \big\{ \liminf_{n \to \infty} f_n(x) \big\} d\mu &\equiv \int \lim_{m \to \infty} g_m d\mu \\ &= \lim_{m \to \infty} \int g_m d\mu \\ &\leq \lim_{m \to \infty} \inf_{n \geq m} \int f_n d\mu \equiv \liminf_{n \to \infty} \int f_n d\mu \,.\end{aligned}$$

La deuxième égalité suit du Théorème de la Convergence Monotone.

$\square$

4. Propriétés vraies presque partout

On a vu que $1_{\mathbb{Q}}$ est nulle presque partout (par rapport à la mesure de Lebesgue), et que par conséquent $\int 1_{\mathbb{Q}} d\lambda^1 = 0$. Plus généralement,

THÉORÈME 3.18. *Soit $f \in \mathcal{M}^+$. Alors $f = 0$ presque partout si et seulement si $\int f d\mu = 0$.*

DÉMONSTRATION. Supposons que $\int f d\mu = 0$. Soit $B_n := \{f \geq 1/n\}$. Pour tout $n \geq 1$, on a $f \geq 1/n 1_{B_n}$, et donc $0 \leq \mu(B_n)/n \leq \int f d\mu = 0$, ce qui implique $\mu(B_n) = 0$. Par conséquent, $\mu\{f > 0\} = \mu(\bigcup_{n \geq 1} B_n) = 0$. Supposons ensuite que $f = 0$ presque partout. Soit $B := \{f > 0\}$, et $f_n := n 1_B$. On a $f \leq \lim_n f_n$ et $\int f_n d\mu = n\mu(B) = 0$. Donc, par le Lemme de Fatou,

$$0 \leq \int f d\mu \leq \liminf_{n \to \infty} \int f_n d\mu = 0 \,.$$

$\square$

COROLLAIRE 3.19. *Soient $f, g \in \mathcal{M}^+$.*

(1) Si $f \leq g$ presque partout, alors $\int f d\mu \leq \int g d\mu$.

(2) Si $f = g$ presque partout, alors $\int f d\mu = \int g d\mu$.

DÉMONSTRATION. Montrons la première affirmation (la deuxième suit immédiatement de la première). On écrit $f = f 1_{\{f \leq g\}} + f 1_{\{f > g\}}$, et comme $\mu(f > g) = 0$ par hypothèse, le théorème précédent implique $\int f 1_{\{f > g\}} d\mu = 0$ et donc

$$\int f d\mu = \int f 1_{\{f \leq g\}} d\mu \leq \int g 1_{\{f \leq g\}} d\mu \leq \int g d\mu \,.$$

$\square$

Théorème 3.20. *Si $f \in \mathcal{M}^+$, alors*

$$\int f d\mu < \infty \Rightarrow f < \infty \text{ presque partout.}$$

Démonstration. Supposons que $\epsilon := \mu(f = +\infty) > 0$. Comme $\{f \geq n\} \searrow \{f = +\infty\}$, $\mu(f \geq n) \geq \epsilon$ pour tout n. En particulier,

$$\int f d\mu \geq \int_{\{f \geq n\}} f d\mu \geq n\mu(f \geq n) \geq n\epsilon\,,$$

une contradiction. □

Théorème 3.21. *Si $f \in \mathcal{M}^+$, alors la fonction d'ensemble*

$$\nu(A) := \int_A f d\mu \quad \forall A \in \mathcal{F} \tag{3.9}$$

est une mesure sur $(X, \mathcal{F})$. De plus, ν est absolument continue par rapport à *μ : si $A \in \mathcal{F}$ est tel que $\mu(A) = 0$, alors $\nu(A) = 0$.*

Démonstration. Clairement, $\nu(\varnothing) = 0$, et la σ-additivité suit du Lemme 3.16 : si $A_n \in \mathcal{F}$ est une suite disjointe, alors

$$\nu\Big(\bigcup_n A_n\Big) = \int_{\bigcup_n A_n} f d\mu = \int \sum_n f 1_{A_n} d\mu = \sum_n \int_{A_n} f d\mu = \sum_n \nu(A_n)\,.$$

Soit A tel que $\mu(A) = 0$. Alors $f 1_A = 0$ presque partout, et le Théorème 3.18 implique $\nu(A) = \int f 1_A d\mu = 0$. □

Théorème 3.22 (Théorème de la Convergence Monotone, version presque-partout). *Soit $f_n \in \mathcal{M}^+$ une suite telle que $f_n(x) \nearrow f(x)$ pour presque tout x, où $f \in \mathcal{M}^+$. Alors*

$$\lim_{n\to\infty} \uparrow \int f_n d\mu = \int f d\mu\,. \tag{3.10}$$

Démonstration. Soit E l'ensemble de complémentaire négligeable sur lequel $f_n \nearrow f$. Si $\tilde{f}_n := f 1_E$, $\tilde{f} := f 1_E$, alors $\tilde{f}_n = f_n$ (et $\tilde{f} = f$) presque partout et $\tilde{f}_n \nearrow \tilde{f}$ partout. Donc par le Corollaire 3.19 et par convergence monotone,

$$\int f_n d\mu = \int \tilde{f}_n d\mu \nearrow \int \tilde{f} d\mu = \int f d\mu\,.$$

□

5. Intégrer les fonctions mesurables

Finalement, on étend l'intégrale aux fonctions de signe quelconque mesurables, $f \in \mathcal{M}$. Rappelons que $f^+ := \sup\{f, 0\}$, $f^- := \sup\{-f, 0\}$. On a $f^\pm \in \mathcal{M}^+$, et $f^+ - f^- = f$, $f^+ + f^- = |f|$.

DÉFINITION 3.23. *Une fonction $f \in \mathcal{M}$ est* **intégrable (par rapport à μ)** *si $|f|$ a une intégrale finie : $\int |f| d\mu < \infty$. Si f est intégrable, on pose*

$$\int f d\mu := \int f^+ d\mu - \int f^- d\mu \,. \tag{3.11}$$

L'ensemble des fonctions intégrables est noté $\mathcal{L}^1 = \mathcal{L}^1(X, \mathcal{F}, \mu)$.

Observons que si f est intégrable, alors $|f| < \infty$, et f est donc finie μ-presque partout. Comme $f^\pm \leq |f|$, (3.11) est bien définie.

Parfois, on aura besoin de spécifier la variable dont la fonction dépend (ça sera utile lorsque f dépend de plusieurs variables), et on écrira

$$\int f d\mu \equiv \int f(x) \mu(dx) \,.$$

LEMME 3.24. *Si $f, g \in \mathcal{L}^1$, $\alpha \in \mathbb{R}$, alors $f + g \in \mathcal{L}^1$ et $\alpha f \in \mathcal{L}^1$. De plus,*

$$\int (f+g) d\mu = \int f d\mu + \int g d\mu \,, \tag{3.12}$$

$$\Big| \int f d\mu \Big| \leq \int |f| d\mu \,. \tag{3.13}$$

DÉMONSTRATION. Soit $h = f + g$ (définie presque partout). Puisque $|h| \leq |f| + |g|$, on a $h \in \mathcal{L}^1$. De plus, $h^+ - h^- = f^+ - f^- + g^+ - g^-$, donc $h^+ + f^- + g^- = h^- + f^+ + g^+$, et par le point (3) du Lemme 3.14,

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

En réarrangeant les termes, on obtient $\int h d\mu = \int f d\mu + \int g d\mu$. D'autre part,

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

□

On a l'analogue du Corollaire 3.19, qui dit que l'on ne change pas la valeur d'une intégrale en modifiant la fonction sur un ensemble de mesure nulle :

LEMME 3.25. *Soient $f, g \in \mathcal{L}^1$.*

(1) Si $f \leq g$ presque partout, alors $\int f d\mu \leq \int g d\mu$.

(2) Si $f = g$ presque partout, alors $\int f d\mu = \int g d\mu$.

Énonçons maintenant le principal théorème de convergence pour l'intégrale de Lebesgue.

THÉORÈME 3.26 (Théorème de la Convergence Dominée). *Soient $f_n \in \mathcal{L}^1$ une suite telle que $|f_n(x)| \leq |g(x)|$ presque partout, où $g \in \mathcal{L}^1$. Si $f_n \to f$ presque partout, $f \in \mathcal{M}$, alors $f \in \mathcal{L}^1$, et*

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu \,. \tag{3.14}$$

DÉMONSTRATION. Pour simplifier, on suppose d'abord que $|f_n| \leq g$ et $f_n \to f$ *partout*. On a alors $|f| \leq g$, et donc $f \in \mathcal{L}^1$. De plus on a $g + f_n \geq 0$, et par le Lemme de Fatou :

$$\int (g+f)d\mu \leq \liminf_{n\to\infty} \int (g+f_n)d\mu = \int g d\mu + \liminf_{n\to\infty} \int f_n d\mu \,,$$

ce qui implique $\int f d\mu \leq \liminf_{n\to\infty} \int f_n d\mu$. De même, $g - f_n \geq 0$ et donc

$$\int (g-f)d\mu \leq \liminf_{n\to\infty} \int (g-f_n)d\mu = \int g d\mu - \limsup_{n\to\infty} \int f_n d\mu \,,$$

ce qui implique $\int f d\mu \geq \limsup_{n\to\infty} \int f_n d\mu$.
Dans le cas général, on procède comme dans la preuve du Théorème 3.22, en introduisant l'ensemble E (de complémentaire négligeable) sur lequel $f_n \to f$ et $|f_n| \leq g$. On obtient le résultat en appliquant l'argument précédent à $\tilde{f}_n := f_n 1_E$ et $\tilde{f} := f 1_E$. □

Considérons par exemple la suite $f_n(x) := x^n$ sur $([0,1], \mathcal{B}([0,1]), \lambda^1)$. Comme $|f_n| \leq 1$, qui est intégrable par rapport à λ^1, et comme $f_n(x) \to 0$ pour presque tout x (en fait, pour tous les x différents de 1), on a

$$\lim_{n\to\infty} \int f_n d\lambda^1 = 0 \,.$$

Observons qu'on a fait ce calcul sans parler de *primitive* de la fonction x^n.

5.1. À propos de l'intégration de fonctions complexes. La théorie de l'intégration développée jusqu'à présent s'applique également au cas des fonctions à valeurs complexes, $f : X \to \mathbb{C}$. Puisque chaque fonction de ce type peut s'écrire $f = \mathrm{Re} f + i \mathrm{Im} f$, où $\mathrm{Re}, \mathrm{Im} : X \to \mathbb{R}$

sont réelles, l'extension est immédiate. On dit que f est intégrable si $\mathrm{Re}f$ et $\mathrm{Im}f$ sont intégrables, et alors l'intégrale de f est définie par

$$\int f d\mu := \int \mathrm{Re} f d\mu + i \int \mathrm{Im} f d\mu \,.$$

L'ensemble des fonctions intégrables à valeurs complexes est généralement noté $\mathcal{L}^1_{\mathbb{C}}(X, \mathcal{F}, \mu)$.

6. Intégrer par rapport à la mesure de Lebesgue

Nous avons construit la mesure de Lebesgue sur $\mathbb{R}^d$, notée λ^d, dans la Section 6 du Chapitre 2. Comme il y a deux tribus naturelles sur $\mathbb{R}^d$, celle des Boréliens et celle des ensembles mesurables au sens de Lebesgue, on distinguera deux notions de mesurabilité pour une fonction $f : \mathbb{R}^d \to \mathbf{R}$. f est mesurable (au sens de Borel) si $f^{-1}(B) \in \mathcal{B}(\mathbb{R}^d)$ pour tout $B \in \mathcal{B}(\mathbf{R})$, et mesurable au sens de Lebesgue si pour tout $B \in \mathcal{B}(\mathbf{R})$, $f^{-1}(B)$ est mesurable au sens de Lebesgue [1].

On utilisera souvent le fait suivant : si $f : \mathbb{R}^d \to \mathbb{R}$ est continue, alors elle est mesurable.

L'espace des fonctions $f : \mathbb{R}^d \to \mathbf{R}$ intégrables par rapport à λ^d est $\mathcal{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$. La mesure de Lebesgue peut aussi être restreinte à des sous-ensembles de $\mathbb{R}^d$, comme par exemple aux intervalles de la droite, dans quel cas l'espace en question est $\mathcal{L}^1([a,b], \mathcal{B}([a,b]), \lambda^1)$ pour un intervalle fini, ou encore $\mathcal{L}^1(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \lambda^1)$ pour un intervalle infini.

6.1. Lien avec l'Intégrale de Riemann. Considérons une fonction $f : [a,b] \to \mathbb{R}$ et comparons les deux notions d'intégrale associées, celle de Riemann et celle de Lebesgue que nous venons d'introduire.

Rappelons comment se calcule l'intégrale de Riemann de f (voir l'introduction) : pour une partition π de $[a,b]$, $x_0 = a < x_1 < \cdots < x_{n-1} < x_n \equiv b$, on considère deux approximations de f par des fonctions en escaliers :

$$\varphi_\pi^- = \sum_{i=0}^{n-1} m_i 1_{[x_i, x_{i+1})} \,, \quad \varphi_\pi^+ = \sum_{i=0}^{n-1} M_i 1_{[x_i, x_{i+1})} \,. \tag{3.15}$$

1. Observons qu'il existe des fonctions mesurables au sens de Lebesgue mais pas au sens de Borel [**7**].

On a $\varphi_\pi^- \leq f \leq \varphi_\pi^+$. En définissant les sommes de Riemann [2]

$$I(\varphi_\pi^-) := \sum_{i=0}^{n-1} m_i \ell([x_i, x_{i+1}))\,, \quad I(\varphi_\pi^+) := \sum_{i=0}^{n-1} M_i \ell([x_i, x_{i+1}))\,,$$

on dit que f est intégrable au sens de Riemann si

$$-\infty < \sup_\pi I(\varphi_\pi^-) = \inf_\pi I(\varphi_\pi^+) < \infty\,.$$

Lorsque ce nombre existe, on le note

$$\int_a^b f(x)dx\,.$$

D'autre part, la mesure de Lebesgue intègre f en l'approximant par des fonctions *étagées*. On notera son intégrale temporairement par

$$\int_{[a,b]} f d\lambda^1\,.$$

THÉORÈME 3.27. *Si $f : [a,b] \to \mathbb{R}$ est bornée et intégrable au sens de Riemann, alors elle est mesurable au sens de Lebesgue, et on a*

$$\int_a^b f(x)dx = \int_{[a,b]} f d\lambda^1\,. \tag{3.16}$$

Ceci permet de calculer certaines intégrales en utilisant le Théorème Fondamental. Par exemple,

$$\int_{[0,1]} x^2 \lambda^1(dx) = \int_0^1 x^2 dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}\,.$$

On sait que le contraire n'est pas vrai : $1_{\mathbb{Q}_{[0,1]}}$ est intégrable au sens de Lebesgue mais pas au sens de Riemann.

DÉMONSTRATION. Puisque f est bornée, on peut toujours supposer qu'elle est positive. Par la définition de l'intégrale de Riemann, il existe une suite de fonctions en escalier $\varphi_n^- : [a,b] \to \mathbb{R}$, $\varphi_n^- \leq f$, telle que $I(\varphi_n^-) \nearrow \int_a^b f(x)dx$, ainsi qu'une suite $\varphi_n^+ : [a,b] \to \mathbb{R}$, $\varphi_n^+ \geq f$, telle que $I(\varphi_n^+) \searrow \int_a^b f(x)dx$. Observons que comme les fonctions en escaliers sont aussi étagées, on a $\int_{[a,b]} \varphi_n^\pm d\lambda^1 = I(\varphi_n^\pm)$.

On peut toujours supposer que φ_n^- (resp. φ_n^+) est monotone croissante (resp. décroissante) [3]. Dans ce cas, les limites $f^- := \lim_n \uparrow \varphi_n^- \leq f$ et

2. On voit ici que si f n'est pas bornée, alors au moins une de ces sommes est infinie.

3. Par exemple, si φ_n^+ n'est pas décroissante, on peut redéfinir $g_n^+ := \min\{\varphi_1^+, \dots, \varphi_n^+\}$.

$f^+ := \lim_n \downarrow \varphi_n^+ \geq f$ existent, et sont mesurables. Par convergence dominée,

$$\int_{[a,b]} f^- d\lambda^1 = \lim_{n\to\infty} \uparrow \int_{[a,b]} \varphi_n^- d\lambda^1 = \lim_{n\to\infty} I(\varphi_n^-) = \int_a^b f(x)dx\,,$$
$$\int_{[a,b]} f^+ d\lambda^1 = \lim_{n\to\infty} \downarrow \int_{[a,b]} \varphi_n^+ d\lambda^1 = \lim_{n\to\infty} I(\varphi_n^+) = \int_a^b f(x)dx\,.$$

On a donc $\int_{[a,b]}(f^+ - f^-)d\lambda^1 = 0$, Comme $f^+ - f^- \geq 0$, ceci implique que $f^+ = f^- = f$ presque partout (Théorème 3.18). Comme $([a,b], \mathcal{M}^*, \lambda^1)$ est complet, f est mesurable (par le Théorème 3.4). Évidemment, (3.16) est vérifiée. □

On sait qu'il existe des fonctions réelles qui sont intégrables au sens de Lebesgue et pas au sens de Riemann, comme l'indicatrice $1_{\mathbb{Q}}$. Mais au-delà de simplement pouvoir intégrer plus de fonctions, l'intégrale de Lebesgue transforme l'intégration en un outil de travail bien plus souple que celle de Riemann. Ce sont en particulier ses théorèmes de convergence (Convegence Monotone, Lemme de Fatou, Convergence Dominée), présentés plus haut, qui la rendent adaptée à de nombreux problèmes d'analyse, où l'approximation d'une fonction intervient de façon naturelle.

À moindre coût, on peut aussi donner la démonstration du résultat suivant, qui donne une caractérisation complète de l'intégrabilité au sens de Riemann. Il fut prouvé pour la première fois par Lebesgue en 1904 :

THÉORÈME 3.28. *$f : [a,b] \to \mathbb{R}$ est intégrable au sens de Riemann si et seulement si elle est continue presque partout, c'est à dire si*

$$\lambda^1(\{x \in [a,b] : f \text{ est discontinue en } x\}) = 0\,.$$

DÉMONSTRATION. Soit

$$g^+(x) := \limsup_{y\to x} f(y) \equiv \lim_{\epsilon\to 0^+} \sup_{y\in B_\epsilon(x)} f(y)\,,$$
$$g^-(x) := \liminf_{y\to x} f(y) \equiv \lim_{\epsilon\to 0^+} \inf_{y\in B_\epsilon(x)} f(y)\,,$$

et soient $\varphi_n^\pm$, $f^\pm$, les fonctions introduites dans la preuve du Théorème 3.27. On peut sans autre supposer que pour tout n, la partition $\pi_n^\pm$ associée à $\varphi_n^\pm$ est obtenue en raffinant $\pi_{n-1}^\pm$, c'est à dire en divisant certains de ses intervalles en sous-intervalles. Soit ensuite $\partial\pi_n^\pm$ l'ensemble des extrémités des intervalles de la partition $\pi_n^{\pm}$, et $\partial^{\pm} := \bigcup_n \partial\pi_n^{\pm}$. Si $x \in [a,b]\backslash\partial^{\pm}$, alors $g^{\pm}(x) = f^{\pm}(x)$. Par le Théorème 3.4, les fonctions $g^{\pm}$ sont donc mesurables. On peut alors écrire

$$\int_{[a,b]} g^{\pm} d\lambda^1 = \int_{[a,b]} f^{\pm} d\lambda^1 = \int_a^b f(x)dx\,.$$

Donc $\int_{[a,b]}(g^+ - g^-)d\lambda^1 = 0$, et comme $g^+ \geq g^-$, on en conclut que $g^+ = g^-$ presque partout. Or les points où g^+ et g^- coïncident sont justement ceux où f est continue. □

Les premiers succès remportés par l'intégrale de Lebesgue furent dans la théorie des séries trigonométriques. Plus tard, elle se développa rapidement et trouva de nombreuses applications en analyse. En particulier, Riesz y contribua de façon significative, en remarquant que ses méthodes n'étaient pas restreintes à la droite et pouvaient s'appliquer sur des espaces plus abstraits.

À propos du niveau d'abstraction requis pour développer sa théorie, Lebesgue commente, dans la préface de son livre **[11]** :

> "[...] On peut se demander, il est vrai, s'il y a quelque intérêt à s'occuper de telles complications et s'il ne vaut pas mieux se borner à l'étude des fonctions qui ne nécessitent que des définitions simples. Cela n'a guère que des avantages quand il s'agit d'un Cours élémentaire ; mais, comme on le verra dans ces Leçons, si l'on voulait toujours se limiter à la considération de ces bonnes fonctions, il faudrait renoncer à résoudre bien des problèmes à énoncés simples posés depuis longtemps. C'est pour la résolution de ces problèmes, et non par amour des complications, que j'ai introduit dans ce livre une définition de l'intégrale plus générale que celle de Riemann et comprenant celle-ci comme cas particulier."

Par le Théorème 3.27, les méthodes développées pour l'intégration de Riemann (intégration par parties, changement de variable, etc.) peuvent être utilisées pour calculer des intégrales de fonctions par rapport à la mesure de Lebesgue. Ceci est aussi vrai pour les intégrales impropres.

THÉORÈME 3.29. *Soit $f : \mathbb{R} \to \mathbb{R}_+$ telle que sur tout intervalle symmétrique $[-n, n]$, f est bornéee et intégrable au sens de Riemann. Alors f est intégrable par rapport à la mesure de Lebesgue (sur tout $\mathbb{R}$) si et seulement si la suite d'intégrales de Riemann $\int_{-n}^{n} f(x)dx$ converge dans la limite $n \to \infty$, et alors*

$$\int f d\lambda^1 = \lim_{n\to\infty} \int_{-n}^{n} f(x)dx\,. \tag{3.17}$$

DÉMONSTRATION. Par le Théorème 3.27, puisqu'elle est intégrable au sens de Riemann, $f_n := f 1_{[-n,n]}$ est mesurable au sens de Lebesgue et $\int f_n d\lambda^1 = \int_{-n}^{n} f(x)dx$. De plus, $f_n \nearrow f$, donc f est mesurable. Par convergence monotone,

$$\int f(x)\lambda^1(dx) = \lim_{n\to\infty} \int f_n d\lambda^1 = \lim_{n\to\infty} \int_{-n}^{n} f(x)dx\,. \tag{3.18}$$

□

Par exemple, considérons $f(x) = \frac{1}{1+x^2}$ pour $x \in \mathbb{R}$. Puisqu'elle est continue, elle est intégrable au sens de Riemann sur tout intervalle borné. Or

$$\lim_{n\to\infty} \int_{-n}^{n} \frac{1}{1+x^2} dx = \lim_{n\to\infty} 2\arctan(n) = \pi < \infty\,,$$

ceci montre que f est intégrable par rapport à la mesure de Lebesgue sur $\mathbb{R}$ et que son intégrale vaut π.

Il faut souligner l'importance de la condition de positivité de f dans l'énoncé ci-dessus. En effet, considérons la fonction $\frac{\sin x}{x}$, à laquelle on donne la valeur 1 au point $x = 0$. Alors il est facile de voir (exercice) que l'intégrale impropre au sens de Riemann

$$\int_{-\infty}^{\infty} \frac{\sin x}{x} dx = \lim_{n\to\infty} \int_{-n}^{n} \frac{\sin x}{x} dx$$

existe, mais que f n'est pas intégrable au sens de Lebesgue, puisque

$$\begin{aligned}
\int_{\mathbb{R}} \Big|\frac{\sin x}{x}\Big| \lambda^1(dx) &= 2\sum_{k=0}^{\infty} \int_{[k\pi,(k+1)\pi]} \Big|\frac{\sin x}{x}\Big| \lambda^1(dx) \\
&\geq 2\sum_{k=0}^{\infty} \frac{1}{(k+1)\pi} \int_{[k\pi,(k+1)\pi]} |\sin x| \lambda^1(dx) \\
&= \frac{2}{\pi}\Big(\int_0^{\pi} |\sin x| dx\Big) \sum_{k\geq 1} \frac{1}{k} = \infty
\end{aligned}$$

On a un résultat similaire pour les intégrales impropres sur un intervalle ouvert (la preuve est laissée en exercice) :

THÉORÈME 3.30. *Soit $f : (a, b] \to \mathbb{R}_+$ telle que l'intégrale de Riemann impropre*

$$\int_{a^+}^{b} f(x)dx := \lim_{\epsilon \to 0^+} \int_{a+\epsilon}^{b} f(x)dx$$

existe (i.e. soit finie). Alors f est mesurable au sens de Lebesgue, et

$$\int_{(a,b]} f d\lambda^1 = \int_{a^+}^{b} f(x)dx\,.$$

Par exemple, puisque

$$\int_{0^+}^{1} \frac{1}{\sqrt{x}}dx = \lim_{\epsilon \to 0^+} \int_{\epsilon}^{1} \frac{1}{\sqrt{x}}dx = \lim_{\epsilon \to 0^+} 2(1 - \sqrt{\epsilon}) = 2\,,$$

on a

$$\int_{(0,1]} \frac{1}{\sqrt{x}}\lambda^1(dx) = 2\,.$$

Il est important de souligner que les trois théorèmes présentés dans cette section ne mentionnent jamais l'existence d'une *primitive* pour la fonction intégrée. Dans le Chapitre 8, on fera le lien avec le Théorème Fondamental de l'Analyse : on s'intéressera à la possibilité de définir une primitive d'une fonction intégrable f par l'expression classique

$$F(x) := \int_{[a,x]} f d\lambda^1\,.$$

6.2. La formule du changement de Variable. Dans la pratique, certaines intégrales (par rapport à la mesure de Lebesgue) sont plus simples à étudier lorsqu'elles peuvent être exprimées en fonction de variables particulières. Typiquement, on voudra transformer une intégrale d'une fonction $f = f(x)$, $x \in D$, où D est un ouvert de $\mathbb{R}^d$, en une intégrale d'une nouvelle fonction dont la variable $u \in U$. On aura besoin de supposer que le changement de variable est suffisament lisse.

DÉFINITION 3.31. *Soient U, D deux ouverts de $\mathbb{R}^d$.*

(1) *Une application $\varphi : U \to D$ bijective, de classe C^1 et telle que φ^{-1} est aussi de classe C^1, est appelée* difféomorphisme de classe C^1.

(2) *Si φ est un difféomorphisme de classe C^1, son* Jacobien *est*

$$J_\varphi(u) := \det \nabla\varphi(u)\,,$$

où $\nabla\varphi$ est la matrice $d \times d$ des dérivées partielles $(\nabla\varphi)_{ij} := \frac{\partial \varphi_i}{\partial u_j}$.

Observons qu'étant dérivables, et donc continues, les fonctions $u \mapsto (\nabla\varphi)_{ij}(u)$ sont mesurables.

Théorème 3.32. *Si $\varphi : U \to D$ est un difféomorphisme de classe C^1, alors pour toute fonction $f \in \mathcal{L}^1(D, \mathcal{B}(D), \lambda^d)$,*

$$\int_D f(x)\lambda^d(dx) = \int_U f(\varphi(u))|J_\varphi(u)|\lambda^d(du)\,. \tag{3.19}$$

Démonstration. On commence par montrer le résultat pour les fonctions indicatrices. Soit donc $B \subset D$ un borélien. La formule (3.19) que l'on veut montrer devient, pour $f = 1_B$,

$$\lambda^d(B) = \int_{\varphi^{-1}(B)} |J_\varphi(u)|\lambda^d(du)\,. \tag{3.20}$$

On peut interpréter cette expression en disant que le transport de la mesure $|J_\varphi|\lambda^d$ par φ est exactement la mesure de Lebesgue. Puisque φ est inversible, on peut de façon équivalente prendre un borélien $A \subset U$ et montrer que

$$\lambda^d(\varphi(A)) = \int_A |J_\varphi(u)|\lambda^d(du)\,. \tag{3.21}$$

Montrons d'abord une version locale du résultat, en étudiant comment la mesure d'un petit cube de U change sous le difféomorphisme. Le résultat suivant montre qu'il existe une constante de proportionnalité entre la mesure du cube et celle de son image, donnée par la valeur du Jacobien au centre du cube :

Lemme 3.33. *Soit $K \subset U$ un compact. Pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que pour tout pavé cubique P_0 centré en un point $u_0 \in K$ et dont les côtés sont de taille inférieure à δ,*

$$(1-\epsilon)|J_\varphi(u_0)|\lambda^d(P_0) \leq \lambda^d(\varphi(P_0)) \leq (1+\epsilon)|J_\varphi(u_0)|\lambda^d(P_0) \tag{3.22}$$

Démonstration. Pour tout $u \in U$ suffisamment proche de u_0, on peut écrire

$$\varphi(u) = \varphi(u_0) + \{\nabla\varphi(u_0) + R(u_0, u)\} \cdot (u - u_0)\,,$$

où [4] $\|R(u_0, u)\|_\infty \to 0$ lorsque $\|u - u_0\|_\infty \to 0$. On peut donc écrire

$$\begin{aligned}\varphi(u) - \varphi(u_0) &= T_{u_0}(u - u_0) + R(u_0, u)(u - u_0)\\ &= T_{u_0}\big(u - u_0 + {T_{u_0}}^{-1}R(u_0, u)(u - u_0)\big)\,,\end{aligned}$$

4. Ici, on utilise la norme $\|x\|_\infty := \max_{i=1,\dots,d}|x_i|$. Si A est une matrice, $\|A\|_\infty := \max_{1\leq i,j\leq d}|A_{ij}|$.

où $T_{u_0} \in GL(d, \mathbb{R})$ est l'application linéaire inversible $v \mapsto \nabla\varphi(u_0)v$. Soit $\tilde{\epsilon} > 0$ tel que $(1+\tilde{\epsilon})^d \equiv 1+\epsilon$, et soit $\delta > 0$ tel que

$$\sup_{u_0 \in K} \sup_{u:\|u-u_0\|_\infty \leq \delta} \|T_{u_0}{}^{-1} R(u_0, u)\|_\infty \leq \tilde{\epsilon}\,.$$

Un tel δ existe par continuité uniforme des fonctions $\nabla\varphi$ et $\nabla\varphi^{-1}$ sur le compact K.

Soit P_0 un pavé cubique centré en u_0, dont les côtés ont taille $l \leq \delta$. Par ce qui précède,

$$\varphi(P_0) \subset \varphi(u_0) + T_{u_0} P_0'\,,$$

où P_0' est le pavé centré à l'origine donc les côtés ont taille $l' = (1+\tilde{\epsilon})l$. On a donc, par l'invariance de la mesure de Lebesgue sous translations (Théorème 2.35) et par le Théorème 2.37,

$$\begin{aligned}
\lambda^d(\varphi(P_0)) &\leq \lambda^d(\varphi(u_0) + T_{u_0} P_0') \\
&= \lambda^d(T_{u_0} P_0') \\
&= |\det T_{u_0}| \lambda^d(P_0') \\
&= |\det T_{u_0}| (1+\tilde{\epsilon})^d \lambda^d(P_0) \\
&\equiv (1+\epsilon) |J_\varphi(u_0)| \lambda^d(P_0)\,.
\end{aligned}$$

On obtient la borne inférieure en procédant de la même manière, en écrivant $\varphi(P_0) \supset \varphi(u_0) + T_{u_0} P_0''$, où P_0'' est un cube de côté $l' = (1-\tilde{\epsilon})l$, où $(1-\tilde{\epsilon})^d \equiv 1-\epsilon$. □

Soit $P \subset U$ un pavé fermé (donc compact), et $\epsilon > 0$. Soit $\delta > 0$ pris comme dans le lemme. Comme $u \mapsto |J_\varphi(u)|$ est continue, elle est uniformément continue sur P. On peut donc supposer en plus que $\|u-u'\|_\infty \leq \delta$ implique $1-\epsilon \leq |J_\varphi(u)|/|J_\varphi(u')| \leq 1+\epsilon$. Soit P_j, $j = 1, \dots, m$ une famille disjointe de pavés de côtés de taille $\leq \delta$, dont les centres $u_j \in P$, et telle que $\bigcup_{j=1}^m P_j = P$. En appliquant le lemme à chaque P_j,

$$\begin{aligned}
\lambda^d(\varphi(P)) = \sum_{j=1}^m \lambda^d(\varphi(P_j)) &\leq (1+\epsilon) \sum_{j=1}^m |J_\varphi(u_j)| \lambda^d(P_j) \\
&\leq (1+\epsilon)^2 \sum_{j=1}^m \int_{P_j} |J_\varphi(u)| \lambda^d(du) \\
&= (1+\epsilon)^2 \int_{P_0} |J_\varphi(u)| \lambda^d(du)\,.
\end{aligned}$$

En procédant de même pour la borne inférieure, ceci montre (3.21) pour les pavés dont la fermeture est contenue dans U. Pour montrer que (3.21) est vrai pour tout borélien, il suffit de voir que

$$\mathcal{D} := \{A \subset U : (3.21) \text{ est vérifiée}\}$$

est une classe de Dynkin. Puisqu'elle contient les pavés et que ceux-ci sont stables sous intersections, le Théorème 2.21 montre que $\mathcal{D}$ coïncide avec la famille de tous les boréliens contenus dans U, ce qui termine la preuve de (3.21) et (3.20).

Pour une fonction $f : D \to \mathbb{R}$ intégrable, on procède de façon standard, en approximant f par une suite de fonctions étagées et en utilisant de façon répétée le Théorème de la Convergence Dominée. $\square$

Par exemple, on peut calculer l'intégrale

$$\int_0^2 x \sin(x^2) \lambda^1(dx)$$

en introduisant le changement de variable $\varphi : (0,4) \to (0,2)$ donné par $\varphi(u) := \sqrt{u}$. Dans ce cas, $J_\varphi(u) = 1/2\sqrt{u}$, et donc par la formule du changement de variable,

$$\int_0^2 x \sin(x^2) \lambda^1(dx) = \frac{1}{2} \int_0^4 \sin(u) \lambda^d(du) = \frac{1 - \cos 4}{2}\,.$$

Il est très utile de pouvoir transformer une intégrale en une autre intégrale, mais il nous manque encore un outil essentiel pour pouvoir intégrer explicitement des fonctions de plusieurs variables. On verra dans le Chapitre 6 comment profiter de la formule de changement de variable donnée dans le Théorème 3.32, et du *Théorème de Fubini*, pour calculer des intégrales dans le plan ou dans l'espace, en utilisant des changements de variables appropriés. Décrivons certains changements de variables souvent utilisés dans $\mathbb{R}^2$ et $\mathbb{R}^3$.

6.2.1. *Coordonnées polaires sur* $\mathbb{R}^2$. Soit $U_0 := (0,\infty) \times (0,2\pi)$. Les points de U_0 sont souvent notés $(r,\theta) \in U_0$. Soit alors $U \subset U_0$ un ouvert, et $\varphi : U_0 \to \mathbb{R}^2$ le difféomorphisme défini par

$$\varphi(r,\theta) := (r\cos\theta, r\sin\theta)\,, \tag{3.23}$$

dont le Jacobien est donné par :

$$J_\varphi(r,\theta) = \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix} = r$$

Soit ensuite $D = \varphi(U)$. La formule du changement de variable devient, pour f intégrable,

$$\int_D f(x,y)\lambda^2(dx,dy) = \int_U f(\varphi(r,\theta))r\lambda^2(dr,d\theta)$$

6.2.2. *Coordonnées cylindriques sur $\mathbb{R}^3$.*

6.2.3. *Coordonnées sphériques sur $\mathbb{R}^3$.*

7. Intégrales dépendant d'un paramètre

Une application utile des théorèmes généraux de convergence présentés plus haut est l'étude d'intégrales dépendant d'un paramètre.

THÉORÈME 3.34. *Soit $(X,\mathcal{F},\mu)$ un espace mesuré et $f : I \times X \to \mathbb{R}$, où I est un intervalle*[a] *de $\mathbb{R}$ (pas forcément borné). Supposons que pour tout $t \in I$, $f(t,\cdot) \in \mathcal{L}^1(X,\mathcal{F},\mu)$. Alors $\varphi(t) := \int f(t,x)\mu(dx)$ est bien définie sur I. Si, de plus,*

(1) pour μ-presque tout $x \in X$, $f(\cdot,x)$ est continue sur I,

(2) il existe une fonction $g \in \mathcal{L}^1(X,\mathcal{F},\mu)$ telle que $|f(t,x)| \leq g(x)$ pour tout $(t,x) \in I \times X$,

alors $\varphi(t)$ est continue sur I.

[a] Ici, I peut en fait être remplacé par n'importe quel espace métrique.

DÉMONSTRATION. Soit $t_* \in I$, t_n une suite de I convergeant vers t_*, et $f_n(x) := f(t_n,x)$. Alors pour tout n on a $f_n \in \mathcal{L}^1(X,\mathcal{F},\mu)$, $|f_n| \leq g$, et $\lim_n f_n(x) = f(t_*,x)$ pour μ-presque tout x. Donc, par convergence dominée,

$$\lim_{n\to\infty} \varphi(t_n) = \lim_{n\to\infty} \int f_n(x)\mu(dx) = \int f(t_*,x)\mu(dx) = \varphi(t_*)\,,$$

et donc φ est continue en t_*. $\square$

EXEMPLE 3.35. Si $f \in \mathcal{L}^1(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda^1)$, alors le résultat précédent montre que la transformée de Fourier de f, définie par

$$\widehat{f}(k) := \int e^{ikx} f(x)\lambda(dx)\,,$$

est une fonction continue.

EXEMPLE 3.36. Soit $f \in \mathcal{L}^1(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda^1)$. Comme $t \mapsto 1_{(-\infty,x]}(t)f(x)$ est continue partout sauf en $t = x$, elle est continue λ^1-presque partout. Donc la fonction

$$F(x) := \int_{(-\infty,x]} f d\lambda^1$$

est continue partout.

THÉORÈME 3.37. *Soit $(X, \mathcal{F}, \mu)$ un espace mesuré et $f : I \times X \to \mathbb{R}$, où I est un intervalle de $\mathbb{R}$ (pas forcément borné). Supposons que :*

(1) Pour tout $t \in I$, $f(t, \cdot) \in \mathcal{L}^1(X, \mathcal{F}, \mu)$.

(2) Pour μ-presque tout $x \in X$, $f(\cdot, x)$ est dérivable en $t_* \in I$.

(3) Il existe une fonction $g \in \mathcal{L}^1(X, \mathcal{F}, \mu)$ telle que $|\frac{\partial f}{\partial t}(t, x)| \leq g(x)$ pour tout $(t, x) \in I \times X$.

Alors $\varphi(t) := \int f(t, x)\mu(dx)$ est bien définie et dérivable en t_*. De plus,

$$\varphi'(t_*) := \int \frac{\partial f}{\partial t}(t_*, x)\mu(dx)\,.$$

DÉMONSTRATION. Soit $t_n \neq t_*$ une suite de I convergeant vers t_*. Soit

$$D_n(x) := \frac{f(t_n, x) - f(t_*, x)}{t_n - t_*}\,.$$

Alors $D_n \in \mathcal{L}^1(X, \mathcal{F}, \mu)$, et $\lim_n D_n(x) = \frac{\partial f}{\partial t}(t_*, x)$ pour μ-presque tout x. Par le Théorème de la Valeur Moyenne, il existe $\tilde{t}$ (dépendant de t_*, t_n, x) tel que $D_n(x) = \frac{\partial f}{\partial t}(\tilde{t}, x)$, et donc $|D_n| \leq g$. Par convergence dominée,

$$\begin{aligned}\lim_{n\to\infty} \frac{\varphi(t_n) - \varphi(t_*)}{t_n - t_*} &= \lim_{n\to\infty} \int D_n(x)\mu(dx)\\ &= \int \lim_{n\to\infty} D_n(x)\mu(dx) = \int \frac{\partial f}{\partial t}(t_*, x)\mu(dx)\,,\end{aligned}$$

ce qui prouve l'affirmation. □

8. Parenthèse : intégration et probabilités

En théorie des probabilités, l'*intégration d'une fonction* s'interprète comme le calcul de l'*espérance d'une variable aléatoire*.

Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité, modélisant une expérience aléatoire donnée, et $X : \Omega \to \mathbb{R}$ une variable aléatoire, c'est-à-dire une fonction mesurable (telle que $X^{-1}(B) \in \mathcal{F}$ pour tout $B \in \mathcal{B}(\mathbb{R})$). En termes probabilistes, $X(\omega)$ est un nombre donnant une certaine information sur la réalisation ω de l'expérience.

Souvent, on s'intéresse plus aux valeurs prises par la variable aléatoire, $X(\omega) \in \mathbb{R}$, que par la réalisation ω. Il est donc intéressant de transporter l'aléa de ω (contenu dans P) sur la droite réelle, en définissant la loi de la variable X, qui est la mesure de probabilité μ_X sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$,

$$\mu_X := P \circ X^{-1}\,.$$

On voit que μ_X n'est autre que le *transport de P par X* (voir exercices). On dit par exemple que X suit une loi normale si

$$\mu_X(B) = \mu_{\mathcal{N}(0,1)}(B) := \int_B \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \lambda^1(dx)\,.$$

pour tout Borélien B.

Lorsque, comme dans l'exemple précédent, la loi de X peut s'écrire comme

$$\mu_X(B) = \int_B \rho(x) \lambda^1(dx)\,,$$

où $\rho : \mathbb{R} \to \mathbb{R}_+$ est mesurable, $\int \rho d\lambda^1 = 1$, on dit que X est une variable aléatoire de densité ρ. Dans ce cas, $\mu_X \ll \lambda^1$: μ_X est *absolument continue* par rapport à la mesure de Lebesgue (Théorème 3.21). On verra au Chapitre 7 que toute variable aléatoire dont la loi est absolument continue par rapport à la mesure de Lebesgue possède une densité.

On a aussi vu dans le Théorème 2.32 qu'une mesure de probabilité sur la droite, ici μ_X, est entièrement déterminée par sa fonction de répartition :

$$F_X(x) := \mu_X((-\infty, x])\,.$$

F_X est non-décroissante, continue à droite, et possède les limites

$$\lim_{x\to-\infty} F_X(x) = 0\,, \quad \lim_{x\to\infty} F_X(x) = 1\,.$$

Si X est intégrable, c'est-à-dire si $\int |X| dP < \infty$, on définit son espérance mathématique par

$$E[X] := \int X dP\,.$$

On peut alors montrer (exercice) que si X est intégrable, son espérance peut être calculée à l'aide de sa loi, par

$$E[X] = \int x \mu_X(dx)\,.$$

Dans le cas particulier où X possède une densité ρ (Exercice 22),

$$E[X] = \int x \rho(x) \lambda^1(dx)\,.$$

Par exemple, si X est discrète, c'est-à-dire si X est *étagée*, prenant ses valeurs dans un ensemble fini, $X(\omega) \in \{x_1, x_2, \dots, x_k\}$, alors sa loi est entièrement déterminée par la connaissance des nombres $p_k := \mu_X(\{x_k\}) = P(\{\omega : X(\omega) = x_k\})$, et l'espérance est donnée par (voir exercice)

$$E[X] = \sum_{j=1}^{k} x_j p_j \,,$$

une expression bien connue du cours de probabilités.

CHAPITRE 4

Les espaces $\mathcal{L}^p$ et L^p

Nous avons défini les fonctions intégrables $\mathcal{L}^1(X, \mathcal{F}, \mu)$ comme étant l'ensembles des fonctions mesurables $f : X \to \mathbf{R}$ telles que $\int |f| d\mu < \infty$. Dans le présent chapitre nous fixerons un nombre $1 \leq p < \infty$ et distinguerons l'ensemble des fonctions $f : X \to \mathbf{R}$ telles que $\int |f|^p d\mu < \infty$. Cet ensemble a l'avantage de pouvoir être muni d'une structure d'espace vectoriel et d'une topologie, ce qui en fait un des espaces de fonctions les plus utilisés de l'analyse fonctionnelle.

1. Les espaces $\mathcal{L}^p$

Soit $(X, \mathcal{F}, \mu)$ un espace mesuré, que nous fixerons jusqu'à la fin du chapitre. Si $1 \leq p < \infty$, définissons

$$\mathcal{L}^p = \mathcal{L}^p(X, \mathcal{F}, \mu) := \left\{ f \in \mathcal{M}(X, \mathcal{F}) : \int |f|^p d\mu < \infty \right\}.$$

Le cas $p = 1$ correspond bien-sûr à l'espace des fonctions intégrables, défini au chapitre précédent. Par exemple, sur $[1, \infty)$ avec la mesure de Lebesgue, $e^{-x} \in \mathcal{L}^p$ pour tout $p \geq 1$, alors que $\frac{1}{|\log x|}$ n'est dans aucun $\mathcal{L}^p$. De manière générale, sur un espace fini, toutes les fonctions mesurables bornées sont dans $\mathcal{L}^p$, puisque

$$\int |f|^p d\mu \leq \sup_{x \in X} |f(x)|^p \cdot \mu(X).$$

Exemple 4.1. Considérons $X = \mathbb{N}$ avec la mesure de comptage. Comme une fonction $f : \mathbb{N} \to \mathbb{R}$ peut être identifiée avec une suite $a_n := f(n)$, $\mathcal{L}^p$ peut être identifié avec l'ensemble ℓ^p des suites $(a_n)_{n \geq 1}$ telles que $\sum_n |a_n|^p < \infty$.

Commençons par remarquer que si $f, g \in \mathcal{L}^p$, alors f et g sont finies presque partout, $f + g$ est donc bien définie presque partout, et

$$\begin{aligned} |f+g|^p &\leq (|f| + |g|)^p \leq (2 \sup\{|f|, |g|\})^p \\ &= 2^p (\sup\{|f|^p, |g|^p\}) \leq 2^p (|f|^p + |g|^p) \end{aligned}$$

Donc si $f, g \in \mathcal{L}^p$, alors $f + g \in \mathcal{L}^p$. Il est clair que si $f \in \mathcal{L}^p$, alors $\alpha f \in \mathcal{L}^p$ pour tout réel α.

Si $f \in \mathcal{L}^p$, on définit

$$\|f\|_p := \Big(\int |f|^p d\mu \Big)^{1/p} . \tag{4.1}$$

Nous allons voir que $\|\cdot\|_p$ possède plusieurs propriétés qui rappellent celle d'une norme. Pourtant, on peut déjà remarquer que si $\|f\|_p = 0$, alors f n'est pas nécessairement identiquement nulle, mais seulement presque partout. On pourra faire de $\|\cdot\|_p$ une *norme*, en modifiant un peu la notion de fonction, dans la Section 2.

Mais avant ça, étudions deux inégalités fondamentales satisfaites par $\|\cdot\|_p$. La première est l'*Inégalité de Hölder*.

LEMME 4.2. *Soient $f \in \mathcal{L}^p$, $g \in \mathcal{L}^q$, où $1 < p < \infty$ et où q est l'***exposant conjugué de** *p, défini comme l'unique $q \geq 1$ satisfaisant*

$$\frac{1}{p} + \frac{1}{q} = 1 .$$

*(On écrira parfois $q \leftrightarrow p$.) Alors on a l'***Inégalité de Hölder** *:*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q . \tag{4.2}$$

En particulier, si $f \in \mathcal{L}^p$ et $g \in \mathcal{L}^q$, où p et q sont conjugués, alors $fg \in \mathcal{L}^1$.

DÉMONSTRATION. Si $\|f\|_p$ (resp. $\|g\|_q$) est nul, alors l'énoncé est trivial. En effet, $\|f\|_p = 0$ implique $f = 0$ presque partout (Corollaire 3.18), et donc le membre de gauche est nul aussi. Supposons donc que $\|f\|_p$ et $\|g\|_q$ sont différents de zéro. Il est facile de vérifier que pour toute paire de nombres positifs[1] a, b,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} .$$

Avec $a = |f(x)|/\|f\|_p$, $b = |g(x)|/\|g\|_q$, on obtient

$$\frac{|f(x)g(x)|}{\|f\|_p \|g\|_q} \leq \frac{|f(x)|^p}{p\|f\|_p^p} + \frac{|g(x)|^q}{q\|g\|_q^q} ,$$

On obtient (4.2) en intégrant par rapport à μ :

$$\frac{\|fg\|_1}{\|f\|_p \|g\|_q} \leq \frac{1}{p} + \frac{1}{q} = 1 .$$

□

1. En effet, à $b > 0$ fixé, la fonction $t \mapsto t^p/p - tb + b^q/q$ est non-négative.

Le seul nombre $p > 1$ qui soit conjugué à lui-même est $p = 2$, donc le cas particulier de l'inégalité de Hölder où $p = q = 2$ correspond à l'Inégalité de Cauchy-Schwartz :

$$\|fg\|_1 \leq \|f\|_2 \|g\|_2 \,.$$

En général, il n'y a pas de relations d'inclusion entre les espaces $\mathcal{L}^p$. Par exemple, sur la droite, $e^{-x^2} \in \mathcal{L}^1$ et $\mathcal{L}^2$, alors que

$$\frac{e^{-x^2}}{\sqrt{|x|}} \in \mathcal{L}^1, \text{ mais } \notin \mathcal{L}^2 \,,$$

et

$$\frac{1}{\sqrt{1+x^2}} \in \mathcal{L}^2, \text{ mais } \notin \mathcal{L}^1.$$

Par contre la comparaison peut se faire sur des espaces finis :

LEMME 4.3. *Si $\mu(X) < \infty$ (par exemple si $(X, \mathcal{F}, \mu)$ est un espace de probabilité), alors $\mathcal{L}^\beta \subset \mathcal{L}^\alpha$ dès que $\alpha < \beta$.*

DÉMONSTRATION. On applique l'inégalité de Hölder avec les exposants conjugués $\alpha' = \beta/\alpha$, $\beta' = \beta/(\beta - \alpha)$:

$$\begin{aligned}
\int |f|^\alpha d\mu &= \int |f|^\alpha \cdot 1 d\mu \\
&\leq \Big(\int (|f|^\alpha)^{\beta/\alpha} d\mu \Big)^{\alpha/\beta} \Big(\int 1^{\beta/(\beta-\alpha)} d\mu \Big)^{(\beta-\alpha)/\beta} \\
&= \mu(X)^{(\beta-\alpha)/\beta} \Big(\int |f|^\beta d\mu \Big)^{\alpha/\beta} \,.
\end{aligned}$$

Donc

$$\|f\|_\alpha \leq \mu(X)^{\frac{1}{\alpha} - \frac{1}{\beta}} \|f\|_\beta \,, \tag{4.3}$$

ce qui prouve l'affirmation. □

Insistons sur le fait que le résultat ci-dessus n'est pas vrai en général si l'espace n'est pas fini. Par exemple, sur la droite, on a $\mathcal{L}^1 \not\subset \mathcal{L}^p$ $(\forall p > 1)$, que l'on comprend facilement en considérant $f(x) = \frac{1}{|x|} 1_{\{|x|>1\}}$.

Passons à la deuxième inégalité fondamentale :

LEMME 4.4. *Soit $1 \leq p < \infty$. Si $f, g \in \mathcal{L}^p$, on a l'Inégalité de Minkowski :*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p \,. \tag{4.4}$$

DÉMONSTRATION. Le cas $p = 1$ est trivial. Or puisque $|f+g|^{(p-1)q} = |f+g|^p$, on a aussi $|f+g|^{p-1} \in \mathcal{L}^q$ lorsque p et q sont conjugués. Donc en appliquant deux fois l'inégalité de Hölder,

$$\begin{aligned}\int |f+g|^p d\mu &= \int |f+g| \cdot |f+g|^{p-1} d\mu \\ &\leq \int |f| \cdot |f+g|^{p-1} d\mu + \int |g| \cdot |f+g|^{p-1} d\mu \\ &\leq \|f\|_p \|f+g\|_p^{p/q} + \|g\|_p \|f+g\|_p^{p/q} ,\end{aligned}$$

ce qui donne (4.4). □

$\|\cdot\|_p$ permet de définir une notion de convergence sur $\mathcal{L}^p$:

DÉFINITION 4.5. *Une suite $f_n \in \mathcal{L}^p$ converge au sens de $\mathcal{L}^p$ vers $f \in \mathcal{L}^p$ (et on notera $f_n \overset{\mathcal{L}^p}{\to} f$) si $\|f_n - f\|_p \to 0$ lorsque $n \to \infty$.*

On dit aussi, en anticipant un peu, que f_n converge vers f dans la norme p si $f_n \overset{\mathcal{L}^p}{\to} f$.

La convergence $\overset{\mathcal{L}^p}{\to}$ permet de définir une topologie sur $\mathcal{L}^p$, c'est à dire une notion de proximité entre les points de $\mathcal{L}^p$. Par exemple, un sous ensemble $\mathcal{C} \subset \mathcal{L}_1$ est fermé si toute suite convergente dans $\mathcal{C}$ converge vers un élément de $\mathcal{C}$: si $f_n \in \mathcal{C}$, et si $f_n \overset{\mathcal{L}^p}{\to} f$, alors $f \in \mathcal{C}$. Aussi, un ensemble $\mathcal{D} \subset \mathcal{L}^p$ est dense si pour tout $f \in \mathcal{L}^p$ et pour tout $\epsilon > 0$ il existe $g \in \mathcal{D}$ tel que $\|f - g\|_p \leq \epsilon$. Un ensemble dense $\mathcal{D}$ permet d'approximer $f \in \mathcal{L}^p$ par une suite $g_n \in \mathcal{D}$ telle que $\|g_n - f\|_p \to 0$. Par exemple, on a vu en exercice que $\mathcal{E}$ est dense dans $\mathcal{L}^1$. Nous reviendrons plus en détails sur l'approximation des fonctions par des ensembles denses dans le chapitre suivant.

2. Les espaces L^p

Nous l'avons dit dans la section précédente, la fonction identiquement nulle n'est pas la seule à satisfaire $\|f\|_p = 0$, et donc $\|\cdot\|_p$ n'est pas une *norme* dans le sens habituel du terme, que nous rappelons ici :

DÉFINITION 4.6. *Sur un espace vectoriel réel V, une application $v \mapsto \|v\| \in \mathbb{R}$ est une norme si*

(1) $\|v\| \geq 0$ pour tout $v \in V$, avec égalité si et seulement si $v = 0$,

(2) $\|\alpha v\| = |\alpha| \|v\|$ pour tout $\alpha \in \mathbb{R}$, $v \in V$,

(3) pour tout $v, w \in V$, $\|v+w\| \leq \|v\| + \|w\|$ (inégalité triangulaire).

La paire $(V, \|\cdot\|)$ est un espace normé.

Pour l'instant, l' application $f \mapsto \|f\|_p$ définie en (4.1) sur les fonctions $f \in \mathcal{L}^p$ satisfait toutes les conditions qui définissent une norme, sauf que $\|f\|_p = 0$ n'implique pas $f = 0$, mais seulement $f = 0$ presque partout. Autrement dit, $\|f - g\|_p = 0$ si et seulement si $f = g$ presque partout. Donc, du point de vue de l'intégrale, il est inutile de distinguer ces fonctions, et rien ne nous empêche d'*identifier* les fonctions qui sont égales presque partout.

Définissons donc la relation suivante sur $\mathcal{L}^p$:

$$f \sim g \Leftrightarrow f = g \text{ presque partout}. \tag{4.5}$$

Il est facile de voir que $\sim$ définit une relation d'équivalence[2] sur $\mathcal{L}^p$. Par exemple, $\|f\|_p = 0$ implique que $f \sim 0$ (la fonction identiquement nulle). La classe d'équivalence de f,

$$[f] := \{g \in \mathcal{L}^p : g \sim f\},$$

représente donc l'ensemble des fonctions qui lui sont égales presque partout. Sur la droite, on a par exemple $1_{\mathbb{Q}} \in [0]$ (car $[0]$ contient toutes les fonctions qui sont égales à zéro presque partout), $e^{-x^2} 1_{\mathbb{Q}^c} \in [e^{-x^2}]$.

La relation d'équivalence induit donc une partition de $\mathcal{L}^p$ en classes d'équivalences, et il ne reste qu'à identifier ces classes comme les points d'un nouvel ensemble. Formellement, ceci revient à considérer l'espace quotient

$$L^p(X, \mathcal{F}, \mu) := \mathcal{L}^p(X, \mathcal{F}, \mu)/ \sim .$$

On munit facilement $L^p = L^p(X, \mathcal{F}, \mu)$ d'une structure d'espace vectoriel, en définissant $[f] + [g] := [f + g]$ (bien définie puisque f et g sont finies presque partout), et $\alpha[f] := [\alpha f]$. Pour une classe d'équivalence, on définit

$$\|[f]\|_p := \|f\|_p .$$

On a alors la structure recherchée :

THÉORÈME 4.7. *$(L^p, \|\cdot\|_p)$ est un espace vectoriel normé, appelé espace de Lebesgue.*

CONVENTION 4.1. Dans la suite, pour alléger les notations, on omettra la plupart du temps de spécifier que l'on travaille avec des classes d'équivalences, et on écrira simplement f au lieu de $[f]$, étant sous-entendu que f est un représentant quelconque de sa classe d'équivalence.

2. Une relation d'équivalence $\sim$ sur un ensemble E est une relation binaire 1) *réflexive* ($x \sim x$ pour tout $x \in E$), 2) *symétrique* ($x \sim y$ implique $y \sim x$), et 3) *transitive* (si $x \sim y$ et $y \sim z$, alors $x \sim z$).

EXEMPLE 4.8. Considérons $X = \mathbb{N}$ avec la mesure de comptage. Comme vu dans l'exemple 4.1, L^p peut être identifié avec les suites $(a_n)_{n\geq 1}$ telles que $\sum_n |a_n|^p < \infty$. Remarquons que dans ce cas il n'existe pas d'ensembles non-vides de mesure nulle, et donc L^p coïncide avec $\mathcal{L}^p$.

Avec une structure d'espace normé, on peut introduire une notion de *convergence* [3] :

DÉFINITION 4.9. *Soit $f_n \in L^p$.*

(1) *f_n* **converge dans la norme** *L^p si il existe $f \in L^p$ telle que $\|f_n - f\|_p \to 0$ lorsque $n \to \infty$. Notation : $f_n \xrightarrow{L^p} f$.*
(2) *Une suite $f_n \in L^p$ est une* **suite de Cauchy** *si $\|f_n - f_m\|_p \to 0$ lorsque $m, n \to \infty$.*

Observons que toute suite de Cauchy est **bornée en norme** : $\sup_n \|f_n\|_p < \infty$ [4], et que toute suite convergente est une suite de Cauchy. Une propriété fondamentale des espaces L^p, comme pour les nombres réels, est que le contraire est aussi vrai :

THÉORÈME 4.10. *Pour tout $1 \leq p < \infty$, $(L^p, \|\cdot\|_p)$ est* **complet** *: toute suite de Cauchy converge.*

DÉMONSTRATION. Soit $f_n \in L^p$ une suite de Cauchy : $\|f_n - f_m\|_p \to 0$ lorsque $m, n \to \infty$. On peut alors extraire une suite d'entiers $n_1 < n_2 < \dots$ telle que $\|f_{n_{k+1}} - f_{n_k}\|_p \leq 2^{-k}$. On étudie la série $\sum_{k\geq 1}(f_{n_{k+1}} - f_{n_k})$. Calculons

$$\begin{aligned}
\int \Big(\sum_{k\geq 1} |f_{n_{k+1}} - f_{n_k}|\Big)^p d\mu &= \lim_{K\to\infty} \int \Big(\sum_{k=1}^{K} |f_{n_{k+1}} - f_{n_k}|\Big)^p d\mu \\
&= \lim_{K\to\infty} \Big\| \sum_{k=1}^{K} |f_{n_{k+1}} - f_{n_k}| \Big\|_p^p \\
&\leq \lim_{K\to\infty} \Big(\sum_{k=1}^{K} \|f_{n_{k+1}} - f_{n_k}\|_p\Big)^p \\
&= \Big(\sum_{k\geq 1} \|f_{n_{k+1}} - f_{n_k}\|_p\Big)^p \leq \Big(\sum_{k\geq 1} 2^{-k}\Big)^p < \infty
\end{aligned}$$

3. On pourrait aussi introduire la métrique $d_{L^p}(f, g) := \|f - g\|_p$.
4. En effet, soit f_n une suite de Cauchy dans $\mathcal{L}^p$. Alors il existe n_0 tel que $\|f_n - f_m\| \leq 1$ pour tout $n, m \geq n_0$. Pour tout $n \geq n_0$, on peut écrire $\|f_n\|_p \leq \|f_{n_0}\|_p + \|f_n - f_{n_0}\|_p \leq \|f_{n_0}\|_p + 1$.

(La première égalité suit de la convergence monotone, l'inégalité suit de l'inégalité de Minkowski.) En particulier, $\sum_{k\geq 1}|f_{n_{k+1}}-f_{n_k}|$, et a fortiori $\sum_{k\geq 1}(f_{n_{k+1}}-f_{n_k})$, convergent presque sûrement. Puisque $f_{n_k}=f_{n_1}+\sum_{j=1}^{k-1}(f_{n_{j+1}}-f_{n_j})$, on a donc existence, presque partout, de la limite $\lim_{k\to\infty} f_{n_k}$. On peut alors poser

$$f := \begin{cases} \lim_{k\to\infty} f_{n_k} & \text{si la limite existe}\,, \\ 0 & \text{sinon,} \end{cases}$$

qui est mesurable. Le Lemme de Fatou permet de montrer que $f \in L^p$:

$$\int |f|^p d\mu = \int \lim_{k\to\infty} |f_{n_k}|^p d\mu \leq \liminf_{k\to\infty} \int |f_{n_k}|^p d\mu \leq \sup_{n\geq 1} \int |f_n|^p d\mu < \infty\,.$$

En effet, puisque c'est une suite de Cauchy, f_n est bornée en norme. Finalement, pour montrer que $f_n \stackrel{L^p}{\to} f$, on utilise à nouveau le Lemme de Fatou :

$$\begin{aligned} \|f-f_{n_j}\|_p{}^p &= \int \lim_{k\to\infty} |f_{n_k}-f_{n_j}|^p d\mu \\ &\leq \liminf_{k\to\infty} \int |f_{n_k}-f_{n_j}|^p d\mu \\ &= \liminf_{k\to\infty} \|f_{n_k}-f_{n_j}\|_p{}^p \end{aligned}$$

Mais on peut utiliser $k-j-1$ fois l'inégalité de Minkowski, et obtenir

$$\|f_{n_k}-f_{n_j}\|_p \leq \sum_{i=j}^{k-1} \|f_{n_{i+1}}-f_{n_i}\|_p \leq \sum_{i=j}^{k-1} 2^{-i} < \sum_{i\geq j} 2^{-i} = 2^{-j+1}\,.$$

Donc $\|f-f_{n_j}\|_p \leq 2^{-j+1}$, ce qui montre que f_{n_j} converge vers f en norme L^p. Puisque $\|f_n-f\|_p \leq \|f_n-f_{n_j}\|_p + \|f_{n_j}-f\|_p$, ceci conclut la preuve. □

Un espace normé qui est complet pour sa norme est appelé espace de Banach. La possibilité de construire des espaces fonctionnels L^p qui soient des espaces de Banach est un atout important de l'intégrale au sens de Lebesgue. Ici, à nouveau, l'intégrale de Riemann ne permet *pas* d'obtenir la complétude : on montre (en exercice) que $C[a,b]$, l'espace des fonctions continues sur $[a,b]$ muni de la norme

$$\|f\|_R := \int_a^b |f(x)|dx$$

est bien un espace normé, mais qu'il n'est pas complet.

3. Remarques sur le cas $p = 2$

Le cas $p = 2$ correspond à l'espace des fonctions de carré sommable : $\int |f|^2 d\mu < \infty$. La particularité de cet espace est qu'il permet de définir un produit scalaire : pour toute paire $f, g \in L^2$,

$$\langle f, g \rangle := \int f g d\mu \,.$$

La norme peut alors s'exprimer par

$$\|f\|_2 = \sqrt{\langle f, f \rangle}\,. \tag{4.6}$$

On vérifie facilement que $\langle \cdot, \cdot \rangle$ a bien les propriétés d'un produit scalaire, à savoir :

(1) symétrique : $\langle f, g \rangle = \langle g, f \rangle$.

(2) linéaire : $\langle \alpha f, g \rangle = \alpha \langle f, g \rangle$ pour tout $\alpha \in \mathbb{R}$.

(3) défini positif : $\langle f, f \rangle \geq 0$, avec égalité si et seulement si $f = 0$.

L'inégalité de Cauchy-Schwarz s'écrit

$$|\langle f, g \rangle| \leq \|f\|_2 \|g\|_2 \,.$$

L'utilisation d'un produit scalaire permet d'introduire des notions géométriques sur L^2. Par exemple, on dit que f et g sont orthogonales si $\langle f, g \rangle = 0$. Remarquons que lorsque f et g sont orthogonales, alors

$$\|f + g\|_2^2 = \langle f + g, f + g \rangle = \langle f, f \rangle + \langle g, g \rangle = \|f\|_2^2 + \|g\|_2^2 \,,$$

identité mieux connue sous le nom Théorème de Pythagore. Le produit scalaire permet aussi de définir la *projection sur un sous-espace* (exercice).

Un espace vectoriel muni d'un produit scalaire $\langle \cdot, \cdot \rangle$, qui est complet par rapport à la norme (4.6) est appelé espace de Hilbert.

L'espace de Hilbert L^2 est très utilisé en analyse fonctionnelle, et de grande importance en physique mathématique [5].

5. En mécanique quantique, $L^2_{\mathbb{C}}(\mathbb{R}^3, \mathcal{B}(\mathbb{R}^3), \lambda^3)$ est l'espace de Hilbert qui décrit une particule quantique (sans spin) dans $\mathbb{R}^3$. Un élément $\psi \in L^2_{\mathbb{C}}(\mathbb{R}^3, \mathcal{B}(\mathbb{R}^3), \lambda^3)$ est appelé *fonction d'onde*.

4. Remarques sur le cas $p = \infty$

Bien que la norme $\|\cdot\|_p$ ne fasse sens que pour $1 \leq p < \infty$, le cas $p = \infty$ peut être défini naturellement : si $f \in \mathcal{M}(X, \mathcal{F})$,

$$\begin{aligned}\|f\|_\infty &:= \inf\{M > 0 : \mu(|f| \geq M) = 0\}\\ &= \inf\{M > 0 : |f| \leq M \text{ pour } \mu\text{-presque tout } x\}\end{aligned}$$

avec la convention habituelle : $\inf \varnothing = +\infty$. $\|f\|_\infty$ est parfois appelé supremum essentiel de f, et noté $\operatorname{ess\,sup}_{x \in X}|f(x)|$. On introduit alors

$$\mathcal{L}^\infty = \mathcal{L}^\infty(X, \mathcal{F}, \mu) := \left\{f \in \mathcal{M}(X, \mathcal{F}) : \|f\|_\infty < \infty\right\},$$

et $L^\infty = L^\infty(X, \mathcal{F}, \mu)$ comme précédemment, en identifiant comme en (4.5) les fonctions égales presque partout. En prenant $q \to \infty$ dans l'identité des exposants conjugués $\frac{1}{p} + \frac{1}{q} = 1$, on aimerait affirmer que "$p = 1$ est conjugué à $q = \infty$", et donner l'équivalent de l'inégalité de Hölder :

Lemme 4.11. *Si $f, g \in \mathcal{M}$,*

$$\int |fg| d\mu \leq \|f\|_\infty \|g\|_1 . \tag{4.7}$$

Démonstration. Si $\|f\|_\infty = \infty$, alors il n'y a rien à montrer. Si $M > \|f\|_\infty$, alors $|f(x)| \leq M$ presque partout, et donc

$$\int |fg| d\mu = \int |fg| 1_{|f| \leq M} d\mu \leq M \int |g| d\mu = M \|g\|_1 .$$

On obtient le résultat en prenant l'infimum sur $M > \|f\|_\infty$. □

Théorème 4.12.

(1) $\|\cdot\|_\infty$ est une norme sur L^∞

(2) $\|f_n - f\|_\infty \to 0$ si et seulement si il existe un ensemble $E \in \mathcal{F}$, $\mu(E^c) = 0$, tel que $f_n \to f$ uniformément sur E.

(3) L^∞ est un espace vectoriel complet pour la norme $\|\cdot\|_\infty$.

(4) Les fonctions simples bornées sont denses dans L^∞ (pour sa norme).

La preuve est laissée en exercice.

CHAPITRE 5

Suites de fonctions et approximations

Lorsqu'on étudie des suites de fonctions sur un ensemble, $f_n : X \to \mathbb{R}$, il y a plusieurs sens dans lesquels f_n peut *converger* vers une fonction f lorsque $n \to \infty$. Chaque notion de convergence donne en particulier un moyen d'*approximer* une fonction par d'autres fonctions, en général plus simples.

Dans tout le chapitre, lorsque l'on parle de fonctions sur les réels, il sera sous-entendu que la mesure en question est la mesure de Lebesgue.

1. Convergence ponctuelle et uniforme

Commençons par les deux notions de convergence les plus simples, qui n'ont pas recours à la mesurabilité :

DÉFINITION 5.1. *f_n **converge ponctuellement vers** f si $f_n(x) \to f(x)$ pour tout $x \in X$.*

DÉFINITION 5.2. *f_n **converge uniformément** vers f si $\sup_{x\in X} |f_n(x) - f(x)| \to 0$. Notation : $f_n \xrightarrow{u} f$.*

Il est clair que la convergence uniforme implique la convergence ponctuelle, et que le contraire n'est pas vrai. En effet, la suite $f_n : [0,1] \to \mathbb{R}$, $f_n(x) := x^n$ converge vers $1_{\{1\}}$ ponctuellement mais pas uniformément puisque $\sup_{x\in[0,1]} |f_n(x) - 1_{\{1\}}(x)| = 1$ pour tout n. De même, $f_n : \mathbb{R} \to \mathbb{R}$, $f_n(x) := 1_{[n,n+1]}$, converge vers $f(x) := 0$ ponctuellement mais pas uniformément.

2. Convergence presque-partout et en mesure

Les notions suivantes caractérisent des modes de convergence se référant à une mesure μ.

DÉFINITION 5.3. *f_n converge μ-**presque-partout**[a] vers f si $N = \{x \in X : f_n(x) \not\to f(x)\}$ est négligeable par rapport à μ. Notation : $f_n \xrightarrow{p.p.} f$.*

a. En probabilités, on parle plutôt de convergence *presque-sûre*.

Bien-sûr, toute suite qui converge ponctuellement converge aussi presque-partout. Observons toutefois que pour la convergence presque-partout, la limite n'est plus unique. En prenant par exemple $f_n := 1_{\mathbb{Q}_n}$ ($\mathbb{Q}_n = \{q_1, \dots, q_n\}$ est l'ensemble des n premiers rationnels de la droite, selon un ordre arbitraire choisi) alors f_n converge λ^1-presque partout vers $1_{\mathbb{Q}}$, mais aussi vers la fonction identiquement égale à zéro.

DÉFINITION 5.4. *Soient $f_n, f \in \mathcal{M}(X, \mathcal{F})$. f_n **converge en mesure** vers f si pour tout $\epsilon > 0$, $\mu(|f_n - f| \geq \epsilon) \to 0$. Notation : $f_n \xrightarrow{\mu} f$.*

En peut vérifier (exercice) que :

LEMME 5.5.

(1) Si $f_n \xrightarrow{\mu} f$ et $f_n \xrightarrow{\mu} g$, alors $f = g$ presque partout.

(2) Si $f_n \xrightarrow{\mu} f$, $g_n \xrightarrow{\mu} g$, alors pour tout $a, b \in \mathbb{R}$, $af_n + bg_n \xrightarrow{\mu} af + bg$, et $f_n g_n \xrightarrow{\mu} fg$.

Comparons convergence en mesure avec convergence presque partout. Un exemple classique de suite qui converge en mesure mais pas presque partout est le suivant : sur $[0,1]$, considérer la suite $f_1 := 1$, $f_2 := 1_{[0,1/2]}$, $f_3 := 1_{(1/2,1]}$, $f_4 := 1_{[0,1/4)}$, $f_5 := 1_{[1/4,1/2)}$, etc. D'autre part, on peut considérer la suite $f_n := 1_{[n,n+1]}$, qui converge presque partout mais pas en mesure. Le lemme suivant éclaircit ces deux situations :

LEMME 5.6.

(1) Si $\mu(X) < \infty$, alors $f_n \xrightarrow{p.p.} f$ implique $f_n \xrightarrow{\mu} f$.

(2) Si $f_n \xrightarrow{\mu} f$, alors il existe une sous-suite n_j telle que $f_{n_j} \xrightarrow{p.p.} f$.

La première affirmation n'est pas vraie en général sans l'hypothèse de finitude, comme le montre $f_n(x) = 1_{[n,n+1](x)}$ (sur la droite munie de la mesure de Lebesgue).

PREUVE DU LEMME 5.6 : On a

$$\{f_n \to f\} = \bigcap_{k \geq 1} \bigcup_{n \geq 1} \bigcap_{m \geq n} \{x : |f_m(x) - f(x)| \leq 1/k\},$$

donc

$$\{f_n \not\to f\} = \bigcup_{k \geq 1} \bigcap_{n \geq 1} \bigcup_{m \geq n} \{x : |f_m(x) - f(x)| > 1/k\}.$$

Supposons que $f_n \xrightarrow{p.p.} f$, c'est-à-dire que $\mu(f_n \not\to f) = 0$. En particulier,

$$\mu\Big(\bigcap_{n \geq 1} \bigcup_{m \geq n} \{x : |f_m(x) - f(x)| > 1/k\}\Big) = 0, \quad \forall k \geq 1.$$

Mais pour tout $\epsilon > 0$, il existe un $k \geq 1$ tel que $1/k \leq \epsilon$. Donc, par la propriété de continuité de la mesure (Proposition 2.15), lorsque $n \to \infty$,

$$\mu(|f_n - f| \geq \epsilon) \leq \mu\Big(\bigcup_{m\geq n}\{x : |f_m(x) - f(x)| \geq 1/k\}\Big)$$
$$\searrow \mu\Big(\bigcap_{n\geq 1}\bigcup_{m\geq n}\{x : |f_m(x) - f(x)| \geq 1/k\}\Big) = 0$$

Donc $f_n \xrightarrow{\mu} f$.

Inversément, supposons que $f_n \xrightarrow{\mu} f$. Choisissons une suite $\epsilon_j \searrow 0$ et une suite $\eta_j \searrow 0$ sommable. On peut donc extraire une sous-suite n_j croissante telle que

$$\mu(|f_{n_j} - f| \geq \epsilon_j) \leq \eta_j\,, \quad \forall j\,.$$

Soit alors

$$N := \bigcap_{n\geq 1}\bigcup_{j\geq n}\{x : |f_{n_j}(x) - f(x)| \geq \epsilon_j\}\,.$$

Par continuité de la mesure et comme $\mu(X) < \infty$, on a

$$\begin{aligned}\mu(N) &= \lim_{n\to\infty} \mu\Big(\bigcup_{j\geq n}\{x : |f_{n_j}(x) - f(x)| \geq \epsilon_j\}\Big)\\ &\leq \lim_{n\to\infty} \sum_{j\geq n} \mu(|f_{n_j} - f| \geq \epsilon_j)\\ &\leq \lim_{n\to\infty} \sum_{j\geq n} \eta_j = 0\,.\end{aligned}$$

Mais comme

$$N^c = \bigcup_{n\geq 1}\bigcap_{j\geq n}\{x : |f_{n_j}(x) - f(x)| < \epsilon_j\}\,,$$

alors $f_{n_j}(x) \to f(x)$ pour tout $x \in N^c$. □

Sur L^1, et lorsque l'espace est fini, alors la convergence en mesure est métrisable (la preuve est laissée en exercice) :

Lemme 5.7. *Pour $f, g \in L^1$, on définit*

$$d(f,g) := \int \frac{|f-g|}{1+|f-g|}\,d\mu\,.$$

Si la mesure est finie, $\mu(X) < \infty$, alors pour $f_n, f \in L^1$, $f_n \xrightarrow{\mu} f$ si et seulement si $d(f_n, f) \to 0$.

3. Convergence en norme L^p

On a déjà rencontré la convergence au sens de la norme $\mathcal{L}^p$:

DÉFINITION 5.8. *Soient $f_n, f \in \mathcal{M}(X, \mathcal{F})$. f_n converge vers f en norme L^p si $\|f_n - f\|_p \to 0$. Notation : $f_n \xrightarrow{L^p} f$.*

Si $\mu(X) < \infty$, alors $f_n \xrightarrow{u} f$ implique $f_n \xrightarrow{L^p} f$, puisque

$$\int |f_n - f|^p d\mu \leq \mu(X)(\sup_{x \in X} |f_n(x) - f(x)|)^p .$$

THÉORÈME 5.9. *$f_n \xrightarrow{L^p} f$ implique $f_n \xrightarrow{\mu} f$. En particulier, si $f_n \xrightarrow{L^p} f$, alors il existe une sous-suite n_j telle que $f_{n_j} \xrightarrow{p.p.} f$.*

DÉMONSTRATION.

$$\begin{aligned} \epsilon^p \mu(|f_n - f| \geq \epsilon) &= \int_{\{|f_n - f| \geq \epsilon\}} \epsilon^p d\mu \\ &\leq \int_{\{|f_n - f| \geq \epsilon\}} |f_n - f|^p d\mu \leq \|f_n - f\|_p{}^p . \end{aligned}$$

□

Le contraire n'est pas vrai : sur $[0,1]$ avec la mesure de Lebesgue, prendre $f_n = n^{1/p} 1_{[0,1/n]}$; f_n converge vers $f \equiv 0$ en mesure, mais pas en norme L^p. En passant, pour prouver le théorème précédent, on a obtenu une inégalité très utilisée en probabilités :

LEMME 5.10 (Inégalité de Chebychev). *Si $f \in \mathcal{L}^p$, alors pour tout $a > 0$,*

$$\mu(|f| \geq a) \leq \frac{(\|f\|_p)^p}{a^p} .$$

Plus p est grand, plus la convergence en norme L^p est forte :

LEMME 5.11. *Si $\mu(X) < \infty$ (en particulier si μ est une probabilité), et si $1 \leq p < q < \infty$, alors $f_n \xrightarrow{L^q} f$ implique $f_n \xrightarrow{L^p} f$.*

DÉMONSTRATION. Par l'inégalité (4.3) prouvée dans le Lemme 4.3, $\|f_n - f\|_p \leq \mu(X)^{(q-p)/pq} \|f_n - f\|_q$. □

Si l'espace n'est pas de mesure finie, le résultat n'est plus vrai. En effet, $f_n = n^{-1} 1_{\{|x| \leq n\}}$ converge vers $f \equiv 0$ au sens de la norme L^2 mais pas au sens de la norme L^1.

Sans hypothèses supplémentaires, il n'y a pas d'implications directes entre la convergence presque-sûre et la convergence en norme. L'exemple précédent est un exemple d'une suite qui couverge vers zéro presque partout mais pas en norme L^1, et si on considère comme plus haut la suite $f_1 := 1$, $f_2 = 1_{[0,1/2]}$, $f_3 = (1/2, 1]$, $f_4 = 1_{[0,1/4]}$, etc., alors f_n converge vers 0 dans toutes les normes L^p, mais ne converge nulle part vers zéro.

THÉORÈME 5.12. *Si $f_n \stackrel{p.p.}{\to} f$, et si il existe $g \in L^p$ telle que $|f_n| \leq g$, alors $f \in L^p$ et $f_n \stackrel{L^p}{\to} f$.*

DÉMONSTRATION. Comme $|f_n| \leq g$, on a aussi $|f| \leq g$ presque partout, et donc $f \in L^p$. Puisque $|f_n - f| \leq 2g$, par convergence dominée $\lim_n \int |f_n - f|^p d\mu = \int \lim_n |f_n - f|^p d\mu = 0$. Donc $f_n \stackrel{L^p}{\to} f$. $\square$

4. Théorèmes de Densité et Approximations dans L^p

Rappelons qu'un ensemble $\mathcal{D} \subset L^p$ est dense si pour tout $f \in L^p$ et pour tout $\epsilon > 0$ il existe $g \in \mathcal{D}$ telle que $\|g - f\|_p \leq \epsilon$. En d'autres termes, $\mathcal{D}$ est dense si pour tout $f \in L^p$ il existe une suite $g_n \in \mathcal{D}$ telle que $g_n \stackrel{L^p}{\to} f$.

THÉORÈME 5.13. *L'espace des fonctions étagées $\varphi = \sum_{k=1}^n a_k 1_{E_k} \in \mathcal{E}$ avec $\mu(E_k) < \infty$ pour tout k, est dense dans $L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$.*

DÉMONSTRATION. Soit $f = f^+ - f^- \in L^p$. Par la Proposition 3.10, il existe des suites de fonctions étagées $\varphi_n^\pm \nearrow f^\pm$. Alors $\varphi_n := \varphi_n^+ - \varphi_n^-$ est étagée et $\varphi_n(x) \to f(x)$ partout. De plus, $|\varphi_n| \leq |f|$. Comme $f \in L^p$, alors $\varphi_n \in L^p$, et donc la mesure de chaque ensemble E_k contenu dans φ_n est finie. Mais on a aussi $|\varphi_n - f| \leq 2|f|$, donc par convergence dominée $\|\varphi_n - f\|_p \to 0$. $\square$

THÉORÈME 5.14. *L'espace $C_o(\mathbb{R}^d)$ des fonctions continues à support compact est dense dans $L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$.*

PREUVE DU THÉORÈME 5.14 : Soit $f \in L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$, et $\epsilon > 0$. Par le Théorème 5.13, il existe $\varphi \in \mathcal{E}$ de la forme $\varphi = \sum_j a_j 1_{E_j}$, où les E_j sont des Boréliens de mesure finie, telle que $\|\varphi - f\|_p \leq \epsilon$. Il reste donc à vérifier que φ peut être approximée (dans la norme $\|\cdot\|_p$) par une fonction g continue à support compact. Comme φ est une combinaison linéaire d'indicatrices, on cherche à approximer chaque

1_{E_j} par une fonction g_j continue à support compact. Remarquons que l'on peut supposer E_j borné [1].

LEMME 5.15. *Soit $B \in \mathcal{B}(\mathbb{R}^d)$ un Borélien borné, $\lambda^d(B) < \infty$. Alors pour tout $\epsilon > 0$, il existe $g \in C_o(\mathbb{R}^d)$ telle que $\|g - 1_B\|_p \leq \epsilon$.*

DÉMONSTRATION. Rappelons que la mesure de Lebesgue λ^d est régulière (Théorème 2.38). Considérons donc un compact $K \subset B$ et un ouvert $G \supset B$ tels que $\lambda^d(K \setminus G) \leq \epsilon^p$, et définissons

$$g(x) := \frac{\mathrm{dist}(x, G^c)}{\mathrm{dist}(x, G^c) + \mathrm{dist}(x, K)}, \quad x \in \mathbb{R}^d.$$

Comme B est borné, on peut supposer G borné, et g est donc à support compact. Elle est de plus continue ; en effet, on montre facilement que pour tout $A \subset \mathbb{R}^d$, $x \mapsto d(x, A)$ est continue. Donc $g \in C_o(\mathbb{R}^d)$. On a aussi $|g| \leq 1$ partout, et $g = 1$ sur K. Ainsi,

$$\|g - 1_B\|_p = \Big(\int |g - 1_B|^p d\lambda^d\Big)^{1/p} \leq \lambda^d(G \backslash K)^{1/p} \leq \epsilon. \qquad \square$$

Appliquons le lemme à chaque indicatrice 1_{E_j} : il existe $g_j \in C_o(\mathbb{R}^d)$ telle que $\|g_j - 1_{E_j}\|_p \leq \epsilon/a$, où $a = \sum_j |a_j|$. Définissons $g := \sum_j a_j g_j$. Alors $g \in C_o(\mathbb{R}^d)$, et $\|\varphi - g\|_p \leq \sum_j |a_j| \|1_{E_j} - g_j\|_p \leq \epsilon$. On a donc $\|g - f\|_p \leq \|g - \varphi\|_p + \|\varphi - f\|_p \leq 2\epsilon$. $\square$

On peut reproduire le même genre de résultat dans un cadre plus général, en supposant la régularité de la mesure :

THÉORÈME 5.16. *Soit (X, ρ) un espace métrique, $\mathcal{B}$ la tribu des ensembles Boréliens. Soit μ une mesure sur $(X, \mathcal{B})$,* **régulière**, *c'est-à-dire que pour tout $B \in \mathcal{B}$ et pour tout $\epsilon > 0$ il existe un ouvert $G \supset B$ et un compact $K \subset B$ tel que $\mu(G \backslash K) \leq \epsilon$. Alors l'ensemble $C(X)$ des fonctions continues intégrables est dense dans $L^p(X, \mathcal{B}, \mu)$.*

5. Le Théorème d'Egorov

THÉORÈME 5.17 (Egorov). *Soit $(X, \mathcal{F}, \mu)$, où μ est finie. Si $f_n \overset{p.p.}{\to} f$, alors pour tout $\delta > 0$ il existe un ensemble mesurable $E_\delta \subset X$ tel que $\mu(E_\delta) \geq \mu(X) - \delta$ et tel que sur E_δ, f_n converge uniformément vers f :*

$$\sup_{x \in E_\delta} |f_n(x) - f(x)| \to 0$$

1. En effet, si E_j n'est pas borné, on peut considérer $E_j(n) := E_j \cap [-n, n]^d$, qui satisfait $E_j(n) \nearrow E_j$, et donc $\|1_{E_j} - 1_{E_j(n)}\|_p^p = \lambda^d(E_j \setminus E_j(n)) = \lambda^d(E_j) - \lambda^d(E_j(n)) \to 0$.

Si la mesure n'est pas finie on construit facilement des contre-exemples : sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda^1)$, la suite $f_n = 1_{[n,n+1]}$ converge partout vers zéro, mais il n'existe pas d'ensemble (de mesure finie) sur lequel la convergence soit uniforme.

Démonstration. On définit

$$E_n^m := \bigcap_{k\geq n} \{|f_k - f| \leq 1/m\} \in \mathcal{F}\,.$$

Clairement, $E_n^m \subset E_{n+1}^m$ et $E^m := \bigcup_{n\geq 1} E_n^m$. Puisque $E_n^m \nearrow E^m$, on a $\mu(E_n^m) \nearrow \mu(E^m)$. Comme $\mu(E^m) = \mu(X) < \infty$, il existe donc $n_0 = n_0(m)$ tel que $\mu(E^m \backslash E_{n_0(m)}^m) \leq \delta/2^m$. Soit

$$E_\delta := \bigcap_{m\geq 1} E_{n_0(m)}^m\,.$$

Si $x \in E_\delta$, alors pour tout m, $|f_k(x) - f(x)| \leq 1/m$ pour tout $k \geq n_0(m)$, donc $\lim_{k\to\infty} \sup_{x\in E_\delta} |f_k(x) - f(x)| = 0$. Remarquons alors que

$$\begin{aligned}
\mu(X\backslash E_\delta) &= \mu\Big(X \cap \big(\bigcap_{m\geq 1} E_{n_0(m)}^m\big)^c\Big) \\
&\leq \sum_{m\geq 1} \mu(X\backslash E_{n_0(m)}^m) \\
&= \sum_{m\geq 1} \mu(E^m\backslash E_{n_0(m)}^m) \\
&\leq \delta\,.
\end{aligned}$$

□

5.1. Le Théorème de Lusin. Prenons la fonction $f := 1_{\mathbb{Q}\cap[a,b]}$, qui est mesurable mais discontinue en tout point $x \in [a, b]$. Si l'on modifie f en posant $f(x) := 0$ en tout rationnel $x \in \mathbb{Q} \cap [a, b]$, alors on obtient la fonction identiquement nulle, qui est continue partout.

Une autre façon de rendre cette fonction f continue sur la plupart de son domaine de définition, et qui permet de faire le lien avec le théorème à suivre, est la suivante. Soit $\epsilon > 0$ un nombre arbitrairement petit, fixé. Considérons une énumération des rationnels de $[a, b]$, $\mathbb{Q} \cap [a, b] = \{q_1, q_2, \dots\}$, et soit I_n l'intervalle ouvert de taille $\epsilon/2^n$ centré en q_n. Soit $K := [a, b]\backslash \bigcup_{n\geq 1} I_n$. Alors

$$\lambda^1(K) = \lambda^1([a, b]\backslash \bigcup_n I_n) \geq b - a - \sum_n \lambda^1(I_n) \geq b - a - \epsilon\,.$$

(En particulier, K est non-vide !) De plus, K compact, et ne contient que des irrationnels. En particulier, la restriction de f à K est la fonction identiquement nulle, qui est continue.

Un phénomène semblable est vrai, beaucoup plus généralement : toute fonction mesurable est continue sur une grande partie de son domaine de définition.

THÉORÈME 5.18 (Lusin). *Soit $f : [a, b] \to \mathbb{R}$ une fonction mesurable au sens de Lebesgue, finie. Alors pour tout $\epsilon > 0$ il existe un compact $K \subset [a, b]$ tel que*

(1) $\lambda^1(K) \geq \lambda^1([a, b]) - \epsilon$.

(2) La restriction de f à K est continue.

DÉMONSTRATION. En exercice. □

CHAPITRE 6

Produits

Dans ce chapitre, on se donne deux espaces mesurés $(X, \mathcal{F}, \mu)$, $(Y, \mathcal{G}, \nu)$, et on étudie l'espace produit

$$X \times Y := \{(x, y) : x \in X, y \in Y\}.$$

On associera à $X \times Y$ une structure mesurable et on construira la mesure produit $\mu \otimes \nu$. On s'intéressera en particulier à l'intérgation des fonctions $f : X \times Y \to \mathbb{R}$ par rapport à la mesure produit. Le principal résultat que l'on prouvera dans la Section 3, le *Théorème de Fubini*, donnera des conditions sous lesquelles

$$\begin{aligned}\int f d(\mu \otimes \nu) &= \int_X \Big(\int_Y f(x, y) \nu(dy) \Big) \mu(dx) \\ &= \int_Y \Big(\int_X f(x, y) \mu(dx) \Big) \nu(dy).\end{aligned}$$

Remarquons en passant que les résultats présentés ici s'étendent sans complications majeures au cas d'un nombre fini de facteurs :

$$X_1 \times X_2 \times \cdots \times X_n.$$

On n'abordera pas le problème de la construction de la mesure produit sur un produit infini :

$$X_1 \times X_2 \times \ldots$$

1. Structure mesurable sur $X \times Y$

Soient donc $(X, \mathcal{F}, \mu)$, $(Y, \mathcal{G}, \nu)$ deux espaces mesurables. Rappelons comment on définit une structure mesurable sur le produit cartésien $X \times Y$, comme dans l'Exemple 2.6.

On définit d'abord les rectangles

$$\mathcal{R} := \big\{A \times B : A \in \mathcal{F}, B \in \mathcal{G}\big\}.$$

Alors, $\mathcal{F} \otimes \mathcal{G} := \sigma(\mathcal{R})$ est la tribu produit sur $X \times Y$, et $(X \times Y, \mathcal{F} \otimes \mathcal{G})$ est l'espace produit.

Pour définir la mesure produit, on aura besoin du

LEMME 6.1. *Soit $\mathcal{A} \subset \mathcal{P}(X \times Y)$ la famille des unions finies disjointes de rectangles. Alors $\mathcal{A}$ est une algèbre, et $\sigma(\mathcal{A}) = \mathcal{F} \otimes \mathcal{G}$.*

DÉMONSTRATION. Il est clair que $\mathcal{A}$ est une algèbre, puisque

$$(A \times B)^c = (A^c \times B) \cup (A \times B^c) \cup (A^c \times B^c)\,.$$

Comme $\mathcal{A} \supset \mathcal{R}$, on a $\sigma(\mathcal{A}) \supset \sigma(\mathcal{R})$. D'autre part, puisque $\sigma(\mathcal{R})$ est une tribu, elle doit contenir les unions finies d'éléments de $\mathcal{R}$, donc $\mathcal{A} \subset \sigma(\mathcal{R})$, et donc $\sigma(\mathcal{A}) \subset \sigma(\mathcal{R})$. □

PROPOSITION 6.2. *Il existe une mesure sur $(X \times Y, \mathcal{F} \otimes \mathcal{G})$, notée $\mu \otimes \nu$ et appelée* **mesure produit de μ avec ν**, *telle que*

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B) \quad \forall \text{ rectangle } A \times B \in \mathcal{R}\,. \tag{6.1}$$

Si μ et ν sont σ-finies, alors $\mu \otimes \nu$ est l'unique mesure avec cette propriété.

DÉMONSTRATION. Définissons d'abord une fonction d'ensemble sur les rectangles par

$$\rho(A \times B) := \mu(A)\nu(B)\,.$$

On peut ensuite étendre ρ aux unions finies disjointes de rectangles, $\bigcup_{i=1}^n A_i \times B_i \in \mathcal{A}$, en posant

$$\rho\Big(\bigcup_{j=1}^n A_i \times B_i\Big) := \sum_{j=1}^n \rho(A_i \times B_i)\,.$$

Il est facile de voir que ρ ne dépend pas de la représentation choisie (comme dans la preuve de la Proposition 2.30, on remarque que l'on peut toujours prendre une partition plus fine). Pour voir que ρ est une pré-mesure, considérons une union disjointe $\bigcup_{n\geq 1} A_n \times B_n \in \mathcal{A}$. On peut toujours supposer qu'en fait cette union est un rectangle : $\bigcup_{n\geq 1} A_n \times B_n \equiv A \times B$. On peut alors écrire sa fonction indicatrice

$$1_{A\times B}(x,y) = 1_A(x)1_B(y) = \sum_{n\geq 1} 1_{A_n}(x)1_{B_n}(y)\,.$$

Observons que $1_B(y)$ et chacune des fonctions $1_{B_n}(y)$ est mesurable par rapport à la tribu $\mathcal{G}$. On peut donc maintenir $x \in X$ fixé et intégrer les deux côtés de l'identité ci-dessus par rapport à ν. Par convergence monotone on obtient

$$1_A(x)\nu(B) = \sum_{n\geq 1} 1_{A_n}(x)\nu(B_n)\,.$$

En intégrant ensuite par rapport à μ, on obtient

$$\mu(A)\nu(B) = \sum_{n \geq 1} \mu(A_n)\nu(B_n)\,,$$

c'est-à-dire $\rho(A \times B) = \sum_{n\geq 1} \rho(A_n \times B_n)$, ce qui prouve que ρ est une pré-mesure. Par le Théorème d'Extension de Carathéodory, ρ s'étend à une mesure sur $\mathcal{F} \otimes \mathcal{G} = \sigma(\mathcal{A})$, notée $\mu \otimes \nu$, et cette mesure coïncide avec ρ sur $\mathcal{A}$. Si μ et ν sont σ-finies, alors ρ est σ-finie, et l'extension est unique. □

Nous avons donc défini le produit

$$(X, \mathcal{F}, \mu) \otimes (Y, \mathcal{G}, \nu) := (X \times Y, \mathcal{F} \otimes \mathcal{G}, \mu \otimes \nu)\,.$$

Remarquons que par la construction de l'extension d'une pré-mesure (c.f. preuve du Théorème 2.29), on sait que pour tout $E \in \mathcal{F} \otimes \mathcal{G}$,

$$(\mu \otimes \nu)(E) = \inf \sum_{n} \rho(A_n \times B_n)\,,$$

où l'infimum porte sur tous les recouvrements dénombrables de E par des rectangles.

Pour alléger les notations, on écrira l'ensembles des fonctions intégrables (par rapport à $\mu\otimes\nu$) $\mathcal{L}^1(X\times Y)$ au lieu de $\mathcal{L}^1(X\times Y, \mathcal{F}\otimes\mathcal{G}, \mu\otimes\nu)$. L'intégrale de Lebesgue d'une fonction mesurable $f : X \times Y \to \mathbf{R}$ s'écrira

$$\int_{X\times Y} f d(\mu \otimes \nu)\,, \qquad \text{ou} \int f d(\mu \otimes \nu)\,.$$

Nous nous intéresserons ensuite aux relations existant entre l'intégration d'une fonction $f : X \times Y \to \mathbb{R}$ par rapport à la mesure produit $\mu \otimes \nu$ et les intégrations itérées par rapport aux mesures μ et ν.

1.1. La mesure de Lebesgue sur $\mathbb{R}^d$. Avant de passer à l'intégration des fonctions, montrons que la mesure de Lebesgue sur $\mathbb{R}^d$, $d \geq 2$, que nous avons construit dans la Section 6 à l'aide des pavés, peut être obtenue à partir du produit de la mesure de Lebesgue sur la droite avec elle-même :

Théorème 6.3. $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d) = \bigotimes_{k=1}^{d} (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda^1)\,.$

Démonstration. On peut d'abord vérifier (exercice) que $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$, et que $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R}^{d-1}) \otimes \mathcal{B}(\mathbb{R})$. Ceci montre bien que $\mathcal{B}(\mathbb{R}^d) = \bigotimes_{k=1}^{d} \mathcal{B}(\mathbb{R})$. Montrons ensuite que $\lambda^d = \lambda^1 \otimes \cdots \otimes \lambda^1$ (d fois). Soit $\lambda^{(d)} := \lambda^1 \otimes \cdots \otimes \lambda^1$. Il est clair que $\lambda^{(d)}$ et λ^d coïncident sur les pavés. Mais comme les pavés génèrent $\mathcal{B}(\mathbb{R}^d)$ et sont stables par intersection, ceci implique $\lambda^{(d)} = \lambda^d$. □

2. Intégrer les sections

Si $E \subset X \times Y$, on définit les sections de E :

$$\forall\, x \in X\,, \quad E_x := \{y \in Y : (x,y) \in E\}\,,$$
$$\forall\, y \in Y\,, \quad E_y := \{x \in X : (x,y) \in E\}\,.$$

Par exemple, pour un rectangle $A \times B \in \mathcal{R}$, $(A \times B)_x = B$ si $x \in A$, $\varnothing$ sinon. De même, $(A \times B)_y = B$ si $y \in B$, $\varnothing$ sinon.

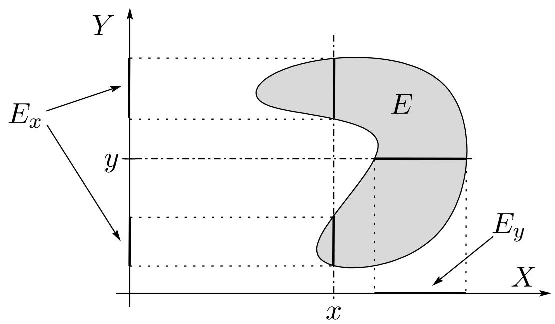


FIGURE 1. Les sections d'un ensemble $E \subset X \times Y$.

Un premier fait non-trivial est que les sections sont mesurables :

LEMME 6.4. *Si $E \in \mathcal{F} \otimes \mathcal{G}$, alors*

$$\forall\, x \in X\,, E_x \in \mathcal{G}\,,$$
$$\forall\, y \in Y\,, E_y \in \mathcal{F}\,.$$

DÉMONSTRATION. Fixons $x \in X$, et considérons $\mathcal{E}_x := \{E \in \mathcal{F} \otimes \mathcal{G} : E_x \in \mathcal{G}\}$. Puisque $(X \times Y)_x = Y \in \mathcal{G}$, que $(E^c)_x = (E_x)^c$, et que pour toute suite $E_n \in \mathcal{F} \otimes \mathcal{G}$, $(\bigcup_n E_n)_x = \bigcup_n (E_n)_x$, $\mathcal{E}_x$ est une tribu. Comme elle contient les rectangles, on a $\mathcal{E}_x = \mathcal{F} \otimes \mathcal{G}$. On définit $\mathcal{E}_y$ de la même façon. □

COROLLAIRE 6.5. *Soit $f : X \times Y \to \mathbb{R}$ une fonction $\mathcal{F} \otimes \mathcal{G}$-mesurable. Alors*

$$\forall\, x \in X\,, \quad y \mapsto f(x,y) \text{ est } \mathcal{G}\text{-mesurable}\,,$$
$$\forall\, y \in Y\,, \quad x \mapsto f(x,y) \text{ est } \mathcal{F}\text{-mesurable}\,.$$

DÉMONSTRATION. Fixons $x \in X$. Alors $\{y \in Y : f(x,y) \leq \alpha\} = \{(x',y) : f(x',y) \leq \alpha\}_x$, qui est mesurable par le Lemme 6.4. □

Le résultat suivant décrit la propriété principale de la mesure produit : $(\mu\otimes\nu)(E)$ peut être calculé en intégrant les mesures des sections.

PROPOSITION 6.6. *Si $(X,\mathcal{F},\mu)$ et $(Y,\mathcal{G},\nu)$ sont σ-finis, alors pour tout $E\in\mathcal{F}\otimes\mathcal{G}$, on a*

$$x\mapsto\nu(E_x)\ \textit{est } \mathcal{F}\textit{-mesurable}\,, \tag{6.2}$$

$$y\mapsto\mu(E_y)\ \textit{est } \mathcal{G}\textit{-mesurable}\,, \tag{6.3}$$

$$(\mu\otimes\nu)(E)=\int_X\nu(E_x)\mu(dx)=\int_Y\mu(E_x)\nu(dy)\,. \tag{6.4}$$

DÉMONSTRATION. Supposons d'abord que μ et ν sont finies. Soit $\mathcal{D}$ la classe des ensembles $E\in\mathcal{F}\otimes\mathcal{G}$ pour lesquels (6.2)-(6.4) sont vérifiées. On montre d'abord que $\mathcal{D}$ contient les rectangles : comme $(A\times B)_x=B$ si $x\in A$, $\varnothing$ sinon, $x\mapsto\nu((A\times B)_x)=1_A(x)\nu(B)$, qui est $\mathcal{F}$-mesurable. De plus,

$$\begin{aligned}(\mu\otimes\nu)(A\times B)=\mu(A)\nu(B)&=\int_X 1_A(x)\nu(B)\mu(dx)\\&\equiv\int_X\nu((A\times B)_x)\mu(dx)\,.\end{aligned}$$

On montre de même que $y\mapsto\mu((A\times B)_y)=1_B(y)\mu(A)$ est $\mathcal{G}$-mesurable et que

$$(\mu\otimes\nu)(A\times B)=\int_Y\mu((A\times B)_y)\nu(dy)\,.$$

On a donc bien $\mathcal{D}\supset\mathcal{R}$. Montrons ensuite que $\mathcal{D}$ est une classe de Dynkin (Définition 2.19). D'abord, puisque $X\times Y$ est un rectangle, $X\times Y\in\mathcal{D}$. Ensuite, si $E,F\in\mathcal{D}$, $E\subset F$, alors $\nu((F\backslash E)_x)=\nu(F_x)-\nu(E_x)$, $\mu((F\backslash E)_y)=\mu(F_y)-\mu(E_y)$ et $(\mu\otimes\nu)(F\backslash E)=(\mu\otimes\nu)(F)-(\mu\otimes\nu)(E)$ (puisque μ et ν sont finies), donc (6.2)-(6.4) sont vérifiées. Finalement, soit $E_n\in\mathcal{F}\otimes\mathcal{G}$, $E_n\nearrow E$. Alors $\nu((E_n)_x)\nearrow\nu(E_x)$, $\mu((E_n)_y)\nearrow\mu(E_y)$, donc (6.2)-(6.3) valent pour E. On vérifie (6.4) pour E par convergence monotone. Par exemple,

$$\begin{aligned}(\mu\otimes\nu)(E)&=\lim_{n\to\infty}\uparrow(\mu\otimes\nu)(E_n)\\&=\lim_{n\to\infty}\uparrow\int_X\nu((E_n)_x)\mu(dx)=\int_X\nu(E_x)\mu(dx)\,.\end{aligned}$$

Donc $\mathcal{D}$ est une classe de Dynkin qui contient les rectangles. Puisque les rectangles sont stables sous intersections, le Théorème 2.21 implique que $\mathcal{D}=\mathcal{F}\otimes\mathcal{G}$.

Supposons ensuite que μ et ν sont σ-finies : il existe $X_n\in\mathcal{F}$, $\mu(X_n)<\infty$, $X_n\nearrow X$, et $Y_n\in\mathcal{G}$, $\nu(Y_n)<\infty$, $Y_n\nearrow Y$. Si $D_n:=X_n\times Y_n$, alors

$D_n \nearrow X\times Y$, et pour tout E on peut écrire $\nu(E_x) = \lim_n \nu((E\cap D_n)_x)$, $\mu(E_y) = \lim_n \mu((E\cap D_n)_y)$, ce qui prouve (6.2)-(6.3) par le paragraphe précédent. Ensuite, $(\mu\otimes\nu)(E) = \lim_n(\mu\otimes\nu)(E\cap D_n)$, ce qui permet de vérifier (6.4). □

3. Le Théorème de Tonelli-Fubini

On peut alors passer à l'intégration des fonctions mesurables non-négatives. Comme avant, pour alléger l'écriture, on écrira $\mathcal{M}^+(X,\mathcal{F}) \equiv \mathcal{M}^+(X)$, $\mathcal{M}^+(Y,\mathcal{G}) \equiv \mathcal{M}^+(Y)$ et $\mathcal{M}^+(X\times Y,\mathcal{F}\otimes\mathcal{G}) \equiv \mathcal{M}^+(X\times Y)$.

Théorème 6.7 (Tonelli). *Si $(X,\mathcal{F},\mu)$ et $(Y,\mathcal{G},\nu)$ sont σ-finis et $f\in\mathcal{M}^+(X\times Y)$, alors*

$$x\mapsto \int_Y f(x,y)\nu(dy) \ \textit{est } \mathcal{F}\textit{-mesurable}\,, \tag{6.5}$$

$$y\mapsto \int_X f(x,y)\mu(dx) \ \textit{est } \mathcal{G}\textit{-mesurable}\,, \tag{6.6}$$

et

$$\int f d(\mu\otimes\nu) = \int_X \Big(\int_Y f(x,y)\nu(dy)\Big)\mu(dx)\,, \tag{6.7}$$

$$= \int_Y \Big(\int_X f(x,y)\mu(dx)\Big)\nu(dy)\,. \tag{6.8}$$

Démonstration. Si $f\in\mathcal{M}^+(X\times Y)$, on considère une suite de fonctions étagées $\varphi_n \nearrow f$. Pour φ_n, les propriétés (6.5)-(6.8) suivent directement de la Proposition 6.6. En effet, $\int_Y 1_E(x,y)\nu(dy) = \nu(E_x)$, $\int_X 1_E(x,y)\mu(dx) = \mu(E_y)$. On étend ensuite ces propriétés à f en utilisant trois fois le Théorème de la Convergence Monotone. D'abord sur $X\times Y$, puis sur X et enfin sur Y :

$$\begin{aligned}
\int f d(\mu\otimes\nu) &= \lim_{n\to\infty}\int \varphi_n d(\mu\otimes\nu)\\
&= \lim_{n\to\infty}\int_X\Big(\int_Y \varphi_n(x,y)\nu(dy)\Big)\mu(dx)\\
&= \int_X \lim_{n\to\infty}\Big(\int_Y \varphi_n(x,y)\nu(dy)\Big)\mu(dx)\\
&= \int_X \Big(\int_Y \lim_{n\to\infty}\varphi_n(x,y)\nu(dy)\Big)\mu(dx)\\
&= \int_X \Big(\int_Y f(x,y)\nu(dy)\Big)\mu(dx)\,,
\end{aligned}$$

ce qui montre (6.7) ; (6.8) se démontre de la même façon. □

Le Théorème de Tonelli est souvent utilisé pour montrer qu'une fonction $f : X \times Y \to \mathbf{R}$ est intégrable par rapport à la mesure produit. En effet, en appliquant le résultat précédent à $|f|$,

$$\int |f| d(\mu \otimes \nu) = \int_X \Big(\int_Y |f(x,y)| \nu(dy) \Big) \mu(dx) \,, \tag{6.9}$$

$$= \int_Y \Big(\int_X |f(x,y)| \mu(dx) \Big) \nu(dy) \,. \tag{6.10}$$

Donc, si l'une quelconque des intégrales doubles du membre de droite est finie, alors l'autre est finie aussi, a la même valeur, et f est intégrable.

EXEMPLE 6.8. La fonction $(x,y) \mapsto \frac{e^{-xy}}{x}$, $(x,y) \in [1,\infty)^2$ est-elle intégrable par rapport à la mesure de Lebesgue λ^2 ? Par le Théorème 6.3, on sait que l'intégration par rapport à λ^2 est équivalente à l'intégration par rapport à la mesure produit $\lambda^1 \otimes \lambda^1$. Or

$$\int_1^\infty \Big(\int_1^\infty \frac{e^{-xy}}{x} \lambda^1(dy) \Big) \lambda^1(dx) = \int_1^\infty \frac{1}{x^2} \lambda^1(dx) < \infty \,.$$

Donc par le Théorème de Tonelli, $\frac{e^{-xy}}{x}$ est intégrable sur $[1,\infty)^2$.

THÉORÈME 6.9 (Théorème de Fubini). *Si $(X, \mathcal{F}, \mu)$ et $(Y, \mathcal{G}, \nu)$ sont σ-finis et $f \in \mathcal{L}^1(X \times Y)$, alors*

$$x \mapsto \int_Y f(x,y) \nu(dy) \text{ est définie } \mu\text{-p.p. et } \mathcal{F}\text{-mesurable} \,, \tag{6.11}$$

$$y \mapsto \int_X f(x,y) \mu(dx) \text{ est définie } \nu\text{-p.p. et } \mathcal{G}\text{-mesurable} \,, \tag{6.12}$$

et

$$\int f d(\mu \otimes \nu) = \int_X \Big(\int_Y f(x,y) \nu(dy) \Big) \mu(dx) \,, \tag{6.13}$$

$$= \int_Y \Big(\int_X f(x,y) \mu(dx) \Big) \nu(dy) \,. \tag{6.14}$$

DÉMONSTRATION. Si $f = f^+ - f^-$ est intégrable, alors f^+ et f^- ont des intégrales finies : $\int f^\pm d(\mu \otimes \nu) < \infty$. Montrons (6.13). Par (6.7),

$$\int f^\pm d(\mu \otimes \nu) = \int_X \Big(\int_Y f^\pm(x,y) \nu(dy) \Big) \mu(dx) \,, \tag{6.15}$$

et donc les fonctions $x \mapsto \phi^{\pm}(x) := \int_Y f^{\pm}(x,y)\nu(dy)$ sont finies μ-presque partout, et $\mathcal{F}$-mesurables par (6.5). Donc

$$\phi(x) := \phi^+(x) - \phi^-(x) = \int_Y f(x,y)\nu(dy)$$

est bien définie μ-presque partout (où elle n'est pas définie, on pose $\phi \equiv 0$), $\mathcal{F}$-mesurable et intégrable par rapport à μ. De plus,

$$\begin{aligned}\int_X \phi d\mu &= \int_X \phi^+ d\mu - \int_X \phi^- d\mu \\ &= \int_{X\times Y} f^+ d(\mu\otimes\nu) - \int_{X\times Y} f^- d(\mu\otimes\nu) \\ &\equiv \int_{X\times Y} f d(\mu\otimes\nu)\,.\end{aligned}$$

Ceci prouve (6.13). On obtient (6.14) en inversant l'ordre d'intégration dans le terme de droite de (6.15) et en procédant de la même manière. □

Montrons que les deux hypothèses du Théorème de Fubini, intégrabilité de f et σ-finitude des mesures μ et ν, sont essentielles. Le premier contre-exemple montre l'importance de l'intégrabilité de f.

Exemple 6.10. Soient $X = Y = \mathbb{N}$ munis de la tribu discrète et $\mu = \nu =$mesure de comptage. Si $m \geq 1$ alors $f(m,m) := 1$ et $f(m+1,m) := -1$. Partout ailleurs, $f(m,n) := 0$. Alors

$$\int_{X\times Y} |f| d(\mu\otimes\nu) = \sum_m \sum_n |f(m,n)| = \infty\,,$$

donc f n'est pas intégrable. Or on vérifie facilement que

$$\int_X \Big(\int_Y f(x,y)\nu(dy)\Big)\mu(dx) = \sum_m \sum_n f(m,n) = 1$$

$$\int_Y \Big(\int_X f(x,y)\mu(dx)\Big)\nu(dy) = \sum_n \sum_m f(m,n) = 0\,.$$

Un exemple semblable, basé sur la même idée :

Exemple 6.11. Sur $X = Y = [0,1]$, avec la mesure de Lebesgue, considérons la fonction

$$f(x,y) := \begin{cases} 2^{2n} & \text{si } (x,y) \in [\frac{1}{2^n}, \frac{1}{2^{n-1}}] \times [\frac{1}{2^n}, \frac{1}{2^{n-1}}), \\ -2^{2n+1} & \text{si } (x,y) \in [\frac{1}{2^{n+1}}, \frac{1}{2^n}) \times [\frac{1}{2^n}, \frac{1}{2^{n-1}}), \\ 0 & \text{sinon.} \end{cases}$$

On vérifie facilement que

$$\int_0^1 \Big(\int_0^1 f(x,y)dy \Big) dx = 1\,, \quad \int_0^1 \Big(\int_0^1 f(x,y)dx \Big) dy = 0\,.$$

On en déduit que f ne peut pas être intégrable, et donc que l'on a nécessairement

$$\int_0^1 \Big(\int_0^1 |f(x,y)|dy \Big) dx = \infty\,,$$

ce qui se vérifie facilement.

Le prochain contre-exemple montre que l'hypothèse de σ-finitude des mesures est essentielle :

EXEMPLE 6.12. Soit $X = (0,1)$, muni de la tribu des Boréliens et de la mesure de Lebesgue λ^1, et $Y = (0,1)$ muni de la tribu discrète et de la mesure de comptage μ_c. Soit $f := 1_\Delta$, où Δ est la diagonale $\Delta := \{(x,x) : x \in (0,1)\}$. Alors f est mesurable (exercice), et

$$\int_X \Big(\int_Y 1_\Delta d\lambda^1 \Big) d\mu_c \neq \int_Y \Big(\int_X 1_\Delta d\mu_c \Big) d\lambda^1\,.$$

Ceci est bien-sûr dû au fait que la mesure de comptage sur $(0,1)$ n'est pas σ-finie.

4. Changements de Variables sur $\mathbb{R}^d$

Même s'il n'est pas utilisé seulement dans ce but, le Théorème de Fubini permet souvent de *calculer* des intégrales de fonctions de plusieurs variables. Nous en avons déjà vu des exemples simples plus haut, et nous passons maintenant à des exemples classiques de son utilisation.

Nous avons vu dans la Section 6.2 une formule permettant de transformer l'intégrale d'une fonction $f = f(x)$, $x \in D \subset \mathbb{R}^d$, en une intégrale portant sur des nouvelles variables, $u = \varphi^{-1}(x)$, $u \in \varphi^{-1}(D)$. Un changement de variable utile se fait en général en fonction des symétries en présence, et le Théorème de Fubini permet parfois de *calculer* l'intégrale écrite en fonction des nouvelles variables.

Par exemple, considérons l'intégrale

$$I = \int e^{-x^2} \lambda^1(dx) = \int_{-\infty}^{\infty} e^{-x^2} dx\,.$$

Comme l'intégrant est positif, le Théorème de Tonelli permet d'écrire I^2 de la façon suivante :

$$\begin{aligned} I^2 &= \Big(\int_{-\infty}^{\infty} e^{-x^2}dx\Big)\Big(\int_{-\infty}^{\infty} e^{-y^2}dy\Big) \\ &= \int_{\mathbb{R}^2} e^{-(x^2+y^2)}dxdy \end{aligned}$$

On peut modifier la région d'intégration sur un ensemble de mesure nulle, sans changer la valeur de l'intégrale : si $D_0 := \mathbb{R}^2 \backslash \{(x,0) \in \mathbb{R}^2 : x \geq 0\}$, alors

$$\varphi(r,\theta) := (r\cos\theta, r\sin\theta)$$

est un difféomorphisme de U_0 dans D_0, où $U_0 := \{(r,\theta) : r > 0, \theta \in (0,2\pi)\}$, de Jacobien $J_\varphi(r,\theta) = r$. La formule du changement de variable (3.23) donne

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)}dxdy = \int_{D_0} e^{-(x^2+y^2)}dxdy = \int_{U_0} e^{-r^2}rdrd\theta$$

Une deuxième utilisation du Théorème de Tonelli et une deuxième utilisation de la formule du changement de variable donne

$$\int_{U_0} e^{-r^2}rdrd\theta = \Big(\int_0^{2\pi} d\theta\Big)\Big(\int_0^{\infty} re^{-r^2}dr\Big) = \pi\,.$$

On a donc montré que

$$\int_{-\infty}^{\infty} e^{-x^2}dx = \sqrt{\pi}\,.$$

CHAPITRE 7

Décomposition et dérivation

Considérons la mesure de Lebesgue sur $\mathbb{R}^d$, λ^d. On sait par le Théorème 2.35 que λ^d est invariante sous translations, et qu'elle "charge" les pavés proportionnellement à leur volume.

Soit ν une autre mesure sur $\mathbb{R}^d$. Supposons que l'on veuille comparer ν avec la mesure de Lebesgue, *localement*. Un moyen naturel de faire cette comparaison est d'étudier

$$f(x) := \lim_{\epsilon \to 0^+} \frac{\nu(B_\epsilon(x))}{\lambda^1(B_\epsilon(x))}, \tag{7.1}$$

où $B_\epsilon(x) := \{y : \|y - x\| < \epsilon\}$, $\|\cdot\|$ étant la norme euclidienne. Lorsque cette limite existe, on dira que $f(x)$ représente la densité de ν par rapport à λ^1 au point x. Pour des raisons évidentes, $f(x)$ sera aussi appelée dérivée de ν par rapport à λ^1.

Pour que la densité existe, ν doit satisfaire une condition de *continuité* par rapport à λ^1 : $\nu(B_\epsilon(x))$ doit tendre vers zéro, et au moins aussi vite que $\lambda^1(B_\epsilon(x))$.

Si l'existence de la densité f est établie, au moins pour λ^1-presque tous les x de la droite, il est naturel d'espérer pouvoir reconstruire ν à partir de λ^1, en intégrant sa densité :

$$\nu(B) = \int_B f d\lambda^1, \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

On étudiera cette question en détails dans ce chapitre, d'abord d'un point de vue beaucoup plus général, sur des espaces ne possédant pas forcément la structure topologique de $\mathbb{R}^d$.

1. Le Théorème de Radon-Nikodým

Commençons avec quelques définitions générales à propos de la comparaison de mesures sur un espace mesurable $(X, \mathcal{F})$.

DÉFINITION 7.1. *Soient μ et ν deux mesures sur $(X, \mathcal{F})$.*

(1) *ν est* absolument continue par rapport à *μ, et on note $\nu \ll \mu$, si $\mu(A) = 0$ ($A \in \mathcal{F}$) implique $\nu(A) = 0$.*
(2) *ν et μ sont* singulières *(ou alors : ν* est singulière par rapport à *μ), et on note $\nu \perp \mu$, si il existe deux ensembles disjoints $E, F \in \mathcal{F}$, $E \cup F = X$, tel que $\nu(E) = 0$, $\mu(F) = 0$.*

Donc deux mesures sont singulières si elles sont supportées par des ensembles disjoints. Par définition, la mesure triviale $\mu \equiv 0$ est absolument continue et singulière par rapport à toute mesure.

On a déjà vu (Théorème 3.21) que si ν est définie à l'aide d'une fonction intégrable $f \in \mathcal{M}^+$ par

$$\nu(A) := \int_A f d\mu \,, \quad \forall A \in \mathcal{F} \,,$$

que l'on note $\nu \equiv f\mu$, alors $\nu \ll \mu$. Le résultat suivant montre que le contraire est aussi vrai, sous une hypothèse de σ-finitude :

THÉORÈME 7.2 (Radon-Nikodým). *Soient μ, ν deux mesures sur $(X, \mathcal{F})$, telles que $\nu \ll \mu$. Si les deux mesures sont σ-finies, alors il existe $f \in \mathcal{M}^+(X, \mathcal{F})$ telle que*

$$\nu(A) = \int_A f d\mu \,, \quad \forall A \in \mathcal{F} \,. \tag{7.2}$$

On appelle f la densité (ou dérivée) de Radon-Nikodým de ν par rapport à *μ, et on la note $f = \frac{d\nu}{d\mu}$. Elle est unique à un ensemble négligeable (par rapport à μ) près.*

La démonstration de ce théorème sera basée sur le Corollaire 7.12, dont la preuve repose sur la notion de charge.

1.1. Charges et décomposition de Hahn. La preuve du Théorème 7.2 repose sur une généralisation de la notion de mesure :

DÉFINITION 7.3. *Une fonction d'ensemble $\eta : \mathcal{F} \to \mathbb{R}$ est une* **charge**[a] *si elle satisfait aux conditions suivantes :*

(1) $\eta(\varnothing) = 0$,

(2) *Si $(A_n)_{n\geq 1}$ est une famille d'ensembles mesurables deux-à-deux disjoints alors*

$$\eta\Big(\bigcup_{n\geq 1} A_n\Big) = \sum_{n\geq 1} \eta(A_n)\,, \tag{7.3}$$

où il est sous-entendu que la série dans (7.3) *converge.*

[a] Dans la littérature, on trouve aussi mesure signée, signed measure.

Il est clair que toute mesure prenant des valeurs finies est une charge. Mais contrairement aux mesures, une charge peut aussi prendre des valeurs négatives. En particulier, une charge *n'est pas nécessairement monotone*. Observons aussi qu'ici $\eta(\cdot)$ prend des valeurs réelles, et non dans la droite achevée (sans quoi la série dans (7.3) pourrait ne pas être définie).

EXEMPLE 7.4. Si μ, ν sont deux mesures finies sur $(X, \mathcal{F})$, alors $\eta := \mu - \nu$ est une charge. On verra qu'en général, toute charge peut être écrite la différence de deux mesures finies (décomposition de Hahn).

On peut facilement vérifier (exercice) que toute mesure signée est continue dans un sens un peu plus faible que celui de la Proposition 2.15 :

LEMME 7.5. *Soit $A_n \in \mathcal{F}$.*

(1) *Si $A_n \nearrow A$, alors $\eta(A_n) \to \eta(A)$.*

(2) *Si $A_n \searrow A$, alors $\eta(A_n) \to \eta(A)$.*

DÉMONSTRATION. Dans le premier cas par exemple, il suffit d'écrire A comme l'union des ensembles $B_n := A_n \backslash A_{n-1}$. □

Puisque son signe peut varier d'un ensemble à l'autre, une charge mène à distinguer les ensembles mesurables de charge positive de ceux de charge négative. Plus précisément, il est intéressant de distinguer les ensembles dont *tous* les sous-ensembles sont ou positifs, ou négatifs :

DÉFINITION 7.6. *Soit η une charge. Un ensemble $E \in \mathcal{F}$ est*

(1) **positif (par rapport à η)** *si $\eta(A) \geq 0$ pour tout $A \subset E$, $A \in \mathcal{F}$.*

(2) **négatif (par rapport à η)** *si $\eta(A) \leq 0$ pour tout $A \subset E$, $A \in \mathcal{F}$.*

LEMME 7.7.

(1) *Si E est positif, alors tout $E' \subset E$ est positif.*

(2) *Si E_n est positif pour tout n, alors $E := \bigcup_n E_n$ est positif.*

DÉMONSTRATION. La première affirmation est triviale. Pour la seconde, on pose $\tilde{E}_k := E_k \cap (E_1 \cup \dots E_{k-1})^c$. Alors les $\tilde{E}_k$ sont disjoints, positifs, et donc pour tout $A \subset E$,

$$\eta(A) = \eta\Big(\bigcup_k A \cap \tilde{E}_k\Big) = \sum_k \eta(A \cap \tilde{E}_k) \geq 0\,,$$

puisque $A \cap \tilde{E}_k \subset \tilde{E}_k$ pour tout k. □

DÉFINITION 7.8. *Une **décomposition de Hahn de X par rapport à η** est une partition de X en une paire (H_+, H_-) d'ensembles mesurables disjoints, $H_+ \cup H_- = X$, telle que H_+ est positif par rapport à η et H_- négatif par rapport à η.*

Une décomposition de Hahn (H_+, H_-) est donc telle que toute partie mesurable de H_+ est chargée positivement, et toute partie mesurable de H_- est chargée négativement.

THÉORÈME 7.9 (Existence d'une Décomposition de Hahn). *Pour toute charge η, il existe une décomposition de Hahn (H_+, H_-) de X par rapport à η. Cette décomposition est unique modulo des ensembles de charge nulle : pour toute autre décomposition (H'_+, H'_-), $\eta(H_+ \triangle H'_+) = \eta(H_- \triangle H'_-) = 0$.*

DÉMONSTRATION. Soit $\mathcal{P}_+$ la classe des ensembles positifs [1]. On a $\mathcal{P}_+ \neq \varnothing$ puisque $\varnothing$ est positif. Soit $\alpha := \sup\{\eta(A) : A \in \mathcal{P}_+\}$. On a $0 \leq \alpha < \infty$ (exercice). Soit E_n une suite d'ensembles positifs telle que $\eta(E_n) \nearrow \alpha$. On peut supposer que E_n est croissante (sinon, on peut toujours redéfinir $E'_n := E_1 \cup \dots \cup E_n$, qui est positive et croissante). Par le Lemme 7.7, l'ensemble

$$H_+ := \bigcup_n E_n$$

est positif. Il reste à vérifier que $H_- := X \backslash H_+$ est négatif.

Supposons que H_- n'est pas négatif. Il doit alors contenir un sous-ensemble mesurable E^* tel que $\eta(E^*) > 0$. Si E^* était positif, alors $\widetilde{H}^+ := H_+ \cup E^*$ serait positif aussi, et $\eta(\widetilde{H}^+) > \alpha$, une contradiction. Donc E^* contient des sous-ensembles de charge *strictement* négative. En particulier, il existe un plus petit nombre entier $n_1 \geq 1$ et un sous

1. Dans cette preuve, la positivité/négativité sera toujours par rapport à η.

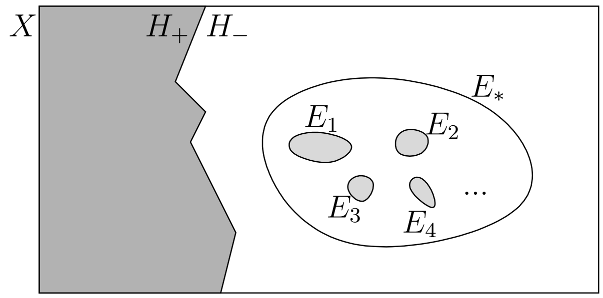


FIGURE 1. Une “vue d’artiste” de la décomposition de Hahn. Les ensembles E_i sont de charge négative.

ensemble $E_1 \subset E^*$ tel que $\eta(E_1) \leq -1/n_1 < 0$. Ensuite, on peut recommencer le même raisonnement avec l’ensemble $E^*\backslash E_1$: il a charge $\eta(E^*\backslash E_1) = \eta(E^*) - \eta(E_1) > \eta(E^*) > 0$, donc ne peut pas être positif, donc il doit exister un plus petit nombre entier $n_2 \geq 1$ et un $E_2 \subset E^*\backslash E_1$ tel que $\eta(E_2) \leq -1/n_2 < 0$. En itérant cet argument, on obtient une suite disjointe E_k de sous ensembles de E^* de charges négatives $\eta(E_k) \leq -1/n_k < 0$. La suite n_k a la propriété suivante : si $n < n_k$, alors il n’existe aucun sous-ensemble $G \subset E^*\backslash(E_1 \cup \cdots \cup E_k)$ tel que $\eta(G) \leq -1/n$. Définissons

$$F := \bigcup_{k\geq 1} E_k \,.$$

Puisque $-\infty < \eta(F) = \sum_k \eta(E_k) \leq -\sum_{k\geq 1} 1/n_k < 0$, ceci implique $n_k \to \infty$.

Montrons maintenant que $E^*\backslash F$ est positif. En effet, s’il était négatif, il existerait $G \subset E^*\backslash F$ avec $\eta(G) < 0$. Mais puisque $n_k \to \infty$, il existe en particulier un k, suffisamment grand, tel que $\eta(G) \leq -1/(n_k - 1)$, en contradiction avec la définition de n_k.

Donc $E^*\backslash F$ est positif. Par conséquent, $H_+ \cup (E^*\backslash F)$ est aussi positif. Mais

$$\eta(H_+ \cup (E^*\backslash F)) = \eta(H_+) + \eta(E^*) - \eta(F) \geq \eta(H_+) + \eta(E^*) > \alpha \,,$$

une contradiction avec la définition de α. Ceci montre que H_- ne peut contenir d’ensembles de charge positive. Donc H_- est négatif, et (H_+, H_-) est bien une décomposition de Hahn.

Pour montrer l’unicité, considérons une autre décomposition (H'_+, H'_-). Alors $H_+\backslash H'_+ \subset H_+$ et $H_+\backslash H'_+ \subset H'_-$. Donc $H_+\backslash H'_+$ est à la fois positif et négatif. On montre ainsi que $\eta(H_+ \triangle H'_+) = \eta(H_- \triangle H'_-) = 0$. $\square$

THÉORÈME 7.10. *Soit η une charge, et soit (H_+, H_-) une décomposition de Hahn de X par rapport à η. Alors les mesures finies*

$$\eta_+(A) := \eta(A \cap H_+)\,, \quad \eta_-(A) := -\eta(A \cap H_-)\,, \quad A \in \mathcal{F} \tag{7.4}$$

sont les uniques mesures sur $(X, \mathcal{F})$ telles que $\eta = \eta_+ - \eta_-$, $\eta_+ \perp \eta_-$.

DÉMONSTRATION. Les propriétés des mesures $\eta_\pm$ découlent de la preuve précédente. Soient μ_+, μ_- deux autres mesures singulières telles que $\eta = \mu_+ - \mu_-$. Soient alors $E, F \in \mathcal{F}$ disjoints tels que $E \cup F = X$, $\mu_+(F) = \mu_-(E) = 0$. Alors (E, F) est une autre décomposition de Hahn de X par rapport à η, et donc pour tout $A \in \mathcal{F}$,

$$\mu_+(A) = \mu_+(A \cap F) = \eta(A \cap F) = \eta(A \cap H_+) = \eta_+(A \cap H_+) = \eta_+(A)\,,$$

donc $\mu_+ = \eta_+$. De même, on montre que $\mu_- = \eta_-$. □

DÉFINITION 7.11. *Les mesures η_+ et η_- construites dans le Théorème 7.10 sont appelées respectivement* variation positive *et* variation négative de *η. La décomposition de η en $\eta_+ - \eta_-$ s'appelle* décomposition de Jordan de *η. La mesure $|\eta| := \eta_+ + \eta_-$ est la* variation totale de *η.*

L'usage que l'on fera des charges dans la preuve du Théorème de Radon-Nikodým est contenue dans le résultat suivant :

COROLLAIRE 7.12. *Si ν et μ sont des mesures finies sur $(X, \mathcal{F})$, alors soit $\nu \perp \mu$ (ce qui n'exclut pas que l'une d'elles soit triviale), soit il existe $\epsilon > 0$ et un ensemble $E \in \mathcal{F}$ tels que $\mu(E) > 0$ et $\nu - \epsilon\mu$ est positive sur E.*

DÉMONSTRATION. Soit $(H_+(n), H_-(n))$ une décomposition de Hahn par rapport à $\eta_n := \nu - \frac{1}{n}\mu$, et soit $H_+ := \bigcup_n H_+(n)$, $H_- := \bigcap_n H_-(n) = H_+^c$. Alors H_- est négatif par rapport à chaque η_n, et donc $0 \leq \nu(H_-) \leq \frac{1}{n}\mu(H_-)$ pour tout n. Par conséquent, $\nu(H_-) = 0$ puisque $\mu(H_-) < \infty$. Si $\mu(H_+) = 0$ alors $\nu \perp \mu$. Si $\mu(H_+) > 0$, alors $\mu(H_+(n_0)) > 0$ pour au moins un n_0. Aussi, $H_+(n_0)$ est positif par rapport à $\eta_{n_0} = \nu - \frac{1}{n_0}\mu$. On peut donc prendre $\epsilon := \frac{1}{n_0}$ et $E := H_+(n_0)$. □

PREUVE DU THÉORÈME 7.2. Soit $\nu \ll \mu$. Supposons d'abord que μ et ν sont finies. Posons

$$\mathcal{K} := \left\{ g \in \mathcal{M}^+ : \int_A g d\mu \leq \nu(A)\,, \forall A \in \mathcal{F} \right\},$$

qui n'est pas vide puisque la fonction identiquement nulle $f \equiv 0$ satisfait à la condition. Soit ensuite

$$\kappa := \sup \left\{ \int g d\mu : g \in \mathcal{K} \right\}.$$

Puisque $\int g d\mu \leq \nu(X) < \infty$ pour tout $g \in \mathcal{K}$, on a $0 \leq \kappa < \infty$. Soit $g_n \in \mathcal{K}$ telle que $\lim_n \int g_n d\mu = \kappa$. Soit $f_n := \sup\{g_1, \dots, g_n\}$. Alors f_n est non-décroissante. De plus, $f_n \in \mathcal{K}$ pour tout n. En effet, pour $n = 1$, $f_1 = g_1 \in \mathcal{K}$. Pour $n = 2$,

$$\begin{aligned}\int_A f_2 d\mu &= \int_{A\cap\{g_1 \geq g_2\}} g_1 d\mu + \int_{A\cap\{g_1 < g_2\}} g_2 d\mu \\ &\leq \nu(A \cap \{g_1 \geq g_2\}) + \nu(A \cap \{g_1 < g_2\}) = \nu(A)\,,\end{aligned}$$

donc $f_2 \in \mathcal{K}$. Par induction, $f_n \in \mathcal{K}$. Soit $f := \lim_{n\to\infty} \uparrow f_n$. Alors f_n est mesurable, et par convergence monotone,

$$\int_A f d\mu = \lim_{n\to\infty} \int_A f_n d\mu\,.$$

Comme c'est vrai pour tout $A \in \mathcal{F}$ on a $f \in \mathcal{K}$. De plus, comme $f_n \geq g_n$, on a $\int f d\mu = \kappa$.

Soit $\tilde{\mu} := \nu - f\mu$, qui est une mesure (finie) puisque $f \in \mathcal{K}$. On va montrer que $\tilde{\mu} = 0$.

Puisque $\nu \ll \mu$, on a aussi $\tilde{\mu} \ll \mu$. Donc à moins d'être identiquement nulle, $\tilde{\mu}$ ne peut pas être singulière par rapport à μ. Si elle n'est pas identiquement nulle, alors par le Corollaire 7.12, il existe $\epsilon > 0$ et $E \in \mathcal{F}$ tels que $\mu(E) > 0$, et tels que $\tilde{\mu} - \epsilon\mu$ est positive sur E. On peut ensuite calculer, pour tout $A \in \mathcal{F}$,

$$\begin{aligned}\int_A (f + \epsilon 1_E) d\mu &= \int_A f d\mu + \epsilon\mu(E \cap A) \\ &\leq \int_A f d\mu + \tilde{\mu}(E \cap A) \\ &= \int_{E^c\cap A} f d\mu + \nu(E \cap A) \\ \text{(puisque } f \in \mathcal{K}) &\leq \nu(E^c \cap A) + \nu(E \cap A) = \nu(A)\,,\end{aligned}$$

ce qui montre que $f + \epsilon 1_E \in \mathcal{K}$. Mais comme

$$\int (f + \epsilon 1_E) d\mu = \kappa + \epsilon\mu(E) > \kappa\,,$$

c'est une contradiction. Ceci montre que $\tilde{\mu} = 0$, et donc prouve que $\nu = f\mu$. Pour voir que la densité f est unique, supposons que

$$\int_A f d\mu = \int_A g d\mu \quad \forall A \in \mathcal{F}\,.$$

En appliquant cette identité une fois avec $A = \{f \geq g\}$, une fois avec $A = \{f < g\}$, on obtient que $f = g$ μ-presque partout. Ceci montre le théorème lorsque les deux mesures sont finies.

Si ν et μ sont σ-finies, il existe $X_n, Y_n \in \mathcal{F}$, $X_n \nearrow X$, $Y_n \nearrow X$, tels que $\mu(X_n) < \infty$, $\nu(Y_n) < \infty$. Soit $Z_n := X_n \cap Y_n$, qui satisfait $Z_n \nearrow X$, et $\mu(Z_n) < \infty$, $\nu(Z_n) < \infty$. Considérons $W_1 := Z_1$, $W_2 := Z_2 \cap W_1^c$, et pour tout $n \geq 2$, $W_n := Z_n \cap (W_1 \cup \cdots \cup W_{n-1})^c$. La suite $(W_n)_{n\geq 1}$ forme une partition de X. De plus, les mesures $\mu_n(\cdot) := \mu(\cdot \cap W_n)$, $\nu_n(\cdot) := \nu(\cdot \cap W_n)$ sont finies, et $\nu_n \ll \mu_n$. On peut donc appliquer la première partie de la preuve : il existe $f_n \in \mathcal{M}^+$ telle que $\nu_n = f_n \mu_n$. Soit $f := \sum_{n\geq 1} f_n \in \mathcal{M}^+$. On a alors, pour tout $A \in \mathcal{F}$,

$$\nu(A) = \sum_{n\geq 1} \nu_n(A) = \sum_{n\geq 1} \int_{A\cap W_n} f_n d\mu = \int_A f d\mu \,.$$

Comme avant, on peut montrer que cette densité f est unique modulo des ensembles de mesure nulle. Ceci termine la preuve du Théorème de Radon-Nikodým. $\square$

La densité $f = \frac{d\nu}{d\mu}$ possède quelques propriétés qui rappellent celles de la dérivée usuelle. Par exemple :

LEMME 7.13. *Soit ν une mesure σ-finie, et μ, θ deux mesures σ-finies telles quel $\nu \ll \mu$, $\mu \ll \theta$. Alors $\nu \ll \theta$, et*

$$\frac{d\nu}{d\theta} = \frac{d\nu}{d\mu}\frac{d\mu}{d\theta} \,.$$

2. La décomposition de Lebesgue

Une première conséquence importante du Théorème de Radon-Nikodým :

THÉORÈME 7.14 (Décomposition de Lebesgue). *Soit μ une mesure σ-finie. Alors pour toute mesure σ-finie ν, il existe deux uniques mesures ν_s et ν_a, telles que*

$$\nu = \nu_s + \nu_a \,, \quad \nu_s \perp \mu \,, \quad \nu_a \ll \mu \,. \tag{7.5}$$

*ν_s est la **composante singulière de** ν, et ν_a est la **composante absolument continue de** ν (par rapport à μ).*

DÉMONSTRATION. Soit $\rho := \nu + \mu$, qui est σ-finie. Alors $\nu \ll \rho$ et $\mu \ll \rho$. Donc par le Théorème 7.2, il existe une densité de Radon-Nikodým $g = \frac{d\mu}{d\rho} \in \mathcal{M}^+$ telle que pour tout $A \in \mathcal{F}$,

$$\mu(A) = \int_A g d\rho \,.$$

Considérons les ensembles disjoints $U := \{g = 0\}$, $V = \{g > 0\}$. On a $X = U \cup V$. Soit

$$\nu_s(A) := \nu(A \cap U), \quad \nu_a(A) := \nu(A \cap V).$$

Puisque $\mu(U) = 0$, on a $\nu_s \perp \mu$. Vérifions ensuite que $\nu_a \ll \mu$. En effet, si $\mu(A) = 0$, alors $\int_A g d\rho = 0$, et donc $g(x) = 0$ pour ρ-presque tout $x \in A$. En particulier, $\rho(A \cap V) = 0$, donc aussi $\nu(A \cap V) = 0$ puisque $\nu \ll \rho$, et donc $\nu_a(A) = \nu(A \cap V) = 0$.

Vérifions l'unicité de la décomposition. Supposons qu'il existe deux décompositions de ν, $\nu_s + \nu_a$ et $\nu_s' + \nu_a'$. Donc il existe A tel que $\mu(A^c) = 0$, $\nu_s(A) = 0$, et il existe A' tel que $\mu(A'^c) = 0$ et $\nu_s(A') = 0$. On a donc, d'une part, que $\mu((A \cap A')^c) = 0$, ce qui implique $\nu_a((A \cap A')^c) = \nu_a'((A \cap A')^c) = 0$. D'autre part, ceci implique $\nu_s(A \cap A') = 0$ et $\nu_s'(A \cap A') = 0$. Soit alors E un mesurable tel que $\nu(E) < \infty$. On peut écrire

$$\begin{aligned}\nu_a(E) &= \nu_a(E \cap (A \cap A')) + \underbrace{\nu_a(E \cap (A \cap A')^c)}_{=0} \\ &= \nu(E \cap (A \cap A')) - \underbrace{\nu_s(E \cap (A \cap A'))}_{=0}.\end{aligned}$$

On montre de même que $\nu_a'(E) = \nu(E \cap (A \cap A'))$, et donc $\nu_a(E) = \nu_a'(E)$. Ceci implique donc $\nu_s(E) = \nu_s'(E)$. Si E est quelconque, il suit de ce qui précède et de la σ-finitude de ν que $\nu_a(E) = \nu_a'(E)$. On a donc montré que $\nu_a = \nu_a'$ et $\nu_s = \nu_s'$. □

3. Parenthèse : l'espérance conditionnelle

Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilités. Les ensembles de la tribu $\mathcal{F}$ sont appelés *événements* plutôt que *ensembles mesurables*.

3.1. À propos des tribus. Travailler avec une tribu $\mathcal{F}$ signifie que pour chaque événement $A \in \mathcal{F}$, et pour chaque réalisation $\omega \in \Omega$ de l'expérience, on peut déterminer si oui ou non ω appartient à A (en d'autres termes, on peut calculer $1_A(\omega)$). Par exemple, considérons

$$\Omega = \{\text{piétons dans les rues de Genève}\},$$

c'est-à-dire que l'expérience consiste à tirer au hasard un piéton dans les rues de Genève. Une tribu possible est

$$\mathcal{F}_1 = \{\varnothing, \Omega, F, M\},$$

où $F = \{\text{piétons de sexe féminin}\}$, $M = \{\text{piétons de sexe masculin}\}$. Cette tribu signifie que l'unique distinction que l'on sera capable de faire après avoir tiré un piéton au hasard, est de déterminer si c'est une femme ou un homme.

Pour la même expérience, une autre tribu possible est

$$\mathcal{F}_2 = \{\varnothing, \Omega, J, V\},$$

où J est l'ensemble des piétons de moins de 60 ans, et V celui des piétons de 60 ans ou plus. On pourrait aussi considérer la tribu qui permet de distinguer le sexe *et* l'âge de la personne, en utilisant une tribu plus *fine* que $\mathcal{F}_1$ et $\mathcal{F}_2$:

$$\mathcal{F}_3 = \mathcal{F}_1 \vee \mathcal{F}_2 := \{\varnothing, \Omega, F \cap J, F \cap V, M \cap J, M \cap V\}.$$

Le choix d'une tribu reflète donc l'*information* révélée par la réalisation d'une expérience aléatoire.

La tribu contenant le moins d'information est la tribu triviale $\{\varnothing, \Omega\}$: le résultat de l'expérience ω ne fournit aucune information. La tribu la plus riche en information est la tribu discrète $\mathcal{P}(\Omega)$. Puisque $\mathcal{P}(\Omega)$ contient aussi les singletons $\{\omega\}$, chaque réalisation permet d'identifier *exactement* quel élément ω a été choisi. Dans l'exemple ci-dessus, cela signifie identifier exactement *quelle* est la/le Genevois choisi (et non seulement son sexe et la tranche d'âge dans laquelle il se trouve).

3.2. Conditionner par rapport à des tribus finies ou dénombrables. Supposons donc qu'une tribu $\mathcal{F}$ soit donnée. Si P est une probabilité sur $(\Omega, \mathcal{F})$, alors $P(A)$ quantifie avec quelle probabilité l'élément ω tiré appartient à $A \in \mathcal{F}$.

Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire intégrable. Pour chaque réalisation ω, $X(\omega)$ est un nombre décrivant une certaine propriété de ω, et son espérance s'écrit

$$E[X] = \int X dP.$$

Dans un certain sens, $E[X]$ est une prédiction raisonnable que l'on peut faire à propos de la valeur de $X(\omega)$ si l'on ne possède aucune information sur ω. En effet, si l'on moyenne les résultats obtenus en répétant l'expérience de manière indépendante, la Loi Forte des Grands Nombres garantit que

$$\frac{X_1 + \cdots + X_n}{n} \to E[X] \quad P - p.s.$$

Se pose donc la question suivante : si l'on sait que le ω tiré appartient à un ensemble B, quelle est la meilleure prédiction possible que l'on peut faire sur $X(\omega)$?

Si l'on sait que $\omega \in B$, alors la probabilité avec laquelle on décrit l'expérience change. Du cours de probabilités, on sait qu' un événement $B \in \mathcal{F}$ de probabilité positive permet de définir une nouvelle mesure $P(\cdot|B)$ sur $(\Omega, \mathcal{F})$, conditionnée par rapport à l'événement B :

$$P(A|B) := \frac{P(A \cap B)}{P(B)} . \tag{7.6}$$

Donc sachant que B a eu lieu, la meilleure prédiction sur la valeur de X est alors

$$E[X|B] := \int X(\omega) P(d\omega|B) = \frac{1}{P(B)} \int_B X \, dP . \tag{7.7}$$

Reprenons l'exemple où l'on tire un piéton au hasard dans les rues de Genève, muni de la tribu discrète. Soit $X(\omega)$ la taille (en centimètres) de l'individu choisi lors de l'expérience ω. Suposons que l'on ait

$$E[X] = 168 \, cm .$$

Si maintenant l'on sait en plus si l'individu est une femme ou un homme, alors on aura peut-être

$$E[X|F] = 163 \, cm , \quad E[X|H] = 172 \, cm, .$$

On peut reconstruire $E[X]$ à partir de $E[X|F]$ et $E[X|H]$, en calculant

$$\begin{aligned} E[X|F]P(F) + E[X|H]P(H) &= \int_F X dP + \int_H X dP \\ &= E[X] . \end{aligned}$$

Donc la paire $(E[X|F], E[X|H])$ contient plus d'information, et il est naturel de définir une nouvelle variable aléatoire

$$E[X|\mathcal{F}_1](\omega) := E[X|F] 1_F(\omega) + E[X|H] 1_H(\omega) ,$$

appelée espérance conditionnelle de X par rapport à la tribu $\mathcal{F}_1$.

On peut être un peu plus général. Soit $\{B_n\}_{n \geq 1}$ une partition dénombrable de Ω, c'est-à-dire une famille $B_n \in \mathcal{F}$ telle que $\bigcup_n B_n = \Omega$, et $B_n \cap B_m = \emptyset$ lorsque $n \neq m$. Une telle partition génère une sous tribu $\mathcal{B} := \sigma(\{B_n\}_{n \geq 1}) \subset \mathcal{F}$, contenant toutes les unions d'ensembles B_n. Soit X une variable aléatoire intégrable. Si $P(B_n) > 0$, on définit $E[X|B_n]$ comme plus haut. Si $P(B_n) = 0$, on définit $E[X|B_n]$

de façon arbitraire, par exemple en posant $E[X|B_n] := 0$. On définit $E[X|\mathcal{B}] : \Omega \to \mathbb{R}$ par

$$E[X|\mathcal{B}](\omega) := \sum_{n\geq 1} E[X|B_n] 1_{B_n}(\omega)\,, \tag{7.8}$$

appelée version de l'espérance conditionnelle de X par rapport à $\mathcal{B}$. Le nom "version" est utilisé pour souligner le fait que l'on peut toujours modifier $E[X|\mathcal{B}]$ sur des ensembles de mesure nulle.

Clairement, $\omega \mapsto E[X|\mathcal{B}](\omega)$ est mesurable par rapport à la tribu $\mathcal{B}$. De plus, pour tout $B \in \mathcal{B}$,

$$\begin{aligned}\int_B E[X|\mathcal{B}]\, dP &= \sum_{n\geq 1} \int_{B\cap B_n} E[X|B_n]\, dP \\ &= \sum_{n\geq 1} P(B\cap B_n) E[X|B_n]\,. \end{aligned} \tag{7.9}$$

Si $B \in \mathcal{B}$ alors B est une union d'éléments B_k, $k \in S$, ce qui implique $P(B \cap B_n) = P(B_n)1_S(n)$, où $1_S(n) = 1$ si $n \in S$, 0 sinon. En utilisant la définition de $E[X|B_n]$, (7.9) vaut

$$\sum_{n\geq 1} 1_S(n) \int_{B_n} X\, dP = \int_B X\, dP\,.$$

Donc la variable $E[X|\mathcal{B}]$ satisfait

$$\int_B E[X|\mathcal{B}]\, dP = \int_B X\, dP\,, \quad \forall B \in \mathcal{B}\,. \tag{7.10}$$

Bien-sûr, toute version de $E[X|\mathcal{B}]$ satisfait (7.10). Le résultat suivant montre que les versions de $E[X|\mathcal{B}]$ sont les seules ayant cette propriété.

LEMME 7.15. *Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et $\mathcal{B} \subset \mathcal{F}$ une sous-tribu. Soient Y_1, Y_2 deux variables mesurables par rapport à $\mathcal{B}$, intégrables par rapport à P. Alors $Y_1 = Y_2$ presque partout si et seulement si*

$$\int_B Y_1\, dP = \int_B Y_2\, dP \quad \forall B \in \mathcal{B}\,. \tag{7.11}$$

DÉMONSTRATION. Suit du Théorème 3.18. □

3.3. Conditionner par rapport à une sous-tribu. Dans le cas d'une tribu $\mathcal{B}$ générée par une partition dénombrable, l'espérance conditionnelle est définie comme en (7.8). On a vu que ses principales propriétés sont

(1) $E[X|\mathcal{B}]$ est mesurable par rapport à $\mathcal{B}$.

(2) La famille de relations (7.10) est satisfaite.

On a aussi vu dans le Lemme 7.15 que $E[X|\mathcal{B}]$ est unique, modulo des modifications de mesure nulle.

Pour étendre l'espérance conditionnelle à des sous-tribus quelconques $\mathcal{G} \subset \mathcal{F}$, le manque de dénombrabilité empêche une définition directe comme en (7.8). En particulier se pose le problème des ensembles $B \in \mathcal{G}$ de mesure nulle, pour lesquels (7.8) ne peut pas être définie pour définir $E[X|\mathcal{G}]$. Malgré tout, on peut utiliser les deux propriétés ci-dessus pour *définir* $E[X|\mathcal{G}]$, et ensuite essayer de montrer son existence et son unicité. Finalement, on peut vérifier la compatibilité de cette définition avec celle obtenue naturellement dans le cas dénombrable (7.8).

DÉFINITION 7.16. *Soit $\mathcal{G} \subset \mathcal{F}$ une sous-tribu et $X : \Omega \to \mathbb{R}$ une variable aléatoire non-négative. Toute variable aléatoire $\mathcal{G}$-measurable $Y : \Omega \to \mathbb{R}$ qui satisfait*

$$\int_B Y\,dP = \int_B X\,dP\,, \quad \forall B \in \mathcal{G}\,, \tag{7.12}$$

est appelée **version de l'espérance conditionnelle de X par rapport à $\mathcal{G}$**, *et généralement notée $E[X|\mathcal{G}]$. Pour une variable intégrable quelconque X, on définit l'espérance conditionnelle comme : $E[X|\mathcal{G}] := E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}]$.*

La principale propriété de l'espérance conditionnelle est donc que

$$\int_B E[X|\mathcal{G}]\,dP = \int_B X\,dP\,, \quad \forall B \in \mathcal{G}\,, \tag{7.13}$$

Observons que quand elle existe, $E[X|\mathcal{G}]$ est intégrable. En effet, en appliquant (7.13) avec $B = \Omega$,

$$E[E[X|\mathcal{G}]] = \int E[X|\mathcal{G}]\,dP = \int X\,dP = E[X]\,. \tag{7.14}$$

Vérifions maintenant que l'espérance conditionnelle existe. Définissons, pour tout $B \in \mathcal{G}$,

$$Q(B) := \int_B X\,dP\,.$$

Comme on suppose que $X \geq 0$ est intégrable, Q est une mesure finie sur $(\Omega, \mathcal{G})$. De plus, elle est absolument continue par rapport à P *sur* $\mathcal{G}$: si $G \in \mathcal{G}$ est tel que $P(G) = 0$, alors $Q(G) = 0$. Par le Théorème de Radon-Nikodým, il existe une variable aléatoire non-négative $\mathcal{G}$-mesurable, Y, telle que

$$Q(B) = \int_B Y\,dP\,, \quad \forall B \in \mathcal{G}\,.$$

Donc Y est une version de l'espérance conditionnelle $E[X|\mathcal{G}]$. On voit que $E[X|\mathcal{G}]$ n'est rien d'autre que la dérivée de Radon-Nikodým $\frac{dQ}{dP}$, et est unique à des ensembles de mesure nulle près (Lemma 7.15).

Exemple 7.17. Vérifions que la définition abstraite coïncide avec celle de la section précédente pour les partitions dénombrables. Soit $\mathcal{B}$ la tribu générée par une partition $\{B_n\}_{n\geq 1}$, et soit $E[X|\mathcal{B}]$ une version de l'espérance conditionnelle au sens de la définition 7.16. Comme $E[X|\mathcal{B}]$ est mesurable par rapport à $\mathcal{B}$, elle est constante sur chaque B_n, c'est-à-dire $E[X|\mathcal{B}](\omega) = b_n$ pour tout $\omega \in B_n$. Donc pour tout n on a

$$b_n P(B_n) = b_n \int_{B_n} dP = \int_{B_n} E[X|\mathcal{B}]\,dP = \int_{B_n} X\,dP\,.$$

Donc, lorsque $P(B_n) > 0$,

$$b_n = \frac{1}{P(B_n)} \int_{B_n} X\,dP\,,$$

qui coïncide avec (7.8) sur les ensembles B_n de mesure positive. Comme les autres ensembles ont mesure nulle et qu'on n'a qu'à considérer un nombre fini d'entre eux, on a bien construit une version de (7.8).

Mentionnons les deux cas extrêmes qui permettent de mieux comprendre la dépendance de $E[X|\mathcal{G}]$ en fonction de $\mathcal{G}$. D'abord, si $\mathcal{G}$ est la plus grande sous-tribu de $\mathcal{F}$, i.e. $\mathcal{G} = \mathcal{F}$, alors clairement $E[X|\mathcal{G}] = X$ presque partout. Ceci vient du fait que X est mesurable par rapport à $\mathcal{F}$, et de l'identité triviale

$$\int_B X\,dP = \int_B X\,dP\,, \quad \forall B \in \mathcal{F}\,.$$

D'autre part, si $\mathcal{G}$ est la tribu triviale, $\mathcal{G} = \{\emptyset, \Omega\}$, alors $E[X|\mathcal{G}] = E[X]$ presque partout. En effet, $E[X|\mathcal{G}]$ doit être constante sur Ω, et cette constante est fixée par la condition (prendre $B = \Omega$ dans (7.13))

$$E[X|\mathcal{G}] = \int E[X|\mathcal{G}]\,dP = \int X\,dP = E[X]\,.$$

Ces deux cas montrent comme $E[X|\mathcal{G}]$ est une approximation de X ; plus $\mathcal{G}$ est fine, plus cette approximation est précise.

L'espérance conditionnelle est un outil de base en théorie des probabilités. Il permet notamment de définir la notion de *martingale adaptée à une filtration* (voir le cours de probabilité).

4. Dualité $L^p - L^q$

Une application importante du Théorème de Radon-Nikodým en analyse fonctionnelle est à propos de la *dualité* entre les espaces L^p et L^q, où p et q sont conjugués.

DÉFINITION 7.18. *Une application $T : L^p \to \mathbb{R}$ est une fonctionnelle linéaire continue si*

(1) $T(f + \alpha g) = T(f) + \alpha T(g)$ pour tout $\alpha \in \mathbb{R}$, $f, g \in L^p$.

(2) $T(f_n) \to T(f)$ lorsque $f_n \stackrel{L^p}{\to} f$.

Si $T : L^p \to \mathbb{R}$, on définit la norme

$$\|T\| := \sup_{\substack{f\in L^p\\ f\neq 0}} \frac{|T(f)|}{\|f\|_p} = \sup_{\substack{f\in L^p\\ \|f\|_p\leq 1}} |T(f)| .$$

On a toujours $|T(f)| \leq \|T\| \cdot \|f\|_p$, et on peut montrer que T est continue si et seulement si $\|T\| < \infty$[2].

On a déjà vu une construction de fonctionnelle linéaire continue dans l'Exercice 42 : si $g \in L^q$, alors

$$T(f) := \int fg d\mu$$

est linéaire et continue, et $\|T\| = \|g\|_q$. Le Théorème de Dualité affirme que toute fonctionelle linaire continue est de cette forme :

THÉORÈME 7.19 (Dualité de Riesz des espaces L^p et L^q). *Soit μ une mesure σ-finie, $1 \leq p < \infty$. Soit q l'exposant conjugué à p (avec la convention $1 \leftrightarrow \infty$). Alors pour toute fonctionnelle linéaire continue $T : L^p \to \mathbb{R}$, il existe $g \in L^q$ telle que*

$$T(f) = \int fg d\mu \quad \forall f \in L^p . \tag{7.15}$$

De plus, $\|T\| = \|g\|_p$.

---

2. En effet, si T est continue, elle est en particulier continue en 0, et donc pour $\epsilon = 1$, il existe un $\delta > 0$ tel que si $\|f\|_p \leq \delta$, alors $|T(f)| \leq 1$. Soit $f \in L^p$ tel que $f \neq 0$. On peut écrire $|T(f)| = |T(\frac{f\delta}{\|f\|_p})|\frac{\|f\|_p}{\delta} \leq \frac{\|f\|_p}{\delta}$. Donc $\|T\| \leq \delta^{-1}$. Inversément, si $\|T\| < \infty$ et $f_n \stackrel{L^p}{\to} f$, alors $|T(f_n) - T(f)| = |T(f_n - f)| \leq \|T\| \cdot \|f_n - f\|_p \to 0$, donc T est continue.

On ne donnera pas la preuve de ce résultat, que le lecteur trouvera dans plusieurs des références données dans l'introduction.

L'espace des fonctionnelles linéaires continues sur un espace vectoriel est appelé le dual de l'espace. Le résultat ci-dessus affirme donc que le dual de L^p peut être identifié à L^q (où $q \leftrightarrow p$). Dans le cas particulier où $p = q = 2$, on dit que L^2 est self-dual.

5. Le Théorème de dérivation de Lebesgue

Le Théorème de Radon-Nykodým est un résultat d'existence : si ν est absolument continue par rapport à μ, alors *il existe* une densité $f = \frac{d\nu}{d\mu}$. Dans cette section, on considère le cas particulier où l'espace mesurable est $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, et la mesure de référence μ est la mesure de Lebesgue λ^d. On considérera une mesure σ-finie, $\nu \ll \lambda^d$, et on montrera comment sa densité $\frac{d\nu}{d\lambda^d}$ peut être *calculée* par un processus naturel de limite, mentionné dans l'introduction du chapitre.

5.1. Dérivabilité.

Rappelons que $B_r(x) = \{z : \|z - x\| < r\}$.

DÉFINITION 7.20. *Soit μ une mesure finie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. On dit que μ est* dérivable *au point $x \in \mathbb{R}^d$ si la limite*

$$D\mu(x) := \lim_{\epsilon \to 0^+} \frac{\mu(B_\epsilon(x))}{\lambda^d(B_\epsilon(x))} \tag{7.16}$$

existe. On appellera $D\mu(x)$ la dérivée de μ par rapport à λ^d au point x.

EXEMPLE 7.21. Si $\mu(B) := \lambda^d(B \cap [-1,1]^d)$, alors μ est dérivable en tout $x \in (-1,1)^d$ et en ces points $D\mu(x) = 1$, ainsi que sur les points de $([-1,1]^d)^c$, où $D\mu(x) = 0$.

Notre objectif est de comparer $D\mu$ avec la dérivée de Radon-Nikodým $\frac{d\mu}{d\lambda^d}$. Dans le reste de la section, la notion de "presque partout" sera toujours en référence à la mesure de Lebesgue λ^d.

THÉORÈME 7.22. *Soit μ une mesure finie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.*

(1) Si $A \in \mathcal{B}(\mathbb{R}^d)$, $\mu(A) = 0$, alors $D\mu(x) = 0$ pour presque tout $x \in A$.

(2) Si $\mu \ll \lambda^d$, alors $D\mu(x)$ existe presque partout, et

$$D\mu(x) = \frac{d\mu}{d\lambda^d} \quad \textit{presque partout}.$$

(3) $D\mu(x)$ existe presque partout.

Pour étudier l'ensemble des points où μ est dérivable, on introduit

$$\overline{D}\mu(x) := \limsup_{\epsilon\to 0^+} \frac{\mu(B_\epsilon(x))}{\lambda^d(B_\epsilon(x))}, \quad \underline{D}\mu(x) := \liminf_{\epsilon\to 0^+} \frac{\mu(B_\epsilon(x))}{\lambda^d(B_\epsilon(x))}.$$

Par définition, μ est dérivable en x si et seulement si $\overline{D}\mu(x) = \underline{D}\mu(x) < \infty$. De plus,

LEMME 7.23. *$x \mapsto \overline{D}\mu(x)$ et $x \mapsto \underline{D}\mu(x)$ sont mesurables.*

DÉMONSTRATION. Pour montrer que $x \mapsto \overline{D}\mu(x)$ est mesurable, il suffit de montrer que $x \mapsto \phi_\delta(x)$ est mesurable, où $\phi_\delta(x) := \sup_{0<r<\delta} \frac{\mu(B_r(x))}{\lambda^d(B_r(x))}$. Fixons $\alpha > 0$ et montrons que $E := \{x : \phi_\delta(x) > \alpha\}$ est ouvert. Soit $x \in E$. Alors il existe $0 < r < \delta$ tel que $\frac{\mu(B_r(x))}{\lambda^d(B_r(x))} > \alpha$. Soit $\eta > 0$ suffisamment petit, tel que $r + \eta < \delta$. Alors $B_{\eta+r}(y) \supset B_r(x)$ $\forall y \in B_\eta(x)$, et donc

$$\frac{\mu(B_{\eta+r}(y))}{\lambda^d(B_{\eta+r}(y))} \geq \frac{\mu(B_r(x))}{\lambda^d(B_{\eta+r}(y))} = \frac{\mu(B_r(x))}{\lambda^d(B_r(x))}\frac{\lambda^d(B_r(0))}{\lambda^d(B_{\eta+r}(0))}.$$

Comme $\lambda^d(B_r(0)) = c_d r^d$ pour une constante $c_d > 0$ (exercice), on a

$$\frac{\lambda^d(B_r(0))}{\lambda^d(B_{\eta+r}(0))} = \left(1 + \frac{\eta}{r}\right)^{-d},$$

dont la limite est 1 lorsque $\eta \to 0$. On peut donc prendre $\eta > 0$ suffisamment petit tel que $\frac{\mu(B_r(x))}{\lambda^d(B_r(x))}\left(1 + \frac{\eta}{r}\right)^{-d} > \alpha$, et donc

$$\phi_\delta(y) \geq \frac{\mu(B_{\eta+r}(y))}{\lambda^d(B_{\eta+r}(y))} > \alpha,$$

ce qui implique $y \in E$. Donc E est ouvert, et comme les ouverts génèrent les Boréliens, ceci montre que ϕ_δ est mesurable. $\square$

Rappelons (voir Proposition 2.39) qu'une mesure finie sur $\mathbb{R}^d$ est régulière.

PROPOSITION 7.24. *Si μ est finie (et donc régulière), alors pour tout $\alpha > 0$ et pour tout Borélien A,*

(1) $\lambda^d(\{x \in A : \overline{D}\mu(x) > \alpha\}) \leq \frac{3^d}{\alpha}\mu(A)$.

(2) $\lambda^d(\{x \in A : \overline{D}\mu(x) > 0\}) = 0$ si $\mu(A) = 0$.

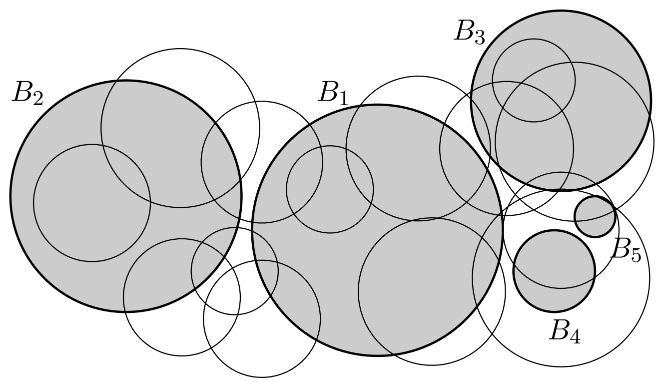
La deuxième partie de la proposition dit que sur un ensemble de mesure nulle par rapport à μ, $D\mu = 0$ λ^d-presque partout. L'ingrédient principal de la preuve sera le

Lemme 7.25. *Soit $\mathcal{B}$ une famille finie de boules (ouvertes ou fermées) sur $\mathbb{R}^d$. Il existe une sous-famille $\{B_1, \dots, B_m\} \subset \mathcal{B}$ telle que*

(1) $B_i \cap B_j = \emptyset$ *si* $i \neq j$,

(2) $\lambda^d(\bigcup_{B\in\mathcal{B}} B) \leq 3^d \lambda^d(\bigcup_{i=1}^m B_i)$.

Démonstration. Soit B_1 la boule de $\mathcal{B}$ de plus grand rayon. Si $B_1, \dots, B_k$ ont déjà été définies, soit B_{k+1} la boule de $\mathcal{B}\backslash\{B_1, \dots, B_k\}$ de plus grand rayon, disjointe de $B_1, \dots, B_k$ (voir Figure). Comme $\mathcal{B}$



est finie, ce procédé termine après avoir sélectionné un nombre fini de boules $B_1, \dots, B_m$. Pour tout $j = 1, \dots, m$, soit $\tilde{B}_j$ la boule de même centre que B_j, mais de rayon trois fois plus grand. Alors $\bigcup_{B\in\mathcal{B}} B \subset \bigcup_{j=1}^m \tilde{B}_j$, et donc

$$\lambda^d\Big(\bigcup_{B\in\mathcal{B}} B\Big) \leq \lambda^d\Big(\bigcup_{j=1}^m \tilde{B}_j\Big) \leq \sum_{j=1}^m \lambda^d(\tilde{B}_j)$$
$$= 3^d \sum_{j=1}^m \lambda^d(B_j) = 3^d \lambda^d\Big(\bigcup_{j=1}^m B_j\Big).$$

□

Preuve de la Proposition 7.24 : Par le Lemme 7.23, $E := \{x \in A : \overline{D}\mu(x) > \alpha\}$ est un Borélien. Soit K un compact et G un ouvert, tels que $K \subset E \subset G$. Pour tout $x \in K$, il existe $r_x > 0$ tel que $B_{r_x}(x) \subset G$ et tel que

$$\frac{\mu(B_{r_x}(x))}{\lambda^d(B_{r_x}(x))} > \alpha\,, \text{ et donc } \lambda^d(B_{r_x}(x)) \leq \frac{1}{\alpha}\mu(B_{r_x}(x))\,.$$

La famille $\{B_{r_x}(x), x \in K\}$ recouvre K. Soit $\mathcal{B} \subset \{B_{r_x}(x), x \in K\}$ un sous-recouvrement fini. Par le Lemme 7.25, il existe $\{B_1, \dots, B_m\} \subset \mathcal{B}$

tel que

$$\begin{aligned}\lambda^d(K) \leq \lambda^d\Big(\bigcup_{B\in\mathcal{B}} B\Big) &\leq 3^d \sum_{j=1}^m \lambda^d(B_j) \\ &\leq \frac{3^d}{\alpha}\sum_{j=1}^m \mu(B_j) \\ &= \frac{3^d}{\alpha}\mu\Big(\bigcup_{j=1}^m B_j\Big) \leq \frac{3^d}{\alpha}\mu(G)\,.\end{aligned}$$

Comme λ^d et μ sont régulières, on peut prendre le supremum sur K et l'infimum sur G, ce qui donne $\lambda^d(E) \leq \frac{3^d}{\alpha}\mu(E) \leq \frac{3^d}{\alpha}\mu(A)$.
Pour la deuxième affirmation, on décompose $\{x \in A : \overline{D}\mu(x) > 0\} = \bigcup_{k\geq 1}\{x \in A : \overline{D}\mu(x) \geq 1/k\}$. Or pour tout k, par la première affirmation, $\lambda^d(\{x \in A : \overline{D}\mu(x) \geq 1/k\}) \leq 3^d k\mu(A) = 0$. □

On considère ensuite les mesures qui ont une densité par rapport à λ^d.

Proposition 7.26. *Soit $f \geq 0$, $f \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$, et soit*

$$\mu(A) := \int_A f d\lambda^d\,, \quad A \in \mathcal{B}(\mathbb{R}^d)\,.$$

Alors μ est une mesure finie, dérivable presque partout, et

$$D\mu(x) = f(x) \quad \textit{presque partout.}$$

Démonstration. Comme f est intégrable, μ est finie, et donc régulière (Proposition 2.39). Pour montrer que $D\mu(x)$ existe et est presque partout égale à $f(x)$, il suffit de vérifier que

$$\lambda^d(\{x : \overline{D}\mu(x) > f(x)\}) = 0\,, \quad \lambda^d(\{x : \underline{D}\mu(x) < f(x)\}) = 0\,.$$

Pour la première égalité (la deuxième se traite de la même façon), on va montrer que

$$\lambda^d(E_\alpha \cap A_\alpha^c) = 0 \quad \text{pour tout } \alpha \in \mathbb{Q}\,, \tag{7.17}$$

où $E_\alpha := \{x : \overline{D}\mu(x) > \alpha\}$ et $A_\alpha := \{x : f(x) \geq \alpha\}$. Fixons $x_0 \in A_\alpha^c$ et étudions $\frac{\mu(B)}{\lambda^d(B)}$, où B est une boule quelconque centrée en x_0. On peut écrire

$$\begin{aligned}\mu(B) &= \alpha\lambda^d(B) + \big(\mu(B) - \alpha\lambda^d(B)\big)\\ &= \alpha\lambda^d(B) + \int_B (f-\alpha)d\lambda^d\\ &\leq \alpha\lambda^d(B) + \int_{B\cap A_\alpha} (f-\alpha)d\lambda^d\,.\end{aligned}$$

Comme $f-\alpha \geq 0$ sur A_α, $\nu(\cdot) := \int_{\cdot\cap A_\alpha}(f-\alpha)d\lambda^d$ est une mesure. Puisque $\|f\|_1 < \infty$, ν est finie. En prenant $B \equiv B_\epsilon(x_0)$, en divisant par $\lambda^d(B_\epsilon(x_0))$ et en prenant la $\limsup_{\epsilon\to 0^+}$, on obtient

$$\overline{D}\mu(x_0) \leq \alpha + \overline{D}\nu(x_0).$$

Mais par construction, $\nu(A_\alpha^c) = 0$. Donc, par la Proposition 7.24, $\overline{D}\nu(x_0) = 0$ pour presque tout $x_0 \in A_\alpha^c$. On a donc montré que $\overline{D}\mu(x_0) \leq \alpha$ pour presque tout $x_0 \in A_\alpha^c$. Ceci montre (7.17). □

Remarque 7.1. Soit f une fonction intégrable et $f = f^+ - f^-$ sa décomposition en partie positive et négative. En appliquant deux fois la Proposition 7.26, on a donc montré que pour presque tout x,

$$\lim_{\epsilon\to 0^+} \frac{1}{\lambda^d(B_\epsilon(x))}\int_{B_\epsilon(x)} f d\lambda^d = f(x)\,.$$

Les points x où la limite ci-dessus existe sont parfois appelés points de Lebesgue de f. Observons que pour que ces limites fassent sens, il suffit de supposer que f est localement intégrable, c'est-à-dire que $f1_K$ est intégrable pour tout compact $K \subset \mathbb{R}^d$.

Preuve du Théorème 7.22 : Soit μ une mesure finie. (1) Par la Proposition 7.24, $\mu(A) = 0$ implique $\lambda^d(\{x \in A : D\mu(x) > 0\}) = \lambda^d(\{x : \overline{D}\mu(x) > 0\}) = 0$. (2) Si $\mu \ll \lambda^d$ alors par le Théorème de Radon-Nikodým, $\mu(A) = \int_A f d\lambda^d$, avec $f = \frac{d\mu}{d\lambda^d}$. Comme μ est finie, $f \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$. Donc par la Proposition 7.26, $D\mu$ existe et est égale à $\frac{d\mu}{d\lambda^d}$ presque partout. (3) Soit $\mu = \mu_a + \mu_s$ la décomposition de Lebesgue de μ par rapport à λ^d : $\mu_a \ll \lambda^d$, $\mu_s \perp \lambda^d$. Il existe donc E tel que $\lambda^d(E) = 0$, $\mu_s(E^c) = 0$. Donc par (1), $D\mu_s = 0$ presque partout dans E^c. Par le Théorème de Radon-Nikodým, $\mu_a(A) = \int_A f_a d\lambda^d$, avec $f_a = \frac{d\mu_a}{d\lambda^d}$. Par (2), Donc $D\mu_a = f_a$ presque partout dans E. Donc $D\mu$ existe presque partout, et vaut $D\mu = D\mu_a = f_a$. □

5.2. Points de densité des ensembles mesurables. Soit A un Borélien de $\mathbb{R}^d$. Un point $x \in A$ est dit point de densité de A si

$$\lim_{\epsilon \to 0^+} \frac{\lambda^d(A \cap B_\epsilon(x))}{\lambda^d(B_\epsilon(x))} = 1 \,.$$

THÉORÈME 7.27. *Si A est un Borélien de mesure de Lebesgue positive, alors presque tous les points de A sont des points de densité.*

DÉMONSTRATION. La fonction $f = 1_A$ est intégrable. On peut donc appliquer le Théorème 7.22 : pour presque tout $x \in \mathbb{R}^d$,

$$\lim_{\epsilon \to 0^+} \frac{\lambda^d(A \cap B_\epsilon(x))}{\lambda^d(B_\epsilon(x))} = \lim_{\epsilon \to 0^+} \frac{1}{\lambda^d(B_\epsilon(x))} \int_{B_\epsilon(x)} 1_A(x) \lambda^d(dx) = 1_A(x) \,.$$

Donc ce nombre est égal à 1 pour presque tous les $x \in \mathbb{A}$. $\square$

CHAPITRE 8

Le Théorème Fondamental de l'Analyse

Le Théorème Fondamental du Calcul Différentiel (1.1) décrit les relations existant entre une fonction et sa dérivée, ou entre une fonction et ses primitives. Dans ce chapitre, on regarde de plus près ces relations, munis de la mesure de Lebesgue et de son intégrale. On considérera exclusivement des fonctions définies sur un intervale $[a, b]$ de la droite. La notion de "presque partout" sera toujours en référence à la mesure de Lebesgue λ^1 sur $[a, b]$.

Soit $f : [a, b] \to \mathbb{R}$ une fonction intégrable. On peut définir, pour tout $x \in [a, b]$, l'intégrale indéfinie

$$F(x) := \int_a^x f d\lambda^1 \equiv \int_{[a,b]} f 1_{[a,x]} d\lambda^1 \, .$$

On sait par le Théorème 3.34 que F est continue. On sait aussi (voir Théorème 1.1) que si f est continue, alors F est dérivable partout sur (a, b), et

$$F'(x) = f(x) \, , \quad \forall x \in (a, b) \, .$$

Lorsque l'on suppose seulement que f est intégrable, cette dernière propriété peut ne plus être vraie. Il suffit de prendre par exemple la fonction

$$f(x) := \begin{cases} 0 & \text{si } x < c \, , \\ 1 & \text{si } x \geq c \, , \end{cases}$$

où $c \in (a, b)$. Cette fonction est intégrable, et son intégrale définie $F(x) = 0$ si $x \leq 0$, $F(x) = x$ si $x > 0$, est dérivable partout sauf en $x = c$. En particuler, $F' = f$ partout sauf en $x = c$. En prenant $f = 1_{\mathbb{Q} \cap [a,b]}$, on obtient $F'(x) = 0$ partout, et donc $F' = f$ presque partout.

Théorème 8.1 (Le Théorème Fondamental de l'Analyse, selon Lebesgue). *Soit f une fonction intégrable sur $[a,b]$. Alors*

$$F(x) := \int_a^x f d\lambda^1$$

est dérivable presque partout, et $F' = f$.

Observons que par définition, $F(x) = \int_a^x f_+ d\lambda^1 - \int_a^x f_- d\lambda^1$, où $f = f_+ - f_-$ est la décomposition de f en ses parties positives et négatives, $f_\pm \geq 0$. On prouvera donc d'abord le théorème dans le cas où $f \geq 0$. Or dans ce cas, $F(x)$ est monotone non-décroissante. On montrera donc le résultat suivant, plus général :

Théorème 8.2 (Lebesgue). *Si $F : [a,b] \to \mathbb{R}$ est continue et monotone, alors elle est dérivable presque partout.*

En fait, l'hypothèse de continuité sur F n'est pas nécessaire, mais elle simplifiera un peu l'exposition. On supposera F non-décroissante.

On introduit les dérivées de Dini :

$$\begin{aligned} {}^{\triangleright}DF(x) &:= \limsup_{y \to x^-} \frac{f(y)-f(x)}{y-x}, & D^{\triangleleft}F(x) &:= \limsup_{y \to x^+} \frac{f(y)-f(x)}{y-x} \\ {}_{\triangleright}DF(x) &:= \liminf_{y \to x^-} \frac{f(y)-f(x)}{y-x}, & D_{\triangleleft}F(x) &:= \liminf_{y \to x^+} \frac{f(y)-f(x)}{y-x} \end{aligned}$$

Comme f est non-décroissante, tous ces nombres sont ≥ 0. De plus, on a évidemment que

$$D_{\triangleleft}F \leq D^{\triangleleft}F, \quad {}_{\triangleright}DF \leq {}^{\triangleright}DF.$$

Pour montrer que F est dérivable en un point x, il faut montrer que ses 4 dérivées de Dini sont égales et finies. Le Théorème 8.2 sera donc une conséquence des Propositions 8.3 et 8.6 ci-dessous :

Proposition 8.3. *Si F est continue et non-décroissante,*

$$\lambda^1\{x \in [a,b] : D^{\triangleleft}F(x) = \infty\} = 0.$$

On commencera par la remarque suivante. Soit $x \in [a,b)$ tel que $D^{\triangleleft}F(x) = +\infty$. Alors pour tout $C > 0$, $D^{\triangleleft}F(x) > C$, et donc il existe $z > x$, $z \in (a,b)$, tel que

$$\frac{f(z)-f(x)}{z-x} > C, \text{ c'est-à-dire } f(x) - Cx < f(z) - Cz.$$

Définition 8.4. *Soit $g : [a,b] \to \infty$ une fonction continue. Un point $x \in (a,b)$ est dit* invisible depuis $+\infty$ (resp. $-\infty$) pour *g si il existe $z > x$ (resp. $z < x$), $z \in (a,b)$, tel que $g(x) < g(z)$. On note $I^+(g)$ (resp. $I^-(g)$) l'ensemble des points de $[a,b]$ invisibles depuis $+\infty$ (resp. $-\infty$) pour g.*

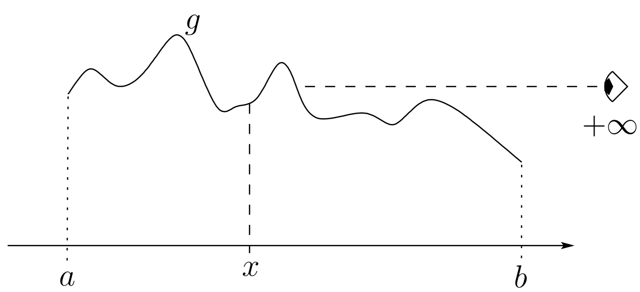


Figure 1. Un point x invisible depuis $+\infty$.

Lemme 8.5. *Si $g : [a,b] \to \infty$ est continue. Alors*

(1) $I^+(g)$ est un ouvert, $I^+(g) = \bigcup_n (a_n, b_n)$, où les $(a_n, b_n) \subset [a,b]$ sont disjoints deux-à-deux, et $g(a_n) \leq g(b_n)$ pour tout n.

(2) $I^-(g)$ est ouvert, $I^-(g) = \bigcup_k (c_k, d_k)$, où les $(c_k, d_k) \subset [a,b]$ sont disjoints deux-à-deux, et $g(d_k) \leq g(c_k)$ pour tout k.

Démonstration. Soit x invisible depuis $+\infty$ pour g : il exists un $z > x$ tel que $g(x) < g(z)$. Pour ce même z, et par la continuité de g, on a $g(y) < g(z)$ pour tous les $y \in (a,b)$ suffisamment proches de x. Ceci prouve que $I^+(g)$ est ouvert, et donc qu'il peut s'écrire comme une union dénombrable d'intervalles ouverts disjoints (a_n, b_n). Supposons, par l'absurde, qu'il existe un (a_n, b_n) tel que $g(a_n) > g(b_n)$. Soit h tel que $g(b_n) < h < g(a_n)$, par exemple $h := \frac{g(a_n)+g(b_n)}{2}$. Soit alors $x_* := \sup\{x \in (a_n, b_n) : g(x) = h\}$. D'une part $a_n < x_* < b_n$, et donc $x_* \in I^+(g)$. D'autre part, $g(x_*) > g(z)$ pour tout $z \in (x_*, b_n]$. Mais, comme $b_n \notin I^+(g)$ par construction, on a aussi $g(b_n) \geq g(z)$ pour tout $z > b_n$. Donc $g(x_*) \geq g(z)$ pour tout $z > x_*$, et donc $x_* \notin I^+(g)$, une contradiction. □

Preuve de la Proposition 8.3 : La fonction $g_C(x) := F(x) - Cx$ est continue, et on a vu que

$$\{D^{\triangleleft} F > C\} \subset I^+(g_C)\,.$$

Par le lemme précédent, on peut écrire $I^+(g_C) = \bigcup_n (a_n, b_n)$, avec $g_C(a_n) \leq g_C(b_n)$, c'est-à-dire $b_n - a_n \leq \frac{1}{C}(F(b_n) - F(a_n))$. Donc

$$\begin{aligned}\lambda^1\{I^+(g_C)\} &\leq \sum_n (b_n - a_n) \\ &\leq \frac{1}{C}\sum_n (F(b_n) - F(a_n)) \\ &\leq \frac{F(b) - F(a)}{C}\,.\end{aligned}$$

Dans la dernière inégalité, on a utilisé la monotonicité de F. On conclut que

$$\lambda^1\{D^{\triangleleft}F = \infty\} \leq \inf_{C>0} \lambda^1\{D^{\triangleleft}F > C\} = 0\,. \qquad \square$$

PROPOSITION 8.6. *Si F est continue et non-décroissante,*

$$\begin{aligned}\lambda^1\{x \in [a,b] : D^{\triangleleft}F(x) > {}_{\triangleright}DF(x)\} = 0\,,\\ \lambda^1\{x \in [a,b] : {}^{\triangleright}DF(x) > D_{\triangleleft}F(x)\} = 0\,.\end{aligned}$$

DÉMONSTRATION. On ne prouvera que la première identité (la deuxième s'obtient de la première appliquée à $\tilde{F}(x) := -F(-x)$, définie sur $[-b,-a]$). Pour commencer, on décompose

$$\{D^{\triangleleft}F > {}_{\triangleright}DF\} = \bigcup_{\substack{p,q\in\mathbb{Q}:\\ 0\leq p<q}} E_{p,q}\,,$$

où $E_{p,q} := \{{}_{\triangleright}DF < p < q < D^{\triangleleft}F\}$. On va montrer que

$$\forall (\alpha,\beta) \subset [a,b]\,, \quad \lambda^1\{(\alpha,\beta) \cap E_{p,q}\} \leq \frac{p}{q}(\beta - \alpha)\,. \tag{8.1}$$

On montre facilement (voir Série 1) que ceci implique $\lambda^1(E_{p,q}) = 0$. Fixons donc un intervalle $(\alpha,\beta) \subset [a,b]$, et écrivons

$$(\alpha,\beta) \cap E_{p,q} = \big\{(\alpha,\beta) \cap \{{}_{\triangleright}DF < p\}\big\} \cap \{D^{\triangleleft}F > q\}\,.$$

Considérons alors la restriction de F à (α,β). Comme précédemment, on observe que si $x \in (\alpha,\beta)$ est tel que ${}_{\triangleright}DF(x) < p$, alors il existe $z \in (\alpha,\beta)$, $z < x$, tel que $\frac{F(z)-F(x)}{z-x} < p$, et donc x est invisible depuis $-\infty$ pour la fonction $g_p(x) := F(x) - px$. On a donc

$$(\alpha,\beta) \cap \{{}_{\triangleright}DF < p\} \subset \bigcup_n (a_n, b_n)\,,$$

où les $(a_n,b_n) \subset (\alpha,\beta)$, avec $g_p(b_n) \leq g_p(a_n)$. Ceci implique

$$F(b_n) - F(a_n) \leq p(b_n - a_n)\,. \tag{8.2}$$

On recommence ensuite le procédé pour chacun des intervalles (a_n, b_n) : en considérant la restriction de F à (a_n, b_n), on remarque qu'un point $x \in (a_n, b_n)$ tel que $D^{\triangleleft}F(x) > q$ est invisible depuis $+\infty$ pour $g_q(x) = F(x) - qx$, et donc

$$(a_n, b_n) \cap \{D^{\triangleleft}F < p\} \subset \bigcup_k (a_{n,k}, b_{n,k})\,,$$

avec $(a_{n,k}, b_{n,k}) \subset (a_n, b_n)$ et

$$b_{n,k} - a_{n,k} \leq \frac{1}{q}(f(b_{n,k}) - f(a_{n,k}))\,. \tag{8.3}$$

En utilisant (8.2) et (8.3), on a donc

$$\begin{aligned}
\lambda^1\{(\alpha, \beta) \cap E_{p,q}\} &\leq \bigcup_{n,k}(b_{n,k} - a_{n,k}) \\
&\leq \frac{1}{q}\sum_n \Big\{\sum_k (f(b_{n,k}) - f(a_{n,k}))\Big\} \\
&\leq \frac{1}{q}\sum_n (f(b_n) - f(a_n)) \\
&\leq \frac{p}{q}\sum_n (b_n - a_n) \leq \frac{p}{q}(\beta - \alpha)\,.
\end{aligned}$$

Ceci prouve (8.1), et par conséquent la première affirmation de la proposition. □

Preuve du Théorème 8.1 : Soit $f = f_+ - f_-$ intégrable. Comme $F(x) = \int_a^x f^+ \, d\lambda^1 - \int_a^x f^- \, d\lambda^1$, et que chacun des termes est une fonction continue et monotone, F est dérivable presque partout par le Théorème 8.2. Il reste à prouver que lorsqu'elle existe, la dérivée de F est égale à f presque partout : $\lambda^1\{F' \neq f\} = 0$. On va d'abord montrer que $\lambda^1\{F' > f\} = 0$. D'abord, décomposons

$$\{F' > f\} = \bigcup_{\substack{p,q \in \mathbb{Q}:\\ p<q}} A_{p,q}\,,$$

où $A_{p,q} = \{f < p < q < F'\}$. (Obs : p et q sont de signes a priori quelconques.) Prenons une paire p, q comme ci-dessus. Fixons $\epsilon > 0$, et considérons δ suffisamment petit, tel que

$$\text{si } \lambda^1(A) \leq \delta\,, \quad \text{alors } \Big|\int_A f \, d\lambda^1\Big| \leq \epsilon\,.$$

(L'existence d'un tel δ a été prouvée en exercice, dans la Série 6). Utilisons ensuite la régularité de la mesure de Lebesgue, et considérons un ouvert G, $(a,b) \supset G \supset A_{p,q}$ tel que $\lambda^1(G) - \delta \leq \lambda^1(A_{p,q}) \leq \lambda^1(G)$. Considérons la restriction de F à G. Un $x \in G$ qui satisfait $F'(x) > q$ est invisible depuis $+\infty$ pour $F(x) - qx$. En particulier, on a

$$G \cap A_{p,q} \subset \bigcup_n (a_n, b_n) \equiv S\,,$$

avec $F(b_n) - F(a_n) \geq q(b_n - a_n)$ pour tout n (rappelons qu'on ne sait rien sur le signe de q, donc on évite de diviser par q). On a alors

$$\begin{aligned} q\lambda^1(S) &= q\sum_n (b_n - a_n) \\ &\leq \sum_n (F(b_n) - F(a_n)) \\ &= \sum_n \int_{[a_n,b_n]} f\,d\lambda^1 = \int_S f\,d\lambda^1 = \int_{A_{p,q}} f\,d\lambda^1 + \int_{S\setminus A_{p,q}} f\,d\lambda^1\,. \end{aligned}$$

Observons que $\lambda^1(S \setminus A_{p,q}) \leq \lambda^1(G \setminus A_{p,q}) \leq \delta$. Donc d'une part $|\int_{S\setminus A_{p,q}} f\,d\lambda^1| \leq \epsilon$, de l'autre

$$\int_{A_{p,q}} f\,d\lambda^1 \leq p\lambda^1(A_{p,q}) = p\lambda^1(S) - p\lambda^1(S \setminus A_{p,q}) \leq p\lambda^1(S) + |p|\delta\,.$$

On a donc montré que

$$\lambda^1(S) \leq \frac{\epsilon + |p|\delta}{q - p}\,,$$

qui peut être rendu arbitrairement petit en choisissant bien ϵ et δ. Comme $\lambda^1(A_{p,g}) \leq \lambda^1(G) = 0$, ceci montre que $\lambda^1\{F' > f\} = 0$ En répétant la même procédure pour $-f$, on montre que $\lambda^1\{F' < f\} = 0$, ce qui conclut la preuve du Théorème Fondamental. $\square$

Bibliographie

1. R. G. Bartle, *The elements of integration*, John Wiley & Sons Inc., New York, 1966.
2. Heinz Bauer, *Probability theory and elements of measure theory*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1981.
3. P. Billingsley, *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1995.
4. V. I. Bogachev, *Measure theory. Vol. I, II*, Springer-Verlag, Berlin, 2007.
5. D. M. Bressoud, *A radical approach do Lebesgue's Theory of Integration*, Cambridge University Press, 2008.
6. C.-J. de la Vallée Poussin, *Intégrales de Lebesgue, fonctions d'ensembles, classes de Baire*, deuxième édition ed., Gauthier-Villars, Paris, 1934.
7. G. B. Folland, *Real analysis*, second ed., Pure and Applied Mathematics (New York), John Wiley & Sons Inc., New York, 1999, Modern techniques and their applications, A Wiley-Interscience Publication.
8. J. Gapaillard, *Intégration pour la licence : cours et exercices corrigés (*2ème *ed.)*, Dunod, 1997.
9. P. R. Halmos, *Measure Theory*, D. Van Nostrand Company, Inc., New York, N. Y., 1950.
10. A. Kolmogorov and S. Fomine, *Eléments de la théorie des fonctions et de l'analyse fonctionnelle*, Éditions Mir, Moscou, 1974.
11. Lebesgue H. L., *Leçons sur l'intégration et la recherche des fonctions primitives*, Gauthier-Villars, 1928.
12. J. Neveu, *Bases mathématiques du calcul des probabilités*, Masson et Cie, Éditeurs, Paris, 1964.
13. Durrett R., *Probability : Theory and examples*, Duxbury Press, 1988.
14. W. Rana, *Introduction to Measure and Integration*, third ed., Springer-Verlag, New York, 2009.
15. W. Rudin, *Real and complex analysis*, third ed., McGraw-Hill Book Co., New York, 1987.
16. A. N. Shiryayev, *Probability*, Graduate Texts in Mathematics, vol. 95, Springer-Verlag, New York, 1984.
17. T. Tao, *An introduction to measure theory*, Graduate Studies in Mathematics, vol. 95, American Mathematical Society, Providence, 2010.