

Théorie de la mesure, Intégration, Probabilités

Classe sino-française, USTC

Stéphane Nonnenmacher*

Février-Avril 2022

Résumé

Ces notes de cours présentent la théorie de la mesure et de l'intégrale de Lebesgue. On suppose connues les bases de la théorie des ensembles et de la topologie. On l'étend ensuite à la théorie des probabilités (non discrètes).

Ces notes reprennent le schéma du cours que Jean-François Le Gall a donné à l'ENS en 2006.

Ce cours est donné aux étudiants de 2e année de la classe sino-française de l'Université des Sciences et Technologies de Chine, à Hefei, durant le printemps 2022.

*<https://www.imo.universite-paris-saclay.fr/~nonnenma/index.html>

Table des matières

I	Théorie de la mesure et Intégration	5
1	Introduction - Motivation	5
1.1	Rappel sur l'intégrale de Riemann	5
1.2	Problèmes de l'intégrale de Riemann	7
1.3	Approche de Lebesgue de l'intégration : « il faut découper dans le sens horizontal »	9
2	Mesures positives, fonctions mesurables	12
2.1	Ensembles mesurables, tribu	12
2.2	Mesure positive définie sur une tribu	15
2.3	Lemme de classe monotone : identification d'une mesure	19
2.4	Fonctions mesurables (entre deux tribus)	23
2.4.1	Définition et premières propriétés des fonctions mesurables	23
2.4.2	Opérations sur les fonctions mesurables	26
2.4.3	Suites de fonctions mesurables à valeurs dans $\overline{\mathbb{R}}$	27
2.4.4	Transport d'une mesure par une application	29
3	Intégration d'une fonction par rapport à une mesure	31
3.1	Intégration d'une fonction mesurable positive	31
3.1.1	Fonctions étagées et définition de l'intégrale de fonctions positives	31
3.1.2	Le théorème de convergence monotone	35
3.2	Fonctions intégrables	42
3.2.1	Intégrale d'une fonction réelle intégrable	42
3.2.2	Fonctions à valeurs complexes	45
3.2.3	Le théorème de convergence dominée	46
3.3	Intégrales dépendant d'un paramètre	48
3.3.1	Continuité d'une intégrale à paramètre	49
3.3.2	Dérivabilité d'une intégrale à paramètre	50
4	Construction et propriétés de la mesure de Lebesgue	53
4.1	Mesures extérieures	53
4.2	Construction de la mesure extérieure de Lebesgue λ^* sur \mathbb{R}	57
4.3	Généralisations en dimension supérieure	63
4.4	Quelques propriétés de la mesure de Lebesgue	66
4.4.1	Invariance par translation	66
4.4.2	Régularité	66

4.4.3	Relation avec l'intégrale de Riemann	67
4.4.4	Il existe un ensemble non mesurable pour la mesure de Lebesgue	69
4.5	Construction plus générale d'une mesure	70
4.6	Mesures finies sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$; intégrale de Stieltjes	72
5	Espaces L^p	80
5.1	Définition des espaces L^p ; inégalité de Hölder	80
5.1.1	Définition de L^p	80
5.1.2	Définition de l'espace L^∞	81
5.1.3	Inégalités de Hölder	82
5.1.4	Cas d'une mesure finie : emboîtement des espaces L^p	84
5.2	L'espace de Banach $L^p(E, \mathcal{A}, \mu)$	86
5.3	Théorèmes de densité dans L^p	93
5.4	Décomposition de mesures : théorème de Radon-Nikodym	97
6	Mesures produits	103
6.1	Espace produit	103
6.2	Construction de la mesure produit	106
6.3	Le théorème de Fubini	110
6.4	Applications du théorème de Fubini	112
6.4.1	Intégration par parties	113
6.4.2	Propriétés de la convolution	114
6.4.3	Une application de la convolution : régularisation de la mesure de Dirac	116
7	Formule de changement de variables	120
7.1	Changement de variable affine	120
7.2	Changement de variable non-linéaire	122
7.3	Application : coordonnées sphériques	126
II	Théorie des probabilités	131
8	Définitions générales	132
8.1	Exemples d'espaces de probabilité	133
8.2	Variables aléatoires	136
8.2.1	Variables aléatoires discrètes	138
8.2.2	Variables aléatoires à densité	139
8.3	Espérance (mathématique)	139
8.3.1	Une expérience aléatoire géométrique : le paradoxe de Bertrand	142

8.3.2	Lois classiques de probabilité	145
8.3.3	Fonction de répartition d'une v.a. réelle	147
8.3.4	Tribu engendrée par une v.a.	148
8.4	Moments d'une v.a.	150
8.4.1	Variance et matrice de covariance	151
8.4.2	Régression linéaire	153
8.4.3	Fonction caractéristique	156
8.4.4	Fonction génératrice d'une variable aléatoire discrète	161
9	Indépendance	163
9.1	Evènements indépendants	163
9.2	Tribus et variables aléatoires et tribus indépendantes	166
9.2.1	Caractérisation de l'indépendance de n v.a. en termes de la loi conjointe	167

Première partie

Théorie de la mesure et Intégration

1 Introduction - Motivation

La théorie de l'intégration a d'abord consisté à calculer des primitives de fonctions explicites, de façon à résoudre l'équation différentielle $u'(x) = f(x)$, avec $f : I \rightarrow \mathbb{R}$ une fonction donnée (explicite), et u la fonction inconnue (ici $I \subset \mathbb{R}$ est un intervalle borné, non réduit à un point ; on pourra prendre par exemple $I = [0, 1]$).

1.1 Rappel sur l'intégrale de Riemann

A partir du 19e siècle, les mathématiciens se sont posé la question d'intégrer des fonctions moins explicites, en généralisant de plus en plus les classes de fonctions pour lesquelles l'intégrale avait un sens. Un pas important a été accompli par B.Riemann, qui a défini l'intégrale de fonctions définies sur \mathbb{R} au moyen d'approximations par des fonctions constantes par morceaux. Sa définition permet alors d'intégrer n'importe quelle fonction $f : I \rightarrow \mathbb{R}$ *continue*, et en fait n'importe quelle fonction *réglée*. Pour définir les fonctions réglées, on commencera par définir une fonction en escalier.

Définition 1.1. Soit $a < b$ deux points de \mathbb{R} , et soit $I = [a, b]$ l'intervalle correspondant. Une fonction $f : I \rightarrow \mathbb{R}$ est une *fonction en escalier* (ou *fonction constante par morceaux*) s'il existe une suite finie de points $a = x_0 < x_1 < x_2 \cdots < x_K = b$, telle que f est constante sur chaque sous-intervalle ouvert $I_k^\circ =]x_{k-1}, x_k[$, $k = 1, \dots, K$ (les valeurs prises par f aux points x_k sont arbitraires).

Pour étudier les suites de fonctions, on utilisera tout d'abord la norme de la convergence uniforme. Pour toute fonction $f : I \rightarrow \mathbb{R}$, on définit la norme sup :

$$\|f\|_{sup} = \sup_{x \in I} |f(x)|.$$

Pour une suite $(f_n)_{n \geq 1}$ de fonctions définies sur I , on dit que la suite (f_n) converge *uniformément* vers f si et seulement si

$$\|f - f_n\|_{sup} \xrightarrow{n \rightarrow \infty} 0.$$

Définition 1.2. Une fonction $f : I \rightarrow \mathbb{R}$ est une fonction réglée s'il existe une suite (f_n) de fonctions en escalier sur I , qui converge uniformément vers f lorsque $n \rightarrow \infty$.

Notons qu'en général, la partition de I en sous-intervalles *dépend* de la fonctions f_n : on notera alors $I = \bigcup_{k=1}^{K(n)} I_k^{(n)}$.

Proposition 1.3. *Toute fonctions f continue est réglée.*

Démonstration. : Supposons pour simplifier que $I = [0, 1]$. On peut par exemple, pour chaque $n > 0$, découper I en n intervalles $I = \bigcup_{k=1}^n \left[\frac{k-1}{n}, \frac{k}{n} \right]$ et construire la fonction en escalier

$$f_n(x) = f\left(\frac{k}{n}\right) \quad \text{si } x \in \left[\frac{k-1}{n}, \frac{k}{n} \right], k = 1, \dots, n \quad \text{et} \quad f_n(1) = f(1).$$

La fonction f étant continue sur l'intervalle compact I , elle y est *uniformément continue* (Théorème de Heine) :

$$\forall \epsilon > 0, \exists n = n(\epsilon) \in \mathbb{N}^*, \forall x, y \in I, |x - y| \leq \frac{1}{n} \implies |f(x) - f(y)| \leq \epsilon.$$

Cette propriété implique alors que

$$\forall m \geq n(\epsilon), \forall x \in I, |f_m(x) - f(x)| \leq \epsilon,$$

donc que les fonctions (f_n) convergent uniformément vers f lorsque $n \rightarrow \infty$. □

Affirmation 1.4. On peut montrer qu'une fonction $f : I \rightarrow \mathbb{R}$ est réglée si et seulement si elle admet, en tout point $x \in I$, une limite à droite et une limite à gauche (ces deux limites peuvent être différentes, et peuvent être toutes deux différentes de $f(x)$).

La définition de l'intégrale d'une fonction en escalier est simple. Soit f une fonction en escalier, telle que

$$f(x) = a_k \quad \text{lorsque } x \in]x_{k-1}, x_k[, \quad \text{pour } k = 1, \dots, K.$$

Alors l'intégrale de f sur I est donnée par

$$\int_I f(x) dx = \sum_{k=1}^K a_k (x_k - x_{k-1}).$$

Proposition 1.5. *Soit une fonction réglée $f : I \rightarrow \mathbb{R}$, et soit (f_n) une suite de fonctions en escalier convergeant uniformément vers f . On montre alors que les intégrales $S_n = \int_I f_n(x) dx$*

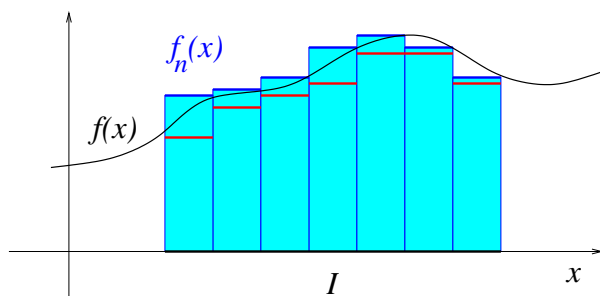


FIGURE 1.1 – Approximation d'une fonction continue f par une fonction en escalier f_n , et représentation de l'intégrale $\int_I f_n(x) dx$ (en bleu). Une seconde fonction en escalier (en rouge) minore f .

ont une limite lorsque $n \rightarrow \infty$. Cette limite définit l'intégrale de la fonction f sur I , notée $\int_I f(x) dx = \int_a^b f(x) dx$.

Démonstration. Le fait que f_n converge vers f uniformément implique que, pour tout $\epsilon > 0$, il existe $n(\epsilon)$ tel que, si $n, m \geq n(\epsilon)$, alors $\|f_n - f_m\|_{sup} \leq \epsilon$. La fonction $f_n - f_m$ est alors une fonction en escalier, associée à la partition formée par l'intersection des partitions $(I_k^{(n)})$ et $(I_j^{(m)})$. Cette fonction satisfaisant $\sup_x |f_n(x) - f_m(x)| \leq \epsilon$, son intégrale vérifie alors l'inégalité

$$|S_n - S_m| = \left| \int_I (f_n(x) - f_m(x)) dx \right| \leq \epsilon (b - a) = \epsilon |I|,$$

où on a noté $|I| = b - a$ la longueur de l'intervall I . Ceci montre que (S_n) est une suite de Cauchy, donc qu'elle converge.

Il reste à vérifier que la limite $\lim_{n \rightarrow \infty} S_n$ ne dépend pas du choix de la suite (f_n) convergeant vers f . En effet, si on considère deux suites (f_n) et (\tilde{f}_n) de fonctions en escalier convergeant uniformément vers f , alors la suite $f_n - \tilde{f}_n$ converge vers zéro uniformément, ce qui implique que $\int f_n(x) dx - \int \tilde{f}_n(x) dx$ tend vers zéro. \square

Graphiquement, l'intégrale de Riemann d'une fonction réglée f représente l'aire algébrique entre le graphe de f au-dessus de I et l'intervalle I , voir la figure .

1.2 Problèmes de l'intégrale de Riemann

Dans la seconde moitié du 19e siècle, les mathématiciens ont introduit des familles de fonctions de plus en plus compliquées, et qu'on ne pouvait pas obtenir comme limites uniformes de fonctions en escalier, mais seulement comme limites simples de fonction en escalier (ou limites

simples de fonctions continues). L'intégrale de Riemann ne permet alors pas de donner un sens à l'intégrale d'une telle fonction. Pouvait-on néanmoins définir, d'une autre manière, l'intégrale d'une telle fonction ?

Exemple 1.6. Sur l'intervalle $I = [0, 1]$, on considère la fonction caractéristique des nombres rationnels, qu'on notera $\mathbb{1}_{\mathbb{Q} \cap I}$. Cette fonction est discontinue en tout point ; elle n'admet de limite à droite ou à gauche en aucun point $x \in I$, donc ce n'est pas une fonction réglée. Cette fonction peut être obtenue comme la double limite :

$$\forall x \in I, \quad \mathbb{1}_{\mathbb{Q} \cap I}(x) = \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \cos(\pi n! x)^{2m} \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} F_n(x),$$

mais aucune des deux limites n'est uniforme. En effet, $F_n(x) = 1$ sur les points $x \in I$ multiples de $\frac{1}{n!}$, et $F_n(x) = 0$ partout ailleurs. Les fonctions F_n sont des fonctions en escalier d'intégrale nulle. Elles convergent simplement vers $\mathbb{1}_{\mathbb{Q} \cap I}$, mais pas uniformément. Peut-on néanmoins donner un sens à l'intégrale de $\mathbb{1}_{\mathbb{Q} \cap I}$?

Intégrale de Riemann - bis

Pour toute fonction bornée $f : I \rightarrow \mathbb{R}$, on peut chercher à assouplir la définition de l'intégrale de Riemann donnée ci-dessus, en considérant uniquement les fonctions en escalier f_n qui minorent f (respectivement qui majorent f) : on peut ainsi définir

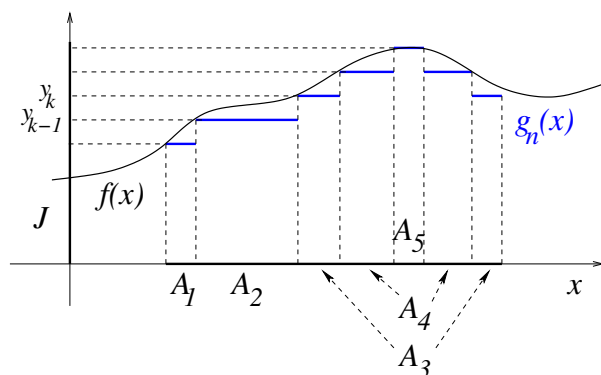
$$S_-(f) \stackrel{\text{def}}{=} \sup_{f_n \leq f} \int_I f_n(x) dx,$$

où on prend le supremum sur toutes les fonctions en escalier f_n minorant f . De même, on définit

$$S_+(f) \stackrel{\text{def}}{=} \inf_{g_n \geq f} \int_I g_n(x) dx.$$

On montre facilement que $S_-(f) \leq S_+(f)$. Si la fonction f est telle que $S_-(f) = S_+(f)$, cette valeur définit l'intégrale de Riemann de f ; une telle fonction f est dite Riemann-intégrable. Les fonctions Riemann-intégrables forment une classe plus large que les fonctions réglées.

Exemple 1.7. 1. On voit facilement que notre fonction $\mathbb{1}_{\mathbb{Q} \cap I}$ n'est pas Riemann-intégrable. En effet, le fait que $\mathbb{R} \setminus \mathbb{Q}$ soit dense dans \mathbb{R} montre que pour toute fonction en escalier $f_n \leq \mathbb{1}_{\mathbb{Q} \cap I}$, $f_n(x)$ doit prendre des valeurs négatives ou nulles sur tout intervalle de sa partition associée, de sorte que $\int_I f_n(x) dx \leq 0$. Tandis que pour toute fonction en escalier $g_n \geq \mathbb{1}_{\mathbb{Q} \cap I}$, du fait de la densité de \mathbb{Q} dans \mathbb{R} , $g_n(x) \geq 1$ sur tout intervalle de sa partition, de sorte que $\int_I g_n(x) dx \geq 1$.

FIGURE 1.2 – Stratégie de Lebesgue : approximation de f par une fonction étagée g_n .

2. Par contre, la fonction $x \mapsto \sin(1/x)$ définie sur l'intervalle $(0,1)$ n'est pas réglée (elle n'admet pas de limite en $x \searrow 0$), mais on peut montrer qu'elle est Riemann-intégrable.

1.3 Approche de Lebesgue de l'intégration : « il faut découper dans le sens horizontal »

L'idée formulée par H. Lebesgue au tout début de 20e siècle, est de calculer l'aire (algébrique) entre le graphe de f et l'intervalle I non pas en découpant la surface par des rectangles fins *verticaux* comme le fait Riemann (v. la Figure 1.1), mais par des « bandes horizontales ». Plutôt que de découper l'intervalle « source » I en sous-intervalles $I_k^{(n)}$ pour fabriquer une fonction en escalier f_n approximant f , on découpe l'intervalle « image » $J \subset \mathbb{R}$ en sous-intervalles $J_k^{(n)} = [y_{k-1}^{(n)}, y_k^{(n)}] \cap J$, par exemple en prenant $y_k^{(n)} = \frac{k}{n}$, et on fabrique une fonction approximante g_n , telle que

$$g_n(x) = y_{k-1}^{(n)} \quad \text{pour } x \in A_k^{(n)} \stackrel{\text{def}}{=} f^{-1} \left(\left[y_{k-1}^{(n)}, y_k^{(n)} \right] \right) \cap I. \quad (1.1)$$

La fonction g_n est appelée une *fonction étagée*. Elle ressemble à une fonction en escalier, mais la différence est que les ensembles $A_k \subset I$ peuvent être beaucoup plus compliqués que des sous-intervalles. Sur la figure, on voit déjà que certains A_k sont des sous-intervalles, d'autres des unions (finies) de sous-intervalles. L'intégrale de la fonction étagée g_n est définie par

$$\int_I g_n(x) dx = \sum_k y_{k-1}^{(n)} |A_k^{(n)}| \quad (1.2)$$

avec $|A_k^{(n)}|$ indiquant la longueur du sous-ensemble (« l'étage ») $A_k^{(n)}$.

Pour une fonction $f : I \rightarrow \mathbb{R}$ continue, on peut montrer que les intégrales $\int g_n(x) dx$ convergent vers l'intégrale de Riemann $\int f(x) dx$, c'est donc une seconde manière d'obtenir la même valeur.

Que gagne-t-on avec cette second manière d'approximer f par un « découpage horizontal » ?

On voit que le calcul de l'intégrale par cette second manière impose de *mesurer* les sous-ensembles $A_k^{(n)}$, définis comme les images réciproques par f des intervalles $J_k^{(n)}$. Que signifie « mesurer » ? Lorsque $A_k^{(n)}$ est une union finie d'intervalles (comme sur la figure), sa longueur est la somme des longueurs des sous-intervalles. Mais que se passe-t-il si $A_k^{(n)}$ est un ensemble plus compliqué ? Peut-on définir de façon raisonnable sa « longueur » ?

La théorie de la Lebesgue va répondre de façon précise à ces questions. Elle aura pour objectif de

1. définir une famille de sous-ensembles $A \subset I$, auxquels on peut assigner une « longueur » ; on appellera cette longueur la *mesure de Lebesgue* de A , et un tel sous-ensemble A sera appelé « ensemble mesurable ».
2. On définira ensuite la une classe de fonctions (dites « fonctions mesurables »), pour lesquelles chaque image réciproque d'un intervalle $f^{-1}([a, b])$ est un ensemble mesurable.
3. Une fonction mesurable pourra être approchée par une famille de fonctions étagées, du type $g_n = \sum_k c_k^{(n)} \mathbb{1}_{A_k^{(n)}}$ avec $A_k^{(n)}$ des ensembles mesurables (la construction de g_n peut être différente de celle donnée en (1.1)). On peut définir l'intégrale de g_n comme en (1.2), et il faudra montrer sous quelles conditions les intégrales $\int g_n(x) dx$ ont une limite.

Ce qu'on gagne par rapport à la théorie de Riemann, c'est que la convergence des g_n vers f n'a plus besoin d'être uniforme, mais elle peut être *simple* : il suffit que

$$\forall x \in I, \quad g_n(x) \xrightarrow{n \rightarrow \infty} f(x).$$

Pour une telle convergence simple des g_n , la convergence des $\int g_n(x) dx$ sera assurée par des théorèmes importants de la théorie de Lebesgue : le *Théorème de convergence monotone*, et le *Théorème de convergence dominée*, ainsi que le *Lemme de Fatou*.

Un autre avantage de la théorie de Lebesgue, c'est qu'on va élargir la notion de « longueur » d'un sous-ensemble $A \subset I$, en la notion plus générale de « mesure » de cet ensemble. La mesure d'un ensemble A sera un « poids » (un nombre réel positif, parfois $+\infty$) associé à

cet ensemble, on notera ce poids $\mu(A)$. On définira alors l'intégrale d'une fonction étagée, relativement à cette mesure μ , par la même formule qu'en (1.2), mais en remplaçant les longueurs $|A_k^{(n)}|$ par les poids $\mu(A_k^{(n)})$. On aura alors les mêmes théorèmes de convergence vers une intégrale « de mesure μ », qu'on notera $\int_I f(x) d\mu(x)$.

Cette généralisation de la notion de « longueur » à celle de « mesure » retire à l'intégrale son caractère géométrique : l'intégrale $\int_I f(x) d\mu(x)$ ne sera plus reliée à l'aire (algébrique) comprise entre le graphe de f et I . Mais ce type d'intégrale aura une interprétation importante en *théorie des probabilités*, que nous étudierons dans la seconde partie de ce cours. Le poids $\mu(A)$ sera alors interprété comme la probabilité de l'ensemble A , avec comme condition que l'ensemble total (ici, l'intervalle I) est de probabilité pleine $\mu(I) = 1$; une telle mesure est appelée mesure de probabilité. Une fonction f sera alors interprétée comme une variable aléatoire, et l'intégrale $\int_I f(x) d\mu(x)$ comme la moyenne (ou l'espérance) de cette variable aléatoire, par rapport à la mesure de probabilité μ .

La théorie de Lebesgue permet aussi de généraliser l'espace portant la mesure, et les fonctions f : on a pris jusqu'à présent $I \subset \mathbb{R}$ un intervalle réel. On pourra considérer à la place de I un ouvert de \mathbb{R}^d , un espace métrique, voire un ensemble E quelconque, auquel cas la mesure d'un sous-ensemble A n'a plus rien à voir avec une « longueur » ou un « volume ».

2 Mesures positives, fonctions mesurables

Dans ce chapitre nous introduisons les éléments de base de la théorie de l'intégration de Lebesgue, sur un ensemble de départ quelconque E , et pour une mesure μ quelconque. On définira la notion de sous-ensemble mesurable, avant celle de la mesure. Les sous-ensembles mesurables forment une famille satisfaisant des propriétés axiomatiques spécifiques, qu'on appelle une *tribu*, ou *σ -algèbre*.

Cette présentation « axiomatique » de la théorie de la mesure trouvera son intérêt dans un cas particulier, celui de la mesure de Lebesgue sur l'espace euclidien $E = \mathbb{R}^d$, qui généralise la notion de volume euclidien. On construira en détail la tribu des sous-ensembles de \mathbb{R}^d sur laquelle cette mesure est définie, qu'on appellera la tribu de Lebesgue.

A une tribu de E , on pourra associer une famille de fonctions à valeurs réelles sur E , les fonctions mesurables. Une fois choisie une mesure μ définie sur cette tribu, on pourra alors définir l'intégrale d'une fonction mesurable. On donnera les théorèmes fondamentaux de cette théorie de l'intégration : les théorèmes de convergence monotone et de convergence dominée.

2.1 Ensembles mesurables, tribu

On se donne un ensemble E quelconque, sur lequel seront définies nos fonctions. Une tribu est une famille de sous-ensembles de E , satisfaisant certaines propriétés axiomatiques.

Définition 2.1. Soit E un ensemble quelconque. Une tribu (aussi appelée *σ -algèbre*) sur E est une famille $\mathcal{A} \subset \mathcal{P}(E)$ de parties de E , telle que :

- (i) l'ensemble E fait partie de la tribu \mathcal{A}
- (ii) si $A \in \mathcal{A}$, alors son complémentaire $A^c \in \mathcal{A}$. (une tribu est stable par prise du complémentaire)
- (iii) Si $(A_n)_{n \in \mathbb{N}}$ est une famille d'éléments de \mathcal{A} , alors leur union $\bigcup_{n \in \mathbb{N}} A_n$ appartient aussi à \mathcal{A} . On dit qu'une tribu est stable par *union dénombrable*.

Une fois qu'on s'est donné une tribu \mathcal{A} sur E , les éléments de \mathcal{A} sont appelés les *ensembles mesurables* (ou \mathcal{A} -mesurables). Le couple (E, \mathcal{A}) est appelé un *espace mesurable*. On déduit facilement les propriétés suivantes de la tribu \mathcal{A} :

1. Par (i) et (ii), on déduit que l'ensemble vide \emptyset est dans \mathcal{A} .

2. Si $(A_n)_{n \in \mathbb{N}}$ est une famille d'éléments de \mathcal{A} , alors leur intersection $\bigcap_{n \in \mathbb{N}} A_n = (\bigcup_n A_n^c)^c$ est aussi dans \mathcal{A} .

3. Les A_n ne sont pas supposés distincts les uns des autres. Si on prend par exemple $A_n = \emptyset$ pour $n \geq N$, $\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n=0}^N A_n$ est effectivement une union finie. Donc \mathcal{A} est aussi stable par unions ou intersections finies.

Si on avait supposé les propriétés (i), (ii) et la stabilité uniquement par union finies, on aurait appelé \mathcal{A} une algèbre. Le fait d'autoriser des unions *dénombrables* est indiqué par le préfixe σ .

Exemple. Les exemples les plus simples de tribus sur un ensemble E sont :

1. la tribu complète $\mathcal{A} = \mathcal{P}(E)$.
2. la tribu minimale, ou triviale $\mathcal{A} = \{\emptyset, E\}$

Ces deux tribus ne sont pas très intéressantes, dans le sens qu'elles ne pourront pas porter les mesures qui vont nous intéresser.

Proposition 2.2. *Si \mathcal{A} et \mathcal{B} sont deux tribus sur E , alors leur intersection $\mathcal{A} \cap \mathcal{B} = \{A \subset E; A \in \mathcal{A} \text{ et } A \in \mathcal{B}\}$ est également une tribu.*

Démonstration. Facile, à partir de la définition d'une tribu. □

Cette propriété d'intersection permet de définir une tribu à partir d'une famille quelconque de parties de E :

Définition 2.3. Soit \mathcal{C} un sous-ensemble de $\mathcal{P}(E)$ (donc une famille de parties de E). Il existe alors une plus petite tribu sur E contenant \mathcal{C} . Cette tribu est unique, définie par

$$\sigma(\mathcal{C}) \stackrel{\text{def}}{=} \bigcap_{\mathcal{A} \supset \mathcal{C}} \mathcal{A}.$$

Dans le membre de droite on prend l'intersection sur toutes les tribus \mathcal{A} sur E qui contiennent \mathcal{C} . Il en existe au moins une, la tribu complète $\mathcal{P}(E)$. La tribu $\sigma(\mathcal{C})$ est appelée la *tribu engendrée par \mathcal{C}* .

Cette notion de tribu engendrée va constituer notre outil majeur pour fabriquer des tribus intéressantes. Il faut pour cela qu'on suppose que l'ensemble E possède des structures supplémentaires. Les cas qui vont nous occuper le plus souvent :

1. E est un espace topologique. Cela signifie qu'on a défini une topologie sur E , qui consiste à définir la famille $\mathcal{O} \subset \mathcal{P}(E)$ l'ensemble des ouverts de E . La tribu engendrée par \mathcal{O} , $\sigma(\mathcal{O})$, est appelée **la tribu des boréliens, ou tribu borélienne de E** . On la notera généralement $\mathcal{B}(E)$. Les éléments de cette tribu sont appelés les (sous-ensembles) boréliens de E . Cette tribu est un cas particulier très important dans ce qui suit.

2. Un cas particulier du cas précédent est celui où E est un espace métrique, donc muni d'une distance $d(\bullet, \bullet)$. Dans ce cas, cette distance permet de définir une topologie sur E , à partir des boules ouvertes $B(x, r) = \{y \in E; d(x, y) < r\}$.

3. Un cas particulier du cas précédent est celui de l'espace euclidien $E = \mathbb{R}^n$, muni de la distance euclidienne $d(x, y) = \|x - y\|$.

Remarque 2.4. On rappelle que pour tout espace topologique E , l'ensemble des ouverts \mathcal{O} contient toujours E et \emptyset , qu'il est stable par intersection finie et par union quelconque (pas forcément dénombrable). Notez les différences subtiles entre ces propriétés, et les propriétés ci-dessus vérifiées par une tribu.

Dans la suite du cours, lorsqu'on parlera de tribu sur un espace topologique E sans préciser (par exemple pour $E = \mathbb{R}^n$), il s'agira implicitement de la tribu des boréliens $\mathcal{B}(E)$.

Exercice 2.5. Montrer que la tribu des boréliens sur l'espace euclidien $E = \mathbb{R}$ est aussi engendrée par l'ensemble des intervalles $\{]-\infty, a[, a \in \mathbb{Q}\}$.

Indication : on pourra se servir du fait que \mathbb{Q} est dense dans \mathbb{R} , et que tout ouvert de \mathbb{R} est donné par une union finie ou dénombrable d'intervalles ouverts disjoints.

Sur la complexité de la tribu des boréliens

Soit E un espace topologique. La tribu des boréliens contient des ensembles assez « compliqués ». Par définition, la tribu $\mathcal{B}(E)$ contient tous les ouverts. Elle contient aussi les intersections dénombrables d'ouverts (qui ne sont en général pas des ouverts); on appelle G_δ l'ensemble de ces intersections dénombrables (ici, le symbole δ signifie « intersection dénombrable ». Par complémentarité, $\mathcal{B}(E)$ contient aussi les fermés, et les unions dénombrables de fermés (qui en général ne sont pas des fermés); on appelle F_σ ce type de sous-ensembles. On peut aussi considérer les unions dénombrables de compacts de E , notés K_σ . A partir de ces nouveaux types d'ensembles, on peut recommencer à prendre des intersections ou union dénombrables : on obtient par exemple des ensembles de type $G_{\delta\sigma}$, $G_{\delta\sigma\delta}$, etc.

Lorsque $E = \mathbb{R}$, on montre qu'à chaque nouvelle opération dénombrable, on obtient effectivement de nouveaux sous-ensembles. Ceci montre la complexité des boréliens, même sur $E = \mathbb{R}$.

Produit de deux espaces mesurables

Il est fréquent de considérer un espace produit cartésien de deux espaces E_1, E_2 . Si chacun de ces espaces est équipé d'une tribu, y a-t-il une tribu naturelle sur $E_1 \times E_2$? Oui, on l'appelle la tribu-produit.

Définition 2.6. Soit (E_1, \mathcal{A}_1) et (E_2, \mathcal{A}_2) deux espaces mesurables. La tribu-produit sur $E_1 \times E_2$ est la tribu définie par

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \stackrel{\text{def}}{=} \sigma(\{A_1 \times A_2; A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}).$$

Remarquons que cette tribu-produit n'est pas seulement constituée des produits cartésiens $A_1 \times A_2$, mais c'est la plus petite tribu contenant tous ces produits.

Exemple 2.7. Pour $E_1 = E_2 = \mathbb{R}$ munis de la tribu des boréliens, la tribu-produit $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ est égale à la tribu borélienne $\mathcal{B}(\mathbb{R}^2)$.

Pour deux espaces topologiques E_1, E_2 quelconques munis de leurs tribus boréliennes, la tribu-produit $\mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$ est toujours contenue dans $\mathcal{B}(E_1 \times E_2)$, mais la réciproque n'est pas toujours vraie.

2.2 Mesure positive définie sur une tribu

On a pour l'instant défini une tribu $\mathcal{A} \subset \mathcal{P}(E)$, qui contient par définition les ensembles *mesurables* de E . « Mesurable » signifie : « qui peut être mesuré ». Mesurer un sous-ensemble, c'est lui allouer un nombre positif, un « poids ». Ce poids doit satisfaire une propriété naturelle d'additivité vis-à-vis de l'union de deux sous-ensembles disjoints. La particularité de la définition ci-dessous, c'est que cette propriété d'additivité s'étend aux unions disjointes dénombrables.

Définition 2.8. Soit (E, \mathcal{A}) un ensemble mesurable. Une mesure positive sur cet ensemble est une application $\mu : \mathcal{A} \rightarrow [0, \infty]$ vérifiant les propriétés suivantes :

- (i) $\mu(\emptyset) = 0$;
- (ii) Pour toute famille $(A_n)_{n \in \mathbb{N}}$ de parties mesurables *disjointes deux à deux*,

$$\mu\left(\bigsqcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n) = \lim_{N \rightarrow \infty} \sum_{n=0}^N \mu(A_n). \quad (\sigma\text{-additivité})$$

L'espace mesurable (E, \mathcal{A}) équipé d'une mesure μ est appelé un *espace mesuré* (E, \mathcal{A}, μ) .

Remarquons que :

0. la mesure d'un ensemble peut valoir $+\infty$.
1. l'union du membre de gauche est bien dans \mathcal{A} , d'après la définition d'une tribu. On aurait pu écrire l'union $\bigcup_{n \in \mathbb{N}} A_n$, mais $\bigsqcup_{n \in \mathbb{N}}$ permet d'insister sur le caractère disjoint des A_n .
2. la limite dans le membre de droite est toujours bien définie, puisque tous ses termes sont ≥ 0 , de sorte que la somme est croissante par rapport à N ; cette limite peut être finie, ou valoir $+\infty$.
3. La propriété (ii) est appelée **σ -additivité** (comme plus haut, le préfixe σ indique le caractère dénombrable de l'union et de la somme). En prenant $A_n = \emptyset$ tous vides pour $n \geq n_0$, on vérifie la propriété d'additivité vis-à-vis d'une union finie de sous-ensembles disjoints A_0, \dots, A_{n_0} .
4. L'additivité ne sera pas vraie pour une union *indénombrable* d'ensembles disjoints. De toute façon, il est difficile de donner à un sens à une somme indénombrable de termes positifs ou nuls. Mais remarquons que la mesure de Lebesgue d'un intervalle $]a, b[$ avec $a < b$ n'est pas égale à la somme des mesures $\mu(\{x\}) = 0$ des points qui le constituent.

Remarque 2.9. Il est important de se souvenir que la mesure μ n'est en général pas définie sur tous les sous-ensembles de E , mais seulement sur ceux appartenant à la tribu \mathcal{A} . Nous montrerons plus tard que c'est le cas en particulier pour la mesure de Lebesgue sur $\mathcal{B}(\mathbb{R})$: il existe des sous-ensembles de \mathbb{R} dont on ne peut pas mesurer la « longueur ».

Propriétés de la mesure :

Tous les ensembles ci-dessous sont supposés être \mathcal{A} -mesurables.

1. Si $A \subset B$, alors $\mu(A) \leq \mu(B)$. Si de plus $\mu(A) < \infty$, alors $\mu(B \setminus A) = \mu(B) - \mu(A)$. (le cas $\mu(A) = \infty$ implique $\mu(B) = \infty$, et on ne peut pas donner un sens à $\infty - \infty$)
2. Pour tous $A, B \in \mathcal{A}$, alors $\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$.
3. Si les $(A_n)_{n \in \mathbb{N}}$ forment une famille croissante ($A_n \subset A_{n+1}$), alors

$$\mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) = \lim_{n \rightarrow \infty} \uparrow \mu(A_n),$$

où la notation $\lim \uparrow$ signifie une limite croissante.

4. Si les $(B_n)_{n \in \mathbb{N}}$ forment une famille décroissante ($B_n \supset B_{n+1}$) et si $\mu(B_0) < \infty$, alors

$$\mu \left(\bigcap_{n \in \mathbb{N}} B_n \right) = \lim_{n \rightarrow \infty} \downarrow \mu(B_n). \quad (2.1)$$

Le résultat reste vrai si $\mu(B_N) = 0$ pour un certain $N \geq 0$.

5. Pour une famille $(A_n)_{n \in \mathbb{N}}$ quelconque d'ensembles mesurables, $\mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) \leq \sum_{n \in \mathbb{N}} \mu(A_n)$.

Démonstration. 1 : on a $B = A \sqcup (B \setminus A)$, donc $\mu(B) = \mu(A) + \mu(B \setminus A)$.

2 : On rappelle les décompositions disjointes $A = (A \setminus B) \sqcup (A \cap B)$, $B = (B \setminus A) \sqcup (A \cap B)$, $A \cup B = (A \setminus B) \sqcup (A \cap B) \sqcup (B \setminus A)$. En passant aux mesures de ces ensembles, on trouve l'égalité.

3. A partir de la famille croissante $(A_n)_{n \in \mathbb{N}}$, on fabrique des ensembles disjoints $C_0 = A_0$, $C_n = A_n \setminus A_{n-1} \in \mathcal{A}$ pour $n \geq 1$, qui vérifient $A_n = \bigsqcup_{j=0}^n C_j$, donc $\mu(A_n) = \sum_{j=0}^n \mu(C_j)$, et $\mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) = \sum_{j=0}^{\infty} \mu(C_j) = \lim \uparrow \mu(A_n)$.

4. En prenant $A_n = B_0 \setminus B_n$, on obtient une suite $(A_n)_{n \in \mathbb{N}}$ croissante. En se servant des propriétés 1. et 3., et du fait que $\mu(B_0) < \infty$, on déduit

$$\begin{aligned} \mu(B_0) - \mu \left(\bigcap_{n \in \mathbb{N}} B_n \right) &= \mu \left(B_0 \setminus \bigcap_{n \in \mathbb{N}} B_n \right) = \mu \left(\bigcup_{n \in \mathbb{N}} A_n \right) = \lim \uparrow \mu(A_n) \\ &= \lim \uparrow (\mu(B_0) - \mu(B_n)) = \mu(B_0) - \lim \downarrow \mu(B_n). \end{aligned}$$

5. On se réduit à une famille d'ensembles disjoints en prenant $C_0 = A_0$, $C_1 = A_1 \setminus A_0$, et $C_n = A_n \setminus \left(\bigcup_{j=0}^{n-1} A_j \right)$ pour tout $n \geq 1$. Ces ensembles sont mesurables, disjoints et vérifient $\bigcup_{j=0}^n A_j = \bigsqcup_{j=0}^n C_j$, donc en prenant la limite $n \rightarrow \infty$, $\bigcup_{j=0}^{\infty} A_j = \bigsqcup_{j=0}^{\infty} C_j$, ce qui amène $\mu \left(\bigcup_{j=0}^{\infty} A_j \right) = \mu \left(\bigsqcup_{j=0}^{\infty} C_j \right) = \sum_{j=0}^{\infty} \mu(C_j) \leq \sum_{j=0}^{\infty} \mu(A_j)$. \square

Après avoir énoncé les propriétés générales d'une mesure, il est temps de présenter des exemples concrets.

Exemples de mesures

1. Pour $E = \mathbb{N}$ et la tribu complète $\mathcal{A} = \mathcal{P}(\mathbb{N})$, on peut considérer la mesure de comptage $\mu(A) = \#A$. On remarque que les ensembles $B_n = \{n, n+1, \dots\}$ forment une suite décroissante ($B_{n+1} \subset B_n$), avec $\mu(B_n) = \infty$ pour tout $n \geq 0$, mais $\bigcap_n B_n = \emptyset$. La

formule (2.1) est donc fausse dans ce cas, ce qui est dû au fait que tous les ensembles B_n sont de mesure infinie.

La mesure de comptage peut être définie pour n'importe quel ensemble mesurable de la forme $(E, \mathcal{P}(E))$, mais pour des ensembles indénombrables elle n'est pas très intéressante, puisqu'elle prend la valeur $+\infty$ sur la plupart des sous-ensembles.

2. Pour un ensemble mesurable quelconque (E, \mathcal{A}) et x un élément de E , on peut définir une mesure δ_x par :

$$\forall A \in \mathcal{A}, \quad \delta_x(A) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A. \end{cases}$$

En se servant des fonctions caractéristiques des ensembles $A \in \mathcal{A}$, on peut réécrire la définition de cette mesure par $\delta_x(A) = \mathbb{1}_A(x)$. On appelle δ_x la mesure de Dirac au point x .

Note historique : Paul Dirac était un physicien théoricien travaillant sur la mécanique quantique, il a utilisé la notation δ_x pour définir une « fonction singulière » sur \mathbb{R}^d qui serait nulle en tout point $y \neq x$, mais telle que son « intégrale » $\int \delta_x(y) dy = 1$. Les mathématiciens ont interprété sa « fonction singulière » comme une mesure définie sur la tribu des boréliens de \mathbb{R}^d , telle que $\delta_x(\{x\}) = 1$ mais $\delta_x(A) = 0$ si $x \notin A$.

3. Pour deux mesures μ_1, μ_2 définies sur une même tribu, et pour $\alpha_1, \alpha_2 \in [0, \infty]$, on peut définir la mesure combinaison linéaire $\mu = \alpha_1\mu_1 + \alpha_2\mu_2$.
4. Mesure de Lebesgue (par exemple sur $E = \mathbb{R}$). Nous en avons déjà parlé dans l'introduction du cours, c'est cette mesure qui permettra de faire le lien avec l'intégrale de Riemann. Cette mesure est définie sur la tribu des boréliens de \mathbb{R} , elle mesure la « longueur » des boréliens. On la note λ (ou parfois m , ou parfois dL , ou parfois dx). Comme on le verra dans l'Exemple 2.17, c'est l'unique mesure sur $\mathcal{B}(\mathbb{R})$ telle que, pour tout intervalle $]a, b[$, on ait $\lambda(]a, b[) = b - a$. La construction rigoureuse de cette mesure (qui semble pourtant très intuitive) va réclamer du travail, nous la ferons dans une section ultérieure. Cette mesure se généralise aux espaces euclidiens \mathbb{R}^d de toute dimension.

Les mesures peuvent être classifiées en plusieurs catégories.

Définition 2.10. 1. μ est dite *finie* si $\mu(E) < \infty$. Le nombre $\mu(E)$ est appelé la masse totale de μ . En particulier, si $\mu(E) = 1$, on appelle μ une *mesure de probabilité*.

2. Une mesure μ est dite *σ -finie* s'il existe une suite croissante $(E_n)_{n \in \mathbb{N}}$ de sous-ensembles mesurables, telle que $E = \bigcup_{n \in \mathbb{N}} E_n$, et $\mu(E_n) < \infty$ pour tout $n \in \mathbb{N}$. Cette définition inclut

formellement le cas où μ est finie, mais on l'utilise surtout dans des cas où $\mu(E) = \infty$. Par exemple, la mesure de comptage sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ est σ -finie, ce qu'on peut voir en prenant $E_n = \llbracket 0, n \rrbracket$. Par contre, si on considère $(E, \mathcal{P}(E))$ avec un ensemble E indénombrable, alors la mesure de comptage sur E n'est pas σ -finie.

3. Un point $x \in E$ est appelé un *atome* de la mesure μ si $\{x\} \in \mathcal{A}$ et $\mu(\{x\}) > 0$.
4. Une mesure μ est dite *diffuse* si elle n'a pas d'atomes.
5. Un ensemble $B \subset E$ tel qu'il existe $A \supset B$ mesurable avec $\mu(A) = 0$, est dit μ -négligeable. Si la mesure μ est sous-entendue, on dira plus simplement que l'ensemble B est négligeable. On dit que la tribu \mathcal{A} est complète (relativement à la mesure μ) si tous les ensembles μ -négligeables sont dans \mathcal{A} . Il est possible de compléter la tribu \mathcal{A} , de manière à obtenir une tribu $\bar{\mathcal{A}} \supset \mathcal{A}$, complète pour μ .

Exemple 2.11. La mesure de Lebesgue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est une mesure diffuse : pour tout $x \in \mathbb{R}$, $\lambda(\{x\}) = 0$. Par σ -additivité, pour tout sous-ensemble $A \subset \mathbb{R}^d$ dénombrable, on aura aussi $\lambda(A) = 0$. En particulier, l'ensemble des rationnels $\mathbb{Q} \subset \mathbb{R}$ est de mesure de Lebesgue nulle.

On peut aussi construire des sous-ensembles indénombrables de \mathbb{R} négligeables pour la mesure de Lebesgue. Par exemple, l'ensemble de Cantor $Can \subset [0, 1]$, composé des nombres réels pouvant s'écrire $x = \sum_{n \geq 1} \frac{k_n}{3^n}$ avec tous les $k_n \in \{0, 2\}$, forme un borélien sur \mathbb{R} , qui est indénombrable. On peut montrer que $\lambda(Can) = 0$.

Exercice 2.12. Sur \mathbb{R}^2 , considérons un morceau fini courbe lisse \mathcal{C} . Montrer qu'elle est de mesure de Lebesgue nulle.

Indication : construire des recouvrements de \mathcal{C} par des unions de petits cubes.

2.3 Lemme de classe monotone : identification d'une mesure

Avant de passer à la définition de l'intégrale, nous introduisons la structure de *classe monotone*, une notion proche de celle de la tribu, qui va nous être utile pour identifier une mesure μ sur une tribu \mathcal{A} (par ex. la mesure de Lebesgue sur $\mathcal{B}(\mathbb{R})$) en ne connaissant que ses valeurs sur une sous-classe $\mathcal{C} \subset \mathcal{A}$ satisfaisant certaines propriétés.

Définition 2.13. Une sous-famille \mathcal{M} de $\mathcal{P}(E)$ est appelée une classe monotone si :

- (i) $E \in \mathcal{M}$;
- (ii) Si $A, B \in \mathcal{M}$ et $A \subset B$, alors $B \setminus A \in \mathcal{M}$;

(iii) Si les $A_n \in \mathcal{M}$ forment une suite croissante ($A_n \subset A_{n+1}$), alors $\bigcup_n A_n \in \mathcal{M}$.

En comparant ces conditions avec celles d'une tribu (Def. 2.1), on voit qu'une tribu est un cas particulier de classe monotone. D'une part, (i) et (ii) impliquent que pour tout $A \in \mathcal{A}$, $A^c \in \mathcal{A}$. Par contre, la condition (iii) ci-dessus est moins contraignante que dans le cas d'une tribu, puisque qu'elle n'est valable que si on considère une suite (A_n) croissante, tandis que pour une tribu. Les classes monotones sont donc « plus nombreuses » que les tribus parmi toutes les sous-familles de $\mathcal{P}(E)$.

Lemme 2.14. *Une classe monotone est une tribu si et seulement si elle est stable par intersections finies.*

Démonstration. Supposons qu'une classe monotone \mathcal{M} est invariante par intersections finies. Comme \mathcal{M} est invariante par passage au complémentaire, elle sera aussi invariante par unions finies. En se servant de la propriété (iii), on montre l'invariance par union dénombrable ; en effet, pour toute suite $(B_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{M} , on peut construire $A_0 = B_0$, $A_1 = A_0 \cup B_1$, et similairement $A_n = A_{n-1} \cup B_n$. Grâce à l'invariance par union finie, on montre alors par récurrence que $A_n \in \mathcal{M}$. De plus, on a forcément $A_n \supset A_{n-1}$, donc la suite (A_n) est croissante. Le point (iii) montre donc que $\bigcup_n A_n = \bigcup_n B_n$ est dans \mathcal{M} . On a donc montré que \mathcal{M} est une tribu □

Si \mathcal{M}_1 et \mathcal{M}_2 sont deux classes monotones, la famille $\mathcal{M}_1 \cap \mathcal{M}_2$ l'est aussi. On peut donc, comme pour les tribus, définir la plus petite classe monotone contenant une sous-famille \mathcal{C} de $\mathcal{P}(E)$; on appelle cette classe $\mathcal{M}(\mathcal{C})$ la classe monotone engendrée par \mathcal{C} :

$$\mathcal{M}(\mathcal{C}) \stackrel{\text{def}}{=} \bigcap_{\mathcal{M} \text{ cl. monot.}, \mathcal{M} \supset \mathcal{C}} \mathcal{M}.$$

Le principal résultat de cette section est le Lemme de classe monotone. Sa preuve utilise déjà les méthodes qui seront utilisées pour construire la mesure de Lebesgue, ce que nous ferons dans un chapitre ultérieur.

Théorème 2.15. *[Lemme de classe monotone] Si $\mathcal{C} \subset \mathcal{P}(E)$ est stable par intersections finies, alors la classe monotone engendrée $\mathcal{M}(\mathcal{C})$ est égale à la tribu $\sigma(\mathcal{C})$.*

Démonstration. Comme les tribus sont toutes des classes monotones, la classe monotone engendrée par \mathcal{C} est incluse dans la tribu engendrée : $\mathcal{M}(\mathcal{C}) \subset \sigma(\mathcal{C})$. On aura égalité si et seulement si on montre que $\mathcal{M}(\mathcal{C})$ est une tribu. D'après le lemme 2.14, pour montrer que

$\mathcal{M}(\mathcal{C})$ est une tribu, il suffit de montrer que $\mathcal{M}(\mathcal{C})$ est stable par intersections finies (on a juste fait l'hypothèse que \mathcal{C} est stable par intersections finies).

Pour un élément $A \in \mathcal{M}(\mathcal{C})$ quelconque, on définit la famille

$$\mathcal{M}_A \stackrel{\text{def}}{=} \{B \in \mathcal{M}(\mathcal{C}) ; A \cap B \in \mathcal{M}(\mathcal{C})\}. \quad (2.2)$$

Notre objectif est alors de montrer que $\mathcal{M}_A = \mathcal{M}(\mathcal{C})$.

Pour commencer, on va supposer que $A \in \mathcal{C}$. Comme \mathcal{C} est stable par intersection, pour tout $B \in \mathcal{C}$ on aura $A \cap B \in \mathcal{C} \subset \mathcal{M}(\mathcal{C})$, ce qui montre que $B \in \mathcal{M}_A$; donc la famille $\mathcal{C} \subset \mathcal{M}_A$. L'astuce consiste à montrer que \mathcal{M}_A est une classe monotone :

- (i) Comme $A \cap E = A$, on a évidemment $E \in \mathcal{M}_A$.
- (ii) Supposons $B \subset B'$ deux éléments de \mathcal{M}_A : on aura alors $(A \cap B) \subset (A \cap B')$ tous deux dans $\mathcal{M}(\mathcal{C})$. Par conséquent, $(A \cap B') \setminus (A \cap B)$ est aussi dans $\mathcal{M}(\mathcal{C})$. Or ce sous-ensemble s'écrit aussi $A \cap (B' \setminus B)$: on a donc montré que $A \cap (B' \setminus B) \in \mathcal{M}(\mathcal{C})$, donc que $B' \setminus B \in \mathcal{M}_A$.
- (iii) Prenons maintenant une suite croissante $(B_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{M}_A . Les ensembles $(A \cap B_n)$ sont alors dans $\mathcal{M}(\mathcal{C})$, et ils forment une suite croissante ; leur union $\bigcup_n (A \cap B_n) = A \cap (\bigcup_n B_n)$ est donc dans $\mathcal{M}(\mathcal{C})$. Par conséquent, $\bigcup_n B_n \in \mathcal{M}_A$.

On a montré que pour $A \in \mathcal{C}$, la famille \mathcal{M}_A satisfait les propriétés d'une classe monotone. Comme $\mathcal{C} \subset \mathcal{M}_A$ et que $\mathcal{M}(\mathcal{C})$ est la classe monotone la plus petite contenant \mathcal{C} , on en déduit $\mathcal{M}_A = \mathcal{M}(\mathcal{C})$. Par conséquent, nous avons montré que :

$$\forall A \in \mathcal{C}, \forall B \in \mathcal{M}(\mathcal{C}), A \cap B \in \mathcal{M}(\mathcal{C}). \quad (2.3)$$

Pour conclure, il faut élargir cette propriété à tout élément $A \in \mathcal{M}(\mathcal{C})$. On reprend alors la même stratégie, en définissant, pour un $A' \in \mathcal{M}(\mathcal{C})$ donné, la famille $\mathcal{M}_{A'}$ comme ci-dessus. La propriété (2.3) que nous venons de montrer, écrite en remplaçant (A, B) par (B, A') , montre que tout élément $B \in \mathcal{C}$ est dans $\mathcal{M}_{A'}$, de sorte que $\mathcal{C} \subset \mathcal{M}_{A'}$. La même preuve que ci-dessus montre que $\mathcal{M}_{A'}$ est une classe monotone, et donc qu'elle est égale à $\mathcal{M}(\mathcal{C})$. On a donc montré que cette dernière est invariante par intersection finie, ce qui achève la preuve du théorème. \square

La principale conséquence de ce lemme est de permettre d'identifier deux mesures lorsqu'elles sont égales sur une sous-classe $\mathcal{C} \subset \mathcal{A}$.

Corollaire 2.16. Soient μ, ν deux mesures sur un espace mesurable (E, \mathcal{A}) . Supposons qu'il existe une classe $\mathcal{C} \subset \mathcal{A}$ stable par intersection finie, telle que $\sigma(\mathcal{C}) = \mathcal{A}$, et telle que pour tout $A \in \mathcal{C}$, on ait $\mu(A) = \nu(A)$. Si une des conditions suivantes est vérifiée :

1. $\mu(E) = \nu(E) < \infty$;
 2. μ et ν sont σ -finies par rapport à une suite croissante $(E_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{C} (c'est-à-dire $E = \bigcup_n E_n$, $\mu(E_n) = \nu(E_n) < \infty$) ;
- alors $\mu = \nu$.

Démonstration. 1. Commençons par le cas de μ, ν finies. Encore une fois, il faut définir une classe astucieusement. On pose $\mathcal{G} \stackrel{\text{def}}{=} \{A \in \mathcal{A} ; \mu(A) = \nu(A)\}$. D'une part, cette classe contient \mathcal{C} ; d'autre part, on vérifie facilement que \mathcal{G} est une classe monotone :

- (i) il est évident que $E \in \mathcal{G}$.
- (ii) si $A \subset B$ sont dans \mathcal{G} , alors $\mu(B \setminus A) = \mu(B) - \mu(A)$ (car $\mu(A) < \infty$), et de même pour ν . On en déduit que $\mu(B \setminus A) = \nu(B \setminus A)$, ce qui montre que $B \setminus A \in \mathcal{G}$.
- (iii) si $(A_n)_{n \geq 0}$ est une suite croissante dans \mathcal{G} , alors $\mu(\bigcup_n A_n) = \lim \uparrow \mu(A_n) = \lim \uparrow \nu(A_n) = \nu(\bigcup_n A_n)$, donc $(\bigcup_n A_n) \in \mathcal{G}$.

Comme \mathcal{C} est stable par intersection finie, le Lemme de classe monotone montre que $\mathcal{A} = \sigma(\mathcal{C}) = \mathcal{M}(\mathcal{C})$. Comme la classe monotone \mathcal{G} contient \mathcal{C} , elle contient $\mathcal{M}(\mathcal{C})$, donc on a forcément $\mathcal{G} = \mathcal{A}$, ce qui revient à dire que $\mu = \nu$ sur la tribu \mathcal{A} entière.

2. Considérons le 2e cas. Pour tout n on définit les mesures $\mu_n(A) = \mu(A \cap E_n)$, $\nu_n(A) = \nu(A \cap E_n)$, (on appelle μ_n la restriction de μ à l'ensemble E_n). Pour tout $A \in \mathcal{C}$ l'ensemble $A \cap E_n$ est aussi dans \mathcal{C} (car \mathcal{C} est stable par intersection), ce qui implique que $\mu_n(A) = \nu_n(A)$. Les mesures μ_n et ν_n satisfont donc l'hypothèse 1. ; d'après la preuve ci-dessus, elle sont donc égales. On montre enfin que pour tout $A \in \mathcal{A}$,

$$\mu(A) = \lim_n \uparrow \mu_n(A) = \lim_n \uparrow \nu_n(A) = \nu(A).$$

□

Exemple 2.17. *Unicité de la mesure de Lebesgue.* Nous n'avons pas encore construit la mesure de Lebesgue sur \mathbb{R} , mais nous avons déjà indiqué une condition qu'elle doit satisfaire : pour tout intervalle borné $]a, b[$, $\lambda(]a, b[) = b - a$. Si \mathcal{C} est la famille de ces intervalles, on a déjà expliqué que la tribu des boréliens est engendrée par cette famille \mathcal{C} . D'autre part, cette famille est stable par intersection finie. Enfin, la suite $(E_n =]-n, n[)_{n \geq 1}$ satisfait bien

$E_n \in \mathcal{C}$, elle est croissante, et $\bigcup_n E_n = \mathbb{R}$. Le corollaire ci-dessus montre alors qu'il existe au plus une seule mesure satisfaisant ces propriétés. Nous montrerons dans un chapitre ultérieur que cette mesure existe effectivement sur toute la tribu $\mathcal{B}(\mathbb{R})$.

Une seconde application sera très utile lorsqu'on s'intéressera aux mesures de probabilité.

Exemple 2.18. *Fonction de répartition d'une mesure finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.* Soit μ une mesure finie sur \mathbb{R} . Pour identifier μ , il suffit de connaître ses valeurs sur la famille $\{]-\infty, a[, a \in \mathbb{R}\}$, ou bien ses valeurs sur la famille $\{]-\infty, a] , a \in \mathbb{R}\}$. Par convention, en théorie des probabilités on utilise plutôt la seconde famille : la fonction $a \mapsto \mu(]-\infty, a])$ est alors appelée la fonction de répartition de μ .

2.4 Fonctions mesurables (entre deux tribus)

Après avoir défini la classe des sous-ensembles mesurables (= la tribu), on peut définir une classe de fonctions reliant deux espaces mesurables, qui sont compatibles avec cette structure mesurable. On les appelle les fonctions mesurables.

Remarque 2.19. Il faudra faire attention au mot « mesurable », qui s'applique à différents objets de natures différentes :

- un espace mesurable (E, \mathcal{A}) , c'est-à-dire un espace muni d'une tribu
- un ensemble mesurable $A \in \mathcal{A}$, élément de la tribu dont il est question
- une fonction mesurable, notion qui fait référence à deux tribus \mathcal{A} et \mathcal{B} .

2.4.1 Définition et premières propriétés des fonctions mesurables

Définition 2.20. Soient (E, \mathcal{A}) et (F, \mathcal{B}) deux espaces mesurables. Une application $f : E \rightarrow F$ est dite mesurable si

$$\forall B \in \mathcal{B}, \quad f^{-1}(B) \in \mathcal{A}.$$

Dans le cas où E et F sont des espaces topologiques munis de leurs tribus boréliennes, on dit qu'une telle fonction f est borélienne. Très souvent l'espace image sera \mathbb{R} , \mathbb{C} ou \mathbb{R}^d (fonction à valeurs réelles, complexes ou vectorielles), dans ce cas \mathcal{B} sera la tribu des boréliens.

Remarque 2.21. On ne demande pas que l'image d'un sous-ensemble mesurable soit mesurable, mais que l'image réciproque d'un mesurable le soit. La situation est analogue au cas d'une fonction entre deux espaces topologiques : la fonction est continue si l'image réciproque

de tout ouvert est un ouvert.

Dans notre introduction, on avait rencontré les ensembles $f^{-1}([y_{k-1}, y_k[)$ dans la définition intuitive de l'intégrale de Lebesgue, avec $f : I \rightarrow \mathbb{R}$; on remarque que les intervalles $[y_{k-1}, y_k[$ sont des boréliens de \mathbb{R} . Il faut donc s'assurer que les $f^{-1}([y_{k-1}, y_k[)$ sont aussi des boréliens.

Proposition 2.22. *La composition de deux fonctions mesurables est encore mesurable.*

Démonstration. Immédiat si on remarque que $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$. \square

En général il est difficile de vérifier la propriété pour tous les éléments de la tribu (par exemple, on a vu la tribu des boréliens sur \mathbb{R} contient des ensembles compliqués). Heureusement, la proposition suivante montre que pour prouver la mesurabilité de f , il suffit de se restreindre à une sous-famille de \mathcal{B} qui engendre la tribu.

Proposition 2.23. *Supposons que la tribu \mathcal{B} est engendrée par une sous-famille $\mathcal{C} \subset \mathcal{B}$. Pour vérifier que $f : E \rightarrow F$ est mesurable, il suffit de vérifier que $f^{-1}(B) \in \mathcal{A}$ pour tout B élément de la famille \mathcal{C} .*

Démonstration. On considère la sous-famille de la tribu \mathcal{B} définie par $\mathcal{G} \stackrel{\text{def}}{=} \{B \in \mathcal{B}; f^{-1}(B) \in \mathcal{A}\}$. On voit que \mathcal{G} a les propriétés d'une tribu :

- (i) $f^{-1}(\emptyset) = \emptyset$,
- (ii) $f^{-1}(B^c) = (f^{-1}(B))^c$. On remarque que cette propriété est fausse si on remplace f^{-1} par f . Cela explique pourquoi la mesurabilité est définie à partir de l'image réciproque, et non de l'image directe.
- (iii) $f^{-1}(\bigcup B_n) = \bigcup f^{-1}(B_n)$.

Supposons maintenant que pour tout B dans la famille \mathcal{C} , $f^{-1}(B) \in \mathcal{A}$. Cela implique que la sous-tribu \mathcal{G} contient la famille \mathcal{C} . Donc, par définition, \mathcal{G} contient la tribu engendrée par \mathcal{C} , qui n'est autre que \mathcal{B} par hypothèse; comme $\mathcal{G} \subset \mathcal{B}$, on en déduit donc $\mathcal{G} = \mathcal{B}$, autrement dit : $f^{-1}(B) \in \mathcal{A}$ pour tout $B \in \mathcal{B}$. On a montré que f est mesurable entre les tribus \mathcal{A} et \mathcal{B} . \square

Corollaire 2.24. *Soit une fonction $f : (E, \mathcal{A}) \rightarrow \mathbb{R}$; comme expliqué plus haut, on considèrera implicitement que la tribu associée à \mathbb{R} est la tribu borélienne $\mathcal{B}(\mathbb{R})$. Donc on est dans le cas $F, \mathcal{B} = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Comme la tribu $\mathcal{B}(\mathbb{R})$ est engendrée par les intervalles $] \alpha, +\infty[$, pour vérifier que f est mesurable (entre \mathcal{A} et $\mathcal{B}(\mathbb{R})$), il suffit donc de montrer que*

$$\forall \alpha \in \mathbb{R}, \quad f^{-1}(] \alpha, +\infty[) \in \mathcal{A}.$$

On aurait pu choisir comme autre critère, par exemple $f^{-1}(]-\infty, \alpha[) \in \mathcal{A}$, ou bien $f^{-1}(]-\infty, \alpha]) \in \mathcal{A}$.

Ce corollaire nous fournit le résultat important suivant, qui relie les propriétés topologiques aux propriétés de mesurabilité :

Corollaire 2.25. *Si E, F sont deux espaces topologiques munis de leurs tribus boréliennes, alors une application $f : E \rightarrow F$ continue est aussi mesurable (entre $\mathcal{B}(E)$ et $\mathcal{B}(F)$).*

Démonstration. On sait que pour $f : E \rightarrow F$ continue, pour tout ouvert $V \subset F$, l'image réciproque $f^{-1}(V)$ est un ouvert de E , en particulier c'est un borélien de E . Comme (par définition de la tribu borélienne) l'ensemble des ouverts de F engendre $\mathcal{B}(F)$, la proposition 2.23 montre que f est mesurable. \square

Tribu définie à partir d'une famille de fonctions

On peut procéder dans le sens inverse : pour un espace E donné, on peut partir d'une famille \mathcal{F} de fonctions $f : E \rightarrow \mathbb{R}$. Y a-t-il alors une tribu \mathcal{A} sur E , telle que toutes les fonctions $f \in \mathcal{F}$ sont mesurables de \mathcal{A} vers $\mathcal{B}(\mathbb{R})$?

Proposition 2.26. *Si on se donne une famille \mathcal{F} de fonctions $f : E \rightarrow \mathbb{R}$, il existe une tribu minimale $\mathcal{A} = \mathcal{A}_{\mathcal{F}}$ sur E , telle que toutes les fonctions $f \in \mathcal{F}$ sont mesurables.*

Démonstration. Considérons la famille de sous-ensembles de E définie par

$$\mathcal{R}_{\mathcal{F}} \stackrel{\text{def}}{=} \{f^{-1}(]-\infty, \alpha[) \subset E, f \in \mathcal{F}, a \in \mathbb{R}\},$$

et soit $\mathcal{A}_{\mathcal{F}} \stackrel{\text{def}}{=} \sigma(\mathcal{R}_{\mathcal{F}})$ la tribu sur E engendrée par $\mathcal{R}_{\mathcal{F}}$. Comme les intervalles $]-\infty, \alpha[$, $\alpha \in \mathbb{R}$ engendrent les boréliens de \mathbb{R} , et que pour toute fonction $f \in \mathcal{F}$, les images réciproques $f^{-1}(]-\infty, \alpha[)$ sont dans $\mathcal{R}_{\mathcal{F}}$, donc dans la tribu $\mathcal{A}_{\mathcal{F}}$, alors la Proposition 2.23 montre que cette fonction f est mesurable sur $\mathcal{A}_{\mathcal{F}}$. On voit aussi que la tribu $\mathcal{A}_{\mathcal{F}}$ est la plus petite ayant cette propriété : n'importe quelle tribu rendant les f mesurables doit forcément contenir $\mathcal{R}_{\mathcal{F}}$, donc contenir la tribu $\mathcal{A}_{\mathcal{F}} = \sigma(\mathcal{R}_{\mathcal{F}})$. \square

Exemple 2.27. Si $E = F = \mathbb{R}$ et $\mathcal{F} = C^0(\mathbb{R}, \mathbb{R})$ la famille des fonctions continues $\mathbb{R} \rightarrow \mathbb{R}$, alors la plus petite tribu rendant ces fonctions mesurables est la tribu des boréliens de \mathbb{R} .

Démonstration. Pour tout intervalle ouvert $]a, b[$, il existe une fonction continue $f : \mathbb{R}^d \rightarrow \mathbb{R}$ telle que $f^{-1}(]-\infty, 0]) =]a, b[$. On peut par exemple prendre la fonction à symétrie sphérique $f(x) = |x - \frac{a+b}{2}| - \frac{b-a}{2}$. Une tribu \mathcal{A} rendant les fonctions f mesurable va donc contenir tous les intervalles $]a, b[$. On a déjà vu qu'une telle tribu doit alors contenir tous les boréliens de \mathbb{R} . D'un autre côté, le corollaire 2.25 montre que les fonctions continues sont mesurables $\mathcal{B}(\mathbb{R}) \rightarrow \mathcal{B}(\mathbb{R})$, donc la tribu \mathcal{A} minimale satisfaisant cette propriété est bien la tribu des boréliens. □

2.4.2 Opérations sur les fonctions mesurables

A partir de deux fonctions *scalaires* (= à valeurs dans \mathbb{R}) $f_1, f_2 : E \rightarrow \mathbb{R}$, on obtient une fonction à valeurs vectorielles $f = (f_1, f_2) : E \rightarrow \mathbb{R}^2$. Qu'en est-il de leur mesurabilité ?

Proposition 2.28. *Soient $f_1 : (E, \mathcal{A}) \mapsto (F_1, \mathcal{B}_1)$ et $f_2 : (E, \mathcal{A}) \mapsto (F_2, \mathcal{B}_2)$ deux fonctions mesurables. Alors l'application produit $f : (E, \mathcal{A}) \mapsto (F_1 \times F_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$ définie par $f(x) = (f_1(x), f_2(x))$ est aussi mesurable. La réciproque est vraie : si $f = (f_1, f_2)$ est mesurable, chacune de ses composantes l'est aussi.*

Démonstration. Par la définition 2.6, la tribu $\mathcal{B}_1 \otimes \mathcal{B}_2$ est engendrée par la famille des $\mathcal{C} = \{B_1 \times B_2, B_i \in \mathcal{B}_i, i = 1, 2\}$. D'après la Proposition 2.23, pour vérifier que f est mesurable il suffit de montrer que les ensembles $f^{-1}(B_1 \times B_2)$ sont dans \mathcal{A} . Or que valent ces ensembles ?

$$\begin{aligned} f^{-1}(B_1 \times B_2) &= \{x \in E ; f(x) \in B_1 \times B_2\} = \{x \in E ; f_1(x) \in B_1 \text{ et } f_2(x) \in B_2\} \\ &= \{x \in E ; f_1(x) \in B_1\} \cap \{x \in E ; f_2(x) \in B_2\} \\ &= f_1^{-1}(B_1) \cap f_2^{-1}(B_2). \end{aligned}$$

Comme f_1 et f_2 sont mesurables, les ensembles $f_i^{-1}(B_i)$ sont dans \mathcal{A} , donc leur intersection l'est aussi. Donc f est bien mesurable.

Inversement, si f est supposée mesurable, alors pour tout $B_1 \in \mathcal{B}_1$, $f^{-1}(B_1 \times F_2) = f_1^{-1}(B_1) \cap f_2^{-1}(F_2) = f_1^{-1}(B_1) \cap E$ est dans \mathcal{A} ; cela montre que f_1 est mesurable. En échangeant les indices, on montre que f_2 est également mesurable. □

Ce résultat sur la mesurabilité de la fonction vectorielle $f = (f_1, f_2)$ va nous permettre des opérations sur les fonctions scalaires.

Corollaire 2.29. Soient $f, g : (E, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ deux fonctions mesurables. Soient $\alpha, \beta \in \mathbb{R}$. Alors les fonctions

$$\alpha f + \beta g, \quad f^+ = \max(f, 0), \quad f^- = \max(-f, 0) \quad \text{sont mesurables.}$$

Démonstration. On se sert du Lemme précédent. On sait que la fonction $x \mapsto (f(x), g(x))$ est mesurable de \mathcal{A} dans $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$. D'autre part, l'application

$$\begin{aligned} \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (a, b) &\mapsto \alpha a + \beta b \end{aligned}$$

est continue, donc elle est mesurable $\mathcal{B}(\mathbb{R}^2) \rightarrow \mathcal{B}(\mathbb{R})$ (cf. le corollaire 2.25). La composée de ces deux applications mesurables est donc mesurable.

Pour les deux autres énoncés, il suffit de remarquer que les applications $\mathbb{R} \rightarrow \mathbb{R}$ données par $a \mapsto \max(a, 0)$, respectivement $a \mapsto \max(-a, 0)$ sont aussi continues, donc mesurables. \square

2.4.3 Suites de fonctions mesurables à valeurs dans $\overline{\mathbb{R}}$

On va maintenant considérer des suites de fonctions. Même si chaque f_n est à valeurs dans \mathbb{R} , il se peut qu'en certains points x , les valeurs $(f_n(x))_n$ admettent comme valeur d'adhérence $+\infty$ ou $-\infty$. On va donc étendre \mathbb{R} en $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\} = [-\infty, +\infty]$, qu'on appelle $\overline{\mathbb{R}}$ la *droite réelle achevée*, et on supposera que les fonctions f_n prennent déjà leurs valeurs dans $\overline{\mathbb{R}}$. Arrêtons-nous un peu sur la topologie de cette droite achevée. $\overline{\mathbb{R}}$ est homéomorphe à l'intervalle $[-1, 1]$, par exemple à travers la fonction $x \mapsto \tanh(x)$. Ainsi, les ouverts de $\overline{\mathbb{R}}$ incluent les ouverts de \mathbb{R} , mais également les intervalles du type $[-\infty, a[$ ou $]a, +\infty]$. À partir de cette topologie, on construit la tribu borélienne $\mathcal{B}(\overline{\mathbb{R}})$.

Exercice 2.30. Montrer que la tribu $\mathcal{B}(\overline{\mathbb{R}})$ est engendrée par les intervalles $\{[-\infty, a[, a \in \mathbb{R}\}$.

On va maintenant s'intéresser aux opérations sur les suites des fonctions. Soit $(a_n)_{n \in \mathbb{N}}$ une suite d'éléments de $\overline{\mathbb{R}}$. Alors la limite supérieure (respectivement limite inférieure) de cette suite sont définies par :

$$\begin{aligned} \limsup_{n \rightarrow \infty} a_n &= \lim_{n \rightarrow \infty} \downarrow \left(\sup_{k \geq n} a_k \right) = \inf_{n \geq 0} \left(\sup_{k \geq n} a_k \right), \\ \text{respectivement} \quad \liminf_{n \rightarrow \infty} a_n &= \lim_{n \rightarrow \infty} \uparrow \left(\inf_{k \geq n} a_k \right) = \sup_{n \geq 0} \left(\inf_{k \geq n} a_k \right). \end{aligned} \quad (2.4)$$

La \limsup et la \liminf existent pour toute suite $(a_n)_{n \geq 0}$, ce sont les plus grande et plus petite valeurs d'adhérence de la suite. La suite $(a_n)_{n \in \mathbb{N}}$ admet une limite en $n \rightarrow \infty$ si et seulement si $\liminf a_n = \limsup a_n$ (la limite est alors égale à la valeur commune). Notons que ces limites supérieure et inférieure prennent leurs valeurs dans $\overline{\mathbb{R}}$.

On s'intéresse maintenant à des suites de fonctions mesurables.

Proposition 2.31. *Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables de (E, \mathcal{A}) vers $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Alors les fonctions*

$$\sup_n f_n, \quad \inf_n f_n, \quad \limsup_n f_n, \quad \liminf_n f_n \quad \text{sont aussi mesurables.}$$

En particulier, si la suite $(f_n)_{n \in \mathbb{N}}$ converge simplement vers une fonction $f : E \rightarrow \overline{\mathbb{R}}$, alors f est mesurable.

Enfin, pour une suite $(f_n)_{n \in \mathbb{N}}$ quelconque, le sous-ensemble de E ,

$$\{x \in E ; f_n(x) \text{ admet une limite lorsque } n \rightarrow \infty\},$$

est mesurable.

Cette proposition nous montre que la notion de mesurabilité s'accommode bien de suites de fonctions admettant une limite *simple*. C'est un des grands atouts de la théorie de Lebesgue.

Démonstration. Définissons la fonction $f \stackrel{\text{def}}{=} \inf_n f_n : E \rightarrow \overline{\mathbb{R}}$. En adaptant le Corollaire 2.24 au cas de $\overline{\mathbb{R}}$, on voit que pour montrer que f est mesurable il suffit de montrer que pour tout $\alpha \in \mathbb{R}$, $f^{-1}([-\infty, \alpha[)$ est mesurable. Remarquons que $\inf_n f_n(x) < \alpha$ implique qu'il existe un $n_0 \in \mathbb{N}$ tel que $f_{n_0}(x) < \alpha$. Ainsi,

$$\begin{aligned} f^{-1}([-\infty, \alpha[) &= \left\{x \in E ; \inf_n f_n(x) < \alpha\right\} = \bigcup_n \{x \in E ; f_n(x) < \alpha\} \\ &= \bigcup_n f_n^{-1}([-\infty, \alpha[). \end{aligned}$$

Par hypothèse chaque $f_n^{-1}([-\infty, \alpha[)$ est mesurable, donc leur union l'est également. Pour traiter le cas de $\sup_n f_n$, on peut noter que $\sup_n f_n = -(\inf_n (-f_n))$.

En combinant ces résultats, on obtient que $g_n \stackrel{\text{def}}{=} (\sup_{k \geq n} f_k)$ est mesurable pour tout $n \geq 0$, et donc que $\inf_n g_n = \limsup_k f_k$ est mesurable.

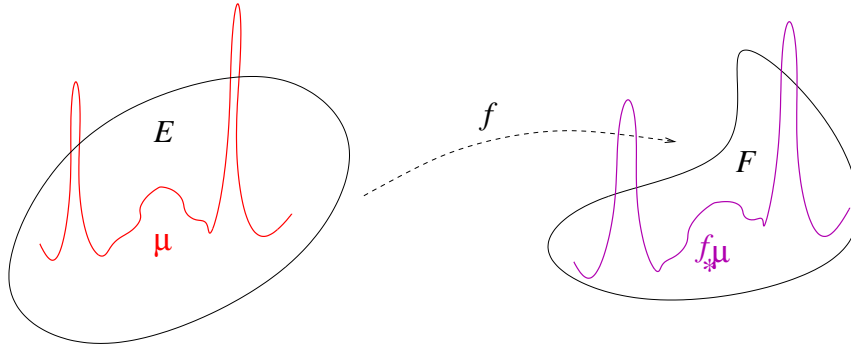


FIGURE 2.1 – Mesure image par une application $f : E \rightarrow F$. Les courbes rouge et violette schématisent des mesures sur E et sur F .

Le sous-ensemble de la dernière assertion peut être défini par l'égalité $\limsup_n f_n(x) = \liminf_n f_n(x)$. Les deux fonctions $\limsup_n f_n$ et $\liminf_n f_n$ sont mesurables, donc l'application $g(x) \stackrel{\text{def}}{=} \limsup_n f_n(x) - \liminf_n f_n(x)$ l'est aussi. L'ensemble que nous cherchons vaut $g^{-1}(\{0\})$, il est mesurable puisque $\{0\}$ est un borélien. \square

2.4.4 Transport d'une mesure par une application

Une façon de fabriquer une mesure sur un espace mesurable (F, \mathcal{B}) est de partir d'une mesure μ sur un espace mesurable (E, \mathcal{A}) , et de la « pousser » vers F au moyen d'une application mesurable $f : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$.

Définition 2.32. Soit $f : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$ une application mesurable, et soit μ une mesure positive sur (E, \mathcal{A}) . La mesure image de μ par f , notée $f_*\mu$ (ou parfois $f(\mu)$), est la mesure positive sur (F, \mathcal{B}) définie par :

$$\forall B \in \mathcal{B}, \quad f_*\mu(B) = \mu(f^{-1}(B)).$$

On vérifie facilement que μ est une mesure grâce aux propriétés suivantes :

- $f^{-1}(\emptyset) = \emptyset$, donc $f_*\mu(\emptyset) = 0$
- $f^{-1}(\bigsqcup_n B_n) = \bigsqcup_n f^{-1}(B_n)$.

La mesure image $f_*\mu$ aura la même masse totale que μ . Par contre, il est possible que μ soit σ -finie, sans que $f_*\mu$ le soit.

Exercice 2.33. On prend la mesure de Lebesgue λ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, et on considère des fonctions $f : \mathbb{R} \rightarrow \mathbb{R}$ mesurables (par rapport à la tribu borélienne sur \mathbb{R}). Décrire la mesure $f_*\lambda$ dans le cas des applications suivantes :

1. $f(x) = 0$ pour tout $x \in \mathbb{R}$. La mesure $f_*\mu$ est-elle σ -finie ?
2. $f(x) = 2x$ pour tout $x \in \mathbb{R}$.

3 Intégration d'une fonction par rapport à une mesure

Nous avons défini une mesure sur une tribu de sous-ensembles mesurables ; nous avons défini ce qu'est une fonction mesurable. Il ne reste plus qu'à combiner les deux notions, en définissant l'intégrale d'une fonction mesurable par rapport à cette mesure. Comme dans le cas de l'intégrale de Riemann, on commencera par définir l'intégrale sur une famille particulière de fonctions : les fonctions étagées, qui sont mesurables mais ne prennent qu'un nombre fini de valeurs. Ceci nous amènera à la définition de l'intégrale (de Lebesgue) d'une fonction mesurable *positive*. La notion de fonction intégrable nous permettra d'étendre la définition aux fonctions changeant de signe.

Ce chapitre nous permettra de présenter les principaux théorèmes de convergence pour l'intégrale de Lebesgue, par rapport à une suite de fonctions mesurables (f_n) convergeant *simplement* vers une fonction limite :

1. Le théorème de convergence monotone
2. Le Lemme de Fatou
3. Le théorème de convergence dominée, qui illustre la « puissance » de la théorie de Lebesgue sur celle de Riemann.

Tout au long de ce chapitre, la mesure d'intégration sera quelconque.

3.1 Intégration d'une fonction mesurable positive

Contrairement à la définition de l'intégrale de Riemann, l'intégrale de Lebesgue d'une présente une forme d'asymétrie, en cela qu'on cherche des fonctions approximantes par valeurs inférieures à f . C'est pourquoi on la définition concerne tout d'abord des fonctions positives.

3.1.1 Fonctions étagées et définition de l'intégrale de fonctions positives

Comme indiqué dans l'introduction, l'idée sera d'approximer f par des *fonctions étagées* (en anglais on dit « simple functions »).

Définition 3.1. Soit (E, \mathcal{A}) un espace mesurable. Une fonction $f : (E, \mathcal{A}) \rightarrow \mathbb{R}$ est dite étagée si elle ne prend qu'un nombre fini de valeurs distinctes $\alpha_1 < \alpha_2 < \dots < \alpha_n$, et que pour chaque valeur α_j , l'ensemble $A_j = f^{-1}(\{\alpha_j\})$ est \mathcal{A} -mesurable. Autrement dit,

$$f(x) = \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j}(x), \quad A_j \in \mathcal{A}, \quad E = \bigsqcup_{j=1}^n A_j. \quad (3.1)$$

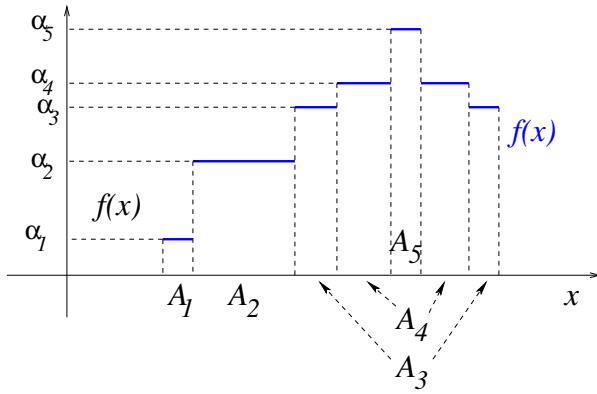


FIGURE 3.1 – Fonction étagée

Cette représentation de f est appelée son *écriture canonique* (on montre facilement qu'elle est unique).

Donnons-nous maintenant une mesure μ définie sur la tribu \mathcal{A} . On peut alors définir l'intégrale d'une fonction étagée, relativement à cette mesure.

Définition 3.2. Supposons la fonction étagée f à valeurs dans \mathbb{R}_+ , définie sur l'espace mesuré (E, \mathcal{A}, μ) . L'intégrale de f par rapport à μ est définie par :

$$\int f d\mu := \sum_{j=1}^n \alpha_j \mu(A_j).$$

On prend la convention que si $\alpha_j = 0$ et $\mu(A_j) = \infty$, alors $\alpha_j \mu(A_j) = 0$. A priori, l'intégrale prend ses valeurs dans $[0, \infty]$.

Ecritures non canoniques d'une fonction étagée

Plus généralement, une fonction étagée f peut être définie par une partition mesurable $E = \bigsqcup_{i=1}^N B_i$, et une suite de valeurs réelles β_1, \dots, β_N , non nécessairement distinctes. On définit alors

$$f = \sum_{i=1}^N \beta_i \mathbb{1}_{B_i}.$$

Pour retrouver l'écriture canonique (3.1), il suffit de regrouper ensemble les β_i qui sont égaux à un α_j donné : les indices i correspondants forment un sous-ensemble $I(\alpha_j) \subset \llbracket 1, N \rrbracket$. On a alors $A_j = \bigsqcup_{i \in I(\alpha_j)} B_i$, et donc $\mathbb{1}_{A_j} = \sum_{i \in I(\alpha_j)} \mathbb{1}_{B_i}$. L'additivité de la mesure μ implique alors

que $\mu(A_j) = \sum_{i \in I(\alpha_j)} \mu(B_i)$, de sorte que

$$\int f d\mu = \sum_{i=1}^N \beta_i \mu(B_i).$$

Ainsi, la formule ci-dessus pour $\int f d\mu$ n'est pas seulement valable pour sa décomposition canonique de la fonction étagée f , mais pour n'importe quelle décomposition (finie).

Proposition 3.3. *Soient f et g deux fonctions étagées positives sur l'espace mesuré (E, \mathcal{A}, μ) .*

1. *Pour tous réels $a, b \geq 0$, $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$.*
2. *Si $f \leq g$, alors $\int f d\mu \leq \int g d\mu$.*

Démonstration. 1. La fonction $af + bg$ est une fonction étagée positive. En effet, si les fonctions f, g se décomposent en $f = \sum_{i=1}^N \beta_i \mathbb{1}_{B_i}$ et $g = \sum_{i=1}^{N'} \beta'_i \mathbb{1}_{B'_i}$, alors les ensembles $\{C_{ij} = B_i \cap B'_j, i = 1, \dots, N, j = 1, \dots, N'\}$ forment une partition de E , sur laquelle f et g peuvent toutes deux se décomposer (on dit que la partition $\sqcup C_{ij}$ est plus « fine » que les partitions $\sqcup B_i$ et $\sqcup B'_j$). On pourra donc écrire $f = \sum_{ij} \beta_i \mathbb{1}_{C_{ij}}$, $g = \sum_{ij} \beta'_j \mathbb{1}_{C_{ij}}$, et donc

$$af + bg = \sum_{ij} \gamma_{ij} \mathbb{1}_{C_{ij}}, \text{ avec les coefficients } \gamma_{ij} = a\beta_i + b\beta'_j.$$

Les fonctions $f, g, af + bg$ sont toutes trois définies sur la même partition $\sqcup C_{ij}$, on peut facilement relier leurs intégrales :

$$\begin{aligned} \int (af + bg) d\mu &= \sum_{ij} \gamma_{ij} \mu(C_{ij}) = \sum_{ij} (a\beta_i + b\beta'_j) \mu(C_{ij}) \\ &= a \sum_i \sum_j \beta_i \mu(C_{ij}) + b \sum_j \sum_i \beta'_j \mu(C_{ij}) \\ &= a \sum_i \beta_i \mu(B_i) + b \sum_j \beta'_j \mu(B'_j). \end{aligned}$$

2. Si $f \geq g$, la fonction $f - g$ est étagée positive, et $f = g + (f - g)$. On retombe sur le cas précédent. Comme $\int (f - g) d\mu \geq 0$, on obtient le résultat. \square

Notations

\mathcal{E}_+ est l'espace des fonctions étagées positives. La proposition ci-dessus utilise l'invariance cet espace par la somme de fonctions, et par leur multiplication par un nombre réel positif.

On dira qu'une fonction f est *mesurable positive* sur (E, \mathcal{A}) si elle est mesurable, et qu'elle prend ses valeurs dans $\overline{\mathbb{R}}_+ = [0, \infty]$.

On peut maintenant définir l'intégrale d'une fonction mesurable positive, à partir de celle des fonctions étagées.

Définition 3.4. [Intégrale d'une fonction mesurable positive] Soit (E, \mathcal{A}, μ) un espace mesuré, et soit $f : (E, \mathcal{A}) \rightarrow \overline{\mathbb{R}}_+$ une fonction mesurable positive. L'intégrale de f est définie par

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

Selon le contexte, l'intégrale d'une fonction peut être notée indifféremment $\int f d\mu = \int f(x) d\mu(x) = \mu(f) = \langle \mu, f \rangle$ (ou parfois $\int f$ lorsque la mesure est évidente).

Proposition 3.5. 1. Si f est une fonction étagée, toute fonction étagée $h \leq f$ vérifiera $\int h d\mu \leq \int f d\mu$, tandis que l'égalité sera atteinte en prenant $h = f$. La définition ci-dessus est donc bien compatible avec celle d'une fonction étagée.

2. Si $f \leq g$ sont mesurables positives, alors la définition implique directement que $\int f d\mu \leq \int g d\mu$.

3. Si f est mesurable positive et $\mu(\{x \in E; f(x) > 0\}) = 0$, alors $\int f d\mu = 0$. En effet, cette hypothèse implique que $\mu(\{x \in E; h(x) > 0\}) = 0$ pour toute fonction étagée positive $h \leq f$, ce qui implique que $\int h d\mu = 0$. On peut noter que cette hypothèse sur f autorise à ce que $f(x) = \infty$ sur un sous-ensemble $A \subset E$ (avec nécessairement $\mu(A) = 0$).

Remarque 3.6. La propriété 3. ci-dessus montre que les valeurs de f sur un sous-ensemble $A \subset E$ de mesure nulle ne sont pas pertinentes dans le calcul de l'intégrale : ces valeurs sont « peu importantes ». Ceci nous amène à la définition d'un ensemble négligeable.

Définition 3.7. [Ensembles négligeables - propriété vraie presque partout] On dira qu'un ensemble $A \subset E$ est μ -négligeable s'il existe un $B \in \mathcal{A}$ tel que $A \subset B$ et $\mu(B) = 0$. Si la mesure μ est implicite, on dira que A est un ensemble négligeable. On dira qu'une propriété est vérifiée μ -presque partout (souvent abrégé par « μ -p.p. », ou « p.p. » lorsque la mesure est connue) si elle est vraie en-dehors d'un sous-ensemble négligeable. On dit aussi qu'une telle propriété est *presque sure*.

La plupart des énoncés de la théorie de l'intégration feront référence à des propriétés vraies p.p.

3.1.2 Le théorème de convergence monotone

La définition de l'intégrale se fait par approximations inférieures de f . Cette définition débouche rapidement sur le premier théorème de convergence pour des suites de fonctions.

Théorème 3.8. [*Théorème de convergence monotone*] Soit $(f_n)_{n \in \mathbb{N}}$ une suite croissante de fonctions mesurables positives (à valeurs dans $[0, \infty]$), et définissons $f = \sup_n f_n = \lim_n \uparrow f_n$. Alors, pour toute mesure μ sur \mathcal{A} , on a :

$$\int f d\mu = \lim_n \uparrow \int f_n d\mu.$$

Remarquons que la limite $f(x) = \lim_n \uparrow f_n(x)$ est une limite simple, pas forcément une limite uniforme.

Démonstration. On sait que déjà que $f = \sup f_n$ est mesurable et positive, donc l'intégrale $\int f d\mu$ est bien définie. Comme (f_n) est croissante, la suite $(\int f_n d\mu)_n$ est également croissante, cf. Proposition 3.5 (2). Comme $f \geq f_n$ pour tout $n \geq 0$, on a aussi $\int f d\mu \geq \int f_n d\mu$, et donc

$$\int f d\mu \geq \lim_n \uparrow \int f_n d\mu.$$

Pour montrer l'égalité, il nous faut montrer l'inégalité inverse. Choisissons une fonction étagée positive $h \leq f$. Fixons $0 < \epsilon < 1$. Comme $f_n(x) \rightarrow f(x)$, on définit pour tout n l'ensemble $E_n = \{x \in E : (1 - \epsilon)h(x) \leq f_n(x)\}$. Comme la suite $(f_n)_{n \in \mathbb{N}}$ est croissante, la suite d'ensembles $(E_n)_n$ est également croissante, et comme $f_n(x) \rightarrow f(x)$, pour tout x on va avoir, à partir d'un certain rang, $f_n(x) \geq (1 - \epsilon)h(x)$; ceci montre que $\bigcup_{n=0}^{\infty} E_n = E$.

Par définition de E_n , on a $f_n \geq (1 - \epsilon)\mathbb{1}_{E_n} h$, de sorte que $\int f_n d\mu \geq (1 - \epsilon) \int \mathbb{1}_{E_n} h d\mu$. Souvenons-nous que h est une fonction étagée : $h = \sum_{k=1}^K \alpha_k \mathbb{1}_{A_k}$, donc

$$\int \mathbb{1}_{E_n} h d\mu = \sum_{k=1}^K \alpha_k \mu(A_k \cap E_n).$$

Comme les $(E_n)_n$ recouvrent tout l'ensemble E , on a $\lim_n \uparrow \mu(A_k \cap E_n) = \mu(A_k)$. On trouve donc, à la limite $n \rightarrow \infty$:

$$\lim_n \uparrow \int f_n d\mu \geq (1 - \epsilon) \lim_n \uparrow \sum_{k=1}^K \alpha_k \mu(A_k \cap E_n) = (1 - \epsilon) \int h d\mu.$$

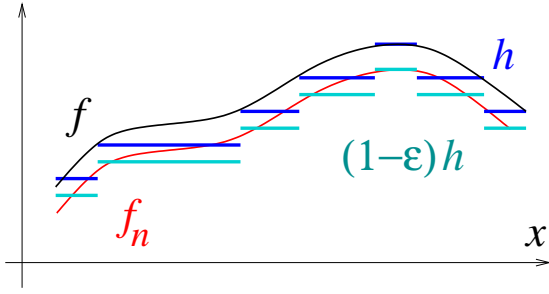


FIGURE 3.2 – Les fonctions $f, f_n, h, (1 - \epsilon)h$ apparaissant dans la preuve du théorème de convergence monotone

Cette équation est vraie pour tout $\epsilon > 0$, donc on peut prendre $\epsilon \rightarrow 0$ et on trouve

$$\lim_n \uparrow \int f_n d\mu \geq \int h d\mu.$$

Cette inégalité est vraie pour toute fonction étagée h minorant f , donc on peut prendre le supremum sur toutes ces fonctions $h \leq f$, et on trouve bien le résultat. \square

Exemple 3.9. Revenons à la fonction $f = \mathbb{1}_{\mathbb{Q} \cap I}$ de l'Exemple 1.6, qui n'admettait pas d'intégrale de Riemann. On avait montré que $f = \lim_n F_n$, avec $F_n = \mathbb{1}_{\frac{1}{n!}\mathbb{N} \cap I}$. La suite $(F_n)_{n \geq 1}$ est croissante, et les fonctions F_n sont étagées par rapport à la tribu de Borel, telles que $\int F_n d\lambda = 0$, puisque l'ensemble $\{x \in I, F_n(x) > 0\}$ est λ -négligeable. Le théorème ci-dessus montre donc que $\int f d\lambda = 0$. On obtient le même résultat en notant que l'ensemble $\{x \in I, f(x) > 0\} = \mathbb{Q} \cap I$ est dénombrable, donc il est borélien, et négligeable pour la mesure λ qui est diffuse.

Le théorème suivant montre qu'on peut approximer (au sens de la convergence simple, et de façon croissante) une fonction mesurable positive par des fonctions étagées.

Théorème 3.10. *[Approximation par des fonctions étagées] Soit f une fonction mesurable positive. Alors il existe une suite croissante $(f_n)_{n \in \mathbb{N}}$ de fonctions positives étagées, telles que $\lim_n \uparrow f_n(x) = f(x)$ pour tout $x \in E$.*

Démonstration. Notons que toute fonction étagée est nécessairement bornée, tandis que f peut valoir $+\infty$ sur un ensemble de E . On va construire les approximations f_n en tronquant leurs valeurs dans l'intervalle $[0, n]$, et en découpant cet intervalle en petits sous-intervalles

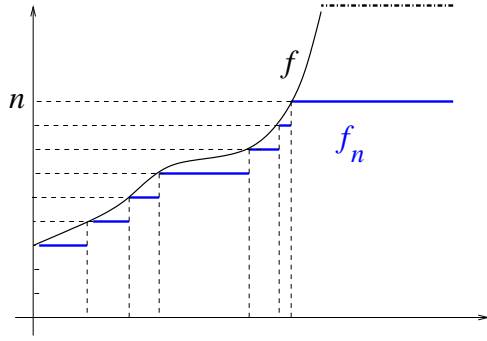


FIGURE 3.3 – Approximation d'une fonction mesurable positive par des fonctions étagées.

de longueur 2^{-n} . Par image réciproque, on définit donc les boréliens :

$$\begin{aligned} A_\infty^{(n)} &= f^{-1}([n, \infty[), \\ A_i^{(n)} &= f^{-1}\left(\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right]\right), \quad i = 0, \dots, n2^n - 1, \end{aligned}$$

et la fonction étagée associée

$$f_n = \sum_{i=0}^{n2^n-1} \frac{i}{2^n} \mathbb{1}_{A_i^{(n)}} + n \mathbb{1}_{A_\infty^{(n)}}.$$

Cette fonction minore clairement f . De plus, on voit facilement que la partition en intervalles $\bigsqcup_i \left[\frac{i}{2^{n+1}}, \frac{i+1}{2^{n+1}}\right[\sqcup [n+1, \infty[$ est plus fine que la partition $\bigsqcup_i \left[\frac{i}{2^n}, \frac{i+1}{2^n}\right[\sqcup [n, \infty[$, ce qui implique que la fonction $f_{n+1} \geq f_n$.

Il reste à montrer que $f_n \nearrow f$. Sur $A^{(n)} := \bigsqcup_i A_i^{(n)} = f^{-1}([0, n])$ on a l'encadrement $f_n(x) \leq f(x) \leq f_n(x) + \frac{1}{2^n}$. Les ensembles $A^{(n)}$ sont croissants, et leur union $\bigsqcup_n A^{(n)} = f^{-1}(\mathbb{R}_+)$, ce qui montre que pour tout $x \in f^{-1}(\mathbb{R}_+)$, $f_n(x) \nearrow f(x)$. Pour tout x sur l'ensemble restant $f^{-1}(\infty) = \bigcap_n A_\infty^{(n)}$, on aura $f_n(x) \geq n$, donc $f_n(x) \nearrow \infty = f(x)$. On a donc montré que $f_n \nearrow f$ sur tout E . \square

Equipés de ce théorème d'approximation, on peut maintenant généraliser les propriétés de « linéarité positive » de l'intégrale des fonctions étagées.

Proposition 3.11. 1. Soient f, g des fonctions mesurables positives, et $a, b \geq 0$ deux réels positifs. Alors

$$\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu. \quad (3.2)$$

2. Si $(f_n)_{n \in \mathbb{N}}$ est une suite quelconque de fonctions mesurables positives, alors

$$\int \left(\sum_{n=0}^{\infty} f_n \right) d\mu = \sum_{n=0}^{\infty} \int f_n d\mu. \quad (3.3)$$

Autrement dit, pour toute suite de fonctions positives, on peut intervertir la somme et l'intégrale.

Attention : la formule (2) est valable car toutes les fonctions f_n sont positives. Elle ne sera pas vraie pour des fonctions f_n générales.

Démonstration. 1. Remarquons que la relation de linéarité (3.2) n'est pas complètement évidente à partir de la définition de l'intégrale. Grâce au théorème d'approximation, on peut approcher f (respectivement g) par une suite croissante de fonctions étagées $(f_n)_{n \in \mathbb{N}}$, et g par une suite similaire $(g_n)_{n \in \mathbb{N}}$. On aura alors automatiquement $af_n(x) + bg_n(x) \rightarrow af(x) + bg(x)$, et la suite $(af_n + bg_n)_n$ est croissante et faite de fonctions positives étagées. Le théorème de convergence monotone montre donc que

$$\int f_n d\mu \xrightarrow{n \rightarrow \infty} \int f d\mu, \quad \int g_n d\mu \xrightarrow{n \rightarrow \infty} \int g d\mu, \quad \int (af_n + bg_n) d\mu \xrightarrow{n \rightarrow \infty} \int (af + bg) d\mu.$$

D'autre part, la Prop.3.3 montre que pour les fonctions étagées,

$$\int (af_n + bg_n) d\mu = a \int f_n d\mu + b \int g_n d\mu,$$

ce qui prouve le résultat.

2. La première partie (linéarité) montre l'égalité pour les sommes finies $F_N = \sum_{n=0}^N f_n$. Ensuite, il suffit de remarquer que $F_N \nearrow F_\infty = \sum_{n=0}^{\infty} f_n$, en particulier F_∞ est bien mesurable. Le théorème de convergence monotone montre donc que

$$\int F_\infty d\mu = \lim_N \uparrow \int F_N d\mu = \lim_N \uparrow \sum_{n=0}^N \int f_n d\mu = \sum_{n=0}^{\infty} \int f_n d\mu.$$

La somme de droite est composée de termes positifs, donc elle a bien une limite. \square

Exercice 3.12. Retrouver le fait que, pour une suite double $(a_{n,k})_{n,k \in \mathbb{N}}$ de nombres positifs, on a l'égalité des sommes doubles

$$\sum_{k \in \mathbb{N}} \left(\sum_{n \in \mathbb{N}} a_{n,k} \right) = \sum_{n \in \mathbb{N}} \left(\sum_{k \in \mathbb{N}} a_{n,k} \right).$$

Une mesure μ peut être pondérée par une fonction positive f , pour former une nouvelle mesure $\nu = f \cdot \mu$.

Définition 3.13. [Mesure modulée par une fonction densité] Soit μ une mesure sur (E, \mathcal{A}) et f une fonction mesurable positive. Alors la mesure $\nu = f \cdot \mu$ est définie comme suit :

$$\forall A \in \mathcal{A}, \quad \nu(A) = \int \mathbb{1}_A f \, d\mu := \int_A f \, d\mu.$$

On peut interpréter la fonction f comme une « densité », qui vient moduler la mesure μ . Par exemple la mesure μ pourrait représenter le nombre de particules (atomes) par unité de volume, et f indiquer la masse de chaque atome, qui dépend de l'espèce chimique dont il s'agit.

Démonstration. Il faut tout de même vérifier, dans la définition ci-dessus, que ν définit bien une mesure sur la tribu \mathcal{A} . On a bien $\nu(\emptyset) = 0$. Si (A_n) forme une suite disjointe de parties mesurables, alors $\nu(\bigsqcup_n A_n) = \int (\sum_n \mathbb{1}_{A_n}) f \, d\mu$ peut se réécrire, grâce à la formule d'échange (3.3),

$$\nu\left(\bigsqcup_n A_n\right) = \sum_n \int \mathbb{1}_{A_n} f \, d\mu = \sum_n \nu(A_n).$$

□

Remarque 3.14. Notons que la mesure ν obtenue par cette modulation n'est pas quelconque : elle vérifie la propriété suivante :

$$\forall A \in \mathcal{A}, \quad \text{si } \mu(A) = 0, \quad \text{alors forcément } \nu(A) = 0.$$

Autrement dit, un ensemble A négligeable pour la mesure μ sera aussi négligeable pour ν . Ceci reste vrai si la fonction f prend la valeur $+\infty$ sur l'ensemble A . (on rappelle la règle $0 \cdot \infty = 0$).

On a défini plus haut la notion d'ensemble négligeable pour une mesure μ , et inversement de propriété vraie μ -presque partout. Voici quelques inégalités très utiles, qui utilisent cette terminologie.

Proposition 3.15. Soit f une fonction mesurable positive sur l'espace mesuré (E, \mathcal{A}, μ) .

1. Pour tout réel $a > 0$,

$$\mu(\{x \in E : f(x) \geq a\}) \leq \frac{1}{a} \int f \, d\mu. \quad (3.4)$$

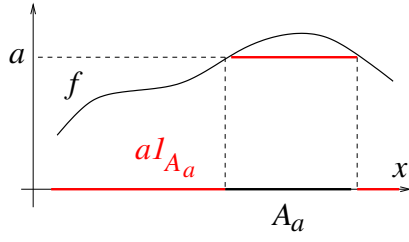


FIGURE 3.4 – Inégalité de Markov

Cette inégalité est très utile en théorie des probabilités, où elle prend le nom d'inégalité de Markov.

2. $\int f d\mu < \infty \implies f < \infty \quad p.p. \text{ (ou } \mu - p.p.)$.
3. $\int f d\mu = 0 \iff f = 0 \quad p.p.$
4. Si g est une autre fonction mesurable positive, alors $g = f \quad p.p. \implies \int f d\mu = \int g d\mu$.

Cette dernière propriété montre que, dans le cadre de l'intégration, il n'est pas nécessaire de connaître une fonction en tout point (« partout ») pour déterminer ses propriétés d'intégration, mais uniquement « presque partout ». Ceci nous amènera à définir les classes d'équivalences de fonctions égales entre elles presque partout.

Démonstration. 1. Notons $A_a = f^{-1}([a, \infty]) = \{x \in E; f(x) \geq a\}$. Comme f est mesurable, A_a l'est aussi. On remarque l'inégalité évidente (« tautologie ») $f \geq a\mathbb{1}_{A_a}$. La positivité de l'intégrale implique alors

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

2. Notons, pour tout $n \geq 1$, $A_n = f^{-1}([n, \infty])$, et $A_\infty = f^{-1}(\{\infty\})$. On remarque que la suite $(A_n)_n$ est décroissante, et que $\bigcap_n A_n = A_\infty$, de sorte que

$$\mu(A_\infty) = \mu\left(\bigcap_n A_n\right) = \lim_n \downarrow \mu(A_n).$$

D'après l'inégalité (3.4), $\mu(A_n) \leq \frac{1}{n} \int f d\mu$, donc si f est d'intégrale finie, on aura $\lim_n \downarrow \mu(A_n) = 0$.

3. On a déjà montré l'implication \Leftarrow dans la Proposition 3.5. Pour montrer \Rightarrow , on va utiliser les ensembles $B_n := f^{-1}([\frac{1}{n}, \infty])$. Ils forment une suite croissante, et leur union $\bigcup_{n \in \mathbb{N}} B_n = f^{-1}(]0, \infty])$. L'inégalité de Markov (3.4) implique $\mu(B_n) \leq n \int f d\mu$ pour tout

$n \geq 1$, donc d'après l'hypothèse, $\mu(B_n) = 0$. Comme Par sous-additivité de la mesure, $\mu(f^{-1}([0, \infty])) = 0$, ce qui revient à dire que f est nulle presque partout.

4. Remarquons qu'on ne sait pas encore intégrer des fonctions changeant de signe, donc on ne peut pas directement essayer de calculer $\int (f - g) d\mu$. On va plutôt considérer les fonctions mesurables $f \vee g := \max(f, g)$ et $f \wedge g := \min(f, g)$. En effet, ces fonctions satisfont $f \vee g \geq f \wedge g$, de sorte que $f \vee g - f \wedge g$ est une fonction mesurable positive. D'autre part, l'hypothèse $f = g$ p.p. implique que $(f \vee g - f \wedge g) = 0$ p.p. Ceci implique que

$$\int f \vee g d\mu = \int f \wedge g d\mu + \int (f \vee g - f \wedge g) d\mu = \int f \wedge g d\mu.$$

Enfin, on a les inégalités $f \wedge g \leq f \leq f \vee g$ et de même en remplaçant f par g . Ces encadrements, avec l'égalité ci-dessus, montrent donc que

$$\int f d\mu = \int f \vee g d\mu \quad \text{et} \quad \int g d\mu = \int f \vee g d\mu,$$

donc ces deux intégrales sont égales. □

Le théorème de convergence monotone concernait exclusivement des suites croissantes de fonctions. Le second résultat important, le Lemme¹ de Fatou, supprime cette hypothèse de monotonie..

Théorème 3.16. [*Lemme de Fatou*] Soit (E, \mathcal{A}, μ) un espace mesuré, et $(f_n)_{n \in \mathbb{N}}$ une suite quelconque de fonctions mesurables positives. Alors

$$\int \left(\liminf_{n \rightarrow \infty} f_n \right) d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu.$$

On remarque que, d'après le théorème de convergence monotone, si la suite $(f_n)_n$ est croissante, alors $\liminf_n f_n = \lim_n f_n$, et on a égalité dans l'expression ci-dessus.

Démonstration. Par définition, $\liminf_n f_n = \lim_n \uparrow (\inf_{k \geq n} f_k)$. Les fonctions $F_n := (\inf_{k \geq n} f_k)$ sont mesurables et forment une suite croissante, donc d'après le théorème de convergence monotone,

$$\lim_n \uparrow \int F_n d\mu = \int \left(\lim_n \uparrow F_n \right) d\mu = \int \left(\liminf_n f_n \right) d\mu.$$

1. En général, un Lemme est un résultat intermédiaire, qui n'a pas d'intérêt en soi. Il se trouve que certains Lemmes sont devenus célèbres, car ils ont eu des conséquences inattendues sur le reste de la théorie. On devrait donc plutôt parler de « Théorème de Fatou », mais on a gardé la dénomination historique de « Lemme ».



FIGURE 3.5 – Illustration du Lemme de Fatou

D'autre part, on a forcément $F_n \leq f_n$, donc $\int F_n d\mu \leq \int f_n d\mu$; en passant à la \liminf_n on obtient

$$\liminf_n \int F_n d\mu \leq \liminf_n \int f_n d\mu.$$

Enfin, on se souvient que la suite $(\int F_n d\mu)_n$ est croissante, de sorte que $\liminf_n \int F_n d\mu = \lim_n \int F_n d\mu$. En combinant avec l'inégalité et l'égalité ci-dessus, on obtient le résultat. \square

Exercice 3.17. 1. On se donne les fonctions $f_n := \mathbb{1}_{[n, n+1]}$ sur l'espace mesuré $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. Calculer les deux côtés de l'inégalité dans le Lemme de Fatou, et vérifier qu'on n'a pas égalité. Dans ce cas, l'inégalité stricte est due au fait que le support de f_n « part à l'infini » lorsque $n \rightarrow \infty$.

2. On définit $f_n = f_0$ si n est pair, $f_n = f_1$ si n est impair, avec $f_0 = \mathbb{1}_{[0, 1/2]}$ et $f_1 = \mathbb{1}_{[1/2, 1]}$. Calculer alors les deux côtés du Lemme de Fatou, et vérifier qu'on a inégalité stricte. Ici, l'inégalité stricte est due à l'« oscillation » du support de f_n .

3. On définit $f_n = n\mathbb{1}_{[0, 1/n]}$. Montrer qu'on a encore une inégalité stricte dans le Lemme de Fatou.

3.2 Fonctions intégrables

Jusqu'à présent on a considéré des fonctions mesurables positives, qui pouvaient prendre la valeur $+\infty$ sur n'importe quel sous-ensemble (mesurable), et dont l'intégrale pouvait prendre la valeur $+\infty$. Afin d'étendre l'intégration aux fonctions prenant des valeurs négatives, ou aux fonctions changeant de signe, on va être obligé d'imposer une condition d'intégrabilité; en effet, on doit éviter de se retrouver face à des expressions du type $+\infty - (+\infty)$.

3.2.1 Intégrale d'une fonction réelle intégrable

Définition 3.18. Soit (E, \mathcal{A}, μ) un espace mesuré, et soit $f : (E, \mathcal{A}) \rightarrow \mathbb{R}$ une fonction mesurable (pas forcément positive). On dit que f est *intégrable par rapport à μ* (ou μ -

intégrable, ou intégrable si la mesure est sous-entendue) si

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

Dans ce cas, on définit l'intégrale de f par

$$\int f d\mu := \int f^+ d\mu - \int f^- d\mu, \quad (3.5)$$

où $f^+ = \max(f, 0)$ et $f^- = \max(-f, 0)$ sont appelées respectivement la partie positive et la partie négative de f . On notera que f^+ et f^- sont mesurables, et que $f = f^+ - f^-$, respectivement $|f| = f^+ + f^-$ (en particulier $|f|$ est bien mesurable et positif).

On note $\mathcal{L}^1(E, \mathcal{A}, \mu)$ l'espace des fonctions intégrables. Notons que cet espace dépend de la mesure μ , et pas uniquement de la tribu \mathcal{A} . On notera $\mathcal{L}_+^1(E, \mathcal{A}, \mu)$ l'espace des fonctions intégrables à valeurs positives.

Remarque 3.19. 1. On a forcément $f^+ \leq |f|$ et $f^- \leq |f|$, donc l'hypothèse d'intégrabilité impose que $\int f^+ d\mu$ et $\int f^- d\mu$ sont finies, de sorte que la différence (3.5) a bien un sens, et donne un nombre fini.

2. Si f est positive, on retrouve la définition précédente.

3. On aurait pu, dans la définition ci-dessus, considérer des fonctions f mesurables à valeurs dans $\overline{\mathbb{R}}$. Cependant, l'intégrabilité de f impose alors que les ensembles $f^{-1}(\{+\infty\})$ et $f^{-1}(\{-\infty\})$ sont de mesure nulle. Donc on peut modifier la valeur de f sur ces ensembles négligeables (par exemple poser $f(x) = 0$ sur ces ensembles), sans changer la valeur de l'intégrale de f . Après ce changement, f prend ses valeurs dans \mathbb{R} .

On va montrer que la plupart des propriétés satisfaites par l'intégrale des fonctions positives s'étendent aux fonctions intégrables.

Proposition 3.20. *a) Pour tout $f \in \mathcal{L}^1(E, \mathcal{A}, \mu)$, on a l'inégalité $|\int f d\mu| \leq \int |f| d\mu$. On qualifie cette inégalité de « triangulaire » : en effet, on la voit comme une généralisation de $|\sum_i a_i| \leq \sum_i |a_i|$, pour une famille (a_1, \dots, a_N) de nombre réels.*

b) $\mathcal{L}^1(E, \mathcal{A}, \mu)$ est un espace vectoriel, et l'application $f \mapsto \int f d\mu$ est une forme linéaire sur cet espace vectoriel. Autrement dit, la formule (3.2) s'étend aux fonctions $f, g \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ et aux réels $a, b \in \mathbb{R}$ quelconques.

c) Si $f, g \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ et $f \leq g$, alors $\int f d\mu \leq \int g d\mu$. On dit parfois que l'intégrale est une forme linéaire positive.

d) Si $f, g \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ et $f = g$ p.p., alors $\int f d\mu = \int g d\mu$.

Démonstration. a) $|\int f| = |\int f^+ - \int f^-| \leq |\int f^+| + |\int f^-| = \int (f^+ + f^-) = \int |f|$ (on a omis d'indiquer les $d\mu$ dans les intégrales). Ici on s'est servi de l'inégalité triangulaire.

b) Bien que la linéarité semble évidente, il faut un peu « ruser » pour le montrer, car on n'a à notre disposition uniquement des résultats pour les fonctions positives. Il faut donc se ramener à des fonctions positives.

Pour $f \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ et $a \in \mathbb{R}$ on a $\int |af| d\mu = |a| \int |f| d\mu < \infty$, donc $af \in \mathcal{L}^1$. Pour montrer la linéarité de l'intégrale par multiplication par le réel a , on sépare les cas $a \geq 0$ et $a \leq 0$:

- si $a \geq 0$, alors $af = (af)^+ - (af)^- = af^+ - af^-$, donc $\int (af) d\mu = \int af^+ d\mu - \int af^- d\mu$. En se servant de (3.2), cette expression vaut $a(\int f^+ - \int f^-) = a \int f d\mu$.

- si $a < 0$, on a $(af)^+ = (-a)f^-$ et $(af)^- = (-a)f^+$. On trouve alors, par les mêmes manipulations, $\int (af) d\mu = (-a)(\int f^- - \int f^+) = (-a)(-\int f) = a \int f d\mu$.

Si f et g sont intégrables, l'inégalité triangulaire ponctuelle $|f(x) + g(x)| \leq |f(x)| + |g(x)|$ implique que $(f + g)$ est aussi intégrable. Pour montrer que $\int (f + g) = \int f + \int g$, il faut décomposer $(f + g)$ en parties positive et négative. A partir de l'identité :

$$f^+ - f^- + g^+ - g^- = f + g = (f + g)^+ - (f + g)^-,$$

on obtient l'identité entre fonctions positives

$$(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+.$$

On prend l'intégrale de cette identité :

$$\int [(f + g)^+ + f^- + g^-] d\mu = \int [(f + g)^- + f^+ + g^+] d\mu,$$

et on utilise la linéarité de l'intégrale pour les fonctions positives, de façon à obtenir :

$$\begin{aligned} \int (f + g)^+ + \int f^- + \int g^- &= \int (f + g)^- + \int f^+ + \int g^+ \\ \implies \int (f + g)^+ - \int (f + g)^- &= \int f^+ - \int f^- + \int g^+ - \int g^-, \end{aligned}$$

ce qui donne $\int (f + g) d\mu = \int f d\mu + \int g d\mu$.

c) Supposons f, g intégrables et $f \leq g$. La fonction $(g - f)$ est intégrable et positive, donc

$0 \leq \int (g - f) d\mu < \infty$. La linéarité montrée en *b)* permet alors d'écrire

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

d) L'égalité presque sûre $f = g$ p.p. implique que $f^+ = g^+$ p.p. et $f^- = g^-$ p.p. A partir du point (4) de la Proposition 3.15, on en déduit que $\int f^+ d\mu = \int g^+ d\mu$ et $\int f^- d\mu = \int g^- d\mu$, et finalement $\int f d\mu = \int g d\mu$. \square

Remarque 3.21. En se servant du *d)*, on peut affaiblir l'hypothèse du *c)* en l'hypothèse : « Si $f, g \in \mathcal{L}^1$ et si $f \leq g$ p.p. ». Comme expliqué plus haut, modifier les valeurs d'une fonction sur un ensemble négligeable ne va pas modifier son intégrale.

3.2.2 Fonctions à valeurs complexes

Pour l'instant on a supposé que les fonctions f prennent des valeurs réelles (ou dans $\overline{\mathbb{R}}$). On peut étendre la théorie aux fonctions à valeurs dans \mathbb{C} , puis dans \mathbb{R}^d pour $d \geq 1$ quelconque.

On rappelle qu'une fonction $f : E \rightarrow \mathbb{C}$ est mesurable si, pour tout borélien $B \subset \mathbb{C}$, $f^{-1}(B)$ est dans la tribu \mathcal{A} . Comme $\mathbb{C} \equiv \mathbb{R}^2$, la Proposition 2.28 montre que f est mesurable si et seulement si les fonctions $\operatorname{Re}(f)$ et $\operatorname{Im}(f)$ sont mesurables.

Définition 3.22. Soit (E, \mathcal{A}, μ) un espace mesuré. Une fonction $f : E \rightarrow \mathbb{C}$ mesurable est dite intégrable si la fonction $|f| : E \rightarrow \mathbb{R}_+$ est intégrable. On note alors $f \in \mathcal{L}_\mathbb{C}^1(E, \mathcal{A}, \mu)$. On définit alors l'intégrale de f par

$$\int f d\mu = \int \operatorname{Re}(f) d\mu + i \int \operatorname{Im}(f) d\mu \in \mathbb{C}. \quad (3.6)$$

Comme $|\operatorname{Re}(f)| \leq |f|$, $|\operatorname{Im}(f)| \leq |f|$ et $|f| \leq |\operatorname{Re}(f)| + |\operatorname{Im}(f)|$, l'intégrabilité de f est équivalente à l'intégrabilité simultanée de ses parties réelle et imaginaire.

Proposition 3.23. Les propriétés *a)*, *b)*, *d)* de la Proposition 3.20 s'étendent aux fonctions intégrables à valeurs complexes.

Démonstration. Les propriétés *b)* et *d)* sont évidentes à étendre, à partir de l'expression (3.6).

Pour montrer l'inégalité triangulaire *a)* il faut travailler un peu plus. On utilise la propriété suivante du module d'un nombre complexe :

$$\forall z \in \mathbb{C}, \quad |z| = \max_{u \in \mathbb{C}, |u|=1} \operatorname{Re}(uz). \quad (3.7)$$

On aura donc

$$\left| \int f d\mu \right| = \max_{u \in \mathbb{C}, |u|=1} \operatorname{Re}(u \int f d\mu).$$

En décomposant $u = \operatorname{Re}(u) + i \operatorname{Im}(u)$, et en prenant la définition de l'intégrale de f , on obtient

$$\left| \int f d\mu \right| = \max_{u \in \mathbb{C}, |u|=1} \left(\operatorname{Re}(u) \int \operatorname{Re}(f) d\mu - \operatorname{Im}(u) \int \operatorname{Im}(f) d\mu \right).$$

Par linéarité de l'intégrale des fonctions réelles, cela donne

$$\left| \int f d\mu \right| = \max_{u \in \mathbb{C}, |u|=1} \left(\int [\operatorname{Re}(u) \operatorname{Re}(f) - \operatorname{Im}(u) \operatorname{Im}(f)] d\mu \right) = \max_{u \in \mathbb{C}, |u|=1} \left(\int \operatorname{Re}(uf) d\mu \right).$$

Finalement, on a pour tout $u \in \mathbb{C}$ de norme 1, l'inégalité entre fonctions intégrables $\operatorname{Re}(uf) \leq |f|$, donc $\int \operatorname{Re}(uf) d\mu \leq \int |f| d\mu$. \square

Exercice 3.24. Définir la notion d'intégrabilité pour une fonction $f : E \rightarrow \mathbb{R}^d$, définir son intégrale en termes de ses composantes $(f_j)_{j=1,\dots,d}$, et montrer la généralisation des propriétés a), c), d) de la proposition 3.20 à ce contexte.

3.2.3 Le théorème de convergence dominée

Le troisième résultat de convergence de la théorie de l'intégration de Lebesgue (probablement le plus important des trois théorèmes présentés dans ce chapitre) est le théorème de convergence dominée.

Théorème 3.25. [Théorème de convergence dominée] Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions dans $\mathcal{L}^1(E, \mathcal{A}, \mu)$ (ou $\mathcal{L}_{\mathbb{C}}^1(E, \mathcal{A}, \mu)$). On fait les hypothèses suivantes :

(1) il existe une fonction f mesurable à valeurs dans \mathbb{R} (ou dans \mathbb{C}) telle que

$$f_n(x) \xrightarrow[n \rightarrow \infty]{} f(x) \quad \mu - p.p.$$

(2) il existe une fonction $g : E \rightarrow \mathbb{R}_+$ mesurable et intégrable ($\int g d\mu < \infty$) telle que

$$\forall n \in \mathbb{N}, \quad |f_n(x)| \leq g(x) \quad \mu - p.p.$$

Alors la fonction $f \in \mathcal{L}^1(E, \mathcal{A}, \mu)$ (ou $f \in \mathcal{L}_{\mathbb{C}}^1(E, \mathcal{A}, \mu)$), et on a les convergences

$$\int f_n d\mu \xrightarrow[n \rightarrow \infty]{} \int f d\mu \tag{3.8}$$

$$\text{et} \quad \int |f - f_n| d\mu \xrightarrow{n \rightarrow \infty} 0. \quad (3.9)$$

La dernière ligne est souvent appelée « convergence \mathcal{L}^1 », ou plus exactement « convergence L^1 ». Le terme « convergence dominée » provient du fait que les fonctions f_n sont toutes « dominées » (bornées) par une même fonction $g \in \mathcal{L}^1$.

Démonstration. Dans un premier temps, on va supposer que les deux hypothèses presque sûres sont vraies partout :

(1)' $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ pour tout $x \in E$,

(2)' pour tout $n \geq 0$ et pour tout $x \in E$, $|f_n(x)| \leq g(x)$.

Ces propriétés impliquent que $|f| \leq g$, donc que la fonction limite f est intégrable (on savait déjà que f était mesurable).

Pour montrer la limite (3.9) on va appliquer le Lemme de Fatou à une suite de fonctions bien choisie. Les propriétés $|f_n| \leq g$ et $|f| \leq g$ impliquent que $|f - f_n| \leq 2g$, donc que la fonction $(2g - |f - f_n|)$ est positive. Cette fonction converge (simplement) vers $2g$, le Lemme de Fatou nous donne donc :

$$\liminf_n \int (2g - |f - f_n|) d\mu \geq \int \liminf_n (2g - |f - f_n|) d\mu = \int 2g d\mu.$$

Ceci peut se réécrire, en utilisant la linéarité de l'intégrale à gauche :

$$\int 2g d\mu - \limsup_n \int |f - f_n| d\mu \geq \int 2g d\mu,$$

donc $\limsup_n \int |f - f_n| d\mu \leq 0$. Comme ces intégrales sont positives, on a donc montré $\lim_n \int |f - f_n| d\mu = 0$.

A partir de cette limite, l'inégalité triangulaire donne

$$\left| \int f d\mu - \int f_n d\mu \right| \leq \int |f - f_n| d\mu \xrightarrow{n \rightarrow \infty} 0,$$

ce qui montre la convergence (3.8). On a donc montré le théorème sous les hypothèses plus fortes (1)' et (2)'.

On fait maintenant les hypothèses (1) et (2). On définit les « bons ensembles » $B_n = \{x \in E; f_n(x) \leq g(x)\}$ et $B_{lim} = \{x \in E; \lim_n f_n(x) = f(x)\}$. Les hypothèses disent que

leurs complémentaires B_n^c et B_{lim}^c sont négligeables, autrement dit ils sont inclus respectivement dans des ensembles N_n et N_{lim} mesurables de mesure nulle (on peut montrer que B_n et B_{lim} sont mesurables, et donc prendre $N_n = B_n^c$, $N_{lim} = B_{lim}^c$, mais on ne s'en servira pas ici). Le « bon ensemble global » $B = B_{lim} \cap (\bigcap_n B_n)$, celui des points x satisfaisant toutes les bonnes propriétés, a pour complémentaire

$$B^c = B_{lim}^c \cup \left(\bigcup_n B_n^c \right) \subset N_{lim} \cup \left(\bigcup_n N_n \right) := N.$$

L'ensemble de droite est de mesure nulle, donc B^c est négligeable. On fabrique alors les fonctions mesurables

$$\tilde{f}_n = f_n \mathbb{1}_{N^c}, \quad \tilde{f} = f \mathbb{1}_{N^c}.$$

Comme $N^c \subset B$, ces fonctions satisfont les propriétés (1)' et (2)', donc elles vérifient la conclusion du théorème. On remarque finalement que, comme N est de mesure nulle, on $f_n = \tilde{f}_n$ p.p. et $f = \tilde{f}$ p.p., donc on peut remplacer dans les intégrales les fonctions \tilde{f}_n, \tilde{f} par f_n, f sans changer les valeurs des intégrales. \square

Exercice 3.26. Discuter les trois exemples de l'Exercice 3.17 : les fonctions $(f_n)_n$ convergent-elles simplement ? Si c'est le cas, sont-elles dominées par une fonction g intégrable ?

3.3 Intégrales dépendant d'un paramètre

On considère à présent la situation où la fonction f dépend de deux variables, qui ont des rôles différents : une variable $x \in E$, sur la quelle on va intégrer ; et une variable u appartenant à un « espace des paramètres » U , sur laquelle on ne va pas intégrer.

Le cas d'une suite de fonctions $(f_n)_{n \in \mathbb{N}}$ est déjà dans ce cadre : ici le paramètre u est l'entier $n \in \mathbb{N}$. On va généraliser au cas où l'espace des paramètres U est un espace métrique quelconque, par exemple un ouvert de l'espace euclidien \mathbb{R}^d . On se pose alors naturellement la question de la variation de l'intégrale $\int f(x, u) d\mu(x)$ par rapport au paramètre u : sous quelles conditions cette intégrale va-t-elle varier continûment par rapport à u ? Plus bas, on s'intéressera aussi à la dérivabilité de l'intégrale par rapport au paramètre $u \in \mathbb{R}^d$. Les résultats de cette section sont des corollaires du TCD.

3.3.1 Continuité d'une intégrale à paramètre

Notre espace mesuré E n'a pas forcément de structure topologique, alors que l'espace métrique en a une. Il s'agit donc de comprendre les relations entre la structure mesurable (par rapport à la variable x) et la structure topologique (continuité, par rapport à la variable u). Ce mélange de deux structures rend les hypothèses dans les théorèmes ci-dessous un peu délicates, les variables u, x ne jouant pas les mêmes rôles.

Théorème 3.27. *Soit (E, \mathcal{A}, μ) un espace mesuré, et (U, d) un espace métrique. On considère une application $f : U \times E \rightarrow \mathbb{R}$, et on fixe un point $u_0 \in U$. On fait les hypothèses suivantes :*

- i) pour tout paramètre $u \in U$, l'application $x \mapsto f(u, x)$ est mesurable ;*
- ii) pour μ -presque tout $x \in E$, l'application $u \mapsto f(u, x)$ est continue en u_0 ;*
- iii) il existe une fonction $g \in \mathcal{L}_+^1(E, \mathcal{A}, \mu)$ telle que, pour tout $u \in U$, $|f(u, x)| \leq g(x)$ $\mu(dx)$ -p.p.*

Alors la fonction $u \mapsto \int f(u, x) d\mu(x)$ est bien définie en tout point $u \in U$, et elle est continue en u_0 .

Remarque 3.28. [Comparaison avec le théorème de convergence dominée (TCD)] Ce théorème généralise le TCD au cas d'un espace des paramètres continu. En effet, en prenant $U = \mathbb{N}$ et sa topologie discrète, on retrouve exactement le TCD (la condition *ii*. de continuité en u_0 est l'analogue de la condition de convergence $\lim_{n \rightarrow \infty} f_n = f$). La nouveauté ici est d'autoriser des familles de fonctions indicées par un ensemble continu.

Démonstration. Les hypothèses *i)* et *iii)* impliquent que pour tout $u \in U$, la fonction $f(u, \cdot)$ est intégrable, de sorte que son intégrale $\int f(u, x) d\mu(x)$ est bien définie. Si on choisit une suite $(u_n)_{n \geq 1}$ une suite de points de U convergeant vers u_0 , alors les hypothèses ci-dessus montrent que les fonctions $(f(u_n, \cdot))_{n \geq 1}$ satisfont les hypothèses du théorème 3.25 de convergence dominée, avec comme fonction limite $f(u_0, \cdot)$. Le TCD montre donc que les intégrales $\int f(u_n, x) d\mu(x)$ convergent vers $\int f(u_0, x) d\mu(x)$ lorsque $n \rightarrow \infty$. \square

Exemple 3.29. [μ -primitive d'une fonction intégrable] On se place sur l'espace mesuré $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ avec μ une mesure diffuse (sans atomes) ; on prend $U = \mathbb{R}$ avec sa structure métrique usuelle. Si $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$, on considère l'intégrale

$$F(u) := \int_{]-\infty, u]} \varphi(x) \mu(dx) = \int_{-\infty}^u \varphi(x) d\mu(x) = \int \varphi \mathbb{1}_{]-\infty, u]} d\mu$$

(les trois écritures sont équivalentes). On montre alors que la fonction $u \mapsto F(u)$ est continue. Attention : cette expression intégrale ressemble à celle de la primitive d'une fonction φ intégrable continue, $\int_{-\infty}^u \varphi(x) dx = \int_{-\infty}^u \varphi(x) \lambda(dx)$. Mais dans notre cas, avec une mesure diffuse μ quelconque, la fonction F n'est pas forcément dérivable, même lorsque φ est continue intégrable.

Montrons la continuité de F . On considère ici la famille de fonctions $f(u, x) = \varphi(x) \mathbb{1}_{]-\infty, u]}(x)$. Pour tout $u \in U$, la fonction $f(u, \cdot)$ (autrement dit $x \mapsto f(u, x)$) est borélienne par rapport à x , et elle est dominée par la fonction $g = |\varphi| \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$; on a donc bien les hypothèses *i)* et *iii)* du théorème. Pour vérifier l'hypothèse *ii)*, il faut remarquer que, la fonction $u \mapsto f(u, x)$ est constante sur $u \in]-\infty, x[$ et sur $u \in]x, \infty[$, donc son seul point de discontinuité est en $u = x$. Donc si on se fixe un point $u_0 \in \mathbb{R}$, alors $u \mapsto f(u, x)$ est continue en $u = u_0$, sauf si $x = u_0$. L'ensemble $\{u_0\} \subset \mathbb{R}$ est négligeable par rapport à la mesure μ , puisque celle-ci est diffuse ; la fonction $u \mapsto f(u, x)$ est donc bien continue en $u = u_0$ pour μ -presque tout point x , ce qui est l'hypothèse *ii)*.

On remarque facilement que ce résultat est faux en général si la mesure μ admet des atomes. Par exemple dans le cas où $\mu = \delta_{u_0}$ la mesure de Dirac en u_0 , on vérifie que $F(u) = 0$ pour $u < u_0$, tandis que $F(u) = \varphi(u_0)$ pour $u \geq u_0$.

Exemple 3.30. *a)* [Transformée de Fourier] Si $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, la fonction $\hat{\varphi}(u) := \int e^{-iux} \varphi(x) \lambda(dx)$ est appelée la transformée de Fourier de φ . Cette fonction est continue et bornée sur \mathbb{R} entier.

b) [Convolution] soit $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ et soit $h : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue et bornée. Alors la fonction

$$u \mapsto h * \varphi(u) = \int h(u - x) \varphi(x) \lambda(dx)$$

est continue et bornée sur \mathbb{R} entier.

3.3.2 Dérivabilité d'une intégrale à paramètre

Pour pouvoir dériver par rapport à $u \in U$, il faut que U ait une structure euclidienne. On prend ici $U = I$ un intervalle ouvert de \mathbb{R} .

Théorème 3.31. *Soit (E, \mathcal{A}, μ) un espace mesuré, et $I \subset \mathbb{R}$ un intervalle ouvert. On considère une application $f : I \times E \rightarrow \mathbb{R}$, et on fixe un point $u_0 \in I$. Supposons que :*

i) pour tout $u \in I$, l'application $x \mapsto f(u, x)$ est dans $\mathcal{L}^1(E, \mathcal{A}, \mu)$;

ii) Pour μ -presque tout x , l'application $u \mapsto f(u, x)$ est dérivable en u_0 , de dérivée notée $\frac{\partial f}{\partial u}(u_0, x)$;

iii) il existe $g \in \mathcal{L}_+^1(E, \mathcal{A}, \mu)$ telle que, pour tout $u \in I$,

$$|f(u, x) - f(u_0, x)| \leq g(x) |u - u_0| \quad \mu(dx) - p.p.$$

Alors la fonction $F(u) = \int f(u, x) \mu(dx)$ est continue et dérivable en u_0 , et sa dérivée en ce point vaut

$$F'(u_0) = \int \frac{\partial f}{\partial u}(u_0, x) \mu(dx).$$

On remarque que la borne uniform $g(x)$ ne s'applique plus directement à la fonction $f(x)$, mais au *taux de variation* de f entre u_0 et u , $\frac{f(u, x) - f(u_0, x)}{u - u_0}$. La dérivée $\frac{\partial f}{\partial u}(u_0, x)$ n'est a priori définie que sur le complémentaire d'un ensemble négligeable, mais on peut la compléter de façon arbitraire en dehors de cet ensemble, de façon à obtenir une fonction définie partout. Il n'est pas évident que cette fonction soit mesurable, la preuve montrera qu'elle est égale p.p. à une fonction mesurable (et intégrable).

Démonstration. Considérons une suite $(u_n)_{n \in \mathbb{N}}$ dans $I \setminus \{x_0\}$, convergeant vers u_0 , et définissons les fonctions $\varphi_n(x) = \frac{f(u_n, x) - f(u_0, x)}{u_n - u_0}$. Comme les fonctions $f(u, \cdot)$ sont toutes mesurables et μ -intégrables, les fonctions φ_n le sont également. De plus, l'hypothèse iii) montre que μ -presque partout, les fonctions $|\varphi_n|$ sont bornées par la fonction intégrable g . On peut donc définir

$$\varphi_\infty(x) = \limsup_{n \rightarrow \infty} \varphi_n(x),$$

une fonction mesurable qui vérifie

$$|\varphi_\infty(x)| \leq g(x) \quad \mu - p.p.,$$

donc $\varphi_\infty \in \mathcal{L}^1(E, \mathcal{A}, \mu)$. D'autre part l'hypothèse ii) montre que

$$\text{pour } \mu\text{-presque tout } x, \quad \varphi_\infty(x) = \lim_{n \rightarrow \infty} \varphi_n(x) = \frac{\partial f}{\partial u}(u_0, x).$$

Le théorème de convergence dominée montre alors que

$$\begin{aligned} \int \varphi_\infty(x) d\mu(x) &= \lim_{n \rightarrow \infty} \int \varphi_n(x) d\mu(x) = \lim_{n \rightarrow \infty} \int \left(\frac{f(u_n, x) - f(u_0, x)}{u_n - u_0} \right) d\mu(x) \\ &= \lim_{n \rightarrow \infty} \frac{1}{u_n - u_0} (F(u_n) - F(u_0)). \end{aligned}$$

Comme $\varphi_\infty(x)$ est égal à $\frac{\partial f}{\partial u}(u_0, x)$ pour presque tout x , l'intégrale du membre de gauche peut s'écrire $\int \frac{\partial f}{\partial u}(u_0, x) d\mu(x)$. Comme ce membre de gauche est indépendant de la suite (u_n) convergeant vers u_0 , l'égalité ci-dessus montre que la fonction F est dérivable en u_0 , de dérivée $F'(u_0) = \int \varphi_\infty(x) d\mu(x)$. \square

Remarque 3.32. Très souvent, les hypothèses *ii)* et *iii)* sont remplacées par des hypothèses plus fortes :

ii)' pour μ -presque tout x , l'application $u \mapsto f(x, u)$ est dérivable sur I ;

iii)' il existe $g \in \mathcal{L}_+^1(E, \mathcal{A}, \mu)$ telle que pour μ -presque tout x ,

$$\forall u \in I, \quad \left| \frac{\partial f}{\partial u}(u, x) \right| \leq g(x).$$

Le théorème des accroissements finis montre que *iii)'* implique *iii)*. Ces hypothèses renforcées impliquent que F est dérivable sur I .

Exercice 3.33. Reprenons les exemples des exercices 3.30 et 3.29.

a) Supposons que $\varphi \in \mathcal{L}^1(R, \mathcal{B}(\mathbb{R}), \lambda)$ est tel que la fonction $x \mapsto x\varphi(x)$ est également dans $\mathcal{L}^1(R, \mathcal{B}(\mathbb{R}), \lambda)$. Montrer alors que la transformée de Fourier $\hat{\varphi}(u)$ est dérivable sur \mathbb{R} , et qu'elle vaut

$$\hat{\varphi}(u) = -i \int x e^{-iux} \varphi(x) \lambda(dx).$$

b) Soit $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ et soit $h : \mathbb{R} \rightarrow \mathbb{R}$ une fonction de classe C^1 , bornée ainsi que sa dérivée. Alors montrer que la convolution $h * \varphi$ est dérivable sur \mathbb{R} , et de dérivée $(h * \varphi)' = h' * \varphi$.

c) Soit μ une mesure diffuse sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et soit $\varphi \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ telle que $x \mapsto x\varphi(x)$ est aussi dans $\mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$. Pour tout $u \in \mathbb{R}$, on pose

$$F(u) := \int (u - x)^+ \varphi(x) \mu(dx).$$

Montrer que F est dérivable sur \mathbb{R} , de dérivée $F'(u) = \int_{]-\infty, u]} \varphi(x) \mu(dx)$.

4 Construction et propriétés de la mesure de Lebesgue

Nous avons pour l'instant évoqué la mesure de Lebesgue sur \mathbb{R} comme étant la mesure « naturelle », celle qui associe à un intervalle sa longueur. Nous avons aussi affirmé qu'elle est bien définie sur toute la tribu des boréliens. Nous avons enfin montré qu'il ne peut exister au plus qu'une telle mesure (grâce au lemme de classe monotone). Dans ce chapitre nous allons construire la mesure de Lebesgue, satisfaisant toutes ces propriétés. On montrera, en particulier, qu'elle permet de faire le lien avec l'intégrale de Riemann sur \mathbb{R} , c'est-à-dire que l'intégrale de Lebesgue d'une fonction réglée est égale à l'intégrale de Riemann correspondante.

Nous nous focaliserons sur la construction de la mesure de Lebesgue, mais notre stratégie de construction pourra en fait s'appliquer à bien d'autres mesures. Cette stratégie consiste à construire tout d'abord un objet moins contraint qu'une mesure, qu'on appelle une *mesure extérieure* (ou mesure σ -sous-additive); cet assouplissement des contraintes permet à cette mesure extérieure d'être sur toutes les parties de E . On introduira aussi la notion d'algèbre, une notion qui anticipe celle de σ -algèbre. Notre mesure finale sera finalement obtenue comme une restriction (sur une certaine tribu) de cette mesure extérieure.

Notre construction d'une mesure μ ira à rebours : on expliquera tout d'abord comment, à partir d'une mesure extérieure μ^* , on définit une tribu $\mathcal{M}(\mu^*)$ et une mesure μ dessus. Dans un second temps, on expliquera comment fabriquer une mesure extérieure μ^* , à partir d'une mesure additive définie sur une algèbre d'ensembles \mathcal{A} (qui sera incluse dans la tribu $\mathcal{M}(\mu^*)$).

4.1 Mesures extérieures

Définition 4.1. Soit E un ensemble quelconque. Une application $\mu^* : \mathcal{P}(E) \rightarrow [0, \infty]$ est appelée une mesure extérieure (ou mesure σ -sous-additive) si :

- i) $\mu^*(\emptyset) = 0$;
- ii) μ^* est croissante : si $A \subset B$, alors $\mu^*(A) \leq \mu^*(B)$;
- iii) μ^* est σ -sous-additive : pour toute suite $(A_k)_{k \in \mathbb{N}}$ d'éléments de $\mathcal{P}(E)$,

$$\mu^* \left(\bigcup_{k \in \mathbb{N}} A_k \right) \leq \sum_{k \in \mathbb{N}} \mu^*(A_k).$$

On pourra comparer ces propriétés avec celles d'une mesure (v. Définition 2.8). Les inégalités ci-dessus sont également vérifiées par une mesure, puisqu'elles découlent de la propriété plus

forte de σ -additivité. Nous n'imposons à μ^* que ces propriétés « faibles » ; par contre, nous imposons à μ^* d'être définie sur toute partie $A \subset E$, et pas uniquement sur une tribu particulière. En fait, la construction de la tribu sur laquelle notre mesure sera définie, va se faire « dynamiquement », à partir de μ^* .

Nous montrerons dans une section ultérieure comment construire une telle mesure ultérieure. Notre objectif présent est de montrer comment, à partir d'une mesure extérieure μ^* donnée, on construit une vraie mesure μ . La première étape est de définir la tribu sur laquelle sera définie μ , ou de façon équivalente, la notion de mesurabilité par rapport à la mesure extérieure μ^* .

Définition 4.2. Une partie $B \subset E$ est dite μ^* -mesurable si,

$$\text{pour toute partie } A \subset E, \quad \mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c). \quad (4.1)$$

On note $\mathcal{M}(\mu^*)$ l'ensemble des parties μ^* -mesurables.

Remarque 4.3. La notion de « μ^* -mesurable » est encore une nouvelle utilisation du mot « mesurable » ! Elle ne signifie pas que $\mu^*(B)$ est bien définie, puisque μ^* est en fait définie sur tout $\mathcal{P}(E)$.

On remarque aussi que l'inégalité $\mu^*(A) \leq \mu^*(A \cap B) + \mu^*(A \cap B^c)$ vraie pour toutes parties B, A , par σ -sous-additivité. L'égalité, par contre, ne sera vraie que pour certaines parties ; pour vérifier l'égalité, il suffira donc de vérifier l'inégalité inverse $\mu^*(A) \geq \mu^*(A \cap B) + \mu^*(A \cap B^c)$. La mesurabilité de la partie B indique que la mesure extérieure μ^* est additive vis-à-vis de la partition de toute partie A en $A = (A \cap B) \sqcup (A \cap B^c)$.

L'intérêt de cette notion est manifeste dans le

Théorème 4.4. 1. L'ensemble $\mathcal{M}(\mu^*)$ est une tribu sur E , qui contient toutes les parties B de E telles que $\mu^*(B) = 0$.

2. La restriction de μ^* à cette tribu $\mathcal{M}(\mu^*)$, qu'on notera μ , est une mesure.

La propriété, pour $\mathcal{M}(\mu^*)$, de contenir tous les ensembles μ^* -négligeables, fait de la mesure μ une mesure complète.

Ce théorème nous a donc fabriqué une tribu et une mesure μ définie sur cette tribu, à partir de la seule mesure extérieure μ^* .

Démonstration. 1. Pour alléger les notations, on écrira dans cette preuve $\mathcal{M} = \mathcal{M}(\mu^*)$. Montrons tout d'abord que la classe \mathcal{M} contient tous les ensembles μ^* -négligeables. Si $\mu^*(B) = 0$, on a pour toute partie A , en utilisant la croissance de μ^* :

$$\mu^*(A) \geq \mu^*(A \cap B^c) \quad \text{et} \quad 0 = \mu^*(B) \geq \mu^*(A \cap B),$$

d'où on déduit $\mu^*(A) \geq \mu^*(A \cap B^c) + \mu^*(A \cap B)$,

et donc l'égalité. On a donc montré que B est μ^* -mesurable.

Montrons que \mathcal{M} satisfait la définition 2.1 d'une tribu. Il est évident que $B = \emptyset$ satisfait la propriété de μ^* mesurabilité, puisque $\mu^*(\emptyset) = 0$; donc l'ensemble vide est bien dans \mathcal{M} . D'autre part, la propriété (4.1) est symétrique par passage au complémentaire $B \rightarrow B^c$, ce qui montre que \mathcal{M} est stable par passage au complémentaire. Il reste à montrer la stabilité de \mathcal{M} par union dénombrable.

Montrons déjà que \mathcal{M} est stable par union finie, en particulier par union de deux de ses éléments. Soient $B_1, B_2 \in \mathcal{M}$. On veut montrer que $(B_1 \cup B_2)$ est également μ^* -mesurable. Pour tout $A \in \mathcal{P}(E)$, on a :

$$\begin{aligned} \mu^*(A \cap (B_1 \cup B_2)) &\stackrel{B_1 \in \mathcal{M}}{=} \mu^*([A \cap (B_1 \cup B_2)] \cap B_1) + \mu^*([A \cap (B_1 \cup B_2)] \cap B_1^c) \\ &= \mu^*(A \cap B_1) + \mu^*(A \cap B_2 \cap B_1^c), \end{aligned}$$

donc

$$\begin{aligned} \mu^*(A \cap (B_1 \cup B_2)) + \mu^*(A \cap (B_1 \cup B_2)^c) &= \mu^*(A \cap B_1) + \mu^*(A \cap B_2 \cap B_1^c) + \mu^*(A \cap B_1^c \cap B_2^c) \\ &\stackrel{B_2 \in \mathcal{M}}{=} \mu^*(A \cap B_1) + \mu^*(A \cap B_1^c) \\ &\stackrel{B_1 \in \mathcal{M}}{=} \mu^*(A). \end{aligned}$$

(on a indiqué l'hypothèse utilisée sur les égalités correspondantes). Ceci montre donc que $(B_1 \cup B_2) \in \mathcal{M}$. La classe \mathcal{M} est donc stable par unions finies, et donc par intersections finies. Et donc aussi par l'opération $B, B' \in \mathcal{M} \implies B \setminus B' \in \mathcal{M}$.

Si on considère une suite $(A_k)_{k \in \mathbb{N}}$ d'éléments de \mathcal{M} , on peut fabriquer la suite $B_0 = A_0$, $B_1 = A_1 \setminus A_0$, $B_2 = A_2 \setminus (A_0 \cup A_1)$, etc. La suite $(B_k)_{k \in \mathbb{N}}$ obtenue sera composée d'éléments de \mathcal{M} qui sont deux à deux disjoints, et tels que pour tout $n \in \mathbb{N}$, on a $\bigcup_{k=0}^n B_k = \bigcup_{k=0}^n A_k$.

En conséquence, $\bigcup_{k=0}^{\infty} B_k = \bigcup_{k=0}^{\infty} A_k$. Pour chaque n , on a la partition :

$$E = B_0 \sqcup B_1 \cdots \sqcup B_n \sqcup \left(\bigcap_{k=0}^n B_k^c \right).$$

Montrons par récurrence que μ^* est additive par rapport à cette partition, c'est-à-dire que

$$\forall A \in \mathcal{P}(E), \quad \mu^*(A) = \sum_{k=0}^n \mu^*(A \cap B_k) + \mu^* \left(A \cap \left(\bigcap_{k=0}^n B_k^c \right) \right). \quad (4.2)$$

La propriété est vraie à l'ordre $n = 0$, puisque $B_0 \in \mathcal{M}$. Supposons qu'elle est vraie à l'ordre n . En se servant du fait que $B_{n+1} \in \mathcal{M}$, on décompose le second terme en :

$$\begin{aligned} \mu^* \left(A \cap \left(\bigcap_{k=0}^n B_k^c \right) \right) &= \mu^* \left(A \cap \left(\bigcap_{k=0}^n B_k^c \right) \cap B_{n+1} \right) + \mu^* \left(A \cap \left(\bigcap_{k=0}^n B_k^c \right) \cap B_{n+1}^c \right) \\ &= \mu^*(A \cap B_{n+1}) + \mu^* \left(A \cap \left(\bigcap_{k=0}^{n+1} B_k^c \right) \right), \end{aligned}$$

où on a utilisé le fait que $B_{n+1} \subset \left(\bigcap_{k=0}^n B_k^c \right)$, puisque $\bigcup_{k=0}^n B_k \subset B_{n+1}^c$ (les B_k sont tous disjoints). En remplaçant cette expression dans l'égalité (4.2), on obtient l'égalité à l'ordre $n + 1$, ce qui montre que la propriété est vraie à tout ordre $n \in \mathbb{N}$.

On voudrait maintenant faire tendre n vers $+\infty$. Comme l'intersection dénombrable $\left(\bigcap_{k=0}^{\infty} B_k^c \right)$ est incluse dans $\left(\bigcap_{k=0}^n B_k^c \right)$, la croissance de μ^* , combinée avec (4.2), montre que pour tout $n \in \mathbb{N}$,

$$\forall A \in \mathcal{P}(E), \quad \mu^*(A) \geq \sum_{k=0}^n \mu^*(A \cap B_k) + \mu^* \left(A \cap \left(\bigcap_{k=0}^{\infty} B_k^c \right) \right).$$

Dans le membre de droite on peut alors faire tendre n vers $+\infty$, et on obtient

$$\begin{aligned} \forall A \in \mathcal{P}(E), \quad \mu^*(A) &\geq \sum_{k=0}^{\infty} \mu^*(A \cap B_k) + \mu^* \left(A \cap \left(\bigcap_{k=0}^{\infty} B_k^c \right) \right) \\ &\geq \mu^* \left(\bigcup_k (A \cap B_k) \right) + \mu^* \left(A \cap \left(\bigcap_{k=0}^{\infty} B_k^c \right) \right), \end{aligned} \quad (4.3)$$

où, dans la dernière ligne, on s'est servi de la propriété de σ -sous-additivité de μ^* . On re-

marque alors que les deux ensembles ci-dessus peuvent s'écrire

$$\bigcup_k (A \cap B_k) = A \cap \left(\bigsqcup_k B_k \right) = A \cap \left(\bigcup_k A_k \right), \quad \text{et} \quad \left(\bigcap_{k=0}^{\infty} B_k^c \right) = \left(\bigsqcup_k B_k \right)^c = \left(\bigcup_k A_k \right)^c.$$

On a donc montré $\mu^*(A) \geq \mu^*(A \cap (\bigcup_k A_k)) + \mu^*(A \cap (\bigcup_k A_k)^c)$; comme on a l'inégalité inverse, on a l'égalité entre les deux membres :

$$\mu^*(A) = \mu \left(A \cap \left(\bigcup_k A_k \right) \right) + \mu^* \left(A \cap \left(\bigcup_k A_k \right)^c \right), \quad (4.4)$$

ce qui montre que $\bigcup_k A_k$ est μ^* -mesurable. On a ainsi montré que la classe \mathcal{M} est une tribu.

2. Notons μ la restriction de μ^* à la tribu \mathcal{M} . Vérifions que μ satisfait bien la Définition 2.8 d'une mesure. On sait déjà que $\mu(\emptyset) = 0$. Soit $(B_k)_{k \in \mathbb{N}}$ une suite d'éléments disjoints de \mathcal{M} . La preuve du 1. ci-dessus nous a montré que $\bigsqcup_k B_k \in \mathcal{M}$. D'autre part, l'inégalité (4.3) combinée avec la σ -sous-additivité, montre que cette inégalité est en fait une égalité :

$$\forall A \in \mathcal{P}(E), \quad \mu^*(A) = \sum_{k=0}^{\infty} \mu^*(A \cap B_k) + \mu^* \left(A \cap \left(\bigsqcup_k B_k \right)^c \right).$$

Si on choisit $A = \bigsqcup_k B_k$, on a alors $A \in \mathcal{M}$, et l'égalité ci-dessus s'écrit $\mu(\bigsqcup_k B_k) = \sum_{k \in \mathbb{N}} \mu(B_k)$. On a montré la σ -additivité de μ . Ceci conclut la preuve que μ est une mesure. \square

4.2 Construction de la mesure extérieure de Lebesgue λ^* sur \mathbb{R}

Avant d'élaborer une théorie générale, on construira la mesure extérieure λ^* sur $\mathcal{P}(\mathbb{R})$, dont la restriction donnera la mesure de Lebesgue λ . On définit λ^* par la formule variationnelle suivante :

$$\forall A \subset E, \quad \lambda^*(A) := \inf \left\{ \sum_{i \in \mathbb{N}} (b_i - a_i) ; A \subset \bigcup_{i \in \mathbb{N}}]a_i, b_i[\right\}. \quad (4.5)$$

A droite on prend l'infimum sur toutes les suites d'intervalles ouverts $(]a_i, b_i[)_{i \in \mathbb{N}}$ tels que $A \subset \bigcup_{i \in \mathbb{N}}]a_i, b_i[$; on inclut a possibilité d'intervalles vides $]a_i, a_i[$, autrement dit on considère aussi les unions finies d'intervalles. Notons que l'infimum peut valoir $+\infty$, si toute famille d'intervalles contenant A satisfait $\sum_{i \in \mathbb{N}} (b_i - a_i) = +\infty$.

Cette définition nous montre déjà que $\lambda^*(A) \in [0, \infty]$ pour toute partie A , et que $\lambda^*(\emptyset) = 0$.

Par contre, il n'est pas complètement évident que $\lambda^*(]a, b]) = b - a$, nous le montrerons ci-dessous.

Il faut vérifier que la fonction λ^* définit bien une mesure extérieure.

Théorème 4.5. *i) λ^* est une mesure extérieure sur \mathbb{R} .*

ii) La tribu $\mathcal{M}(\lambda^)$ contient la tribu borélienne $\mathcal{B}(\mathbb{R})$.*

iii) Pour tous $a \leq b$, $\lambda^([a, b]) = \lambda^*(]a, b]) = b - a$.*

La restriction de λ^* à la tribu $\mathcal{M}(\lambda^*)$ s'appelle la mesure de Lebesgue sur \mathbb{R} , notée λ . Comme sa tribu $\mathcal{M}(\lambda^*)$ contient tous les boréliens, elle contient tous les intervalles ouverts, et leur mesure vaut $\lambda(]a, b]) = b - a$. Comme on l'a montré dans l'Exemple 2.17, λ est l'unique mesure possédant cette propriété.

Démonstration. *i)* la définition montre immédiatement que $\lambda^*(\emptyset) = 0$, et que λ^* est croissante. Il reste à montrer la σ -sous-additivité.

Soit $(A_n)_{n \in \mathbb{N}}$ une suite de parties de \mathbb{R} . Si l'une des parties satisfait $\lambda^*(A_i) = \infty$, il n'y a rien à montrer, donc on supposera que toutes les $\lambda^*(A_i) < \infty$. Choisissons un $\epsilon > 0$. Pour tout $n \in \mathbb{N}$, la définition de $\lambda^*(A_n)$ montre qu'on peut recouvrir A_n par une union d'intervalles

$$A_n \subset \bigcup_{i \in \mathbb{N}}]a_i^{(n)}, b_i^{(n)}[, \quad \text{tels que} \quad \sum_{i \in \mathbb{N}} (b_i^{(n)} - a_i^{(n)}) \leq \lambda^*(A_n) + \frac{\epsilon}{2^n}.$$

Si on somme sur les indices i et n , on obtient une famille *dénombrable* d'intervalles $(]a_i^{(n)}, b_i^{(n)}[)_{i, n \in \mathbb{N}}$, dont l'union recouvre $\bigcup_n A_n$, et telle que

$$\sum_{i, n \in \mathbb{N}} (b_i^{(n)} - a_i^{(n)}) \leq \sum_n \left(\lambda^*(A_n) + \frac{\epsilon}{2^n} \right) = 2\epsilon + \sum_n \lambda^*(A_n).$$

Par définition, $\lambda^*(\bigcup_n A_n)$ est majoré par le membre de gauche :

$$\lambda^*\left(\bigcup_n A_n\right) \leq 2\epsilon + \sum_n \lambda^*(A_n).$$

Enfin, on peut prendre $\epsilon \searrow 0$ dans l'expression ci-dessus, ce qui montre la σ -sous-additivité de λ^* . On a donc montré que λ^* est une mesure extérieure.

ii) On sait déjà que $\mathcal{M}(\lambda^*)$ est une tribu. Pour montrer qu'elle contient les boréliens, il suffit de montrer qu'elle contient une sous-famille \mathcal{C} qui engendre les boréliens. Par exemple, il suffit

de montrer qu'elle contient les intervalles de la forme $] -\infty, a]$, $a \in \mathbb{R}$. Choisissons donc un tel intervalle *Soit* $B =] -\infty, a]$. Pour montrer qu'il appartient à $\mathcal{M}(\lambda^*)$, il faut vérifier que pour toute partie $A \subset \mathbb{R}$,

$$\lambda^*(A) \geq \lambda^*(A \cap B) + \lambda^*(A \cap B^c) \quad (4.6)$$

(l'inégalité inverse étant vraie par la sous-additivité de λ^* qu'on vient de montrer). Il faut donc comparer les recouvrements de A , $A \cap B$ et $A \cap B^c$ par des unions d'intervalles ouverts. Soit $(]a_i, b_i[)_{i \in \mathbb{N}}$ une famille d'intervalles recouvrant A , et telle que

$$\sum_{i \in \mathbb{N}} (b_i - a_i) \leq \lambda^*(A) + \epsilon. \quad (4.7)$$

En rappelant la notation $a \wedge b = \min(a, b)$, $a \vee b = \max(a, b)$, on définit alors les deux familles :

$$\mathcal{F}_1 = (]a_i \wedge a, (b_i \wedge a) + \epsilon 2^{-i}[)_{i \in \mathbb{N}}, \quad \mathcal{F}_2 = (]a_i \vee a, b_i \vee a[)_{i \in \mathbb{N}}.$$

La première famille recouvre $A \cap] -\infty, a]$, la seconde recouvre $A \cap]a, \infty[$. V. la figure pour visualiser ces deux familles. A partir de ces deux familles, on trouve alors les inégalités :

$$\begin{aligned} \lambda^*(A \cap B) &\leq \sum_{i \in \mathbb{N}} [(b_i \wedge a) + \epsilon 2^{-i} - a_i \wedge a] = 2\epsilon + \sum_{i \in \mathbb{N}} [(b_i \wedge a) - a_i \wedge a], \\ \lambda^*(A \cap B^c) &\leq \sum_{i \in \mathbb{N}} [b_i \wedge a - a_i \wedge a]. \end{aligned}$$

En sommant ces deux inégalités et en remarquant que pour tout intervalle $]a_i, b_i[$, on a

$$(b_i \wedge a) - a_i \wedge a + b_i \wedge a - a_i \wedge a = b_i - a_i,$$

on trouve que

$$\lambda^*(A \cap B) + \lambda^*(A \cap B^c) \leq 2\epsilon + \sum_{i \in \mathbb{N}} [b_i - a_i].$$

En comparant avec l'inégalité (4.7), on trouve donc

$$\lambda^*(A \cap B) + \lambda^*(A \cap B^c) \leq \lambda^*(A) + 3\epsilon.$$

Comme $\epsilon > 0$ est arbitrairement petit, on a donc montré l'inégalité (4.6) qu'on voulait, montrant que l'intervalle B est λ^* -mesurable. La tribu $\mathcal{M}(\mu^*)$ contient donc les boréliens.

iii) Comme l'intervalle fermé $[a, b]$ est recouvert par $]a - \epsilon, b + \epsilon[$ pour tout $\epsilon > 0$, on a

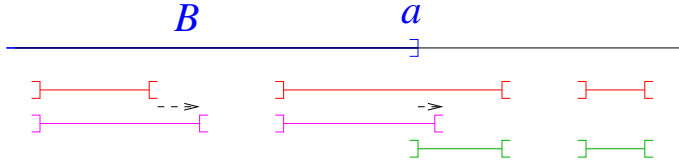


FIGURE 4.1 – Représentation schématique des intervalles $]a_i, b_i[$ (rouge) recouvrant A , et de leurs troncations recouvrant $A \cap B$ (famille \mathcal{F}_1 , violet) et $A \cap B^c$ (famille \mathcal{F}_2 , vert). Le prolongement par $\epsilon 2^{-i}$ est indiqué par une flèche.

$\lambda^*([a, b]) \leq b - a + 2\epsilon$, et donc $\lambda^*([a, b]) \leq b - a$. Montrons l'inégalité inverse. Supposons que $[a, b]$ est recouvert par la famille $(]a_i, b_i[)_{i \in \mathbb{N}}$. Comme $[a, b]$ est compact, le théorème de Borel-Lebesgue montre qu'il est en fait recouvert par une union finie, qu'on peut appeler $\bigcup_{i=0}^N]a_i, b_i[$. Vérifions la propriété intuitive que la somme des longueurs de ces intervalles doit forcément majorer celle de $[a, b]$:

$$b - a \leq \sum_{i=0}^N (b_i - a_i).$$

Quitte à restreindre l'union encore plus, on peut supposer que tous les $]a_i, b_i[$ intersectent $[a, b]$. Un des intervalles (on dira que c'est le premier) doit contenir le point a , donc vérifier $a_0 < a < b_0$. Si $b_0 > b$, on a donc $b_0 - a_0 > b - a$, et la preuve est finie. Si $b_0 < b$, le point b_1 doit être recouvert par un intervalle de la famille (supposons qu'il s'agit de $]a_1, b_1[$) donc tel que $a_1 < b_0 < b_1$. On a alors $b_1 - a_1 + b_0 - a_0 > b_1 - a_0$. Si $b_1 > b$ on a montré la propriété voulue. Si $b_1 < b$, on identifie un troisième intervalle de la famille contenant b_1 ; on aura alors $\sum_{i=0}^2 (b_i - a_i) > b_2 - a_0$. On procède ainsi de suite, jusqu'à ce qu'avoir un intervalle $]a_n, b_n[$ tel que $b_n > b$, et donc tel que

$$\sum_{i=0}^n (b_i - a_i) > b_2 - a_0 > b - a.$$

En rajoutant les intervalles « inutiles », on a donc

$$b - a \leq \sum_{i=0}^{\infty} (b_i - a_i).$$

En prenant l'infimum sur tous les recouvrements, on trouve donc $b - a \leq \lambda^*([a, b])$, et finalement $\lambda^*([a, b]) = b - a$.

Pour calculer la mesure de l'intervalle ouvert $]a, b[$, il suffit de remarquer que les singletons

$\{a\} = [a, a]$ et $\{b\} = [b, b]$ sont de mesure nulle, et que l'additivité de λ^* sur sa tribu implique que $\lambda^*([a, b]) = \lambda^*([a, b]) + \lambda^*([a, b]) + \lambda^*([a, b]) + \lambda^*([a, b])$. \square

Avec cette construction de la mesure extérieure s'achève celle de la mesure de Lebesgue. Le point *ii*) du théorème ci-dessus nous dit que la tribu $\mathcal{M}(\lambda^*)$ sur laquelle λ est définie (qu'on appellera la *tribu de Lebesgue*) contient la tribu borélienne $\mathcal{B}(\mathbb{R})$, mais elle n'est pas forcément égale à celle-ci. Mais la différence entre ces deux tribus est, d'une certaine façon, peu importante. On sait aussi que cette tribu $\mathcal{M}(\lambda^*)$ contient tous les ensembles λ -négligeables, ce qui n'était pas forcément le cas de la tribu borélienne. Nous allons donner une autre caractérisation de la tribu de Lebesgue, comme la complétion de la tribu de Borel par rapport à λ . Commençons par définir la complétée d'une tribu par rapport à une de ses mesures.

Proposition 4.6. *Soit (E, \mathcal{A}, μ) un espace mesuré. La tribu complétée de \mathcal{A} par rapport à μ est définie par $\bar{\mathcal{A}} = \sigma(\mathcal{A} \cup \mathcal{N})$, où \mathcal{N} est l'ensemble des parties μ -négligeables de E . Il existe alors sur $(E, \bar{\mathcal{A}})$ une unique mesure qui prolonge μ .*

Démonstration. Montrons qu'une façon de construire la tribu complétée est de considérer toutes les parties de E qui sont « encadrées » par deux éléments de \mathcal{A} de même mesure. On définit la classe

$$\mathcal{C} := \{A \in \mathcal{P}(E); \exists B, B' \in \mathcal{A}, B \subset A \subset B', \mu(B' \setminus B) = 0\},$$

nous allons montrer que $\mathcal{C} = \bar{\mathcal{A}}$. On montre d'abord que \mathcal{C} est une tribu. Si A est dans \mathcal{C} , le passage au complémentaire montre que $B'^c \subset A^c \subset B^c$, avec B'^c, B^c dans \mathcal{A} , et $B^c \setminus B'^c = B' \setminus B$ est de mesure nulle, donc $A \in \mathcal{C}$. Si les $(A_k)_{k \in \mathbb{N}}$ sont encadrés par $(B_k)_k$ et $(B'_k)_k$, alors $\bigcup_k A_k$ est encadrée par les éléments $\bigcup_k B_k$ et $\bigcup_k B'_k$ de \mathcal{A} , et on vérifie que

$$\begin{aligned} \mu \left(\left(\bigcup_k B'_k \right) \setminus \left(\bigcup_j B_j \right) \right) &= \mu \left(\bigcup_k \left(B'_k \setminus \left(\bigcup_j B_j \right) \right) \right) \leq \mu \left(\bigcup_k (B'_k \setminus B_k) \right) \\ &\leq \sum_k \mu(B'_k \setminus B_k) = 0. \end{aligned}$$

Ceci montre que $\bigcup_k A_k$ est bien dans \mathcal{C} . Comme la classe \mathcal{C} contient à la fois \mathcal{A} et l'ensemble des négligeables \mathcal{N} , elle contient $\bar{\mathcal{A}}$. Montrons l'inclusion inverse $\mathcal{C} \subset \bar{\mathcal{A}}$. Soit un élément $A \in \mathcal{C}$; il existe alors $B, B' \in \mathcal{A}$ encadrant A , donc en particulier $A = B \cup (A \setminus B)$, et $A \setminus B \subset B' \setminus B$ est négligeable. Ceci montre que A est dans l'union $\mathcal{A} \cup \mathcal{N}$, donc dans sa tribu engendrée $\bar{\mathcal{A}}$.

Remarque : On peut voir les éléments de $\mathcal{C} = \bar{\mathcal{A}}$ comme des éléments de \mathcal{A} auxquels on aurait rajouté, ou retiré, de la « poussière » de mesure nulle.

Il est alors facile de prolonger la mesure μ à cette tribu complétée : si $A \in \bar{\mathcal{A}}$ est encadrée par B, B' , comme $\mu(B') = \mu(B) + \mu(B' \setminus B) = \mu(B)$, on posera naturellement $\mu(A) = \mu(B)$. Il est facile de voir que cette définition ne dépend pas du choix de la paire (B, B') encadrant A . En effet, si A est encadrée par une seconde paire (\tilde{B}, \tilde{B}') , on a alors $\mu(\tilde{B}') = \mu(\tilde{B})$, mais aussi $\mu(\tilde{B}') \geq \mu(B)$ puisque $B \subset A \subset \tilde{B}'$, et inversement $\mu(\tilde{B}) \leq \mu(B')$. On en déduit donc que ces quatre ensembles ont la même mesure. Montrons la σ -additivité de cette mesure prolongée. Si $(A_n)_n$ est une suite d'éléments disjoints de $\bar{\mathcal{A}}$, chacun est encadré par une paire (B_n, B'_n) . Les $(B_n)_n$ forment alors une suite disjointe, et on aura alors

$$\sum_n \mu(A_n) = \sum_n \mu(B_n) = \mu\left(\bigcup_n B_n\right).$$

Enfin, on remarque comme ci-dessus que $(\bigcup_n A_n) \setminus (\bigcup_k B_k) = \bigcup_n (A_n \setminus B_n)$ du fait de la disjonction des $(A_n)_n$. Cette dernière union est négligeable comme union d'ensembles négligeables, de sorte que $\mu(\bigcup_n A_n) = \mu(\bigcup_n B_n) = \sum_n \mu(A_n)$. \square

On va pouvoir appliquer cette notion de tribu complétée pour décrire la tribu de Lebesgue à partir de celle de Borel.

Proposition 4.7. *La tribu de Lebesgue $\mathcal{M}(\lambda^*)$ est la complétée de la tribu de Borel $\mathcal{B}(\mathbb{R})$ par rapport à la mesure de Lebesgue. On notera cette tribu complétée $\overline{\mathcal{B}(\mathbb{R})}$, en gardant en mémoire qu'elle est complète pour la mesure de Lebesgue.*

Démonstration. D'après le théorème 4.5, $\mathcal{M}(\lambda^*)$ contient les boréliens, et contient les ensembles λ^* -négligeables de \mathbb{R} , qui sont aussi les ensembles λ -négligeables. Elle contient donc $\overline{\mathcal{B}(\mathbb{R})}$, la tribu complétée par rapport à la mesure de Lebesgue.

Montrons inversement que si $A \in \mathcal{M}(\lambda^*)$, alors $A \in \overline{\mathcal{B}(\mathbb{R})}$. On notera ici λ la mesure de Lebesgue sur $\mathcal{M}(\lambda^*)$.

1. Supposons dans un premier temps que A est borné, c'est-à-dire que $A \subset]-K, K[$ pour un certain $K > 0$. Par croissance de λ on aura alors $\lambda(A) < \infty$. Pour tout $n > 0$, on peut trouver une famille d'intervalles $\left(]a_i^{(n)}, b_i^{(n)}[\right)_{i \in \mathbb{N}}$ qui recouvrent A , en « l'approximant à une erreur $\frac{1}{n}$ près », autrement dit tels que

$$\lambda(A) + \frac{1}{n} \geq \sum_i \left(b_i^{(n)} - a_i^{(n)} \right).$$

Notons $B_n = \bigcup_i]a_i^{(n)}, b_i^{(n)}[$ ce recouvrement de A ; c'est bien sûr un borélien, et il satisfait (par σ -sous-additivité) $\lambda(A) + \frac{1}{n} \geq \lambda^*(B_n)$. Bien que les $(B_n)_n$ ne forment pas forcément une suite décroissante, on imagine qu'ils doivent se « serrer de plus en plus » autour de A lorsque $n \rightarrow \infty$. On peut considérer leur intersection $B' := \bigcap_n B_n$, qui est aussi un borélien contenant A . Cet ensemble vérifie

$$\lambda(A) + \frac{1}{n} \geq \lambda(B'), \text{ pour tout } n, \text{ donc } \lambda(A) \geq \lambda(B'), \text{ donc } \lambda(A) = \lambda(B').$$

En considérant l'ensemble $\tilde{A} =]-K, K[\setminus A$, qui est également dans la tribu $\mathcal{M}(\lambda^*)$, on fabrique de la même manière un borélien $\tilde{B}' \supset \tilde{A}$ (qu'on peut supposer inclus dans $] - K, K[$), et tel que $\lambda(\tilde{B}') = \lambda(\tilde{A})$. Son complémentaire dans cet intervalle, $B =]-K, K[\setminus \tilde{B}'$, satisfait $B \subset A$ et $\lambda(B) = \lambda(A)$ par additivité de la mesure. On a donc identifié deux boréliens (B, B') encadrant A , et tels que $\lambda(B' \setminus B) = 0$. On peut interpréter B comme un « rétrécissement » de A , et B' un « épaississement » de A , de manière à entrer dans la tribu de Borel.

2. On abandonne maintenant l'hypothèse que $A \in \mathcal{M}(\lambda^*)$ est borné. On peut néanmoins obtenir A comme l'union croissante des ensembles $A_K := A \cap]-K, K[$, qui sont tous dans $\mathcal{M}(\lambda^*)$. La preuve ci-dessus, appliquée à A_K , montre que chaque $A_K \in \overline{\mathcal{B}(\mathbb{R})}$. Comme $\overline{\mathcal{B}(\mathbb{R})}$ est une tribu, leur union $A = \bigcup_{K \geq 1} A_K \in \overline{\mathcal{B}(\mathbb{R})}$. \square

Proposition 4.8. *Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction borélienne. Supposons que la fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ est égale à f λ -presque partout. Alors g est mesurable pour la tribu complétée $\overline{\mathcal{B}(\mathbb{R})}$.*

Démonstration. Par hypothèse, il existe un ensemble borélien $A \in \mathcal{B}(\mathbb{R})$ tel que $\lambda(A) = 0$, et tel que $f|_{A^c} = g|_{A^c}$. Pour tout borélien $B \in \mathcal{B}(\mathbb{R})$, l'image réciproque de B par g peut se décomposer en :

$$\begin{aligned} g^{-1}(B) &= (g^{-1}(B) \cap A^c) \cup (g^{-1}(B) \cap A) \\ &= (f^{-1}(B) \cap A^c) \cup (g^{-1}(B) \cap A). \end{aligned}$$

L'ensemble $(f^{-1}(B) \cap A^c)$ est borélien, et $(g^{-1}(B) \cap A) \subset A$ est λ -négligeable, donc leur union est bien dans la tribu complétée $\overline{\mathcal{B}(\mathbb{R})}$. \square

4.3 Généralisations en dimension supérieure

La mesure extérieure de Lebesgue sur \mathbb{R} a été définie en (4.5) à partir des longueurs d'intervalles ouverts. En dimension $d > 1$, il est naturel de remplacer les intervalles par des pavés

ouverts, qui sont des ouverts de \mathbb{R}^d de la forme :

$$P = \prod_{j=1}^d]a_j, b_j[.$$

On peut aussi considérer les pavés fermés $\bar{P} = \prod_{j=1}^d [a_j, b_j]$. Le volume d'un pavé vaut, par définition, $\text{vol}(P) = \prod_{j=1}^d (b_j - a_j)$. On peut alors définir la mesure extérieure extérieure de Lebesgue sur \mathbb{R}^d , par

$$\forall A \subset \mathbb{R}^d, \quad \lambda^*(A) := \inf \left\{ \sum_{i \in \mathbb{N}} \text{vol}(P_i) ; A \subset \bigcup_{i \in \mathbb{N}} P_i \right\},$$

où on prend l'infimum sur tous les recouvrements de A par une union dénombrable de pavés ouverts. On montre alors un théorème analogue au Thme 4.5 :

- i) la fonction λ^* ci-dessus est une mesure extérieure
- ii) sa tribu engendrée $\mathcal{M}(\lambda^*)$ contient la tribu $\mathcal{B}(\mathbb{R}^d)$. On peut pour utiliser le fait que les boréliens sur \mathbb{R}^d peuvent être engendrés par les « cylindres » de la forme

$$B = \mathbb{R} \times \cdots \times]-\infty, a] \times \mathbb{R} \times \cdots \times \mathbb{R},$$

donc il suffira de montrer que ces cylindres sont λ^* -mesurables. Comme on l'a fait en dimension 1, à partir d'un recouvrement de $A \subset \mathcal{P}(\mathbb{R}^d)$ par une famille dénombrable \mathcal{F} de pavés, et pour tout choix de $\epsilon > 0$, on peut fabriquer deux familles $\mathcal{F}_1, \mathcal{F}_2$ de pavés recouvrant respectivement $A \cap B$ et $A \cap B^c$, et tels que

$$\sum_{P \in \mathcal{F}_1} \text{vol}(P) + \sum_{P \in \mathcal{F}_2} \text{vol}(P) \leq \epsilon + \sum_{P \in \mathcal{F}} \text{vol}(P).$$

Si la famille \mathcal{F} recouvre A à ϵ près, on aura donc

$$\lambda^*(A \cap B) + \lambda^*(A \cap B^c) \leq \lambda^*(A) + 2\epsilon, \quad \text{donc} \quad \lambda^*(A \cap B) + \lambda^*(A \cap B^c) = \lambda^*(A),$$

et donc la mesurabilité de B .

- iii) On veut montrer que la mesure du pavé fermé $\lambda^*(\bar{P}) = \text{vol}(\bar{P})$. Il s'agit alors de vérifier que pour tout recouvrement de \bar{P} par une union finie de pavés ouverts

$$\bar{P} \subset \bigcup_{i=1}^n P_i, \quad \text{on a alors} \quad \text{vol}(\bar{P}) \leq \sum_{i=1}^n \text{vol}(P_i).$$

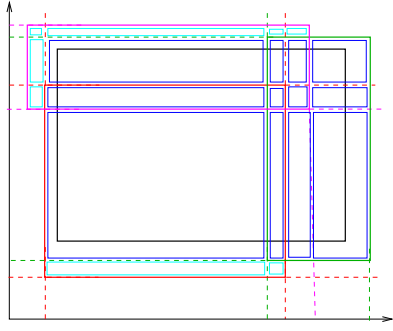


FIGURE 4.2 – Recouvrement d'un pavé P (en noir) en l'union finie de pavés P_i (en rouge, vert, rose). En prolongeant les faces des P_i , on partitionne $\bigcup_i P_i$ en « petits pavés » $P^{(k)}$ (en bleu). Certains sont utilisés pour recouvrir P (bleu foncé), d'autres seront inutiles (bleu clair). Le calcul de $\text{vol}(P_i)$ et de $\text{vol}(P)$ peut se faire en factorisant selon les $d = 2$ dimensions.

Cette vérification n'est pas aussi immédiate qu'en dimension 1. Une façon de procéder est de prolonger les faces de tous les P_i par des hyperplans, de façon à définir des « petits » pavés $(P^{(k)})_{1 \leq k \leq K}$; chaque pavé P_i est partitionné en une union de plusieurs pavés $P^{(k)}$: $P_i = \bigsqcup_{k \in I(i)} P^{(k)}$, certains $P^{(k)}$ pouvant apparaître dans la décomposition de plusieurs P_i . La décomposition forme un « quadrillage », de sorte que la formule

$$\text{vol}(P_i) = \sum_{k \in I(i)} \text{vol}(P^{(k)})$$

se montre en partitionnant chacun des intervalles formant P_i , de façon à factoriser la somme de droite sur les d dimensions. Le pavé P est aussi décomposé dans l'union disjointe d'un certain nombre de pavés, $P = \bigsqcup_{k \in I} P^{(k)}$, et on a encore $\text{vol}(P) = \sum_{k \in I} \text{vol}(P^{(k)})$ par factorisation sur les dimensions. Lorsqu'on compare cette somme à

$$\sum_{i=1}^n \text{vol}(P_i) = \sum_{i=1}^n \sum_{k \in I(i)} \text{vol}(P^{(k)}),$$

on voit que tous les $(P^{(k)})_{1 \leq k \leq K}$ apparaissent dans le membre de droite au moins une fois, et certains peuvent y apparaître plusieurs fois. Cette somme majore donc la somme $\text{vol}(P) = \sum_{k \in I} \text{vol}(P^{(k)})$, dans laquelle chaque $P^{(k)}$ apparaît au plus une fois.

4.4 Quelques propriétés de la mesure de Lebesgue

4.4.1 Invariance par translation

L'espace euclidien \mathbb{R}^d admet une structure de groupe abélien, le groupe des translations. Cette action de groupe permet de caractériser la mesure de Lebesgue.

Théorème 4.9. *La mesure de Lebesgue λ sur \mathbb{R}^d est invariante par translation : pour tout $A \in \mathcal{B}(\mathbb{R}^d)$ et pour tout vecteur $v \in \mathbb{R}^d$, $\lambda(A) = \lambda(A + v)$.*

Inversement, si μ est une mesure sur $\mathcal{B}(\mathbb{R}^d)$ localement finie, et invariante par translation, alors il existe une constante $c \geq 0$ telle que $\mu = c\lambda$.

Ainsi, à une constante multiplicative près, la mesure de Lebesgue est la seule mesure localement finie qui soit invariante par translations.

Démonstration. V. la feuille d'exercices « Week 2 », exercice 1. La première assertion est assez simple à vérifier. La formule $\lambda(A) = \lambda(A + v) := \sigma_v(\lambda)(A)$ est vraie lorsque A est un pavé, par calcul explicite. Le Lemme de classe monotone implique alors que les mesures λ et $\sigma_v(\lambda)$ coïncident sur toute la tribu des boréliens.

Pour montrer la réciproque, on se place sur l'hypercube unité $C = [0, 1]^d$, et on note $c = \mu(C)$. Si on décompose C en un treillis de n^d petits hypercubes $C_k^{(n)}$, on a par invariance par translation et additivité, que $\mu(C_k^{(n)}) = \frac{c}{n^d}$. Pour tout $a_1, \dots, a_d \in]0, 1]$, le pavé $P = \prod_{j=1}^d [0, a_j]$ peut être encadré par deux unions disjointes de petits cubes :

$$\bigsqcup_{k \in I^-} P^{(k)} \subset P \subset \bigsqcup_{k \in I^+} P^{(k)},$$

tels que le nombre $\#I^+ - \#I^- = \mathcal{O}(n^{d-1})$. La différence de volumen entre les deux unions est donc d'ordre $\mathcal{O}(n^{-1})$. En étudiant plus précisément les unions disjointes de droite et de gauche, et en faisant tendre n vers l'infini, on trouve que $\mu(P) = c \prod_{j=1}^d a_j = c\lambda(P)$. \square

4.4.2 Régularité

Lorsqu'on a une mesure μ définie sur la tribu de Borel d'un espace topologique E , on peut étudier le comportement de μ lorsqu'on approche un borélien A par au-dessus par des ouverts, et par en-dessous par des compacts. Le bon comportement de μ vis-à-vis de ces limites définit la notion de régularité de cette mesure.

Définition 4.10. Soit E un espace topologique séparé, et μ une mesure définie sur la tribu de Borel $\mathcal{B}(E)$. Cette mesure est dite

1. extérieurement régulière si, pour tout $A \in \mathcal{B}(E)$, $\mu(A) = \inf \{\mu(O) ; O \text{ ouvert de } E, A \subset O\}$;
2. intérieurement régulière si, pour tout $A \in \mathcal{B}(E)$, $\mu(A) = \sup \{\mu(K) ; K \text{ compact de } E, K \subset A\}$;
3. régulière si elle est à la fois extérieurement régulière et intérieurement régulière.

Proposition 4.11. *La mesure de Lebesgue sur \mathbb{R}^d est régulière.*

Démonstration. On cherche d'abord à montrer la régularité extérieure. Soit $A \in \mathcal{B}(\mathbb{R}^d)$. Si $\lambda(A) = \infty$, le résultat est évident, vu que $\lambda(A) \leq \lambda(O)$ pour tout ouvert $O \supset A$. On suppose maintenant que $\lambda(A) < \infty$. Comme $\lambda(A) = \lambda^*(A)$, on peut trouver un recouvrement de A par des pavés $(P_i)_{i \in \mathbb{N}}$, tel que $\lambda^*(A) + \epsilon \geq \sum_i \text{vol}(P_i)$. Par sous-additivité, l'ouvert $O = \bigcup_i P_i$ vérifiera $\lambda(O) \leq \sum_i \text{vol}(P_i)$, et $\lambda(O) \geq \lambda(A)$ puisqu'il recouvre A . On a donc obtenu

$$\lambda(O) \geq \lambda(A) \geq \lambda(O) - \epsilon,$$

ce qui montre bien la régularité extérieure.

Pour montrer la régularité intérieure on va d'abord supposer que A est contenu dans un hypercube fermé $C = [-n, n]^d$. L'ensemble $C \setminus A$ est un borélien, on peut donc l'approcher par un ouvert O le contenant, de telle façon que $\lambda(C \setminus A) \geq \lambda(O) - \epsilon$. En passant au complémentaire par rapport à C , on en déduit :

$$\lambda(A) \leq \lambda(C \setminus O) + \epsilon.$$

On vérifie que $C \setminus O$ est fermé borné, donc compact. On a ainsi pu approcher A par en-dessous par des compacts.

Si maintenant A n'est pas borné, on construit la suite croissante $A_n := A \cap [-n, n]^d$. On aura alors $\lambda(A) = \lim_{n \rightarrow \infty} \lambda(A_n)$. Pour chaque n , et chaque $\epsilon > 0$, on peut trouver un compact $K_n \subset [-n, n]^d$ tel que $\lambda(K_n) \geq \lambda(A_n) - \epsilon$. Dans les deux cas où $\lambda(A) < \infty$ ou $\lambda(A) = \infty$, on vérifie donc la régularité intérieure. \square

4.4.3 Relation avec l'intégrale de Riemann

Dans l'introduction du cours, nous avons rappelé la définition de l'intégrale de Riemann sur un intervalle $I \subset \mathbb{R}$, d'abord pour les fonctions continues, puis les fonctions réglées,

puis finalement les fonctions dites de Riemann. Dans le cadre le plus général, la définition de l'intégrale de Riemann est donnée de façon variationnelle, mais à partir de fonctions en escalier, qui sont une sous-classe des fonctions étagées. Pour toute fonction $f : I \rightarrow \mathbb{R}$ bornée, on peut définir deux expressions variationnelles :

$$S_-(f) \stackrel{\text{def}}{=} \sup \left\{ \int_I h(x) dx ; h \text{ en escalier, } h \leq f \right\},$$

$$S_+(f) \stackrel{\text{def}}{=} \inf \left\{ \int_I \tilde{h}(x) dx ; \tilde{h} \text{ en escalier, } \tilde{h} \geq f \right\}.$$

Pour une fonction f bornée quelconque, on a forcément $S_-(f) \leq S_+(f)$.

Définition 4.12. Une fonction bornée $f : I \rightarrow \mathbb{R}$ est dite Riemann-intégrable si $S_-(f) = S_+(f)$. Dans ce cas la valeur commune définit l'intégrale de Riemann de f , qu'on notera $S(f)$ ou $I(f)$.

Il est alors naturel de se poser la question suivante :

Une fonction f Riemann-intégrable est-elle borélienne ? Si c'est le cas, son intégrale de Riemann coïncide-t-elle avec son intégrale par la mesure de Lebesgue ?

Théorème 4.13. Soit une fonction $f : I \rightarrow \mathbb{R}$ Riemann-intégrable, d'intégrale de Riemann $S(f)$. Alors f est mesurable pour la tribu complétée $\overline{\mathcal{B}(I)}$, et on a l'égalité

$$S(f) = \int_I f d\lambda.$$

Démonstration. Soit une suite $(h_n)_{n \in \mathbb{N}}$ de fonctions en escalier, telles que $h_n \leq f$ et $\lim_{n \rightarrow \infty} \uparrow S(h_n) = S(f)$. Notons que si on remplace h_n par $h'_n = \max(h_1, \dots, h_n)$, la fonction h'_n est toujours en escalier, elle est toujours majorée par f , donc on aura aussi $\lim_{n \rightarrow \infty} \uparrow S(h'_n) = S(f)$. Autrement dit, on peut supposer que la suite $(h_n)_{n \in \mathbb{N}}$ est croissante.

De façon identique, on peut construire une suite $(\tilde{h}_n)_{n \in \mathbb{N}}$ décroissante de fonctions en escalier, majorant f , et telles que $\lim_{n \rightarrow \infty} \downarrow S(\tilde{h}_n) = S(f)$. Comme les suites $(h_n)_n$ et $(\tilde{h}_n)_n$ sont monotones, on peut définir les fonctions $h_\infty = \lim_n \uparrow h_n$ et $\tilde{h}_\infty = \lim_n \downarrow \tilde{h}_n$. Comme les fonctions h_n, \tilde{h}_n sont en escalier, elles sont boréliennes, donc les fonctions h_∞ et \tilde{h}_∞ sont aussi boréliennes ; elles sont aussi bornées sur I , puisque f l'est, donc elles sont forcément

intégrables par rapport à λ . Le théorème de convergence dominée montre que

$$\begin{aligned}\int h_\infty d\lambda &= \lim_{n \rightarrow \infty} \uparrow \int h_n d\lambda = \lim_{n \rightarrow \infty} \uparrow S(h_n) = S_-(f), \\ \int \tilde{h}_\infty d\lambda &= \lim_{n \rightarrow \infty} \downarrow \int \tilde{h}_n d\lambda = \lim_{n \rightarrow \infty} \downarrow S(\tilde{h}_n) = S_+(f),\end{aligned}$$

où on a utilisé le fait que pour toute fonction en escalier h , l'intégrale $\int_I h(x) dx = \int_I h d\lambda$. Si la fonction f est Riemann-intégrable, ces deux expressions coïncident. On a trouvé une paire de fonctions boréliennes h_∞ et \tilde{h}_∞ qui encadrent f , et telles que $\int h_\infty d\lambda = \int \tilde{h}_\infty d\lambda$. Ceci impose que $\tilde{h}_\infty = h_\infty = f$ λ -p.p.. La proposition 4.8 montre alors que f est $\overline{\mathcal{B}(\mathbb{R})}$ -mesurable. De plus, son intégrale de Lebesgue $\int f d\lambda = \int h_\infty d\lambda = S(f)$. \square

4.4.4 Il existe un ensemble non mesurable pour la mesure de Lebesgue

On a vu que les ensembles de la tribu de Lebesgue peuvent être compliqués : ensembles $G_\delta, F_\sigma, G_{\delta\sigma}, \dots$. On a en plus rajouté aux ensembles boréliens les ensembles λ -négligeables, de manière à former la tribu complétée $\overline{\mathcal{B}(\mathbb{R})} = \mathcal{M}(\lambda^*)$. Il est légitime de se demander si cette tribu complétée, ou même la tribu des boréliens sur \mathbb{R} , ne contient déjà pas toutes les parties de \mathbb{R} .

Nous montrons que ce n'est pas le cas, en exhibant un sous-ensemble de \mathbb{R} qui ne peut pas faire partie de cette tribu. Cet ensemble est « compliqué » : pour le construire on doit faire appel à l'axiome du choix de la théorie des ensembles.

Considérons l'espace \mathbb{R}/\mathbb{Q} des classes d'équivalence des réels modulo les rationnels : $x \sim y$ ssi $x - y \in \mathbb{Q}$. Pour chaque classe d'équivalence $a \in \mathbb{R}/\mathbb{Q}$, on peut choisir un représentant x_a , qu'on peut prendre dans l'intervalle $[0, 1]$. C'est ce choix d'un représentant pour tout a , qui nécessite l'axiome du choix. En effet, \mathbb{R}/\mathbb{Q} est indénombrable. On définit alors l'ensemble de ces représentats :

$$F := \{x_a ; a \in \mathbb{R}/\mathbb{Q}\} \subset [0, 1],$$

qui est un ensemble indénombrable, puisque tous les x_a sont forcément différents.

Théorème 4.14. *L'ensemble F n'est pas contenu dans la tribu de Lebesgue $\overline{\mathcal{B}(\mathbb{R})}$.*

Démonstration. Supposons au contraire que F est mesurable pour la tribu de Lebesgue. Comme la tribu de Lebesgue est invariante par translation, pour tout rationnel q l'ensemble $(q + F)$ sera aussi mesurable. Comme tout $x \in \mathbb{R}$ appartient à une classe $a \in \mathbb{R}/\mathbb{Q}$, et que

chaque classe regroupe tous les éléments $x_a + \mathbb{Q}$, on vérifie que

$$\bigcup_{q \in \mathbb{Q}} (q + F) = \mathbb{R}.$$

Si $\lambda(F) = 0$, alors par invariance par translation on aura pour tout rationnel $\lambda(q + F) = 0$, et par σ -additivité $\lambda(\mathbb{R}) = 0$, ce qui est absurde. Donc on a forcément $\lambda(F) = c > 0$. Comme $F \subset [0, 1]$, on doit cependant avoir $0 < c \leq 1$. L'invariance par translation de λ impose alors que pour tout rationnel q , $\lambda(q + F) = c$.

Par ailleurs, les ensembles $(q + F)$ sont tous disjoints (si $q + x_a = q' + x_{a'}$ alors $x_a \sim x_{a'}$, donc $a = a'$ et $x_a = x_{a'}$). A partir de l'inclusion

$$\bigcup_{q \in \mathbb{Q} \cap [0, 1]} (q + F) \subset [0, 2]$$

on déduit par σ -additivité

$$\sum_{q \in \mathbb{Q} \cap [0, 1]} \lambda(q + F) \leq \lambda([0, 2]) = 2,$$

ce qui est impossible puisque le membre de gauche vaudra $\infty \cdot c = \infty$.

On a donc une contradiction au fait que F est mesurable. Donc F ne peut pas être mesurable. \square

4.5 Construction plus générale d'une mesure

Dans cette section on généralisera la construction d'une mesure extérieure μ^* sur $\mathcal{P}(E)$, à partir d'une mesure μ définie sur une classe restreinte de parties de E . Cette classe restreinte de parties de E doit former une *semi-algèbre* sur un ensemble E . Les preuves sont laissées aux lecteurs, qui pourront s'inspirer des preuves données dans le cas de la mesure de Lebesgue.

Définition 4.15. Soit E un ensemble. Une semi-algèbre \mathcal{S} sur E est un ensemble de parties de E , tel que :

- (i) $\emptyset \in \mathcal{S}$;
- (ii) \mathcal{S} est stable par intersections finies ;
- (iii) $\forall A \in \mathcal{S}$, il existe B_1, \dots, B_n des éléments disjoints de \mathcal{S} , tels que $A^c = \bigsqcup_{i=1}^n B_i$.

Remarquons que cette dernière propriété est une version affaiblie de l'invariance par passage au complémentaire.

Exemple 4.16. La mesure de Lebesgue est au départ définie sur les intervalles ouverts et fermés de \mathbb{R} : $\lambda([a, b[) = \lambda([a, b]) = b - a$ pour tous réels $a < b$. L'ensemble des intervalles (ouverts et fermés) de \mathbb{R} forme bien une semi-algèbre.

Définition 4.17. Une mesure définie sur une semi-algèbre \mathcal{S} est une application $\mu : \mathcal{S} \rightarrow [0, \infty]$ telle que

1. $\mu(\emptyset) = 0$;
2. μ est σ -additive sur la semi-algèbre \mathcal{S} : si $A \in \mathcal{S}$ se décompose en $A = \bigsqcup_{i \in \mathbb{N}} B_i$ avec les $B_i \in \mathcal{S}$, alors $\mu(A) = \sum_{i \in \mathbb{N}} \mu(B_i)$.

La semi-algèbre \mathcal{S} engendre naturellement une algèbre :

Définition 4.18. Soit E un ensemble. Une algèbre \mathcal{A} sur E est un ensemble de parties de E tels que :

- (i) $\emptyset \in \mathcal{A}$;
- (ii) \mathcal{A} est stable par passage au complémentaire
- (iii) \mathcal{A} est stable par union finie (et donc par intersection finie).

Si on ajoute à une sous-algèbre \mathcal{S} toutes les unions finies de ses éléments, on obtient une algèbre, qu'on appelle l'algèbre engendrée par \mathcal{S} .

Remarque : La seule différence entre la notion d'algèbre et celle de tribu, est l'absence de stabilité par unions *dénombrables*.

Proposition 4.19. Soit μ une mesure définie sur une sous-algèbre \mathcal{S} . Alors on peut étendre μ à l'algèbre \mathcal{A} engendrée par \mathcal{S} , en posant :

$$\forall A \in \mathcal{A}, \text{ si } A = \bigsqcup_{i=1}^n B_i \text{ (avec } B_i \in \mathcal{S} \text{ disjoints),} \quad \mu(A) := \sum_{i=1}^n \mu(B_i).$$

La mesure μ étendue est alors σ -additive sur l'algèbre \mathcal{A} (si $A = \bigsqcup_{i \in \mathbb{N}} B_i$ avec $A, B_i \in \mathcal{A}$, alors $\mu(A) = \sum_{i \in \mathbb{N}} \mu(B_i)$).

Exemple 4.20. L'algèbre engendrée par la semi-algèbre des intervalles de \mathbb{R} , est celle des unions finies d'intervalles de \mathbb{R} .

A partir d'une mesure σ -additive sur une algèbre \mathcal{A} , on va pouvoir construire une mesure extérieure.

Théorème 4.21. *Soit μ une mesure σ -additive sur une algèbre $\mathcal{A} \subset \mathcal{P}(E)$. Pour tout $A \in \mathcal{P}(E)$, posons*

$$\mu^*(A) := \inf \left\{ \sum_{i=0}^{\infty} \mu(A_i) ; A_i \subset \mathcal{A}, A \subset \bigcup_{i=0}^n A_i \right\}.$$

Alors μ^ définit une mesure extérieure (ous σ -sous-additive) sur E , qui prolonge μ .*

Exemple 4.22. La mesure extérieure λ^* avait été définie en (4.5) de façon similaire, à partir de recouvrements de A par des unions dénombrables d'intervalles ouverts. On remarque que, partant de la mesure de Lebesgue sur la semi-algèbre des intervalles ouverts ou fermés, la définition ci-dessus inclura également comme possible A_i les intervalles fermés. Comme les singletons $\{a\}$ sont de mesure de Lebesgue nulle, cela ne change pas le résultat de λ^* .

Une fois qu'on dispose d'une mesure extérieure, la construction de la tribu $\mathcal{M}(\mu^*)$ et de la mesure $\mu = \mu^*_{|\mathcal{M}(\mu^*)}$ a été présentée dans la section 4.1.

4.6 Mesures finies sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$; intégrale de Stieltjes

Comme expliqué dans l'Exemple 2.18, on peut décrire toute mesure borélienne finie sur \mathbb{R} par sa fonction de répartition.

Théorème 4.23. *i) Soit μ une mesure finie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Sa fonction de répartition est définie par :*

$$\forall x \in \mathbb{R}, \quad F_\mu(x) = \mu([-\infty, x]).$$

Cette fonction est croissante, bornée, continue à droite, et $\lim_{x \rightarrow -\infty} F_\mu(x) = 0$.

ii) Inversement, soit $F : \mathbb{R} \rightarrow \mathbb{R}$ une fonction croissante, bornée, continue à droite, telle que $\lim_{x \rightarrow -\infty} F_\mu(x) = 0$. Alors il existe une unique mesure borélienne finie μ sur \mathbb{R} telle que $F = F_\mu$.

L'intégrale par rapport à la mesure μ sera parfois notée

$$\int f(x) \mu(dx) = \int f(x) dF(x).$$

Il faut faire un peu attention à cette notation : la fonction F n'est en général pas dérivable au sens « classique ». On appelle cette intégrale *l'intégrale de Stieltjes par rapport à la fonction F* . On aura alors

$$F(b) - F(a) = \mu([a, b]) \quad (4.8)$$

(noter l'asymétrie entre les deux bornes de l'intervalle), tandis que

$$F(b) - F(a-) = \mu([a, b]), \quad \text{où } F(a-) = \lim_{x \rightarrow a-} F(x).$$

Démonstration. i) Les propriétés de croissance, de borne uniforme sont évidentes. La continuité à droite est due à la régularité extérieure de la mesure finie μ , qu'on montre ci-dessous dans la Proposition 4.24. La valeur en $-\infty$ est due au fait que $\bigcap_{a \in \mathbb{R}}]-\infty, a] = \emptyset$, et au fait que

$$0 = \mu\left(\bigcap_{a \in \mathbb{R}}]-\infty, a]\right) = \lim_{a \rightarrow -\infty} \downarrow \mu(]-\infty, a]).$$

ii) L'exemple 2.18 montre l'unicité de la mesure associée à F . Pour montrer son existence, il faut construire la mesure extérieure correspondante. On définit, pour tout $A \subset \mathbb{R}$:

$$\mu^*(A) := \inf \left\{ \sum_{i \in \mathbb{N}} F(b_i) - F(a_i), \quad A \subset \bigcup_{i \in \mathbb{N}}]a_i, b_i] \right\}.$$

On notera les modifications par rapport à la définition de λ^* :

- la longueur $(b_i - a_i)$ des intervalles est remplacée par $F(b_i) - F(a_i)$.
- on utilise des intervalles semi-ouverts $]a_i, b_i]$ au lieu des intervalles ouverts. Ce choix est bien sûr à relier avec la formule (4.8).

Malgré ces différences, la preuve que μ^* est une mesure extérieure est essentiellement identique au cas de λ^* . On montre aussi que les intervalles $]-\infty, a]$ sont dans la tribu $\mathcal{M}(\mu^*)$, montrant que cette tribu contient la tribu de Borel. La restriction de μ^* à cette tribu forme donc une mesure. Vérifions enfin que $\mu([a, b]) = F(b) - F(a)$, ce qui montrera que $F = F_\mu$.

Par le principe variationnel, on a déjà $\mu^*([a, b]) \leq F(b) - F(a)$. Montrons l'inégalité dans l'autre sens. On va considérer un recouvrement de $[a, b]$ par une union d'intervalles $]x_i, y_i]$. Prenons $\varepsilon > 0$ petit. Pour tout $i \in \mathbb{N}$, par continuité à droite de F en y_i , on peut choisir $y'_i > y_i$ tel que

$$F(y'_i) \leq F(y_i) + \varepsilon 2^{-i}.$$

Comme $\bigcup_{i \in \mathbb{N}}]x_i, y_i]$ recouvre $[a, b]$, on voit que $\bigcup_{i \in \mathbb{N}}]x_i, y'_i]$ recouvre l'intervalle compact $[a +$

$\varepsilon, b]$. Donc on peut en extraire un recouvrement fini $\bigcup_{i=0}^N]x_i, y'_i[$, avec $N = N(\varepsilon) < \infty$. On déduit de ce recouvrement fini :

$$\begin{aligned} F(b) - F(a + \varepsilon) &\leq \sum_{i=0}^N (F(y'_i) - F(x_i)) \leq \sum_{i=0}^{\infty} (F(y'_i) - F(x_i)) \\ &\leq 2\varepsilon + \sum_{i=0}^{\infty} (F(y_i) - F(x_i)). \end{aligned}$$

En prenant maintenant $\varepsilon \searrow 0$, on trouve

$$F(b) - F(a) \leq \sum_{i=0}^{\infty} (F(y_i) - F(x_i)).$$

Comme cette inégalité est valable pour tout recouvrement $\bigcup_{i \in \mathbb{N}}]x_i, y_i]$ de $]a, b]$, on peut prendre l'infimum sur les recouvrements, et on obtient l'inégalité voulue, $F(b) - F(a) \leq \mu^*(]a, b])$. \square

Après avoir étudié les mesures finies sur \mathbb{R} , on va donner un résultat général de régularité pour les mesures finies définies sur un espace métrique (E, d) .

Proposition 4.24. *Soit (E, d) un espace métrique, et soit μ une mesure finie sur $(E, \mathcal{B}(E))$. Alors pour tout $A \in \mathcal{B}(E)$,*

$$\begin{aligned} \mu(A) &= \inf \{ \mu(U) : U \text{ ouvert}, A \subset U \} \\ &= \sup \{ \mu(F) : F \text{ fermé}, F \subset A \}. \end{aligned}$$

Ainsi, une telle mesure est extérieurement régulière. La seconde propriété n'est pas exactement la régularité intérieure (on n'approxime pas A par des compacts, mais par des fermés).

Démonstration. Appelons \mathcal{O} l'ensemble des ouverts de E , et soit \mathcal{C} la classe des ensembles boréliens qui vérifient la propriété ci-dessus. Il suffit de montrer que cette classe contient les ouverts \mathcal{O} , et qu'elle forme une tribu.

Si A est ouvert, la première propriété est évidente. Considérons pour tout $n \geq 1$ l'ensemble :

$$F_n := \{x \in E : d(x, A^c) \geq 1/n\}.$$

Comme la fonction $x \mapsto d(x, A^c)$ est continue, F_n est un fermé, inclus dans A . Comme ce dernier peut être défini par $A = \{x \in E : d(x, A^c) > 0\}$, on voit que la suite croissante

$(F_n)_{n \geq 1}$ recouvre A :

$$A = \bigcup_{n \geq 1} F_n = \lim_n \uparrow F_n,$$

donc les propriétés des mesures donnent bien $\mu(A) = \lim_n \uparrow \mu(F_n)$. Ainsi, tout ouvert appartient à la classe \mathcal{C} . On a en particulier $\emptyset \in \mathcal{C}$.

Les propriétés définissant \mathcal{C} sont symétriques entre ouverts et fermés, ce qui montre que si $A \in \mathcal{C}$, son complémentaire $A^c \in \mathcal{C}$.

Finalement, soit $(A_n)_{n \in \mathbb{N}}$ est une famille d'éléments de \mathcal{C} : on veut montrer que leur union $A := \bigcup_n A_n$ est aussi dans la classe. Fixons $\varepsilon > 0$. Pour chaque n on trouve un ouvert $U_n \supset A_n$ tel que $\mu(U_n) \leq \mu(A_n) + \varepsilon/2^n$. On a alors :

$$\mu\left(\bigcup_n U_n \setminus \bigcup_k A_k\right) \leq \sum_n \mu(U_n \setminus A_n) \leq 2\varepsilon.$$

On peut donc approcher A à 2ε près par l'ouvert $\bigcup_n U_n$, donc A vérifie la première propriété.

Comme la mesure est finie, $\mu(A) < \infty$, donc pour $N = N(\varepsilon)$ assez grand,

$$\mu\left(\bigcup_{n=0}^N A_n\right) \geq \mu(A) - \varepsilon.$$

Pour chaque $n \leq N$ on trouve un fermé $F_n \subset A_n$ tel que $\mu(A_n \setminus F_n) \leq \varepsilon 2^{-n}$. L'ensemble fermé $F = \bigcup_{n=0}^N F_n$ satisfait donc :

$$\mu\left(\bigcup_{n=0}^N A_n \setminus F\right) \leq \sum_{n=0}^N \mu(A_n \setminus F_n) \leq 2\varepsilon.$$

On en déduit donc :

$$\mu(A \setminus F) \leq 3\varepsilon,$$

montrant que A satisfait la seconde propriété. On a donc montré que $A \in \mathcal{C}$, donc que \mathcal{C} est une tribu. \square

Extension aux mesures localement finies sur \mathbb{R} : mesures de Radon

On peut également se servir d'une fonction de répartition modifiée pour décrire les *mesures de Radon* sur \mathbb{R} , qui sont les mesures *boréliennes localement finies*. Il n'est plus possible en

général de définir la fonction de répartition en utilisant les intervalles de longueur infinie $] -\infty, a]$, car la mesure de tous ces intervalles peut être infinie (c'est le cas par exemple pour la mesure λ). Il faut alors définir cette fonction en utilisant des intervalles finis, qui ont une mesure finie. On définit par exemple :

$$F(x) = \begin{cases} \mu([0, x]), & x \geq 0, \\ -\mu([x, 0]), & x < 0. \end{cases}$$

La fonction F est toujours croissante, mais elle n'est plus forcément bornée, en particulier sa limite en $-\infty$ peut être égale à $-\infty$. Elle s'annule toujours en $x = 0$, même lorsque la mesure μ est une masse de Dirac en l'origine (on a alors $F(0) = 0$, mais $F(0-) = -1$).

Les mesures de Radon sur \mathbb{R} comme formes linéaires

Il existe une interprétation (très importante en analyse) des mesures de Radon sur \mathbb{R} , comme des formes linéaires sur l'espace des fonctions continues à support compact $C_c(\mathbb{R})$. En effet, soit μ mesure de Radon sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Toute fonction continue f à support compact vérifie $\text{supp } f \subset [-K, K]$ pour un certain $K > 0$. Etant continue, la fonction f est borélienne. De plus, elle est bornée par

$$\forall x \in \mathbb{R}, \quad |f(x)| \leq \|f\|_{\text{sup}} \mathbb{1}_{[-K, K]}(x).$$

Comme μ est finie sur le compact $[-K, K]$, la fonction f est donc μ -intégrable. La Proposition 3.20 montre que l'application

$$f \mapsto J_\mu(f) = \int f d\mu$$

est linéaire. Cette application définit donc une forme linéaire sur l'espace fonctionnel $C_c(\mathbb{R})$. De plus, si f est une fonction positive, cette forme linéaire est positive. Enfin, la borne ci-dessus sur $|f(x)|$ implique :

$$|J_\mu(f)| \leq \|f\|_{\text{sup}} \int \mathbb{1}_{[-K, K]} d\mu = \|f\|_{\text{sup}} \mu([-K, K]),$$

valable pour toute fonction continue supportée dans $[-K, K]$.

Affirmation 4.25. On peut munir l'espace $C_c(\mathbb{R})$ d'une topologie, telle que la forme linéaire $J_\mu : C_c(\mathbb{R}) \rightarrow \mathbb{R}$ est continue.

Le théorème de Riesz-Markov affirme que cette interprétation des mesures de Radon est en

fait un isomorphisme.

Théorème 4.26. (*Riesz-Markov*) Soit J une forme linéaire positive sur $C_c(\mathbb{R})$. Il existe alors une unique mesure de Radon μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, telle que $J = J_\mu$. Autrement dit,

$$\forall f \in C_c(\mathbb{R}), \quad J(f) = \int f d\mu.$$

Exemple 4.27. L'intégrale de Riemann $\int f(x) dx$ définit aussi une forme linéaire

Remarque 4.28. Ce résultat n'est pas spécifique aux mesures de Radon sur \mathbb{R} . On peut le généraliser à tout espace métrique (E, d) séparable et localement compact².

On va maintenant montrer un résultat de régularité des mesures de Radon sur \mathbb{R} , qui se généralise à tout espace métrique (E, d) séparable localement compact. Un tel espace peut être vu comme la « limite » d'une suite croissante d'ensembles ouverts précompacts³.

Lemme 4.29. Soit (E, d) un espace métrique séparable localement compact. Alors il existe une suite croissante de compacts $(L_n)_{n \in \mathbb{N}}$ tels que, pour tout n , $L_n \subset \overset{\circ}{L}_{n+1}$, et $E = \bigcup_{n \geq 0} L_n = \bigcup_{n \geq 0} \overset{\circ}{L}_n$.

Démonstration. Dans le cas de l'espace euclidien $E = \mathbb{R}^d$, on peut bien sûr prendre $L_n = B(0, n)$ la boule centrée en l'origine, de rayon $n \geq 1$.

Dans le cas d'un espace métrique plus général, il faut faire plus attention, car les grandes boules ne sont pas forcément compactes. Montrons que E est une réunion dénombrable d'une suite croissante de compacts. Soit $(x_p)_{p \in \mathbb{N}}$ une suite dense dans E . Le fait que E est localement compact montre que pour tout $x \in E$, si on prend $r > 0$ assez petit, la boule fermée $\bar{B}(x, r)$ est compacte (mais si r est trop grand on peut perdre la compacité). On peut alors définir les couples d'entiers

$$I = \{(p, k) : \bar{B}(x_p, 2^{-k}) \text{ est compacte}\},$$

sachant que pour tout $p \in \mathbb{N}$, $(p, k) \in I$ si k est assez grand. Soit $x \in E$ quelconque, et soit $r > 0$ tel que $\bar{B}(x, r)$ est compacte. Par densité, on peut trouver une suite $(x_{p_n})_{n \geq 1}$ convergeant vers x lorsque $n \rightarrow \infty$. On peut trouver $N \geq 1$ tel que pour tout $n \geq N$, $x_{p_n} \in B(x, r/8)$. L'inégalité triangulaire montre alors que si on prend $k \geq 1$ tel que $r/4 \leq 2^{-k} < r/2$, alors la

2. On rappelle qu'un espace métrique est séparable s'il admet un sous-ensemble dénombrable dense ; en conséquence, il admet aussi une base dénombrable d'ouverts. Un espace est localement compact si tout point admet un voisinage compact.

3. Un sous-ensemble A d'un espace topologique E est dit précompact si sa fermeture \bar{A} est compacte.

boule $\bar{B}(x_{p_n}, 2^{-k}) \subset \bar{B}(x, r)$, donc elle est compacte, et elle contient x . Ceci montre que

$$E = \bigcup_{(p,k) \in I} \bar{B}(x_p, 2^{-k}).$$

Comme l'ensemble d'indices I est dénombrable, on peut l'obtenir comme union dénombrable d'ensembles finis croissants $(I_n)_{n \in \mathbb{N}}$. Si on pose

$$K_n := \bigcup_{(p,k) \in I_n} \bar{B}(x_p, 2^{-k}),$$

on obtient bien une suite croissante de compacts, dont l'union recouvre E . Pour obtenir une suite (L_n) de compacts satisfaisant la propriété $L_n \subset \overset{\circ}{L}_{n+1}$, il faut modifier un peu la construction. On procède par récurrence : on commence par $L_1 = K_1$. Supposant qu'on ait construit L_n , on peut recouvrir le compact $K_{n+1} \cup L_n$ par une union

$$\bigcup_{x \in K_{n+1} \cup L_n} V_x,$$

où chaque V_x est un voisinage ouvert précompact de x . Par compacité de $K_{n+1} \cup L_n$ et Borel-Lebesgue, on peut restreindre l'union à une union finie $V_1 \cup \dots \cup V_m$ de voisinages ouverts précompacts, et prendre $L_{n+1} = \bar{V}_1 \cup \bar{V}_2 \cup \dots \cup \bar{V}_m$, qui est compact et satisfait bien $L_n \subset \overset{\circ}{L}_{n+1}$. \square

En utilisant ce lemme, on montre facilement la propriété suivante de régularité des mesures de Radon :

Proposition 4.30. *Soit (E, d) un espace métrique séparable localement compact, et μ une mesure de Radon sur $(E, \mathcal{B}(E))$. Alors cette mesure est régulière.*

Démonstration. Soit A un ensemble borélien de E . On va utiliser le recouvrement de E par une suite croissante $(\overset{\circ}{L}_n)_{n \geq 0}$ comme dans le Lemme ci-dessus. Restreignons à la fois A et μ à l'ouvert $\overset{\circ}{L}_n$. Comme la mesure μ est localement compacte, la mesure $\mu_n := \mu|_{\overset{\circ}{L}_n}$ est finie. La Proposition 4.24 nous dit qu'on peut approximer le borélien $A_n := A \cap \overset{\circ}{L}_n$ par des ouverts $O_n \subset \overset{\circ}{L}_n$, tels que $A_n \subset O_n$, et

$$\mu(O_n \setminus A_n) \leq \varepsilon 2^{-n}.$$

La réunion $O = \bigcup_{n \geq 0} O_n$ est un ouvert de E , qui satisfait :

$$\mu(O \setminus A) = \mu\left(\bigcup_n O_n \setminus \bigcup_k A_k\right) \leq \sum_{n \geq 0} \mu(O_n \setminus A_n) \leq 2\varepsilon.$$

Ceci montre que μ est extérieurement régulière.

De même, comme $\mu(A_n)$ est fini pour tout n , on peut trouver un fermé $F_n \subset A_n$, donc compact puisque inclus dans L_n , tel que

$$\mu(A_n \setminus F_n) \leq \varepsilon 2^{-n}.$$

Si $\mu(A) = \infty$, alors $\mu(A_n) \rightarrow \infty$ puisque $\mu(A) = \lim_n \uparrow \mu(A_n)$; la propriété ci-dessus montre que $\mu(F_n) \rightarrow \infty$ aussi.

Si $\mu(A) < \infty$, on montre comme dans la Proposition 4.24 que pour N assez grand,

$$\mu(A) \leq \mu(A_N) + \varepsilon,$$

On a alors

$$\mu\left(\bigcup_{n=0}^N A_N \setminus \bigcup_{n=0}^N F_n\right) \leq \sum_{n=0}^N \mu(A_n \setminus F_n) \leq 2\varepsilon,$$

ce qui montre que

$$\mu\left(A \setminus \bigcup_{n=0}^N F_n\right) \leq 3\varepsilon.$$

Ceci montre la régularité intérieure de μ . □

Remarque 4.31. Pour tout ouvert $U \subset \mathbb{R}$, on a les caractérisations

$$\begin{aligned} \mu(U) &= \sup \{ \mu(K), K \subset U \text{ compact} \} \\ &= \sup \{ J_\mu(f), f \in C_c(\mathbb{R}), 0 \leq f \leq \mathbb{1}_U \}. \end{aligned}$$

5 Espaces L^p

Nous avons déjà défini l'espace $\mathcal{L}^1(E, \mathcal{A}, \mu)$ des fonctions intégrables sur un certain espace mesuré : il s'agit des fonctions $f : (E, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurables telles que $\int |f| d\mu < \infty$. On va, dans ce chapitre, généraliser cette notion à celle des fonctions mesurables f telles qu'une certaine puissance $|f|^p$ est intégrable (ici $p > 1$ n'est pas nécessairement entier).

Pourquoi cette généralisation ? D'une part, les équations aux dérivées partielles non-linéaires font souvent apparaître des puissances de la fonction inconnue, donc il peut être utile d'imposer, ou de montrer, des propriétés d'intégrabilité de cette puissance. D'autre part, l'échelle des espaces fonctionnels \mathcal{L}^p correspondants nous fournira certaines valeurs « spécialement intéressantes ». Par exemple, l'espace \mathcal{L}^2 des fonctions de carré intégrable (on dit souvent « de carré sommable ») sera naturellement muni d'une structure hilbertienne, qui le rendra très utile. Pour les valeurs $p \neq 2$, on aura au moins une structure d'espace de Banach, déjà intéressante pour montrer des théorèmes d'existence, ou de convergence.

Ces structures d'espaces fonctionnels (Banach, Hilbert) ne s'appliqueront en fait pas directement aux espaces \mathcal{L}^p , mais aux espaces L^p de classes d'équivalence de fonctions égales entre elles μ -presque partout. Au départ ce passage aux classes d'équivalence de fonctions peut sembler compliqué, inconfortable, mais il présente tellement d'avantages qu'on apprend vite à manipuler les espaces L^p comme s'il s'agissait de simples espaces de fonctions, en oubliant parfois que chacun de leurs éléments est une classe d'équivalence de fonctions. On identifie souvent un élément de L^p avec un certain représentant dans la classe d'équivalence.

5.1 Définition des espaces L^p ; inégalité de Hölder

5.1.1 Définition de L^p

Dans tout ce chapitre on se place sur un espace mesuré (E, \mathcal{A}, μ) . Pour tout réel $p \geq 1$, considérons l'espace fonctionnel

$$\mathcal{L}^p(E, \mathcal{A}, \mu) := \left\{ f : E \rightarrow \mathbb{R} \text{ mesurable, telle que } \int |f|^p d\mu < \infty \right\}.$$

On notera, pour tout f mesurable :

$$\|f\|_p := \left(\int |f|^p d\mu \right)^{1/p},$$

en se souvenant que $\|f\|_p$ dépend de la mesure μ . Plus loin on va vouloir montrer que cette expression a les propriétés d'une norme. Cependant, dans la proposition 3.15 on a vu que $\|f\|_p = 0$ n'est pas équivalent au fait que $f = 0$, mais seulement que $|f|^p = 0$ p.p., donc que $f = 0$ p.p. Autrement dit, $\|\bullet\|_p$ ne peut pas constituer une norme sur \mathcal{L}^p , mais au mieux une semi-norme.

C'est une des raisons pour lesquelles on est tenté de définir une relation d'équivalence sur \mathcal{L}^p :

$$f, g \in \mathcal{L}^p(E, \mathcal{A}, \mu) \text{ sont équivalentes : } f \sim g, \text{ ssi } f = g \text{ } \mu\text{-p.p.}$$

Il est clair que cette propriété est une relation d'équivalence : si $f = g$ sur $E \setminus N_1$ avec N_1 négligeable, et $g = h$ sur $E \setminus N_2$ avec N_1, N_2 μ -négligeables, alors $f = h$ sur $E \setminus (N_1 \cup N_2)$, et on a bien sûr $\mu(N_1 \cup N_2) = 0$.

On est alors en capacité de définir l'espace quotient :

$$L^p(E, \mathcal{A}, \mu) := \mathcal{L}^p(E, \mathcal{A}, \mu) / \sim,$$

chaque classe d'équivalence $[f] \in L^p$ étant constituée de toutes les fonctions mesurables $g = f$ μ -p.p.. On pourrait en fait assouplir la définition à toutes les fonctions g égales à f p.p., sans nécessairement être mesurables.

Dans la suite on adoptera un léger abus de langage : on ne notera pas les éléments de L^p par $[f]$ (classe d'équivalence de la fonction $f \in \mathcal{L}^p$), mais plutôt par f , autrement dit on choisira une fonction représentant cette classe. On aura donc des expressions comme

$$f \in L^p, \quad \text{au lieu de } [f] \in L^p.$$

5.1.2 Définition de l'espace L^∞

Lorsque $p \gg 1$, l'intégrale $\int |f|^p d\mu$ va être dominée par les grandes valeurs de $|f|$; cependant, pas forcément par les valeurs maximales de $|f|$, lorsque celles-ci sont prises sur un ensemble μ -négligeable. Ceci nous amène à la définition de l'espace des fonctions « essentiellement bornées », ou « bornées presque partout » :

$$\mathcal{L}^\infty(E, \mathcal{A}, \mu) := \{f : E \rightarrow \mathbb{R} \text{ mesurable, } \exists C \in \mathbb{R}_+, |f| \leq C \text{ } \mu\text{-p.p.}\},$$

et on définira de même la quantité

$$\|f\|_\infty := \inf \{C \in \mathbb{R}_+, |f| \leq C \mu - p.p.\}.$$

Il faut bien comprendre que :

- en général $\|f\|_\infty$ est strictement plus petit que $\sup_{x \in E} |f(x)|$
- la valeur de $\|f\|_\infty$ ne dépend pas seulement de la fonction $f : E \rightarrow \mathbb{R}$, mais également de la mesure μ . Par exemple pour la mesure de Dirac $\mu = \delta_0$, $\|f\|_\infty = |f(0)|$. La dépendance dans la mesure sera implicite, elle n'est pas indiquée dans la notation $\|\bullet\|_\infty$.
- si $f \sim g$ μ -p.p., on aura $\|f\|_\infty = \|g\|_\infty$.

Cette dernière propriété nous incite, comme au paragraphe précédent, à définir l'espace des classes d'équivalence de fonctions bornées presque partout : on notera

$$L^\infty(E, \mathcal{A}, \mu) = \mathcal{L}^\infty(E, \mathcal{A}, \mu) / \sim.$$

Remarque 5.1. Dans chaque classe $[f] \in L^\infty$, on peut trouver un représentant $\tilde{f} \in [f]$ tel que $|\tilde{f}| \leq \|f\|_\infty$ sur tout E . En effet, l'ensemble N des $x \in E$ tels que $|f(x)| > \|f\|_\infty$ est μ -négligeable, on vérifie que la fonction $\tilde{f} = f \mathbb{1}_{E \setminus N}$ est bien équivalente à f .

Dans la suite on traitera parfois l'espace L^∞ en même temps que les L^p avec $p \in [1, \infty[$, c'est-à-dire qu'on considèrera les exposants $p \in [1, \infty]$. Mais on verra que l'espace L^∞ diffère néanmoins des L^p sur plusieurs points.

5.1.3 Inégalités de Hölder

Le fait de considérer une famille (ou « échelle ») d'espaces fonctionnels ($L^p, p \in [1, \infty]$) va nous induire à essayer de comparer ces espaces les uns avec les autres : sont-ils inclus les uns dans les autres ? Peut-on obtenir des informations sur une norme $\|f\|_p$, à partir d'informations sur une autre norme $\|f\|_{p'}$?

Le premier résultat en ce sens, relativement simple à prouver mais fondamental dans la suite, est l'inégalité de Hölder, qui relie les normes sur L^p et sur L^q , où q est l'exposant conjugué à p , vérifiant :

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

Cet exposant conjugué est parfois noté p' . Ici on peut prendre $p, q \in [1, \infty]$.

Théorème 5.2. (*Inégalité de Hölder*) Soient $p, q \in [1, \infty]$ des exposants conjugués. Alors, si f, g sont deux fonctions mesurables de (E, \mathcal{A}) vers $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on a l'inégalité :

$$\int |fg| d\mu \leq \|f\|_p \|g\|_q. \quad (5.1)$$

En particulier, si $f \in L^p(E, \mathcal{A}, \mu)$ et $g \in L^q(E, \mathcal{A}, \mu)$, alors leur produit $fg \in L^1(E, \mathcal{A}, \mu)$.

Démonstration. 1. Tout d'abord, si $\|f\|_p = 0$, cela implique que $f = 0$ p.p., par conséquent $fg = 0$ p.p., et donc $\int |fg| d\mu = 0$, donc l'inégalité est vraie. C'est aussi le cas si $\|g\|_q = 0$.

2. On va maintenant supposer que $\|f\|_p$ et $\|g\|_q$ sont non nuls. On va distinguer plusieurs cas, dépendant de la valeur de l'indice p .

Si $p = 1$, et donc $q = \infty$, on a de façon évidente $|f(x)g(x)| \leq \|g\|_\infty |f(x)|$ p.p. (par définition de $\|g\|_\infty$), donc en intégrant des deux côtés, et en utilisant le fait que l'intégrale préserve l'inégalité, et sa linéarité, on trouve donc

$$\int |fg| d\mu \leq \int \|g\|_\infty |f(x)| d\mu \leq \|g\|_\infty \int |f(x)| d\mu = \|g\|_\infty \|f\|_1.$$

Bien sûr, le cas $p = \infty, q = 1$ est identique.

3. Considérons à présent les cas $1 < p < \infty$. On prend alors $\alpha = \frac{1}{p} \in]0, 1[$, et on va se servir des propriétés de la fonction $\varphi_\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}$ définie par

$$\varphi_\alpha(t) = t^\alpha - \alpha t.$$

La dérivée première $\varphi'_\alpha(t) = \alpha(t^{\alpha-1} - 1)$ est strictement décroissante sur tout \mathbb{R}_+ , elle ne s'annule qu'en $t = 1$, montrant que φ_α admet un maximum en $t = 1$. On a donc

$$\forall t \in \mathbb{R}_+, \quad \varphi_\alpha(t) \leq \varphi_\alpha(1) = 1 - \alpha.$$

On va « homogénéiser » cette inégalité en prenant $t = \frac{u}{v}$, où $u \geq 0, v > 0$: on trouve alors une inégalité homogène de degré 1 par rapport à u, v :

$$u^\alpha v^{1-\alpha} \leq \alpha u + (1 - \alpha)v.$$

On vérifie que cette inégalité reste vraie si $v = 0$, ou bien si u ou v sont égaux à $+\infty$.

Remarque : cette inégalité est une forme d'inégalité arithmético-géométrique : en effet, le membre de droite est le barycentre (« arithmétique ») de u, v , pondérés par les poids $\alpha, (1 - \alpha)$;

tandis que le membre de gauche est un « barycentre géométrique » de u, v , pondéré par les mêmes poids.

4. On se souvient que $\alpha = \frac{1}{p}$, et $(1 - \alpha) = \frac{1}{q}$. Pour $x \in E$ donné, on applique cette inégalité aux valeurs suivantes :

$$u = \frac{|f(x)|^p}{\|f\|_p^p}, \quad v = \frac{|g(x)|^q}{\|g\|_q^q}.$$

On remarque que ces deux quantités sont homogènes de degré zéro par rapport à f et g . On obtient alors les inégalités ponctuelles suivante, valables pour tout $x \in E$:

$$\frac{|f(x)g(x)|}{\|f\|_p\|g\|_q} \leq \frac{1}{p} \frac{|f(x)|^p}{\|f\|_p^p} + \frac{1}{q} \frac{|g(x)|^q}{\|g\|_q^q}.$$

En intégrant cette inégalité par rapport à μ , on trouve

$$\begin{aligned} \frac{1}{\|f\|_p\|g\|_q} \int |f(x)g(x)| d\mu(x) &\leq \frac{1}{p} \frac{\int |f(x)|^p d\mu(x)}{\|f\|_p^p} + \frac{1}{q} \frac{\int |g(x)|^q d\mu(x)}{\|g\|_q^q} \\ &\leq \frac{1}{p} + \frac{1}{q} = 1, \end{aligned}$$

ce qui est le résultat voulu. □

Exercice 5.3. Montrer que pour avoir égalité dans (5.1), il faut et il suffit que $|f|^p$ et $|g|^q$ soient proportionnelles presque partout : il existe deux réels $a, b \geq 0$, non simultanément nuls, tels que $a|f|^p = b|g|^q$ p.p.

La valeur $p = 2$ est particulière, en cela que l'exposant conjugué $q = p = 2$. Dans ce cas, l'inégalité de Hölder (5.1) se réduit à l'inégalité de *Cauchy-Schwarz* :

$$\int |fg| d\mu \leq \left(\int |f|^2 d\mu \int |g|^2 d\mu \right)^{1/2}. \quad (5.2)$$

On verra ci-dessous que l'espace $L^2(E, \mathcal{A}, \mu)$ peut être muni d'un produit scalaire, l'inégalité ci-dessus pourra également être montrée à partir de cette structure.

5.1.4 Cas d'une mesure finie : emboîtement des espaces L^p

Supposons que $\mu(E) < \infty$. Si on prend g égale à la fonction constante, l'inégalité de Hölder montre alors que pour toute fonction f mesurable,

$$\int |f| d\mu \leq \|f\|_p \|1\|_q = \|f\|_p \mu(E)^{1/q}. \quad (5.3)$$

Donc, pour tout $p \in]1, \infty]$, si la fonction $f \in L^p(E, \mathcal{A}, \mu)$, alors nécessairement f est intégrable : $f \in L^1(E, \mathcal{A}, \mu)$. Ceci montre que les espaces satisfont, dans ce cas :

$$\forall p \in [1, \infty], \quad L^p(E, \mathcal{A}, \mu) \subset L^1(E, \mathcal{A}, \mu).$$

Si on se donne un exposant $r \in]1, \infty[$, une fonction mesurable g , et on applique l'inégalité (5.3) à la fonction $f = |g|^r$, on obtient :

$$\begin{aligned} \int |g|^r d\mu &\leq \mu(E)^{1/q} \left(\int |g|^{pr} d\mu \right)^{1/p} \\ \implies \|g\|_r &\leq \mu(E)^{1/qr} \|g\|_{pr} = \mu(E)^{\frac{1}{r} - \frac{1}{r'}} \|g\|_{r'}, \end{aligned} \quad (5.4)$$

où on a noté $r' = pr > r$.

Ainsi, pour tous indices $1 \leq r \leq r' \leq \infty$, on a montré que les espaces $L^{r'} \subset L^r$: plus l'exposant r est grand, plus l'espace L^r est « petit », au sens où il contient moins de fonctions que l'espace $L^{r-\epsilon}$. Les espaces $(L^r)_{r \in [1, \infty]}$ forment une suite décroissante d'espaces fonctionnels.

Remarque 5.4. Lorsque μ est une mesure de probabilité, on a les inégalités :

$$\|f\|_r \leq \|f\|_{r'}, \quad \text{pour tous exposants } 1 \leq r \leq r' \leq \infty. \quad (5.5)$$

Cette inégalité entre normes L^r est un cas particulier d'une inégalité de convexité, due à Jensen.

Théorème 5.5. (*Inégalité de Jensen*) Soit μ une mesure de probabilité sur (E, \mathcal{A}) , et soit $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$ une fonction convexe. Alors, pour tout $f \in L^1(E, \mathcal{A}, \mu)$, on a :

$$\int \varphi \circ f d\mu \geq \varphi \left(\int f d\mu \right).$$

Démonstration. Vérifions que le membre de gauche est bien défini. φ est convexe, donc continue, donc borélienne entre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. La fonction $\varphi \circ f$ est donc mesurable, positive, donc son intégrale par rapport à μ est bien définie.

Une fonction convexe a pour propriété de pouvoir être définie de façon variationnelle, à partir de fonctions affines (de la forme $at + b$). Définissons l'ensemble des droites affines majorées par φ :

$$\mathcal{E}_\varphi = \{ (a, b) \in \mathbb{R}^2; \forall t \in \mathbb{R}, \varphi(t) \geq at + b \}.$$

La fonction φ peut alors s'exprimer de façon variationnelle, à partir de ces fonctions affines majorées :

$$\forall t \in \mathbb{R}, \varphi(t) = \sup_{(a,b) \in \mathcal{E}_\varphi} (at + b).$$

Géométriquement, le graphe de φ est « enveloppé » par les droites affines maximisant l'expression ci-dessus.

Comme l'intégrale respecte l'ordre entre les fonctions, on peut se servir de cette caractérisation de φ pour calculer le membre de gauche dans le théorème. Pour tout $(a,b) \in \mathcal{E}_\varphi$, la fonction $\varphi \circ f \geq af + b$. Le membre de gauche est mesurable et positif, le membre de droite est intégrable, on en déduit (du fait que l'intégrale préserve les inégalités) :

$$\int \varphi \circ f \, d\mu \geq \int (af + b) \, d\mu.$$

Ensuite, la linéarité de l'intégrale donne

$$\int \varphi \circ f \, d\mu \geq a \int f \, d\mu + b\mu(E) = a \int f \, d\mu + b$$

On peut alors prendre le supremum sur $(a,b) \in \mathcal{E}_\varphi$:

$$\int \varphi \circ f \, d\mu \geq \sup_{(a,b) \in \mathcal{E}_\varphi} \left(a \int f \, d\mu + b \right) = \varphi \left(\int f \, d\mu \right).$$

□

L'inégalité (5.5) montre que pour toute fonction mesurable f , la fonction $r \mapsto \|f\|_r$ est croissante. On peut donc prendre la limite $r \rightarrow \infty$, et on trouve l'inégalité

$$\lim_{r \rightarrow \infty} \|f\|_r \leq \|f\|_\infty.$$

Exercice 5.6. Montrer qu'on a égalité : $\lim_{r \rightarrow \infty} \|f\|_r = \|f\|_\infty$. Montrer que cette égalité reste vraie dans le cas d'une mesure finie.

5.2 L'espace de Banach $L^p(E, \mathcal{A}, \mu)$

Une raison importante de considérer l'espace L^p des classes d'équivalence, plutôt que l'espace \mathcal{L}^p des fonctions, est que $\|f\|_p$ va avoir la structure d'une norme sur L^p , alors qu'elle n'est

qu'une semi-norme sur \mathcal{L}^p . En effet, on a vu que $\|f\|_p = 0$ si et seulement si $f = 0$ p.p., autrement dit si $[f] = [0]$. Ceci est la première condition pour être une norme.

Il faut vérifier les deux autres propriétés d'une norme :

1) homogénéité : pour tout $\alpha \in \mathbb{R}$, $\|\alpha f\|_p = |\alpha| \|f\|_p$. Cette identité découle trivialement de l'homogénéité de l'intégrale des fonctions mesurables positives.

2) inégalité triangulaire : celle-ci est moins triviale, elle résulte de l'inégalité de Minkowski, montrée dans le théorème ci-dessous :

Théorème 5.7. (*Inégalité de Minkowski*) Soit $p \in [1, \infty]$, et soient $f, g \in L^p(E, \mathcal{A}, \mu)$. Alors $f + g \in L^p(E, \mathcal{A}, \mu)$, et

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Démonstration. L'inégalité triangulaire ponctuelle $|f + g| \leq |f| + |g|$ permet de montrer l'inégalité de Minkowski dans les cas $p = 1$ et $p = \infty$.

Supposons maintenant que $p \in]1, \infty[$. La convexité de la fonction $t \mapsto t^p$ sur \mathbb{R}_+ implique que pour tous $u, v \in \mathbb{R}_+$,

$$\left(\frac{u+v}{2}\right)^p \leq \frac{u^p + v^p}{2}.$$

En prenant $u = |f(x)|$, $v = |g(x)|$, il en résulte l'inégalité ponctuelle

$$|f + g|^p \leq 2^{p-1} (|f|^p + |g|^p).$$

Comme les deux fonctions du membre de droite sont intégrables, le membre de gauche est également intégrable, donc $f + g \in L^p$. Cependant, l'inégalité ci-dessus ne permet pas de retrouver l'inégalité de Minkowski par simple intégration.

A partir de l'inégalité triangulaire simple $|f + g| \leq |f| + |g|$, on trouve :

$$\begin{aligned} |f + g|^p &= |f + g|^{p-1} |f + g| \\ &\leq |f| |f + g|^{p-1} + |g| |f + g|^{p-1}. \end{aligned}$$

En intégrant, on obtient

$$\int |f + g|^p d\mu \leq \int |f| |f + g|^{p-1} d\mu + \int |g| |f + g|^{p-1} d\mu.$$

On va appliquer l'inégalité de Hölder aux deux intégrales de droite, pour les paramètres p et

$$q = \frac{p}{p-1} :$$

$$\begin{aligned} \int |f + g|^p d\mu &\leq \|f\|_p \| |f + g|^{(p-1)} \|_q + \|g\|_p \| |f + g|^{(p-1)} \|_q \\ &\leq (\|f\|_p + \|g\|_p) \left(\int |f + g|^{(p-1)q} d\mu \right)^{1/q} \\ &\leq (\|f\|_p + \|g\|_p) \left(\int |f + g|^p d\mu \right)^{\frac{p-1}{p}}. \end{aligned}$$

Dans le cas où $\int |f + g|^p d\mu = 0$, l'inégalité de Minkowski est évidente. Dans le cas contraire, on peut diviser l'inégalité ci-dessus par $\left(\int |f + g|^p d\mu \right)^{\frac{p-1}{p}}$, et on trouve l'inégalité voulue. \square

On a montré que $\|\bullet\|_p$ est une norme sur l'espace $L^p(E, \mathcal{A}, \mu)$, donc que cet espace est un espace vectoriel normé. La propriété suivante est extrêmement importante pour les applications à l'analyse de ces espaces fonctionnels. En effet, la complétude d'un espace vectoriel permet de montrer l'existence d'une solution à un problème (équation différentielle, équation algébrique), à partir de l'existence d'une suite de Cauchy de solutions approchées.

Théorème 5.8. (*Riesz*) Pour tout $p \in [1, \infty]$, l'espace $L^p(E, \mathcal{A}, \mu)$ muni de la norme $\|\bullet\|_p$ est un espace vectoriel normé complet, qu'on appelle aussi un espace de Banach.

Démonstration. 1. La preuve de ce théorème est simple dans le cas $p = \infty$. En effet, donnons-nous une suite $(f_n)_{n \in \mathbb{N}}$ de fonctions dans L^∞ , qui forme une suite de Cauchy :

$$\|f_n - f_m\|_\infty \rightarrow 0 \quad \text{lorsque } n, m \rightarrow \infty.$$

Quitte à modifier les fonctions f_n sur un ensemble μ -négligeable, on peut supposer que la norme sup $\|f_n - f_m\|_{\text{sup}}$ tend vers zéro. En tout point x , la suite $(f_n(x))_{n \in \mathbb{N}}$ est de Cauchy dans \mathbb{R} , et admet donc une limite $f(x)$. La fonction f , limite d'une suite de fonctions mesurables bornées, est mesurable et bornée, donc dans L^∞ . On voit facilement que la suite (f_n) converge uniformément vers f , donc converge dans L^∞ .

2. Passons maintenant au cas $p \in [1, \infty[$. Considérons une suite $(f_k)_{k \in \mathbb{N}}$ de Cauchy dans L^p (c'est-à-dire une suite de Cauchy pour la norme $\|\bullet\|_p$). On peut alors extraire une sous-suite $(k_n)_{n \geq 0} \subset \mathbb{N}$ strictement croissante, telle que la sous-suite correspondante de fonctions $(f_{k_n})_{n \geq 0}$ satisfasse :

$$\forall n \geq 0, \quad \|f_{k_{n+1}} - f_{k_n}\|_p \leq \frac{1}{2^n}. \quad (5.6)$$

Pour simplifier les notations on prendra $g_n := f_{k_n}$. L'idée va être de contrôler la fonction définie par la série

$$G(x) = \sum_{n=0}^{\infty} |g_{n+1}(x) - g_n(x)| = \lim_{N \rightarrow \infty} \sum_{n=0}^N |g_{n+1}(x) - g_n(x)|.$$

Notons que la fonction G est mesurable positive (à valeurs dans $[0, \infty]$). On peut donc calculer l'intégrale $\int G(x)^p d\mu(x)$. Le théorème de convergence monotone nous dit que

$$\begin{aligned} \int G(x)^p d\mu &\stackrel{\text{TCM}}{=} \lim_{N \rightarrow \infty} \int \left(\sum_{n=0}^N |g_{n+1}(x) - g_n(x)| \right)^p d\mu \\ &= \lim_{N \rightarrow \infty} \left\| \sum_{n=0}^N |g_{n+1} - g_n| \right\|_p^p \\ &\stackrel{\text{Minkowski}}{\leq} \lim_{N \rightarrow \infty} \left(\sum_{n=0}^N \|g_{n+1} - g_n\|_p \right)^p \end{aligned}$$

(on applique l'inégalité de Minkowski à une somme de $N + 1$ termes). Les estimées (5.6) montrent que le membre de droite est borné par 2^p , en particulier il est fini. On a donc montré que la fonction G est dans L^p , ce qui implique en particulier que

$$G(x) = \sum_{n=0}^{\infty} |g_{n+1}(x) - g_n(x)| < \infty \quad p.p.$$

Plus précisément, $G(x)$ converge en-dehors d'un ensemble négligeable $N \in \mathcal{A}$. Ceci nous permet de dire que la somme

$$h(x) = g_1(x) + \sum_{n \geq 1} (g_{n+1}(x) - g_n(x))$$

converge absolument pour tout $x \in E \setminus N$; sur N on peut alors définir $h(x) = 0$. On obtient donc une fonction h finie en tout point. De plus, sur $x \in E \setminus N$ la série définissant $h(x)$ est *télescopique*, de sorte que

$$h(x) = \lim_{K \rightarrow \infty} \mathbb{1}_{E \setminus N}(x) \left(g_1(x) + \sum_{n=0}^K (g_{n+1}(x) - g_n(x)) \right) = \lim_{K \rightarrow \infty} \mathbb{1}_{E \setminus N}(x) g_K(x).$$

Toutes les fonctions $\mathbb{1}_{E \setminus N} g_K$ sont mesurables, donc h est mesurable comme limite de fonctions

mesurables. Le lemme de Fatou, appliqué aux fonctions $|\mathbb{1}_{E \setminus N} g_K|^p$, dit que

$$\begin{aligned} \int |h|^p d\mu &= \int \liminf_{K \rightarrow \infty} |\mathbb{1}_{E \setminus N} g_K|^p d\mu = \int \liminf_{K \rightarrow \infty} |g_K|^p d\mu \\ &\stackrel{\text{Fatou}}{\leq} \liminf_{K \rightarrow \infty} \int |g_K|^p d\mu. \end{aligned}$$

Comme la suite (g_n) est de Cauchy, elle est en particulier bornée dans L^p , donc les intégrales $\int |g_K|^p d\mu$ sont toutes bornées par un $C > 0$. Ceci montre que la fonction h est dans L^p .

Il reste à montrer que $g_n \rightarrow h$ dans L^p . Pour tout $n \geq 0$, on a :

$$\int |h - g_n|^p d\mu = \int \lim_{K \rightarrow \infty} |g_K - g_n|^p d\mu \stackrel{\text{Fatou}}{\leq} \liminf_{K \rightarrow \infty} \int |g_K - g_n|^p d\mu = \liminf_{K \rightarrow \infty} \|g_K - g_n\|_p^p.$$

Dès que $K \geq n$,

$$\|g_K - g_n\|_p \stackrel{\text{Mink.}}{\leq} \sum_{k=n}^{K-1} \|g_{k+1} - g_k\|_p \leq \sum_{k=n}^{K-1} \frac{1}{2^k} \leq \frac{2}{2^n},$$

et donc $\int |h - g_n|^p d\mu \leq 2^{(1-n)p}$, autrement dit $\|h - g_n\|_p \leq 2^{1-n}$. Cette estimée montre la convergence $g_n \rightarrow h$ dans L^p . Une fois qu'on a montré la limite pour les fonctions $g_n = f_{k_n}$, il est aisé de montrer que la suite complète $(f_k)_{k \in \mathbb{N}}$ converge vers h dans L^p . \square

Exemple 5.9. Si on prend $E = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$ et μ la mesure de comptage, pour $p \in [1, \infty[$ l'espace $L^p(E, \mathcal{P}(E), \mu)$ est l'espace des suites de réels $a = (a_n)_{n \in \mathbb{N}}$ tels que la série

$$\sum_{n=0}^{\infty} |a_n|^p < \infty.$$

Cet espace vectoriel est muni de la norme

$$\|a\|_p = \left(\sum_{n=0}^{\infty} |a_n|^p \right)^{1/p}.$$

Pour $p = \infty$, l'espace $L^\infty(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$ est simplement l'espace des suites $a = (a_n)_{n \in \mathbb{N}}$ bornées. La norme sur cet espace est simplement la norme sup $\|a\|_\infty = \sup_{n \in \mathbb{N}} |a_n|$.

Notons qu'il n'existe pas d'ensemble négligeable pour la mesure de comptage, donc deux suites a, b égales μ -presque partout seront en fait égales partout. L'espace L^p est donc égal à l'espace \mathcal{L}^p , chaque classe d'équivalence ne contenant qu'un seul élément.

L'espace $L^p(E, \mathcal{P}(E), \mu)$ est habituellement noté $\ell^p(\mathbb{N})$.

Revenons à la preuve du Théorème de Riesz. En partant d'une suite de Cauchy $(f_n)_{n \in \mathbb{N}}$ pour la norme $\|\bullet\|_p$, on construit une fonction h comme limite simple d'une sous-suite f_{k_n} en-dehors d'un ensemble négligeable. En appelant f la fonction limite h , on a donc montré la propriété suivante, qui relie deux types de convergence, la convergence dans L^p et la convergence presque partout :

Proposition 5.10. *Soit $p \in [1, \infty[$ et $(f_k)_{k \in \mathbb{N}}$ une suite qui converge vers f dans $L^p(E, \mathcal{A}, \mu)$. Il existe alors une sous-suite $(f_{k_n})_{n \in \mathbb{N}}$ qui converge μ -p.p. vers f .*

Exercice 5.11. Montrer qu'en général, il est nécessaire d'extraire une sous-suite pour obtenir la convergence p.p. On pourra par exemple construire une suite de fonctions $f_k : [0, 1] \rightarrow \mathbb{R}$ qui convergent vers la fonction nulle dans L^p , mais qui présentent des « pics fins » autour de points x_k , tels que ces pics « balayent l'intervalle indéfiniment » lorsque $k \rightarrow \infty$.

Remarque 5.12. Le cas $p = \infty$ est différent. Si $(f_k)_{k \in \mathbb{N}}$ converge vers f dans $L^\infty(E, \mathcal{A}, \mu)$, alors les fonctions $(f_k)_{k \in \mathbb{N}}$ convergent vers f μ -p.p., sans qu'il soit besoin d'extraire de sous-suite.

La proposition ci-dessus montre que convergence dans L^p implique la convergence p.p. pour une sous-suite. Il est facile de voir que la propriété inverse est fautive : pour un $p \geq 1$ donné, on peut facilement construire une suite $(f_n)_{n \in \mathbb{N}}$ de fonctions $\mathbb{R} \rightarrow \mathbb{R}$ qui convergent simplement vers la fonction nulle, mais telle que $\|f_n\|_p \rightarrow \infty$. Par contre, en imposant une condition de domination, on obtient le résultat suivant :

Proposition 5.13. *Soit $p \in [1, \infty[$. Supposons qu'une suite de fonctions $(f_n)_{n \in \mathbb{N}}$ vérifie :*

i) $f_n \rightarrow f$ μ -p.p.

ii) *Il existe une fonction positive $g \in L^p$, telle que pour tout $n \in \mathbb{N}$, $|f_n| \leq g$ μ -p.p.*

Alors f_n converge vers g dans L^p .

Pour $p = 1$ il s'agit exactement du théorème 3.25 de convergence dominée. Le cas $p > 1$ se montre de façon similaire.

On peut trouver une autre forme de condition de domination, en cas de mesure finie.

Proposition 5.14. *Supposons que $\mu(E) < \infty$. Soit $p \in [1, \infty[$. Supposons que $(f_n)_{n \in \mathbb{N}}$ vérifie :*

i) $f_n \rightarrow f$ μ -p.p.

ii) *Il existe un exposant $r > 1$ et $C > 0$ tel que pour tout $n \in \mathbb{N}$, $\int |f_n|^r d\mu \leq C$.*

Alors f_n converge vers f dans L^p pour tout $1 \leq p < r$.

Démonstration. A faire en exercice. □

On se souvient qu'en cas de mesure finie, les espaces L^p sont emboîtés les uns dans les autres. Remarquons que la suite (f_n) ne converge pas forcément pour la norme $\|\bullet\|_r$: il n'y a convergence que pour des normes $\|\bullet\|_p$ strictement plus faibles que la norme $\|\bullet\|_r$.

Cas $p = 2$: L^2 est un espace de Hilbert

Le cas particulier du théorème de Riesz dans le cas $p = 2$ fournit la structure très riche d'espace de Hilbert.

Théorème 5.15. *L'espace $L^2(E, \mathcal{A}, \mu)$ muni du produit scalaire*

$$\langle f, g \rangle = \int fg \, d\mu$$

forme un espace de Hilbert (réel), autrement dit un espace muni d'un produit scalaire euclidien, et complet vis-à-vis de la norme associée.

Démonstration. L'inégalité de Cauchy-Schwarz (5.2) montre que si $f, g \in L^2$, la fonction fg est intégrable, donc le produit scalaire est bien défini. Ce produit scalaire est bilinéaire par rapport au couple (f, g) , et il est symétrique et défini positif. La norme associée est bien sûr la norme $\|f\|_2$. La complétude de L^2 a été montrée dans le théorème de Riesz. □

Comme pour tout espace de Hilbert, l'espace dual (formé par les formes linéaires continues) s'identifie avec l'espace lui-même, par le biais du produit scalaire.

Théorème 5.16. *(Théorème de représentation de Riesz) Pour toute forme linéaire continue $\Phi : L^2(E, \mathcal{A}, \mu) \rightarrow \mathbb{R}$, il existe un unique élément $g = g_\Phi \in L^2(E, \mathcal{A}, \mu)$ tel que*

$$\forall f \in L^2(E, \mathcal{A}, \mu), \quad \Phi(f) = \langle f, g \rangle.$$

La norme de la forme linéaire Φ ,

$$\|\Phi\| := \sup_{0 \neq f \in L^2} \frac{|\Phi(f)|}{\|f\|_2} = \sup_{\|f\|_2=1} |\Phi(f)|.$$

est égale à $\|g\|_2$.

Démonstration. Pour tout $g \in L^2$, il est clair que l'application $\Phi_g : f \mapsto \langle f, g \rangle$ est une forme linéaire, et l'inégalité de Cauchy-Schwarz $|\langle f, g \rangle| \leq \|g\|_2 \|f\|_2$ montre que $\|\Phi_g\| \leq \|g\|$. En prenant $f = g$, on montre que $\|\Phi_g\| = \|g\|$. Il est évident que l'application $g \mapsto \Phi_g$ est injective : l'égalité $\Phi_g = \Phi_{g'}$ implique que pour tout $f \in L^2$, $\langle f, g - g' \rangle = 0$, en particulier pour $f = g - g'$, donc $g - g' = 0$.

Réciproquement, si on se donne une forme linéaire continue Φ non triviale, le sous-espace vectoriel $\ker(\Phi)$ est un sous-espace fermé de codimension 1, donc un hyperplan. L'orthogonal de ce hyperplan, $\ker(\Phi)^\perp$, est une droite, qui peut être engendrée par un vecteur $\tilde{g} \neq 0$ normalisé ($\|\tilde{g}\|_2 = 1$). Tout élément f peut alors se décomposer de façon unique en :

$$f = \langle f, \tilde{g} \rangle \tilde{g} + h, \quad h \in \ker(\Phi).$$

La linéarité de Φ donne alors $\Phi(f) = \langle f, \tilde{g} \rangle \Phi(\tilde{g}) = \langle f, g \rangle$, en posant $g = \Phi(\tilde{g})\tilde{g}$. \square

5.3 Théorèmes de densité dans L^p

Comme expliqué plus haut, l'intérêt d'un espace complet, réside en la possibilité de construire une solution d'un problème donné, en se servant d'une suite de solutions approchées, formant une suite de Cauchy. Il est souvent pratique de construire ces approximations parmi des espaces de fonctions beaucoup plus « régulières » que les fonctions mesurables, à condition que ces fonctions « régulières » soient denses dans L^p . Nous allons donc montrer la densité de certaines familles de fonctions.

Bien que les applications concernent souvent les fonctions définies sur des espaces euclidiens \mathbb{R}^d , nous énoncerons les théorèmes suivants dans le cas plus général où E est un espace métrique, donc muni d'une distance $d(x, y)$. On pourra toujours prendre comme exemple le cas $E = \mathbb{R}^d$, muni de sa distance euclidienne $d(x, y) = \|x - y\|$.

Notre théorème principal impliquera les fonctions lipschitziennes

Définition 5.17. Sur un espace métrique (E, d) , une fonction $f : E \rightarrow \mathbb{R}$ est dite lipschitzienne s'il existe une constante $K > 0$ telle que

$$\forall x, y \in E, \quad |f(x) - f(y)| \leq K d(x, y).$$

Théorème 5.18. Soit $p \in [1, \infty[$.

1) L'espace des fonctions étagées intégrables est dense dans $L^p(E, \mathcal{A}, \mu)$.

2) Si (E, d) est un espace métrique et si μ est extérieurement régulière sur $(E, \mathcal{B}(E))$, le sous-espace de L^p formé de fonctions lipschitziennes bornées est dense dans L^p .

3) Si (E, d) est un espace métrique localement compact séparable, et μ une mesure de Radon sur E , alors l'espace des fonctions Lipschitziennes à support compact est dense dans L^p .

Les fonctions évoquées dans ces 3 points sont de plus en plus contraintes, donc formant des espaces de « plus en plus petits ». Pour que ces espaces « de plus en plus petits » restent denses dans $L^p(E)$, il devient nécessaire de faire des hypothèses de plus en plus fortes sur l'espace E : en 1) E est quelconque, en 2) c'est un espace métrique, en 3) un espace métrique localement compact et séparable. Remarquons que l'espace euclidien \mathbb{R}^d est localement compact et séparable.

Démonstration : 1) En décomposant f en $f = f^+ - f^-$, il suffit d'étudier le cas de fonctions $f \in L^p$ positives. On sait qu'il existe une suite croissante $(\varphi_n)_{n \in \mathbb{N}}$ de fonctions étagées positives, telles que $\lim_{n \rightarrow \infty} \uparrow \varphi_n = f$. On a donc, comme $0 \leq \varphi_n \leq f$,

$$\int |\varphi_n|^p d\mu \leq \int |f|^p d\mu < \infty,$$

donc $\varphi_n \in L^p$. Puisque $|f - \varphi_n|^p \leq |f|^p$, le TCD nous dit que

$$\lim_n \int |f - \varphi_n|^p d\mu = \int \lim_n |f - \varphi_n|^p d\mu = 0,$$

donc que $(\varphi_n)_n$ converge vers f dans L^p .

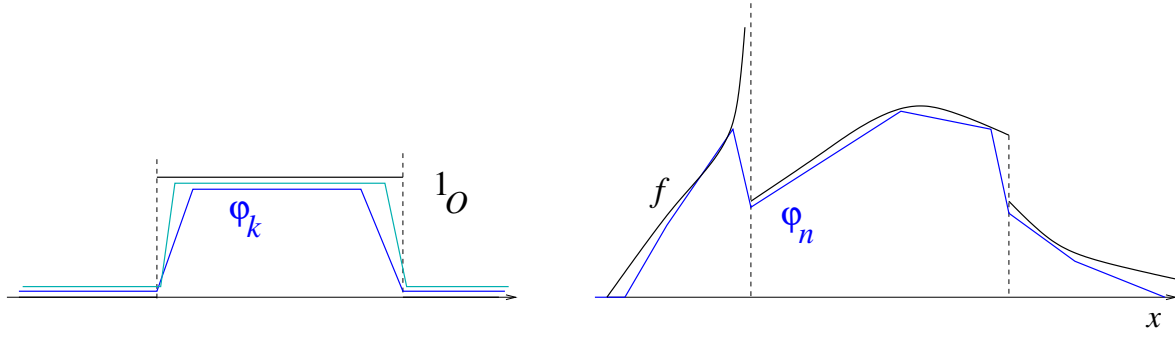
2) On procède par deux approximations successives : comme on vient de montrer que les fonctions étagées sont denses, il suffit de montrer que toute fonction étagée peut être approchée (dans L^p) par des fonctions lipschitziennes bornées. L'écriture canonique d'une fonction étagée est une combinaison linéaire de fonctions $f = \mathbb{1}_A$, avec $A \in \mathcal{A} = \mathcal{B}(E)$; comme f est dans L^p , on a forcément $\mu(A) < \infty$. Prenons un petit $\varepsilon > 0$. Comme μ est extérieurement régulière, on peut trouver un ouvert O contenant A , et tel que $\mu(O \setminus A) \leq (\varepsilon/2)^p$. On aura alors

$$\|\mathbb{1}_O - \mathbb{1}_A\|_p \leq \varepsilon/2,$$

donc on peut remplacer $\mathbb{1}_A$ par la fonction caractéristique de l'ouvert O . Pour $k \geq 1$, on fabrique alors la fonction

$$\varphi_k(x) = \min(1, kd(x, O^c)),$$

qui est supportée sur \overline{O} , et qui « remonte » vers la valeur 1 lorsque x pénètre à l'intérieur de

FIGURE 5.1 – Approximation dans L^p par des fonctions lipschitziennes

O . Cette fonction est lipschitzienne : en utilisant l'inégalité triangulaire, on vérifie que

$$|\varphi_k(x) - \varphi_k(y)| \leq kd(x, y).$$

On a $\varphi_k(x) = 1$ dès que $d(x, O^c) \geq 1/k$, ce qui montre que pour tout $x \in O$, $\lim_{k \rightarrow \infty} \varphi_k(x) = 1$. On a donc $\varphi_k \rightarrow \mathbb{1}_O$. Par le TCD,

$$\int |\varphi_k - \mathbb{1}_O|^p d\mu \xrightarrow[k \rightarrow \infty]{} 0.$$

Donc en choisissant k assez grand, on trouve $\|\varphi_k - \mathbb{1}_O\|_p^p d\mu \leq \varepsilon/2$, donc par inégalité triangulaire :

$$\|\varphi_k - \mathbb{1}_A\|_p \leq \varepsilon.$$

Par linéarité, on pourra approcher une fonction $\sum_{j=1}^J \alpha_j \mathbb{1}_{A_j}$ par une combinaison linéaire de fonctions lipschitziennes, qui est encore lipschitzienne, et bornée.

3) Soit f une fonction dans L^p . La Proposition 4.30 nous montre que la mesure de Radon μ sur un espace métrique séparable localement compact est régulière. En appliquant la partie 2), on peut donc approcher f dans L^p , par des fonctions lipschitziennes bornées. Il reste donc à montrer qu'une telle fonction f_{Lip} bornée et dans L^p , peut elle-même être approchée (toujours dans L^p) par des fonctions lipschitziennes à supports compacts. La propriété $\int |f_{Lip}|^p d\mu < \infty$ et le recouvrement $E = \bigcup_n \mathring{L}_n$ implique, par convergence dominée, que

$$\int |f_{Lip}|^p \mathbb{1}_{\mathring{L}_n} d\mu \xrightarrow[n \rightarrow \infty]{} \int |f_{Lip}|^p d\mu,$$

autrement dit

$$\int |f_{Lip}|^p \mathbb{1}_{(\mathring{L}_n)^c} d\mu = \int \left| f_{Lip} \mathbb{1}_{(\mathring{L}_n)^c} \right|^p d\mu \xrightarrow[n \rightarrow \infty]{} 0,$$

ce qui montre que les fonctions tronquées $f_{Lip}\mathbb{1}_{\dot{L}_n}$ convergent dans L^p vers f_{Lip} . Il reste à modifier ces fonctions pour les rendre lipschitziennes. Si on remplace la fonction caractéristique $\mathbb{1}_{\dot{L}_n}$ par une fonction « lissée »

$$\varphi_{n,k}(x) = \min \left(1, kd(x, (\dot{L}_n)^c) \right),$$

on a $\varphi_{n,k} \leq \mathbb{1}_{\dot{L}_n}$, en particulier $\varphi_{n,k} \in L^p$. De plus, par convergence dominée on voit que pour tout indice n , les fonctions $f_{Lip}\varphi_{n,k}$ convergent vers $f_{Lip}\mathbb{1}_{\dot{L}_n}$ dans L^p , lorsque $k \rightarrow \infty$. En se servant du découpage :

$$\|f_{Lip} - f_{Lip}\varphi_{n,k}\|_p \leq \|f_{Lip} - f_{Lip}\mathbb{1}_{\dot{L}_n}\|_p + \|f_{Lip}\mathbb{1}_{\dot{L}_n} - f_{Lip}\varphi_{n,k}\|_p,$$

on fixe d'abord n assez grand pour que le premier terme soit $\leq \varepsilon$, puis on choisit k assez grand pour que le second terme soit $\leq \varepsilon$: on a alors approché f_{Lip} par la fonction lipschitzienne à support compact $f_{Lip}\varphi_{n,k}$, à 2ε près dans L^p . \square

L'approximation par des fonctions Lipschitziennes nous permet de déduire d'autres résultats de compacité dans le cas de l'espace euclidien $E = \mathbb{R}^d$.

Corollaire 5.19. *Pour $p \in [1, \infty[$:*

i) *Pour toute mesure de Radon μ sur \mathbb{R}^d , l'espace $C_c(\mathbb{R}^d)$ des fonctions continues à support compact est dense dans $L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$. C'est le cas en particulier pour la mesure de Lebesgue λ .*

ii) *L'espace des fonctions en escalier à support compact est dense dans $L^p(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$.*

Démonstration. Le i) est évident, les fonctions lipschitziennes étant continues. Pour le ii), il faut vérifier que toute fonction $f \in C_c(\mathbb{R}^d)$ peut être approchée dans L^p par des fonctions en escalier (à support compact). En prenant les approximations

$$f_n(x) = \sum_{k \in \mathbb{Z}} f\left(\frac{k}{n}\right) \mathbb{1}_{[\frac{k}{n}, \frac{k+1}{n}[},$$

on vérifie aisément que $f_n \rightarrow f$ uniformément, d'où on déduit que $f_n \rightarrow f$ dans $L^p(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. \square

Remarque 5.20. Ces résultats d'approximation sont faux dans L^∞ , la convergence dans L^∞ s'apparentant plus à une convergence uniforme.

Une application de ces résultats d'approximation concerne la transformée de Fourier d'une fonction intégrable.

Corollaire 5.21. *Si $f \in L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, alors sa transformée de Fourier $\hat{f}(\xi)$ converge vers 0 lorsque $|\xi| \rightarrow \infty$.*

Démonstration. On vient de montrer que f est approchée dans L^1 par des fonctions en escalier f_n . On a déjà la borne uniforme :

$$\sup_{\xi \in \mathbb{R}} |\hat{f}(\xi) - \hat{f}_n(\xi)| = \sup_{\xi \in \mathbb{R}} \left| \int (f(x) - f_n(x)) e^{-ix\xi} dx \right| \leq \|f - f_n\|_1,$$

qui tend vers zéro lorsque $n \rightarrow \infty$. D'une autre côté, pour chaque fonction en escalier $\varphi = \sum_{j=1}^p \lambda_j \mathbb{1}_{[x_j, x_{j+1}[}$, on a la formule explicite

$$\hat{\varphi}(\xi) = i \sum_{j=1}^p \lambda_j \left(\frac{e^{-i\xi x_{j+1}} - e^{-i\xi x_j}}{\xi} \right),$$

qui converge vers zéro lorsque $|\xi| \rightarrow \infty$. □

Remarque 5.22. L'hypothèse $f \in L^1$ est la plus faible possible pour construire la transformée de Fourier. Dans le cas où f est plus régulière (par exemple si f est de régularité C^k et à support compact), on peut obtenir plus de précision sur la façon donc $\hat{f}(\xi)$ décroît lorsque $|\xi| \rightarrow \infty$, en utilisant des intégrations par parties successives. Pour une fonction f qui est « juste L^1 », la preuve ci-dessus ne fournit pas de vitesse de décroissance : bien qu'on ait le taux de décroissance de $\hat{f}_n(\xi)$ pour chaque fonction en escalier f_n , on ne contrôle pas cette décroissance uniformément le long de la suite $(\hat{f}_n)_{n \in \mathbb{N}}$.

5.4 Décomposition de mesures : théorème de Radon-Nikodym

Jusqu'à présent on s'est intéressé aux fonctions intégrables, ou aux fonctions L^p , par rapport à une mesure μ fixée. Dans cette section, on va essayer de « comparer » deux mesures μ, ν définies sur la même tribu. Dans cette section, il sera très important de conserver les notations relatives aux mesures : ainsi, on distinguera la propriété μ -p.p. de la propriété ν -p.p., les espaces $L^p(\mu)$ et $L^p(\nu)$, etc.

Dans la Définition 3.13, on avait construit une mesure modulée $\nu = f \cdot \mu$, à partir d'une mesure μ sur (E, \mathcal{A}) , et d'une fonction mesurable positive f . On avait remarqué (Remarque 3.14)

qu'une propriété de la mesure de ν , est de s'annuler sur tous les ensembles μ -négligeables. On peut alors se poser la question suivante :

Etant donné deux mesures μ et ν sur un espace mesurable, comment peut-on décider si ν est une « modulation » de μ , au sens ci-dessus ?

Comme on va le voir ci-dessous, modulo des conditions de σ -finitude, on va obtenir une « réciproque » à la Remarque 3.14. Pour cela, introduisons les définitions suivantes, permettant de comparer deux mesures.

Définition 5.23. Soient μ et ν deux mesures sur un espace mesurable (E, \mathcal{A}) . On dit que :

i) ν est *absolument continue par rapport à μ* (notation $\nu \ll \mu$) si :

$$\forall A \in \mathcal{A}, \quad \mu(A) = 0 \implies \nu(A) = 0.$$

ii) ν est *étrangère à μ* ($\nu \perp \mu$) s'il existe $N \in \mathcal{A}$ tel que $\mu(N) = 0$ et $\nu(N^c) = 0$.

Le résultat de la Remarque 3.14 est donc que pour toute fonction mesurable f , la mesure modulée $f \cdot \mu$ est absolument continue par rapport à μ . On remarque que les deux propriétés ci-dessus sont exclusives : la seule mesure ν qui soit à la fois absolument continue par rapport à μ et étrangère à μ , est la mesure nulle $\nu = 0$.

On remarque aussi qu'une mesure ν n'est en général ni absolument continue par rapport à μ , ni étrangère à μ . Par contre, le théorème ci-dessus montre que toute mesure ν peut se scinder en ces deux types de mesure.

Théorème 5.24. (*Radon-Nikodym*) Soient μ et ν deux mesures σ -finies sur (E, \mathcal{A}) . Il existe alors un unique couple (ν_a, ν_s) de mesures σ -finies sur (E, \mathcal{A}) , telles que :

$$1) \quad \nu = \nu_a + \nu_s ;$$

$$2) \quad \nu_a \ll \mu \text{ et } \nu_s \perp \mu.$$

De plus, il existe une fonction mesurable $g : E \rightarrow \mathbb{R}$ telle que $\nu_a = g \cdot \mu$. La fonction g est unique, à un ensemble de μ -mesure nulle près (de façon équivalente, la fonction g définit une unique classe d'équivalence $[g]$).

Ainsi, si μ est σ -finie, toute mesure ν σ -finie et absolument continue par rapport à μ pourra s'écrire $\nu = g \cdot \mu$: on a donc une réciproque (partielle) à la Remarque 3.14.

Démonstration. 1. On commence par supposer que μ et ν sont des mesures finies. Supposons tout d'abord que nos deux mesures sont telles que $\mu \leq \nu$ (comme fonctions $\mathcal{A} \rightarrow \overline{\mathbb{R}}_+$). On en déduit facilement que pour toute fonction g \mathcal{A} -mesurable positive, $\int g d\nu \leq \int g d\mu$. (Cette propriété a comme conséquence directe que $\nu \ll \mu$). Montrons alors que l'application

$$\begin{aligned} \Phi : L^2(E, \mathcal{A}, \mu) &\rightarrow \mathbb{R} \\ f &\mapsto \Phi(f) = \int f d\nu, \end{aligned} \quad \text{est bien définie.}$$

ATTENTION : on intègre par rapport à la mesure ν , une fonction $f \in L^2(\mu)$.

L'hypothèse $\nu \leq \mu$ montre que

$$\int |f| d\nu \leq \int |f| d\mu,$$

et la finitude de μ , qui implique que $\mathcal{L}^2(\mu) \subset \mathcal{L}^1(\mu)$, nous montre que le membre de droite est fini si $f \in \mathcal{L}^2(\mu)$, donc c'est aussi le cas du membre de gauche. De plus, si deux fonctions mesurables f, \tilde{f} satisfont $f = \tilde{f}$ μ -p.p., alors la relation $\nu \leq \mu$ implique qu'on a aussi $f = \tilde{f}$ ν -p.p., et donc que $\int f d\nu = \int \tilde{f} d\nu$. Ceci montre que la valeur $\Phi(f)$ ne dépend pas du choix de représentant dans la classe $[f] \in L^2(\mu)$, donc que Φ est donc bien définie sur $L^2(\mu)$.

L'inégalité de Cauchy-Schwarz, couplée à la relation d'ordre, nous montre que :

$$|\Phi(f)| \leq \left(\int |f|^2 d\nu \right)^{1/2} \nu(E)^{1/2} \leq \left(\int |f|^2 d\mu \right)^{1/2} \nu(E)^{1/2} = \|f\|_{L^2(\mu)} \nu(E)^{1/2}.$$

Ceci montre que Φ définit une forme linéaire continue sur l'espace de Hilbert $L^2(\mu)$. Par le théorème 5.16 de représentation de Riesz, il existe alors une fonction $h \in L^2(E, \mathcal{A}, \mu)$ telle que

$$\forall f \in L^2(\mu), \quad \Phi(f) = \langle f, h \rangle_{L^2(\mu)} = \int f h d\mu.$$

En particulier, en prenant pour f la fonction caractéristique d'un ensemble mesurable $f = \mathbb{1}_A$, on trouve

$$\nu(A) = \int h d\mu,$$

ce qui montre que $\nu = h \cdot \mu$ est une modulation de μ .

Montrons que l'hypothèse $\nu \leq \mu$ implique que la fonction h est bornée par $0 \leq h \leq 1$ μ -p.p.. En effet, pour tout $\varepsilon > 0$, on a les inégalités suivantes sur l'ensemble des points tels que

$h(x) \geq 1 + \varepsilon :$

$$\begin{aligned} \mu(\{x : h(x) \geq 1 + \varepsilon\}) &\geq \nu(\{x : h(x) \geq 1 + \varepsilon\}) \\ &\geq \int_{\{x : h(x) \geq 1 + \varepsilon\}} h \, d\mu \\ &\geq (1 + \varepsilon)\mu(\{x : h(x) \geq 1 + \varepsilon\}), \end{aligned}$$

d'où on tire que $\mu(\{x : h(x) \geq 1 + \varepsilon\}) = 0$, et en prenant une suite décroissante de $\varepsilon > 0$, on trouve finalement que $\mu(\{x : h(x) > 1\}) = 0$. Par un argument identique, on vérifie que $\mu(\{x : h(x) < 0\}) = 0$. Par conséquent, on peut choisir un représentant \tilde{h} dans la classe $[h]$, tel que $0 \leq \tilde{h} \leq 1$ partout, et tel que $\nu = \tilde{h} \cdot \mu$.

2. On considère à présent le cas d'une mesure ν finie quelconque, et on applique la première partie de la preuve aux mesures $\tilde{\nu} = \nu$ et $\tilde{\mu} = \nu + \mu$, qui vérifient évidemment $\tilde{\nu} \leq \tilde{\mu}$. D'après la preuve ci-dessus, il existe une fonction mesurable $0 \leq h \leq 1$ telle que $\tilde{\nu} = h \cdot \tilde{\mu} = h \cdot (\mu + \nu)$. Ainsi, pour toute fonction $f \in L^2(\mu + \nu)$, on a :

$$\int f \, d\nu = \int f h \, d(\mu + \nu) = \int f h \, d\mu + \int f h \, d\nu,$$

ce qui peut se réécrire

$$\int f(1 - h) \, d\nu = \int f h \, d\mu. \quad (5.7)$$

Les fonctions dans ces intégrales sont toutes positives ou nulles. Pour toute fonction F mesurable positive, le théorème de convergence monotone nous permet d'approcher les intégrales $\int F(1 - h) \, d\nu$ et $\int F h \, d\mu$ en utilisant une suite croissante de fonctions étagées positives $f_n \uparrow F$. Chacune des fonctions étagées f_n satisfait l'égalité ci-dessus, donc en passant à la limite, l'égalité ci-dessus est vraie aussi pour la fonction F . L'égalité ci-dessus donne un rôle particulier aux points x tels que $h(x) = 1$. Appelons-donc l'ensemble $N := \{x \in E : h(x) = 1\}$, qui est \mathcal{A} -mesurable car h l'est. En considérant la fonction $f = \mathbb{1}_N$, on voit que le membre de gauche dans (5.7) s'annule, ce qui nous donne l'équation

$$0 = \int \mathbb{1}_N h \, d\mu = \int \mathbb{1}_N \, d\mu = \mu(N),$$

autrement dit l'ensemble N est μ -négligeable. Définissons alors les mesures

$$\nu_s := \mathbb{1}_N \cdot \nu, \quad \nu_a := \mathbb{1}_{N^c} \cdot \nu.$$

La première vérifie trivialement $\nu_s(N^c) = 0$. La propriété $0 = \mu(N)$ nous montre alors que $\nu_s \perp \mu$.

Pour analyser la seconde composante ν_a , prenons maintenant la fonction $f = \mathbb{1}_{N^c} \frac{F}{1-h}$ dans l'équation (5.7), pour F une fonction mesurable positive quelconque (remarquons que cette fonction est bien définie, puisque $1-h$ ne s'annule pas sur l'ensemble N^c). On trouve alors :

$$\int F \mathbb{1}_{N^c} d\nu = \int F \mathbb{1}_{N^c} \frac{h}{1-h} d\mu = \int F g d\mu,$$

après avoir défini la fonction mesurable positive $g = \mathbb{1}_{N^c} \frac{h}{1-h}$. Cette équation signifie que $\nu_a = g \cdot \mu$, donc que ν_a est absolument continue par rapport à μ .

3. Montrons l'unicité de la décomposition obtenue. Supposons que ν se décompose également en $\nu = \nu'_a + \nu'_s$, où ν'_a et ν'_s satisfont les mêmes propriétés 1) et 2) par rapport à μ . On a donc, pour tout ensemble mesurable A :

$$\nu_a(A) - \nu'_a(A) = \nu_s(A) - \nu'_s(A).$$

Comme les mesures singulières ν_s, ν'_s sont supportées respectivement par des ensembles μ -négligeables N, N' , on a pour tout A :

$$\begin{aligned} \nu_s(A) - \nu'_s(A) &= \nu_s(A \cap (N \cup N')) - \nu'_s(A \cap (N \cup N')) \\ &= \nu_a(A \cap (N \cup N')) - \nu'_a(A \cap (N \cup N')) \\ &= 0, \end{aligned}$$

où pour la dernière ligne on se sert du fait que $\mu(N \cup N') = 0$ et l'hypothèse d'absolue continuité de ν_a, ν'_a . Cette égalité montre que $\nu_s = \nu'_s$, et donc également $\nu_a = \nu'_a$.

Vérifions également que la fonction g est essentiellement unique. Soit \tilde{g} une seconde fonction telle que $\nu_a = \tilde{g} \cdot \mu$. On veut alors mesurer l'ensemble mesurable $N_{>} := \{x : \tilde{g}(x) > g(x)\}$:

$$\int_{N_{>}} \tilde{g} d\mu = \nu_a(N_{>}) = \int_{N_{>}} g d\mu,$$

d'où on tire directement

$$\int_{N_{>}} (\tilde{g} - g) d\mu = 0.$$

La fonction $(\tilde{g} - g)$ étant positive sur $N_{>}$, l'égalité ci-dessus implique que $\tilde{g} - g = 0$ μ -p.p. sur l'ensemble $N_{>}$, ce qui impose que l'ensemble $N_{>}$ est lui-même négligeable. En échangeant

g, \tilde{g} , on montre donc que $\tilde{g} = g$ μ -p.p. Ceci achève la preuve du théorème dans le cas de mesures μ, ν finies.

4. Considérons à présent le cas de mesures μ, ν σ -finies. Cela signifie qu'on peut recouvrir E par une suite croissante d'ensembles mesurables $(E_n)_{n \in \mathbb{N}}$ tels que $\mu(E_n) < \infty$, respectivement une suite croissante $(\tilde{E}_n)_{n \in \mathbb{N}}$ telle que $\nu(\tilde{E}_n) < \infty$. La suite $(E'_n := E_n \cap \tilde{E}_n)_{n \in \mathbb{N}}$ est alors croissante, elle recouvre E , et on a à la fois $\mu(E'_n) < \infty$ et $\nu(E'_n) < \infty$. On peut considérer $F_0 = E'_0$, $F_n = E'_n \setminus E'_{n-1}$ pour $n \geq 1$: les $(F_n)_{n \in \mathbb{N}}$ forment alors une partition mesurable de E , telle que les mesures $\mu_n = \mu|_{F_n}$ et $\nu_n = \nu|_{F_n}$ sont finies. Pour chacune de ces mesures finies, la preuve ci-dessus permet de décomposer ν_n en :

$$\nu_n = \nu_a^n + \nu_s^n,$$

où ν_s^n est étrangère à μ_n , tandis que $\nu_a^n = g_n \cdot \mu_n$ pour une certaine fonction mesurable positive g_n . On peut choisir la fonction g_n telle qu'elle s'annule sur le complémentaire de F_n (puisque ν_n et μ_n s'annulent sur F_n^c). La propriété de partition $E = \bigsqcup_n F_n$ permet de vérifier que les mesures

$$\nu_a := \sum_{n \in \mathbb{N}} \nu_a^n, \quad \nu_s := \sum_{n \in \mathbb{N}} \nu_s^n$$

sont σ -finies, qu'elles sont respectivement absolument continue par rapport à μ et étrangère à μ , et que $\nu = \nu_a + \nu_s$. Enfin, la fonction $g = \sum_n g_n$ est bien définie, mesurable, et elle vérifie bien $\nu_a = g \cdot \mu$.

Remarquons que la partition $E = \bigsqcup_n F_n$ peut être choisie de différentes manières ; on a alors $\nu_a^n = \nu_a|_{F_n}$, $\nu_s^n = \nu_s|_{F_n}$ et $g_n = g|_{F_n}$. L'unicité des restrictions à chacun des F_n impose alors l'unicité de ν_a, ν_s et g . \square

6 Mesures produits

Nous avons construit séparément la mesure de Lebesgue sur \mathbb{R} , puis sur \mathbb{R}^d . Dans ce chapitre nous allons montrer comment, à partir de la mesure λ sur \mathbb{R} , on peut construire une mesure sur l'espace $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, appelée la mesure produit $\lambda \otimes \lambda$; nous montrerons alors que ce produit de deux mesures de Lebesgue est égal à la mesure λ que nous avons construite sur \mathbb{R}^2 . Cette construction se généralise à toute paire d'espaces mesurés.

Ces mesures produits permettent de calculer l'intégrale d'une fonction par étapes, en calculant d'abord l'intégrale de la fonction $f(x, y)$ par rapport à la première variable / mesure, puis en intégrant le résultat par rapport à la seconde mesure : ce processus est décrit par le théorème de Fubini. Ce théorème est très important en pratique, il permet de choisir l'ordre des intégrations successives.

6.1 Espace produit

Avant de définir la mesure produit, il faut déjà définir la tribu sur laquelle elle sera définie. Pour deux espaces mesurables (E_1, \mathcal{A}_1) et (E_2, \mathcal{A}_2) , la *tribu produit* $\mathcal{A}_1 \otimes \mathcal{A}_2$ sur l'espace $E_1 \times E_2$ a déjà été décrite dans la Définition 2.6 :

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \stackrel{\text{def}}{=} \sigma(\{A_1 \times A_2; A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}).$$

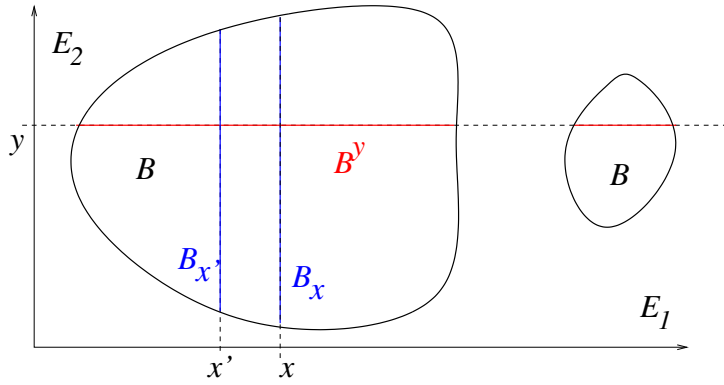
Parmi les éléments de cette tribu, les éléments produits $A_1 \times A_2$ sont appelés des *pavés mesurables*, par analogie avec les pavés de \mathbb{R}^2 , qui sont des produits d'intervalles.

Affirmation 6.1. La tribu produit est également la plus petite tribu qui rende mesurables les deux projections canoniques $\pi_1 : E_1 \times E_2 \rightarrow E_1$, $\pi_2 : E_1 \times E_2 \rightarrow E_2$.

De même, pour deux fonction mesurables $f_1 : (F, \mathcal{B}) \rightarrow (E_1, \mathcal{A}_1)$ et $f_2 : (F, \mathcal{B}) \rightarrow (E_2, \mathcal{A}_2)$, nous avons défini dans la Proposition 2.28 l'application produit $f : (F, \mathcal{B}) \rightarrow (E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ par $f(x) = (f_1(x), f_2(x))$; la proposition a montré que l'application produit f est mesurable ssi ses deux composantes f_1 et f_2 le sont.

On peut bien sûr itérer l'opération produit : pour n espaces mesurables $(E_1, \mathcal{A}_1), \dots, (E_n, \mathcal{A}_n)$, on définit la tribu produit sur $E_1 \times E_2 \times \dots \times E_n$ par :

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \dots \otimes \mathcal{A}_n \stackrel{\text{def}}{=} \sigma(\{A_1 \times \dots \times A_n; A_i \in \mathcal{A}_i\}).$$

FIGURE 6.1 – Les tranches B_x, B^y d'une partie $B \subset E_1 \times E_2$

On vérifie facilement la propriété d'associativité du produit :

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \mathcal{A}_3 = (\mathcal{A}_1 \otimes \mathcal{A}_2) \otimes \mathcal{A}_3 = \mathcal{A}_1 \otimes (\mathcal{A}_2 \otimes \mathcal{A}_3).$$

Notations. Pour toute partie $B \subset E_1 \times E_2$, on note, pour $x \in E_1$, la « tranche horizontale »

$$B_x := \{y \in E_2 : (x, y) \in B\}$$

et pour $y \in E_2$, la « tranche verticale »

$$B^y := \{x \in E_1 : (x, y) \in B\}.$$

Pour f une fonction $E_1 \times E_2 \rightarrow F$, on notera pour $x \in E_1$, la fonction $f_x : E_2 \rightarrow F$ définie par $f_x(y) = f(x, y)$. De même, $f^y : E_1 \rightarrow F$. On notera parfois ces fonctions $f_x = f(x, \bullet)$, respectivement $f^y = f(\bullet, y)$.

Théorème 6.2. *i) Supposons que $B \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Alors, pour tout $x \in E_1$, la tranche $B_x \in \mathcal{A}_2$. De même, pour tout $y \in E_2$, la tranche $B^y \in \mathcal{A}_1$.*

ii) Soit $f : (E_1 \times E_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \rightarrow (F, \mathcal{B})$ une application mesurable. Alors pour tout $x \in E_1$, l'application $f_x : (E_2, \mathcal{A}_2) \rightarrow (F, \mathcal{B})$ est mesurable. Idem pour $f^y : (E_1, \mathcal{A}_1) \rightarrow (F, \mathcal{B})$ pour tout $y \in E_2$.

Ce théorème est la première pierre du théorème de Fubini : la mesurabilité des fonctions f_x va nous permettre d'intégrer ces fonctions le long de E_2 , avant d'intégrer le résultat le long de E_1 ; ou vice-versa.

Démonstration. *i)* Comme on l'a déjà plusieurs fois, la stratégie consiste à définir astucieu-

sement une classe, et vérifier qu'elle est une tribu. Fixons $x \in E_1$, et définissons la classe :

$$\mathcal{C}_x := \{B \in \mathcal{A}_1 \otimes \mathcal{A}_2 : C_x \in \mathcal{A}_2\} \subset E_1 \times E_2.$$

Il est facile de vérifier que cette classe est une tribu : l'ensemble vide est évidemment dans \mathcal{C}_x ; si $B \in \mathcal{C}_x$, alors $(B^c)_x = (B_x)^c \in \mathcal{A}_2$, donc $B^c \in \mathcal{C}_x$; enfin, si $(B_n)_{n \in \mathbb{N}}$, alors $(\bigcup_n B_n)_x = \bigcup_n (B_n)_x \in \mathcal{A}_2$, donc $\bigcup_n B_n \in \mathcal{C}_x$. Pour tout pavé mesurable $B = A_1 \times A_2$, on a $B_x = A_2$ si $x \in A_1$, et $B_x = \emptyset$ sinon. Dans les deux cas, $B_x \in \mathcal{A}_2$, ce qui montre que le pavé $B = A_1 \times A_2 \in \mathcal{C}_x$. La tribu \mathcal{C}_x contient donc tous les pavés, donc elle contient la tribu produit.

ii) Pour tout $B \in \mathcal{B}$, son image réciproque par la fonction f_x vaut :

$$f_x^{-1}(B) = \{y \in E_2 : f(x, y) \in B\} = (f^{-1}(B))_x,$$

qui est mesurable d'après la mesurabilité de f et le résultat i). □

Cas des tribus de Borel sur deux espaces métriques.

Proposition 6.3. *Si E_1 et E_2 sont deux espaces métriques séparables, on a :*

$$\mathcal{B}(E_1 \times E_2) = \mathcal{B}(E_1) \otimes \mathcal{B}(E_2).$$

Démonstration. Pour tout ouvert $U_1 \subset E_1$, son image réciproque par π_1 vaut $U_1 \times E_2$, qui est ouvert dans $E_1 \times E_2$: ainsi, l'application $\pi_1 : E_1 \times E_2 \rightarrow E_1$ est continue, et donc mesurable entre $\mathcal{B}(E_1 \times E_2)$ et $\mathcal{B}(E_1)$. C'est aussi le cas de $\pi_2 : E_1 \times E_2 \rightarrow E_2$. D'après l'Affirmation 6.1, la tribu produit $\mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$ est la plus petite rendant π_1 et π_2 mesurables, donc elle est forcément incluse dans $\mathcal{B}(E_1 \times E_2)$.

Montrons l'inclusion inverse dans le cas d'espaces E_i mesurables. Soit $(x_k)_{k \in \mathbb{N}}$ une suite de points dense dans E_1 . Considérons l'ensemble des boules $\mathcal{U}_1 := \{B(x_k, r), k \in \mathbb{N}, r \in \mathbb{Q}\}$, qui est évidemment dénombrable. On vérifie facilement que tout ouvert de E_1 peut s'obtenir comme union d'éléments de \mathcal{U}_1 . On considère de même un ensemble dénombrable \mathcal{U}_2 de boules de E_2 , tel que tout ouvert de E_2 est l'union de boules de \mathcal{U}_2 . L'ensemble produit $E_1 \times E_2$ est également séparable. Tout ouvert de $E_1 \times E_2$ peut être recouvert par une union dénombrable de boules produits de la forme $B(x_k, r) \times B(y_j, r')$, ce qui montre que cet ouvert appartient à la tribu $\mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$. Comme les ouverts engendrent la tribu borélienne, on déduit que $\mathcal{B}(E_1 \times E_2) \subset \mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$. □

6.2 Construction de la mesure produit

Comme on va devoir considérer des suites d'ensembles et de mesures, dans cette section nous appellerons nos deux espaces mesurés (E, \mathcal{A}, μ) et (F, \mathcal{B}, ν) . Sous certaines conditions, nous allons construire une mesure sur l'espace produit $(E \times F, \mathcal{A} \otimes \mathcal{B})$.

Théorème 6.4. *Soient μ et ν deux mesures σ -finies respectivement sur (E, \mathcal{A}) et (F, \mathcal{B}) .*

i) Il existe une unique mesure m sur $(E \times F, \mathcal{A} \otimes \mathcal{B})$ telle que :

$$\forall A \in \mathcal{A}, \forall B \in \mathcal{B}, \quad m(A \times B) = \mu(A)\nu(B)$$

(avec comme d'habitude $0 \cdot \infty = 0$). Cette mesure est appelée la mesure produit, elle est σ -finie. On la note $m = \mu \otimes \nu$.

ii) Pour tout $C \in \mathcal{A} \otimes \mathcal{B}$,

$$\mu \otimes \nu(C) = \int_E \nu(C_x) \mu(dx) = \int_F \mu(C^y) \nu(dy).$$

Le second résultat est un cas particulier du théorème de Fubini (pour la fonction caractéristique $f = \mathbb{1}_C$).

Démonstration. Commençons par montrer l'*unicité* de la mesure produit. Pour montrer l'unicité on va se servir du Corollaire 2.16 au Lemme de classe monotone, plus précisément de la 2e condition. On choisit comme classe \mathcal{C} la famille des pavés mesurables

$$\mathcal{C} = \{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\} = \mathcal{A} \times \mathcal{B}.$$

Cette classe est stable par intersection :

$$(A \times B) \cap (A' \times B') = (A \cap A') \times (B \cap B').$$

Comme μ et ν sont σ -finies, il existe des suites croissantes de sous-ensembles $(A_n \in \mathcal{A})_{n \in \mathbb{N}}$ recouvrant E , respectivement $(B_m \in \mathcal{B})_{m \in \mathbb{N}}$ recouvrant F , telles que $\mu(A_n) < \infty$, resp. $\nu(B_m) < \infty$. Les pavés mesurables $(A_n \times B_m)$ recouvrent alors $E \times F$. Il est possible d'extraire une sous-famille de pavés $(A_{n_k} \times B_{m_k})_{k \in \mathbb{N}}$ qui soit croissante, et qui recouvre $E \times F$. Si m, m' sont deux mesures satisfaisant les propriétés de *i*), elles vérifient alors

$$m(A_{n_k} \times B_{m_k}) = m'(A_{n_k} \times B_{m_k}) = \mu(A_{n_k})\nu(B_{m_k}) < \infty,$$

et ces deux mesures sont égales sur la classe \mathcal{C} . Le Corollaire nous dit alors que $m = m'$.

Montrons à présent l'*existence* de la mesure m , en se servant des expressions du *ii*). Pour $C \in \mathcal{A} \otimes \mathcal{B}$, montrons que l'expression

$$m(C) = \int_E \nu(C_x) \mu(dx) \quad (6.1)$$

a un sens. On sait déjà que pour tout $x \in E$, la tranche $C_x \in \mathcal{B}$, donc la mesure ν est bien définie sur cette tranche. Il faut néanmoins vérifier que la fonction $x \mapsto \nu(C_x)$ est mesurable par rapport à \mathcal{A} .

On va d'abord supposer que la mesure ν est finie. Comme on l'a déjà fait plusieurs fois, on fabrique une classe adaptée :

$$\mathcal{G} := \{C \in \mathcal{A} \otimes \mathcal{B} : x \mapsto \nu(C_x) \text{ est mesurable}\} \subset \mathcal{A} \otimes \mathcal{B}.$$

Comme dans la preuve précédente, on remarque que tout pavé $C = A \times B$ est dans cette classe, puisque la fonction $\nu(C_x) = \nu(B)\mathbb{1}_A(x)$ est intégrable. On va ensuite montrer que \mathcal{G} est une classe monotone (v. Définition 2.13). L'ensemble totale $E \times F$ est dans \mathcal{G} , puisque $(E \times F)_x = \nu(F)$ pour tout $x \in E$. Si $C \supset C'$ sont dans \mathcal{G} , alors $(C \setminus C')_x = C_x \setminus C'_x$, et comme ν est finie et $C_x \supset C'_x$, on a

$$\nu(C_x \setminus C'_x) = \nu(C_x) - \nu(C'_x);$$

cette différence de deux fonctions mesurables est bien mesurable (v. Corollaire 2.29). Enfin, si $(C_n)_{n \in \mathbb{N}}$ forme une suite croissante d'éléments de \mathcal{G} , pour tout x la suite $(C_n)_x$ est également croissante, donc

$$\forall x \in E, \quad \nu \left(\left(\bigcup_n C_n \right)_x \right) = \lim_{n \rightarrow \infty} \uparrow \nu((C_n)_x)$$

et cette limite de fonctions mesurables est mesurable. On a donc montré que la classe \mathcal{G} est une classe monotone, et qu'elle contient la classe monotone $\mathcal{M}(\mathcal{C})$ engendrée par les pavés. Comme de plus la famille \mathcal{C} des pavés est invariante par intersections finies, le Lemme de classe monotone nous informe que $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C}) = \mathcal{A} \otimes \mathcal{B}$, donc que $\mathcal{G} = \mathcal{A} \otimes \mathcal{B}$.

On a donc montré que pour tout $C \in \mathcal{A} \otimes \mathcal{B}$, la fonction $x \mapsto \nu(C_x)$ est mesurable sur (E, \mathcal{A}) . Cette fonction mesurable est positive (et bornée par $\nu(F)$), on peut donc l'intégrer sur E , ce qui définit $m(C)$ par (6.1).

Relaxons à présent l'hypothèse que $\nu(E) < \infty$, en ne supposant plus que la σ -finitude. On

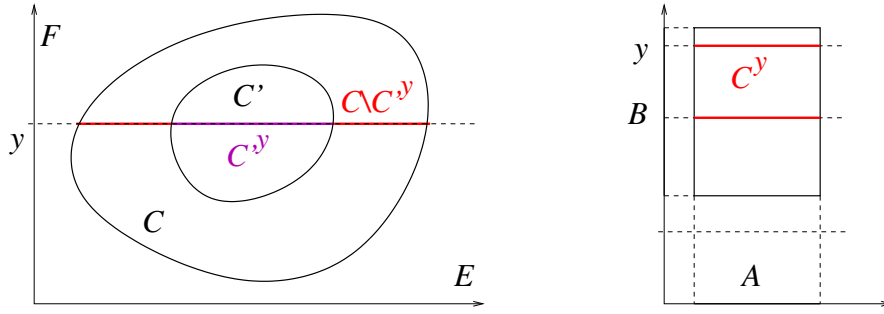


FIGURE 6.2 – Illustration de la construction de la mesure produit à partir de la seconde formule. Gauche : comparaison de C^y et $(C \setminus C')^y$ pour $C \supset C'$. Droite : tranches C^y pour $C = A \times B$ un pavé.

peut alors se servir de la suite croissante de mesurables $(B_n \in \mathcal{B})_{n \in \mathbb{N}}$ comme ci-dessus, tels que $F = \bigcup_n B_n$ et $\nu(B_n) < \infty$. Ceci nous permet de définir une famille croissante de mesures finies $\nu_n(B) = \nu(B \cap B_n)$, telles que $\nu = \lim_{n \rightarrow \infty} \uparrow \nu_n$. Pour tout $C \in \mathcal{A} \otimes \mathcal{B}$, chaque fonction $x \mapsto \nu_n(C_x)$ est mesurable, de sorte que $\nu(C_x) = \lim_{n \rightarrow \infty} \uparrow \nu_n(C_x)$ l'est également. On peut donc encore définir $m(C)$ par la formule (6.1).

Montrons que la fonction m est bien une mesure. On a évidemment $m(\emptyset) = 0$. Si $(C_n)_{n \in \mathbb{N}}$ est une suite de parties disjointes de $\mathcal{A} \otimes \mathcal{B}$, alors pour tout $x \in E$, les $((C_n)_x)_{n \in \mathbb{N}}$ forment une famille de parties mesurables disjointes de F , de sorte que

$$\nu \left(\left(\bigsqcup_n C_n \right)_x \right) = \nu \left(\bigsqcup_n (C_n)_x \right) = \sum_{n \in \mathbb{N}} \nu((C_n)_x).$$

En intégrant cette égalité sur E , on trouve

$$\begin{aligned} m \left(\bigsqcup_n C_n \right) &= \int_E \nu \left(\bigsqcup_n (C_n)_x \right) \mu(dx) \\ &= \int_E \left(\sum_{n \in \mathbb{N}} \nu((C_n)_x) \right) \mu(dx) \\ &= \sum_{n \in \mathbb{N}} \int_E \nu((C_n)_x) \mu(dx) \\ &= \sum_{n \in \mathbb{N}} m(C_n). \end{aligned}$$

La troisième égalité utilise la Proposition 3.11 2., qui découle du TCM. Ceci confirme que m est bien une mesure définie sur la tribu $\mathcal{A} \otimes \mathcal{B}$. Elle vérifie, pour tout pavé, $m(A \times B) = \mu(A)\nu(B)$.

On peut de même construire la fonction

$$m'(C) = \int_F \mu(C^y) \nu(dy) \quad \text{pour tout } C \in \mathcal{A} \otimes \mathcal{B},$$

qui définit une mesure sur $\mathcal{A} \otimes \mathcal{B}$, et telle que $m'(A \times B) = \mu(A)\nu(B)$ pour tout pavé $A \times B$. Le résultat d'unicité montré ci-dessus nous indique alors que $m = m'$, ce qui achève la preuve du *ii*). \square

Remarque 6.5. *i)* Si les mesures μ ou ν ne sont pas σ -finies, on perd en général la propriété *ii*). Par exemple, si les deux espaces sont $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, la première mesure $\mu = \lambda$ (qui est σ -finie) et la seconde ν est la mesure de comptage (non σ -finie), alors la mesure de la « diagonale » $C = \{(x, x) : x \in \mathbb{R}\}$ ne donne pas le même résultat selon qu'on intègre d'abord selon μ ou selon ν :

$$\begin{aligned} \int \nu(C_x) \lambda(dx) &= \int 1 d\lambda = \infty, \\ \text{tandis que} \quad \int \lambda(C^y) \nu(dy) &= \int 0 d\nu = 0. \end{aligned}$$

ii) On peut itérer la procédure de produit : si on a n mesures σ -finies μ_i sur des espaces mesurables (E_i, \mathcal{A}_i) , $i = 1, \dots, n$, on peut définir la mesure produit

$$\mu_1 \otimes (\mu_2 \otimes (\dots \otimes (\mu_{n-1} \otimes \mu_n))) \quad \text{sur l'espace } E_1 \times \dots \times E_n.$$

Cette mesure associe aux pavés $A_1 \times \dots \times A_n$ la valeur $\mu_1(A_1)\mu_2(A_2)\dots\mu_n(A_n)$; cela montre que l'ordre dans lequel on a réalisé les produits n'a pas d'importance. Cette mesure produit peut donc être notée

$$\mu_1 \otimes \mu_2 \otimes \dots \otimes \mu_n.$$

Exemple 6.6. Une fois qu'on a construit la mesure de Lebesgue $\lambda^{\mathbb{R}}$ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, on peut donc construire la mesure produit $\lambda^{\mathbb{R}} \otimes \lambda^{\mathbb{R}}$ sur \mathbb{R}^2 , définie sur la tribu $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^2)$ (car l'espace métrique \mathbb{R} est séparable). Pour tout pavé $]a, b[\times]c, d[$, cette mesure produit donne la valeur $(b - a)(d - c)$, qui est exactement l'aire du pavé pour la mesure de Lebesgue $\lambda^{\mathbb{R}^2}$ sur \mathbb{R}^2 . D'après le corollaire au Lemme de classe monotone, cette coïncidence des valeurs suffit à montrer que $\lambda^{\mathbb{R}} \otimes \lambda^{\mathbb{R}} = \lambda^{\mathbb{R}^2}$. Par itération, on montre ainsi que pour tout $d \geq 1$, $\lambda^{\mathbb{R}^d} = \underbrace{\lambda^{\mathbb{R}} \otimes \dots \otimes \lambda^{\mathbb{R}}}_d$.

6.3 Le théorème de Fubini

Le théorème 6.4 nous a montré que pour tout $C \in \mathcal{A} \otimes \mathcal{B}$, l'intégrale d'une fonction caractéristique $\mathbb{1}_C$ par la mesure produit $\mu \otimes \nu$ peut se calculer de plusieurs façons possibles. Nous allons maintenant généraliser ce résultat à toute fonction mesurable positive sur $E \times F$.

Théorème 6.7. (*Fubini-Tonelli*) Soient μ et ν deux mesures σ -finies respectivement sur (E, \mathcal{A}) et (F, \mathcal{B}) , et soit $f : E \times F \rightarrow [0, \infty]$ une fonction mesurable.

i) Les fonctions

$$\begin{aligned} x \in E &\mapsto \int_F f(x, y) \nu(dy), \\ y \in F &\mapsto \int_E f(x, y) \mu(dx) \end{aligned}$$

sont respectivement \mathcal{A} -mesurable et \mathcal{B} -mesurable.

ii) On a les identités :

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

Démonstration. Le Théorème 6.4 montre que les résultats sont valables pour $f = \mathbb{1}_C$, pour $C \in \mathcal{A} \otimes \mathcal{B}$. Par linéarité, les résultats i) et ii) sont valables pour toute fonction étagée positive. Une fonction mesurable positive peut être obtenue comme limite simple d'une suite croissante de fonctions étagées positives : $f = \lim_{n \rightarrow \infty} \uparrow f_n$. On aura alors, pour tout $x \in E$,

$$\begin{aligned} \int_F f(x, y) \nu(dy) &= \int \lim_n \uparrow f_n(x, y) \nu(dy) \\ &\stackrel{TCM}{=} \lim_n \uparrow \int f_n(x, y) \nu(dy). \end{aligned}$$

Comme les fonctions $x \mapsto \int f_n(x, y) \nu(dy)$ sont toutes mesurables positives, leur limite l'est

aussi. On répète l'argument pour montrer que

$$\begin{aligned}
 \int_E \left(\int_F f(x, y) \nu(dy) \right) \mu(dx) &= \int_E \left(\lim_n \uparrow \int_F f_n(x, y) \nu(dy) \right) \mu(dx) \\
 &\stackrel{TCM}{=} \lim_n \uparrow \int_E \left(\int_F f_n(x, y) \nu(dy) \right) \mu(dx) \\
 &= \lim_n \uparrow \int_{E \times F} f_n d\mu \otimes \nu \\
 &\stackrel{TCM}{=} \int_{E \times F} f d\mu \otimes \nu.
 \end{aligned}$$

On a les mêmes égalités si on intègre d'abord sur E , puis sur F . \square

Exemple 6.8. Si deux fonctions positives $f \in L^1(E, \mu)$ et $g \in L^1(F, \nu)$, alors la fonction $\varphi(x, y) = f(x)g(y)$ est dans $L^1(E \times F, \mu \otimes \nu)$, et

$$\int_{E \times F} f(x)g(y) \mu(dx) \otimes \nu(dy) = \left(\int_E f \mu \right) \left(\int_F g \nu \right).$$

Passons maintenant au « vrai » théorème de Fubini, qui concerne les fonctions intégrables.

Théorème 6.9. (Fubini-Lebesgue) Soient μ, ν deux mesures σ -finies sur (E, \mathcal{A}) , respectivement (F, \mathcal{B}) . Soit $f \in \mathcal{L}^1(E \times F, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$. Alors :

a) Pour μ -presque tout $x \in E$, la fonction $f_x : y \mapsto f(x, y)$ est dans $\mathcal{L}^1(F, \mathcal{B}, \nu)$

Pour ν -presque tout $y \in F$, la fonction $f^y : x \mapsto f(x, y)$ est dans $\mathcal{L}^1(E, \mathcal{A}, \mu)$.

b) La fonction $x \mapsto \int_F f(x, y) \nu(dy)$ (respectivement $y \mapsto \int_E f(x, y) \mu(dx)$), bien définie pour μ -presque tout x (resp. pour ν -presque tout y), est dans $\mathcal{L}^1(E, \mathcal{A}, \mu)$ (resp. dans $\mathcal{L}^1(F, \mathcal{B}, \nu)$).

c) On a les identités :

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

Démonstration. a) Le théorème précédent s'applique à la fonction $|f|$, donc on a :

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

La finitude de cette intégrale implique que $\int_F |f(x, y)| \nu(dy) < \infty$ pour μ -presque tout x . Ceci montre que la fonction mesurable $f_x : y \mapsto f(x, y)$ est dans $\mathcal{L}^1(F, \mathcal{B}, \nu)$ pour μ -presque tout

x . Mutatis mutandis, on montre que $f^y : x \mapsto f(x, y)$ est dans $\mathcal{L}^1(E, \mathcal{A}, \mu)$ pour ν -presque tout y .

b) On sépare f en ses parties positive et négative f^\pm , qui sont toutes deux mesurables et intégrables : $f = f^+ - f^-$. Le théorème précédent montre que les deux fonctions $x \mapsto \int_F f^\pm(x, y) \nu(dy)$ sont mesurables, et prennent des valeurs finies pour μ -presque tout x . En dehors d'un ensemble négligeable, la fonction

$$x \mapsto \int_F f^+(x, y) \nu(dy) - \int_F f^-(x, y) \nu(dy)$$

est donc bien définie et prend une valeur finie, qui définit $\int_F f(x, y) \nu(dy)$. On peut décider que $\int_F f(x, y) \nu(dy) = 0$ sur l'ensemble μ -négligeable des $x \in E$ tels que l'intégrale $\int_F |f(x, y)| \nu(dy)$ diverge.

Une fois définie l'intégrale partielle $\int_F f(x, y) \nu(dy)$, l'inégalité triangulaire pour tout $x \in E$ nous donne :

$$\int_E \left| \int_F f(x, y) \nu(dy) \right| \mu(dx) \leq \int_E \left(\int_F |f(x, y)| \nu(dy) \right) \mu(dx) = \int_{E \times F} |f| d\mu \otimes \nu < \infty,$$

ce qui montre que la fonction $x \mapsto \int_F f(x, y) \nu(dy)$ est intégrable.

c) D'après le théorème précédent, ces égalités sont vérifiées pour les fonctions mesurables positives f^\pm , et donnent des valeurs finies à

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

On peut alors prendre la différence de ces deux identités, et ainsi obtenir l'identité pour la fonction f . □

Remarque 6.10. Pour que l'égalité entre les différentes expressions de $\int f d\mu \otimes \nu$ soit réalisée, il est crucial que f soit intégrable sur $E \times F$. On pourra par exemple étudier le cas de la fonction $f(x, y) = 2e^{-2xy} - e^{-xy}$ sur $(x, y) \in]0, \infty] \times]0, 1[$.

6.4 Applications du théorème de Fubini

Les résultats de cette section seront spécifiques à la mesure de Lebesgue λ sur \mathbb{R}^d . On va d'abord montrer la formule (bien connue) d'intégration par parties sur \mathbb{R} , dans le cadre fonctionnel des fonction Lebesgue-intégrables.

6.4.1 Intégration par parties

On considère $f, g : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions boréliennes, *localement intégrables* (on notera $f, g \in L^1_{loc}(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$), ce qui signifie que pour tout intervalle borné $[a, b] \subset \mathbb{R}$, alors $\int_{[a, b]} |f| d\lambda < \infty$, et de même pour g . On peut alors définir des primitives de ces fonctions :

$$F(x) = \int_0^x f(t) dt = \begin{cases} \int_{[0, x]} f d\lambda, & \text{si } x \geq 0 \\ -\int_{[x, 0]} f d\lambda, & \text{si } x < 0, \end{cases} \quad \text{et de même } G(x) = \int_0^x g(t) dt.$$

L'Exemple 3.29 nous indique que les primitives F, G sont des fonctions continues, donc en particulier mesurables et localement intégrables.

Proposition 6.11. *Sous les hypothèses ci-dessus, on a, pour tout intervalle $[a, b] \subset \mathbb{R}$:*

$$F(b)G(b) = F(a)G(a) + \int_a^b f(t)G(t) dt + \int_a^b F(t)g(t) dt,$$

ou de façon équivalente :

$$\int_a^b f(t) (G(t) - G(a)) dt = \int_a^b (F(b) - F(t)) g(t) dt.$$

Démonstration. Comme F, G sont continues, les produits fG et Fg sont mesurables et localement intégrables, donc les intégrales ci-dessus ont un sens. On va montrer la seconde égalité :

$$\begin{aligned} \int_a^b f(t) (G(t) - G(a)) dt &= \int_a^b f(t) \left(\int_a^t g(s) ds \right) dt \\ &= \int_a^b \left(\int_a^b f(t) \mathbb{1}_{\{s \leq t\}} g(s) ds \right) dt \\ &\stackrel{\text{Fubini}}{=} \int_a^b \left(\int_a^b f(t) \mathbb{1}_{\{s \leq t\}} g(s) dt \right) ds \\ &= \int_a^b g(s) \left(\int_a^b f(t) \mathbb{1}_{\{s \leq t\}} dt \right) ds \\ &= \int_a^b g(s) \left(\int_s^b f(t) dt \right) ds \\ &= \int_a^b g(s) (F(b) - F(s)) ds. \end{aligned}$$

On a appliqué le théorème de Fubini-Lebesgue à la fonction $\varphi(s, t) = f(t) \mathbb{1}_{\{s \leq t\}} g(s)$, qui est

bien dans $L^1([a, b] \times [a, b], \lambda)$:

$$\int_{[a,b]^2} |\varphi(s, t)| ds dt \leq \int_{[a,b]^2} |f(s)g(t)| ds dt;$$

comme f et g sont intégrables sur $[a, b]$, l'exemple 6.8 après le théorème de Fubini-Tonnelli montre que l'intégrale ci-dessus est égale à $\int |f(s)| ds \times \int |g(t)| dt < \infty$. \square

6.4.2 Propriétés de la convolution

Soient $f, g : \mathbb{R}^d \rightarrow \mathbb{R}$ deux fonctions mesurables. Pour tout $x \in \mathbb{R}^d$, l'intégrale de convolution

$$f * g(x) = \int f(x - y)g(y) dy$$

est bien définie si la fonction $y \mapsto f(x - y)g(y)$ est intégrable. La mesure de Lebesgue est invariante par translation, ainsi que par symétrie $y \mapsto -y$, donc elle l'est aussi par la transformation $y \mapsto x - y$. Ceci montre que $g * f(x) = f * g(x)$.

Remarque 6.12. Une modification de f ou g sur un ensemble λ -négligeable ne modifie pas la valeur de l'intégrale $f * g(x)$.

On cherche des conditions d'intégrabilité sur les fonctions individuelles f, g , telles que $f * g$ est bien définie, dans un sens plus large.

Proposition 6.13. *Supposons que $f, g \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$. Alors, pour λ -presque tout $x \in \mathbb{R}^d$, la convolution $f * g(x)$ est bien définie. De plus, $f * g \in L^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$, et on a l'inégalité :*

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1.$$

Ceci est un premier exemple de l'inégalité de Young.

La proposition ci-dessus indique que $f * g(x)$ est soit non défini (pour x dans un ensemble λ -négligeable), soit défini, et sa valeur est alors indépendante du choix de représentants f, g dans leurs classes L^1 respectives. Sur les points indéfinis, on peut décider que $f * g$ prend une valeur arbitraire (par exemple zéro).

Démonstration. Le Théorème de Fubini-Tonnelli montre que :

$$\begin{aligned}
 \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-t)| |g(t)| dt \right) dx &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-t)| |g(t)| dx \right) dt \\
 &= \int_{\mathbb{R}^d} |g(t)| \left(\int_{\mathbb{R}^d} |f(x-t)| dx \right) dt \\
 &= \int_{\mathbb{R}^d} |g(t)| \left(\int_{\mathbb{R}^d} |f(x)| dx \right) dt \\
 &= \left(\int_{\mathbb{R}^d} |g(t)| dt \right) \left(\int_{\mathbb{R}^d} |f(x)| dx \right) < \infty,
 \end{aligned}$$

où à la troisième égalité on a utilisé l'invariance par translation de λ . Ceci montre que la fonction $x \mapsto \int_{\mathbb{R}^d} |f(x-t)| |g(t)| dt$ est finie λ -p.p., ce qui fournit la première affirmation. En ces points où la fonction est définie, la convolution $f * g(x)$ est bien définie et vérifie

$$|f * g(x)| \leq \int_{\mathbb{R}^d} |f(x-t)| |g(t)| dt = |f| * |g|(x)$$

Les égalités ci-dessus nous montrent que le membre de droite, et donc aussi de gauche, sont intégrables sur \mathbb{R}^d , et qu'on a :

$$\int |f * g(x)| dx \leq \int_{\mathbb{R}^d} (|f| * |g|(x)) dx \leq \|f\|_1 \|g\|_1.$$

□

En jouant avec l'échelle des espaces L^p , on obtient une seconde forme d'inégalité de Young.

Proposition 6.14. *Soit $p \in [1, \infty[$ et $q \in]1, \infty]$ son exposant conjugué ($\frac{1}{p} + \frac{1}{q} = 1$). Soient $f \in L^p(\mathbb{R}^d, \lambda)$ et $g \in L^q(\mathbb{R}^d, \lambda)$. Alors, pour tout $x \in \mathbb{R}^d$, la convolution $f * g(x)$ est bien définie, et la fonction $f * g$ est uniformément continue et bornée sur \mathbb{R}^d .*

Preuve. Pour tout $x \in \mathbb{R}^d$, on utilise l'inégalité de Hölder pour les fonctions $f(x - \bullet)$ et $g(\bullet)$: les hypothèses impliquent que la fonction produit est intégrable, et que :

$$\int |f(x-y)g(y)| dy \leq \|f(x - \bullet)\|_p \|g\|_q = \|f\|_p \|g\|_q,$$

où la dernière égalité est due à l'invariance de λ par la transformation $t \mapsto x - t$.

Pour montrer la continuité uniforme, on se sert du Lemme suivant, qui décrit une famille de fonctions translatées les unes des autres :

Lemme 6.15. *Notons l'opérateur de translation $\sigma_x(y) = y - x$. Pour $f \in L^p(\mathbb{R}^d, \lambda)$, d'exposant $p \in [1, \infty]$, l'application $x \mapsto f \circ \sigma_x$ de \mathbb{R}^d dans L^p est uniformément continue.*

Finissons la preuve de la Proposition en se servant du Lemme. Pour $x, x' \in \mathbb{R}^d$,

$$\begin{aligned} |f * g(x) - f * g(x')| &\leq \int |f(x - y) - f(x' - y)| |g(y)| dy \\ &\stackrel{\text{Hölder}}{\leq} \|f(x - \bullet) - f(x' - \bullet)\|_p \|g\|_q \\ &\leq \|f(x + \bullet) - f(x' + \bullet)\|_p \|g\|_q \\ &\leq \|f \circ \sigma_{-x} - f \circ \sigma_{-x'}\|_p \|g\|_q. \end{aligned}$$

Le lemme montre que le membre de droite tend vers zéro uniformément lorsque $|x - x'| \rightarrow 0$.
□

Preuve du Lemme. On utilise la densité des fonctions $C_c(\mathbb{R}^d)$ dans $L^p(\mathbb{R}^d, \lambda)$. Si $f \in C_c(\mathbb{R}^d)$, alors

$$\int |f \circ \sigma_x - f \circ \sigma_y|^p d\lambda = \int |f(t - x) - f(t - y)|^p dt = \int |f(z) - f(z - (y - x))|^p dz,$$

en se servant de l'invariance de λ sous l'action de σ_x . Comme f est continue à support compact, elle est uniformément continue, donc $\|f - f(\bullet - (y - x))\|_{\text{sup}}$ tend vers zéro lorsque $(y - x) \rightarrow 0$, et donc l'intégrale $\|f \circ \sigma_x - f \circ \sigma_y\|_p^p$ également.

Démonstration. Si maintenant $f \in L^p$, on considère une suite (f_n) de fonctions continues à support compact, qui convergent vers f dans L^p . En utilisant l'inégalité de Minkowski, on obtient :

$$\begin{aligned} \|f \circ \sigma_x - f \circ \sigma_y\|_p &\leq \|f \circ \sigma_x - f_n \circ \sigma_x\|_p + \|f_n \circ \sigma_x - f_n \circ \sigma_y\|_p + \|f \circ \sigma_y - f_n \circ \sigma_y\|_p \\ &\leq 2\|f - f_n\|_p + \|f_n \circ \sigma_x - f_n \circ \sigma_y\|_p. \end{aligned}$$

Pour $\varepsilon > 0$ fixé, on choisit $n \in \mathbb{N}$ tel que $\|f - f_n\|_p \leq \varepsilon/3$; comme f_n est absolument continue, on trouve $\delta = \delta(n, \varepsilon) > 0$ tel que $\|f_n \circ \sigma_x - f_n \circ \sigma_y\|_p < \varepsilon/3$ pour tous les x, y tels que $|x - y| \leq \delta$. Cette condition implique donc que $\|f \circ \sigma_x - f \circ \sigma_y\|_p \leq \varepsilon$. □

6.4.3 Une application de la convolution : régularisation de la mesure de Dirac

Une suite de fonctions $(\varphi_n \in C_c(\mathbb{R}^d, \mathbb{R}))_{n \in \mathbb{N}}$ est appelée une approximation de la mesure de Dirac en zéro, si :

- il existe un compact $K \subset \mathbb{R}^d$ tel que $\text{supp } \varphi_n \subset K$ pour tout $n \in \mathbb{N}$;
- pour tout n , $\varphi_n \geq 0$ et vérifie $\int_{\mathbb{R}^d} \varphi_n dx = 1$;
- pour tout $\delta > 0$, $\lim_{n \rightarrow \infty} \int_{\{|x| > \delta\}} \varphi_n dx = 0$. Cette dernière propriété montre que la fonction φ_n se concentrent de plus en plus autour de l'origine.

Une façon de construire une telle suite (φ_n) est d'utiliser des dilatations d'une fonction $\varphi \in C_c(\mathbb{R}^d, \mathbb{R}_+)$ donnée, telle que $\int \varphi(x) dx = 1$: on définit pour tout $n \geq 1$:

$$\varphi_n(x) := n^d \varphi(nx), \quad \forall x \in \mathbb{R}^d.$$

On peut choisir φ (et donc ses dilatées φ_n) de classe C^∞ , par exemple en prenant :

$$\varphi(x) = C \exp\left(-\frac{1}{1-|x|^2}\right) \mathbb{1}_{\{|x| < 1\}},$$

avec la constante $C > 0$ choisie de façon à ce que l'intégrale $\int \varphi d\lambda = 1$. On vérifie que cette fonction est lisse sur la sphère $\{|x| = 1\}$, et donc sur \mathbb{R}^d entier.

En quoi ces fonctions φ_n approximent-elles la mesure de Dirac en l'origine ? Pour une fonction mesurable quelconque f , on a, pour tout $x \in \mathbb{R}^d$:

$$f(x) = \int f(x-y) \delta_0(dy),$$

ce qui ressemble à une formule de convolution de f par la mesure δ_0 . Les fonctions φ_n , continues à support compact, fournissent des approximations à cette « convolution ».

Proposition 6.16. *Soit $(\varphi_n)_{n \in \mathbb{N}}$ une approximation de δ_0 .*

i) *Si $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est continue, alors $\varphi_n * f(x) \rightarrow f(x)$ lorsque $n \rightarrow \infty$, uniformément sur tout compact.*

ii) *Si $f \in L^p(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$, avec $p \in [1, \infty[$, on a $\varphi_n * f \rightarrow f$ dans L^p lorsque $n \rightarrow \infty$.*

Démonstration. En utilisant la propriété de localisation de φ_n , on décompose :

$$\varphi_n * f(x) = \int_{\{|y| \leq \delta\}} f(x-y) \varphi_n(y) dy + \int_{\{|y| > \delta\}} f(x-y) \varphi_n(y) dy,$$

et ces deux intégrales sont de toute façon localisées dans le compact K . On suppose que $x \in K'$ un compact de \mathbb{R} . Comme $y \in K$, la fonction f , restreinte au compact $K' - K$, est

uniformément continue ; si on choisit un $\varepsilon > 0$, en prenant $\delta > 0$ assez petit et $|y| \leq \delta$, on aura $|f(x - y) - f(x)| \leq \varepsilon$, donc la première intégrale vérifie :

$$\left| \int_{\{|y| \leq \delta\}} f(x - y) \varphi_n(y) dy - f(x) \int_{\{|y| \leq \delta\}} \varphi_n(y) dy \right| \leq \varepsilon \int_{\{|y| \leq \delta\}} \varphi_n(y) dy \leq \varepsilon.$$

Comme $|f| \leq C$ sur le compact $K' - K$, la seconde intégrale vérifie

$$\left| \int_{\{|y| > \delta\}} f(x - y) \varphi_n(y) dy \right| \leq C \int_{\{|y| > \delta\}} \varphi_n(y) dy.$$

En prenant n assez grand, le membre de droite est plus petit que ε . On a donc montré que pour n assez grand,

$$|\varphi_n * f(x) - f(x)| \leq 2\varepsilon, \quad \forall x \in K',$$

ce qui prouve le *i*). Pour prouver le *ii*), prenons deux fonctions $f, g \in L^p(\mathbb{R}^d, \lambda)$. Comme $\varphi_n \in C_c(K)$ avec K un compact, on a en particulier $\varphi_n \in L^q$, donc $\varphi_n * f(x)$ et $\varphi_n * g(x)$ sont bien définies et uniformément continues. Comparons ces deux fonctions :

$$\begin{aligned} \int |\varphi_n * f(x) - \varphi_n * g(x)|^p dx &= \int \left| \int \varphi_n(x - y) (f(y) - g(y)) dy \right|^p dx \\ &\leq \int \left(\int \varphi_n(x - y) |f(y) - g(y)| dy \right)^p dx \\ &\stackrel{\text{Jensen}}{\leq} \int \left(\int \varphi_n(x - y) |f(y) - g(y)|^p dy \right) dx \\ &\stackrel{\text{Fubini}}{\leq} \int |f(y) - g(y)|^p \left(\int \varphi_n(x - y) dx \right) dy \\ &\leq \|f - g\|_p^p. \end{aligned}$$

L'inégalité de Jensen est due au fait que $\int \varphi_n(x - y) dy = 1$, autrement dit $\varphi_n(x - \bullet) \cdot dy$ est une mesure de probabilité.

Le Corollaire 5.19 nous enseigne que $f \in L^p(\mathbb{R}^d, \lambda)$ peut être approximée dans L^p par des fonctions $f_k \in C_c(\mathbb{R}^d)$. On choisit une telle approximation f_k telle que $\|f - f_k\|_p \leq \varepsilon$. Comme f_k est à support compact, c'est aussi le cas de $\varphi_n * f_k$, qui sont supportées sur $\text{supp } f_k - K$. Le *i*) nous montre que n assez grand, $|\varphi_n * f_k(x) - f_k(x)|^p \leq \frac{\varepsilon^p}{\lambda(\text{supp } f_k - K)}$ sur tout $x \in \text{supp } f_k - K$, de sorte que $\|\varphi_n * f_k - f_k\|_p \leq \varepsilon$. En utilisant l'inégalité ci-dessus, on trouve donc, pour n

assez grand :

$$\begin{aligned}\|\varphi_n * f - f\|_p &\leq \|\varphi_n * f - \varphi_n * f_k\|_p + \|\varphi_n * f_k - f_k\|_p \\ &\leq \|f - f_k\|_p + \|\varphi_n * f_k - f_k\|_p \leq 2\varepsilon.\end{aligned}$$

□

Cette proposition nous permet donc de construire explicitement des approximations continues à support compact de fonctions L^p : cette construction est plus simple que celle du Théorème 5.18, mais elle est spécifique à la mesure de Lebesgue sur \mathbb{R}^d .

Application de l'application : le théorème de Weierstrass

En dimension $d = 1$, les fonctions

$$\varphi_n(x) = c_n(1 - x^2)^n \mathbb{1}_{\{|x| \leq 1\}}, \quad \text{pour un } c_n > 0 \text{ bien choisi,}$$

forment bien une suite approximant la mesure δ_0 . Soit $[a, b] \subset]0, 1[$ un intervalle, et $f : [a, b] \rightarrow \mathbb{R}$ une fonction continue sur cet intervalle. On peut facilement prolonger f en une fonction $\tilde{f} \in C_c(]0, 1[)$. La Proposition ci-dessus montre alors que les fonctions $\varphi_n * \tilde{f}$ convergent vers \tilde{f} uniformément sur $[0, 1] - [-1, 1] = [-1, 2]$. D'autre part, lorsque $x \in]0, 1[$, l'intégrale

$$\varphi_n * \tilde{f}(x) = \int \varphi_n(x - y) \tilde{f}(y) dy = c_n \int (1 - (x - y)^2)^n \tilde{f}(y) dy,$$

puisque pour $y \in \text{supp } \tilde{f} \subset]0, 1[$ et $x \in]0, 1[$, on a forcément $(x - y) \in]-1, 1[$. En développant le polynôme de l'intégrand, on s'aperçoit que le membre de droite est un polynôme en x . On a donc retrouvé le célèbre théorème de Weierstrass, qui nous dit que la fonction continue $f : [a, b] \rightarrow \mathbb{R}$ peut être approchée uniformément sur $[a, b]$ par des polynômes.

7 Formule de changement de variables

Dans le chapitre précédent nous avons introduit plusieurs outils permettant de calculer explicitement les intégrales sur \mathbb{R}^d , grâce au théorème de Fubini. Nous avons également montré la formule d'intégration par parties, qui est un outil majeur dans le calcul d'intégrales. Il nous reste à introduire un troisième outil très important, la formule de changement de variables, qui permet souvent de simplifier les expressions.

7.1 Changement de variable affine

On commence par considérer une application affine sur \mathbb{R}^d , définie par une matrice réelle M de dimension $d \times d$ (on notera parfois $M \in \mathbb{R}^{d \times d}$, et un vecteur $b \in \mathbb{R}^d$. La transformation affine est donnée par

$$f(x) = Mx + b, \quad \forall x \in \mathbb{R}^d.$$

On se restreindra au cas où cette transformation est inversible, autrement dit M est une matrice inversible.

Proposition 7.1. *Soit $f(x) = Mx + b$ une transformation affine sur \mathbb{R}^d . Alors, pour tout borélien $A \in \mathcal{B}(\mathbb{R}^d)$, la mesure de Lebesgue de $f(A)$ est donnée par*

$$\lambda(f(A)) = |\det(M)| \lambda(A).$$

Cette expression peut se réécrire en termes de la mesure $f^{-1}(\lambda)$, image de λ par l'application affine f^{-1} .

Démonstration. Comme f est inversible et lisse, on vérifie que $f(A)$ est Borélien si A l'est : en effet, $f(A) = (f^{-1})^{-1}(A)$, avec f^{-1} une transformation affine, donc continue, donc mesurable. On sait aussi que la mesure de Lebesgue est invariante par translation, donc

$$\lambda(f(A)) = \lambda(M(A) + b) = \lambda(M(A)).$$

Pour simplifier, on supposera donc que $b = 0$, autrement dit f est une transformation linéaire. Que sait-on sur la mesure image $f^{-1}(\lambda)$? Pour tout vecteur $a \in \mathbb{R}^d$, on a :

$$f^{-1}(\lambda)(A + a) := \lambda(M(A + a)) = \lambda(M(A) + Ma) = \lambda(M(A)) = f^{-1}(\lambda)(A),$$

où l'invariance par translation de λ a été utilisée dans la 3e inégalité. On a donc montré que $f^{-1}(\lambda)$ est également invariante par translation. D'après les Théorème 4.9, on doit alors avoir

$$f^{-1}(\lambda) = c_f \lambda,$$

pour un certain coefficient $c_f \geq 0$. Pour montrer la proposition, il faut montrer que $c_f = |\det(M)|$.

Supposons tout d'abord que M est une *matrice orthogonale*. La boule unité $B_d = \{x \in \mathbb{R}^d, |x| < 1\}$ est invariante par l'application linéaire f correspondante, donc on doit avoir $c_f = 1$, qui est effectivement égal à $|\det(M)|$.

Supposons ensuite que M est une matrice symétrique définie positive (donc inversible) : il existe alors une matrice orthogonale P diagonalisant M : $M = PD^tP$, avec D une matrice diagonale à coefficients diagonaux $\alpha_i > 0$. Soit $C = [0, 1]^d$ l'hypercube unité. Son image par la matrice orthogonale P vérifie :

$$f(P(C)) = MP(C) = PD^tPP(C) = PD(C),$$

avec $D(C) = \prod_{i=1}^d [0, \alpha_i[$ est un pavé. En utilisant la remarque ci-dessus dans le cas d'une transformation orthogonale on peut calculer la mesure de Lebesgue de cet élément :

$$\lambda(P(D(C))) = 1 \times \lambda(D(C)) = \prod_{i=1}^d \alpha_i.$$

En remarquant que $\lambda(P(C)) = \lambda(C) = 1$, on a donc montré que

$$f^{-1}(\lambda)(P(C)) = \prod_{i=1}^d \alpha_i = \left(\prod_{i=1}^d \alpha_i \right) \lambda(P(C)),$$

donc pour cette matrice M , la constante de multiplication c_f vaut

$$c_f = \prod_{i=1}^d \alpha_i = \det(M).$$

Enfin, une matrice inversible réelle admet une décomposition polaire de la forme $M = PS$, où $S = \sqrt{^tMM}$ est symétrie définie positive, et $P = MS^{-1}$ est orthogonale. La transformation f associée à M est donc la composée des deux types de transformations étudiées ci-dessus.

On trouve donc dans ce cas une constante

$$c_f = \det(S) = |\det(M)|.$$

□

7.2 Changement de variable non-linéaire

On considère à présent U et D deux ouverts de \mathbb{R}^d , reliés l'un à l'autre par un difféomorphisme φ de régularité C^1 , ainsi que son inverse. En tout point $u \in U$, on notera $\varphi'(u)$ l'application tangente (= la dérivée de φ) au point u . C'est une matrice inversible $d \times d$, et on appellera

$$J_\varphi(u) = \det(\varphi'(u))$$

le Jacobien de φ au point u . Le théorème de changement de variable nous fait passer des variables $x \in D$ (variables « images » par φ) aux variables $u \in U$ (variable « antécédents »).

Théorème 7.2. *Soit $\varphi : U \rightarrow D$ un difféomorphisme C^1 . Alors, pour toute fonction borélienne positive $f : D \rightarrow \mathbb{R}_+$ ou intégrable $f \in \mathcal{L}^1(D, \lambda)$, on a :*

$$\int_D f(x) dx = \int_U f(\varphi(u)) |J_\varphi(u)| du.$$

De façon équivalente, la mesure $\varphi^{-1}(\lambda)$ sur U est donnée par la mesure modulée $|J_\varphi| \cdot \lambda$.

Remarque. Si $\varphi(u) = Mu + b$ est une transformation affine, on a en tout point $J_\varphi(u) = \det(M)$, donc on retrouve la formule de la section précédente.

Preuve : Par les arguments habituels, on se ramène au cas où $f = \mathbb{1}_A$ pour un borélien $A \subset D$. On veut donc montrer que

$$\lambda(A) = \int_{\varphi^{-1}(A)} |J_\varphi(u)| du,$$

on de façon équivalente, pour le borélien $B = \varphi^{-1}(A) \in \mathcal{B}(U)$,

$$\lambda(\varphi(B)) = \int_B |J_\varphi(u)| du \iff \varphi^{-1}(\lambda)(B) = \int_B |J_\varphi| d\lambda.$$

Remarquons que si A est un borélien quelconque de D , alors B est un borélien quelconque de U .

Démonstration. L'idée va consister à approximer *localement* φ par une transformation affine, pour laquelle on connaît la formule de changement de variables, puis à recoller tous ces morceaux locaux. On commence par contrôler l'image de « petits cubes » : \square

Lemme 7.3. *Soit $K \Subset U$ un compact, et soit $\varepsilon > 0$. Alors il existe $\delta > 0$ tel que, pour tout cube C parallèle aux axes, centré en un point $u_0 \in K$ et de côté $\leq \delta$, on a l'encadrement :*

$$(1 - \varepsilon) |J_\varphi(u_0)| \lambda(C) \leq \lambda(\varphi(C)) \leq (1 + \varepsilon) |J_\varphi(u_0)| \lambda(C).$$

Autrement dit, à un facteur $(1 \pm \varepsilon)$ près, $\lambda(\varphi(C)) \approx |J_\varphi(u_0)| \lambda(C)$.

Démonstration. On se sert du développement limité de φ basé en u_0 :

$$\varphi(u) = \varphi(u_0) + \varphi'(u_0) \cdot (u - u_0) + o(|u - u_0|) \quad \text{lorsque } u \rightarrow u_0.$$

On note $f(v) = \varphi(u_0) + \varphi'(u_0) \cdot v$ la partie affine, et $r(u) = f(u - u_0) - \varphi(u)$ le terme de reste. On peut choisir $\delta > 0$ tel que, si $|u - u_0| < d\delta$, alors $|r(u)| \leq \varepsilon|u - u_0|$. En définissant $g(u) := \varphi'(u_0)^{-1} \cdot r(u)$, l'expression ci-dessus peut se réécrire :

$$\varphi(u) = \varphi(u_0) + \varphi'(u_0) \cdot ((u - u_0) + g(u)) = f((u - u_0) + g(u)), \quad (7.1)$$

et on a l'estimation

$$|u - u_0| < d\delta \implies |g(u)| \leq \varepsilon a |u - u_0|, \quad \text{où } a := \sup \{ \|\varphi'(v)^{-1}\|_{\text{sup}}, v \in K \}.$$

Considérons un cube C centré en u_0 , et de côté $\leq \delta$, et appelons $\tilde{C} = C - u_0$ le même cube centré en l'origine. Les estimées ci-dessus montrent que pour tout $u \in C$, d'une part $|u - u_0| < d\delta$, d'autre part le point $(u - u_0) + g(u)$ sera contenu dans le cube $(1 + ad\varepsilon)\tilde{C}$. On aura donc :

$$\varphi(C) \subset f\left((1 + ad\varepsilon)\tilde{C}\right),$$

et donc au niveau de la mesure de Lebesgue :

$$\begin{aligned} \lambda(\varphi(C)) &\leq \lambda\left(f\left((1 + ad\varepsilon)\tilde{C}\right)\right), \\ &\leq |\det \varphi'(u_0)| \lambda\left((1 + ad\varepsilon)\tilde{C}\right) \\ &\leq |J_\varphi(u_0)| (1 + ad\varepsilon)^d \lambda(C). \end{aligned}$$

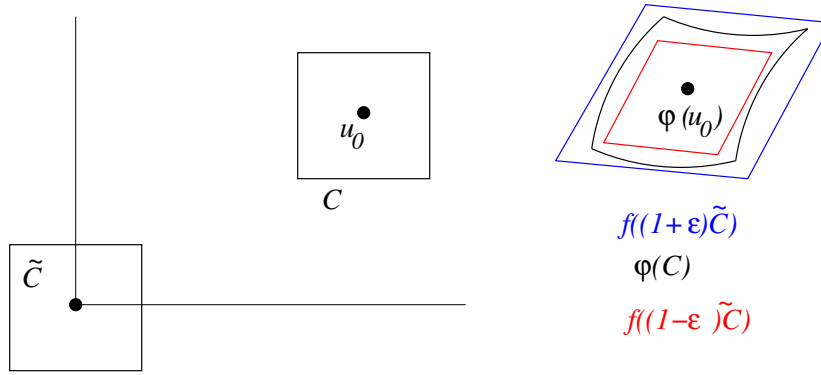


FIGURE 7.1 – Les « petits cubes » et leurs images

En prenant $\varepsilon' = ad\varepsilon$, on retrouve bien la borne supérieure voulue, avec ε' en place de ε . Inversement, on vérifie que l'ensemble

$$\{(u - u_0) + g(u), u \in C\}$$

contient le cube $(1 - ad\varepsilon)\tilde{C}$, de sorte que

$$\begin{aligned} \lambda(\varphi(C)) &\geq \lambda\left(f\left((1 - ad\varepsilon)\tilde{C}\right)\right) \\ &\geq |J_\varphi(u_0)| (1 - ad\varepsilon)^d \lambda(C), \end{aligned}$$

ce qui montre la borne inférieure du Lemme. \square

Revenons maintenant à la preuve du Théorème. Pour tout $n \geq 1$, on peut paver l'espace \mathbb{R}^d de n -cubes de la forme

$$C = \prod_{i=1}^d \left[\frac{k_i}{2^n}, \frac{k_i + 1}{2^n} \right], \quad k_i \in \mathbb{Z}.$$

On note \mathcal{C}_n la classe formée par ces n -cubes.

Démonstration. Choisissons C_0 un n_0 -cube, tel que son adhérence $\overline{C_0}$ est contenue dans l'ouvert U , et fixons $\varepsilon > 0$. On peut alors choisir $n \geq n_0$ tel que

- le Lemme ci-dessus s'applique en prenant le compact $K = \overline{C_0}$ et $\delta = 2^{-n}$;
- pour tous $u, v \in C_0$ tels que $|u - v| < d\delta$, on aura aussi :

$$(1 - \varepsilon)|J_\varphi(u)| \leq |J_\varphi(v)| \leq (1 + \varepsilon)|J_\varphi(u)|, \quad (7.2)$$

où on utilise le fait que J_φ est une fonction continue, donc uniformément continue sur le

compact K . On vérifie facilement que le « grand cube » C_0 peut se partitionner en une union disjointe de $2^{(n-n_0)d}$ « petits cubes » $C \in \mathcal{C}_n$. Pour chacun de ces « petits cubes » C , notons u_C le centre du cube. Pour calculer $\lambda(\varphi(C_0))$, on décompose C_0 en les petits cubes :

$$\begin{aligned} \lambda(\varphi(C_0)) &= \sum_{\mathcal{C}_n \ni C \subset C_0} \lambda(\varphi(C)) \\ &\leq \sum_{\mathcal{C}_n \ni C \subset C_0} (1 + \varepsilon) |J_\varphi(u_C)| \lambda(C) \\ &\leq \sum_{\mathcal{C}_n \ni C \subset C_0} (1 + \varepsilon)^2 \int_C |J_\varphi(u)| d\lambda \\ &\leq (1 + \varepsilon)^2 \int_{C_0} |J_\varphi(u)| d\lambda. \end{aligned}$$

La première inégalité est une application du Lemme à chacun des petits cubes C , la seconde résulte de l'estimée (7.2) avec $v = u_C$, qui revient à montrer que la somme de Riemann approche l'intégrale.

En utilisant les bornes inférieures dans le Lemme et dans (7.2), on montre de même

$$\lambda(\varphi(C_0)) \geq (1 - \varepsilon)^2 \int_{C_0} |J_\varphi(u)| d\lambda.$$

Comme le paramètre $\varepsilon > 0$ peut être choisi arbitrairement petit, on a donc montré que pour le carré C_0 ,

$$\lambda(\varphi(C_0)) = \int_{C_0} |J_\varphi(u)| d\lambda.$$

On a donc montré la formule voulue dans le cas où $A = C_0$, d'adhérence contenue dans U . \square

Le théorème exprime le fait que les deux mesures suivantes sur U sont égales :

1. la mesure $\mu = \varphi^{-1}(\lambda)$, image par φ^{-1} de la mesure de Lebesgue sur D , donc telle que pour tout $A \subset U$ mesurable, $\mu(A) = \lambda(\varphi(A))$;
2. la mesure $\tilde{\mu} = |J_\varphi| \cdot \lambda$.

Jusqu'à présent on a montré qu'elles coïncident sur tous les cubes C (d'ordre n quelconque) tels que $\overline{C} \subset U$. On veut appliquer le Corollaire 2.16 du Lemme de classe montone pour conclure que $\mu = \tilde{\mu}$. Appelons $\mathcal{C}(U)$ la classe formée par tous ces cubes $\overline{C} \subset U$, ainsi que les unions finies de tels cubes. Tout élément de $\mathcal{C}(U)$ peut s'écrire comme une union disjointe finie de n -cubes, pour un certain n ; par additivité des mesures, on a donc $\mu_{\mathcal{C}(U)} = \tilde{\mu}_{\mathcal{C}(U)}$. Cette classe $\mathcal{C}(U)$ engendre la tribu des boréliens sur U (chaque ouvert de U contient un

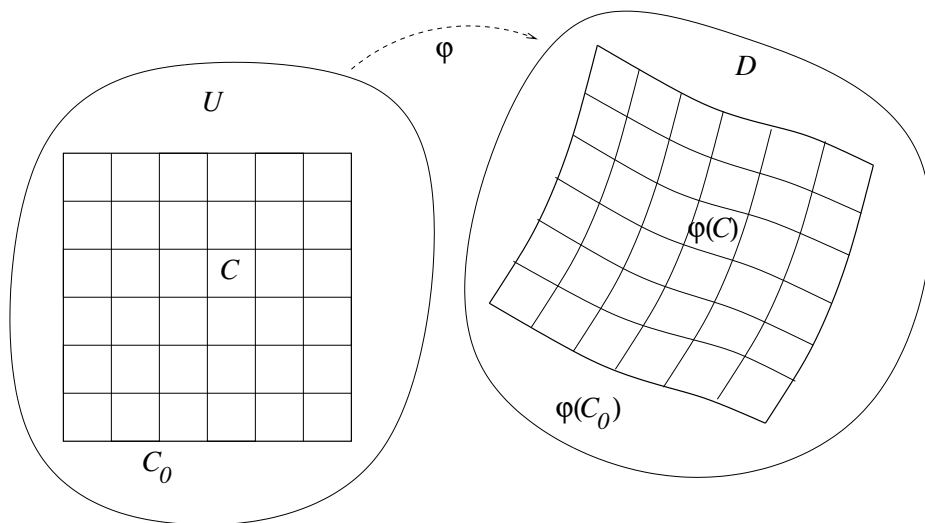


FIGURE 7.2 – Décomposition du « grand cube » C_0 en une union de « petits cubes » C d'ordre n .

de ces cubes), et elle est invariante par intersection, car l'intersection de deux cubes est un cube. D'autre part, si on appelle U_n l'union des cubes $C \in \mathcal{C}(U)$ d'ordre n , qui sont aussi contenus dans la boule $\{|u| \leq n\}$, alors $(U_n)_{n \geq 1}$ est une suite croissante d'ensembles, telle que $\lim_{n \rightarrow \infty} U_n = U$; comme chaque $\overline{U_n}$ est compact les résultats ci-dessus montrent que $\mu(U_n) = \tilde{\mu}(U_n) < \infty$. On est donc en position pour appliquer le point 2. du Corollaire au Lemme de classe monotone, et montrer ainsi que $\mu = \tilde{\mu}$. \square

7.3 Application : coordonnées sphériques

Pour les applications pratiques du calcul d'intégrales, par exemple à la physique, un changement de coordonnées très important consiste à passer des coordonnées euclidiennes (x, y) sur \mathbb{R}^2 , aux coordonnées polaires (r, θ) . On ne peut pas passer du premier système au second par un difféomorphisme défini sur \mathbb{R}^2 tout entier; nous devons donc nous restreindre à un sous-ensemble dense de \mathbb{R}^2 . On prendra ainsi les ouverts

$$U =]0, \infty[\times]0, 2\pi[, \quad D = \mathbb{R}^2 \setminus \{(x, 0); x \leq 0\},$$

et le difféomorphisme $\varphi(r, \theta) = (r \cos \theta, r \sin \theta)$, défini de U vers D .

Ce difféomorphisme $\varphi : U \rightarrow D$ est C^1 (en fait, C^∞), son application tangente vaut

$$\varphi'(r, \theta) = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}, \quad \text{de déterminant } J_\varphi(r, \theta) = r.$$

On voit que ce changement de coordonnées devient singulier lorsque $r \searrow 0$, ce qui explique pourquoi on a exclu ce point de U . Le théorème 7.2 montre alors que pour toute fonction borélienne positive $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$,

$$\begin{aligned} \int_D f(x, y) dx dy &= \int_U f(r \cos \theta, r \sin \theta) r dr d\theta \\ &= \int_0^\infty \int_{-\pi}^\pi f(r \cos \theta, r \sin \theta) r dr d\theta. \end{aligned}$$

Comme la droite $\{(x, 0); x \leq 0\}$ est Lebesgue-négligeable, on peut la rajouter à l'intégrale de gauche sans changer sa valeur, de sorte que :

$$\int_{\mathbb{R}^2} f(x, y) dx dy = \int_0^\infty \int_{-\pi}^\pi f(r \cos \theta, r \sin \theta) r dr d\theta.$$

Le cercle unité $S^1 = \{(x, y) \in \mathbb{R}^2, \sqrt{x^2 + y^2} = 1\}$ peut être paramétré par $\{z_\theta = (\cos \theta, \sin \theta), \theta \in [0, 2\pi[\}$. On peut interpréter $d\theta$ dans la formule ci-dessus comme une mesure sur S^1 , induite par la mesure de Lebesgue sur \mathbb{R}^2 . Le cercle S^1 est un espace métrique, on peut donc considérer sa tribu de Borel $\mathcal{B}(S^1)$. Formellement, la mesure ω_2 d'un borélien $A \subset S^1$ est définie en considérant le cône de \mathbb{R}^2 basé sur A :

$$\Gamma(A) := \{rz; z \in A, r \in [0, 1]\}.$$

On va directement énoncer un résultat en dimension $d \geq 2$. On note $S^{d-1} \subset \mathbb{R}^d$ la sphère unité sur \mathbb{R}^d .

Théorème 7.4. *Pour tout $A \in \mathcal{B}(S^{d-1})$, on définit*

$$\omega_d(A) := d\lambda_d(\Gamma(A)),$$

où le cône $\Gamma(A)$ est défini comme ci-dessus. Alors ω_d est une mesure finie sur S^1 , invariante par les isométries vectorielles de \mathbb{R}^d . Pour toute fonction f borélienne positive sur \mathbb{R}^2 , on a la formule

$$\int_{\mathbb{R}^d} f(x) d\lambda(x) = \int_0^\infty \int_{S^{d-1}} f(rz) r^{d-1} dr d\omega_d(z). \quad (7.3)$$

La masse totale de S^1 est $\omega_2(S^1) = 2\pi$.

La masse totale de ω_d est $\omega_d(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)}$, où $\Gamma(\bullet)$ est la fonction Gamma d'Euler.

En dimension $d = 2$, la mesure ω_2 est l'image de la mesure $d\theta$ par la paramétrisation $\theta \in$

$[0, 2\pi[\rightarrow z_\theta \in S^1$.

Démonstration. Notons que chaque cône $\Gamma(A)$ est une « tranche » de la boule unité B_d . La projection $p : x \in B_d \setminus \{0\} \mapsto \frac{x}{|x|} \in S^{d-1}$ permet de définir le cône $\Gamma(A)$ (privé de l'origine) comme l'image réciproque $\Gamma(A) = p^{-1}(A)$. La mesure ω_d est donc la mesure image de $d\lambda_d|_{B_d \setminus 0}$ à travers la projection p . La mesure totale de ω_d est alors égale à $\omega_d(S^{d-1}) = d\lambda_d(B_d \setminus 0) = d\lambda_d(B_d)$. L'invariance de λ_d par les isométries linéaires (transformations linéaires par des matrices orthogonales) montre que, pour tout $A \in \mathcal{B}(S^{d-1})$ et toute matrice P orthogonale,

$$\omega_d(P(A)) = d\lambda_d(\Gamma(P(A))) = d\lambda_d(P(\Gamma(A))) = d\lambda_d(\Gamma(A)) = \omega_d(A),$$

où on s'est servi du fait que le cône transformé $P(\Gamma(A))$ est égal au cône basé sur $P(A)$. Ceci montre que ω_d est invariante par les transformations orthogonales.

Pour montrer la formule (7.3), on va se contenter de considérer les fonctions caractéristiques $f = \mathbb{1}_B$ sur des boréliens $B \subset \mathbb{R}^d \setminus \{0\}$. On veut donc montrer que $\lambda_d(B)$ est égal à l'expression :

$$\mu(B) = \int_0^\infty \int_{S^{d-1}} \mathbb{1}_B(rz) r^{d-1} dr d\omega_d(z)$$

On va considérer un borélien de la forme :

$$B = \{x \in \mathbb{R}^d \setminus \{0\}, a < |x| \leq b, p(x) \in A\}, \quad \text{avec } A \text{ un borélien de } S^{d-1}, 0 < a < b, \quad (7.4)$$

donc de la forme d'une « couche » d'épaisseur $b - a$ et de base A . Pour un tel borélien, la fonction $\mathbb{1}_B(x) = \mathbb{1}_{]a,b]}(|x|) \mathbb{1}_A\left(\frac{x}{|x|}\right)$ se factorise, ce qui permet de calculer l'intégrale :

$$\begin{aligned} \mu(B) &= \int_0^\infty \int_{S^{d-1}} \mathbb{1}_{]a,b]}(r) \mathbb{1}_A(z) r^{d-1} dr d\omega_d(z) = \omega_d(A) \frac{b^d - a^d}{d}. \\ &= \int_0^\infty \mathbb{1}_{]a,b]}(r) r^{d-1} dr \left(\int_{S^{d-1}} \mathbb{1}_A(z) d\omega_d(z) \right) \\ &= \frac{b^d - a^d}{d} \omega_d(A). \end{aligned}$$

On va calculer la mesure de Lebesgue de cette « couche » B en la reliant à la mesure de $\Gamma(A)$. Pour cela on utilise des dilatations successives. Si on définit $\alpha = \frac{a}{b} \in]0, 1[$, et les couches

$$\Gamma_n(A) = B = \{x \in \mathbb{R}^d \setminus \{0\}, \alpha^{n+1} < |x| \leq \alpha^n, p(x) \in A\} \quad n \geq 0,$$

alors $B = b\Gamma_0(A)$; en vertu du changement de variables par transformations linéaires, cela

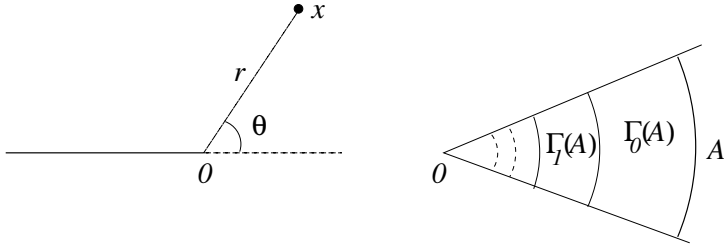


FIGURE 7.3 – Gauche : coordonnées polaires du plan euclidien. Droite : les « couches » $\Gamma_n(A)$ engendrées par $A \subset S^{d-1}$.

implique que $\lambda_d(B) = b^d \lambda_d(\Gamma_0(A))$, donc il nous faut calculer la mesure de $\Gamma_0(A)$. De même, comme $\Gamma_n(A) = \alpha^n \Gamma_0(A)$, on aura $\lambda_d(\Gamma_n(A)) = \alpha^{nd} \lambda_d(\Gamma_0(A))$. On se sert alors de la partition :

$$\Gamma(A) \setminus \{0\} = \bigsqcup_{n=0}^{\infty} \Gamma_n(A)$$

pour montrer que

$$\lambda_d(\Gamma(A)) = \lambda_d(\Gamma(A) \setminus 0) = \sum_{n=0}^{\infty} \lambda_d(\Gamma_n(A)) = \sum_{n=0}^{\infty} \alpha^{nd} \lambda_d(\Gamma_0(A)),$$

$$\text{et donc} \quad \lambda_d(\Gamma(A)) = \frac{1}{1 - \alpha^d} \lambda_d(\Gamma_0(A)).$$

D'après la définition de ω_d , on en déduit :

$$(1 - \alpha^d) \frac{\omega_d(A)}{d} = \lambda_d(\Gamma_0(A)),$$

et finalement

$$\lambda_d(B) = b^d (1 - \alpha^d) \frac{\omega_d(A)}{d} = \frac{b^d - a^d}{d} \omega_d(A).$$

On a donc montré que $\lambda_d(B) = \mu(B)$ pour toute couche (7.4). On note que chaque couche a la forme d'un « pavé mesurable » vis-à-vis du produit $\mathbb{R}_+^* \times S^{d-1}$. La classe \mathcal{C} formée par ces couches est invariante par intersections finies, et il est simple de vérifier qu'elle engendre la tribu des boréliens, et que les couches $B_n = \{x \in \mathbb{R}^2, n^{-1} < |x| \leq n\}$ forment une suite croissante recouvrant $\mathbb{R}^2 \setminus 0$. On est alors en droit d'utiliser le corollaire du Lemme de classe monotone, pour affirmer que $\lambda_d = \mu$.

Il reste à calculer la mesure totale de la sphère S^{d-1} . Une façon pour le faire est de calculer

de différentes manières l'intégrale de la fonction gaussienne

$$G_d(x) = \exp(-|x|^2) = \prod_{i=1}^d \exp(-x_i^2).$$

La factorisation de G_d nous montre, par le théorème de Fubini, que

$$I_d = \int_{\mathbb{R}^d} G_d d\lambda_d = \left(\int_{\mathbb{R}} G_1 d\lambda_1 \right)^d, \quad \text{avec} \quad I_1 = \int_{\mathbb{R}} G_1 d\lambda_1 = \int_{\mathbb{R}} e^{-t^2} dt.$$

En utilisant la formule (7.3), on arrive à :

$$\begin{aligned} I_d &= \int_0^\infty \int_{S^{d-1}} G_d(rz) r^{d-1} dr d\omega_d(z) = \int_0^\infty \int_{S^{d-1}} e^{-r^2} r^{d-1} dr d\omega_d(z) \\ &= \int_0^\infty e^{-r^2} r^{d-1} dr \left(\int_{S^{d-1}} d\omega_d(z) \right) = \omega_d(S^{d-1}) \int_0^\infty e^{-r^2} r^{d-1} dr. \end{aligned}$$

Il reste à évaluer les intégrales

$$\int_0^\infty e^{-r^2} r^{d-1} dr = \frac{1}{2} \int_0^\infty e^{-t} t^{\frac{d}{2}-1} dt = \frac{1}{2} \Gamma(d/2),$$

en utilisant le changement de variable $t = r^2$, puis la définition de la fonction Γ d'Euler. Il reste à calculer la valeur de I_1 . Pour cela on se sert de l'expression simple de I_2 , et du fait que $\omega_2(S^1) = \int_0^{2\pi} d\theta = 2\pi$:

$$I_1^2 = I_2 = \frac{1}{2} \omega_2(S^1) \int_0^\infty e^{-t} dt = \frac{1}{2} \omega_2(S^1) = \pi,$$

$$\text{donc} \quad I_1 = \sqrt{\pi}.$$

On a donc, pour tout $d \geq 2$: $I_d = \pi^{d/2} = \frac{1}{2} \omega_d(S^{d-1}) \Gamma(d/2)$, et donc

$$\omega_d(S^{d-1}) = \frac{2\pi^{d/2}}{\Gamma(d/2)}.$$

Remarquons que les valeurs de $\Gamma(d/2)$ peuvent se calculer par récurrence, en intégrant par parties plusieurs fois l'intégrale $\int_0^\infty e^{-t} t^{\frac{d}{2}-1} dt$. \square

Deuxième partie

Théorie des probabilités

La théorie moderne des probabilités pourrait être interprétée comme une partie de la théorie de l'intégration de Lebesgue. Il s'agit en effet de définir une mesure (dite mesure de probabilité) sur un espace mesurable, et d'étudier certaines classes de fonctions mesurables sur cet espace. Mais c'est surtout l'interprétation de ces objets qui change, ainsi que leurs noms, et leurs notations. Ainsi, les fonctions mesurables sont appelées *variables aléatoires*, et elles jouent un rôle plus important que dans la théorie de la mesure ; au contraire, la structure de l'espace mesuré (noté Ω) est moins centrale que précédemment, et sera souvent omis dans les énoncés. La mesure de probabilité \mathbb{P} sur Ω sera souvent remplacée par ses mesures images à travers les variables aléatoires (qu'on appellera la *loi* de la variable aléatoire).

Vous avez déjà suivi un cours de probabilités discrètes (le résultat d'une expérience ne peut prendre qu'un nombre fini ou dénombrable de valeurs). Ces situations discrètes seront naturellement des cas particuliers de notre théorie générale.

Intégration	Probabilités
Espace mesurable (E, \mathcal{A})	Espace de probabilité (Ω, \mathcal{A})
Mesure de probabilité μ	Mesure de probabilité \mathbb{P}
Sous-ensemble mesurable $A \in \mathcal{A}$	Evènement $A \in \mathcal{A}$
Fonction mesurable $f : (E, \mathcal{A}) \rightarrow (F, \mathcal{B})$	Variable aléatoire X à valeur dans (F, \mathcal{B})
Poids $\mu(A)$	Probabilité $\mathbb{P}(A)$
Intégrale $\int f d\mu$	Espérance $\mathbb{E}[X]$
Mesure image $f_*\mu$ sur (F, \mathcal{B})	Loi P_X de la variable aléatoire X

TABLE 1 – Dictionnaire entre théorie de l'intégration et théorie des probabilités

8 Définitions générales

Nous commencerons ce chapitre par établir un « dictionnaire » entre les notions utilisées précédemment, et les notions de la théorie des probabilités. Définissons (Ω, \mathcal{A}) un espace mesurable, et \mathbb{P} une mesure de probabilité sur cet espace (donc une mesure positive de masse totale égale à 1). $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé un espace de probabilité. En théorie des probabilités, ces objets servir à définir une expérience comportant une *composante aléatoire*, au sens de « non déterministe » : avant de réaliser l'expérience, je n'ai pas de moyen d'en deviner le résultat au vu de mes connaissances présentes.

1. Chaque point $\omega \in \Omega$ représente toutes les éventualités possibles, lorsqu'on réalise l'expérience. Si je connaissais ω , je connaîtrais le résultat de mon expérience, quelle que soit la quantité (ou variable) que je mesure ;
2. la tribu \mathcal{A} est constituée par les *événements*, ce sont les parties de Ω dont je peux déterminer la probabilité. Un événement $A \in \mathcal{A}$ a souvent la signification d'une certaine propriété concrète du système physique qu'on observe : par exemple, un événement pour être défini par la valeur d'une certaine fonction réelle f qu'on va mesurer (disons, la température, ou la quantité d'énergie du système) : $A = f^{-1}(I)$ pour un intervalle $I \subset \mathbb{R}$.
3. pour un événement $A \in \mathcal{A}$, $\mathbb{P}(A)$ est la probabilité que cet événement soit réalisé lors d'une expérience. Que signifie cette probabilité ? Supposons qu'on refait la même expérience un grand nombre N de fois ; à chaque fois, on regarde si l'événement A est réalisé ou non. Dans la limite $N \rightarrow \infty$, on s'imagine que la fraction du nombre d'expériences pour lesquelles A a été réalisé tend vers une limite asymptotique : celle-ci définit alors la probabilité de A :

$$\mathbb{P}(A) = \lim_{N \rightarrow \infty} \frac{N_A}{N}.$$

Il y a bien sûr une difficulté dans l'expression ci-dessus : en général, le nombre N_A est lui-même aléatoire, donc c'est aussi le cas du rapport $\frac{N_A}{N}$; en quelle mesure un tel nombre aléatoire va-t-il converger vers un nombre $\mathbb{P}(A)$ déterministe ? Cette interprétation « expérimentale » de $\mathbb{P}(A)$ sera précisée plus loin.

8.1 Exemples d'espaces de probabilité

1) Tirages finis d'un dé

Le cas le plus simple consiste en un espace de probabilité fini. Supposons qu'on lance un dé à 6 faces : on a alors 6 résultats possibles, donc il est naturel de prendre comme espace de probabilité $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$, avec la tribu totale $\mathcal{P}(\Omega)$. Si le dé n'est pas truqué, chaque $\omega \in \Omega$ a comme probabilité $\mathbb{P}(\omega) = \frac{1}{6}$. Un évènement peut être « je tire le numéro 1 », ou « je tire un numéro pair ».

Supposons qu'on tire le dé 2 fois de suite, et qu'on s'intéresse à toutes les combinaisons possibles. Un évènement sera par exemple « j'ai tiré deux fois 1 », ou « j'ai d'abord tiré 1, puis 2 », ou « la somme de mes deux tirages vaut 5 ». Dans ce cas, l'espace de probabilité naturel est $\Omega_2 = \{1, 2, \dots, 6\}^2$, muni de sa tribu totale. C'est l'espace le plus petit permettant de distinguer tous ces évènements. Si le dé n'est pas truqué, chaque élément $(i, j) \in \Omega_2$ a une probabilité $\frac{1}{6^2}$.

Remarquons que l'espace Ω_2 permet aussi de décrire l'expérience précédente, où je ne réalise qu'un seul tirage du dé : l'évènement « je tire 1 » est alors représenté par la partie $A_1 = \{(1, 1), (1, 2), \dots, (1, 6)\} \subset \Omega_2$. Pour décrire cet unique tirage, l'espace Ω_2 semble « trop gros », et moins naturel que Ω_1 . On verra plus loin qu'il est parfois utile d'utiliser un « gros » espace de probabilité.

2) Un expérience de tirage de dé nécessitant un espace de probabilité indénombrable

Je réalise maintenant l'expérience suivante : je tire mon dé jusqu'à ce que j'obtienne un 6 ; à quel moment ce 6 va-t-il tomber ? Quel pourrait être Ω ? A priori, le résultat de mon expérience est un nombre naturel (le moment n pour lequel mon 6 apparaît pour la première fois), ou éventuellement $+\infty$ (si, par immense malchance, je ne tire jamais le 6). Donc on pourrait chercher à décrire ce problème en prenant comme espace $\Omega = \mathbb{N}^* \cup \{\infty\}$, et en allouant une probabilité à chaque « temps » $n \in \Omega$. Cependant, pour calculer cette probabilité, il s'avère nécessaire d'utiliser comme espace de probabilité contenant l'information sur les résultats possibles de tous mes tirages (avant et après avoir obtenu un 6), donc d'une infinité de tirages ; on est obligé d'inclure une infinité de tirages, car on ne peut pas savoir, a priori, si on va tirer un 6 avant le temps $n = 1000$, ou avant $n = 10^{10}$, ou encore si on ne le tirera

jamais). Ainsi, l'espace de probabilité le plus naturel est en fait

$$\Omega = \{1, 2, \dots, 6\}^{\mathbb{N}^*},$$

autrement dit l'espace des suites infinies de chiffres

$$\Omega = \{\omega = \omega_1\omega_2\cdots, \omega_i \in \{1, 2, \dots, 6\}\}.$$

Cet ensemble est *indénombrable* : en effet, l'application $\varphi : \Omega \rightarrow [0, 1]$

$$\varphi(\omega) = \sum_{k=1}^{\infty} \frac{\omega_k - 1}{6^k} \quad (8.1)$$

fournit une surjection $\Omega \rightarrow [0, 1]$, qui est essentiellement la représentation d'un nombre réel $x \in [0, 1]$ en base 6. Quelle tribu est-il naturel d'imposer sur cet ensemble Ω ? Une expérience naturelle consiste à recenser les résultats des n premiers tirages ; chaque suite de résultats (i_1, i_2, \dots, i_n) définit donc un évènement

$$A_{i_1, \dots, i_n} = \{\omega \in \Omega ; \omega_1 = i_1, \omega_2 = i_2, \dots, \omega_n = i_n\} \subset \Omega.$$

Il est naturel de demander que cet évènement soit mesurable. On va alors considérer la tribu \mathcal{A} engendrée par ces évènements (avec $n \in \mathbb{N}$ quelconque). Cette tribu est aussi la plus petite qui rende l'application $\varphi : (\Omega, \mathcal{A}) \rightarrow ([0, 1], \mathcal{B}([0, 1]))$ mesurable. Si les tirages ne sont pas truqués, la probabilité de chacun de ces évènements de longueur n doit être la même, donc $\mathbb{P}(A_{i_1, \dots, i_n}) = \frac{1}{6^n}$. La classe définie par tous ces évènements est invariante par intersection ; le corollaire au Lemme de classe monotone nous montre alors que la probabilité \mathbb{P} satisfaisant ces conditions est unique.

Qu'en est-il de son existence ? L'application (8.1) envoie l'évènement A_{i_1, \dots, i_n} vers l'intervalle fermé

$$I_{i_1, \dots, i_n} = \left[x_{i_1, \dots, i_n}, x_{i_1, \dots, i_n} + \frac{1}{6^n} \right], \quad \text{avec} \quad x_{i_1, \dots, i_n} = \sum_{k=1}^n \frac{i_k - 1}{6^k}.$$

On remarque que $\frac{1}{6^n} = \lambda(I_{i_1, \dots, i_n})$. Autrement dit, si la mesure \mathbb{P} existe, on s'attend à ce que λ soit son image par l'application φ .

Peut-on au contraire construire \mathbb{P} à partir de λ ? Il faudrait pour cela inverser l'application φ , c'est-à-dire envoyer tout point $x \in [0, 1]$ vers une suite ω . On sait que φ n'est pas injective, mais elle est « presque » injective : les seuls points $x \in [0, 1]$ admettant 2 suites $\omega \neq \omega'$

telles que $x = \varphi(\omega) = \varphi(\omega') = x$ sont les points de la forme x_{i_1, \dots, i_n} , qui admettent comme antécédents $\omega = i_1 i_2 \dots i_n 111 \dots$ et (en supposant $i_n \neq 1$) $\omega' = i_1 i_2 \dots (i_n - 1) 6666 \dots$. Ces points forment un ensemble dénombrable $N \subset [0, 1]$, donc un ensemble négligeable pour λ . On peut donc définir un quasi-inverse de φ comme suit : si $x \notin N$, on choisit $s(x) = \omega$ son unique antécédent par φ ; si $x \in N$, on choisit le premier antécédent $\omega = i_1 i_2 \dots i_n 111 \dots$. Cette application $s : [0, 1] \rightarrow \Omega$ (qui est la représentation canonique du point $x \in [0, 1]$ en base 6) permet de définir la mesure image $s_* \lambda$ sur (Ω, \mathcal{A}) . On vérifie que cette mesure image $s_* \lambda$ associe à tout événement A_{i_1, i_2, \dots, i_n} la valeur $\frac{1}{6^n} = \lambda(I_{i_1, i_2, \dots, i_n})$. C'est donc la mesure de probabilité \mathbb{P} que nous cherchons. Elle est définie sur la tribu image de $\mathcal{B}([0, 1])$ par s .

3) Lois continues : Probabilité de présence d'une particule quantique

La mécanique quantique nous apprend qu'il est très difficile de « localiser » les objets microscopiques, comme les atomes ou les molécules. En général, à un instant t_0 donné, on ne peut pas dire « l'atome d'hydrogène que j'observe se trouve au point $x_0 \in \mathbb{R}^3$ ». Mais en travaillant bien, le physicien pourra déterminer une loi de probabilité de présence de la particule à travers l'espace, à cet instant t_0 : pour tout ouvert $U \subset \mathbb{R}^3$, l'atome aura une probabilité $\mathbb{P}_{t_0}(U)$ de se trouver dans l'ouvert U . La mesure de probabilité \mathbb{P}_{t_0} est alors définie sur la tribu borélienne $\mathcal{B}(\mathbb{R}^3)$, l'espace de probabilité étant $\Omega = \mathbb{R}^3$. En général la mesure \mathbb{P}_{t_0} est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^3 . Si la particule est dans un état stationnaire, sa probabilité de présence \mathbb{P}_{t_0} est indépendante du temps.

4) Trajectoires d'une particule soumise à forces aléatoires

En physique classique, on peut suivre le mouvement d'une particule au cours du temps. Mais pour cela il faut connaître précisément l'ensemble des forces auxquelles est soumise la particule. Or, si celle-ci est très petite (un grain de poussière) et se trouve dans un liquide à température ambiante, elle va être soumise à une multitude de forces dues aux molécules voisines du liquide. En raison du nombre gigantesque de molécules, on ne connaît pas précisément les forces exercées, on parle alors d'« agitation thermique », et on considère que la particule est soumise à une succession de forces aléatoires. La particule décrit alors une trajectoire continue, mais irrégulière, et qu'on peut observer (au microscope), mais pas prévoir. Si on regarde le mouvement d'une particule pendant une minute, on obtient une fonction continue $\omega : [0, 1] \rightarrow \mathbb{R}^3$. L'espace de probabilité naturel est alors l'espace fonction-

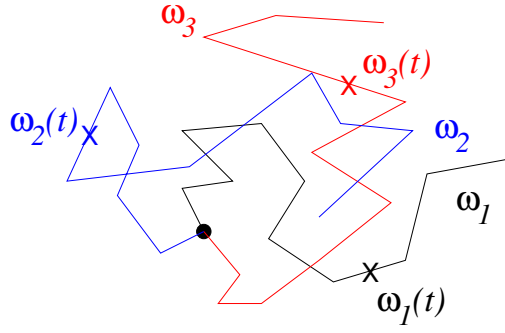


FIGURE 8.1 – 3 trajectoires aléatoires (chacune d’une couleur différente), et leurs positions au temps t

nel $\Omega = C([0, 1], \mathbb{R}^3)$ de toutes les « trajectoires continues » dans l’intervalle de temps $[0, 1]$.⁴ L’observateur va pouvoir détecter la position de la particule à chaque instant $t \in [0, 1]$. On veut donc que les événements du type « au temps t_0 la particule se trouve dans la boule $B_0 \in \mathbb{R}^3$ » soit mesurable. Autrement dit, la fonction « observation au temps t_0 » :

$$\omega \in \Omega \mapsto \omega(t_0) \in \mathbb{R}^3$$

doit être mesurable, et ce pour n’importe quel temps t_0 . On identifie donc la tribu minimale \mathcal{A} de Ω , qui rende toutes ces applications mesurables. A défaut de pouvoir prévoir les trajectoires individuelles, le physicien peut néanmoins décrire les propriétés statistiques (ou aléatoires) de ces trajectoires ; ces propriétés (qui sont l’objet de la physique statistique) dépendent de grandeurs physiques comme la température ou la pression. Mathématiquement, la physique statistique nous fournit ainsi une mesure de probabilité sur l’espace fonctionnel Ω . Nous ne décrirons pas plus en détail ce type de mesures ; la plus « simple » est sans doute la mesure de Wiener, qui décrit une classe de trajectoires appelées « mouvement brownien ».

8.2 Variables aléatoires

Définition 8.1. Soit (E, \mathcal{E}) un espace mesurable. Une application mesurable $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$ est appelée une *variable aléatoire* (en abrégé, v.a.) à valeurs dans E . Si $E = \mathbb{R}$, on parle de *variable aléatoire réelle*.

Typiquement, une v.a. est une fonction qu’on va pouvoir observer expérimentalement, donc l’espace image est souvent un espace euclidien, les grandeurs physiques étant généralement

4. Pour être plus réaliste, il faudrait supposer que le liquide, et donc la particule, sont confinés dans un récipient $D \subset \mathbb{R}^3$, donc il faudrait regarder $C([0, 1], D)$, mais ceci nous obligerait à analyser ce qui se passe lorsque la particule touche le bord du récipient D .

décrites par des coordonnées euclidiennes. L'image peut également être un espace discret, s'il s'agit de compter des objets, ou le nombre d'occurrences d'un certain phénomène. Par convention, les v.a. sont notées par la lettre X , ou des versions indicées X_1, X_2, \dots .

Revenons aux 4 exemples du paragraphe précédent.

Exemple. 1) Tirage de 2 dés : comme Ω_2 est discret, on peut considérer la variable aléatoire discrète $X((i, j)) = i + j$ qui donne la somme des deux tirages. Elle prend ses valeurs dans $\{2, \dots, 12\}$.

2) Tirage d'une infinité de dés. Le moment où tombe le premier 6, $X(\omega) = \inf \{n \in \mathbb{N}^*, \omega_n = 6\}$, avec $X(\omega) = \infty$ si ω ne contient aucun indice $\omega_j = 6$. Cette v.a. est définie dans $\overline{\mathbb{N}^*} = \mathbb{N}^* \cup \{\infty\}$. Cette fonction est-elle mesurable $\mathcal{A} \rightarrow \mathcal{P}(\overline{\mathbb{N}^*})$? Il suffit de vérifier que pour tout singleton $k \in \overline{\mathbb{N}^*}$, la préimage $X^{-1}(\{k\})$ est dans la tribu \mathcal{A} . Or

$$\forall k \in \mathbb{N}^*, \quad X^{-1}(\{k\}) = \{\omega \in \Omega; \omega_1 \neq 6, \omega_2 \neq 6, \dots, \omega_{k-1} \neq 6, \omega_k = 6\} = \bigsqcup_{i_1, i_2, \dots, i_{k-1}} A_{i_1 i_2 \dots i_{k-1} 6}, \quad (8.2)$$

il s'agit d'une union finie des événements $A_{i_1 i_2 \dots i_k}$, qui est donc dans \mathcal{A} . Il faut aussi vérifier que $X^{-1}(\{\infty\})$ est dans \mathcal{A} . En repassant à travers l'application φ , on observe que $\varphi(X^{-1}(\{\infty\}))$ est un ensemble de Cantor de $[0, 1]$, construit à partir des intervalles $I_{i_1 i_2 \dots i_k}$. A chaque niveau de construction n , l'ensemble K_n est donné par l'union des intervalles $I_{i_1 i_2 \dots i_n}$ tels que tous les $i_j \neq 6$. L'ensemble $K = \bigcap_{n \geq 1} K_n$ est obtenu après un ensemble dénombrable d'opérations, c'est donc un borélien. Son antécédent $X^{-1}(\{\infty\})$ par φ est donc dans la tribu \mathcal{A} .

3) Revenons au cas de la particule quantique dans \mathbb{R}^3 , en un temps t_0 . Comme $\omega \in \Omega$ est déjà une position, la variable aléatoire pertinente est donc l'identité : $\forall \omega \in \Omega, X(\omega) = \omega \in \mathbb{R}^3$. Cette v.a. est évidemment mesurable.

4) Lorsqu'on observe le mouvement d'une particule dans le liquide, on va également chercher à mesurer sa position en un temps t . On aura donc comme v.a. $X(\omega) = \omega(t) \in \mathbb{R}^3$, la position de la particule au temps t .

A chaque variable aléatoire on peut associer sa *loi*, qui est une information probabiliste essentielle.

Définition 8.2. La loi de la variable aléatoire X est la mesure image de \mathbb{P} par X . C'est la mesure de probabilité sur (E, \mathcal{E}) , que nous noterons P_X , définie par

$$P_X(B) = \mathbb{P}(X^{-1}(B)), \quad \forall B \in \mathcal{E}.$$

On écrira également :

$$P_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega; X(\omega) \in B\}).$$

La loi P_X est donc obtenue à partir de la mesure de probabilité \mathbb{P} , a priori elle contient moins d'informations que cette dernière. Cependant, la loi P_X suffit à déterminer la probabilité des évènements $A \in \mathcal{A}$ qui ne « dépendent que de X », c'est-à-dire des évènements de la forme $A = X^{-1}(B)$, avec $B \in \mathcal{E}$.

Comment interpréter cette loi P_X sur E , l'espace image de la v.a. X ? Chaque point $\omega \in \Omega$ détermine une valeur $X(\omega) \in E$; $P_X(B)$ est la probabilité que la valeur $X(\omega)$ tombe dans B . La loi P_X détermine la distribution des valeurs de X , par rapport à l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

Remarque 8.3. 1) Comme les v.a. X correspondent aux grandeurs observables expérimentalement, les lois P_X sont, en pratique, plus importantes que la loi globale \mathbb{P} sur Ω . L'expérimentateur peut mesurer (ou tester) les lois P_X , mais en général il n'a pas accès à la mesure globale \mathbb{P} sur Ω .

2) A partir d'une mesure de probabilité μ sur (E, \mathcal{E}) , il est aisé de construire une v.a. X ayant pour loi $P_X = \mu$. Il suffit de choisir $(\Omega, \mathcal{A}, \mathbb{P}) = (E, \mathcal{E}, \mu)$, et de définir la v.a. « tautologique » $X(\omega) = \omega$.

Différents types de variables aléatoires

8.2.1 Variables aléatoires discrètes

Si E est un ensemble dénombrable (avec $\mathcal{E} = \mathcal{P}(E)$), la loi de toute v.a. X à valeurs dans E peut s'écrire :

$$P_X = \sum_{x \in E} p_x \delta_x, \quad \text{avec } p_x = \mathbb{P}(X = x), \quad \text{et } \delta_x \text{ est le masse de Dirac en } x.$$

En effet, pour tout $B \subset E$,

$$P_X(B) = \mathbb{P}\left(\bigsqcup_{x \in B} \{X = x\}\right) = \sum_{x \in B} \mathbb{P}(X = x) = \sum_{x \in E} \mathbb{P}(X = x) \mathbb{1}_B(x) = \sum_{x \in E} \mathbb{P}(X = x) \delta_x(B),$$

où on a remarqué les identités $\mathbb{1}_B(x) = \mathbb{1}_{x \in B} = \delta_x(B)$.

Exemple. Dans l'exemple 2) ci-dessus, la v.a. $X(\omega) = \min \{j; \omega_j = 6\}$ prend ses valeurs dans $\overline{\mathbb{N}^*}$, elle est donc discrète. On a décrit plus haut la mesure \mathbb{P} sur Ω . On peut alors en extraire la loi P_X : pour tout $k \in \mathbb{N}^*$, l'évènement $X^{-1}(\{k\}) = \{X = k\}$ est décrit par la formule (8.2). Comme chaque évènement $A_{i_1 i_2 \dots i_{k-1} 6}$ a comme probabilité $\frac{1}{6^k}$, on trouve

$$\mathbb{P}(X = k) = \frac{5^{k-1}}{6^k} = \frac{1}{6} \left(\frac{5}{6}\right)^{k-1}, \quad \forall k \in \mathbb{N}^*.$$

On remarque que $\sum_{k \in \mathbb{N}^*} \mathbb{P}(X = k) = 1$, ce qui montre que $\mathbb{P}(X = \infty) = 0$. Ceci ne signifie nullement que l'ensemble $X^{-1}(\{\infty\})$ est vide, mais cet ensemble est négligeable pour \mathbb{P} . La discussion suivant (8.2) nous a montré que cet ensemble peut être mis en correspondance avec un ensemble de Cantor sur $[0, 1]$, qui a effectivement une mesure de Lebesgue nulle.

8.2.2 Variables aléatoires à densité

Définition 8.4. Une variable aléatoire X à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est appelée *variable aléatoire à densité* si sa loi P_X est absolument continue par rapport à la mesure de Lebesgue λ .

Le théorème de Radon-Nikodym nous montre alors qu'il existe une fonction borélienne $p : \mathbb{R}^d \rightarrow \mathbb{R}_+$ telle que

$$\forall B \in \mathcal{B}(\mathbb{R}^d), \quad P_X(B) = \int_B p(x) d\lambda(x).$$

En particulier, on a $\int_{\mathbb{R}^d} p(x) d\lambda(x) = 1$, ce qui montre que p est intégrable. La fonction p est unique dans $L^1(\mathbb{R}^d, \lambda)$ (donc unique, à un ensemble λ -négligeable près). On l'appelle la *densité* de la loi de X (ou, pour faire court, la densité de X).

En dimension $d = 1$, on a pour tout $\alpha \leq \beta$:

$$\mathbb{P}(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} p(x) dx.$$

8.3 Espérance (mathématique)

En théorie des probabilités, l'intégrale d'une variable aléatoire est appelée l'*espérance* de la variable aléatoire.

Définition 8.5. Soit X une v.a. réelle (c'est-à-dire à valeurs dans \mathbb{R}). On note alors

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) \mathbb{P}(d\omega),$$

une intégrale qui est bien définie si l'une des deux conditions ci-dessous est valide :

- 1) si $X \geq 0$, auquel cas $\mathbb{E}[X] \in [0, \infty]$;
- 2) si X est intégrable, ce qui signifie que $\mathbb{E}[|X|] = \int |X| d\mathbb{P} < \infty$.

On étend la définition aux v.a. vectorielles : si $X = (X_1, \dots, X_d)$ est à valeurs dans \mathbb{R}^d , alors $\mathbb{E}[X] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_d])$, qui a un sens si tous les $\mathbb{E}[X_i]$ sont définis.

Remarque 8.6. En choisissant comme v.a. une fonction caractéristique $X = \mathbb{1}_A$ pour une partie $A \subset \Omega$ mesurable, on trouve $\mathbb{E}[X] = \mathbb{P}(A)$, donc les espérances permettent de remonter à toute la mesure \mathbb{P} .

L'espérance d'une variable aléatoire s'interprète comme la *moyenne* de cette variable, pondérée par la mesure de probabilité \mathbb{P} . Par exemple, dans le cas de l'exemple 1) où $\Omega = \{1, 2, \dots, 6\}$ et $\mathbb{P}(i) = \frac{1}{6}$, on a pour toute v.a. $\mathbb{E}[X] = \frac{1}{6} \sum_{i=1}^6 X(i)$, qui est la moyenne des valeurs $X(i)$ au sens habituel.

Proposition 8.7. Soit X une v.a. à valeurs dans (E, \mathcal{E}) . Pour toute fonction mesurable $f : E \rightarrow \overline{\mathbb{R}}_+$, la v.a. $f(X)$ a comme espérance :

$$\mathbb{E}[f(X)] = \int_E f(x) P_X(dx).$$

Cette espérance est bien définie, puisque la v.a. $f(X)$ est positive.

Démonstration. Si on prend $f = \mathbb{1}_B$ pour un $B \in \mathcal{E}$, on obtient $\mathbb{E}[\mathbb{1}_B(X)] = \mathbb{P}(X^{-1}(B)) = P_X(B) = \int_E \mathbb{1}_B dP_X$. Par linéarité, le résultat est valable pour toute fonction f étagée positive sur E . Enfin, le théorème de convergence monotone permet de l'étendre aux limites de suites de fonctions étagées positives, donc aux fonctions mesurables positives. \square

Remarque 8.8. 1) Si la fonction f n'est pas de signe constant, la proposition reste vraie si l'intégrale est bien définie, autrement dit si $\mathbb{E}[|f(X)|] < \infty$ ($f(X)$ est intégrable pour \mathbb{P} , ou $f \in \mathcal{L}^1(E, \mathcal{E}, P_X)$).

1) Par rapport à la Définition 8.5, la formule ci-dessus est exprimée en termes de la loi P_X sur E , et non de la mesure de probabilité \mathbb{P} . Elle montre bien que seule la loi P_X est nécessaire pour comprendre les propriétés statistiques de v.a. de la forme $f(X)$ (on dit qu'une telle v.a.

est *totale*ment dépendante de la v.a. X). Inversement, si, pour toute fonction f « suffisamment générale », on peut écrire

$$\mathbb{E}[f(X)] = \int f d\nu$$

pour une certaine mesure de probabilité ν sur \mathbb{R} , alors on pourra dire que $\nu = P_X$.

La proposition qui suit donne un exemple de ce principe. Elle permet d'extraire les lois de v.a. individuelles réelles X_j , à partir de la loi globale de la v.a. vectorielle (X_1, \dots, X_d) , en supposant cette dernière à densité.

Proposition 8.9. *Soit $X = (X_1, \dots, X_d)$ une v.a. à valeurs dans \mathbb{R}^d , et on la suppose à densité $p(x_1, \dots, x_d)$. Alors, pour tout $j = 1, \dots, d$, la loi de X_j est aussi à densité, donnée par*

$$p_j(x) = \int_{\mathbb{R}^{d-1}} p(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_d) dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_d.$$

Autrement dit, p_j est obtenu en opérant les intégrales partielles de p le long des autres variables. Les lois $P_{X_j} = p_j dx$ sont appelées les lois marginales de X .

Démonstration. Soit π_j la projection de \mathbb{R}^d sur la j -ième composante. Pour toute fonction borélienne $f : \mathbb{R} \rightarrow \mathbb{R}_+$, le théorème de Fubini-Tonnelli nous permet d'écrire :

$$\begin{aligned} \mathbb{E}[f(X_j)] &= \mathbb{E}[f \circ \pi_j(X)] = \int_{\mathbb{R}^d} f(x_j) p(x_1, \dots, x_d) dx_1 \cdots dx_d \\ &= \int_{\mathbb{R}} f(x_j) \left(\int_{\mathbb{R}^{d-1}} p(x_1, \dots, x_d) dx_1 \cdots dx_{j-1} dx_{j+1} \cdots dx_d \right) dx_j \\ &= \int_{\mathbb{R}} f(x_j) p_j(x_j) dx_j, \end{aligned}$$

le théorème nous disant aussi que la fonction $p_j(x)$ est définie Lebesgue-p.p. et est Lebesgue-intégrable. \square

Plus généralement, comme $X_j = \pi_j(X)$, la loi de X_j est déterminée par la loi de X : $P_{X_j} = \pi_{j*} P_X$. La loi P_X « contient » donc toutes les lois P_{X_j} .

Mais la réciproque est fautive : connaître les lois individuelles de X_1, X_2, \dots, X_d n'est pas suffisant pour reconstituer la loi de $X = (X_1, \dots, X_d)$. Par exemple, considérons une v.a. vectorielle (X_1, X_2) ayant une loi de densité $p(x, y)$ à symétrie circulaire : $p(x, y) = \tilde{p}(\sqrt{x^2 + y^2})$. Les lois marginales de X_1 et X_2 sont alors identiques, égales à $p_1(x) dx$, où $p_1(x) = \int p(x, y) dy$. Considérons maintenant la v.a. vectorielle $X' = (X'_1, X'_2)$ supportée sur la diagonale $\{x = y\}$, ayant pour loi $p_1(x) \delta_{x=y}$. X' n'est plus une v.a. à densité, mais les

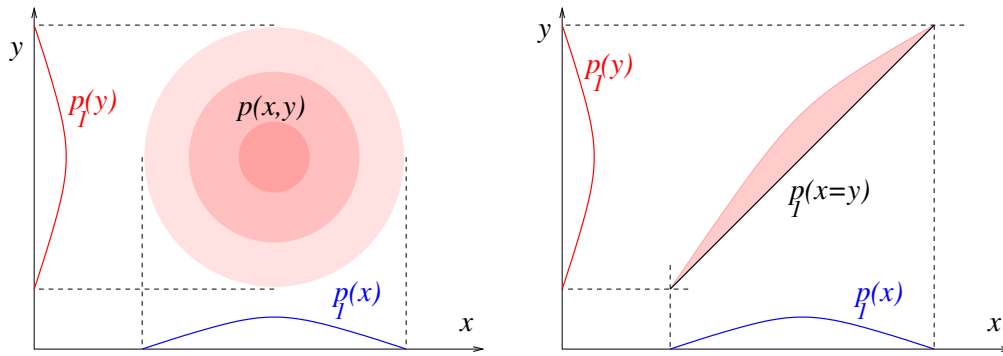


FIGURE 8.2 – Deux v.a. vectorielles admettant les mêmes marginales

variables X'_1 et X'_2 admettent toutes deux comme loi $\pi_{1*}P_{X'} = p_1(x) dx$, comme dans le cas précédent. On voit donc que deux v.a. vectorielles très différentes peuvent admettre les mêmes marginales.

8.3.1 Une expérience aléatoire géométrique : le paradoxe de Bertrand

On propose d'étudier un exemple de problème aléatoire géométrique, qui va illustrer l'importance de bien définir les termes de l'expérience aléatoire qu'on considère. On se pose la question suivante : si dans le cercle unité S^1 on choisit une corde au hasard (un segment reliant deux points du cercle) ; quelle est la probabilité que la longueur de la corde soit supérieure au côté d'un triangle équilatéral inscrit dans S^1 ? (un peu de trigonométrie montre que cette longueur vaut $\sqrt{3}$). Pour répondre à cette question, il faut d'abord comprendre ce que signifie « choisir une corde au hasard ». Bertrand a proposé deux façons de procéder à ce choix.

(a) On choisit au hasard, et indépendamment l'un de l'autre, les deux extrémités A, B de la corde. Une fois qu'on a choisi A , on peut paramétrer le point B par la valeur de l'angle \widehat{AOB} , qu'on choisit distribué uniformément dans $[0, 2\pi[$. On voit que la corde sera plus longue que $\sqrt{3}$ ssi l'angle \widehat{AOB} est dans l'intervalle $]2\pi/3, 4\pi/3[$, ce qui correspond à $1/3$ de la longueur du cercle. Donc la probabilité cherchée vaut $1/3$.

(b) Une corde peut également être paramétrée par la position de son centre, qui doit se trouver quelque part dans le disque unité ouvert. Choisir une corde aléatoirement peut donc aussi être effectué en choisissant au hasard son centre C dans le disque unité. On remarque que les centres des côtés du triangle équilatéral sont à distance $1/2$ du centre du cercle. Par conséquent, la corde aura une longueur supérieure à $\sqrt{3}$ ssi son centre est à distance $|OC| < 1/2$ de l'origine. En comparant les surfaces du disque $D(0, 1/2)$ et du disque unité,

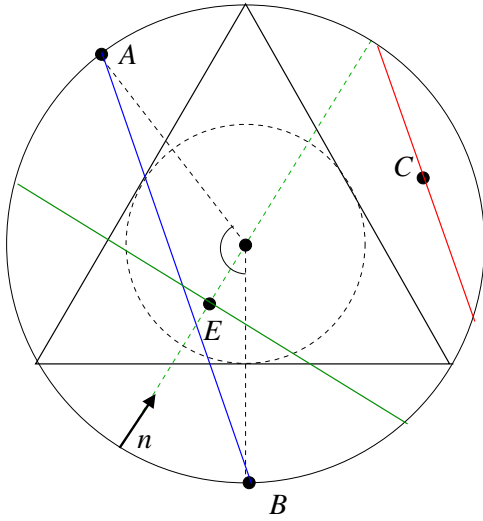


FIGURE 8.3 – Trois façons de choisir une corde aléatoire

on voit que la probabilité que la corde soit de longueur $> \sqrt{3}$ vaut $\frac{1}{4}$. ????

Pourquoi les deux modes de calcul ne donnent-ils pas le même résultat ?

Parce que dans les deux cas, l'expérience aléatoire, l'espace de probabilité, et donc la variable aléatoire, ne sont pas les identiques. Même si la quantité physique qu'on cherche à mesurer est la même (la longueur de la corde), on va voir que la loi de la v.a. correspondante n'est pas la même dans les deux situations.

(a) On choisit au hasard les deux extrémités de la corde. On peut repérer ces deux points au moyen d'angles $\theta, \theta' \in [0, 2\pi[$, donc l'espace de probabilité naturel est $\Omega_1 = [0, 2\pi[^2$: chaque choix de corde est déterminé par $\omega = (\theta, \theta')$. La mesure naturelle sur Ω est la mesure de Lebesgue normalisée : $\mathbb{P}(d\omega) = \frac{1}{(2\pi)^2} d\theta d\theta'$. En fonction de ces angles, la longueur de la corde $X_1(\omega) = 2 \sin\left(\frac{\theta - \theta'}{2}\right)$. On peut alors calculer la loi de X_1 , en étudiant l'espérance de n'importe

quelle fonction de X_1 :

$$\begin{aligned}
 \mathbb{E}[f(X_1)] &= \int_{\Omega_1} f(X_1(w)) \mathbb{P}(d\omega) \\
 &= \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} f\left(2 \sin\left(\frac{\theta - \theta'}{2}\right)\right) d\theta d\theta' \\
 &= \frac{1}{2\pi} \int_0^{2\pi} f\left(2 \sin\left(\frac{u}{2}\right)\right) du \\
 &= \frac{1}{\pi} \int_0^\pi f\left(2 \sin\left(\frac{u}{2}\right)\right) du \\
 &= \frac{1}{\pi} \int_0^2 f(x) \frac{1}{\sqrt{1-x^2/4}} dx,
 \end{aligned}$$

où on a effectué les changements de variables $(\theta, \theta') \mapsto (u = \theta - \theta', \theta')$ de jacobien unité, puis $x = 2 \sin(u/2)$ de jacobien $\frac{\partial x}{\partial u} = \cos(u/2) = \sqrt{1-x^2/4}$, donc $\frac{\partial u}{\partial x} = \frac{1}{\sqrt{1-x^2/4}}$. La dernière équation montre que la loi de X_1 est donnée par

$$P_{X_1} = p_1(x) dx, \quad \text{avec} \quad p_1(x) = \frac{1}{\pi \sqrt{1-x^2/4}} \mathbb{1}_{[0,2]}(x) dx.$$

On retrouve le fait que $\mathbb{P}(X_1 \geq \sqrt{3}) = \int_{\sqrt{3}}^2 p(x) dx = \frac{1}{3}$.

(b) Comme une corde peut être déterminée par la position de son milieu, quelque part dans le disque unité, on choisit comme espace de probabilité

$$\Omega_2 = D(0, 1) = \{\omega = (y, z) \in \mathbb{R}^2, y^2 + z^2 < 1\},$$

muni de la mesure de Lebesgue normalisée $\mathbb{P}(d\omega) = \frac{1}{\pi} \mathbb{1}_{\Omega_2}(y, z) dy dz$. La longueur de la corde est donnée par $X_2(\omega) = 2\sqrt{1-y^2-z^2}$. On calcule la loi de X_2 de la même façon :

$$\begin{aligned}
 \mathbb{E}[f(X_2)] &= \frac{1}{\pi} \int_{D(0,1)} f\left(2\sqrt{1-y^2-z^2}\right) dy dz \\
 &= 2 \int_0^1 f\left(2\sqrt{1-r^2}\right) r dr \\
 &= \frac{1}{2} \int_0^2 f(x) x dx,
 \end{aligned}$$

en passant d'abord en coordonnées polaires, puis en prenant $x = 2\sqrt{1-r^2}$, de jacobien

$\frac{\partial x}{\partial r} = \frac{-2r}{\sqrt{1-r^2}} = \frac{-4r}{x}$, donc $\frac{\partial r}{\partial x} = -\frac{x}{4r}$. On a donc abouti à la loi

$$P_{X_2} = p_2(x) dx, \quad \text{avec} \quad p_2(x) = \frac{x}{2} \mathbb{1}_{[0,2]}(x) dx.$$

On retrouve bien $\mathbb{P}(X_2 \geq \sqrt{3}) = \int_0^2 p_2(x) dx = \frac{1}{4}$. On observe que les deux densités p_1, p_2 ont des comportements différents aux extrémités de l'intervalle $[0, 2]$: p_1 diverge lorsque $x \nearrow 2$, alors que p_2 reste bornée. Cela explique que P_{X_1} « favorise » les longues cordes par rapport à P_{X_2} , ce qui pourrait expliquer pourquoi $\mathbb{E}[X_1] > \mathbb{E}[X_2]$. D'un autre côté, on remarque aussi que $p_2(x) \rightarrow 0$ lorsque $x \searrow 0$, tandis que $p_1(x)$ reste positif.

Exercice 8.10. Etudier la 3e manière de choisir une corde au hasard : en choisissant au hasard la direction du rayon orthogonal à la corde, puis au hasard le centre de la corde uniformément le long de ce rayon.

8.3.2 Lois classiques de probabilité

Vous connaissez déjà probablement les lois discrètes suivantes.

Lois discrètes

(a) *Loi uniforme* sur un ensemble fini. Si E est un ensemble de cardinal $n \in \mathbb{N}^*$, une v.a. X est de loi uniforme sur E vérifie

$$\mathbb{P}(X = x) = \frac{1}{n}, \quad \forall x \in E.$$

Le cas standard correspond à prendre $E = \{1, 2, \dots, n\}$.

(b) *Loi de Benoulli* de paramètre $p \in [0, 1]$. C'est la loi d'une v.a. à valeurs dans $\{0, 1\}$, telle que :

$$\mathbb{P}(X = 1) = p, \text{ et donc } \mathbb{P}(X = 0) = 1 - p.$$

On retrouve une loi uniforme lorsque $p = 1/2$, ce qui correspond à la loi lorsqu'on tire une pièce à pile ou face. Si la pièce est truquée, on peut obtenir une loi avec une probabilité $p \neq 1/2$ d'obtenir « pile ».

(c) *Loi binomiale* $\mathcal{B}(n, p)$, pour $n \in \mathbb{N}^*$, $p \in [0, 1]$. C'est la loi d'une v.a. à valeurs dans $\{0, 1, 2, \dots, n\}$, telle que

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \leq k \leq n.$$

Cette loi est induite par la loi de Bernoulli : si on tire n fois une pièce de loi de Bernoulli p , alors $\mathbb{P}(X = k)$ est la probabilité qu'on ait obtenu k fois « pile ».

(d) *Loi géométrique* de paramètre $p \in]0, 1[$. C'est la loi d'une v.a. X à valeurs dans \mathbb{N} , telle que

$$\mathbb{P}(X = k) = (1 - p)p^k, \quad k \in \mathbb{N}.$$

Toujours dans le contexte d'une suites de tirage à pile ou face de loi de Bernoulli de paramètre p , cette loi décrit le nombre de piles obtenus avant le premier « face ». C'est une variation sur l'exemple 2. étudié au début du chapitre.

(e) Loi de Poisson de paramètre $\lambda > 0$. C'est encore une v.a. à valeurs dans \mathbb{N} , mais de probabilités

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}.$$

Cette loi décrit les occurrences d'évènements rares pendant un période longue (ou un grand nombre d'expériences répétées). On pourra l'obtenir à partir d'une suite de v.a. X_n de lois binomiales $\mathcal{B}(n, p_n = \frac{\lambda}{n})$:

$$\lim_{n \rightarrow \infty} P(X_n = k) = \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Tirer n fois la pièce, avec une probabilité λ/n d'obtenir « pile » à chaque fois, induit le fait que le nombre de « pile » va typiquement rester borné même lorsque $n \gg 1$: on fait beaucoup de tirages, mais chaque tirage a une probabilité très faible.

Lois continues

Venons-en maintenant aux v.a. continues. Dans les exemples ci-dessous, la v.a. X prend ses valeurs dans \mathbb{R} , et admet une loi à densité.

(a) *Loi uniforme* sur un intervalle borné $[a, b]$: la densité vaut $p(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$.

(b) *Loi exponentielle* de paramètre $\lambda > 0$:

$$p(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x).$$

Ces lois héritent d'une propriété intéressante de l'exponentielle : pour tous $a, b > 0$:

$$\mathbb{P}(X > a + b) = \mathbb{P}(X > a) \mathbb{P}(X > b).$$

Cette propriété s'interprète de la façon suivante : pour que la v.a. X dépasse $a + b$, il faut d'abord qu'elle dépasse a , ce qu'elle fait avec une probabilité $\mathbb{P}(X > a)$; puis, une fois en a , il lui faut encore franchir une distance b pour arriver en $a + b$. La probabilité d'aller de $a \rightarrow a + b$ est identique à celle d'aller de $0 \rightarrow b$, qui est donnée par $\mathbb{P}(b)$. Une fois arrivé en a , le système ne se souvient pas de son passé, il recommence à avancer comme s'il était au début de l'expérience ; aussi, il se comporte comme s'il était en 0 et devait parcourir une distance b . Cette absence de mémoire implique que la probabilité totale est donnée par le produit $\mathbb{P}(X > a) \mathbb{P}(X > b)$. Ce produit de deux probabilités est un signe de *l'indépendance* de ces deux étapes aléatoires (v. le chapitre suivant). Cette loi exponentielle décrit par exemple le temps de vie d'une particule radioactive ; l'espérance $\mathbb{E}[X] = \lambda^{-1}$ est appelé le temps de demi-vie de la particule.

(c) Loi gaussienne, ou loi normale, $\mathcal{N}(m, \sigma^2)$, avec $m \in \mathbb{R}$ le centre de la gaussienne, σ^2 sa variance. C'est une v.a. distribuée sur tout \mathbb{R} , de densité

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Sa densité est donc une fonction gaussienne, de centre m et de largeur σ . Les paramètres m, σ s'obtiennent par :

$$m = \mathbb{E}[X], \quad \sigma^2 = \mathbb{E}[(X - m)^2].$$

La variable translatée $X - m$ suit la loi centrée $\mathcal{N}(0, \sigma^2)$, tandis que la variable $\frac{X-m}{\sigma}$ suit la loi centrée réduite $\mathcal{N}(0, 1)$. Plus généralement, si X suit la loi $\mathcal{N}(m, \sigma^2)$, alors pour tous $\lambda, \mu \in \mathbb{R}$, la v.a. $\lambda X + \mu$ suit la loi $\mathcal{N}(\lambda m + \mu, \lambda^2 \sigma^2)$.

Avec la loi de Poisson, la loi normale est la plus importante en théorie des probabilités, et en physique en général. Elle apparaît par exemple comme loi limite, lorsqu'on prend la somme d'un grand nombre de v.a. indépendantes identiquement distribuées (c'est le contenu du Théorème Central Limite, un des théorèmes importants de la théorie des probabilités).

8.3.3 Fonction de répartition d'une v.a. réelle

On rappelle la définition donnée dans l'Exemple 2.18, dans le cas particulier d'une mesure de probabilité.

Si X est une v.a. réelle, on définit sa fonction de répartition par la fonction $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par :

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

Cette fonction est croissante, continue à droite, et converge vers 0 en $t \rightarrow -\infty$, et vers 1 en $t \rightarrow \infty$.

Le Théorème 4.23 nous a appris que, pour toute fonction F ayant ces propriétés, il lui correspond une unique mesure de probabilité μ sur \mathbb{R} , telle que $F(t) = \mu(]-\infty, t])$ pour tout $t \in \mathbb{R}$. On peut donc interpréter F comme la fonction de répartition d'une v.a. réelle X , de loi μ .

Le Théorème 4.23 nous a donc montré que la fonction F_X caractérise la loi P_X de la v.a. X , en particulier on a :

$$\begin{aligned}\mathbb{P}(a \leq X \leq b) &= F_X(b) - F_X(a-), & \text{si } a \leq b \\ \mathbb{P}(a < X < b) &= F_X(b-) - F_X(a), & \text{si } a < b.\end{aligned}$$

Les points de discontinuité de F_X correspondent aux atomes de la loi P_X .

8.3.4 Tribu engendrée par une v.a.

L'espace de probabilité Ω est équipé d'une tribu \mathcal{A} . Pour un problème donné, on peut se poser la question de la « meilleure tribu » adaptée au problème. Dans la Proposition 2.26 on a introduit une tribu engendrée par une famille de fonctions. Restreignons-nous maintenant à une unique v.a. X , à valeur dans un espace mesurable (E, \mathcal{E}) . La tribu engendrée par X , qu'on notera $\sigma(X)$, est la plus petite tribu sur Ω qui rende X mesurable. On vérifie facilement qu'elle est donnée par les éléments suivants :

$$\sigma(X) = \{A = X^{-1}(B) : B \in \mathcal{E}\}.$$

Remarque 8.11. Si on se donne une famille quelconque de v.a. $(X_i)_{i \in I}$, avec chaque X_i une v.a. à valeurs dans (E_i, \mathcal{E}_i) , alors on peut définir la tribu engendrée par :

$$\sigma((X_i)_{i \in I}) = \sigma(\{X_i^{-1}(B) ; B \in \mathcal{E}_i, i \in I\}).$$

La proposition suivante donne un critère pour qu'une v.a. Y soit totalement dépendante d'une autre v.a. X .

Proposition 8.12. *Soit X une v.a. à valeurs dans (E, \mathcal{E}) , et soit Y une v.a. réelle. Il y a équivalence entre les deux énoncés suivants :*

i) la v.a. Y est $\sigma(X)$ -mesurable (autrement dit, $\sigma(Y)$ est une sous-tribu de $\sigma(X)$);

ii) il existe une fonction mesurable $f : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, telle que $Y = f(X)$.

On résume par le schéma suivant les différentes fonctions :

$$\begin{array}{ccc} \Omega & \xrightarrow{X} & E \\ & \searrow Y & \downarrow f \\ & & \mathbb{R} \end{array}$$

Démonstration. L'implication ii) \rightarrow i) est évidente, puisque, pour tout borélien $A \in \mathbb{R}$, $f^{-1}(A)$ dans \mathcal{E} car f est mesurable, de sorte que $X^{-1}(f^{-1}(A)) = Y^{-1}(A) \in \sigma(X)$.

Pour montrer i) \rightarrow ii), supposons que Y est $\sigma(X)$ -mesurable. Commençons par le cas d'une fonction Y étagée, de représentation canonique :

$$Y = \sum_{i=1}^n \lambda_i \mathbb{1}_{A_i}.$$

La mesurabilité de Y nous dit que $A_i \in \sigma(X)$ pour tout $i = 1, \dots, n$. Pour chaque indice i , il existe donc un $B_i \in \mathcal{E}$ tel que $A_i = X^{-1}(B_i)$, ou de façon équivalente, $\mathbb{1}_{A_i} = \mathbb{1}_{B_i} \circ X$. On remarque que les B_i sont forcément disjoints, car les A_i le sont. En sommant sur les i , on obtient la représentation de Y :

$$Y = \sum_{i=1}^n \lambda_i \mathbb{1}_{B_i} \circ X = f \circ X,$$

avec la fonction $f = \sum_{i=1}^n \lambda_i \mathbb{1}_{B_i} : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable (car chaque $B_i \in \mathcal{E}$).

Soit maintenant Y une v.a. quelconque, qu'on suppose $\sigma(X)$ -mesurable. En appliquant le Théorème 3.10 aux parties positive et négative de Y , on voit que Y est la limite simple d'une suite de v.a. Y_n étagées et $\sigma(X)$ -mesurables. D'après l'étape ci-dessus, chacune de ces v.a. Y_n peut s'écrire $Y_n = f_n(X)$, pour une certaine fonction mesurable $f_n : E \rightarrow \mathbb{R}$. La convergence simple des Y_n nous montre que $f_n(x)$ a une limite lorsque $n \rightarrow \infty$ dès que x est dans l'image de X . On définit alors la fonction f comme suit :

$$f(x) = \begin{cases} \lim_{n \rightarrow \infty} f_n(x), & \text{si la limite existe,} \\ 0, & \text{dans le cas contraire.} \end{cases}$$

On a déjà vu que cette fonction est $\sigma(X)$ -mesurable, parce que toutes les f_n le sont. Le premier cas concerne au moins tous les x dans l'image de X , donc pour tout $\omega \in \Omega$, $f(X(\omega)) =$

$$\lim_n f_n(X(\omega)) = \lim_n Y_n(\omega) = Y(\omega).$$

□

8.4 Moments d'une v.a.

Définition 8.13. Soit X une v.a. réelle, et soit $p \in \mathbb{N}^*$. Le moment d'ordre p de X est la quantité $\mathbb{E}[X^p]$. Comme on l'a vu dans le cas de l'espérance, cette quantité n'est définie que si X est une v.a. positive, ou que $\mathbb{E}[|X|^p] < \infty$.

On remarque que le moment d'ordre 1 n'est autre que l'espérance de X . On dit qu'une v.a. réelle est *centrée* si elle est intégrable et si $\mathbb{E}[X] = 0$.

Quelques rappels du cours d'intégration, adaptés à l'espérance

Comme l'espérance d'une v.a. réelle n'est autre que l'intégrale d'une fonction mesurable par rapport à une mesure finie, on rappelle les propriétés suivantes :

Convergence monotone : $X_n \geq 0$, $X_n \uparrow X \implies \mathbb{E}[X_n] \uparrow \mathbb{E}[X]$

Lemme de Fatou : $X_n \geq 0 \implies \mathbb{E}[\liminf_n X_n] \leq \liminf_n \mathbb{E}[X_n]$

Convergence dominée : $|X_n| \leq Z$ et $\mathbb{E}[Z] < \infty$, $X_n \rightarrow X$ p.p. $\implies \mathbb{E}[X_n] \rightarrow \mathbb{E}[X]$.

En probabilités, l'expression « presque partout » est remplacée par « presque sûrement », qu'on abrège en p.s.

Les espaces $L^p(\Omega, \mathcal{A}, \mathbb{P})$ sont définis comme dans la section 5 du cours d'intégration. L'inégalité de Hölder s'écrit :

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{1/p} \mathbb{E}[|Y|^q]^{1/q},$$

avec $p, q \in [1, \infty]$ une paire d'exposants conjugués. Comme on est sur une mesure de probabilité, on trouve $\|X\|_1 \leq \|X\|_p$ pour tout $p \geq 1$, ainsi que $\|X\|_r \leq \|X\|_p$ pour tout $r \leq p$. Le cas particulier de l'inégalité de Cauchy-Schwartz,

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^2]^{1/2} \mathbb{E}[|Y|^2]^{1/2}, \quad \text{donne aussi l'inégalité} \quad \mathbb{E}[|X|]^2 \leq \mathbb{E}[X^2].$$

Après la définition de l'espérance d'une v.a. (qui existe si $X \in L^1$), voici celui de la variance.

8.4.1 Variance et matrice de covariance

Définition 8.14. Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. La variance de la v.a. X est définie par :

$$\text{var}(X) = \mathbb{E} [(X - \mathbb{E}[X])^2],$$

et son écart-type par $\sigma_X = \sqrt{\text{var}(X)}$.

Intuitivement, l'écart-type d'une v.a. X (lorsqu'il est fini) mesure la dispersion de la v.a. autour de sa moyenne $\mathbb{E}[X]$. En particulier, si la variance $\text{var}(X) = 0$, alors $X = \mathbb{E}[X]$ presque sûrement.

Dans le cas d'une loi normale $\mathcal{N}(m, \sigma)$, le paramètre σ est égal à l'écart-type σ_X , qui représente (à un facteur près) la « largeur à mi-hauteur » de la courbe gaussienne :

$$\frac{p(x)}{p_{\max}} \leq e^{-1/2} \iff |x| \leq \sigma, \quad \text{ou bien} \quad \frac{p(x)}{p_{\max}} \leq \frac{1}{2} \iff |x| \leq \sigma\sqrt{2 \ln 2}.$$

On montre les caractérisations suivantes de la variance :

Proposition 8.15. Soit $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. La variance peut aussi s'exprimer par $\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$. Pour tout $a \in \mathbb{R}$,

$$\mathbb{E} [(X - a)^2] = \text{var}(X) + (\mathbb{E}[X] - a)^2,$$

d'où on tire l'expression variationnelle :

$$\text{var}(X) = \inf_{a \in \mathbb{R}} \mathbb{E} [(X - a)^2],$$

dont le minimum est atteint en $a = \mathbb{E}[X]$.

Démonstration. On développe simplement

$$\begin{aligned} \mathbb{E} [(X - a)^2] &= \mathbb{E} [X^2 - 2aX + a^2] = \mathbb{E}[X^2] - 2a\mathbb{E}[X] + a^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 + (\mathbb{E}[X] - a)^2 \\ &= \text{var}(X) + (\mathbb{E}[X] - a)^2. \end{aligned}$$

Pour obtenir la première identité, il suffit de prendre $a = 0$. □

En adaptant la Proposition 3.15 aux notations probabilistes, on obtient

l’Inégalité de Markov :

Pour une v.a. $X \geq 0$ et tout $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{1}{a} \mathbb{E}[X].$$

Inégalité de Bienaymé-Tchebicheff (ou Bienaymé-Chebychev) :

Pour une v.a. $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ et tout $a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{1}{a^2} \text{var}(X).$$

Cette inégalité résulte de l’inégalité de Markov appliquée la v.a. $Y = (X - \mathbb{E}[X])^2$.

Comme il est souvent plus facile de calculer la moyenne ou la variance d’une v.a., ces deux inégalités permettent d’obtenir des bornes supérieures sur les probabilités des « grandes valeurs » de X .

Covariance entre plusieurs v.a.

La variance d’une v.a. réelle donne une indication sur la dispersion (ou la « largeur ») de la loi P_X . Lorsqu’on a plusieurs v.a. réelles sur un même espace de probabilité, on peut chercher à comparer celles-ci, ou à estimer leurs « dispersions relatives ». Pour cela on dispose d’un indicateur généralisant la variance, qu’on appelle la covariance.

Définition 8.16. Soient $X, Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. La *covariance* de X et Y est définie par :

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[X(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Plus généralement, pour une v.a. vectorielle $X = (X_1, \dots, X_d)$ à valeurs dans \mathbb{R}^d , dont toutes les composantes $X_i \in L^2(\Omega, \mathbb{P})$, la matrice de covariance de X est donnée par

$$K_X := (\text{cov}(X_i X_j))_{1 \leq i, j \leq d}.$$

On remarque les propriétés suivantes :

1) si $X = Y$, la covariance $\text{cov}(X, X) = \text{var}(X)$, qui prend des valeurs positives ou nulles. Par contre, pour deux v.a. distinctes, la covariance $\text{cov}(X, Y)$ peut prendre des valeurs négatives.

Par exemple, si $Y = -X$, on aura $\text{cov}(X, -X) = -\text{var}(X) \leq 0$. L'inégalité de Cauchy-Schwarz implique une inégalité entre covariance de la paire (X, Y) et variances des variables séparées :

$$|\text{cov}(X, Y)| \leq \sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}.$$

3) L'application $(X, Y) \mapsto \text{cov}(X, Y)$ est une forme bilinéaire sur $L^2(\Omega, \mathbb{P})$.

4) Dans le cas vectoriel $X = (X_1, \dots, X_d)$, la matrice de covariance est symétrique positive : pour tout $\lambda = (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$, on a :

$$\langle \lambda, K_X \lambda \rangle = \sum_{i,j=1}^d \lambda_i \lambda_j K_X(i, j) = \text{var} \left(\sum_{i=1}^d \lambda_i X_i \right) \geq 0.$$

Exercice 8.17. Si A est une matrice $n \times d$ et $Y = AX$, vérifier que la matrice de covariance de Y vaut $K_Y = AK_X^t A$.

Intuitivement, la covariance est une mesure de la dépendance, ou la *corrélation*, entre les variables X et Y . Lorsque $Y \approx X$, les variables (X, Y) sont « positivement corrélées », ce qui se traduit par $\text{cov}(X, Y) \geq 0$. Au contraire, si $Y \approx -X$, les variables sont « négativement corrélées ». On verra dans le prochain chapitre que si les deux variables sont indépendantes, alors $\text{cov}(X, Y) = 0$ (**mais la réciproque est fausse !**)

8.4.2 Régression linéaire

On considère dans cette section des v.a. réelles dans $L^2(\Omega, \mathcal{A}, \mathbb{P})$.

On se donne une famille de v.a. Y_1, \dots, Y_n , qu'on voit comme des « v.a. de référence ». Si on se donne ensuite une autre v.a. X , on cherche à approximer X par une combinaison affine des Y_1, \dots, Y_n , donc par une v.a. de la forme :

$$\beta_0 + \beta_1 Y_1 + \dots + \beta_n Y_n, \quad \text{avec des coefficients } \beta_i \in \mathbb{R}.$$

Pour optimiser la « ressemblance » entre X et une telle combinaison, on va chercher à minimiser l'espérance :

$$\mathbb{E} [(X - (\beta_0 + \beta_1 Y_1 + \dots + \beta_n Y_n))^2],$$

sur tous les choix du $(n+1)$ -uplet $(\beta_i)_{i=0,\dots,n}$. Comme toutes les v.a. sont dans L^2 , cette espérance est finie, quel que soit (β_i) . Ce problème de minimisation est appelé une *régression linéaire*.

Proposition 8.18. *Ce problème de minimisation admet une unique solution :*

$$\inf_{(\beta_i) \in \mathbb{R}^{n+1}} \mathbb{E} [(X - (\beta_0 + \beta_1 Y_1 + \cdots + \beta_n Y_n))^2] = \mathbb{E} [(X - Z)^2],$$

où la v.a. Z minimisante est donnée par

$$Z = \mathbb{E}[X] + \sum_{i=1}^n \alpha_i (Y_i - \mathbb{E}[Y_i]),$$

les coefficients $(\alpha_i)_{i=1, \dots, n}$ étant une solution (n'importe laquelle) du système linéaire

$$\sum_{i=1}^n \alpha_i \text{cov}(Y_i, Y_k) = \text{cov}(X, Y_k), \quad 1 \leq k \leq n.$$

En particulier, si la matrice K_Y est inversible, on a une unique solution $\alpha = \text{cov}(X, Y) K_Y^{-1}$, où α et $\text{cov}(X, Y)$ sont notés comme des vecteurs ligne.

Démonstration. On ajoute au n -uplet de variables aléatoires la fonction constante 1, pour former un $(n+1)$ -uplet $(1, Y_1, \dots, Y_n)$ de v.a. On appelle H le sous-espace vectoriel de $L^2(\Omega, \mathbb{P})$ engendré par ce $(n+1)$ -uplet. Remarquons que H est aussi engendré par le $(n+1)$ -uplet formé par la fonction constante 1 et les v.a. centrées $\tilde{Y}_i = Y_i - \mathbb{E}[Y_i]$, $i = 1, \dots, n$. On remarque que la fonction 1 est orthogonale aux \tilde{Y}_i , de sorte que $H = \mathbb{R}1 \oplus \text{Vect}\{(\tilde{Y}_j)_{j=1, \dots, n}\}$. On cherche alors à minimiser

$$\mathbb{E} [(X - V)^2] = \int_{\Omega} |X - V|^2 d\mathbb{P} = \|X - V\|_2^2,$$

sur tous les $V \in H$. Un tel problème de minimisation est bien connu : le minimum est la distance euclidienne entre X et le sous-espace H , et il est atteint par la projection orthogonale de X sur H , donnée par un unique vecteur $Z \in H$. Cette projection $Z \in H$ peut se décomposer sous la forme :

$$Z = \alpha_0 + \sum_{i=1}^n \alpha_i \tilde{Y}_i = \alpha_0 + \sum_{i=1}^n \alpha_i (Y_i - \mathbb{E}[Y_i]). \quad (8.3)$$

Cet élément $Z \in H$ est caractérisé par le fait que $(X - Z)$ est orthogonal à H , pour le produit scalaire euclidien. Ceci implique :

$$\mathbb{E} [(X - Z) \cdot 1] = 0, \quad \text{ce qui est équivalent à } \mathbb{E}[X] = \alpha_0,$$

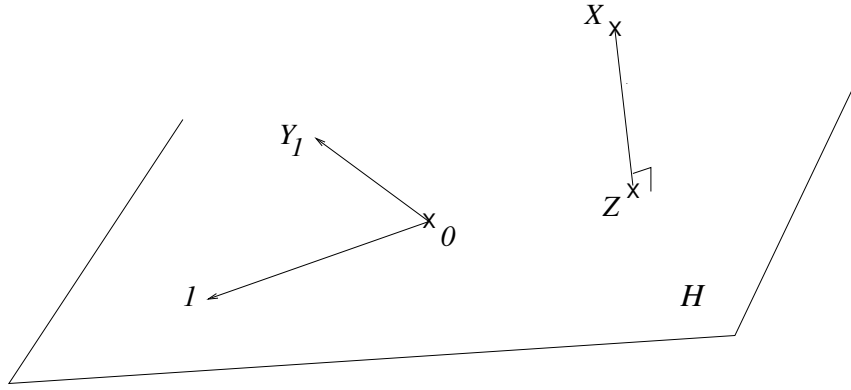


FIGURE 8.4 – Projection orthogonale de X sur l'hyperplan H engendré par $1, Y_1, \dots, Y_n$ (ici on a représenté le cas $n = 1$).

et pour tout $k = 1, \dots, n$,

$$\begin{aligned} \mathbb{E}[(X - Z) \cdot \tilde{Y}_k] &= 0 \\ \iff \text{cov}(Z, Y_k) &= \text{cov}(X, Y_k) \\ \iff \sum_{j=1}^n \alpha_j \text{cov}(Y_j, Y_k) &= \text{cov}(X, Y_k). \end{aligned}$$

Inversement, pour tout n -vecteur (α_j) vérifiant les conditions ci-dessus, le vecteur Z donné par (8.3) vérifie $(X - Z) \perp H$, c'est donc la projection de X sur H .

Si la matrice K_Y est inversible, le système linéaire $\alpha K_Y = \text{cov}(X, Y_k)$ admet une unique solution $(\alpha_i)_{i=1, \dots, n}$. Cette inversibilité de K_Y indique que les v.a. centrées $(\tilde{Y}_i)_{i=1, \dots, n}$ sont linéairement indépendantes. Comme elles sont aussi indépendantes de la fonction constante 1, cela implique que $\dim H = n + 1$, donc que la représentation (8.3) est unique. Au contraire, si $\det K_Y = 0$, les (\tilde{Y}_j) forment une famille dépendante, l'espace H sera de dimension $< n + 1$, et la représentation (8.3) de Z ne sera pas unique.

□

Exemple 8.19. Si on cherche à approcher X par une unique v.a. Y (qui n'est pas constante p.s., donc de variance > 0), la meilleure approximation est donnée par la fonction affine de Y :

$$Z = \mathbb{E}[X] + \frac{\text{cov}(X, Y)}{\text{var}(Y)} (Y - \mathbb{E}[Y]).$$

On appelle cette fonction affine la droite de régression de X en Y .

8.4.3 Fonction caractéristique

Pour décrire la loi d'une v.a. réelle X , on s'est servi de ses moments $\mathbb{E}[X^p]$ (lorsqu'ils sont définis). Ces moments donnent une idée de la « localisation » de la v.a. X . L'existence même de tous ses moments implique que X est suffisamment localisée. Suivant l'inégalité de Bienaymé-Tchebicheff, cette « localisation » se traduit par la propriété suivante : la probabilité que X prenne des « grandes valeurs » est faible.

On peut chercher à décrire P_X en considérant l'espérance d'une autre famille de fonctions de X , les exponentielles complexes $e^{i\xi X}$, de paramètres $\xi \in \mathbb{R}$. Cette définition s'étend facilement à la dimension supérieure.

Définition 8.20. Si X est une v.a. à valeurs dans \mathbb{R}^d , la *fonction caractéristique* de X est la fonction $\Phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ définie par :

$$\Phi_X(\xi) = \mathbb{E} [\exp(i\xi \cdot X)], \quad \xi \in \mathbb{R}^d.$$

On peut donc écrire

$$\Phi_X(\xi) = \int_{\mathbb{R}^d} e^{i\xi \cdot x} P_X(dx),$$

ce qui permet d'interpréter Φ_X comme la transformée de Fourier⁵ de la loi P_X , on écrit alors $\Phi_X(\xi) = \hat{P}_X(\xi)$. Comme la fonction $x \mapsto e^{i\xi \cdot x}$ est bornée et continue, elle est intégrable pour la mesure de probabilité P_X . Le TCD montre que Φ_X est continue et bornée sur \mathbb{R}^d :

$$|\Phi_X(\xi)| \leq \int_{\mathbb{R}^d} 1 P_X(dx) = 1, \quad \forall \xi \in \mathbb{R}^d.$$

Le paramètre $\xi \in \mathbb{R}^d$ est appelé le « paramètre de Fourier », le « vecteur de Fourier », ou le « vecteur d'onde ».

Nous allons à présent montrer que la fonction Φ_X *caractérise* la loi P_X , au sens où l'application $P_X \mapsto \Phi_X$ est injective. Pour cela, on étudie tout d'abord la classe des v.a. gaussiennes.

Lemme 8.21. *Soit X une v.a. de loi gaussienne $\mathcal{N}(0, \sigma^2)$. Alors*

$$\Phi_X(\xi) = \exp\left(-\frac{\sigma^2 \xi^2}{2}\right), \quad \xi \in \mathbb{R}.$$

5. On avait défini la transformée de Fourier d'une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ par $\hat{f}(\xi) = \int e^{-i\xi \cdot x} f(x) dx$. On a ici le facteur $e^{i\xi \cdot x}$ au lieu de $e^{-i\xi \cdot x}$, c'est juste une question de convention.

Démonstration. Considérons tout d'abord le cas $\sigma = 1$. On veut calculer l'intégrale

$$\Phi_X(\xi) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2 + i\xi x} dx.$$

Par le changement de variable $y = x/\sigma$ on se ramène au cas $\sigma = 1$, qu'on considère dorénavant. Ensuite, un argument de parité montre que $\int e^{-x^2/2} \sin(\xi x) dx = 0$, de sorte que la partie imaginaire de Φ_X est identiquement nulle. Sa partie réelle vaut

$$\Phi_X(\xi) = f(\xi) = \frac{1}{\sqrt{2\pi}} \int e^{-x^2/2} \cos(\xi x) dx.$$

D'après la section 3.3.2, et le fait que la fonction $|x|e^{-x^2/2}$ est intégrable, on a le droit de dériver par rapport à ξ sous le signe intégral :

$$f'(\xi) = -\frac{1}{\sqrt{2\pi}} \int e^{-x^2/2} \sin(\xi x) x dx.$$

En intégrant par parties, on trouve

$$f'(\xi) = -\frac{1}{\sqrt{2\pi}} \int e^{-x^2/2} \xi \cos(\xi x) x dx = -\xi f(\xi).$$

La fonction f est donc solution de l'équation différentielle $f'(\xi) = -\xi f(\xi)$, avec la condition initiale $f(0) = 1$. La seule solution de cette équation différentielle est $f(\xi) = \exp(-\xi^2/2)$.

Pour finir, le cas $\sigma \neq 1$ se traite ramène au cas précédent par le changement de variables $y = x/\sigma$:

$$\Phi_X(\xi) = \int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2 + i\xi x} dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2 + i\xi\sigma y} dy = f(\sigma\xi).$$

□

La famille des fonctions gaussiennes est donc invariante par passage à la transformée de Fourier.

On en vient au résultat principal de cette section :

Théorème 8.22. *La fonction caractéristique Φ_X d'une v.a. X à valeurs dans \mathbb{R}^d caractérise la loi P_X . Autrement dit, si, pour deux v.a. X et Y , on a $\Phi_X = \Phi_Y$, alors les lois $P_X = P_Y$.*

Démonstration. Traitons tout d'abord le cas unidimensionnel. Appelons g_σ la densité de la

loi gaussienne $\mathcal{N}(0, \sigma^2)$:

$$g_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

On va se servir de g_σ comme d'un noyau de convolution. Pour μ une mesure de probabilité sur \mathbb{R} , on considère la mesure régularisée

$$\begin{aligned} \mu_\sigma(dx) &= f_\sigma(x) dx, \quad \text{avec} \\ f_\sigma(x) &= g_\sigma * \mu(x) := \int g_\sigma(x-y) \mu(dy). \end{aligned}$$

Par cette convolution, on a transformé la mesure de probabilité quelconque μ , en une mesure régulière μ_σ (à densité dans $C^\infty(\mathbb{R})$).

Pour montrer le théorème, il faut montrer que :

1. μ_σ est déterminée par la transformée de Fourier $\hat{\mu}$;
2. pour tout fonction $\varphi \in C_b(\mathbb{R})$, les intégrales $\int \varphi(x) \mu_\sigma(x) \rightarrow \int \varphi(x) \mu(dx)$ lorsque $\sigma \searrow 0$.

Comme toute mesure de probabilité μ sur \mathbb{R} est *régulière* (c'est un cas particulier de la Proposition 4.30), connaître toutes les intégrales $\int \varphi d\mu$, pour $\varphi \in C_b(\mathbb{R})$, suffit à identifier la mesure μ .

Etablissons tout d'abord le point 1. Le Lemme ci-dessus montre (en échangeant σ et $1/\sigma$) que :

$$\forall x \in \mathbb{R}, \quad \sigma\sqrt{2\pi}g_\sigma(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right) = \int_{\mathbb{R}} e^{i\xi x} g_{1/\sigma}(\xi) d\xi. \quad (8.4)$$

En injectant cette égalité dans la convolution f_σ , il vient :

$$\begin{aligned} f_\sigma(x) &= \int_{\mathbb{R}} g_\sigma(x-y) \mu(dy) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{i\xi(x-y)} g_{1/\sigma}(\xi) d\xi \right) \mu(dy) \\ &\stackrel{\text{Fubini}}{=} \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\xi x} g_{1/\sigma}(\xi) \left(\int_{\mathbb{R}} e^{-i\xi y} \mu(dy) \right) d\xi \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\xi x} g_{1/\sigma}(\xi) \hat{\mu}(-\xi) d\xi. \end{aligned}$$

L'application du théorème de Fubini-Lebesgue est possible, puisque la fonction $(y, \xi) \mapsto e^{i\xi(x-y)} g_{1/\sigma}(\xi)$ est bien dans $\mathcal{L}^1(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \mu(dy) \times d\xi)$. On a donc exprimé explicitement la densité f_σ de μ_σ , à partir de la fonction $\hat{\mu}$.

Pour montrer le point 2., on développe l'intégrale $\int \varphi d\mu_\sigma$, et « fait passer » la convolution

de la mesure μ vers la fonction φ :

$$\begin{aligned} \int \varphi d\mu_\sigma &= \int \varphi(x) \left(\int g_\sigma(x-y) \mu(dy) \right) dx \stackrel{\text{Fubini}}{=} \int \left(\int \varphi(x) g_\sigma(x-y) dx \right) \mu(dy) \\ &= \int (g_\sigma * \varphi) d\mu. \end{aligned}$$

On va ensuite utiliser le fait que les gaussiennes g_σ ont des propriétés similaires aux approximations de δ étudiées dans la section 6.4.3. Les noyaux g_σ satisfont en effet :

$$\begin{aligned} \int g_\sigma(x) dx &= 1, \quad \forall \sigma > 0, \\ \lim_{\sigma \searrow 0} \int_{\{|x| > \varepsilon\}} g_\sigma(x) dx &= 0, \quad \forall \varepsilon > 0, \end{aligned}$$

la différence principale étant le support non compact des g_σ . Pour une fonction $\varphi \in C_b(\mathbb{R})$, on peut montrer un résultat similaire à celui de la Proposition 6.16, *i*), c'est-à-dire la convergence ponctuelle :

$$\forall y \in \mathbb{R}, \quad \lim_{\sigma \searrow 0} g_\sigma * \varphi(y) = \varphi(y),$$

et la convergence est uniforme sur tout compact. La preuve de ce résultat est similaire à celle de la Proposition (point *i*)). L'absence de compacité de g_σ étant remplacée par le caractère borné de φ , de sorte que l'intégrale « loin du pic de g_σ »,

$$\left| \int_{\{|x| \geq \varepsilon\}} \varphi(y-x) g_\sigma(x) dx \right| \leq \|\varphi\|_{\text{sup}} \left| \int_{\{|x| \geq \varepsilon\}} g_\sigma(x) dx \right|$$

converge vers zéro lorsque $\sigma \rightarrow 0$, uniformément par rapport à y .

Comme $|g_\sigma * \varphi| \leq \|\varphi\|_{\text{sup}}$, et que μ est une mesure de probabilité, le théorème de convergence dominée nous montre alors que

$$\int \varphi d\mu_\sigma = \int g_\sigma * \varphi d\mu \xrightarrow{\sigma \searrow 0} \int \varphi d\mu,$$

ce qui achève de prouver le point 2.

Le cas de la dimension supérieure est très similaire. Le noyau de convolution g_σ est remplacé par un noyau d -dimensionnel

$$g_\sigma^{(d)}(x_1, \dots, x_d) = \prod_{j=1}^d g_\sigma(x_j).$$

En intégrant séparément chaque variable, on obtient une identité similaire à (8.4) :

$$\forall x \in \mathbb{R}^d, \quad \left(\sigma\sqrt{2\pi}\right)^d g_\sigma^{(d)}(x) = \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right) = \int_{\mathbb{R}^d} e^{i\xi \cdot x} g_{1/\sigma}^{(d)}(\xi) d\xi.$$

□

Les fonctions caractéristiques permettent donc d'encoder l'information sur une loi P_X de v.a. vectorielle (donc une mesure de probabilité sur \mathbb{R}^d), à travers une fonction continue. Cette fonction permet aussi de retrouver les moments de la variable X , lorsque ceux-ci existent. Le résultat ci-dessus concerne les v.a. dans $L^2(\Omega, \mathbb{P})$, qui admettent des moments d'ordres $p = 1, 2$.

Proposition 8.23. *Soit $X = (X_1, \dots, X_d)$ une v.a. à valeurs dans \mathbb{R}^d , et de carré intégrable. Alors Φ_X est de classe C^2 , et admet comme développement de Taylor :*

$$\Phi_X(\xi) = 1 + i \sum_{j=1}^d \xi_j \mathbb{E}[X_j] - \frac{1}{2} \sum_{k=1}^d \sum_{j=1}^d \xi_j \xi_k \mathbb{E}[X_j X_k] + o(|\xi|^2),$$

quand $\xi = (\xi_1, \dots, \xi_d)$ tend vers 0.

Démonstration. Par hypothèse, la fonction $\|x\|^2$ est intégrable par rapport à la mesure P_X , d'intégrale donnant $\mathbb{E}[|X|^2]$. Cela implique que les fonctions x_j^2 sont intégrables, et par Cauchy-Schwarz, de même pour les fonctions x_j et $x_j x_k$ pour tout $j, k \in \{1, \dots, d\}$. Or $\Phi_X(\xi)$ est donnée par l'intégrale de l'exponentielle $e^{i\xi \cdot x}$. Les dérivées d'ordre 1 et 2 de cette fonction par rapport à ξ donnent $ix_j e^{i\xi \cdot x}$ et $-x_j x_k e^{i\xi \cdot x}$, des fonctions intégrables. On est donc en mesure d'appliquer le théorème de dérivée sous le signe intégrale, et on obtient :

$$\frac{\partial \Phi_X}{\partial \xi_j}(\xi) = i \int x_j e^{i\xi \cdot x} P_X(dx) = i \mathbb{E} [X_j e^{i\xi \cdot X}].$$

En appliquant une seconde fois le théorème de dérivation sous le signe intégral, on voit que

$$\frac{\partial^2 \Phi_X}{\partial \xi_j \partial \xi_k}(\xi) = - \int x_j x_k e^{i\xi \cdot x} P_X(dx) = - \mathbb{E} [X_j X_k e^{i\xi \cdot X}].$$

Le théorème de continuité sous le signe intégrale assure que les fonctions ci-dessus sont continues par rapport à ξ . Ceci montre que Φ_X est une fonction C^2 , on a donc le droit d'écrire son développement de Taylor à l'ordre 2 basé à l'origine $\xi = 0$. □

Remarque 8.24. Si on suppose que $|X|^p$ est intégrable, avec $p \geq 1$ entier, le même raisonnement montre que la fonction Φ_X est de classe C^p , et que son développement de Taylor d'ordre p en $\xi = 0$ fournit les moments d'ordres $\leq p$ de X . Ainsi, les propriétés de localisation d'une v.a X se traduisent en propriétés de régularité de sa fonction caractéristique Φ_X .

8.4.4 Fonction génératrice d'une variable aléatoire discrète

Pour une v.a. à valeurs dans \mathbb{N} , la fonction caractéristique est remplacée par une *fonction génératrice*, de paramètre $r \in [0, 1]$.

Définition 8.25. Soit X une v.a. à valeurs dans \mathbb{N} . La fonction génératrice de X est la fonction g_X définie sur l'intervalle $r \in [-1, 1]$ par la série entière :

$$g_X(r) = \sum_{n=0}^{\infty} \mathbb{P}(X = n) r^n = \mathbb{E}[r^X]$$

La fonction g_X est bien définie pour $r \in [-1, 1]$, elle est croissante sur $[0, 1]$, avec pour valeurs limites $g_X(0) = \mathbb{P}(X = 0)$ et $g_X(1) = 1$. Le TCD montre qu'elle est continue sur $[-1, 1]$. Le rayon de convergence de la série entière est donc ≥ 1 . Les poids $\mathbb{P}(X = n)$ apparaissent naturellement dans le développement de Taylor en $r = 0$ de g_X , ce qui montre que g_X contient toute l'information sur la loi P_X .

La dérivée de r^n donne nr^n , qui tend vers zéro lorsque $n \rightarrow \infty$ si $|r| < 1$. Cela montre (par le théorème de dérivation sous le signe somme) que g_X est dérivable, et en fait infiniment dérivable, sur l'intervalle $] -1, 1[$. En $r = 1$ elle admet une dérivée à gauche, qui peut éventuellement diverger :

$$g'_X(r = 1) = \mathbb{E}[X] \in [0, \infty].$$

Plus généralement, pour tout $p \in \mathbb{N}^*$,

$$\lim_{r \uparrow 1} g_X^{(p)}(r) = \mathbb{E}[X(X-1)(X-2) \cdots (X-p+1)].$$

Comme dans le cas de la fonction caractéristique, on peut donc retrouver tous les moments de X en étudiant le comportement de sa fonction génératrice, cette fois-ci en $r = 1$.

Remarque 8.26. On peut faire un lien entre la fonction génératrice et la fonction caractéristique, en étendant g_X aux arguments complexes $z \in \{|z| \leq 1\}$. Sur le cercle unité, on trouve

$$g_X(e^{i\theta}) = \sum_{n=0}^{\infty} \mathbb{P}(X = n) e^{in\theta} = \mathbb{E}[e^{i\theta X}],$$

qui est donc la fonction caractéristique de la v.a. discrète X . Les dérivées par rapport à θ de $g_X(e^{i\theta})$ en $\theta = 0$ sont directement reliées aux dérivées de $g_X(r)$ en $r = 1$, ce qui explique pourquoi on retrouve les moments de X .

9 Indépendance

En théorie des probabilités, la notion d'indépendance entre deux évènements, ou entre deux variables aléatoires, est facile à comprendre intuitivement : elle signifie que le premier évènement n'influence pas le second ; respectivement que la valeur prise par la première v.a. n'influence pas la valeur prise par le 2e. Un exemple typique est la répétition d'un tirage de dé : lorsqu'on tire un dé plusieurs fois, le résultat de 1er tirage n'influencera pas le résultat du 2e tirage, ni du 3e tirage. Cette indépendance peut donc ici signifier l'absence de mémoire du processus de tirage répété d'un dé.

Un résultat important de ce chapitre concerne le tirage de N dés identiques, plus généralement la répétition un grand nombre de fois de la même expérience aléatoire, avec absence de mémoire. Mathématiquement, il s'agit donc de N variables aléatoires indépendantes, chacune ayant la même loi. La *loi des grands nombres* nous fournit alors un lien entre la définition mathématique d'une distribution de probabilité, et la définition heuristique (ou « fréquentiste ») d'une telle distribution, qui est celle de la fréquence d'apparition d'un évènement, lorsqu'on répète l'expérience un grand nombre de fois.

Ce caractère intuitif de la notion d'indépendance entre deux v.a. X_1, X_2 , cette propriété n'est pas forcément facile à se représenter lorsqu'on voit X_1, X_2 comme deux fonctions sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Au-delà des évènements et des v.a., on verra que l'indépendance peut également concerner des sous-tribus de \mathcal{A} , et que cette notion d'indépendance de sous-tribus est en fait plus « fondamentale » que celle des évènements ou des v.a.

9.1 Evènements indépendants

Définition 9.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Deux évènements $A, B \in \mathcal{A}$ sont dits indépendants si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B). \quad (9.1)$$

Cette propriété est toujours vérifiée si A ou B sont négligeables. Supposons que $\mathbb{P}(B) > 0$. On peut alors définir la *probabilité conditionnelle*

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{aussi appelée la « probabilité de } A, \text{ sachant } B \text{ »}.$$

On remarque que $\mathbb{P}(\bullet|B)$ est aussi une mesure de probabilité sur \mathcal{A} . Elle mesure la probabilité qu'un évènement A ait lieu, lorsqu'on sait que B a lieu ; ou la probabilité de A , conditionnée par le fait que B a lieu.

Proposition 9.2. *Supposons $\mathbb{P}(B) > 0$. Alors l'évènement A est indépendant de l'évènement B si*

$$\mathbb{P}(A|B) = \mathbb{P}(A).$$

Autrement dit, la connaissance de B ne modifie pas la probabilité que A ait lieu.

La proposition ci-dessus donne des rôles différents à A et B . Cependant, la propriété d'indépendance est symétrique : « A est indépendant de B » est équivalent à « B est indépendant de A ».

Exemple 9.3. (i) Lancer de deux dés (ou 2 lancers successifs du même dé). Comme on l'a vu, l'espace de probabilité est alors $\Omega_2 = \{1, \dots, 6\}^2$, et la mesure de probabilité naturelle est uniforme, $\mathbb{P}(\omega) = \frac{1}{36}$ pour tout $\omega \in \Omega_2$. L'évènement « le premier dé tombe sur 6 » correspond à

$$A = \{6\} \times \{1, \dots, 6\} = \{(6, 1), (6, 2), \dots, (6, 6)\},$$

tandis que « le second dé tombe sur 4 » vaut $B = \{1, \dots, 6\} \times \{4\}$. Chacun de ces évènements a une probabilité $\frac{6}{6^2} = \frac{1}{6}$. On vérifie que l'évènement $A \cap B = \{(6, 4)\}$, de probabilité $\frac{1}{36} = \frac{1}{6} \times \frac{1}{6}$ de sorte que la relation (9.1) est vérifiée : ces deux évènements sont bien indépendants. On avait en fait défini la mesure uniforme \mathbb{P} de manière à ce que deux tirages consécutifs soient indépendants, v. la section 8.1.

(ii) Si on ne lance le dé qu'une seule fois, certaines paires d'évènements de Ω_1 peuvent être indépendants. Par exemple, les évènements $A = \{1, 2\}$ et $B = \{1, 3, 5\}$ sont indépendants, Par contre, $A = \{1, 2\}$ et $C = \{3, 4, 5\}$ ne sont pas indépendants, ils sont même corrélés (négativement) : si A est réalisé, alors on sait à coup sûr que B ne l'est pas. L'indépendance entre A et B est ici assez subtile, et pas très intuitive.

L'indépendance de 2 évènements peut se généraliser à n évènements.

Définition 9.4. [Evènements indépendants] On dit que n évènements $A_1, A_2, \dots, A_n \in \mathcal{A}$ sont indépendants si, pour tout sous-ensemble non vide d'indices $\{j_1, \dots, j_p\} \subset \{1, 2, \dots, n\}$, on a

$$\mathbb{P}(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_p}) = \mathbb{P}(A_{j_1})\mathbb{P}(A_{j_2}) \dots \mathbb{P}(A_{j_p}). \quad (9.2)$$

Il s'agit donc d'une propriété de factorisation des probabilités.

Remarque 9.5. 1. La propriété unique $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \dots \mathbb{P}(A_n)$ ne suffit pas pour caractériser l'indépendance. Par exemple, si on considère un évènement $A \in \mathcal{A}$ tel

que $0 < \mathbb{P}(A) < 1$, alors les évènements $A_1 = A$, $A_2 = A^c$, $A_3 = \emptyset$ vérifient cette unique équation, mais ils ne sont pas indépendants.

2. Inversement, il ne suffit pas que les paires A_i, A_j soient indépendantes deux à deux, pour que les $(A_i)_{i=1,\dots,n}$ le soient globalement. Si on tire à pile ou face 2 fois, les évènements suivants sont indépendants deux à deux, mais pas globalement :

$$A = \{\text{pile au 1er lancer}\}, \quad B = \{\text{pile au 2e lancer}\}, \quad C = \{\text{même résultat aux deux lancers}\}.$$

Le manque d'indépendance globale se voit par le fait que, si on connaît le résultat des deux lancers (donc on sait si A et B sont réalisés ou non), alors on sait parfaitement si C est réalisé ou non.

Malgré la remarque 1., on peut caractériser l'indépendance par une propriété similaire à cette unique égalité.

Proposition 9.6. *Les n évènements A_1, \dots, A_n sont indépendants ssi*

$$\mathbb{P}(B_1 \cap B_2 \cap \dots \cap B_n) = \mathbb{P}(B_1)\mathbb{P}(B_2) \dots \mathbb{P}(B_n), \quad (9.3)$$

pour tout choix B_i dans la sous-tribu $\sigma(A_i) = \{\emptyset, A_i, A_i^c, \Omega\}$, $i \in \{1, \dots, n\}$. (Il ne s'agit donc pas d'une seule équation, mais de 4^n équations à vérifier).

Démonstration. La preuve de \Leftarrow est simple. Pour un sous-ensemble d'indices $J = \{j_1, \dots, j_p\}$, prenons $B_i = A_i$ si $i \in J$ et $B_i = \Omega$ sinon ; la relation (9.2) découle alors de (9.3).

Montrons à présent \Rightarrow . On peut supposer que $B_i \neq \emptyset$ pour tout i (sinon l'équation (9.3) est évidente). Prenons alors l'ensemble d'indices $J = \{i : B_i \neq \Omega\}$, autrement dit $J = \{i : B_i = A_i \text{ ou } B_i = A_i^c\}$. On veut alors montrer, pour $J = \{j_1, \dots, j_p\}$:

$$\mathbb{P}(B_{j_1} \cap \dots \cap B_{j_p}) = \mathbb{P}(B_{j_1}) \dots \mathbb{P}(B_{j_p}) \quad \text{pour } B_{j_k} = A_{j_k} \quad \text{ou } B_{j_k} = A_{j_k}^c.$$

L'indépendance des (A_i) montre que cette égalité est déjà vraie si tous les $B_{j_k} = A_{j_k}$.

Montrons maintenant que si A_1, \dots, A_n sont indépendants, alors A_1^c, A_2, \dots, A_n sont également indépendants. Pour tout sous-ensemble $\{i_1, \dots, i_q\} \subset \{1, \dots, n\}$, si tous les $i_j \neq 1$ alors

la relation (9.2) est vérifiée. Si $i_1 = 1$, alors

$$\begin{aligned}\mathbb{P}(A_1^c \cap A_{i_2} \cap \cdots \cap A_{i_q}) &= \mathbb{P}((A_{i_2} \cap \cdots \cap A_{i_q}) \setminus (A_1 \cap A_{i_2} \cap \cdots \cap A_{i_q})) = \mathbb{P}(A_{i_2} \cap \cdots \cap A_{i_q}) - \mathbb{P}(A_1 \cap A_{i_2} \cap \cdots \cap A_{i_q}) \\ &= \mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_q}) - \mathbb{P}(A_1) \mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_q}) \\ &= (1 - \mathbb{P}(A_1)) \mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_q}) \\ &= \mathbb{P}(A_1^c) \mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_q}).\end{aligned}$$

Par itération, on montre que les 2^n n -uplets $A_1^{(c)}, A_2^{(c)}, \dots, A_n^{(c)}$ sont indépendants. La relation ci-dessus entre les B_{j_k} est donc vraie, quelque soit le choix pour les B_{j_k} . \square

9.2 Tribus et variables aléatoires et tribus indépendantes

On passe naturellement de la notion d'évènements indépendants, à celle de sous-tribus indépendantes. La Proposition (9.6) donne déjà un exemple de ce phénomène.

Définition 9.7. [Sous-tribus indépendantes] Soient $\mathcal{B}_1, \dots, \mathcal{B}_n$ n sous-tribus de \mathcal{A} . On dit que ces sous-tribus sont indépendantes ssi

$$\forall A_1 \in \mathcal{B}_1, \dots, \forall A_n \in \mathcal{B}_n, \quad \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2) \cdots \mathbb{P}(A_n).$$

La Proposition (9.6) montre qu'un n -uplet d'évènements A_1, \dots, A_n est indépendant ssi le n -uplet des tribus engendrées $\sigma(A_1), \dots, \sigma(A_n)$ est indépendant. Ceci nous amène aussi naturellement à la notion d'indépendance entre n variables aléatoires; c'est cette notion qui nous sera en fait la plus utile.

Définition 9.8. [Variables aléatoires indépendantes] Soient X_1, X_2, \dots, X_n n variables aléatoires à valeurs dans des espaces mesurables $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$. Ces variables sont dites indépendantes ssi les tribus qu'elles engendrent, $\sigma(X_1), \sigma(X_2), \dots, \sigma(X_n)$, sont indépendantes. Comme $\sigma(X_i) = \{X_i^{-1}(F), F \in \mathcal{E}_i\}$, l'indépendance est équivalente à la propriété suivante :

$$\forall F_1 \in \mathcal{E}_1, \dots, \forall F_n \in \mathcal{E}_n, \quad \mathbb{P}(\{X_1 \in F_1\} \cap \cdots \cap \{X_n \in F_n\}) = \mathbb{P}(X_1 \in F_1) \cdots \mathbb{P}(X_n \in F_n).$$

On retrouve l'idée de l'indépendance donnée dans l'introduction à la section : la connaissance de la valeur d'une v.a. X_1 ne donne aucun renseignement quant aux valeurs des autres X_i .

Remarque 9.9. Supposons les n sous-tribus $\mathcal{B}_1, \dots, \mathcal{B}_n$ indépendantes. Alors, si pour tout $i = 1, \dots, n$ la v.a. X_i est \mathcal{B}_i -mesurable, alors les variables X_1, X_2, \dots, X_n sont indépendantes.

9.2.1 Caractérisation de l'indépendance de n v.a. en termes de la loi conjointe

Pour n v.a. X_1, X_2, \dots, X_n à valeurs dans E_1, \dots, E_n , on peut considérer la v.a. vectorielle (X_1, X_2, \dots, X_n) , qui prend ses valeurs dans l'espace produit $E_1 \times \dots \times E_n$, muni de la tribu produit $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$. On va caractériser l'indépendance des $(X_i)_{i=1, \dots, n}$ en terme de la loi de la v.a. (X_1, X_2, \dots, X_n) , qu'on appelle aussi la loi conjointe des v.a. X_i .

Théorème 9.10. *Les v.a. X_1, \dots, X_n sont indépendantes ssi la loi du n -uplet (X_1, \dots, X_n) est le produit des lois de X_1, X_2, \dots, X_n :*

$$P_{(X_1, \dots, X_n)} = P_{X_1} \otimes P_{X_2} \otimes \dots \otimes P_{X_n}.$$

De plus, pour toutes fonctions $f_i : E \rightarrow \mathbb{R}_+$ mesurables positives, $i = 1, \dots, n$, on a alors la factorisation des espérances :

$$\mathbb{E} \left[\prod_{i=1}^n f_i(X_i) \right] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)].$$

Démonstration. Prenons des ensembles mesurables $F_i \in \mathcal{E}_i$, $i = 1, \dots, n$. La probabilité que tous les évènements $(X_i \in F_i)$ soient réalisés simultanément vaut :

$$P_{(X_1, \dots, X_n)}(F_1 \times \dots \times F_n) = \mathbb{P}(\{X_1 \in F_1\} \cap \{X_2 \in F_2\} \cap \dots \cap \{X_n \in F_n\}).$$

D'après la Définition 9.8, les $(X_i)_{i=1, \dots, n}$ sont des v.a. indépendantes ssi, pour tout choix des F_i , cette probabilité se factorise :

$$P_{(X_1, \dots, X_n)}(F_1 \times \dots \times F_n) = \mathbb{P}(X_1 \in F_1) \dots \mathbb{P}(X_n \in F_n).$$

D'autre part, on sait que la mesure produit $P_{X_1} \otimes \dots \otimes P_{X_n}$ sur $\prod_i E_i$ est caractérisée par ses valeurs sur les pavés mesurables :

$$\forall F_i \in \mathcal{E}_i, i = 1, \dots, n, P_{X_1} \otimes \dots \otimes P_{X_n}(F_1 \times \dots \times F_n) = \mathbb{P}(X_1 \in F_1) \dots \mathbb{P}(X_n \in F_n).$$

Ceci suffit donc à montrer que les mesures $P_{(X_1, \dots, X_n)}$ et $P_{X_1} \otimes \dots \otimes P_{X_n}$ coïncident sur toute la tribu produit $\mathcal{E}_1 \otimes \dots \otimes \mathcal{E}_n$.

Le second point est une conséquence directe du théorème de Fubini-Tonnelli, appliqué sur le

produit $\prod_{i=1}^n E_i$:

$$\begin{aligned}
 \mathbb{E} \left[\prod_{i=1}^n f_i(X_i) \right] &= \int_{\prod E_i} \prod_{i=1}^n f_i(x_i) P_{(X_1, \dots, X_n)}(dx_1 \cdots dx_n) \prod_{i=1}^n \mathbb{E}[f_i(X_i)]. \\
 &= \int_{\prod E_i} \prod_{i=1}^n f_i(x_i) P_{X_1} \otimes \cdots \otimes P_{X_n}(dx_1 \cdots dx_n) \\
 &= \int_{\prod E_i} \prod_{i=1}^n f_i(x_i) P_{X_1}(dx_1) \cdots P_{X_n}(dx_n) \\
 &\stackrel{\text{F-T}}{=} \prod_{i=1}^n \int_{E_i} f_i(x_i) P_{X_i}(dx_i) \\
 &= \prod_{i=1}^n \mathbb{E}[f_i(X_i)].
 \end{aligned}$$

□

Si on se donne n lois de probabilité μ_1, \dots, μ_n , ce théorème permet de construire un espace de probabilité portant n v.a. indépendantes Y_i de lois respectives μ_1, \dots, μ_n . Prenons par exemple des v.a. réelles, auquel cas les μ_i sont des mesures de probabilité sur \mathbb{R} . On sait construire la v.a. vectorielle (Y_1, \dots, Y_n) à valeurs dans \mathbb{R}^n , dont la loi est le produit $\mu_1 \otimes \cdots \otimes \mu_n$. Le théorème ci-dessus montre que les composantes Y_i de ce vecteur aléatoire sont des v.a. indépendantes, de lois respectives μ_1, \dots, μ_n .

Remarque 9.11. Ce théorème s'étend aux fonctions f_i réelles de signe variable, à condition que les v.a. $f_i(X_i)$ soient intégrables : $\mathbb{E}[|f_i(X_i)|] < \infty$ pour $i = 1, \dots, n$. On a alors aussi

$$\mathbb{E} \left[\left| \prod_{i=1}^n f_i(X_i) \right| \right] = \prod_{i=1}^n \mathbb{E}[|f_i(X_i)|] < \infty,$$

ce qui donne un sens à l'espérance jointe $\mathbb{E}[\prod_{i=1}^n f_i(X_i)] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)]$.

Par exemple, si les X_i sont des v.a. réelles indépendantes et intégrables, la v.a. produit $X_1 X_2 \cdots X_n$ est aussi intégrable, et d'espérance

$$\mathbb{E}[X_1 X_2 \cdots X_n] = \prod_{i=1}^n \mathbb{E}[X_i]. \quad (9.4)$$

ATTENTION : le fait que $X_1 \cdots X_n$ soit intégrable dépend fortement de l'hypothèse d'indépendance des X_i , et est faux en général sans cette hypothèse. Par exemple, si on considère

une v.a. $X_1 \in L^1$ telle que $X_1 \notin L^2$, et on prend $X_2 = X_1$, la variable $X_1 X_2 = X_1^2$ n'est pas intégrable.

Corollaire 9.12. *Si X_1, X_2 sont deux v.a. réelles indépendantes et dans L^2 , alors on a $\text{cov}(X_1, X_2) = 0$.*

Démonstration. Evident à partir de (9.4), puisque $\text{cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$. \square

ATTENTION : la réciproque à ce corollaire est fausse. La propriété de covariance nulle entre deux v.a. est beaucoup plus faible que la propriété d'indépendance. Prenons par exemple une v.a. réelle X_1 , à densité $p(x)$ donnée par une fonction paire ($p(x) = p(-x)$ pour tout $x \in \mathbb{R}$). Ensuite, considérons la variable de Bernoulli $\varepsilon \in \{-1, 1\}$ où les deux signes sont équiprobables, et supposons que ε et X_1 sont indépendantes. Considérons alors la v.a. $X_2 = \varepsilon X_1$. Quelles sont alors les corrélations entre X_1 et X_2 ? On vérifie facilement que $\text{cov}(X_1, X_2) = 0$. Cependant, on voit que pour tout intervalle $I \subset \mathbb{R}$, l'évènement $X_1 \in I$ implique que $X_2 \in I \cup (-I)$. En particulier, si I est petit, l'information $X_2 \in I \cup (-I)$ est très contraignante, ce qui montre que l'information sur X_1 influence fortement les valeurs prises par X_2 .

Une autre manière de voir les choses est de raisonner par l'absurde. Si X_1 et X_2 étaient indépendantes, alors $|X_1|$ et $|X_2|$ le seraient également. Or ces deux v.a. sont égales! Le lemme suivant impose alors que $|X_1|$ est constante p.s., ce qui contredit le fait que $|X_1|$ est la variable à densité $2p(x)\mathbb{1}_{\mathbb{R}_+}(x)$.

Lemme 9.13. *Si une v.a. réelle est indépendante d'elle-même, alors elle doit être constante p.s., autrement dit sa loi est une mesure de Dirac en un point $x_0 \in \mathbb{R}$.*

Démonstration. Supposons que X_1 est indépendante d'elle-même. On a alors, pour tout borélien $B \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}(X_1 \in B) = \mathbb{P}(\{X_1 \in B\} \cap \{X_1 \in B\}) = \mathbb{P}(X_1 \in B)^2,$$

d'où on déduit que pour tout borélien B , la probabilité $\mathbb{P}(X_1 \in B)$ vaut 0 ou 1. En particulier, pour tout $t \in \mathbb{R}$, la fonction de répartition $F_X(t) = \mathbb{P}(X \leq t)$ vaut 0 ou 1. Cette fonction croissante a donc un unique point de discontinuité, noté x_0 , où elle « saute » de 0 à 1 : en utilisant la continuité à droite, on a alors $F_X(t) = \mathbb{1}_{[x_0, \infty[}(t)$, ce qui correspond à la loi $P_{X_1} = \delta_{x_0}$. \square

Dans le cas où les lois des X_i sont à densité, on obtient une expression simple de la loi jointe de variables indépendantes :

Corollaire 9.14. *Soient X_1, \dots, X_n des v.a. réelles.*

(i) *Supposons que pour chaque $i = 1, \dots, n$, la v.a. X_i est à densité $p_i(x)$, et que les v.a. $(X_i)_{i=1, \dots, n}$ sont indépendantes. Alors la loi du n -uplet (X_1, \dots, X_n) sur \mathbb{R}^n est à densité, donnée par*

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p_i(x_i), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n.$$

(ii) *Inversement, supposons que la loi du n -uplet a une densité de forme factorisée*

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n q_i(x_i), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n, \quad (9.5)$$

pour des fonctions q_i boréliennes positives. Alors les v.a. X_1, \dots, X_n sont indépendantes, et pour tout $i = 1, \dots, n$, la loi de X_i est à densité, donnée par $p_i = C_i q_i$, avec $C_i > 0$ des constantes adaptées pour normaliser p_i .

Démonstration. (i) est une application directe du Théorème 9.10, le théorème de Fubini-Tonnelli nous montrant que la mesure $P_{X_1} \otimes \dots \otimes P_{X_n}$ est de loi de densité $\prod_i p_i(x_i)$ sur \mathbb{R}^d .

Ce même théorème de Fubini-Tonnelli nous montre que l'hypothèse (9.5) implique la contrainte suivante sur les fonctions q_i :

$$\prod_{i=1}^n \left(\int_{\mathbb{R}} q_i(x) dx \right) = 1,$$

ce qui impose que $K_i := \int_{\mathbb{R}} q_i(x) dx \in]0, \infty[$ pour chaque $i = 1, \dots, n$. D'après la Proposition 8.9, la connaissance de la loi conjointe permet de retrouver la loi de chaque X_i :

$$\begin{aligned} p_i(x) &= \int_{\mathbb{R}^{d-1}} p(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_d) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_d \\ &= q_i(x) \prod_{j \neq i} \left(\int_{\mathbb{R}} q_j(x) dx \right) = q_i(x) \prod_{j \neq i} K_j = \frac{1}{K_i} q_i(x). \end{aligned}$$

La loi de X_i est donc de densité $p_i(x) = \frac{1}{K_i} q_i(x)$. La loi jointe est donnée par la densité $\prod_{i=1}^n p_i(x_i)$, c'est donc bien la loi produit $P_{X_1} \otimes \dots \otimes P_{X_n}$, ce qui montre l'indépendance des X_i . \square