Intégration, probabilités et analyse de Fourier

E. Kowalski

ETH Zürich
kowalski@math.ethz.ch

Do not infest your mind with beating on
The strangeness of this business; at pick'd leisure
Which shall be shortly, single I'll resolve you,
Which to you shall seem probable, of every
These happen'd accidents; till when, be cheerful
And think of each thing well.

N'occupez pas votre esprit à lutter
Contre l'étrangeté de ces choses ; le temps venu,
Qui sera proche, je résoudrai pour vous,
Et vous en verrez la probabilité, chacun
Des incidents survenus ; jusque là, soyez joyeux
Et n'en pensez que du bien.

W. Shakespeare, **The Tempest**

Table des matières

Préface

Ce cours a été rédigé originellement pour le cours d'intégration, analyse de Fourier et probabilités de Licence à l'Université Bordeaux I, que j'ai donné au second semestre 2001/2002, et il a ensuite été utilisé pour le même cours les deux années suivantes avec quelques modifications.

Après le texte proprement dit on trouvera les sujets des examens donnés durant les années 2001/2002 et 2002/2003 (examen partiel, puis examen de juin et examen de rattrapage de septembre) ; tout ces examens sont corrigés, ce qui pourra rendre ce texte utile pour une étude indépendante.

Par rapport à d'autres cours d'intégration de niveau comparable, la seule spécificité éventuelle de celui-ci est d'incorporer « en continu » la théorie des probabilités dans le développement, au lieu de la traiter indépendamment lors d'une partie séparée. Je reste persuadé que ce point de vue est une façon fructueuse d'initier un(e) étudiant(e) au langage et à la façon de penser probabiliste, mais il n'est cependant pas certain que cela soit la meilleure manière de procéder en pratique. En particulier, en l'absence (déplorable) d'un index très complet, cela rend difficile d'utiliser le texte comme outil d'apprentissage des probabilités à partir de la connaissance de la théorie de la mesure et de l'intégration.[1]

Un certain nombre d'erreurs présentes dans la première version de ce cours ont été corrigées. Une bonne part m'a été signalée par les étudiants et étudiantes qui ont assisté aux cours, et d'autres par Chantal Menini en particulier.

Il reste sans doute de nombreuses erreurs (au moins de détail) ; il est probable qu'elles resteront présentes jusqu'à ce que je reprenne ce texte pour base d'un cours, et lors d'une éventuelle traduction en anglais pour un enseignement ultérieur à ETH Zürich – mais cela ne pourra arriver avant plusieurs années...

Je dédie avec plaisir ce qu'il y de bien dans ce cours à André Goldman, en souvenir toujours émerveillé de ses cours de probabilités de maîtrise et D.E.A que j'ai suivis il y a bien longtemps...

[1] Mais peut-être cet index verra-t-il le jour ultérieurement, ce qui diminuera l'importance de ce bémol...

Introduction

L'intégrale d'une fonction est définie d'abord pour des fonctions relativement régulières définies sur un intervalle fini. Elle a deux propriétés essentielles : l'intégrale indéfinie donne l'opération inverse de la dérivation, et l'intégrale définie d'une fonction continue positive est l'aire de la région du plan placée sous le graphe de la fonction. Pourquoi chercher à remplacer cette notion d'intégrale ?

Le fait est que la notion d'intégrale enseignée avant la licence (appelée « intégrale de Riemann »), possède bien des défauts dès lors qu'on aborde des situations un peu plus complexes que celle d'une fonction continue $f : [a,b] \to \mathbf{C}$ définie sur un intervalle borné de $\mathbf{R}$.

Rappelons la définition : si on se donne une subdivision

$$a = y_0 < y_1 < \cdots < y_N = b \tag{0.1}$$

de $[a,b]$, on pose

$$S^+(f) = \sum_{i=1}^{N} (y_i - y_{i-1}) \sup_{y_{i-1} \leqslant x \leqslant y_i} f(x) \tag{0.2}$$

$$S^-(f) = \sum_{i=1}^{N} (y_i - y_{i-1}) \inf_{y_{i-1} \leqslant x \leqslant y_i} f(x), \tag{0.3}$$

et f est intégrable au sens de Riemann si

$$\lim S^+(f) = \lim S^-(f),$$

la limite étant prise quand le pas $\max |y_i - y_{i-1}|$ de la subdivision tend vers 0. Si c'est le cas, la limite commune est appelée l'intégrale de f entre a et b.

Voici quelques exemples des problèmes posés par cette définition, dont vous verrez que la « nouvelle » intégrale résout la plupart.

– Problèmes de passage à la limite : soit (f_n) une suite de fonctions continues sur $[a,b]$ « convergeant » vers une fonction f. Selon le sens de « convergence » (c'est à dire selon la topologie utilisée), il n'est pas toujours vrai que

$$\lim_{n \to +\infty} \int_a^b f_n(t)dt = \int_a^b f(t)dt, \tag{0.4}$$

ni même que f est intégrable.

Cette formule est valide, et facilement vérifiée, si $f_n \to f$ uniformément sur $[a,b]$: dans ce cas on sait que la limite f est continue, et que (0.4) est vraie. Mais cette restriction est trop limitante, car la convergence uniforme n'est pas si fréquente, et est souvent difficile à vérifier puisque c'est une propriété globale, et non pas ponctuelle. Par contre, si l'on suppose seulement qu'il y a convergence simple, c'est à dire $f_n(x) \to f(x)$ pour tout $x \in [a,b]$, on construit aisément des suites où pour $n \geqslant 1$ on a $\int f_n(x)dx = 1$ mais $\int f(x)dx = 0$. Par exemple, si $[a,b] = [0,1]$ et

$$f_n(x) = \begin{cases} 2n^2x & \text{si } 0 \leqslant x \leqslant \frac{1}{2n} \\ 2n - 2n^2x & \text{si } \frac{1}{2n} \leqslant x \leqslant \frac{1}{n} \\ 0 & \text{si } \frac{1}{n} \leqslant x \leqslant 1 \end{cases} \tag{0.5}$$

on a $f_n(x) \to 0$ pour tout $x \in [0,1]$ (car $f_n(0) = 0$ et la suite est constante pour $n > x^{-1}$ si $x \in]0,1]$) mais

$$\int_0^1 f_n(x)dx = 1$$

pour tout n. De tels exemples illustrent la necessité d'au moins certaines conditions pour pouvoir échanger limite et intégrale.

Un problème évidemment lié est que les différentes normes associées à l'intégration, notablement celle donnée par le produit scalaire

$$(f \mid g) = \int_a^b f(t)\overline{g(t)}dt$$

ne donnent pas un espace métrique complet sur l'espace des fonctions intégrables au sens ci-dessus.

– Problèmes pour les dimensions supérieures : de même que l'intégrale

$$\int_a^b f(t)dt$$

a une interprétation géométrique quand $f(t) \geqslant 0$ sur $[a,b]$ comme aire du domaine plan située entre l'axe des x et le graphe de f, on a besoin de calculer le volume d'un solide (par exemple, une boule), ou l'aire de la surface qu'il borde (par exemple, une sphère). La définition de l'intégrale de Riemann n'est pas du tout adaptée à ce cadre, puisqu'il n'est pas possible de trouver une manière « raisonnable » et universelle de faire des subdivisions finies, par exemple, d'un solide arbitraire.

Toujours dans ce cadre se pose le problème de l'interversion des intégrales multiples : est-ce que

$$\int_a^b \int_c^d f(x,y)dxdy = \int_c^d \int_a^b f(x,y)dydx \ ? \tag{0.6}$$

– Interprétation probabiliste : si l'on veut donner un sens mathématique précis à des énoncés intuitifs tels que « un nombre pris au hasard entre 0 et 1 à une probabilité $1/2$ d'être $> 1/2$ », on est vite tenté de dire que la probabilité que $x \in [0,1]$ vérifie une propriété $\mathcal{P}$ est

$$(b-a) = \int_a^b dx$$

si l'ensemble $I(\mathcal{P})$ des $x \in [0,1]$ vérifiant $\mathcal{P}$ est égal à $[a,b]$. Mais très vite on arrive à construire des propriétés d'apparence naturelle pour lesquelles $I(\mathcal{P})$ n'est pas un intervalle, ni même une réunion disjointe finie d'intervalles ! Par exemple, soit $\mathcal{P}$ la propriété : « il n'y a pas de 1 dans le développement $x = 0, d_1 d_2 \cdots$ de x en base 3, $d_i \in \{0,1,2\}$.» Quelle est la probabilité que cela soit vrai ?

– Intégrales infinies : dans la méthode de Riemann, on définit les intégrales « impropres » du type

$$\int_a^{+\infty} f(t)dt$$

a posteriori comme limites d'intégrales finies de a à b quand $b \to +\infty$. Cela pose ensuite bien des problèmes puisque toutes les propriétés désirées doivent être vérifiées séparément à partir de cette définition.

Lebesgue lui-même a mis l'accent sur le second point. L'idée de sa solution est très simple : selon ses propres mots, il y a deux manières de compter combien on a de monnaie en poche. Soit on sort chaque pièce l'une après l'autre, et on additionne au fur et à mesure, soit on sort toutes les pièces, on les groupe en tas suivant leurs valeurs, et on ajoute la contribution de chaque tas.

En d'autres termes, si $a_1, \ldots, a_m$ sont m nombres réels pouvant prendre un nombre fini $v_1, \ldots, v_r$ de valeurs, on a la formule

$$S = \sum_{n=1}^{m} a_n = \sum_{i=1}^{r} v_i N_i \tag{0.7}$$

où $N_i = |\{n \mid a_n = v_i\}|$.[2]

La première expression est vue comme analogue de l'intégrale de Riemann, alors que la seconde sera analogue de celle de Lebesgue. Une différence essentielle est que la première dépend *a priori* de l'ordre dans lequel les a_n interviennent (c'est l'analogue de la subdivision d'un intervalle qui utilise l'ordre dans $\mathbf{R}$), alors que la seconde n'en dépend pas. Il suffit de savoir « compter » combien de pièces ont une valeur donnée.

L'idée de l'intégrale de Lebesgue, pour une fonction $f : X \to \mathbf{C}$ (où X peut être un ensemble quelconque) prenant un nombre fini de valeurs v_i, sera la formule

$$\int_X f(x)dx = \sum_{i=1}^{r} v_i I_i \tag{0.8}$$

où

$$I_i = \text{la « mesure » de l'ensemble des } x \in X \text{ tels que } f(x) = v_i.$$

remplace le nombre N_i ci-dessus.

La première étape nécessaire pour donner un sens à une telle formule est d'expliciter ce que le mot « mesure » sous-entend. En quelque sorte, on déplace le problème de définir l'intégrale de la « verticale », l'espace des valeurs de la fonction, vers « l'horizontale », l'espace des points où la fonction est définie.

Notations usuelles

Dans la théorie de la mesure on sera amené à faire de l'arithmétique avec des éléments $x \in [0, +\infty]$, c'est à dire soit $x \geqslant 0$ est un nombre réel, soit $x = +\infty$. Les règles sont les suivantes :

(1) L'addition et la multiplication de $x \in \mathbf{R}$ sont celles habituelles.
(2) On a $x + \infty = \infty + x = \infty$ pour tout $x \geqslant 0$.
(3) On a $x \cdot \infty = \infty \cdot x = \infty$ *si* $x > 0$.
(4) On a $0 \cdot \infty = \infty \cdot 0 = 0$.

La règle (4) en particulier est surprenante de prime abord, on en verra rapidement la justification.

On vérifie toute de suite que ces règles donnent sur $[0, +\infty]$ deux opérations commutatives, associatives et que la multiplication est distributive par rapport à l'addition. De plus la relation d'ordre usuelle s'étend aussi avec $a \leqslant +\infty$ pour tout $a \in [0, +\infty]$ et $a < +\infty$ si et seulement si $a \in [0, +\infty[\subset \mathbf{R}$. On a les régles

$$a \leqslant b, c \leqslant d \text{ impliquent } a + c \leqslant b + d \text{ et } ac \leqslant bd$$

si a, b, c et $d \in [0, +\infty]$. Par contre, $a + b = a + c$ ou $ab = ac$ n'impliquent pas forcément $b = c$ si $a = +\infty$.

Remarque. Attention ! Il est interdit d'utiliser $-$ et $/$ tant qu'on opère avec des éléments pouvant être $= +\infty$!

Il y a sur $[0, +\infty]$ une topologie évidente pour laquelle la « convergence vers $+\infty$ » a son sens habituel : les voisinages de $+\infty$ sont les complémentaires des ensembles bornés $X \subset \mathbf{R}$.

Il est pratique de remarquer que, si $a_n \in [0, +\infty]$, la série $\sum a_n$ converge toujours dans $[0, +\infty]$, et sa somme est réelle (et non pas infinie) si et seulement si les sommes partielles sont bornées : autrement dit, « une série à coefficients positifs converge toujours dans $[0, +\infty]$ ».

[2]. Dans ce texte, la notation $|X|$, si X est un ensemble, signifie le cardinal de X.

Rappelons la définition de la limite supérieure et inférieure d'une suite (a_n) de nombres réels

$$\limsup_n a_n = \lim_{k\to+\infty} \sup_{n\geqslant k} a_n, \quad \liminf_n a_n = \lim_{k\to+\infty} \inf_{n\geqslant k} a_n \ ;$$

ces limites existent dans $[-\infty, +\infty]$ comme limites de suites monotones (décroissante pour la limite supérieure, croissante pour la limite inférieure). La suite (a_n) converge vers x si et seulement si $\limsup_n a_n = x = \liminf_n a_n$.

Enfin, si $a_n \leqslant b_n$ on a

$$\limsup_n a_n \leqslant \limsup_n b_n \text{ et } \liminf_n a_n \leqslant \liminf_n b_n. \tag{0.9}$$

Les limites supérieures et inférieures permettent souvent de raisonner avec une suite réelle (a_n) « comme si » elle convergeait, le but étant souvent de montrer que c'est effectivement le cas, mais en évitant de parler de la limite de la suite avant de savoir qu'elle existe effectivement...

Par exemple, le « théorème des gendarmes » s'obtient aisément par le formalisme ci-dessus : si $b_n \leqslant a_n \leqslant c_n$ avec $\lim b_n = \lim c_n = a \in \mathbf{R}$, si l'on ne sait pas que a_n converge on peut néanmoins prendre la limite supérieure pour déduire

$$a = \lim b_n = \limsup b_n \leqslant \limsup a_n \leqslant \limsup c_n = \lim c_n = a,$$

et la limite inférieure

$$a = \lim b_n = \liminf b_n \leqslant \liminf a_n \leqslant \liminf c_n = \lim c_n = a,$$

de sorte que $a = \liminf a_n = \limsup a_n$, ce qui veut dire que (a_n) converge vers a.

Le lecteur constatera que l'on n'a pas cherché une consistance parfaite dans la manière de raisonner avec les limites, préférant illustrer suivant l'humeur du moment des techniques un peu diverses.

La notion de relation d'équivalence et d'ensemble quotient apparaît dans la théorie des espaces L^p. Rappelons que si X est un ensemble, une relation d'équivalence $\sim$ sur X est une relation entre éléments de X, notée $x \sim y$, telle que les propriétés

$$x \sim x, \quad x \sim y \text{ si et seulement si } y \sim x, \quad x \sim y \text{ et } y \sim z \text{ impliquent } x \sim z$$

caractéristiques de la relation d'égalité sont vérifiées. Un exemple est fourni, si Y est un ensemble quelconque et $f : X \to Y$ une application par la relation $x \sim_f y$ si et seulement si $f(x) = f(y)$.

Étant donnée une relation d'équivalence $\sim$ et $x \in X$, la *classe d'équivalence* $\pi(x)$ de x est

$$\pi(x) = \{y \in X \mid y \sim x\}.$$

On note $X/\sim$ et on appelle espace quotient de X par $\sim$ l'ensemble des classes d'équivalences, et alors π devient une application $\pi : X \to X/\sim$, dont la propriété fondamentale est que $\pi(x) = \pi(y)$ *si et seulement si* $x \sim y$, autrement dit π permet de transcrire $\sim$ en relation d'égalité. De plus, π est surjective.

Pour construire une application $f : Y \to X/\sim$, où Y est un ensemble quelconque, il *suffit* de construire une application $f_1 : Y \to X$ et de poser $f = \pi \circ f_1$. Pour construire une application de source $X/\sim$, $g : X/\sim \to Y$, il est équivalent de construire $g_1 : X \to Y$ telle que $x \sim y$ implique $g_1(x) = g_1(y)$. En effet, dans ce cas la valeur $g_1(x)$ ne dépend que de la classe d'équivalence $\pi(x)$ et l'on peut poser $g(\pi(x)) = g(x)$ pour $\pi(x) \in X/\sim$. L'application g est dite *induite* par g_1, ou bien on dit que g_1 se factorise par $X/\sim$.

Si V est un espace vectoriel, et si $W \subset V$ est un sous-espace vectoriel on note V/W l'espace quotient de V par la relation $x \sim y$ si et seulement si $x - y \in W$. Les applications induites par

$$+ : (x, y) \mapsto x + y \text{ et } \cdot : (\lambda, x) \mapsto \lambda x$$

sont bien définies car W est un espace vectoriel, et munissent le quotient V/W d'une structure d'espace vectoriel tel que l'application $\pi : V \to V/W$ est une application linéaire surjective.

On utilisera les notations suivantes :

(1) Si X est un ensemble, $|X| \in [0, +\infty]$ est son cardinal, avec $|X| = \infty$ si X est infini. On ne distingue pas ici entre les différents cardinaux infinis.

(2) Si E et F sont des espaces vectoriels sur $\mathbf{R}$ ou $\mathbf{C}$, on note $L(E, F)$ l'espace vectoriel des applications linéaires de E dans F et $L(E) = L(E, E)$.

(3) Le coefficient binomial donnant le nombre de sous-ensembles à k éléments d'un ensemble à n éléments est noté
$$\binom{n}{k}$$
et correspond donc à ce qui est aussi souvent noté C_n^k.

CHAPITRE 1

Théorie de la mesure

Soit X un ensemble quelconque. On veut donc fournir un sens à la « mesure » $\mu(Y)$ d'un sous-ensemble $Y \subset X$. La théorie de la mesure fournit le cadre pour ce type de questions. De même que la généralisation de la notion de convergence a amené à introduire les espaces métriques, puis les espaces topologiques, qui font appel à des structures supplémentaires sur l'ensemble donné X, soumises à certains axiomes, la mise en place rigoureuse de la notion de mesure se fait dans un cadre axiomatique qui met en valeur le fait qu'il n'y a pas de définition nécessairement « canonique » de la mesure : pour un même espace, il peut exister plusieurs interprétations (comme il peut exister plusieurs métriques). On tournera rapidement cette abondance à notre avantage.

Un point délicat à première vue est la nécessité de postuler également l'existence de certaines parties de X qui sont mesurables : le plus souvent, il n'est pas possible de définir $\mu(Y)$ pour tout $Y \subset X$. En particulier, il n'est pas possible de définir la mesure ($\geqslant 0$) de tout sous-ensemble $Y \subset \mathbf{R}$ de sorte que les conditions suivantes soient vérifiées :

– On a $\mu([a, b]) = b - a$;
– On a $\mu(Y) = \mu(Z)$ si Z est obtenu à partir de Y par translation ;
– On a

$$\mu\Big(\bigcup_{n \geqslant 1} Y_n\Big) = \sum_n \mu(Y_n)$$

si les Y_n sont des sous-ensembles disjoints de $\mathbf{R}$.

En se limitant à certains sous-ensembles, dont les intervalles font partie, on parviendra néanmoins à avoir ces trois conditions.

Une interprétation du besoin de spécifier les ensembles mesurables provient des probabilités : si l'on veut que $\mu(Y)/\mu(X)$ corresponde à la probabilité qu'un point $x \in X$ pris au hasard soit dans Y, on peut bien s'imaginer qu'on sache si peu de choses sur Y qu'il ne soit pas possible d'attribuer une probabilité à cet événement. Par exemple, si un dispositif expérimental est limité par sa précision à mesurer des valeurs > 1 d'une quantité X, il ne permettra jamais de répondre à la question « Quelle est la probabilité que $X < 0.5$? », ce qu'on pourra modéliser en disant que cette condition définit un ensemble non-mesurable.

Dans la suite, nous introduirons le langage probabiliste en même temps que le langage « analytique » dans la théorie de l'intégration. Cela apporte souvent une intuition supplémentaire qui peut être fort utile.

1.1. Tribus, algèbres, etc

Définition 1.1.1. Soit X un ensemble.

(1) Une tribu sur X, aussi appelée σ-algèbre, est une famille $\mathcal{M}$ de sous-ensembles de X tels que

(i) On a $\emptyset \in \mathcal{M}$, $X \in \mathcal{M}$.
(ii) Si $Y \in \mathcal{M}$, alors le complémentaire $X - Y \in \mathcal{M}$.
(iii) Si $Y_n \in \mathcal{M}$ pour $n \geqslant 1$, alors

$$\bigcup_{n \geqslant 1} Y_n \in \mathcal{M} \text{ et } \bigcap_{n \geqslant 1} Y_n \in \mathcal{M}. \tag{1.1}$$

Un ensemble $Y \in \mathcal{M}$ est dit *mesurable* pour $\mathcal{M}$. Le couple $(X, \mathcal{M})$ est appelé un *espace mesurable.*

(2) Soit $(X, \mathcal{M})$ et $(X', \mathcal{M}')$ des ensembles munis de tribus. Une *application mesurable* $f : X \to X'$ est une application telle que

$$f^{-1}(Y) \in \mathcal{M} \text{ pour tout } Y \in \mathcal{M}'.$$

REMARQUE 1.1.2. En plus des propriétés ci-dessus, noter que si Y, Z sont mesurables, alors $Y - Z = Y \cap (X - Z)$ est également mesurable.

Comme pour les topologies, il existe d'innombrables exemples. Par contre, l'exemple « intéressant » est beaucoup plus difficile à construire.

Notons tout de suite un lemme évident mais important:

LEMME 1.1.3. (1) *Soit $(X, \mathcal{M})$ un espace mesurable. Alors l'application identité $(X, \mathcal{M}) \to (X, \mathcal{M})$ est mesurable.*

(2) *Toute application constante est mesurable, quelles que soient les tribus à la source et au but.*

(3) *Soient $X \xrightarrow{f} X'$ et $X' \xrightarrow{g} X''$ des applications mesurables. Alors la composée $g \circ f : X \to X''$ est mesurable.*

DÉMONSTRATION. (1) et (2) sont trivialissimes, et pour (3) il suffit de remarquer que

$$(g \circ f)^{-1}(Y'') = f^{-1}(g^{-1}(Y'')).$$

$\square$

EXEMPLE 1.1.4. (1) Si X est quelconque, on peut prendre $\mathcal{M} = \{X, \emptyset\}$.

(2) On peut aussi prendre pour $\mathcal{M}$ l'ensemble de toutes les parties de X. Dans ce cas, toute application $X \to (X', \mathcal{M}')$ est mesurable.

(3) Si $\mathcal{M}'$ est une tribu sur X', et $f : X \to X'$ est une application quelconque, alors

$$\mathcal{M} = f^{-1}(\mathcal{M}') = \{f^{-1}(Y) \mid Y \in \mathcal{M}'\}$$

est une tribu, dite image réciproque sur X. En effet, on a

$$\begin{gathered} f^{-1}(\emptyset) = \emptyset, \quad f^{-1}(X' - Y') = X - f^{-1}(Y'), \\ f^{-1}\Big(\bigcup_{i \in I} Y_i\Big) = \bigcup_i f^{-1}(Y_i), \quad f^{-1}\Big(\bigcap_{i \in I} Y_i\Big) = \bigcap_i f^{-1}(Y_i) \end{gathered} \tag{1.2}$$

pour tout ensemble I. (Noter que l'image directe $f(\mathcal{M})$ n'est pas, en général, une tribu sur X' ; par exemple $f(X)$ n'est pas X' si f n'est pas surjective, donc $X' \notin f(\mathcal{M})$ dans ce cas.)

Noter que f est par définition mesurable si X est muni de $f^{-1}(\mathcal{M}')$.

Comme cas particulier, soit $(X, \mathcal{M})$ un espace mesurable et $X' \subset X$. Alors X' muni de $\mathcal{M}' = \{Y \cap X' \mid Y \in \mathcal{M}\}$ est un espace mesurable. Il s'agit en effet de la tribu image réciproque $i^{-1}(\mathcal{M})$ où $i : X' \hookrightarrow X$ est l'inclusion. Si $X' \in \mathcal{M}$, on vérifie aussitôt qu'on a plus simplement

$$\mathcal{M}' = \{Y \in \mathcal{M} \mid Y \subset X'\},$$

mais ce n'est pas vrai si $X' \notin \mathcal{M}$.

(4) Si $(\mathcal{M}_i)_{i \in I}$ sont des tribus sur X alors l'intersection $\bigcap_i \mathcal{M}_i$ est une tribu sur X.

(5) Si $Y \subset X$ est un sous-ensemble quelconque, alors Y est mesurable si et seulement si la *fonction caractéristique* $\chi_Y : X \to \{0, 1\}$ de Y

$$\chi_Y(x) = \begin{cases} 1 & \text{si } x \in Y \\ 0 & \text{sinon} \end{cases}$$

est mesurable, pour la tribu triviale $\{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$ sur $\{0, 1\}$. Il s'agit pratiquement d'une tautologie puisque $\chi_Y^{-1}(\{0\}) = X - Y$ et $\chi_Y^{-1}(\{1\}) = Y$.

Remarque 1.1.5. Dans le langage probabiliste, on note souvent $X = \Omega$ et $\mathcal{M} = \Sigma$. Un élément $\omega \in \Omega$ est appelé une expérience, et $A \subset \Sigma$ un événement, correspondant aux expériences vérifiant certaine propriété.

L'intersection de deux événements correspond au « et » logique, et la réunion au « ou ». Ainsi, par exemple, si A_n, $n \geqslant 1$ sont des événements/propriétés, on peut dire que l'événement « tout les A_n sont vrais » en est encore un puisque l'intersection des A_n est mesurable, et de même l'événement « au moins un des A_n est vrai » est mesurable.

Des exemples plus intéressants proviennent, assez implicitement, de la construction suivante.

Définition 1.1.6. (1) Soit X un ensemble et $\mathcal{A}$ une famille quelconque de parties de X. La *tribu engendrée* par $\mathcal{A}$ est la plus petite tribu contenant $\mathcal{A}$, c'est à dire l'intersection de toutes les tribus $\mathcal{M}$ telles que $\mathcal{A} \subset \mathcal{M}$. On note cette tribu $\sigma(\mathcal{A})$.

(2) Soit $(X, \mathcal{T})$ un espace topologique. La *tribu borélienne* $\mathcal{B}$ sur X est la tribu engendrée par la topologie $\mathcal{T}$, c'est à dire par l'ensemble des ouverts de X.

(3) Si $(X, \mathcal{M})$ et $(X', \mathcal{M}')$ sont des ensembles mesurés, la tribu produit sur $X \times X'$ est la tribu $\mathcal{M} \otimes \mathcal{M}'$ engendrée par les parties de la forme $Y \times Y'$ avec $Y \in \mathcal{M}$ et $Y' \in \mathcal{M}'$.

Remarque 1.1.7. (1) Si $(X, \mathcal{T})$ est un espace topologique, on vérifie aussitôt que $\mathcal{B}$ est engendrée, indifféremment, par les ouverts ou par les fermés. Si $X = \mathbf{R}$ avec la topologie ordinaire, on peut remarquer que $\mathcal{B}$ contient tout les intervalles $[a, b]$, $[a, b[$, $[a, +\infty[$, etc. Par exemple

$$[a, b] = \mathbf{R} - (] - \infty, a[\cup]b, +\infty[), \text{ et }]a, b] = [a, b]\cap]a, +\infty[. \tag{1.3}$$

De plus, la tribu borélienne sur $\mathbf{R}$ est engendrée par les intervalles ouverts ou fermés, voire simplement par les intervalles de la forme $]a, \infty[$ pour $a \in \mathbf{R}$. Cela provient du fait que tout ouvert est une réunion dénombrable d'intervalles (ses composantes connexes), et de relations telles que (1.3), de sorte que la tribu engendrée par ces intervalles contient tout les ouverts, donc la tribu borélienne.

De manière similaire, la tribu borélienne sur $[0, +\infty]$ est engendrée par les $]a, +\infty]$ pour tout $a \in [0, +\infty]$.

En général, si X est un espace topologique et qu'il n'est pas fait explicitement mention d'une tribu, les notions de mesurabilité seront relatives à $\mathcal{B}$. En particulier, pour tout espace mesurable $(X, \mathcal{M})$, une fonction mesurable réelle est une application mesurable $f : X \to \mathbf{R}$ où $\mathbf{R}$ est muni de la tribu borélienne. De même pour $\mathbf{C}$, etc...

(2) Si $(X, \mathcal{M})$ et $(X', \mathcal{M}')$ sont des espaces mesurés, on peut voir que $\mathcal{M} \otimes \mathcal{M}'$ est la plus petite tribu sur $X \times X'$ telle que les projections

$$p_1 : X \times X' \to X \text{ et } p_2 : X \times X' \to X'$$

soient mesurables.

En effet, ces deux projections sont mesurables pour une tribu $\mathcal{N}$ sur $X \times X'$ si et seulement si $X \times Y' \in \mathcal{N}$ et $Y \times X' \in \mathcal{N}$ pour $Y \in \mathcal{M}$ et $Y' \in \mathcal{M}'$. Mais la tribu engendrée par les ensembles de cette forme est la tribu $\mathcal{M} \otimes \mathcal{M}'$ puisque

$$Y \times Y' = (Y \times X') \cap (X \times Y').$$

(3) La tribu borélienne sur $\mathbf{C} = \mathbf{R}^2$ est identique à la tribu $\mathcal{B} \otimes \mathcal{B}$ où $\mathcal{B}$ est la tribu borélienne sur $\mathbf{R}$: cela provient du fait que tout ouvert dans $\mathbf{C}$ est réunion dénombrable du type $I_1 \times I_2$ où $I_i \subset \mathbf{R}$ est un intervalle ouvert dont les extrémités sont rationnelles. De plus on vérifie aussitôt que la restriction de la tribu borélienne sur $\mathbf{C}$ à $\mathbf{R} \subset \mathbf{C}$ est la tribu borélienne sur $\mathbf{R}$ car l'inclusion $\mathbf{R} \hookrightarrow \mathbf{C}$ est continue (cf.le Corollaire 1.1.9 ci-dessous).

(4) Lorsqu'une fonction mesurable $(X, \mathcal{M}) \to \mathbf{C}$ est donnée sans précision supplémentaire sur la tribu sur le but $\mathbf{C}$, celle-ci est toujours supposée être la tribu borélienne. En particulier,

en probabilité, une application mesurable $\Omega \to \mathbf{C}$ est appelée une *variable aléatoire*, $\mathbf{C}$ étant muni de la tribu borélienne.[1]

La définition d'une tribu engendrée, et en particulier de la tribu borélienne, est très implicite, mais elle est néanmoins assez maniable, par exemple grâce au lemme suivant.

Lemme 1.1.8. (1) *Soit $(X, \mathcal{M})$ et $(X', \mathcal{M}')$ des espaces mesurés tels que $\mathcal{M}'$ est engendrée par $\mathcal{A}$. Alors $f : X \to X'$ est mesurable si et seulement si $f^{-1}(\mathcal{A}) \subset \mathcal{M}$.*

(2) *En particulier, si $(X, \mathcal{M})$ est un espace mesurable, une fonction $f : X \to \mathbf{R}$ est mesurable si et seulement si $f^{-1}(]a, +\infty[) \in \mathcal{M}$ pour tout $a \in \mathbf{R}$.*

(3) *Si $(X, \mathcal{M})$, $(X', \mathcal{M}')$ et $(X'', \mathcal{M}'')$ sont des espaces mesurés alors $f : X'' \to X \times X'$ est mesurable pour la tribu produit sur $X \times X'$ si et seulement si $p_1 \circ f : X'' \to X$ et $p_2 \circ f : X'' \to X'$ sont mesurables.*

Démonstration. (2) est un cas particulier de (1). Pour ce premier point, il suffit de vérifier que

$$f^{-1}(\sigma(\mathcal{A})) = \sigma(f^{-1}(\mathcal{A})), \tag{1.4}$$

puisqu'alors on aura

$$f^{-1}(\mathcal{M}') = f^{-1}(\sigma(\mathcal{A})) = \sigma(f^{-1}(\mathcal{A})) \subset \sigma(\mathcal{M}) = \mathcal{M}.$$

Le membre de gauche de (1.4) est une tribu (image réciproque) contenant $f^{-1}(\mathcal{A})$ donc contenant le membre de droite qui est la tribu engendrée par cette famille.

Pour la réciproque, on remarque que

$$\mathcal{M}'' = \{Y \mid f^{-1}(Y) \in \sigma(f^{-1}(\mathcal{A}))\}$$

est une tribu sur X' (cf. (1.2)), et qu'elle contient $\mathcal{A}$ donc $\sigma(\mathcal{A})$. De sorte que

$$f^{-1}(\sigma(\mathcal{A})) \subset f^{-1}(\mathcal{M}'') \subset \sigma(f^{-1}(\mathcal{A})),$$

comme désiré.

Pour (3), par composition $p_1 \circ f$ et $p_2 \circ f$ sont mesurables si f l'est (cf. Remarque 1.1.7, (2)). Réciproquement, il suffit de vérifier que $f^{-1}(Y \times Y') \in \mathcal{M}''$ pour tout $Y \in \mathcal{M}$, $Y' \in \mathcal{M}'$ d'après le point (1). Or $f(x) = (p_1 \circ f(x), p_2 \circ f(x))$ et

$$f^{-1}(Y \times Y') = (p_1 \circ f)^{-1}(Y) \cap (p_2 \circ f)^{-1}(Y')$$

d'où le résultat. □

Corollaire 1.1.9. (1) *Soit $f : X \to X'$ une application continue entre espaces topologiques. Alors f est mesurable pour les tribus boréliennes sur X et X'.*

(2) *Soit $(X, \mathcal{M})$ un espace mesurable et f, $g : X \to \mathbf{C}$ des applications mesurables. Alors $f \pm g$ et fg sont mesurables, et si g ne s'annule pas, $1/g$ est mesurable. En particulier, l'ensemble des fonctions mesurables complexes sur X est un espace vectoriel.*

(3) *Une fonction $f : X \to \mathbf{C}$ est mesurable pour la tribu borélienne sur $\mathbf{C}$ si et seulement si $\mathrm{Re}(f)$ et $\mathrm{Im}(f)$ sont mesurables comme fonctions $X \to \mathbf{R}$.*

Démonstration. Le point (1) provient directement du Lemme 1.1.8, (1). Quand à (2), on a par exemple $f + g = p \circ (f \times g)$ où

$$p : \mathbf{C} \times \mathbf{C} \to \mathbf{C}$$

est l'addition et $f \times g : x \mapsto (f(x), g(x))$. D'après le Lemme 1.1.8, (3), l'application $f \times g$ est mesurable, et d'après (1), l'application p est mesurable puisque continue. Donc par composition $f + g$ est également mesurable, et similairement pour $f - g$, fg et $1/g$.

Enfin, (3) est un cas particulier de (2) d'après la remarque 1.1.7, (3). □

[1]. Dans la littérature mathématique une variable aléatoire est souvent définie comme étant à valeurs réelles, mais cela ne fait pas une grande différence d'après la remarque précédente.

Exemple 1.1.10. Dans le langage probabiliste, si $X : \Omega \to \mathbf{C}$ est une variable aléatoire, un événement du type $X^{-1}(Y)$ est noté plutôt simplement $\{X \in Y\}$. On note similairement, par exemple,

$$\{X > a\} = \{\omega \mid X(\omega) > a\}$$

pour $a \in \mathbf{R}$, et si on a $f : \mathbf{C} \to \mathbf{C}$ est une fonction mesurable on écrira volontiers $f(X)$ la variable aléatoire $Y = f \circ X$.

Jusqu'à présent, la présence de l'axiome (iii) dans la définition d'une tribu n'est pas intervenu de manière essentielle. Sa présence s'explique par l'importance des suites dans l'analyse. On en déduit en effet le lemme suivant qui est fondamental :

Lemme 1.1.11. *Soit $(X, \mathcal{M})$ un espace mesurable.*

(1) *Si (f_n), $n \geqslant 1$, est une suite de fonctions réelles mesurables telles que $f_n(x) \to f(x)$ pour tout $x \in X$, c'est à dire que f_n converge ponctuellement vers f, alors la fonction limite f est mesurable.*

(2) *Plus généralement,* $\limsup f_n$, $\liminf f_n$, $\sup f_n$ *et* $\inf f_n$ *sont mesurables.*

On notera que même si les f_n sont continues, il n'y a en général aucune raison que leur limite simple f le soit ! La mesurabilité de f pour la tribu borélienne est donc une vraie propriété.

Démonstration. Puisque la limite de $f_n(x)$ est égale à sa limite supérieure si elle existe, il suffit de montrer (2). Puisque de plus on a

$$\limsup f_n(x) = \lim_k \sup_{n \geqslant k} f_n(x) = \inf_k \sup_{n \geqslant k} f_n(x),$$

(limite décroissante) il suffit de montrer le résultat pour $\inf f_n$ et $\sup f_n$ en toute généralité, et en remplaçant f_n par $-f_n$, il suffit de traiter la borne supérieure.

Soit $g(x) = \sup_{n \geqslant 1} f_n(x) \in [0, +\infty]$. D'après le Lemme 1.1.8, (1) et la remarque 1.1.7, (1), il suffit de montrer que pour tout $a \in [0, +\infty]$ on a

$$E_a = \{x \mid g(x) > a\} \in \mathcal{M}.$$

Mais on peut écrire cet ensemble comme suit :

$$E_a = \bigcup_n \{x \mid f_n(x) > a\}$$

(autrement dit, $g(x) = \sup f_n(x) > a$ si et seulement si il existe k tel que $f_k(x) > a$). Comme f_n est mesurable et que $\mathcal{M}$ est stable par union dénombrable (1.1), on a $E_a \in \mathcal{M}$. □

Remarque 1.1.12. (1) Si f et g sont des fonctions mesurables sur $(X, \mathcal{M})$, il découle du lemme en particulier que $\sup(f, g)$ et $\inf(f, g)$ sont également mesurables.

À partir de là, on voit que si f est à valeurs réelles sa valeur absolue

$$|f| = \sup(f, -f) \tag{1.5}$$

est mesurable, et que

$$f^+ = \sup(f, 0), \text{ et } f^- = -\inf(f, 0) \tag{1.6}$$

le sont également. Noter que $f^+ \geqslant 0$, $f^- \geqslant 0$ et on a

$$f = f^+ - f^- \text{ et } |f| = f^+ + f^- \tag{1.7}$$

une expression pratique de f (resp. $|f|$) comme différence (resp. somme) de deux fonctions positives mesurables.

Alternativement, les fonctions $(x, y) \mapsto \sup(x, y)$ et $(x, y) \mapsto \inf(x, y)$ sont continues sur $\mathbf{R}^2$, donc mesurables pour la tribu borélienne (Corollaire 1.1.9, (1)), et la mesurabilité de $\sup(f, g)$ ou $\inf(f, g)$ en découle par composition (Lemme 1.1.3, (3)).

(2) Si (f_n) est une suite quelconque de fonctions mesurables complexes sur X, alors

$$Y = \{x \in X \mid f_n(x) \text{ converge }\}$$

est mesurable. En effet, on peut écrire

$$Y = \bigcap_{k\geqslant 1} \bigcup_{N\geqslant 1} \bigcap_{n,m\geqslant N} \{x \mid |f_n(x) - f_m(x)| < 1/k\} \tag{1.8}$$

en recopiant le critère de Cauchy (appliqué aux $\varepsilon = 1/k$) dans $\mathbf{C}$, puisque $|f_n - f_m|$ est mesurable.

1.2. Mesure sur une tribu

Ayant défini les ensembles mesurables, sous la forme de tribus, la notion de mesure de ces ensembles n'est également pas définie de façon unique, mais par d'autres axiomes.

Définition 1.2.1. (1) Soit $(X, \mathcal{M})$ un espace mesurable. Une mesure μ sur $(X, \mathcal{M})$ est une application

$$\mu : \mathcal{M} \to [0, +\infty]$$

telle que $\mu(\emptyset) = 0$ et

$$\mu\Big(\bigcup_{n\geqslant 1} Y_n\Big) = \sum_{n\geqslant 1} \mu(Y_n) \tag{1.9}$$

pour toute famille dénombrable d'ensembles $Y_n \in \mathcal{M}$ deux à deux disjoints. Le triplet $(X, \mathcal{M}, \mu)$ est appelé un espace mesuré.

(2) Une mesure μ est dite *finie* si $\mu(X) < +\infty$, et σ-finie si on peut écrire X comme réunion dénombrable d'ensembles de mesure finie :

$$X = \bigcup_{n\geqslant 1} Y_n, \text{ avec } \mu(Y_n) < +\infty.$$

(3) Une mesure de probabilité est une mesure μ telle que $\mu(X) = 1$.

Remarque 1.2.2. Si μ est une mesure finie et vérifie la condition $\mu(X) > 0$, $\mu'(Y) = \mu(Y)/\mu(X)$ pour $Y \in \mathcal{M}$ donne une mesure de probabilité. Cela signifie que la théorie des mesures finies est pratiquement équivalente à celle des mesures de probabilité.

Dans le langage probabiliste, on écrit plutôt P la mesure et on dit que $P(E) \in [0, 1]$, pour $E \in \Sigma$, est la probabilité de l'événement E. Le triplet (Ω, Σ, P) est appelé un espace probabilisé.

Les mesures usuelles sont σ-finies : cela s'avère important pour les applications aux intégrales multiples, par exemple (cf. 4).

On déduit aussitôt de la définition les propriétés suivantes :

Proposition 1.2.3. *Soit μ une mesure sur $(X, \mathcal{M})$. On a :*

(1) *Si Y, $Z \in \mathcal{M}$ avec $Y \subset Z$, alors $\mu(Y) \leqslant \mu(Z)$, et plus précisément*

$$\mu(Z) = \mu(Y) + \mu(Z - Y).$$

(2) *Si Y, $Z \in \mathcal{M}$, on a*

$$\mu(Y \cup Z) + \mu(Y \cap Z) = \mu(Y) + \mu(Z).$$

(3) *Si $Y_1 \subset \cdots \subset Y_n \subset \cdots$ est une suite croissante d'ensembles mesurables, alors*

$$\mu\Big(\bigcup_{n} Y_n\Big) = \lim_{n\to+\infty} \mu(Y_n).$$

(4) *Si $Y_1 \supset \cdots \supset Y_n \supset \cdots$ est une suite décroissante d'ensembles mesurables, et si $\mu(Y_1) < +\infty$, alors*

$$\mu\Big(\bigcap_{n} Y_n\Big) = \lim_{n\to+\infty} \mu(Y_n).$$

(5) *Si (Y_n) est une suite d'ensembles mesurables, alors*

$$\mu\Big(\bigcup_{n\geqslant 1} Y_n\Big) \leqslant \sum_{n} \mu(Y_n).$$

Démonstration. Pour chacun de ces quatre points, un dessin pattatoïdique bien fait, laissé aux soins du lecteur, en dira plus long que la preuve très exacte qui suit.

(1) : il suffit d'écrire la réunion disjointe

$$Z = Y \cup (Z - Y)$$

pour déduire $\mu(Z) = \mu(Y) + \mu(Z - Y) \geqslant \mu(Y)$ puisque μ ne prend que des valeurs positives.

(2) : similairement, on a les unions disjointes

$$Y \cup Z = Y \cup (Z - Y) \text{ et } Z = (Z - Y) \cup (Z \cap Y),$$

donc $\mu(Y \cup Z) = \mu(Y) + \mu(Z - Y)$ et $\mu(Z) = \mu(Z - Y) + \mu(Z \cap Y)$, ce qui donne bien $\mu(Y \cup Z) = \mu(Y) + \mu(Z) - \mu(Z \cap Y)$.

(3) : posons $Z_1 = Y_1$ puis $Z_n = Y_n - Y_{n-1}$ pour $n \geqslant 2$. On a donc

$$Y = Y_1 \cup Y_2 \cup \cdots = Z_1 \cup Z_2 \cup \cdots$$

et les Z_i sont disjoints puisque la suite (Y_n) est croissante : pour $j > 0$ on a

$$Z_{i+j} \cap Z_i \subset Z_{i+j} \cap Y_i \subset Z_{i+j} \cap Y_{i+j-1} = \emptyset.$$

Donc par (1.9)

$$\mu(Y) = \sum_n \mu(Z_n) = \lim_{k \to +\infty} \sum_{1 \leqslant n \leqslant k} \mu(Z_n);$$

comme on a de même

$$\mu(Y_k) = \sum_{1 \leqslant n \leqslant k} \mu(Z_n),$$

le résultat suit.

(4) : ce résultat est dual du précédent, mais attention la condition $\mu(Y_1) < +\infty$ est nécessaire (prendre par exemple $Y_n = \{k \geqslant n\}$ dans $X = \{n \geqslant 1\}$ avec la mesure de comptage décrite ci-dessous ; l'intersection des Y_n est vide, bien que chaque Y_n ait une mesure infinie).

On pose $Z_n = Y_1 - Y_n$; les Z_n sont croissants et mesurables donc

$$\mu\Bigl(\bigcup_n Z_n\Bigr) = \lim_n \mu(Y_1 - Y_n);$$

à gauche on a

$$\bigcup_n Z_n = Y_1 - \bigcap_n Y_n$$

À droite on a $\mu(Y_1 - Y_n) = \mu(Y_1) - \mu(Y_n)$ car $Y_n \subset Y_1$, et donc le résultat en découle en simplifiant $\mu(Y_1)$ (ce qui n'est pas possible si $\mu(Y_1) = +\infty$).

(5) : La réunion Y des Y_n peut se voir comme la réunion disjointe des Z_n définis par récurrence par $Z_1 = Y_1$ et

$$Z_{n+1} = Y_{n+1} - \bigcup_{1 \leqslant i \leqslant n} Y_i$$

pour $n \geqslant 1$. On $\mu(Z_n) \leqslant \mu(Y_n)$ par monotonie (1), et par (1.9) on en déduit

$$\mu(Y) = \sum_n \mu(Z_n) \leqslant \sum_n \mu(Y_n).$$

□

Voici les exemples simples de mesures.

Exemple 1.2.4. (1) Soit X un ensemble, $\mathcal{M}$ la tribu contenant toutes les parties de X. Alors

$$\mu(Y) = |Y|$$

définit une mesure sur $(X, \mathcal{M})$, dite mesure de comptage. La propriété d'additivité dénombrable (1.9) est immédiate dans ce cas.

(2) Soit X un ensemble, $\mathcal{M}$ la tribu contenant toutes les parties de X, et $x_0 \in X$ fixé. Alors la mesure δ_{x_0} définie par

$$\delta_{x_0}(Y) = \begin{cases} 1 & \text{si } x_0 \in Y \\ 0 & \text{sinon} \end{cases}$$

est une mesure sur $(X, \mathcal{M})$ appelée mesure de Dirac en x_0. La formule (1.9) résulte alors de ce qu'un au plus des ensembles disjoints Y_n peut contenir x_0.

(3) Si X est un ensemble fini, la mesure

$$\mu(Y) = \frac{|Y|}{|X|}$$

définie sur la tribu de tout les sous-ensembles de X, est une mesure de probabilité. C'est la base des probabilités dites discrètes ; noter que pour cette mesure tout singleton $\{x\}$ a la même mesure

$$\mu(\{x\}) = \frac{1}{|X|},$$

on parle d'équiprobabilité de chacune des expériences.

(4) Soit $X = \mathbf{N}$, avec la tribu $\mathcal{M}$ formée de tout les sous-ensembles de $\mathbf{N}$ (c'est aussi, trivialement, la tribu induite par la tribu borélienne sur $\mathbf{R}$). Alors il n'existe pas de mesure de probabilité uniforme sur X, c'est à dire qu'il n'existe pas de mesure P sur $\mathcal{M}$ telle que $P(\mathbf{N}) = 1$ et $P(n + A) = P(A)$ pour tout $A \subset \mathbf{N}$, où $n + A = \{n + a \mid a \in A\}$ est le translaté de A par n.

En effet, si P existe, on aurait $P(\{n\}) = P(n + \{0\}) = P(0)$ pour tout n, et alors par additivité

$$P(A) = \sum_{a \in A} P(\{a\}) = |A| P(0)$$

pour tout A fini. Si $P(0) \neq 0$, en choisissant A assez grand on aurait $P(A) > 1$, ce qui est impossible pour une mesure de probabilité, et si $P(0) = 0$, on aurait $P(A) = 0$ pour tout A fini, donc $P(A) = 0$ pour tout A par additivité dénombrable.

Bien entendu, il existe toutefois beaucoup de mesures de probabilité sur $\mathbf{N}$. Plus exactement, réécrivant de manière plus positive ce qui précède, on voit qu'il est équivalent de se donner une telle mesure P sur $(\mathbf{N}, \mathcal{M})$ ou bien une suite $(p_k)_{k \in \mathbf{N}}$ de nombres réels $p_k = in[0, 1]$ tels que

$$\sum_{k \geqslant 0} p_k = 1,$$

la correspondance étant donnée par $p_k = P(\{k\})$ dans un sens et

$$P(Y) = \sum_{y \in Y} p_k$$

dans le sens opposé. Pour donner un exemple concret, $p_k = 2^{-k-1}$ convient.

Soit $(X, \mathcal{M}, \mu)$ un espace mesuré. Les ensembles de mesure (ou de probabilité) nulle sont importants. Notons

$$\mathcal{M}_0(\mu) = \{Y \in \mathcal{M} \mid \mu(Y) = 0\}$$

(qui dépend de μ).

Bien que $\mathcal{M}_0(\mu)$ ne soit pas une tribu (puisque $\mu(X) \neq 0$ en général !), elle vérifie les conditions de stabilités par union et intersection dénombrable (voir la Proposition 1.2.3, (5) pour l'union). Intuitivement, un ensemble de mesure nulle est un ensemble exceptionnel qui n'a pas d'incidence sur le comportement global des objets étudiés du point de vue de l'intégration. Conformément à cette intuition, on fait la définition suivante :

Définition 1.2.5. Soit $(X, \mathcal{M}, \mu)$ un espace mesuré. Un ensemble μ-négligeable est un ensemble $Y \in \mathcal{M}_0(\mu)$. Si $\mathcal{P}(x)$ est une propriété ayant un point $x \in X$ comme paramètre, on dit que $\mathcal{P}$ est vraie presque partout, ou que $\mathcal{P}$ est presque sûre en langage probabiliste, si

$$\{x \in X \mid \mathcal{P}(x) \text{ est faux}\} \in \mathcal{M}_0(\mu).$$

Remarque 1.2.6. Il paraît vite souhaitable, et conforme à l'intuition, que si $Y \in \mathcal{M}_0(\mu)$ est de mesure nulle et $Z \subset Y$, on ait $Z \in \mathcal{M}_0(\mu)$ (un sous-ensemble d'un ensemble négligeable devrait être lui-même négligeable). Cela n'est pas vrai en général. Une mesure μ vérifiant cette condition est dite *complète*.

On a le résultat de complétion suivant, qui permet de supposer presque toujours que l'on travaille avec une mesure complète : soit

$$\begin{aligned} \mathcal{M}'_0 &= \{Y \subset X \mid Y \subset Z \text{ avec } Z \in \mathcal{M}_0(\mu)\} \\ \mathcal{M}' &= \{Y \subset X \mid Y = Y_1 \cup Y_0 \text{ avec } Y_0 \in \mathcal{M}'_0 \text{ et } Y_1 \in \mathcal{M}\}, \end{aligned}$$

et posons $\mu'(Y) = \mu(Y_1)$ pour $Y = Y_0 \cup Y_1 \in \mathcal{M}'$. Alors on vérifie aisément que $\mathcal{M}'$ est une tribu contenant $\mathcal{M}$ et μ' est une mesure sur $\mathcal{M}'$ étendant μ, qui est complète.

Par exemple, pour vérifier que le complémentaire d'un ensemble $Y = Y_0 \subset Y_1$ est dans $\mathcal{M}'$, avec $Y_0 \subset Z_0$, $Z_0 \in \mathcal{M}_0(\mu)$, on se ramène au cas où $Z_0 \cap Y_1 = \emptyset$ en remplaçant si besoin est Y_0 et Z_0 par $Y_0 - Y_1$ et $Z_0 - Z_1$, et alors on a

$$X - Y = (X - (Y_1 \cup Z_0)) \cup (Z_0 - Y_0) \in \mathcal{M}'.$$

Pour vérifier que μ' est bien définie, si $Y = Y_1 \cup Y_0 = Y'_1 \cup Y'_0$, $Y_0 \subset Z_0 \in \mathcal{M}_0(\mu)$ et $Y'_0 \subset Z'_0 \in \mathcal{M}_0(\mu)$, on a $Y_1 \subset Y'_1 \cup Z'_0 \in \mathcal{M}$ donc par monotonie (les deux ensembles sont mesurables), on déduit

$$\mu(Y_1) \leqslant \mu(Y'_1) + \mu(Z'_0) = \mu(Y'_1)$$

et de même $\mu(Y'_1) \leqslant \mu(Y_1)$.

On peut construire des mesures à partir d'autres.

Proposition 1.2.7. *Soit $(X, \mathcal{M}, \mu)$ un espace mesurable.*

(1) *Si $\mu_1, \ldots, \mu_n$ sont des mesures sur $(X, \mathcal{M})$ et $\alpha_i \in [0, +\infty[$ des nombres réels positifs, alors $\mu = \sum \alpha_i \mu_i$ définie par*

$$\mu(Y) = \sum_{1 \leqslant i \leqslant n} \alpha_i \mu_i(Y)$$

pour $Y \in \mathcal{M}$ est une mesure sur $(X, \mathcal{M})$.

(2) *Soit $f : X \to X'$ une application mesurable. Posons*

$$f_*(\mu)(Y) = \mu(f^{-1}(Y)) \text{ pour } Y' \in \mathcal{M}'.$$

Alors $f_*(\mu)$ est une mesure sur X', appelée image directe de μ, et notée parfois $f(\mu)$. Pour $g : X' \to X''$ également mesurable, on a

$$(g \circ f)_*(\mu) = g_*(f_*(\mu)). \tag{1.10}$$

(3) *Soit $Y \subset X$ un ensemble mesurable. La mesure μ restreinte aux sous-ensembles mesurables $Z \subset Y$ est une mesure sur Y pour la tribu restreinte à Y.*

Démonstration. La vérification de ces propriétés est immédiate. Pour le second point, on utilise encore (1.2) pour vérifier (1.9). La formule (1.10) provient de la relation supplémentaire

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

pour $Y'' \subset X''$. ∎

Remarque 1.2.8. Dans le langage probabiliste, si $X : \Omega \to \mathbf{C}$ est une variable aléatoire à valeurs complexes, la mesure image directe $X(P)$ est appelée la *loi de probabilité* de la variable X. Cette notion est capitale car l'espace Ω est souvent laissé assez implicite, et l'on se donne seulement un certain nombre de variables aléatoires X_i (sensées, par exemple, modéliser un phénomène réel) avec des hypothèses sur leurs lois μ_i. Celles-ci sont des mesures sur l'espace topologique $\mathbf{C}$ qui est beaucoup plus concret, et l'essentiel est que la connaissance de la loi μ_i de X_i suffit à répondre à toutes les questions de probabilité concernant les valeurs de X_i : par définition

$$P(X_i \text{ vérifie } \mathcal{P}) = \mu_i(\{z \in \mathbf{C} \mid \mathcal{P}(z) \text{ est vraie}\})$$

pour toute propriété $\mathcal{P}(z)$ pouvant s'appliquer à $z \in \mathbf{C}$ (telle que l'ensemble à droite soit mesurable, bien entendu).

La notion suivante est purement probabiliste.

Définition 1.2.9. Soit (Ω, Σ, P) un espace probabilisé, A et $B \in \Sigma$ des événements.

(1) La probabilité conditionnelle de A sachant B est

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

si $P(B) \neq 0$.

(2) Les événements A et B sont dit indépendants si $P(A \cap B) = P(A)P(B)$.

(3) Si X_1, X_2 sont des variables aléatoires sur Ω à valeurs dans $\mathbf{C}$ alors X_1 et X_2 sont dites indépendantes si $\{X_1 \in C\} = X_1^{-1}(C)$ et $\{X_2 \in D\} = X_2^{-1}(D)$ sont indépendants pour tout C, $D \in \mathcal{B}_{\mathbf{C}}$.

(4) Plus généralement, des événements A_i, $i \in I$ sont indépendants si pour toute famille finie $A_{i(1)}, \ldots, A_{i(n)}$ on a

$$P(A_{i(1)} \cap \ldots \cap A_{i(n)}) = P(A_{i(1)}) \cdots P(A_{i(n)}),$$

et des variables aléatoires X_i, $i \in I$, sont dites indépendantes si pour tout $C_i \in \mathcal{B}_{\mathbf{C}}$, les événements $(\{X_i \in C_i\})_i$ sont indépendants.

Remarque 1.2.10. Si $P(B) = 0$, on a $P(A \cap B) = 0$ par monotonie de la mesure, et donc B est indépendant de A pour tout choix de A.

Pour simplifier la vérification et la construction d'événements et de variables aléatoires indépendantes, on utilise souvent le résultat élémentaire suivant :

Proposition 1.2.11. (1) *Soit (Ω, Σ, P) un espace probabilisé et $\mathcal{A} \subset \Sigma$ un sous-ensemble d'événements tels que $\sigma(\mathcal{A}) \supset \Sigma$. Dans les définitions ci-dessus, il suffit de vérifier les conditions d'indépendance pour des événements dans $\mathcal{A}$.*

(2) *Soit (X_n), $n \leqslant N$ des variables aléatoires indépendantes, et $\varphi_n : \mathbf{C} \to \mathbf{C}$ des applications mesurables. Alors les variables aléatoires $Y_n = \varphi(X_n)$ sont indépendantes.*

Démonstration. Le point (1) est élémentaire : pour prendre l'exemple de deux variables aléatoires X et Y dont on veut montrer l'indépendance, on considère l'ensemble des paires (C, D) d'événements tels que $\{X \in C\}$ et $\{Y \in D\}$ sont indépendants. Pour C (resp. D) fixé, on vérifie que c'est une tribu contenant $\mathcal{A}$, donc contenant $\sigma(\mathcal{A}) \supset \Sigma$, et le résultat en découle.

Quand à (2), on a tout simplement pour tout événements C_n

$$(Y_1, \ldots, Y_N) \in C_1 \times \cdots \times C_N$$

si et seulement si

$$(X_1, \ldots, X_N) \in \varphi_1^{-1}(C_1) \times \cdots \times \varphi_N^{-1}(C_N)$$

et le résultat provient directement de l'indépendance des X_n. □

Exercice 1.2.12. Généraliser ce dernier résultat à des situations plus complexes :

(1) Soient X_1, ..., X_4 des variables aléatoires indépendantes. Montrer que

$$P((X_1, X_2) \in C \text{ et } (X_3, X_4) \in D) = P((X_1, X_2) \in C)P((X_3, X_4) \in D)$$

pour tout C et D dans la tribu produit $\Sigma \otimes \Sigma$. (Indication : procéder comme pour la Proposition 1.2.11, (1)).

(2) Soient $\varphi_i : \mathbf{C}^2 \to \mathbf{C}$, $i = 1, 2$, des applications mesurables. Montrer que $\varphi_1(X_1, X_2)$ et $\varphi_2(X_3, X_4)$ sont indépendantes.

(3) Énoncer et démontrer une généralisation utilisant un nombre arbitraire de variables. (Voir la Section 4.2 pour la « bonne » approche de ce type de résultats).

1.3. La mesure de Lebesgue

Le théorème fondamental de la théorie de la mesure, sans lequel ce ne serait guère qu'une gentille plaisanterie, est l'existence de la mesure de Lebesgue dans $\mathbf{R}$. Sous forme préliminaire, le résultat peut s'énoncer ainsi :

Théorème 1.3.1. *Il existe une tribu $\mathcal{M}$ sur $\mathbf{R}$, contenant la tribu borélienne $\mathcal{B}_{\mathbf{R}}$, et une mesure λ sur $\mathcal{M}$ tels que $\lambda([a, b]) = b - a$ pour tout réels $a \leqslant b$. La mesure λ est appelée la mesure de Lebesgue. C'est une mesure complète.*

On raffinera cet énoncé dans la suite (unicité, propriétés de régularité, etc...) On peut déjà voir que λ n'est pas une mesure finie, mais qu'elle est σ-finie puisque

$$\mathbf{R} = \bigcup_{n \in \mathbf{Z}} [n, n + 1]$$

La mesure de Lebesgue restreinte à $[0, 1]$ fournit un premier exemple intéressant de mesure de probabilité.

Exemple 1.3.2. (1) On a $\lambda(\mathbf{N}) = \lambda(\mathbf{Q}) = 0$. En effet, par définition on a

$$\lambda(\{x\}) = \lambda([x, x]) = 0 \text{ pour tout } x \in \mathbf{R},$$

et donc par additivité de la mesure

$$\lambda(\mathbf{N}) = \sum_{n \in \mathbf{N}} \lambda(\{n\}) = 0,$$

et de même pour $\mathbf{Q}$ puisque $\mathbf{Q}$ est dénombrable : il existe une suite r_n, $n \geqslant 1$, tel que chaque rationnel $x = a/b$ est de la forme $r_{n(x)}$ pour un unique $n(x) \geqslant 1$.

(2) Soit $X = [0, 1]$ avec la mesure de Lebesgue λ restreinte à X. Soit $b \geqslant 2$ un entier fixé. Pour tout $x \in [0, 1]$ il existe un développement de x en base b

$$x = \sum_{i=1}^{+\infty} d_i b^{-i} \text{ avec } d_i \in \{0, \ldots, b - 1\},$$

noté $x = 0, d_1 d_2 \ldots$.

Ce développement n'est pas unique si $x > 0$ est rationnel à dénominateur une puissance de b, mais on suppose choisi dans ce cas l'un des deux développements possibles. Le choix n'a pas d'importance du point de vue de la théorie de la mesure puisque ces rationnels forment un ensemble de mesure nulle.

Pour tout $i \geqslant 1$, on obtient ainsi une application

$$X_i \quad \begin{cases} [0, 1] \to \{0, \ldots, b - 1\} \\ x \mapsto d_i \end{cases}$$

qui est mesurable : en effet, l'image réciproque par X_i de $d \in \{0, \ldots, b - 1\}$ est simplement une réunion de b^{i-1} intervalles, chacun de longueur b^{-i}.

Vues comme variables aléatoires sur $([0,1], \mathcal{B}, \lambda)$, les X_i donnent un exemple important d'une suite de variables aléatoires indépendantes. De plus, celles-ci sont identiquement distribuées, c'est à dire que la loi de X_i, $(X_i)_*(\lambda)$ ne dépend pas de i, plus précisément $(X_i)_*(\lambda)$ est la mesure uniforme P sur $\{0, \ldots, b-1\}$, telle que

$$P(d) = \frac{1}{b} \text{ pour tout } d.$$

On peut maintenant répondre à la question mentionnée dans l'introduction : « Quelle est la probabilité p qu'il n'y ait pas de 1 dans le développement en base 3 d'un réel $x \in [0,1]$ pris au hasard ?»

L'interprétation que l'on en fait est que $p = \lambda(C)$ où

$$C = \{x \in [0,1] \mid X_i(x) \neq 1 \text{ pour tout } i \geqslant 1\}.$$

Pour calculer p on remarque que

$$C = \bigcap_n C_n \text{ où } C_n = \{x \in [0,1] \mid X_i(x) \neq 1 \text{ pour tout } i \leqslant n\}.$$

La suite C_n est une suite décroissante de boréliens, donc $p = \lambda(C) = \lim \lambda(C_n)$, et d'après ce qui a été dit précédemment, C_n est réunion de 2^n intervalles de longueur 3^{-n}, donc d'après les propriétés de la mesure de Lebesgue on a $\lambda(C_n) = (2/3)^n$. Ainsi $p = 0$.

L'ensemble C est connu sous le nom d'ensemble de Cantor. C'est un compact (intersection de compacts) non dénombrable totalement discontinu.

Nous renvoyons à plus tard la démonstration du Théorème 1.3.1. Il est intéressant d'observer combien l'essentiel de la difficulté du cours d'intégration est entièrement cachée dans cet énoncé. Tout le reste, et en particulier la construction de l'intégrale, sera relativement simple et direct.

Cette approche est aussi justifiée parce que le Théorème 1.3.1 est une « boîte noire » remarquablement efficace : si on l'admet, il n'est pas besoin de connaître les détails de la construction de la mesure λ pour bâtir une théorie extrêmement utile. Les propriétés générales de toute mesure, et bien entendu surtout (1.9), suffisent pour la plupart des besoins à ce niveau...

Tout aussi bien, la preuve étant plutôt difficile et subtile, elle est plus aisée à comprendre avec un bagage technique et une habitude du type de raisonnements nécessaires qui ferait défaut à ce point du cours, alors qu'il se construira naturellement lors de la suite.

1.4. Mesures boréliennes et propriétés de régularité

Nous terminons ce chapitre assez abstrait par quelques dernières définitions qui introduisent des notions relatives aux mesures définies sur les tribus boréliennes des espaces topologiques.

Définition 1.4.1. (1) Soit X un espace topologique. Une mesure borélienne sur X est une mesure μ pour la tribu borélienne $\mathcal{B}_X$.

(2) Soit μ une mesure borélienne sur X. On dit que μ est *régulière* si, pour tout borélien $Y \subset X$ on a

$$\mu(Y) = \inf\{\mu(U) \mid U \text{ est un ouvert contenant } Y\} \tag{1.11}$$

$$\mu(Y) = \sup\{\mu(K) \mid K \text{ est un compact contenu dans } Y\}. \tag{1.12}$$

La régularité, quand elle est vraie, permet de relier la mesure d'un ensemble mesurable quelconque avec celle, souvent plus facile à calculer, des ouverts et des compacts.

La construction de la mesure de Lebesgue fournira la précision utile suivante :

Théorème 1.4.2. *La mesure de Lebesgue λ est régulière.*

Remarque 1.4.3. Si l'on admet cela, on obtient une caractérisation de la mesure de Lebesgue : on a pour U ouvert

$$\lambda(U) = \sum_i (b_i - a_i) \text{ si } U = \bigcup_{i \geqslant 1}]a_i, b_i[$$

est la décomposition de U en composantes connexes, et $\lambda(Y)$ est donnée par (1.11) pour $Y \in \mathcal{M}$. On peut effectivement démontrer le Théorème 1.3.1 en prenant cette définition et en *posant* $Y \in \mathcal{M}$ si le nombre $\lambda(Y)$ ainsi défini vérifie de plus (1.12).

La vérification que cela donne une tribu et une mesure dénombrablement additive est bien entendu loin d'être évidente.

Plus généralement, on a un critère très simple de régularité.

THÉORÈME 1.4.4. *Soit X un espace localement compact tel que tout ouvert $U \subset X$ est réunion dénombrable de compacts. Si μ est une mesure borélienne sur X telle que $\mu(K) < +\infty$ pour tout $K \subset X$ compact, alors μ est régulière.*

Le résultat annoncé sur la mesure de Lebesgue est un cas particulier, puisque si $K \subset \mathbf{R}$ est compact on a $K \subset [-M, M]$ pour un certain $M \in [0, +\infty[$, donc $\lambda(K) \leqslant \lambda([-M, M]) = 2M$.

CHAPITRE 2

Intégration

La construction de l'intégrale va maintenant procéder en toute généralité à partir de la donnée d'un espace mesuré $(X, \mathcal{M}, \mu)$. La progression se fera depuis les fonctions (mesurables) prenant comme valeurs 0 et 1 (fonctions caractéristiques d'ensembles mesurables, pour lesquelles la mesure donne l'intégrale), vers les fonctions prenant un nombre fini de valeurs (positives), vers les fonctions positives par passage à la limite, vers les fonctions générales par décomposition en parties réelles et imaginaires et (1.7) pour les fonctions réelles. Cette dernière étape seulement donnera une restriction aux fonctions dites « intégrables ».

Dans ce chapitre, toute fonction est supposée mesurable sauf mention du contraire.

2.1. Fonctions étagées et leur intégrale

DÉFINITION 2.1.1. Une fonction étagée sur un ensemble X est une fonction ne prenant qu'un nombre fini de valeurs.

En d'autres termes, une fonction étagée est une fonction $s : X \to \mathbf{C}$ qui peut s'écrire

$$s = \sum_{i=1}^{n} \alpha_i \chi_{Y_i} \tag{2.1}$$

où les $\alpha_i \in \mathbf{C}$ sont les différentes valeurs de f et $Y_i = \{x \in X \mid s(x) = \alpha_i\}$ sont des sous-ensembles disjoints de X. Cette écriture est clairement unique et $s \geqslant 0$ si et seulement si $\alpha_i \geqslant 0$ pour tout i.

Si X est un espace mesurable, s est mesurable si et seulement si $Y_i \in \mathcal{M}$ pour tout i (cf. Exemple 1.1.4, (5)).

L'ensemble $E(X)$ des fonctions étagées mesurables est une $\mathbf{C}$-algèbre puisque

$$\chi_Y \chi_Z = \chi_{Y\cap Z} \text{ et } \chi_Y + \chi_Z = \chi_{Y\cap Z} + \chi_{Y\cup Z} = 2\chi_{Y\cap Z} + \chi_{(Y\cup Z)-(Y\cap Z)}. \tag{2.2}$$

Procédant de la manière suggérée par Lebesgue dans l'introduction (cf. (0.7)), on définit l'intégrale de s par :

DÉFINITION 2.1.2. Soit $(X, \mathcal{M}, \mu)$ un espace mesuré, $s : X \to \mathbf{R}$ une fonction étagée positive donnée par (2.1). Soit $Y \in \mathcal{M}$, on pose

$$\int_Y s(x)d\mu(x) = \sum_{i=1}^{n} \alpha_i \mu(Y_i \cap Y) \in [0, +\infty]$$

l'intégrale de s par rapport à μ sur Y.

Comme autres notations pour l'intégrale, on pourra trouver :

$$\int_Y s d\mu, \quad \int_Y s(x)d\mu, \quad \int_Y s(x)\mu, \text{ voire } \int_Y s(x)dx$$

s'il n'y a pas d'ambiguité sur μ dans le contexte (c'est le plus souvent le cas pour la mesure de Lebesgue sur $\mathbf{R}$ par exemple).

Noter que par définition on peut aussi écrire (voir (2.2))

$$\int_Y s(x)d\mu(x) = \int_X s(x)\chi_Y(x)d\mu(x)$$

L'avantage pratique d'écrire $d\mu(x)$ est d'expliciter la variable d'intégration lorsque f dépend de paramètres supplémentaires.

Remarque 2.1.3. Cette intégrale est dores et déjà beaucoup plus générale que celle de Riemann dans le cas où $X = [a, b]$ est un intervalle. L'exemple typique est le suivant, dû à Dirichlet : soit $X = [0,1]$ et f la fonction caractéristique de $\mathbf{Q} \cap [0,1]$. Du point de vue de l'intégrale de Riemann, c'est une fonction non-intégrable puisque pour toute subdivision (0.1) on a évidemment $S^+(f) = 1$ et $S^-(f) = 0$ car tout intervalle $[y_i, y_{i+1}]$ contient des points rationnels où $f(x) = 1$ et des points irrationnels où $f(x) = 0$. Les sommes de Riemann n'ont donc pas de limite commune.

Par contre, pour la théorie de Lebesgue, cette fonction est simplement une fonction étagée et donc

$$\int_{[0,1]} f(x) d\lambda(x) = \lambda(\mathbf{Q} \cap [0,1]) = 0$$

(Exemple 1.3.2, (1)).

Proposition 2.1.4. (1) *L'application* $E(X) \to [0, +\infty]$

$$\Lambda \,:\, s \mapsto \int_X s(x) d\mu(x)$$

vérifie $\Lambda(\alpha s + \beta t) = \alpha\Lambda(s) + \beta\Lambda(t)$ *pour* s, t *étagées positives et* α, $\beta \geqslant 0$. *De plus* $\Lambda(s) \geqslant 0$ *et* $\Lambda(s) = 0$ *si et seulement si* s *est nulle presque partout.*

(2) *Supposons* $s \geqslant 0$. *L'application*

$$\mu_s \,:\, Y \mapsto \int_Y s(x) d\mu(x)$$

est une mesure sur $\mathcal{M}$, *et pour* s, $t \geqslant 0$ *on a*

$$\mu_{s+t} = \mu_s + \mu_t \text{ c'est à dire } \int_Y (s+t) d\mu = \int_Y s d\mu + \int_Y t d\mu.$$

De plus si $\mu(Y) = 0$, *alors* $\mu_s(Y) = 0$.

Ces deux faits auront, bien évidemment, des généralisations exactement analogues pour des fonctions intégrables plus générales lorsque celles-ci seront définies.

On note aussi $\mu_s = s d\mu$.

Démonstration. (1) Comme $E(X)$ est engendré par les fonctions χ_Y avec $Y \in \mathcal{M}$, il suffit de prouver que

$$\int_X (\alpha\chi_Y + \beta\chi_Z) d\mu = \alpha \int_X \chi_Y d\mu + \beta \int_X \chi_Z d\mu,$$

autrement dit que

$$(\alpha + \beta)\mu(Y \cap Z) + \alpha\mu(Y - Z) + \beta\mu(Z - Y) = \alpha\mu(Y) + \beta\mu(Z),$$

ce qui découle de l'additivité de la mesure puisque on a les unions disjointes

$$Y = (Y \cap Z) \cup (Y - Z), \text{ et } Z = (Y \cap Z) \cup (Z - Y).$$

La positivité de Λ est évidente, et $\Lambda(s) = 0$ si s est nulle presque partout. Réciproquement, si $\Lambda(s) = 0$ on a

$$0 = \sum_i \alpha_i \mu(Y_i) \geqslant \alpha_i \mu(Y_i)$$

et donc $\mu(Y_i) = 0$ si $\alpha_i > 0$, de sorte que

$$\mu(\{x \mid s(x) \neq 0\}) = \sum_{\alpha_i > 0} \mu(Y_i) = 0.$$

(2) L'application μ_s est évidemment positive et $\mu_s(\emptyset) = 0$. Soient (Z_k), $k \geqslant 1$, une suite de sous-ensembles disjoints de X et Z leur réunion. On a par additivité dénombrable de μ :

$$\mu_s(Z) = \sum_{i=1}^{n} \alpha_i \mu(Z \cap Y_i) = \sum_{i=1}^{n} \alpha_i \sum_{k \geqslant 1} \mu(Z_k \cap Y_i)$$
$$= \sum_{k \geqslant 1} \sum_{i=1}^{n} \mu(Z_k \cap Y_i) = \sum_{k \geqslant 1} \mu_s(Z_k) \in [0, +\infty],$$

l'échange des deux sommes étant permis puisque la somme sur i est finie.

L'identité $\mu_{s+t}(Y) = \mu_s(Y) + \mu_t(Y)$ est simplement l'identité $\Lambda(s+t) = \Lambda(s) + \Lambda(t)$ de (1) (appliquée à Y au lieu de X).

Finalement, si $\mu(Y) = 0$, on a

$$\mu_s(Y) = \sum_i \alpha_i \mu(Y \cap Y_i) = 0.$$

□

2.2. Intégration de fonctions positives

Soit maintenant $f : X \to [0, +\infty]$ une fonction positive mesurable. On définit l'intégrale de f par approximation par des fonctions étagées positives, plus précisément :

Définition 2.2.1. On pose pour tout $Y \in \mathcal{M}$

$$\int_Y f d\mu = \sup\Big\{ \int_Y s d\mu \mid s \leqslant f \text{ est une fonction étagée} \Big\} \in [0, +\infty]$$

Ainsi l'intégrale de toute fonction positive est définie, mais elle peut être $+\infty$, et on remarque aussi que si f est elle-même étagée, les deux définitions données de l'intégrale coïncident.

Commençons par les propriétés les plus élémentaires.

Proposition 2.2.2. *On a :*

(1) On a $\int f d\mu = 0$ si et seulement si $f(x) = 0$ presque partout et $\int f d\mu < +\infty$ implique que $f(x) < +\infty$ presque partout.

(2) Si $\mu(Y) = 0$, alors

$$\int_Y f d\mu = 0 \text{ même si } f = +\infty \text{ sur } Y.$$

(3) On a

$$\int_Y f d\mu = \int_X f(x) \chi_Y(x) d\mu.$$

(4) Si $0 \leqslant f \leqslant g$, on a

$$\int_X f d\mu \leqslant \int_X g d\mu. \tag{2.3}$$

(5) Si $Y \subset Z$, alors

$$\int_Y f d\mu \leqslant \int_Z f d\mu.$$

(6) Si $\alpha \in [0, +\infty[$, alors

$$\int_Y \alpha f d\mu = \alpha \int_Y f d\mu.$$

Démonstration. (1) : si f est nulle sauf sur $Z \in \mathcal{M}_0(\mu)$, alors toute $s \leqslant f$ est nulle sauf sur Z, donc

$$\int_Y s(x) d\mu(x) = 0,$$

de sorte que l'intégrale de f est nulle. Réciproquement, si f n'est pas nulle presque partout, il existe un entier $k \geqslant 1$ tel que $Y_k = \{x \mid f(x) \geqslant 1/k\}$ soit de mesure $\mu(Y_k) > 0$ (sinon

$$\mu(\{x \mid f(x) > 0\}) = \lim_{k \to +\infty} \mu(Y_k) = 0$$

par Proposition 1.2.3), et alors la fonction $s = k^{-1}\chi_{Y_k}$ est une fonction étagée telle que $s \leqslant f$ et $\int s d\mu = k^{-1}\mu(Y_k) > 0$.

Si $f(x) = +\infty$ pour tout $x \in Y$ avec $\mu(Y) > 0$, les fonctions étagées $s_n = n\chi_Y$ vérifient $0 \leqslant s_n \leqslant f$ et $\int s_n d\mu = n\mu(Y) \to +\infty$ donc par contraposition, si $\int f d\mu < +\infty$, on doit avoir $\mu(\{x \mid f(x) = +\infty\}) = 0$.

Le point (2) est évident.

Pour (3), on a pour toute $s \leqslant f$

$$\int_Y s d\mu = \int_X s\chi_Y d\mu,$$

par définition et d'autre part toute fonction étagée $t \leqslant f\chi_Y$ est de la forme $t = s\chi_Y$ avec $s \leqslant f$ (prendre $s = t\chi_Y$), d'où le résultat.

Pour (4), il suffit d'observer que $s \leqslant f$ implique $s \leqslant g$ et pour (5) d'appliquer (3) et (4) avec $0 \leqslant f\chi_Y \leqslant f\chi_Z$.

Et pour (6), le résultat est vrai pour $\alpha = 0$, et si $\alpha > 0$ on a $s \leqslant f$ si et seulement si $\alpha s \leqslant \alpha f$ avec (Proposition 2.1.4, (1))

$$\int_Y (\alpha s) d\mu = \alpha \int_Y s d\mu.$$

□

Le premier résultat important est le théorème de convergence monotone, aussi appelé théorème de Beppo Levi.

Théorème 2.2.3. *Soit* (f_n), $n \geqslant 1$, *une suite* croissante *de fonctions positives mesurables, et* $f(x) = \lim f_n(x)$. *Alors* f *est mesurable et*

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

Démonstration. La mesurabilité de f est juste un rappel du Lemme 1.1.11. Puisque f_n est une suite croissante on a $f_n \leqslant f_{n+1} \leqslant f$ donc d'après (3) de la proposition précédente

$$\int_X f_n d\mu \leqslant \int_X f_{n+1} d\mu \leqslant \int_X f d\mu$$

ce qui montre que la suite $(\int f_n d\mu)$ est croissante, et que sa limite dans $[0, +\infty]$ vérifie

$$\lim_{n \to +\infty} \int_X f_n d\mu \leqslant \int_X f d\mu. \tag{2.4}$$

Pour montrer la réciproque, soit s une fonction étagée telle que $0 \leqslant s \leqslant f$. Soit $c \in]0, 1[$ un paramètre fixé. Notons

$$X_n = \{x \in X \mid f_n(x) \geqslant cs(x)\}.$$

pour $n \geqslant 1$. Les X_n sont des ensembles mesurables, croissants puisque (f_n) est croissante, et recouvrant X puisque $f_n(x) \to f(x)$ pour tout x (si $f(x) = 0$, $x \in X_1$, et sinon $cs(x) < f(x)$ donc il existe n tel que $f_n(x) > cs(x)$).

On a alors d'après les propriétés élémentaires ci-dessus

$$\int_X f_n d\mu \geqslant \int_{X_n} f_n d\mu \geqslant c \int_{X_n} s d\mu = c\mu_s(X_n).$$

Puisque μ_s est une mesure et que (X_n) est une suite croissante d'ensembles mesurables, on a $\mu_s(X) = \lim \mu_s(X_n)$ par la Proposition 1.2.3, (3), d'où

$$c\mu_s(X) \leqslant \lim_{n \to +\infty} \int_X f_n d\mu.$$

Faisant alors tendre c vers 1, on déduit

$$\mu_s(X) = \int_X s d\mu \leqslant \lim_{n\to+\infty} \int_X f_n d\mu$$

et en prenant la borne supérieure sur $s \leqslant f$, on a l'inégalité réciproque de (2.4) et le théorème. $\square$

Ce premier théorème permet de fournir une meilleure approximation de l'intégrale.

PROPOSITION 2.2.4. *Soit $f : X \to [0, +\infty]$ une fonction mesurable positive.*

(1) *Il existe une suite croissante (s_n) de fonctions étagées positives telle que $s_n(x) \to f(x)$ pour tout $x \in X$.*

(2) *Pour toute suite (s_n) comme en* (1), *on a*

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

DÉMONSTRATION. La seconde partie est juste une application directe du théorème de convergence monotone.

Pour construire la suite (s_n), fixons un entier $n \geqslant 1$, et posons alors simplement

$$s_n(x) = \begin{cases} n & \text{si } f(x) \geqslant n \\ \frac{i-1}{2^n} & \text{si } \frac{i-1}{2^n} \leqslant f(x) < \frac{i}{2^n} \text{ avec } 1 \leqslant i \leqslant n2^n. \end{cases}$$

Il est clair que s_n est une fonction étagée positive, mesurable car f est mesurable donc

$$f^{-1}([n, +\infty[) \in \mathcal{M} \text{ et } f^{-1}\Big(\Big[\frac{i-1}{2^n}, \frac{i}{2^n}\Big[\Big) \in \mathcal{M}$$

et que pour $n \geqslant 1$ on a $s_n \leqslant s_{n+1} \leqslant f$. D'autre part pour tout $x \in X$, si $f(x) = +\infty$, on a $s_n(x) = n \to +\infty$, et si $f(x) < +\infty$, on a

$$0 \leqslant f(x) - s_n(x) \leqslant \frac{1}{2^n}$$

pour $n > f(x)$, donc $s_n(x) \to f(x)$. $\square$

REMARQUE 2.2.5. On peut observer la simplicité enfantine de la construction de la suite s_n et sa grande généralité, qui est permise par la possibilité de considérer les ensembles *a priori* assez arbitraires $f^{-1}([(i-1)2^{-n}, i2^{-n}[)$. On peut comparer cela avec la structure beaucoup plus rigide nécessaire pour obtenir des fonctions en escalier comme celles utilisées pour l'intégrale de Riemann.

COROLLAIRE 2.2.6. (1) *Soient (f_n), $n \geqslant 1$, des fonctions mesurables positives, et soit*

$$f(x) = \sum_{n\geqslant 1} f_n(x) \in [0, +\infty].$$

Alors on a

$$\int_X f d\mu = \sum_{n\geqslant 1} \int_X f_n d\mu.$$

(2) *En particulier, si f et g sont mesurables et positives alors*

$$\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu. \tag{2.5}$$

(3) *Soit f une fonction mesurable positive, et soit*

$$\mu_f(Y) = \int_Y f d\mu$$

pour $Y \in \mathcal{M}$. Alors μ_f est une mesure sur $\mathcal{M}$ telle que $\mu(Y) = 0$ implique $\mu_f(Y) = 0$ et

$$\int_Y g d\mu_f = \int_Y g f d\mu \tag{2.6}$$

pour $Y \in \mathcal{M}$ et $g \geqslant 0$ mesurable.

(4) *Soit $\varphi : X \to X'$ une fonction mesurable. Pour toute fonction positive mesurable g sur X', et tout $Y \in \mathcal{M}'$*

$$\int_{\varphi^{-1}(Y)} (g \circ \varphi) d\mu = \int_Y g d\varphi_*(\mu). \tag{2.7}$$

Démonstration. Soient d'abord f et g positives. D'après la proposition précédente il existe des suites croissantes (s_n) et (t_n) de fonctions étagées positives telles que $s_n(x) \to f(x)$ et $t_n(x) \to g(x)$. La suite croissante $u_n = s_n + t_n$ converge alors vers $f + g$. Puisque s_n et t_n sont étagées on a

$$\int_X (s_n + t_n) d\mu = \int_X s_n d\mu + \int_X t_n d\mu$$

par la Proposition 2.1.4, (2). D'après le théorème de convergence monotone, on en déduit quand $n \to +\infty$

$$\int_X (f + g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

Par récurrence, cela se généralise et montre que

$$\int_X \sum_{n=1}^{N} f_n(x) d\mu(x) = \sum_{n=1}^{N} \int_X f_n(x) d\mu(x).$$

Comme les sommes partielles convergent en croissant vers les séries correspondantes quand $N \to +\infty$, une nouvelle application du théorème de convergence monotone prouve le (1).

Pour le (2), on sait que $\mu_f \geqslant 0$ et $\mu_f(\emptyset) = 0$ (Proposition 1.2.3). Il reste à montrer l'additivité dénombrable. Mais soient Y_n, $n \geqslant 1$, des ensembles mesurables disjoints et Y leur union. La suite

$$g_n = \sum_{1 \leqslant i \leqslant n} f \chi_{Y_i}$$

converge en tout point vers $f(x)\chi_Y(x)$ puisque les Y_n sont disjoints, et elle est croissante. D'après le théorème de convergence monotone on a donc

$$\mu_f(Y) = \int_Y f d\mu = \int_X f(x)\chi_Y(x) d\mu(x) = \lim_{n \to +\infty} \int_X g_n d\mu = \lim_{n \to +\infty} \sum_{1 \leqslant i \leqslant n} \mu_f(Y_i).$$

Si $\mu(Y) = 0$, on a $\mu_s(Y) = 0$ pour toute fonction étagée $s \leqslant f$, et donc $\mu_f(Y) = \sup \mu_s(Y) = 0$.

Quand à (2.6), on remarque que la formule est vraie par définition si g est une fonction étagée. Si s_n est une suite croissante convergeant vers g, la suite $t_n = s_n f \chi_Y$ converge en croissant vers $f g \chi_Y$, et donc d'après le théorème de convergence monotone (appliqué pour μ_f et μ)

$$\int_Y g d\mu_f = \lim_{n \to +\infty} \int_Y s_n d\mu_f = \lim_{n \to +\infty} \int_Y s_n f d\mu = \int_Y g f d\mu.$$

Enfin pour (4), si $s_n \to g$ en croissant, alors $s_n \circ \varphi \to g \circ \varphi$ en croissant, donc il suffit de supposer g étagée, ou par additivité $g = \chi_Z$ avec $Z \in \mathcal{M}'$. On a alors

$$\int_{\varphi^{-1}(Y)} (\chi_Z \circ \varphi) d\mu = \int_{\varphi^{-1}(Y)} \chi_{\varphi^{-1}(Z)} d\mu = \mu(\varphi^{-1}(Y \cap Z)) = \int_{Y \cap Z} d\varphi_*(\mu) = \int_Y \chi_Z d\varphi_*(\mu)$$

par définition de la mesure $\varphi_*(\mu)$. $\square$

2.3. Intégration de fonctions générales

On peut maintenant définir enfin les fonctions intégrables. Rappelons pour $f : X \to \mathbf{R}$ la décomposition $f = f^+ - f^-$ décrite en (1.6), pour laquelle $|f| = f^+ + f^-$.

Définition 2.3.1. Soit $(X, \mathcal{M}, \mu)$ un espace mesuré.

(1) Soit $f : X \to \mathbf{R}$ une fonction mesurable. On dit que f est intégrable si la fonction positive $|f| = f^+ + f^-$ vérifie

$$\int_X |f| d\mu < +\infty,$$

et on pose alors

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu \in \mathbf{R}.$$

(2) Soit $f : X \to \mathbf{C}$ une fonction mesurable. On dit que f est intégrable si $|f| \geqslant 0$ est intégrable et on pose

$$\int_X f d\mu = \int_X \mathrm{Re}(f) d\mu + i \int_X \mathrm{Im}(f) d\mu \in \mathbf{C},$$

de sorte que par définition

$$\mathrm{Re}\Big(\int_X f d\mu\Big) = \int_X \mathrm{Re}(f) d\mu \text{ et } \mathrm{Im}\Big(\int_X f d\mu\Big) = \int_X \mathrm{Im}(f) d\mu. \tag{2.8}$$

(3) On note $L^1(\mu)$ l'ensemble des fonctions complexes intégrables sur X.

Noter que ces définitions donnent effectivement des nombres, car $0 \leqslant f^{\pm} \leqslant |f|$ donc par monotonie (2.3) on a

$$\int_X f^{\pm} d\mu \leqslant \int_X |f| d\mu < +\infty,$$

si f est intégrable, et de même si f est à valeurs complexes, on a

$$|\mathrm{Re}(f)| \leqslant |f|, \text{ et } |\mathrm{Im}(f)| \leqslant |f|,$$

donc les parties réelles et imaginaires de f sont intégrables.

On peut remarquer aussi que

$$\int_X \bar{f} d\mu = \overline{\int_X f d\mu}. \tag{2.9}$$

Remarque 2.3.2. Si (Ω, Σ, P) est un espace probabilisé et X est une variable aléatoire complexe sur Ω, on note $E(X)$ l'intégrale de X sur Ω, et on dit que c'est *l'espérance* de X.

Si X a pour loi la mesure $\mu = X(P)$ sur $\mathbf{C}$ (cf. Remarque 1.2.8), on a (voir 2.7)

$$E(X) = \int_{\mathbf{C}} x d\mu$$

Proposition 2.3.3. (1) *L'ensemble $L^1(\mu)$ est un espace vectoriel, sur lequel l'application*

$$\begin{cases} L^1(\mu) \to [0, +\infty[\\ f \mapsto \displaystyle\int_X |f| d\mu \end{cases}$$

est une semi-norme, notée $\|f\|_1$, telle que $\|f\|_1 = 0$ si et seulement si f est nulle presque partout.

(2) *L'application*

$$\begin{cases} L^1(\mu) \to \mathbf{C} \\ f \mapsto \displaystyle\int_X f d\mu \end{cases}$$

est une forme linéaire de norme 1, positive au sens où $f \geqslant 0$ implique $\int f d\mu \geqslant 0$.

(3) *Pour toute application mesurable $\varphi : X \to X'$, et toute fonction intégrable g sur X' on a*

$$\int_{\varphi^{-1}(Y)} (g \circ \varphi) d\mu = \int_Y g d\varphi_*(\mu). \tag{2.10}$$

En termes concrets, cela signifie d'une part que $\|0\|_1 = 0$, $\|f+g\|_1 \leqslant \|f\|_1 + \|g\|_1$ et $\|\alpha f\|_1 = |\alpha| \|f\|_1$ pour f, $g \in L^1(\mu)$ et $\alpha \in \mathbf{C}$, et d'autre part que

$$\int_X (\alpha f + \beta g) d\mu = \alpha \int_X f d\mu + \beta \int_X g d\mu$$

pour f, $g \in L^1(\mu)$ et α, $\beta \in \mathbf{C}$, et enfin que

$$\Big| \int_X f d\mu \Big| \leqslant \int_X |f| d\mu, \tag{2.11}$$

propriétés qui bien entendu semblent parfaitement naturelles et dont il semblerait fort étrange de ne pas disposer.

Remarque 2.3.4. On note également parfois $\|f\|_1 = \int |f| d\mu \in [0, +\infty]$ pour f quelconque, y compris $f : X \to [0, +\infty]$. Notons que la condition $\|f\|_1 < +\infty$ implique que $|f|$, donc f aussi, ne peut prendre la valeur $+\infty$ sur un ensemble de mesure > 0.

Démonstration. Remarquons d'abord que

$$|\alpha f + \beta g| \leqslant |\alpha||f| + |\beta||g|$$

de sorte que par additivité et monotonie de l'intégrale pour les fonctions positives on a

$$\int_X |\alpha f + \beta g| d\mu \leqslant \int_X |\alpha||f| d\mu + \int_X |\beta||g| d\mu$$

pour toutes $f, g \in L^1(\mu)$ et $\alpha, \beta \in \mathbf{C}$. Cela montre que $\alpha f + \beta g \in L^1(\mu)$ et l'inégalité triangulaire (prendre $\alpha = \beta = 1$). Comme l'inégalité devient une égalité $|\alpha f| = |\alpha||f|$ si $g = 0$, on a aussi $\|\alpha f\|_1 = |\alpha| \|f\|_1$ par Proposition 2.2.2, (6).

Pour terminer avec (1), les fonctions telles que $\|f\|_1 = 0$ sont celles qui sont nulles presque partout par application directe du fait correspondant pour les fonctions positives, à savoir la Proposition 2.2.2, (1).

Pour montrer la linéarité de l'intégrale, il faut être un peu soigneux car les opérations $f \mapsto f^{\pm}$ ne sont *pas* linéaire...

On montre séparément pour $\alpha \in \mathbf{C}$ et f, $g \in L^1(\mu)$

$$\int \alpha f d\mu = \alpha \int f d\mu \tag{2.12}$$

$$\int (f+g) d\mu = \int f d\mu + \int g d\mu \tag{2.13}$$

D'après ce qui précède, toutes les fonctions sous le signe $\int$ sont effectivement intégrables.

Soit $\alpha \in \mathbf{R}$ et f à valeurs réelles. Si $\varepsilon \in \{-1, 1\}$ est le signe de α, avec la signification évidente pour $\varepsilon\pm$, on a

$$(\alpha f)^+ = (\varepsilon\alpha) f^{\varepsilon+}, \text{ et } (\alpha f)^- = (\varepsilon\alpha) f^{\varepsilon-}.$$

L'identité (2.12) en découle dans ce cas : par définition

$$\int_X \alpha f d\mu = \int_X (\varepsilon\alpha) f^{\varepsilon+} d\mu - \int_X (\varepsilon\alpha) f^{\varepsilon-} d\mu,$$

donc par Proposition 2.2.2, (6) il vient

$$\int_X \alpha f d\mu = \varepsilon\alpha \int_X f^{\varepsilon+} d\mu - \varepsilon\alpha \int_X f^{\varepsilon-} d\mu = \varepsilon\alpha \int_X (f^{\varepsilon+} - f^{\varepsilon-}) d\mu = \alpha \int_X f d\mu.$$

Quand à (2.13), posons $h = f + g$. Alors $h^+ - h^- = f^+ - f^- + g^+ - g^-$, que l'on réarrange pour avoir des sommes de fonctions positives $h^+ + f^- + g^- = f^+ + g^+ + h^-$, d'où par (2.5)

$$\int h^+ d\mu + \int f^- d\mu + \int g^- d\mu = \int f^+ d\mu + \int g^+ d\mu + \int h^- d\mu, \text{ donc } \int f d\mu + \int g d\mu = \int h d\mu.$$

Il reste à considérer le cas où $\alpha \in \mathbf{C}$ et f à valeurs complexes, qui s'obtiennent aisément par des manipulations du même type.

Finalement il faut montrer (2.11). Si f est à valeurs réelles, c'est évident puisque $|f| = f^+ + f^-$. Si f est à valeurs complexes, il faut une petite astuce. Notons $\theta \in \mathbf{R}$ un réel tel que

$$\int_X f d\mu = e^{i\theta} \Big| \int_X f d\mu \Big|.$$

On a alors (utilisant (2.12) et (2.8)

$$\Big| \int_X f d\mu \Big| = \int_X e^{-i\theta} f d\mu = \mathrm{Re}\Big(\int_X e^{-i\theta} f d\mu \Big) = \int_X \mathrm{Re}(e^{-i\theta} f) d\mu.$$

Mais $\mathrm{Re}(z) \leqslant |z|$ donc la dernière intégrale vérifie bien

$$\int_X \mathrm{Re}(e^{-i\theta} f) d\mu \leqslant \int_X |f| d\mu$$

par monotonie (2.3). Cela montre, strictement parlant, que l'application linéaire $f \mapsto \int f d\mu$ est de norme $\leqslant 1$. Mais on a égalité dans (2.11) si $f \geqslant 0$ est intégrable et pas presque partout nulle[1] donc la norme est bien 1.

L'identité (2.10) est la généralisation de (2.7) qui s'obtient directement par linéarité à partir de celle-ci. □

Nous avons donné tout les détails (ou presque...) de ces calculs assez fastidieux pour indiquer qu'il y a des choses à vérifier. La définition à l'aide de f^+ et f^- de l'intégrale va maintenant s'évanouir à peu près complètement pour être remplacée par des manipulations aisées des propriétés de linéarité usuelles.

Exemple 2.3.5. Considérons les mesures simples de l'Exemple 1.2.4 ; le cas de la mesure de Lebesgue est traité dans la Section 2.4.

(1) Si μ est la mesure de comptage sur X, on vérifie aussitôt que les fonctions intégrables sont les f telles que la famille $(f(x))$, $x \in X$ est (absolument) *sommable* et

$$\int_X f(x) d\mu = \sum_{x \in X} f(x).$$

En particulier si $X = \mathbf{N}$, cela donne une une interprétation des séries complexes absolument convergentes comme intégrale. Dans ce langage, le Corollaire 2.2.6 montre que si $a_{i,j} \geqslant 0$, alors

$$\sum_{i \geqslant 1} \sum_{j \geqslant 1} a_{i,j} = \sum_{j \geqslant 1} \sum_{i \geqslant 1} a_{i,j}.$$

(2) Si $\mu = \delta_{x_0}$ est la mesure de Dirac en x_0, toute fonction $f : X \to \mathbf{C}$ est intégrable et

$$\int_X f(x) d\delta_{x_0}(x) = f(x_0). \tag{2.14}$$

Plus généralement, si $x_1, \ldots, x_n \in X$, on peut former la mesure de probabilité

$$\delta = \frac{1}{n} \sum_{1 \leqslant i \leqslant n} \delta_{x_i}$$

telle que

$$\int_X f(x) d\delta(x) = \frac{1}{n} \sum_{1 \leqslant i \leqslant n} f(x_i),$$

[1]. Pour être pédant, notons que f n'existe que s'il existe Y tel que $\mu(Y) \notin \{0, +\infty\}$.

que l'on peut interpréter comme une mesure « d'échantillonnage ».

Exercice 2.3.6. Soit (Ω, Σ, P) un espace probabilisé. On va montrer que si X et Y sont des variables aléatoires complexes intégrables indépendantes (Définition 1.2.9), on a $XY \in L^1$ et

$$E(XY) = E(X)E(Y). \tag{2.15}$$

(1) Montrer que si X et Y sont positives et étagées, alors (2.15) est vrai en utilisant la définition de variables indépendantes.

(2) Si X et Y sont positives, et S_n et T_n sont les fonctions étagées construites dans la preuve de la Proposition 2.2.4, alors S_n et T_n sont indépendantes pour tout $n \geqslant 1$. (Attention ! Il ne s'agit pas de dire que la *suite* des S_n est indépendante de celle des T_n, seulement de l'énoncé pour chaque n fixé).

(3) En déduire que (2.15) est vraie pour X et Y positives en appliquant le théorème de convergence monotone deux fois.

(4) En déduire le résultat général.

On verra (Exemple 4.3.4) que la construction des mesures produits fournit une autre preuve plus simple de cette propriété importante.

Nous arrivons maintenant au premier col dans la théorie de l'intégration : le théorème de convergence dominée de Lebesgue, qui offre une condition très souple pour échanger limite et intégrale.

Théorème 2.3.7. *Soit $(X, \mathcal{M}, \mu)$ un espace mesuré, (f_n) une suite de fonctions complexes intégrables. On suppose que pour tout $x \in X$, $f_n(x) \to f(x)$.*

Alors, s'il existe $g \in L^1(\mu)$ telle que

$$|f_n(x)| \leqslant g(x) \textit{ pour tout } n \geqslant 1 \textit{ et } x \in X, \tag{2.16}$$

la fonction limite f est intégrable et

$$\int_X f(x)d\mu(x) = \lim_{n\to+\infty} \int_X f_n(x)d\mu(x). \tag{2.17}$$

De plus, on a alors $f_n \to f$ dans $L^1(\mu)$, c'est à dire

$$\int_X |f_n - f|d\mu \to 0 \textit{ quand } n \to +\infty. \tag{2.18}$$

Remarque 2.3.8. La condition de domination (2.16) n'est pas une condition nécessaire et suffisante pour avoir (2.17), mais c'est la condition nécessaire la plus simple, qui se trouve être très souvent vérifiée en pratique.[2] Par exemple, si μ est finie (en particulier, pour une mesure de probabilité), la condition est vraie si toutes les f_n sont uniformément bornées puisque $1 \in L^1(\mu)$ pour μ finie. Même dans ce cas, avec de plus f_n continue, il est très délicat de démontrer (2.16) au sens de l'intégrale de Riemann !

Pour la preuve on utilise un lemme concernant les limites de fonctions positives qui est également utile dans d'autres contextes (on l'appelle le lemme de Fatou).

Lemme 2.3.9. *Soit (f_n) une suite de fonctions mesurables $f_n : X \to [0, +\infty]$. Alors*

$$\int_X (\liminf_{n\to+\infty} f_n)d\mu \leqslant \liminf_{n\to+\infty} \int_X f_n d\mu.$$

2. Noter que le contre-exemple (0.5) de l'introduction *reste valable* : l'intégrale de Lebesgue ne peut pas faire de miracles et permettre l'interversion quand celle-ci est fausse, mais elle permet de vérifier qu'elle est permise beaucoup plus facilement.

Démonstration. On a par définition

$$\liminf_{n\to+\infty} f_n = \lim_{n\to+\infty} g_n \text{ avec } g_n(x) = \inf_{k\geqslant n} f_k(x),$$

et c'est une limite croissante, avec $g_n \geqslant 0$ mesurable par le Lemme 1.1.11. D'après le théorème de convergence monotone on a donc

$$\int_X (\liminf_{n\to+\infty} f_n)d\mu = \lim_{n\to+\infty} \int_X g_n d\mu,$$

or de plus $g_n \leqslant f_n$, donc $\int g_n d\mu \leqslant \int f_n d\mu$ et le résultat provient alors de (0.9). □

Preuve du théorème 2.3.7. On a $|f| \leqslant g$ par passage à la limite dans (2.16) donc $f \in L^1(\mu)$ (c'est cette condition qui est souvent mise en défaut dans les suites donnant un contre-exemple à (2.17)). Considérons la suite $h_n = 2g - |f_n - f|$; on a $h_n \geqslant 0$ et $h_n \to 2g$. Le lemme de Fatou montre

$$(2.19) \qquad 2\int_X g d\mu = \int_X (\lim_{n\to+\infty} h_n) d\mu \leqslant \liminf_{n\to+\infty} \int_X h_n d\mu.$$

Comme g ne dépend pas de n, on a par linéarité

$$\int_X h_n d\mu = 2\int_X g d\mu - \|f - f_n\|_1,$$

donc (puisque $\liminf(-a_n) = -\limsup a_n$)

$$\liminf_{n\to+\infty} \int_X h_n d\mu = 2\int_X g d\mu - \limsup_{n\to+\infty} \|f - f_n\|_1.$$

Comparant avec (2.19), on déduit que

$$\limsup_{n\to+\infty} \|f - f_n\|_1 \leqslant 0 \leqslant \liminf_{n\to+\infty} \|f - f_n\|_1 \,;$$

puisqu'il s'agit d'une suite de nombres positifs, on a $\lim\|f - f_n\|_1 = 0$, c'est à dire (2.18).

Finalement, par (2.11) on a

$$\Big|\int_X f_n d\mu - \int_X f d\mu\Big| = \Big|\int_X (f_n - f) d\mu\Big| \leqslant \|f - f_n\|_1 \to 0,$$

d'où l'interversion de limite et intégrale (2.17). □

Voici un premier exemple d'application simple mais utile.

Lemme 2.3.10. *Soit $X_n \in \mathcal{M}$ une suite croissante d'ensembles mesurables recouvrant X, soit $Y_n = X - X_n$ les ensembles complémentaires. Alors pour toute $f \in L^1(\mu)$ on a*

$$\lim_{n\to+\infty} \int_{X_n} f(x)d\mu(x) = \int_X f d\mu$$

$$\lim_{n\to+\infty} \int_{Y_n} f(x)d\mu(x) = 0.$$

Démonstration. Soit $f_n = f\chi_{X_n}$. Puisque $X_n \subset X_{n+1}$, on voit que $f_n(x) \to f(x)$ pour tout $x \in X$. Comme de plus $|f_n(x)| \leqslant |f(x)|$ qui est intégrable par hypothèse, on en déduit par le Théorème 2.3.7 que

$$\int_{X_n} f(x)d\mu(x) = \int_X f_n(x)d\mu(x) \to \int_X f(x)dx.$$

Le second énoncé est simplement complémentaire du précédent puisque

$$\int_{X_n} f(x)d\mu(x) + \int_{Y_n} f(x)d\mu(x) = \int_X f(x)d\mu(x)$$

pour $n \geqslant 1$. □

2.4. Comparaison avec l'intégrale de Riemann

On a déjà vu l'exemple de fonctions (même étagées) intégrables au sens de Lebesgue mais pas au sens de Riemann. On va voir maintenant plus précisément le rapport entre les deux théories. L'espace mesuré considéré ci-après est donc $(\mathbf{R}, \mathcal{B}_\mathbf{R}, \lambda)$, où λ est la mesure de Lebesgue. On réserve ici la notation

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

à l'intégrale de Riemann.

<u>1er cas</u> : soit $I = [a, b]$ un intervalle compact et $f : I \to \mathbf{R}$ une fonction intégrable au sens de Riemann. Admettons temporairement que f est mesurable. Alors $f \in L^1(\lambda)$ et

$$\int_a^b f(x)dx = \int_I f(x)d\lambda(x). \tag{2.20}$$

En effet, pour toute subdivision

$$\mathsf{s} : a = y_0 < y_1 < \cdots < y_n = b$$

on a aussitôt, avec les notations (0.2) et (0.3)

$$S^-(f) \leqslant \sum_{i=0}^{n-1} \int_{[y_i, y_{i+1}]} f(x)d\lambda(x) = \int_{[a,b]} f(x)d\lambda(x) \leqslant S^+(f)$$

de sorte que (2.20) est conséquence de la définition de l'intégrale de Riemann.

En particulier, si f est continue, ou continue par morceaux, elle est certainement mesurable et le résultat est acquis.

En général on a besoin du petit lemme suivant :

Lemme 2.4.1. *Soit $f : [a, b] \to \mathbf{C}$ une fonction intégrale au sens de Riemann. Alors f coincide presque partout avec une fonction mesurable.*

Démonstration. Pour $x \in [a, b]$, posons

$$m(x) = \liminf_{y \to x} f(x)$$
$$M(x) = \limsup_{y \to x} f(x).$$

On a donc

$$m(x) \leqslant f(x) \leqslant M(x) \tag{2.21}$$

en tout point. Pour une subdivision s donnée, on a

$$S^-(f) = \int_{[a,b]} m_\mathsf{s}(x)d\lambda(x), \text{ et } S^+(f) = \int_{[a,b]} M_\mathsf{s}(x)d\lambda(x)$$

où m_s et M_s sont les fonctions en escalier telles que pour $y_i \leqslant x < y_{i+1}$ on a $m_\mathsf{s}(x) = \inf\{f(x) \mid x \in [y_i, y_{i+1}]\}$ et $M_\mathsf{s}(x) = \sup\{f(x) \mid x \in [y_i, y_{i+1}]\}$.

Quand on raffine une subdivision, le minorant m_s croît et la majorant M_s décroît. De plus quand le pas tend vers 0 on a

$$m_\mathsf{s} \to m, \text{ et } M_\mathsf{s} \to M.$$

Cela montre en particulier que m et M sont mesurables, puis par convergence monotone que

$$\int_a^b f(x)dx = \lim S^-(f) = \int_{[a,b]} m(x)d\lambda(x) \leqslant \int_{[a,b]} M(x)d\lambda(x) = \lim S^+(f) = \int_a^b M(x)dx.$$

On trouve donc que

$$\int_{[a,b]} (M(x) - m(x))d\lambda(x) = 0$$

de sorte que par (2.21) et la Proposition 2.2.2, (1), on a $m(x) = f(x) = M(x)$ presque partout. $\square$

Puisque $m(x) = M(x)$ si et seulement si f est continue en x, on a ainsi montré la moitié du théorème suivant :

THÉORÈME 2.4.2. *Soit $f : [a, b] \to \mathbf{C}$. Alors f est intégrable au sens de Riemann si et seulement si f est bornée et l'ensemble des points de discontinuité de f est de mesure de Lebesgue nulle.*

<u>2nd cas</u> : Si $I = [a, +\infty[$ et $f : I \to \mathbf{C}$ est telle que l'intégrale de Riemann

$$\int_a^{+\infty} f(x)dx$$

converge absolument, alors $f \in L^1(d\lambda)$ et

$$\int_a^{+\infty} f(x)dx = \int_I f(x)d\lambda(x).$$

En effet, posons

$$f_n = f\chi_{[n,+\infty[}$$

pour $n \geqslant a$. Les fonctions f_n sont mesurables et $f_n \to f$ en tout point. De plus $|f_n| \leqslant |f_{n+1}| \to |f|$, donc le théorème de convergence monotone et (2.20) montrent que

$$\int_I |f(x)|d\lambda(x) = \lim_n \int_I |f_n|d\lambda(x) = \lim_n \int_a^n |f(x)|dx = \int_a^{+\infty} |f(x)|dx < +\infty$$

par hypothèse, donc $f \in L^1(\mu)$. Alors la majoration $|f_n| \leqslant |f|$, et $f_n \to f$ en tout point, impliquent par le théorème de convergence dominée cette fois que

$$\int_a^n f(x)dx = \int_{[a,n]} f(x)d\lambda(x) = \int_I f_n(x)d\lambda(x) \to \int_I f(x)d\lambda(x)$$

quand $n \to +\infty$. Mais par définition la limite ci-dessus est aussi l'intégrale impropre de f sur I au sens de Riemann.

Un raisonnement similaire vaut pour une fonction définie sur $[a, b[$ telle que l'intégrale de Riemann

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

converge absolument.

CHAPITRE 3

Premières applications de l'intégrale

Ce chapitre contient quelques premiers exemples d'applications simples mais importantes de l'intégrale de Lebesgue. Certains seront développés davantage ultérieurement.

3.1. Espaces L^p

Soit $(X, \mathcal{M}, \mu)$ un espace mesuré. La Proposition 2.3.3 montre que l'ensemble noté $L^1(\mu)$ a presque la structure d'un espace vectoriel normé par $\|\cdot\|_1$. Le seul obstacle est que la condition $\|f\|_1 = 0$ n'implique pas $f = 0$, mais seulement $f = 0$ presque partout. Pour corriger cela, on est amené à changer la définition.

Définition 3.1.1. L'espace $L^1(\mu)$ est l'espace vectoriel quotient

$$\{f : X \to \mathbf{C} \mid f \text{ est intégrable}\}/N$$

où N est le sous-espace vectoriel

$$N = \{f \mid f \text{ est mesurable et nulle presque partout}\} = \Big\{f \mid \int_X |f| d\mu = 0\Big\}. \tag{3.1}$$

C'est un espace normé pour la norme

$$\|f\|_1 = \int_X |f| d\mu.$$

À partir de ce point, la notation $L^1(\mu)$ réfèrera exclusivement à cette définition

On note aussi parfois $L^1(X, \mu)$ ou bien simplement $L^1(X)$ s'il n'y a pas d'ambiguïté dans le contexte. Par exemple, $L^1(\mathbf{R})$ désigne toujours $L^1(\lambda)$ où λ est la mesure de Lebesgue sur $\mathbf{R}$.

Concrètement (cf. le rappel dans la section Notations), un élément de $L^1(\mu)$ est une (classe de) fonction intégrable f, et $f = g$ dans $L^1(\mu)$ si et seulement si $f = g$ presque partout. Pour construire une application définie sur $L^1(\mu)$ il suffit de le faire pour les fonctions intégrables f et de s'assurer que le résultat de la construction ne change pas si f est remplacée par g qui coïncide avec f presque partout. Par exemple, la norme $\|f\|_1$ est effectivement bien définie pour $f \in L^1(\mu)$ car si $f = g$ presque partout on a $|f| = |g|$ presque partout donc $\int |f| d\mu = \int |g| d\mu$.

Similairement, une relation comme $f \leqslant g$, pour $f, g \in L^1(\mu)$ signifie que l'inégalité est vraie presque partout seulement, etc...

Il devient aussi très vite naturel de considérer comme éléments de $L^1(\mu)$ des fonctions f qui ne sont vraiment définies que presque partout (par exemple la somme d'une série ne convergeant que presque partout, voir la preuve de la Proposition 3.1.3 ci-dessous). Cela est légitime, et une justification formelle est donnée dans l'exercice suivant que le lecteur encore peu à l'aise avec la notion d'ensemble quotient est encouragé à résoudre en détail.

Exercice 3.1.2. Soit L l'ensemble des couples (Y, f) où $Y \in \mathcal{M}$ vérifie $\mu(X - Y) = 0$ et $f : Y \to \mathbf{C}$ est une fonction mesurable telle que

$$\int_Y |f| d\mu < +\infty.$$

Soit L_1 l'ensemble quotient $L/\sim$ où $\sim$ est la relation d'équivalence

$$(Y, f) \sim (Y', f') \text{ si } f(x) = f'(x) \text{ pour } x \in Y \cap Y'.$$

(1) Montrer que $\sim$ est effectivement une relation d'équivalence.

(2) Montrer que L_1 est un espace vectoriel normé pour l'addition et la multiplication scalaire induites par

$$(Y, f) + (Y', f') = (Y \cap Y', f + f'), \text{ et } \lambda(Y, f) = (Y, \lambda f),$$

et pour la norme induite par

$$\|(Y, f)\| = \int_Y |f| d\mu.$$

(3) Montrer que l'application

$$\begin{cases} L^1(\mu) \to L_1 \\ f \mapsto (X, f) \end{cases}$$

est une isométrie d'espaces normés dont l'inverse est

$$(Y, f) \mapsto g \text{ telle que } g(x) = \begin{cases} f(x) & \text{si } x \in Y \\ 0 & \text{sinon.} \end{cases}$$

(4) Montrer que le théorème de convergence dominée implique le résultat suivant : si (f_n) est une suite d'éléments $f_n \in L_1$ tels que $f_n \to f$ presque partout, et si $|f_n| \leqslant g$ avec $g \in L_1$, alors f_n converge vers f en norme L_1 et

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

En pratique, on s'habitue très vite à négliger ces formalités. Pour le premier résultat ci-dessous, nous donnerons une preuve explicitant les relations entre « vraies » fonctions et éléments de $L^1(\mu)$. Il s'agit seulement de convaincre le lecteur de la facilité qu'il y a à manipuler ces notions. Après cela, nous utiliserons les fonctions L^1 de façon beaucoup plus souple : la justification de nos petits abus de langage serait immédiate mais fort répétitive.

Proposition 3.1.3. *Soit $(X, \mathcal{M}, \mu)$ un espace mesuré, (f_n) une suite de fonctions $f_n \in L^1(\mu)$. Supposons que la série $\sum f_n$ converge normalement dans $L^1(\mu)$, c'est à dire*

$$\sum_{n \geqslant 1} \|f_n\|_1 < +\infty.$$

Alors la série

$$\sum_{n \geqslant 1} f_n(x)$$

converge presque partout vers $g(x)$, de plus $g \in L^1(\mu)$ et

$$\int_X g d\mu = \sum_{n \geqslant 1} \int_X f_n d\mu. \tag{3.2}$$

Enfin, la convergence a également lieu pour la norme $\|\cdot\|_1$.

Démonstration. Soit

$$h(x) = \sum_{n \geqslant 1} |f_n(x)|$$

pour $x \in X$. Si l'on change f_n sur un ensemble X_n de mesure 0, alors h est changée sur l'ensemble

$$X = \bigcup_{n \geqslant 1} X_n$$

qui est de mesure 0 : cela montre que $h \geqslant 0$ est bien définie presque partout. En particulier son intégrale est bien définie et d'après le Corollaire 2.2.6 au théorème de convergence monotone on a

$$\int_X h d\mu = \sum_{n \geqslant 1} \int_X |f_n| d\mu = \sum_{n \geqslant 1} \|f\|_1 < +\infty \text{ par hypothèse},$$

en particulier h est finie presque partout (Proposition 2.2.2, (1)).

Si x vérifie $h(x) < \infty$, la série $\sum f_n(x)$ converge absolument et donc a une somme $g(x)$ telle que $|g(x)| \leqslant h(x)$. On peut voir g soit comme une fonction définie presque partout, comme dans l'exercice ci-dessus, ou bien on peut l'étendre arbitrairement à X en posant $g(x) = 0$ si $x \in Y$, où $Y = \{x \mid h(x) = +\infty\}$).

Comme $|g| \leqslant h$ dans tout les cas, on voit que $g \in L^1(\mu)$, et on peut alors appliquer le théorème de convergence dominée à la suite des sommes partielles

$$u_n(x) = \sum_{1 \leqslant i \leqslant n} f_i(x) \text{ si } x \notin Y, \quad u_n(x) = 0 \text{ sinon}$$

telles que $u_n(x) \to g(x)$ pour tout x, et conclure que

$$\int_X g d\mu = \lim_n \int_X u_n d\mu = \lim_n \sum_{1 \leqslant i \leqslant n} \int_X f_i d\mu = \sum_{n \geqslant 1} \int_X f_n d\mu.$$

□

Proposition 3.1.4. (1) *Soit $(X, \mathcal{M}, \mu)$ un espace mesuré. Alors l'espace $L^1(\mu)$ est un espace de Banach pour la norme $\|\cdot\|_1$, c'est à dire que c'est un espace vectoriel complet.*

(2) *Plus précisément, si (f_n) est une suite de Cauchy dans $L^1(\mu)$, alors il existe $f \in L^1(\mu)$ telle que $f_n \to f$ dans $L^1(\mu)$, c'est à dire $\|f - f_n\|_1 \to 0$, et de plus il existe une sous-suite (f_{n_k}), $k \geqslant 1$, telle que*

$$\lim_{k \to +\infty} f_{n_k}(x) = f(x) \textit{ pour presque tout } x.$$

Remarque 3.1.5. On s'aperçoit rapidement que l'énoncé (2) ne peut être amélioré : il est possible qu'il n'existe pas de sous-suite de f_n qui converge *partout* vers f (un exemple est donné dans l'Exercice 3.1.6) ou que la limite des f_n ne soit définie que presque partout (la suite $f_n(x)$ divergeant en certains points). Cela montre qu'il s'agit véritablement d'un énoncé de type L^1, qu'il serait peu élégant de vouloir exprimer purement en termes de fonctions.

Démonstration. La condition de Cauchy est : pour tout $\varepsilon > 0$, il existe $N(\varepsilon) \geqslant 1$ tel que pour tout $n \geqslant N$, $m \geqslant N$, on a

$$\|f_n - f_m\|_1 < \varepsilon. \tag{3.3}$$

Prenons $\varepsilon = 2^{-k}$, $k \geqslant 1$. Alors par récurrence sur k on voit qu'il existe une suite (n_k) tel que $n_{k+1} > n_k$ et

$$\|f_{n_{k+1}} - f_{n_k}\|_1 < 2^{-k}.$$

Cette sous-suite répond au point (2). Pour le vérifier, soit $g_k = f_{n_{k+1}} - f_{n_k} \in L^1(\mu)$ et considérons alors la série $\sum g_k$.

On a

$$\sum_{n \geqslant 1} \|g_k\|_1 \leqslant \sum_{k \geqslant 1} 2^{-k} < +\infty.$$

D'après la Proposition 3.1.3, la série converge donc presque partout vers $g \in L^1(\mu)$. Mais les sommes partielles vérifient

$$\sum_{k=1}^{K} g_k = f_{n_{K+1}} - f_{n_1}$$

donc cela signifie que la suite extraite (f_{n_k}) converge presque partout, disons vers $f \in L^1(\mu)$. Toujours d'après la Proposition, la convergence de la série est également vraie dans $L^1(\mu)$, c'est à dire que

$$\lim_{k \to +\infty} \|f - f_{n_k}\|_1 = 0. \tag{3.4}$$

Cette dernière formule dit que f est valeur d'adhérence de la suite de Cauchy (f_n), et l'on sait qu'une suite de Cauchy admettant une valeur d'adhérence converge vers celle-ci. Répétons l'argument qui le montre : pour tout n et tout k, on a

$$\|f - f_n\|_1 \leqslant \|f - f_{n_k}\|_1 + \|f_{n_k} - f_n\|_1$$

Par (3.3), si n et n_k sont $\geqslant N(\varepsilon)$, on a $\|f_n - f_{n_k}\|_1 < \varepsilon$. De plus par (3.4) pour $k > K(\varepsilon)$ on a $\|f - f_{n_k}\|_1 < \varepsilon$. Fixons un $k > K(\varepsilon)$ tel que de plus $n_k > N(\varepsilon)$. Alors, pour tout $n > N(\varepsilon)$, on a $\|f - f_n\|_1 < 2\varepsilon$, ce qui montre la convergence de f_n vers f en norme. □

Exercice 3.1.6. On considère la fonction $f : [1, +\infty[\to \mathbf{R}$ définie de la manière suivante : si $n \geqslant 1$ est l'entier tel que $n \leqslant x < n + 1$, si de plus $n = 2^k + j$ avec $k \geqslant 0$ et $0 \leqslant j < 2^k$, alors

$$f(x) = \begin{cases} 1 & \text{si } j2^{-k} \leqslant x < (j+1)2^{-k} \\ 0 & \text{sinon.} \end{cases}$$

(1) Esquisser le graphe de f. Quelle image rappelle-t-il ?

(2) Pour $n \geqslant 1$, soit $f_n(x) = f(x + n)$ pour $x \in X = [0, 1]$. Montrer que $f_n \in L^1(X, d\lambda)$ et que $f_n \to 0$ dans $L^1(X)$.

(3) Montrer que pour tout $x \in [0, 1]$, la suite $f_n(x)$ n'a pas de limite.

(4) Expliciter une sous-suite (f_{n_k}) telle que f_{n_k} converge presque partout vers 0.

Le raisonnement amenant au Théorème 3.1.10 s'applique sans changement aux espaces L^p qui sont définis comme suit.

Définition 3.1.7. Soit $(X, \mathcal{M}, \mu)$ un espace mesuré, $p \geqslant 1$ un nombre réel. L'espace $L^p(\mu)$ est l'espace vectoriel quotient

$$\{f : X \to \mathbf{C} \mid |f|^p \text{ est intégrable}\}/N$$

où N est comme en (3.1). Cet espace est muni de la norme

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

Comme pour L^1, on note aussi $L^p(X, \mu)$ ou $L^p(X)$, et en particulier $L^p(\mathbf{R}) = L^p(\lambda)$.

Pour vérifier que cette définition a un sens, il faut s'assurer que l'on a bien un espace vectoriel et que $\|\cdot\|_p$ est effectivement une norme. Cela découle des inégalités bien connues de Hölder et de Minkowski.

Rappelons qu'une fonction $\varphi : [0, +\infty[\to [0, +\infty[$ est convexe si on a

$$\varphi(ra + sb) \leqslant r\varphi(a) + s\varphi(b) \tag{3.5}$$

pour tout a, b tels que $0 \leqslant a \leqslant b$ et tout r, $s \geqslant 0$ tels que $r + s = 1$. Si φ est de class C^2, alors elle est convexe si et seulement si $\varphi'' \geqslant 0$.

Proposition 3.1.8. (1) *Soit* $\varphi : [0, +\infty[\to [0, +\infty[$ *une fonction continue, croissante et* convexe. *Soit* $(X, \mathcal{M}, \mu)$ *un espace mesurable avec* $\mu(X) = 1$ *et* $f : X \to [0, +\infty[$ *une fonction mesurable. On a alors*

$$\varphi\Big(\int_X f d\mu\Big) \leqslant \int_X (\varphi \circ f) d\mu. \tag{3.6}$$

(2) *Soit* $(X, \mathcal{M}, \mu)$ *un espace mesuré,* $p > 1$ *un entier,* $q > 1$ *tel que* $p^{-1} + q^{-1} = 1$. *Alors pour toutes* f, $g : X \to [0, +\infty]$ *mesurables, on a*

$$\int_X fg d\mu \leqslant \Big(\int_X f^p d\mu\Big)^{1/p} \Big(\int_X g^q d\mu\Big)^{1/q} = \|f\|_p \|g\|_q. \tag{3.7}$$

(3) *Soit* $(X, \mathcal{M}, \mu)$ *un espace mesuré,* $p > 1$ *un entier. Alors pour toutes* f, $g : X \to [0, +\infty]$ *mesurables, on a*

$$\|f + g\|_p = \Big(\int_X (f + g)^p d\mu\Big)^{1/p} \leqslant \Big(\int_X f^p d\mu\Big)^{1/p} + \Big(\int_X g^p d\mu\Big)^{1/p} = \|f\|_p + \|g\|_p. \tag{3.8}$$

Démonstration. (1) : l'inégalité (3.6) se réduit à (3.5) pour $f \geqslant 0$ ne prenant que deux valeurs a et b puisque

$$\mu\{f(x) = a\} + \mu\{f(x) = b\} = \mu(X) = 1.$$

Par récurrence, (3.6) reste vraie pour toute fonction étagée $s \geqslant 0$. Pour $f \geqslant 0$ quelconque, soit s_n une suite croissante de fonctions étagées positives telles que $s_n(x) \to f(x)$ pour tout x. On a $\int s_n d\mu \to \int f d\mu$ et de même, comme φ est croissante, $\varphi \circ s_n \to \varphi \circ f$ en croissant, donc $\int (\varphi \circ s_n) d\mu \to \int (\varphi \circ f) d\mu$ par convergence monotone.

Par continuité de φ on a donc

$$\varphi\Big(\int_X f d\mu\Big) = \lim_{n\to+\infty} \varphi\Big(\int_X s_n d\mu\Big) \leqslant \lim_{n\to+\infty} \int_X (\varphi \circ s_n) d\mu = \int_X (\varphi \circ f) d\mu.$$

(2) : si $1/p + 1/q = 1$, l'inégalité de convexité pour $\varphi(x) = e^x$ donne, après un changement de variable $x \mapsto p \log x$, $y \mapsto p \log y$, l'inégalité de Young

$$xy \leqslant \frac{x^p}{q} + \frac{y^q}{q}, \text{ pour } x \geqslant 0,\ y \geqslant 0.$$

Pour montrer (3.7), il suffit de considérer le cas où f et g vérifient

$$0 < \|f\|_p < +\infty, \quad 0 < \|f\|_g < +\infty,$$

les autres étant évidents. En considérant alors $f/\|f\|_p$, $g/\|g\|_q$, on se ramène par homogénéité au cas où $\|f\|_p = \|g\|_q = 1$. Dans ce cas l'inégalité de Young $fg \leqslant \frac{f^p}{p} + \frac{g^q}{q}$ implique en intégrant

$$\int_X fg d\mu \leqslant \frac{\|f\|_p}{p} + \frac{\|f\|_q}{q} = 1,$$

ce qui est le résultat voulu.

(3) L'inégalité est vraie si $\|f\|_p = +\infty$ ou $\|g\|_p = +\infty$. Sinon, soit $h = (f+g)^{p-1}$. Notons que

$$(f+g)^p \leqslant (2\max(f,g))^p \leqslant 2^p(f^p + g^p)$$

donc $\|f+g\|_p < +\infty$. On écrit $(f+g)^p = fh + gh$, donc par (3.7) on trouve

$$\int_X (f+g)^p d\mu = \int_X fh d\mu + \int_X gh d\mu \leqslant \|f\|_p \|h\|_q + \|g\|_p \|h\|_q.$$

Or

$$\|h\|_q = \Big(\int_X (f+g)^{q(p-1)} d\mu\Big)^{1/q} = \|f+g\|_p^{p/q} \text{ car } q(p-1) = p,$$

et de plus $1 - 1/q = 1/p$ donc (en traitant séparément le cas où $h = 0$ presque partout) l'inégalité précédente donne en divisant par $\|f+g\|_p^{p/q} < +\infty$:

$$\|f+g\|_p = \Big(\int_X (f+g)^p d\mu\Big)^{p(1-1/q)} \leqslant \|f\|_p + \|g\|_p$$

$\square$

Exemple 3.1.9. Pour $(\mathbf{R}, \mathcal{B}, \lambda)$, la fonction f définie par

$$f(x) = \frac{\sin(x)}{x} \text{ si } x \neq 0 \text{ et } f(0) = 1$$

est dans $L^2(\mathbf{R})$, mais pas dans $L^1(\mathbf{R})$ (car l'intégrale ne converge pas absolument). En particulier, on remarque que l'intégrale d'une fonction dans L^p, $p \neq 1$, n'existe pas toujours !

Le théorème suivant généralise la Proposition 3.1.3 ; il est connu sous le nom de Théorème de Riesz-Fisher.

Théorème 3.1.10. (1) *Soit* $(X, \mathcal{M}, \mu)$ *un espace mesuré,* $p \geqslant 1$*,* (f_n) *une suite de fonctions* $f_n \in L^p(\mu)$*. Si* $\sum \|f_n\|_p < +\infty$*, alors la série*

$$\sum_{n \geqslant 1} f_n$$

converge presque partout et en norme L^p *vers une fonction* $g \in L^p(\mu)$*.*

(2) *Pour* $p \geqslant 1$*, l'espace* $L^p(\mu)$ *est un espace de Banach pour la norme* $\|\cdot\|_p$*, et plus précisément, si* (f_n) *est une suite de Cauchy dans* $L^p(\mu)$*, il existe* $f \in L^p(\mu)$ *telle que* $f_n \to f$ *dans* $L^p(\mu)$*, et de plus il existe une sous-suite* (f_{n_k})*,* $k \geqslant 1$*, telle que*

$$\lim_{k \to +\infty} f_{n_k}(x) = f(x) \text{ pour presque tout } x.$$

(3) *En particulier, si* $p = 2$*,* $L^2(\mu)$ *est un espace de Hilbert pour le produit scalaire*

$$(f \mid g) = \int_X f(x)\overline{g(x)} d\mu.$$

Démonstration. Si (1) est prouvé, l'argument utilisé pour montrer la Proposition 3.1.4 à partir de la Proposition 3.1.3 se transcrit ligne par ligne en remplaçant simplement les normes $\|\cdot\|_1$ par $\|\cdot\|_p$.

Pour montrer (1), on recopie la preuve de la Proposition 3.1.3 : soit $h(x) = \sum |f_n(x)|$. On a par continuité

$$h^p = \lim_{N \to +\infty} \Bigl(\sum_{n \leqslant N} |f_n(x)|\Bigr)^p.$$

C'est une limite croissante donc par convergence monotone et l'inégalité triangulaire il vient

$$\Bigl(\int_X h^p d\mu\Bigr)^{1/p} = \lim_{N \to +\infty} \Bigl(\int_X \Bigl(\sum_{n \leqslant N} |f_n|\Bigr)^p d\mu\Bigr)^{1/p} = \lim_{N \to +\infty} \Bigl\|\sum_{n \leqslant N} |f_n|\Bigr\|_p \leqslant \sum_{n \geqslant 1} \|f_n\|_p < +\infty$$

par hypothèse. Donc h^p, et h également par conséquent, est finie presque partout.

Comme pour le cas de $L^1(\mu)$, cela montre que la série $g(x) = \sum f_n(x)$ converge absolument presque partout et $|g(x)| \leqslant h(x)$, donc $g \in L^p(\mu)$. Comme

$$\Bigl\|g - \sum_{n \leqslant N} f_n\Bigr\|_p = \Bigl\|\sum_{n > N} f_n\Bigr\|_p \leqslant \sum_{n > N} \|f_n\|_p \to 0,$$

la convergence a également lieu dans L^p. □

Remarque 3.1.11. Si (Ω, Σ, P) est un espace probabilisé, et X une variable aléatoire, alors

$$\sigma^2(X) = V(X) = E(|X - E(X)|^2) = E(|X|^2) - |E(X)|^2$$

est appelé la *variance* de X. Elle est bien définie pour $X \in L^2(\Omega)$ (voir ci-dessous), et mesure intuitivement l'écart moyen entre X et sa valeur moyenne : une variance « petite » (dans un sens dépendant du contexte) indiquera que la variable aléatoire est concentrée autour de la moyenne.

Pour vérifier la seconde formule ci-dessus on calcule

$$E(|X - E(X)|^2) = E(|X|^2 - XE(\bar{X}) - \bar{X}E(X) + |E(X)|^2) = E(|X|^2) - 2|E(X)|^2 + |E(X)|^2$$

puisque $E(1) = 1$ pour un espace probabilisé.

La racine carrée $\sigma(X) = \sqrt{V(X)} = \|X - E(X)\|_2$ de la variance est appelée *l'écart-type* de X.

Variance et écart-type ne dépendent que de la loi μ de X, et l'on parle aussi de la variance ou de l'écart-type de μ : on a par (2.7)

$$E(X) = \int_{\mathbf{C}} x d\mu(x) \text{ et } V(X) = \int_{\mathbf{C}} (x - E(X))^2 d\mu(x).$$

En général, la variance d'une somme $X+Y$ de variables aléatoires ne peut être déterminée simplement à partir des lois de X et Y. Une exception importante est le cas de variables indépendantes.

Proposition 3.1.12. *Soit (Ω, Σ, P) un espace probabilisé, soient X et Y des variables aléatoires L^2 indépendantes. Alors*

$$V(X+Y) = V(X) + V(Y). \tag{3.9}$$

Démonstration. On a $E(X+Y) = E(X) + E(Y)$ donc, utilisant la seconde formule, on a

$$\begin{aligned} V(X+Y) &= E(|X|^2 + 2\,\mathrm{Re}(XY) + |Y|^2) - (|E(X)|^2 + 2\,\mathrm{Re}(E(X)E(Y)) + |E(Y)|^2) \\ &= V(X) + V(Y) + 2\,\mathrm{Re}(E(XY) - E(X)E(Y)) = V(X) + V(Y) \end{aligned}$$

par (2.15). □

Notons aussi que $V(aX) = |a|^2 V(X)$.

Un dernier espace noté L^∞ joue un rôle important comme « limite » des espaces L^p pour $p \to +\infty$. Il s'agit essentiellement des fonctions bornées, mais la définition requière un peu de soin pour la raison habituelle...

Définition 3.1.13. Soit $f : X \to \mathbf{C}$ une fonction mesurable. On dit que f est *essentiellement bornée* par M si

$$\mu(\{x \mid |f(x)| > M\}) = 0. \tag{3.10}$$

Proposition 3.1.14. *Soit f une fonction mesurable, et*

$$\|f\|_\infty = \inf\{M \mid f \text{ est essentiellement bornée par } M\} \in [0, +\infty].$$

Alors f est essentiellement bornée par $\|f\|_\infty$ et l'espace vectoriel quotient

$$L^\infty(\mu) = \{f \mid \|f\|_\infty < +\infty\}/N,$$

où N est le sous-espace (3.1) *des fonctions nulles presque partout, est un espace vectoriel normé par $\|\cdot\|_\infty$.*

De plus si $f \in L^1(\mu)$ et $g \in L^\infty(\mu)$, on a $fg \in L^1(\mu)$ et

$$\int_X |fg| d\mu \leqslant \|f\|_1 \|g\|_\infty. \tag{3.11}$$

Attention à ce que $\|f\|_\infty \leqslant \sup\{f(x)\}$ avec une inégalité stricte en général, même si la borne supérieure est atteinte. Par exemple, si $X = (\mathbf{R}, \mathcal{B}, \lambda)$ et $f = \chi_{\mathbf{Q}}$ on a $\|f\|_\infty = 0$ bien que la fonction prenne la valeur 1.

On utilise la définition communément sous la forme

$$|f(x)| \leqslant M \text{ presque partout si et seulement si } M \geqslant \|f\|_\infty.$$

Démonstration. Pour vérifier que $M = \|f\|_\infty$ vérifie la condition (3.10), il suffit de remarquer qu'il existe une suite décroissante M_n de nombres vérifiant (3.10) telle que $M_n \to M$. On a alors

$$\mu(\{x \mid |f(x)| > M\}) = \mu\Big(\bigcup_n \{x \mid |f(x)| > M_n\}\Big) = 0.$$

Il est immédiat alors que $\|f\|_\infty = 0$ équivaut à f nulle presque partout puisque la condition devient

$$\mu(\{x \mid f(x) \neq 0\}) = \mu(\{x \mid |f(x)| > 0\}) = 0.$$

Les autres axiomes pour un espace vectoriel normé sont tout aussi simples à vérifier, et (3.11) est également évident par monotonie si on remarque que l'on peut supposer en remplaçant g sur un ensemble de mesure nulle que $|g(x)| \leqslant \|g\|_\infty$ pour tout $x \in X$. □

L'analogue du théorème de Riesz-Fisher est vrai pour l'espace L^∞ mais la preuve est différente.

PROPOSITION 3.1.15. (1) *Soit* $(X, \mathcal{M}, \mu)$ *un espace mesuré,* (f_n) *une suite de fonctions* $f_n \in L^\infty(\mu)$. *Si* $\sum \|f_n\|_\infty < +\infty$, *alors la série*

$$\sum_{n \geqslant 1} f_n$$

converge presque partout et en norme L^∞ *vers une fonction* $g \in L^\infty(\mu)$.

(2) *L'espace* $L^\infty(\mu)$ *est un espace de Banach, et plus précisément, si* (f_n) *est une suite de Cauchy dans* $L^\infty(\mu)$, *il existe* $f \in L^\infty(\mu)$ *telle que* $f_n \to f$ *dans* $L^\infty(\mu)$ *et presque partout.*

DÉMONSTRATION. (1) : il s'agit de la même méthode que précédemment. On pose

$$h(x) = \sum_{n \geqslant 1} |f_n(x)|, \quad g(x) = \sum_{n \geqslant 1} f_n(x)$$

et on vérifie que pour presque tout x les deux séries convergent absolument. Comme

$$|g(x)| \leqslant h(x) \leqslant \sum_{n \geqslant 1} \|f_n\|_\infty \text{ presque partout}$$

on a $g \in L^\infty(\mu)$, et enfin $\sum f_n \to g$ dans L^∞ car

$$\Big\| g - \sum_{n \leqslant N} f_n \Big\|_\infty = \Big\| \sum_{n > N} f_n \Big\|_\infty \leqslant \sum_{n > N} \|f_n\|_\infty \to 0.$$

(2) : la preuve est plus simple que pour L^p, $p < \infty$ (et il n'est pas besoin d'avoir une sous-suite pour avoir convergence presque partout).

En effet, si (f_n) est une suite de Cauchy dans $L^\infty(\mu)$, on a

$$|f_n(x) - f_m(x)| \leqslant \|f_n - f_m\|_\infty$$

pour presque tout x ; soit $A_{n,m}$ l'ensemble de mesure nulle en dehors duquel cette inégalité est vraie, et A la réunion des $A_{n,m}$, $\mu(A) = 0$.

Pour $x \notin A$, la suite $(f_n(x))$ est de Cauchy dans $\mathbf{C}$ (puisque (f_n) est de Cauchy dans L^∞), et donc converge, disons vers $f(x) \in \mathbf{C}$. Il reste à vérifier que $f \in L^\infty(\mu)$ (prolongée par 0 par exemple).

La suite des normes $(\|f_n\|_\infty)$ est elle-même de Cauchy grâce à l'inégalité triangulaire, et converge donc vers $M \geqslant 0$. Soit B_n l'ensemble des x tels que $|f_n(x)| > \|f_n\|_\infty$, de sorte que $\mu(B_n) = 0$, et B la réunion de ces ensembles. On a $\mu(A \cup B) = 0$ également.

Pour $x \notin A \cup B$, on a

$$|f_n(x)| \leqslant \|f_n\|_\infty \text{ pour tout } n \geqslant 1 \text{ et } f_n(x) \to f(x)$$

donc par passage à la limite il vient

$$|f(x)| \leqslant M \text{ presque partout,}$$

et $\|f\|_\infty \leqslant M < +\infty$. Cela étant acquis, soit $\varepsilon > 0$ et N tel que

$$\|f_n - f_m\|_\infty < \varepsilon$$

pour n, $m \geqslant N$. On a alors $|f_n(x) - f_m(x)| < \varepsilon$ pour $x \notin A \cup B$ et n, $m \geqslant N$. Soit $m \geqslant N$ quelconque. En faisant $n \to +\infty$ il vient

$$|f(x) - f_m(x)| < \varepsilon \text{ pour presque tout } x,$$

c'est à dire $\|f - f_m\|_\infty < \varepsilon$ pour tout $m > N$, donc $f_n \to f$ dans $L^\infty(\mu)$. $\square$

Remarque 3.1.16. Il n'y a pas forcément de relations évidentes entre les différents espaces $L^p(\mu)$, $p \geqslant 1$. Si on compare le Théorème 3.1.10, (1) avec la Proposition 3.1.3, dans le cas de séries de fonctions dans $L^p(\mu)$, la formule (3.2) n'est pas forcément vraie, simplement parce qu'il se peut que $g \notin L^1(\mu)$.

Par exemple, considérons $X = \mathbf{R}$ avec la mesure de Lebesgue. La fonction $f(x) = \inf(1, 1/|x|)$ est dans L^2, mais pas dans L^1, alors que $g(x) = x^{-1/2}\chi_{[0,1]}$ est dans L^1 mais pas dans L^2.

Par contre si la mesure μ est finie (par exemple une mesure de probabilité) on a des relations telles que $L^2 \subset L^1$ (de sorte que si la variance d'une variance aléatoire existe, son espérance également). Cette inclusion provient de l'inégalité de Hölder (3.7) : si $f \in L^2(\mu)$, on a

$$\int_X |f| d\mu = \int_X (|f| \cdot 1) d\mu \leqslant \|f\|_2 \|1\|_2 = \mu(X)^{1/2} \|f\|_2.$$

Plus généralement, on a

Proposition 3.1.17. *Soit $(X, \mathcal{M}, \mu)$ un espace mesuré avec μ finie. Soit $p \geqslant p' \geqslant 1$ des réels. Alors l'inclusion naturelle*

$$L^p(\mu) \hookrightarrow L^{p'}(\mu)$$

est une application linéaire continue entre espaces normés, de norme $\leqslant \mu(X)^{1/p'-1/p}$.

Démonstration. Il s'agit de montrer qu'il existe $K \geqslant 0$ tel que

$$\|f\|_{p'} \leqslant K \|f\|_p$$

pour $f \in L^p(\mu)$. Soit $r = p/p' \geqslant 1$, on a

$$\int_X |f|^{p'} d\mu \leqslant \Big(\int_X |f|^{rp'} d\mu \Big)^{1/r} \mu(X)^{1/s}$$

avec $1/r + 1/s = 1$ et (puisque $rp' = p$ et $\frac{1}{sp'} = \frac{1}{p'} - \frac{1}{p}$)

$$\|f\|_{p'} \leqslant \mu(X)^{1/p'-1/p} \|f\|_p.$$

□

De plus, toujours si μ est finie, on a $L^\infty \subset L^p$ et même

$$\|f\|_p \leqslant \mu(X)^{1/p} \|f\|_\infty \text{ pour tout } p \geqslant 1 \tag{3.12}$$

par (3.11). La proposition suivante justifie le fait de considérer L^∞ comme cas limite des L^p.

Proposition 3.1.18. *Sous ces hypothèses, on a*

$$\lim_{p \to +\infty} \|f\|_p = \|f\|_\infty.$$

Démonstration. Si $f = 0$ dans L^p (ou dans L^∞), c'est évident. Si $f \neq 0$, on peut se ramener par linéarité au cas $\|f\|_\infty = 1$. En changeant f sur un ensemble de mesure nulle, on peut aussi supposer que $|f(x)| \leqslant 1$ pour tout $x \in X$.

Soit $\varepsilon > 0$ et $Y_\varepsilon = \{x \mid |f(x)| \geqslant 1 - \varepsilon\}$. On a par monotonie

$$\int_X |f|^p d\mu \geqslant (1-\varepsilon)^p \mu(Y_\varepsilon)$$

donc

$$\|f\|_p \geqslant (1-\varepsilon)\mu(Y_\varepsilon)^{1/p}.$$

Comme $\|f\|_\infty = 1$, on a $\mu(Y_\varepsilon) > 0$, donc $\mu(Y_\varepsilon)^{1/p} \to 1$ quand $p \to +\infty$. Prenant donc la limite inférieure dans l'inégalité il vient $\liminf \|f\|_p \geqslant 1 - \varepsilon$. Faisant alors $\varepsilon \to 0$, on déduit que $\liminf \|f\|_p \geqslant 1$.

Comme d'après (3.12) on a $\limsup \|f\|_p \leqslant \|f\|_\infty = 1$, le résultat en découle. □

Exercice 3.1.19. Soit μ finie et $f \in L^\infty(\mu)$, $\|f\|_\infty \leqslant 1$. Montrer que

$$\lim_{p\to+\infty} \int_X |f|^p d\mu = \mu(\{x \mid |f(x)| > 0\}).$$

Terminons cette section avec le résultat suivant, qui justifie pleinement l'importance donnée aux espaces L^p. C'est seulement un particulier d'un énoncé d'approximation beaucoup plus général qui sera démontré plus tard.

Proposition 3.1.20. *Soit $X = \mathbf{R}$ avec la mesure de Lebesgue, $p \in [1, +\infty[$. Alors $L^p(\mathbf{R})$ est l'espace vectoriel complété de l'espace vectoriel normé $(C_c(\mathbf{R}), \|\cdot\|_p)$ où $C_c(\mathbf{R})$ est l'espace des fonctions continues à support compact sur $\mathbf{R}$ et la norme est*

$$\|f\|_p = \Big(\int_{\mathbf{R}} |f(t)|^p dt\Big)^{1/p}.$$

Remarque 3.1.21. Ceci est faux si $p = +\infty$. Dans ce cas (topologie de la convergence uniforme), l'adhérence de $(C_c(\mathbf{R}), \|\cdot\|_\infty)$ est l'espace des fonctions continues bornées sur $\mathbf{R}$.

Rappelons que le support $F = \mathrm{supp}(f)$ d'une fonction continue $f : X \to \mathbf{C}$, où X est un espace topologique, est « le plus grand fermé sur lequel f est non nulle », plus précisément $F = \overline{\{x \mid f(x) \neq 0\}}$.

La norme est donc juste la restriction de la norme L^p à $C_c(\mathbf{R})$. Elle est bien définie car si $\mathrm{supp}(f)$ est compact, $\mathrm{supp}(f) \subset [-M, M]$ pour un certain $M \geqslant 0$ et $|f(x)| \leqslant N < +\infty$ pour $x \in [-M, M]$ (N existe car f est continue), on a

$$\int_{\mathbf{R}} |f|^p d\lambda = \int_{[-M,M]} |f|^p d\lambda \leqslant 2MN^p < +\infty.$$

Noter aussi qu'il n'est pas nécessaire, pour les fonctions continues, de considérer un espace quotient : si $f = g$ presque partout et f, g sont continues, on a $f = g$. En effet, le complémentaire $Z = \mathbf{R} - Y$ d'un ensemble de mesure de Lebesgue nulle Y est *dense* dans $\mathbf{R}$. (Preuve : soit $x \in \mathbf{R}$, U un voisinage ouvert de x ; il existe $\varepsilon > 0$ tel que $]x-\varepsilon, x+\varepsilon[\subset U$, donc $\mu(U) \geqslant 2\varepsilon > 0$, et alors $U \cap Z \neq \emptyset$ car sinon $U \subset Y$). En particulier, $\|f\|_p = 0$ implique $f = 0$ *partout* si f est continue.

3.2. Exemples probabilistes : le lemme de Borel-Cantelli et la loi des grands nombres

Dans cette section nous présentons des premiers résultats proprement probabilistes. On considère fixé un espace probabilisé (Ω, Σ, P).

Le premier résultat est le lemme de Borel-Cantelli. Il s'agit de répondre à la question intuitive suivante : si l'on a une suite d'événements indépendants A_n, $n \geqslant 1$, comment savoir s'il existe une probabilité > 0 qu'une infinité des A_n soit vérifiés ?

Soit A l'événement correspondant. On a

$$A = \bigcap_{N\geqslant 1} \bigcup_{n\geqslant N} A_n \in \Sigma$$

donc A est effectivement mesurable. (Pour vérifier cette formule, remarquer que ω appartient au membre de droite si et seulement pour tout N il existe $n \geqslant N$ tel que $\omega \in A_n$, et ceci équivaut à dire que $\sup\{n \mid \omega \in A_n\} = +\infty$, ou que ω appartient à une infinité des A_n).

Proposition 3.2.1. *Soient A_n, $n \geqslant 1$, des événements, A comme ci-dessus. Posons*

$$p = \sum_{n\geqslant 1} P(A_n) \in [0, +\infty]. \tag{3.13}$$

(1) *Quels que soient les A_n, si $p < +\infty$, alors $P(A) = 0$.*
(2) *Supposons que les A_n sont indépendants. Alors si $p = +\infty$, on a $P(A) = 1$.*

Démonstration. Le point (1) est immédiat : on a par monotonie

$$P(A) \leqslant P\Big(\bigcup_{n\geqslant N} A_n\Big) \leqslant \sum_{n\geqslant N} P(A_n),$$

pour tout $N \geqslant 1$, et l'hypothèse $p < +\infty$ montre que cette quantité tend vers 0 quand $N \to +\infty$, donc $P(A) = 0$.

Pour le point (2), on remarque que $\omega \in A$ si et seulement si

$$\sum_{n\geqslant 1} \chi_{A_n}(\omega) = +\infty, \text{ si et seulement si } \exp\Big(-\sum_{n\geqslant 1} \chi_{A_n}(\omega)\Big) = 0. \tag{3.14}$$

Considérons donc, pour $N \geqslant 1$, l'intégrale

$$\int \exp\Big(-\sum_{n\leqslant N} \chi_{A_n}\Big)dP = \int \prod_{n\leqslant N} \exp(-\chi_{A_n})dP.$$

Or les A_n, $n \leqslant N$, sont indépendants, et cela implique aussitôt que

$$\int \prod_{n\leqslant N} \exp(-\chi_{A_n})dP = \prod_{n\leqslant N} \int \exp(-\chi_{A_n})dP.$$

Chacune des intégrales est aisée à calculer séparément :

$$\int \exp(-\chi_{A_n})dP = e^{-1}P(A_n) + 1 - P(A_n).$$

Soit $c = 1 - e^{-1} \in]0, 1[$. Il vient

$$\log \prod_{n\leqslant N} \int \exp(-\chi_{A_n})dP = \sum_{n\leqslant N} \log(1 - cP(A_n)).$$

On sait que $\log(1 - x) \leqslant -x$ pour $x \geqslant 0$, et donc

$$\log \prod_{n\leqslant N} \int \exp(-\chi_{A_n})dP \leqslant -c \sum_{n\leqslant N} P(A_n) \to -\infty \text{ quand } N \to +\infty$$

d'après l'hypothèse $p = +\infty$. Retraçant le chemin, cela signifie (après avoir appliqué le théorème de convergence dominée) que

$$\int \exp\Big(-\sum_{n\geqslant 1} \chi_{A_n}\Big)dP = 0,$$

ce qui veut dire que (3.14) est vrai presque partout. □

Remarque 3.2.2. Dans (2), la condition d'indépendance est nécessaire : si $A_n = A_1$ pour tout $n \geqslant 1$ et $0 < P(A_1) < 1$, alors la série (3.13) diverge, mais $P(A) = P(A_1) < 1$.

Notre second résultat probabiliste sera une version simple (mais non-triviale) de la loi (forte) des grands nombres. Ici, la situation est la suivante : on a une suite (X_n), $n \geqslant 1$, de variables aléatoires *indépendantes* et *uniformément distribuées*, c'est à dire que la loi $\mu = X_n(P)$ de X_n est indépendante de n. Cette condition implique en particulier, si X_n est intégrable, (resp. de carré intégrable) que

$$E(X_n) = E(X_1) = \int x d\mu \text{ et } V(X_n) = V(X_1) = \int (x - E(X_1))^2 d\mu \text{ pour } n \geqslant 1,$$

l'espérance (resp. la variance) des X_n est donc constante.

Un modèle mathématique est donné par l'exemple des chiffres du développement en base b d'un $x \in [0, 1] = \Omega$ (voir l'Exemple 1.3.2, (2)), et un modèle intuitif est celui de la répétition infinie d'une expérience, et de la mesure d'une certain quantité associée, de sorte que chaque expérience soit indépendante des autres, par exemple une partie de pile-ou-face infinie. Dans ce dernier cas, fort classique, X_n prend deux valeurs $X_n \in \{\mathsf{p}, \mathsf{f}\}$, et la loi de X_n est le plus souvent supposée être uniforme

$$P(X_n = \mathsf{p}) = P(X_n = \mathsf{f}) = \frac{1}{2}. \tag{3.15}$$

Toutefois le cas d'une loi « biaisée »

$$P(X_n = \mathsf{p}) = p, \quad P(X_n = \mathsf{f}) = 1 - p$$

avec $p \in]0, 1[$ fixé, est également intéressant.

Noter, dans le cas uniforme, la ressemblance avec le cas précédent du développement en base $b = 2$, si l'on fait la correspondance, par exemple, $\mathsf{p} \mapsto 0$, $\mathsf{f} \mapsto 1$.

L'intuition dit que pour un grand nombre n d'expériences répétées à l'identique et sans dépendances entre elles, la « moyenne empirique »

$$\frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n}$$

devrait être proche de la véritable moyenne, à savoir l'espérance commune (si X_n est intégrable)

$$E(X_n) = E(X_1) = \int x d\mu.$$

Cela semble d'autant plus raisonnable que, d'une part, on a $E(S_n/n) = E(X_1)$ par linéarité et indépendance, et d'autre part d'après (3.9), la variance de S_n/n est

$$V\Big(\frac{S_n}{n}\Big) = \frac{V(S_n)}{n^2} = \frac{1}{n^2}(V(X_1) + \cdots + V(X_n)) = \frac{V(X_1)}{n} \tag{3.16}$$

qui tend vers 0, montrant que S_n/n devrait se concentrer vers cette moyenne.

Il est possible de confirmer cette intuition sous diverses formes. Commençons par la loi faible des grands nombres, qui nous permet d'introduire la notion de convergence en probabilité.

Définition 3.2.3. Soit (Ω, Σ, P) un espace probabilisé, (X_n), $n \geqslant 1$, une suite de variables aléatoires et X une variable aléatoire. On dit que X_n converge vers X en probabilité si pour tout $\varepsilon > 0$ on a

$$\lim_{n \to +\infty} P(|X_n - X| > \varepsilon) = 0.$$

Remarque 3.2.4. Plus généralement, si (f_n) est une suite de fonctions mesurables sur un espace mesuré $(X, \mathcal{M}, \mu)$, on dit que (f_n) converge en mesure vers f si pour tout $\varepsilon > 0$ on a

$$\mu(\{x \in X \mid |f_n(x) - f(x)| > \varepsilon\}) \to 0$$

quand $n \to +\infty$.

Proposition 3.2.5. *Soit (Ω, Σ, P) un espace probabilisé, (X_n), $n \geqslant 1$, une suite de variables aléatoires indépendantes et uniformément distribuées telles que $X_n \in L^2$. Alors S_n/n converge en probabilité vers la fonction constante $E(X_1)$, c'est à dire*

$$P\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - E(X_1)\Big| > \varepsilon\Big) \to 0$$

quand $n \to +\infty$, pour tout $\varepsilon > 0$.

La preuve est extrêmement simple, mais permet d'introduire une inégalité utile, dite inégalité de Chebychev–Byenaimé–Markov, que nous énonçons sous deux formes :

Proposition 3.2.6. (1) *Soit (Ω, Σ, P) un espace probabilisé, X une variable aléatoire et $p \geqslant 1$. Si $X \in L^p(\Omega)$, alors on a pour tout $\varepsilon > 0$*

$$P(|X| \geqslant \varepsilon) \leqslant \varepsilon^{-p} E(|X|^p). \tag{3.17}$$

(2) *Soit $(X, \mathcal{M}, \mu)$ un espace mesuré, $p \geqslant 1$ et $f \in L^p(\mu)$. Alors on a pour tout $\varepsilon > 0$*

$$\mu(\{x \mid |f(x)| \geqslant \varepsilon\}) \leqslant \varepsilon^{-p} \int_X |f|^p d\mu = \varepsilon^{-p} \|f\|_p^p. \tag{3.18}$$

Démonstration. Les deux formes sont évidemment équivalentes, prenons la première par exemple (on vérifiera que la preuve ne dépend pas de la condition $P(\Omega) = 1$). Soit $A = \{|X| > \varepsilon\}$. On a alors tautologiquement

$$\varepsilon\chi_A \leqslant |X|$$

donc aussitôt

$$\varepsilon^p P(A) \leqslant E(|X|^p).$$

□

Remarque 3.2.7. Cette inégalité montre, en particulier, que si $f_n \to f$ dans L^1, alors f_n converge vers f en mesure :

$$\mu(\{x \mid |f_n(x) - f(x)| > \varepsilon\}) \leqslant \varepsilon^{-1}\|f_n - f\|_1.$$

Par contre, si f_n converge vers f presque partout, il n'est pas forcément vrai que f_n converge vers f en mesure : la suite $f_n = \chi_{]n,n+1[}$ de fonctions sur $\mathbf{R}$ vérifie $f_n \to 0$ (presque) partout, mais $\lambda(\{x \mid |f_n(x)| \geqslant 1\}) = 1$ pour tout n.

Par contre, si μ est finie, en particulier dans le cas probabiliste, la convergence presque partout implique la convergence en mesure (resp. en probabilité). En effet, soit $\varepsilon > 0$ fixé et considérons

$$A_n = \{x \mid |f_n(x) - f(x)| > \varepsilon\}$$
$$B_k = \bigcup_{n\geqslant k} A_n \text{ et } B = \bigcap_{k\geqslant 1} B_k.$$

On a $B_{k+1} \subset B_k$ et l'hypothèse implique $\mu(B) = 0$ (comparer avec (1.8)). Comme $\mu(X) < +\infty$, on a donc

$$\lim_{k\to+\infty} \mu(B_k) = 0$$

d'après la Proposition 1.2.3, (4), et comme $\mu(A_n) \leqslant \mu(B_n)$ par monotonie, on en déduit que $\mu(A_n) \to 0$, c'est à dire que (f_n) converge en mesure vers f.

Preuve de la Proposition 3.2.5. Puisque $X_n \in L^2$, on a $S_n \in L^2$, et on peut appliquer (3.17) avec $p = 2$, ce qui donne, pour $\varepsilon > 0$

$$P\Bigl(\Bigl|\frac{S_n}{n} - E(X_1)|\Bigr| > \varepsilon\Bigr) \leqslant \varepsilon^{-2}E\Bigl(\Bigl|\frac{S_n}{n} - E(X_1)\Bigr|^2\Bigr) = \varepsilon^{-2}E\Bigl(\Bigl|\frac{S_n}{n} - E\Bigl(\frac{S_n}{n}\Bigr)\Bigr|^2\Bigr)$$
$$= \varepsilon^{-2}V(S_n/n) = \varepsilon^{-2}n^{-1}V(X_1) \to 0 \text{ quand } n \to +\infty,$$

d'après (3.16). □

En général la convergence en probabilité est significativement plus faible que la convergence presque sûre. Par exemple, la suite (f_n) de l'Exercice 3.1.6 converge en mesure vers 0 car on a $P(|f_n| \geqslant \varepsilon) \leqslant 2^{-k}$ si $n \geqslant 2^k$, mais ne converge pas presque partout.

La loi forte des grands nombres indique que la convergence presque sûre est vraie pour S_n/n si X_n est intégrable. Une preuve simple est possible sous une hypothèse un peu plus forte, mais néanmoins souvent vérifiée.

Théorème 3.2.8. *Soit (X_n) une suite de variables aléatoires indépendantes uniformément distribuées de loi μ. Supposons que les X_n sont bornées presque sûrement. Alors on a*

$$\frac{X_1 + \cdots + X_n}{n} \to E(X_1) = \int_{\mathbf{C}} x d\mu$$

presque sûrement.

Démonstration. On va montrer le résultat sous une hypothèse sensiblement plus faible : on suppose que $X_n \in L^4$ (cette condition ne dépend que de la loi des X_n, comme la précédente). Bien entendu, si $|X_n| \leqslant M$ pour presque tout ω, on a $E(|X_n|^4) \leqslant M^4$, donc il s'agit bien d'un affaiblissement.

La méthode de démonstration est similaire à celle du lemme de Borel-Cantelli. On considère la série

$$\sum_{n\geqslant 1}\Big|\frac{X_1+\cdots+X_n}{n}-E(X_1)\Big|^4,$$

qui est une variable aléatoire à valeurs dans $[0,+\infty]$. Le résultat voulu sera démontré si l'on parvient à démontrer que la série converge presque sûrement : en effet, dans ce cas le terme général doit converger vers 0 presque sûrement, ce qui est la conclusion du théorème. Similairement, la série converge presque sûrement si

$$\int\sum_{n\geqslant 1}\Big|\frac{X_1+\cdots+X_n}{n}-E(X_1)\Big|^4 dP=\sum_{n\geqslant 1}\int\Big|\frac{X_1+\cdots+X_n}{n}-E(X_1)\Big|^4 dP<+\infty. \tag{3.19}$$

Posons $Y_n=X_n-E(X_n)=X_n-E(X_1)$, donc $E(Y_n)=0$. Le terme général de la série (3.19) est

$$\frac{1}{n^4}E(|Y_1+\cdots+Y_n|^4)=\frac{1}{n^4}E((Y_1+\cdots+Y_n)^2(\bar{Y}_1+\cdots+\bar{Y}_n)^2)$$

En développant la puissance quatrième on fait apparaître par linéarité des termes du type $E(Y_pY_q\bar{Y}_r\bar{Y}_s)$. Or pour des variables aléatoires indépendantes X et Y on a $E(XY)=E(X)E(Y)$ (Exercice 2.3.6) et de plus X^a et $\bar{Y}^b$ sont indépendantes pour tout $a\geqslant 1$, $b\geqslant 1$ (Proposition 1.2.11, (2)), et même Y_pY_q et $\bar{Y}_r\bar{Y}_s$ sont indépendantes si $\{p,q\}\cap\{r,s\}=\emptyset$ (Exercice 1.2.12).

Ainsi, par exemple, si p, q, r et s sont distincts on a

$$E(Y_pY_q\bar{Y}_r\bar{Y}_s)=E(Y_p)E(Y_q)E(\bar{Y}_r)E(\bar{Y}_s)=0,$$

et si $p\neq q$, $r=s$, on a

$$E(Y_pY_q|Y_r|^2)=E(Y_p)E(Y_q)E(|Y_r|^2)=0,\quad \text{etc}$$

et finalement tout les termes $E(Y_pY_q\bar{Y}_r\bar{Y}_s)$ sont nuls sauf ceux du type $E(Y_i^2\bar{Y}_j^2)$ ou $E(|Y_iY_j|^2)$ avec $i\neq j$ et ceux du type $E(|Y_i|^4)$. Comme les Y_i sont uniformément distribuées, on a

$$E(|Y_i|^4)=E(|Y_1|^4),\quad E(Y_i^2\bar{Y}_j^2)=|E(Y_1^2)|^2 \text{ et } E(|Y_iY_j|^2)=E(|Y_1|^2)^2$$

et il vient finalement en comptant les différents types de termes restants

$$\begin{aligned}\frac{1}{n^4}E(|Y_1+\cdots+Y_n|^4)&=\frac{1}{n^4}\Big(\sum_i E(|Y_i|^4)+2\sum_{i\neq j}E(Y_i^2\bar{Y}_j^2)+4\sum_{i\neq j}E(|Y_iY_j|^2)\Big)\\&=\frac{1}{n^4}\Big(nE(Y_1^4)+n(n-1)|E(Y_1^2)|^2+2n(n-1)E(|Y_1|^2)^2\Big).\end{aligned}$$

Par comparaison avec les séries convergentes $\sum n^{-2}$ et $\sum n^{-3}$, on en déduit que la série (3.19) converge. ☐

Exemple 3.2.9. (1) Voici d'abord un exemple d'utilisation du lemme de Borel-Cantelli (assez artificiel ?) Considérons le jeu de pile ou face répété un grand nombre de fois, comment décrit ci-dessus, modélisé par une suite de variables aléatoires X_n, supposées avoir comme loi commune la loi de probabilité donnée par (3.15). Divisons les expériences $X_1,\ldots,X_n,\ldots$ en blocs successifs de longueur $2k$, $k\geqslant 1$. Le premier bloc est donc donné par X_1, X_2, le second par $X_3,\ldots,X_6$, etc...

Considérons l'événement $A_k=\{\text{il y a autant de pile que de face dans le bloc } k\}$. Les A_k sont indépendants parce qu'ils sont définis en fonction des valeurs de X_n pour n appartenant à des ensembles disjoints pour différentes valeurs de k.

Quelle est la probabilité de A_k ? Si l'on considère la suite $(Y_1,\ldots,Y_{2k})$ des variables aléatoires définissant A_k, on a $\omega\in A_k$ si et seulement si

$$|\{i\leqslant 2k \mid Y_k(\omega)=\mathsf{p}\}|=k.$$

Chacune des suites de $2k$ symboles étant équiprobable, on a donc

$$P(A_k)=\frac{a_k}{2^{2k}}$$

où a_k est le nombre de sous-ensembles à k éléments dans un ensemble de $2k$ éléments, donc $a_k = \binom{2k}{k}$.

Comme

$$2^{2k} = (1+1)^{2k} = \sum_{i=0}^{2k} \binom{2k}{i}$$

et $\binom{2k}{i} \leqslant \binom{2k}{k}$ pour $0 \leqslant i \leqslant 2k$, on a

$$2^{2k} \leqslant 2k\binom{2k}{k}$$

et donc $P(A_k) \geqslant (2k)^{-1}$.

En particulier la série $\sum P(A_k)$ diverge et on conclut qu'il existe des blocs arbitrairement longs de tirages qui sont exactement équilibrés.

(2) Soit $b \geqslant 2$ et X_n la suite de variables aléatoires correspondant au développement en base b. Les X_n sont bornées donc le Théorème 3.2.8 peut être appliqué. Considérons plutôt un chiffre d_0 fixé et soit χ la fonction caractéristique de $d_0 \in \mathbf{R}$. Les variables aléatoires $Y_n = \chi(X_n)$ à valeurs dans $\{0,1\}$ sont encore indépendantes (Proposition 1.2.11) et de même loi donnée par

$$P(Y_n = 1) = \frac{1}{b}, \text{ et } P(Y_n = 0) = \frac{b-1}{b}.$$

Dans ce cas on $Y_1 + \cdots + Y_n = |\{k \leqslant n \mid X_n = d_0\}|$ et $E(Y_n) = b^{-1}$ donc la loi des grands nombres dit que presque sûrement on a

$$\lim_{n\to+\infty} \frac{|\{k \leqslant n \mid X_n = d_0\}|}{n} = \frac{1}{b},$$

c'est à dire, traduit en termes concrets, pour presque tout $x \in [0,1]$, il y a la même proportion $1/b$ de chaque chiffre $0, \ldots, b-1$, dans le développement de x en base b.

Puisque une intersection d'événements presque sûrs l'est également, on peut même dire que presque tout x possède cette propriété *dans toute base* $b \geqslant 2$. Un tel $x \in [0,1]$ est dit normal en toute base b.

Noter que de ce point de vue, on peut juger de la complexité de l'ensemble exceptionnel des réels $x \in [0,1]$ pour lesquels cette propriété est fausse... Tout les rationnels en font évidemment partie (dans une base $b = 10^k$ assez grande, par exemple, leur développement ne comporte qu'un seul chiffre à partir d'un certain rang...)

Il est d'ailleurs significatif qu'il soit extrêmement difficile de fournir un seul exemple *explicite* de nombre réel normal en toute base b, bien que l'on ait montré que presque tous possèdent cette propriété ; bien que l'on conjecture que de nombreuses constantes naturelles (par exemple, $\pi - 3$) soit normales en toute base, *aucun* exemple « concret » n'est connu à l'heure actuelle.

De manière générale, la loi des grands nombres suggère encore de nombreuses questions. La première est de savoir si l'on peut affaiblir l'hypothèse $X_n \in L^4$, puisque l'énoncé ne requière que l'existence de l'espérance. C'est en effet le cas, comme l'a démontré Kolmogorov : la démonstration est cependant nécessairement plus délicate.[1]

Une autre question est : à quelle vitesse est-ce que les sommes $S_n/n = (X_1 + \cdots + X_n)/n$ s'approchent de l'espérance $E(X_1)$? Supposons, quitte à considérer $Y_n = E_n - E(X_n)$, que $E(X_n) = 0$. La question devient : quel est l'ordre de grandeur de S_n/n quand $n \to +\infty$?

On peut voir que la preuve du Théorème 3.2.8 fournit en fait un énoncé plus fort : toujours si $E(X_n) = 0$, on a

$$\frac{S_n}{n^\alpha} \to 0 \text{ presque sûrement}$$

[1]. Un élément essentiel de la démonstration est *l'inégalité de Kolmogorov* qui est le sujet de l'Exercice 3 de l'examen partiel.

pour tout $\alpha > 1/2$ (il faut $4\alpha > 2$). Il se trouve que ce n'est pas un hasard et que le bon ordre de grandeur de S_n est effectivement $\sqrt{n}$, en un certain sens, plus faible que la convergence presque sûre, et différent également de la convergence en probabilité.

Remarquons que $E(n^{-1/2}S_n) = 0$ et $V(n^{-1/2}S_n) = V(X_1)$, ce qui montre que $n^{-1/2}S_n$ est d'espérance 0 et variance constante pour tout $n \geqslant 1$, et suggère l'existence d'une loi de probabilité ayant cette espérance et cette variance dont S_n devient « proche »! C'est là le contenu du théorème central-limite.[2]

Théorème 3.2.10. *Soit (X_n) une suite de variables aléatoires réelles indépendantes et uniformément distribuées, telles que $X_n \in L^2(\Omega)$ et dont l'espérance est $E(X_n) = 0$. Soit $\sigma = V(X_n)$ la variance des X_n. Pour tout $a \in \mathbf{R}$, on a*

$$P\Big(\frac{X_1 + \cdots + X_n}{\sqrt{n}} \leqslant a\Big) \to \frac{1}{\sqrt{2\pi\sigma}} \int_{[-\infty,a]} e^{-\frac{t^2}{2\sigma}} dt$$

quand $n \to +\infty$.

Remarque 3.2.11. La loi de probabilité $\mu_{0,\sigma}$ sur $\mathbf{R}$ donnée par

$$\mu_{0,\sigma} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{t^2}{2\sigma}} d\lambda$$

est appelée la *loi normale centrée de variance σ*, ou loi gaussienne centrée de variance σ. Il n'est pas évident que $\int d\mu_{0,\sigma} = 1$, cela sera démontré plus tard, cf. Chapitre 4.

Le type de convergence dans ce théorème est appelé la *convergence en loi*. Il s'agit d'une notion plus faible que la convergence en probabilité (dans la situation du théorème central-limite, il est possible de montrer que $S_n/n - E(X_1)$ ne converge jamais en probabilité). Cette notion est développée dans la Section 5.6.

Le théorème de la limite centrale sera démontré dans la Section 9.2 finale du cours.

3.3. Fonctions définies par une intégrale

Une des utilisations les plus fréquentes de l'intégrale est pour définir d'autres fonctions en intégrant une fonction de deux variables par rapport à l'une d'entre elles seulement. En raison de l'effet de moyenne provoqué par l'intégration (en général), ces fonctions ont des propriétés de régularité souvent très agréables.

Soit (X, d) un espace métrique, $\mathcal{B}_X$ la tribu borélienne associée, $(Y, \mathcal{M}, \mu)$ un espace mesuré. Soit

$$h : X \times Y \to \mathbf{C}$$

une application mesurable (pour la tribu produit $\mathcal{B}_X \otimes \mathcal{M}$) telle que pour tout $x \in X$ fixé, la fonction $h_x : y \mapsto h(x, y)$ est μ-intégrable. On peut alors définir une fonction $f : X \to \mathbf{C}$ par intégration sur Y :

$$f(x) = \int_Y h(x, y) d\mu(y).$$

La fonction $h(x, y)$ est parfois appelée le « noyau ».

Pour établir des propriétés de régularité de f, il est nécessaire de renforcer l'hypothèse d'intégrabilité des fonctions h_x en introduisant une condition plus forte de domination similaire à celle qui intervient dans le théorème de convergence dominée : on suppose qu'il existe $g \in L^1(\mu)$ telle que

$$|h(x, y)| \leqslant g(y) \tag{3.20}$$

pour tout $(x, y) \in X \times Y$.

Dans ces conditions on a d'abord un énoncé simple de continuité sous le signe somme.

[2] Historiquement (comme me l'a indiqué E. Charpentier), en allemand, le théorème central de la théorie des probabilités concernant la limite etc...

Proposition 3.3.1. *Soit $x_0 \in X$ et supposons que pour presque tout $y \in Y$, la fonction*

$$h(\cdot, y) \,:\, x \mapsto h(x, y)$$

sur l'espace métrique X soit continue en x_0. Alors f est continue en x_0.

Démonstration. Puisque X est un espace métrique, la continuité peut être testée à l'aide des suites, et il suffit de montrer que pour toute suite $x_n \to x_0$, on a $f(x_n) \to f(x_0)$.

Pour cela, soit $u_n \,:\, Y \to \mathbf{C}$ la fonction $u_n(y) = h(x_n, y)$. Par hypothèse on a $u_n(y) = h(x_n, y) \to h(x_0, y)$ pour presque tout y, et de plus $|u_n(y)| \leqslant g(y)$ avec $g \in L^1(\mu)$. On peut donc appliquer le théorème de convergence dominée et obtenir

$$\lim_{n\to+\infty} \int_Y u_n(y)d\mu(y) = \int_Y h(x_0, y)d\mu(y), \text{ c'est à dire } \lim_{n\to+\infty} f(x_n) = f(x_0),$$

d'où le résultat. □

Souvent la fonction $h(\cdot, y)$ sera continue sur tout X, et alors f sera continue automatiquement sur X, mais le cas général peut être utile.

Exemple 3.3.2. Prenons $Y = [0, 1]$ muni de la mesure de Lebesgue notée dy, et pour $f \in L^1(Y)$, soit $g \,:\, [0, 1] \to \mathbf{C}$ la fonction « primitive »

$$g(x) = \int_0^x f(y)dy.$$

On peut écrire

$$g(x) = \int_Y h(x, y)dy$$

avec $h(x, y) = f(y)\chi_{[0,x]}(y)$. La fonction h vérifie

$$|h(x, y)| \leqslant |f(y)| \text{ pour tout } x \in [0, 1],$$

or $f \in L^1([0, 1])$ donc (3.20) est vérifiée. De plus pour y fixé on a

$$h(x, y) = f(y)\chi_{[0,x]}(y) = \begin{cases} f(y) & \text{si } 0 \leqslant y \leqslant x \\ 0 & \text{sinon} \end{cases}$$

qui est continue en x_0 si $x_0 \neq y$ (elle est alors constante au voisinage de x_0), en particulier pour presque tout y. D'après la proposition, la fonction g est donc continue sur $[0, 1]$. Noter que f n'est pas forcément bornée et donc la méthode usuelle pour borner $f(x+h) - f(x)$ en majorant f n'est pas applicable.

Si $X = I$ est un intervalle de $\mathbf{R}$ (avec la métrique induite, bien évidemment) on peut demander de plus si f est dérivable. Il faut bien sûr ajouter une hypothèse pour cela (en plus de (3.20), toujours supposée vraie).

Proposition 3.3.3. *Supposons que $X = I \subset \mathbf{R}$ est un intervalle ouvert et qu'il existe $g_1 \in L^1(Y, \mu)$ telle que pour presque tout $y \in Y$, la fonction*

$$h(\cdot, y) \,:\, I \to \mathbf{C}$$

est dérivable et sa dérivée vérifie

$$\Big|\frac{d}{dx}h(x, y)\Big| \leqslant g_1(y). \tag{3.21}$$

Alors f est dérivable sur I et on a

$$f'(x) = \int_Y \frac{d}{dx}h(x, y)d\mu(y).$$

Autrement dit, sous une hypothèse raisonnable, on peut « dériver sous le signe d'intégration ».

Démonstration. Notons

$$f_1(x) = \int_Y \frac{d}{dx} h(x,y) d\mu(y),$$

qui est bien définie d'après l'hypothèse.

Soit $x_0 \in I$ fixé. Il existe un intervalle ouvert $J =]-\alpha, \alpha[$ tel que $x_0 + \delta \in I$ pour $\delta \in J$. Considérons la fonction $\psi : J \times Y \to \mathbf{C}$ telle que

$$\psi(\delta, y) = \begin{cases} \dfrac{h(x_0 + \delta, h) - h(x_0, y)}{\delta} & \text{si } \delta \neq 0 \\ \dfrac{d}{dx} h(x_0, y) & \text{si } \delta = 0. \end{cases}$$

Pour presque tout $y \in Y$ fixé la fonction $\psi(\delta, y)$ est continue en $\delta = 0$ par définition même de la dérivabilité de $h(\cdot, y)$ en x_0.

Vérifions alors l'hypothèse (3.20) : pour $\delta = 0$, on a

$$|\psi(0, y)| \leqslant g_1(y)$$

d'après (3.20) et pour $\delta \neq 0$, il existe $\eta \in [0, 1]$ tel que

$$\begin{aligned} |\psi(\delta, y)| &= \Big| \frac{h(x_0 + \delta, y) - h(x_0, y)}{\delta} \Big| \\ &= \Big| \frac{d}{dx} h(x_0 + \eta\delta, y) \Big| \leqslant g_1(y) \end{aligned}$$

d'après le théorème des accroissements finis. Ainsi (3.21) est valide pour ψ avec $g = g_1$.

Par la proposition précédente, la fonction

$$\varphi(\delta) = \int_Y \psi(\delta, y) d\mu(y)$$

est donc définie et continue en 0. Comme

$$\varphi(0) = \int_Y \frac{d}{dx} h(x_0, y) d\mu(y) = f_1(x_0),$$

et pour $\delta \neq 0$

$$\varphi(\delta) = \frac{f(x_0 + \delta) - f(x_0)}{\delta},$$

la limite $\varphi(\delta) \to \varphi(0)$ signifie que f est dérivable en x_0 avec $f'(x_0) = f_1(x_0)$. □

3.4. Un exemple : la transformée de Fourier

Les résultats de la section précédente s'appliquent souvent très facilement. Ici nous donnons simplement la définition et les propriétés les plus simples de la *transformée de Fourier* d'une fonction intégrable sur $\mathbf{R}$, qui sera étudiée plus en détail dans les Chapitres 7 et 9.

On définit d'abord la fonction $e : \mathbf{C} \to \mathbf{C}$ par

$$e(z) = e^{2i\pi z}. \tag{3.22}$$

Cette fonction cousine de l'exponentielle vérifie

$$e(x + y) = e(x)e(y), \; e(x + 1) = e(x) \text{ et } |e(x)| = 1 \text{ si } x \in \mathbf{R}$$

(c'est pour avoir une fonction de période 1 que l'on modifie ainsi l'exponentielle). De plus

$$e(x) = 1 \text{ si et seulement si } x \in \mathbf{Z}.$$

Définition 3.4.1. Soit $f \in L^1(\mathbf{R})$. La transformée de Fourier de f est la fonction $\hat{f} : \mathbf{R} \to \mathbf{C}$ telle que

$$\hat{f}(t) = \int_{\mathbf{R}} f(y) e(-yt) dy.$$

Remarque 3.4.2. D'autres normalisations sont aussi utilisées, comme

$$\int_{\mathbf{R}} f(x)e^{ixt}dx, \quad \int_{\mathbf{R}} f(x)e^{-ixt}dx, \quad \frac{1}{\sqrt{2\pi}}\int_{\mathbf{R}} f(x)e^{-ixt}dx, \text{ etc...}$$

C'est la dernière qui apparaît dans [**R**]. La normalisation choisie ici a l'avantage de donner une formule d'inversion simple (cf. Section 7.3). Les trois autres ci-dessus se retrouvent facilement : elles sont de la forme, respectivement

$$\hat{f}(-t/2\pi), \quad \hat{f}(t/2\pi), \quad \frac{1}{\sqrt{2\pi}}\hat{f}(t/2\pi).$$

Proposition 3.4.3. (1) *Soit $f \in L^1(\mathbf{R})$. Alors $\hat{f}$ est une fonction continue bornée sur $\mathbf{R}$ telle que*

$$\|\hat{f}\|_\infty \leqslant \|f\|_1. \tag{3.23}$$

(2) *Si la fonction $g(x) = xf(x)$ est une fonction intégrable, alors $\hat{f}$ est de classe C^1 sur $\mathbf{R}$ et*

$$\hat{f}'(t) = -2i\pi\hat{g}(t).$$

(3) *Si f est classe C^1, $f(x) \to 0$ quand $|x| \to +\infty$ et $g = f' \in L^1(\mathbf{R})$, alors*

$$\hat{g}(t) = 2i\pi t\hat{f}(t).$$

Démonstration. La transformée de Fourier est une fonction définie par une intégrale comme à la section précédente, avec $h(t,y) = f(y)e(-yt)$. Comme $|h(t,y)| = |f(t)|$ pour $x \in \mathbf{R}$, la condition (3.20) est vérifiée pour $f \in L^1(\mathbf{R})$. De plus h est continue sur $\mathbf{R}$ pour tout y fixé (c'est une constante multipliée par une fonction exponentielle) et la Proposition 3.3.1 montre donc que $\hat{f}$ est définie et continue sur $\mathbf{R}$.

De plus on a (3.23) comme application directe de (3.11) et $|e(x)| = 1$.

Pour (2), la fonction h est également dérivable sur $\mathbf{R}$ pour tout y fixé avec

$$\frac{d}{dt}h(t,y) = -2i\pi yf(y)e(yt) \text{ donc } \left|\frac{d}{dt}h(t,y)\right| \leqslant 2\pi|yf(y)| = 2\pi|g(y)|.$$

Si $g \in L^1(\mathbf{R})$ on peut donc appliquer cette fois la Proposition 3.3.3 qui montre que $\hat{f}$ est dérivable avec

$$\hat{f}'(t) = \int_{\mathbf{R}} 2i\pi yf(y)e(yt)dy = -2i\pi\hat{g}(t).$$

Pour (3), si $g = f'$ existe et est intégrable et $f \to 0$ à l'infini, alors on peut intégrer par partie et obtenir

$$\hat{g}(t) = \int_{\mathbf{R}} f'(y)e(-yt)dy = 2i\pi t\int_{\mathbf{R}} f(y)e(-yt)dy$$

(comme ici f est dérivable, l'intégration par partie relève de l'intégrale de Riemann ; voir la Proposition 5.6.5 pour une version plus générale). □

Exemple 3.4.4. Soit $f : \mathbf{R} \to \mathbf{C}$ une fonction continue à support compact. Alors f est intégrable, et de plus pour tout $k \geqslant 1$, la fonction $x \mapsto x^k f(x)$ est également à support compact donc intégrable. Cela montre par récurrence que $\hat{f}$ est une fonction C^∞ et

$$\hat{f}^{(k)}(t) = (-2i\pi)^k\int_{\mathbf{R}} y^k f(y)e(-yt)dy.$$

Plus généralement ce résultat s'applique si la condition ci-dessus est vraie pour tout k, par exemple pour $f(y) = e^{-|y|}$ ou bien $f(y) = e^{-y^2}$.

Remarque 3.4.5. L'inégalité (3.23) montre que l'application « transformée de Fourier »

$$\begin{cases} L^1(\mathbf{R}) \to C_b(\mathbf{R}) \\ f \mapsto \hat{f} \end{cases}$$

est continue et de norme $\leqslant 1$, où $C_b(\mathbf{R})$, l'espace des fonctions continues bornées sur $\mathbf{R}$, est muni de la norme $\|f\|_\infty$. En particulier, si $f_n \to f$ dans L^1, alors $\hat{f}_n \to \hat{f}$ en norme L^∞, c'est à dire uniformément.

Remarque 3.4.6. Soit (Ω, Σ, P) un espace probabilisé et X une variable aléatoire réelle sur Ω. La *fonction caractéristique* de X est la fonction $\varphi : \mathbf{R} \to \mathbf{C}$ définie par

$$\varphi(t) = \int_\Omega e^{itX} dP = E(e^{itX}).$$

Comme $|e^{itX}| = 1$ puisque X est à valeurs réelles, φ est définie et la Proposition 3.3.1 montre que c'est une fonction continue.

C'est une variante de la transformée de Fourier : si $\mu = X(P)$ est la loi de X, on a

$$\varphi(t) = \int_{\mathbf{R}} e^{ity} d\mu(y)$$

par (2.10). En particulier, si la loi μ est de la forme $\mu = f d\lambda$ pour une fonction $f : \mathbf{R} \to [0, +\infty[$ (nécessairement intégrable puisque

$$\int_{\mathbf{R}} f d\mu = \int_\Omega dP = 1$$

par (2.10)), on a

$$\varphi(t) = \hat{f}\Big(-\frac{t}{2\pi}\Big).$$

Voir le Chapitre 9 pour les propriétés de la fonction caractéristique et ses applications.

CHAPITRE 4

Mesure et intégration sur les espaces produits

Parmi les quatre problèmes significatifs de l'intégrale de Riemann mentionnés dans le chapitre introductif, nous pouvons voir maintenant que trois sont résolus de manière essentiellement satisfaisante : les opérations de passage à la limite, y compris de dérivation sous le signe $\int$, se font le plus souvent aisément, les intégrales sur $\mathbf{R}$ sont traitées sur un pied d'égalité avec celles sur un intervalle compact, et une théorie des probabilités extrêmement puissante découle non seulement de l'usage de la théorie de la mesure, mais de toute la panoplie d'outils déjà mentionnés. Il reste le second point de la discussion introductive, à savoir les intégrales en dimension supérieure. La théorie de l'intégrale étant générale, il s'agit d'abord d'un problème de construction de mesures intéressantes sur $X \times Y$, étant données des mesures sur X et Y.

Nous allons donc présenter ici la résolution, caractéristiquement élégante, de ce problème.

4.1. Mesure produit

Soient $(X, \mathcal{M}, \mu)$ et $(Y, \mathcal{N}, \nu)$ des espaces mesurés. L'espace produit $X \times Y$ est muni de la tribu produit $\mathcal{M} \otimes \mathcal{N}$ (cf. Définition 1.1.6) engendrée par les *rectangles mesurables* $A \times B$ avec $(A, B) \in \mathcal{M} \times \mathcal{N}$.

On désire munir $X \times Y$ d'une mesure $\mu \otimes \nu$ telle que, pour $f : X \times Y \to \mathbf{C}$ intégrable, la formule d'interversion analogue de (0.6)

$$\int_X \int_Y f(x,y) d\nu(y) d\mu(x) = \int_Y \int_X f(x,y) d\mu(x) d\nu(y) \tag{4.1}$$

soit valide. Appliquée à la fonction caractéristique f d'un ensemble mesurable $C \in \mathcal{M} \otimes \mathcal{N}$, cela suggère (et requière) la formule suivante

$$(\mu \otimes \nu)(C) = \int_X \nu(t_x(C)) d\mu(x) = \int_Y \mu(t^y(C)) d\nu(y) \tag{4.2}$$

où

$$t_x(C) = \{y \mid (x,y) \in C\} = C \cap (\{x\} \times Y) \subset Y \tag{4.3}$$

$$t^y(C) = \{x \mid (x,y) \in C\} = C \cap (X \times \{y\}) \subset X \tag{4.4}$$

sont les « tranches » horizontales et verticales (si l'on veut) de l'ensemble C, cela provenant des formules

$$\int_Y f(x,y) d\nu(y) = \nu(t_x(C)) \text{ pour tout } x \in X$$

$$\int_X f(x,y) d\mu(x) = \mu(t^y(C)) \text{ pour tout } y \in Y$$

pour les intégrales de $f = \chi_C$ à x ou y fixé.

Ainsi (4.2) fournit a priori une définition de la mesure désirée. Il faut cependant vérifier qu'elle a un sens, c'est à dire que les fonctions positives $x \mapsto \nu(t_x(C))$ et $y \mapsto \mu(t^y(C))$ sont mesurables pour $\mathcal{M}$ et $\mathcal{N}$ respectivement ; et que les deux définitions de ce qui devrait être la même mesure sont effectivement équivalentes. On note déjà que si $C = A \times B$ est l'un des « rectangles » engendrant $\mathcal{M} \otimes \mathcal{N}$, on a bien l'identité voulue puisque

$$\int_X \int_Y \chi_C(x,y) d\nu(y) d\mu(x) = \int_Y \int_X \chi_C(x,y) d\mu(x) d\nu(y) = \mu(A)\nu(B)$$

dans ce cas.

Avant d'énoncer le premier résultat, rappelons la définition d'une mesure σ-finie, qui est indispensable dans la théorie des mesures produits (cf. Définition 1.2.1) : μ sur X est σ-finie s'il existe une suite (X_n) de sous-ensembles mesurables de mesure $\mu(X_n) < +\infty$ dont la réunion est X. Voir l'Exemple 4.1.7 ci-dessous pour constater que (4.2) ne peut fournir une mesure convenable sans hypothèse supplémentaire.

Les propriétés élémentaires suivantes de l'opération « tranche » peuvent être notées tout de suite. On les vérifie immédiatement, et l'on peut aussi noter que par définition on a $t_x(C) = i_x^{-1}(C)$ (resp. $t^y(C) = j_y^{-1}(C)$) où $i_x : Y \to X \times Y$ est donnée par $y \mapsto (x, y)$ (resp. $j_y : X \to X \times Y$ est donnée par $x \mapsto (x, y)$), et appliquer alors (1.2).

Lemme 4.1.1. *Pour tout $x \in X$ fixé on a*

$$t_x((X \times Y) - C) = Y - t_x(C) \tag{4.5}$$

$$t_x\Bigl(\bigcup_{i\in I} C_i\Bigr) = \bigcup_{i\in I} t_x(C_i) \tag{4.6}$$

$$t_x\Bigl(\bigcap_{i\in I} C_i\Bigr) = \bigcap_{i\in I} t_x(C_i). \tag{4.7}$$

Proposition 4.1.2. *Soient $(X, \mathcal{M}, \mu)$ et $(Y, \mathcal{N}, \nu)$ des espaces mesurés.*

(1) *Si $C \in \mathcal{M} \otimes \mathcal{N}$, la tranche $t_x(C) \subset Y$ appartient à la tribu $\mathcal{N}$ pour tout $x \in X$, ou de manière équivalente, pour tout $x \in X$, l'application $i_x : Y \to X \times Y$ est mesurable.*

Supposons de plus que Y est σ-fini.

(2) *Dans ce cas la fonction positive*

$$x \mapsto \nu(t_x(C))$$

est $\mathcal{M}$-mesurable pour tout $C \in \mathcal{M} \otimes \mathcal{N}$.

(3) *L'application*

$$\begin{cases} \mathcal{M} \otimes \mathcal{N} \to [0, +\infty] \\ C \mapsto \displaystyle\int_X \nu(t_x(C)) d\mu(x) \end{cases} \tag{4.8}$$

qui est bien définie pour ces raisons est une mesure sur $\mathcal{M} \otimes \mathcal{N}$.

Démonstration. Si (1) et (2) sont acquis, le point (3) est une conséquence immédiate. En effet posons

$$\pi(C) = \int_X \nu(t_x(C)) d\mu(x) \text{ pour } C \in \mathcal{M} \otimes \mathcal{N}.$$

Clairement $\pi(\emptyset) = 0$, et si (C_n), $n \geqslant 1$, est une suite de parties mesurables disjointes d'union C dans la tribu produit, par le lemme et l'additivité de ν il vient

$$\nu(t_x(C)) = \sum_{n\geqslant 1} \nu(t_x(C_n))$$

pour tout x, et par le théorème de convergence monotone

$$\pi(C) = \int_X \sum_{n\geqslant 1} \nu(t_x(C_n)) d\mu(x) = \sum_{n\geqslant 1} \int_X \nu(t_x(C_n)) d\mu(x) = \sum_{n\geqslant 1} \pi(C_n).$$

Pour montrer (1), on considère x fixé. Dire que i_x est mesurable signifie que $i_x^{-1}(C) \in \mathcal{N}$ pour tout $C \in \mathcal{M} \otimes \mathcal{N}$. Par le lemme 1.1.8, il suffit (par définition de la tribu produit) de vérifier que $i_x^{-1}(C) \in \mathcal{N}$ si $C = A \times B$ est un rectangle mesurable. Mais alors

$$i_x^{-1}(C) = t_x(C) = \begin{cases} \emptyset & \text{si } x \notin A \\ B & \text{si } x \in A \end{cases} \tag{4.9}$$

qui appartient toujours à $\mathcal{N}$.

L'idée pour montrer (2), similaire à celle de la preuve du Lemme 1.1.8, est de considérer l'ensemble des parties $C \in \mathcal{M} \otimes \mathcal{N}$ vérifiant les conclusions demandées, et de montrer qu'il s'agit d'une tribu contenant les rectangles mesurables. Soit donc

$$\mathcal{O} = \{C \in \mathcal{M} \otimes \mathcal{N} \mid x \mapsto \nu(t_x(C)) \text{ est } \mathcal{M}\text{-mesurable}\}.$$

Les rectangles mesurables $C = A \times B$ appartiennent à $\mathcal{O}$ d'après (4.9) puisque

$$\nu(t_x(C)) = \begin{cases} 0 & \text{si } x \notin A \\ \nu(B) & \text{si } x \in A \end{cases}$$

donc $x \mapsto \nu(t_x(A \times B))$ est même étagée.

Si $C \in \mathcal{O}$ et $D = X \times Y - C$ est son complémentaire on a $t_x(D) = Y - t_x(C)$: la mesure de cet ensemble n'est pas calculable directement si $\nu(Y) = \nu(t_x(C)) = +\infty$.

1ère étape : On suppose d'abord que $\nu(Y) < +\infty$.

Il vient donc $\nu(t_x(D)) = \nu(Y) - \nu(t_x(C))$, une fonction mesurable de x, donc $\mathcal{O}$ est stable par complémentaire. La stabilité par union, même finie, n'est pas non plus évidente. On constate que $\mathcal{O}$ vérifie les conditions suivantes :

(1) $\mathcal{O}$ contient les rectangles mesurables ;
(2) $\mathcal{O}$ est stable par complémentaire ;
(3) $\mathcal{O}$ est stable par réunion dénombrable disjointe ;
(4) $\mathcal{O}$ est stable par union (resp. intersection) dénombrable *croissante* (resp. décroissante).

Le point (3) provient de la formule

$$\nu\Big(t_x\Big(\bigcup_{n\geqslant 1} C_n\Big)\Big) = \sum_{n\geqslant 1} \nu(t_x(C_n)) \text{ si les } C_n \text{ sont disjoints}$$

et de la mesurabilité des limites de fonctions mesurables. De même (4) provient de la formule

$$\nu\Big(t_x\Big(\bigcup_{n\geqslant 1} C_n\Big)\Big) = \lim_{n\to+\infty} \nu(t_x(C_n))$$

si $C_1 \subset C_2 \subset \ldots \subset C_n \ldots$ est une suite croissante d'éléments de $\mathcal{O}$, le cas de l'intersection s'en déduisant par passage au complémentaire en utilisant l'hypothèse $\nu(Y) < +\infty$.

En particulier, avec la terminologie du Lemme 4.1.4 ci-dessous, $\mathcal{O}$ est une *classe monotone* contenant l'ensemble $\mathcal{E}$ des réunions finies de rectangles mesurables[1] ; ce dernier ensemble est une *algèbre d'ensembles*, c'est à dire qu'il est stable par complémentaire, union finie et intersection finie : si $C_1 = A_1 \times B_1$ et $C_2 = A_2 \times B_2$ sont des rectangles, on a (faire un dessin !)

$$C_1 \cap C_2 = (A_1 \cap A_2) \times (B_1 \cap B_2) \in \mathcal{E} \tag{4.10}$$

$$X \times Y - C_1 = ((X - A_1) \times (Y - B_1)) \cup ((X - A_1) \times B_1) \cup (A_1 \times (Y - B_1)) \in \mathcal{E} \tag{4.11}$$

$$C_1 \cup C_2 = (X \times Y) - \{(X \times Y - C_1) \cap (X \times Y - C_2)\} \in \mathcal{E}, \tag{4.12}$$

et donc le lemme appliqué à $\mathcal{O}$ et $\mathcal{A} = \mathcal{E}$ implique que $\mathcal{O} \supset \sigma(\mathcal{E}) = \mathcal{M} \otimes \mathcal{N}$, comme désiré.

2ème étape : On suppose seulement que Y est un espace mesuré σ-fini. Soient $Y_n \in \mathcal{N}$, $n \geqslant 1$, tels que

$$Y = \bigcup_{n\geqslant 1} Y_n \text{ et } \nu(Y_n) < +\infty.$$

On peut évidemment supposer que les Y_n sont disjoints. Pour tout $C \in \mathcal{M} \otimes \mathcal{N}$ et $x \in X$, on a une réunion disjointe

$$t_x(C) = \bigcup_{n\geqslant 1} t_x(C \cap Y_n) \text{ donc } \nu(t_x(C)) = \sum_{n\geqslant 1} \nu(t_x(C \cap Y_n)).$$

[1]. Une union finie de rectangles peut se réécrire comme une union finie disjointe de rectangles, cf. les formules ci-dessus...

D'après la première étape appliquée à $X \times Y_n$ (avec sur Y_n la mesure qui est induite par ν, donc finie) au lieu de $X \times Y$, chacune des fonctions $x \mapsto \nu(t_x(C \cap Y_n))$ est mesurable, et donc la fonction $x \mapsto \nu(t_x(C))$ et mesurable. $\square$

Rappelons aussi que l'on a déjà vu que si $C = A \times B$ est un rectangle mesurable, on a

$$\int_X \nu(t_x(C))d\mu(x) = \mu(A)\nu(B). \tag{4.13}$$

Corollaire 4.1.3. *Avec les hypothèse de la proposition, soit $f : X\times Y \to \mathbf{C}$ une application mesurable pour la tribu produit. Alors pour tout $x \in X$ fixé l'application $t_x(f) : y \mapsto f(x, y)$ est mesurable pour $\mathcal{N}$.*

Démonstration. On a $t_x(f) = f \circ i_x$ donc par composition, $t_x(f)$ est mesurable. $\square$

Voilà maintenant le lemme technique utilisé ci-dessus (appelé souvent théorème de la classe monotone).

Lemme 4.1.4. *Soit X un ensemble, $\mathcal{O}$ un ensemble de parties de X telle que*
(1) *Si $A_n \in \mathcal{O}$, et $A_n \subset A_{n+1}$ pour tout n, alors $\bigcup A_n \in \mathcal{O}$;*
(2) *Si $A_n \in \mathcal{O}$, et $A_n \supset A_{n+1}$ pour tout n, alors $\bigcap A_n \in \mathcal{O}$.*
Une telle partie est appelée une classe monotone. *Alors si $\mathcal{O} \supset \mathcal{A}$, et si $\mathcal{A}$ est une algèbre, c'est à dire stable par complémentaire, réunion finie et intersection finie, on a*

$$\mathcal{O} \supset \sigma(\mathcal{A}).$$

Démonstration. Soit $\mathcal{O}' \subset \mathcal{O}$ l'intersection de toutes les classes monotones contenant $\mathcal{A}$: il est immédiat qu'il s'agit encore d'une classe monotone (engendrée par $\mathcal{A}$). Il suffit de montrer que $\mathcal{O}' \supset \sigma(\mathcal{A})$. Pour cela il suffit de vérifier que $\mathcal{O}'$ est elle-même une algèbre : en effet, si tel est le cas, on peut transformer une union dénombrable en réunion croissante :

$$\bigcup_{n\geqslant 1} C_n = \bigcup_{N\geqslant 1} \Bigl(\bigcup_{1\leqslant n\leqslant N} C_n \Bigr) \in \mathcal{O}'$$

pour tout $C_n \in \mathcal{O}'$, ce qui montre que $\mathcal{O}'$ est une tribu contenant $\mathcal{A}$.

La vérification des trois propriétés de stabilité d'une algèbre est similaire et fastidieuse : tout d'abord, soit

$$\mathcal{G} = \{A \subset X \mid X - A \in \mathcal{O}'\} ;$$

on vérifie (parce que le complémentaire échange union croissante et intersection décroissante !) que $\mathcal{G} \supset \mathcal{A}$ est une classe monotone, donc $\mathcal{G} \supset \mathcal{O}'$, et la stabilité par complémentaire de $\mathcal{O}'$ en découle.

Reste à montrer que pour A, $B \in \mathcal{O}'$ on a $A \cup B \in \mathcal{O}'$. On procède par bonds : soit

$$\mathcal{G}_1 = \{A \subset X \mid A \cup B \in \mathcal{O}' \text{ pour tout } B \in \mathcal{A}\}.$$

On a $\mathcal{G}_1 \supset \mathcal{A}$ puisque $\mathcal{A}$ est stable par union finie, et on vérifie de nouveau que $\mathcal{G}_1$ est une classe monotone grâce aux formules

$$\Bigl(\bigcup_{n\geqslant 1} A_n\Bigr) \cup B = \bigcup_{n\geqslant 1} (A_n \cup B)$$
$$\Bigl(\bigcap_{n\geqslant 1} A_n\Bigr) \cup B = \bigcap_{n\geqslant 1} (A_n \cup B) ;$$

donc $\mathcal{G}_1 \supset \mathcal{O}'$, ce qui signifie que $\mathcal{O}'$ est stable par réunion *avec un élément de $\mathcal{A}$*.

Enfin on peut donc poser

$$\mathcal{G}_2 = \{A \subset X \mid A \cup B \in \mathcal{O}' \text{ pour tout } B \in \mathcal{O}'\}.$$

L'inclusion $\mathcal{G}_2 \supset \mathcal{A}$ provient de l'étape précédente, et on vérifie que $\mathcal{G}_2$ est toujours une classe monotone (ce sont les mêmes formules que ci-dessus): donc $\mathcal{G}_2 \supset \mathcal{O}'$, ce qui (avec une récurrence immédiate) montre la stabilité de $\mathcal{O}'$ par union finie... $\square$

Appliquant la Proposition 4.1.2 après avoir échangé l'ordre des facteurs X et Y, on voit que si X et Y sont tout deux σ-finis, alors $t^y(C) \in \mathcal{M}$ pour tout y, $y \mapsto \mu(t^y(C))$ est mesurable et

$$C \mapsto \int_Y \mu(t^y(C))d\nu(y) \tag{4.14}$$

est une mesure.

Proposition 4.1.5. *Soient $(X, \mathcal{M})$ et $(Y, \mathcal{N})$ des espaces mesurables.*

(1) *Soient μ_1 et μ_2 des mesures sur $(X \times Y, \mathcal{M} \otimes \mathcal{N})$ telles que $\mu_1(C) = \mu_2(C)$ pour tout rectangle mesurable $C = A \times B$, et telles qu'il existe une suite disjointe de rectangles mesurables $C_n \times D_n$ tels que*

$$X \times Y = \bigcup_{n \geqslant 1} (C_n \times D_n) \text{ et } \mu_i(C_n \times D_n) < +\infty.$$

Alors on a $\mu_1 = \mu_2$.

(2) *En particulier, si $(X \mathcal{M}, \mu)$ et $(Y, \mathcal{N}, \nu)$ sont des espaces σ-finis, les mesures définies sur $\mathcal{M} \otimes \mathcal{N}$ par (4.8) et (4.14) coïncident.*

Définition 4.1.6. Soient $(X, \mathcal{M}, \mu)$ et $(Y, \mathcal{N}, \nu)$ des espaces mesurés σ-finis. La mesure produit $\mu \otimes \nu$ est la mesure sur $\mathcal{M} \otimes \mathcal{N}$ telle que

$$(\mu \otimes \nu)(C) = \int_X \nu(t_x(C))d\mu(x) = \int_Y \mu(t^y(C))d\nu(y)$$

pour $C \in \mathcal{M} \otimes \mathcal{N}$.

Preuve de la Proposition 4.1.5. L'argument pour le point (1) est similaire à celui utilisé pour prouver la partie (2) de la Proposition précédente. Supposons tout d'abord que $\mu_1(X \times Y) = \mu_2(X \times Y) < +\infty$.

Soit

$$\mathcal{O} = \{C \in \mathcal{M} \otimes \mathcal{N} \mid \mu_1(C) = \mu_2(C)\},$$

qui contient donc par hypothèse les rectangles mesurables.

Les formules (4.10), (4.11) et (4.12) montrent que $\mathcal{O} \supset \mathcal{E}$, où $\mathcal{E}$ est l'algèbre des réunions finies de rectangles mesurables. Enfin, si C_n est une suite croissante de parties dans $\mathcal{O}$, on a par continuité de la mesure

$$\mu_1\Big(\bigcup_{n \geqslant 1} C_n\Big) = \lim_{n \to +\infty} \mu_1(C_n) = \lim_{n \to +\infty} \mu_2(C_n) = \mu_2\Big(\bigcup_{n \geqslant 1} C_n\Big),$$

et de même pour les intersections décroissantes (en utilisant $\mu_i(C_1) \leqslant \mu_i(X \times Y) < +\infty$.)

En définitive, $\mathcal{O}$ est une classe monotone contenant $\mathcal{E}$, et donc $\mathcal{O} \supset \mathcal{M} \otimes \mathcal{N}$ par le Lemme 4.1.4 encore.

Dans le cas général restant à traiter, on a par hypothèse

$$X \times Y = \bigcup (C_n \times D_n),$$

les $C_n \times D_n$ étant disjoints et de mesure (μ_1 ou μ_2) finie. Si $C \subset C_n \times D_n$ est un ensemble mesurable, le premier cas implique $\mu_1(C) = \mu_2(C)$. Pour tout $C \in \mathcal{M} \otimes \mathcal{N}$, on a la réunion disjointe dénombrable

$$C = \bigcup_{n \geqslant 1} (C \cap (C_n \times D_n))$$

et donc par additivité des mesures il vient

$$\mu_1(C) = \sum_n \mu_1(C \cap (C_n \times D_n)) = \sum_n \mu_2(C \cap (C_n \times D_n)) = \mu_2(C).$$

Le point (2) est une application directe de ce qui précède : posons

$$\mu_1(C) = \int_X \nu(t_x(C))d\mu(x)$$

$$\mu_2(C) = \int_Y \nu(t^y(C))d\nu(y)$$

pour $C \in \mathcal{M} \otimes \mathcal{N}$. D'après la Proposition 4.1.2, il s'agit bien de mesures, et de plus si $C = A \times B$ est un rectangle mesurable, on a $\mu_1(C) = \mu(A)\nu(B) = \mu_2(C)$. Le point (1) permet donc de conclure que $\mu_1 = \mu_2$. □

EXEMPLE 4.1.7. Il est effectivement nécessaire de supposer que X et Y vérifient une propriété de finitude pour obtenir une bonne théorie de la mesure produit. Prenons en effet $X = [0, 1]$ avec la tribu borélienne et la mesure de Lebesgue λ et $Y = [0, 1]$, mais avec la mesure de comptage ν sur la tribu $\mathcal{N}$ formée de toutes les parties de $[0, 1]$, de sorte que ν n'est pas σ-finie.

Soit $D = \{(x, x) \mid x \in [0, 1]\} \subset X \times Y$ la diagonale. C'est un ensemble $\mathcal{B} \otimes \mathcal{N}$-mesurable (car $\mathcal{B} \otimes \mathcal{N} \supset \mathcal{B} \otimes \mathcal{B}$ par exemple, et cette dernière tribu s'identifie, comme dans la Remarque 1.1.7, (3), à la tribu borélienne sur $[0, 1]^2$, donc contient le fermé D).

Les tranches $t_x(D)$ et $t^y(D)$ sont les singletons $\{(x, x)\}$ et $\{(y, y)\}$. Ils sont bien mesurables mais on a

$$\mu(t^y(D)) = 0 \text{ donc } \int_Y \mu(t^y(D))d\nu(y) = 0,$$

tandis que

$$\nu(t_x(D)) = |t_x(D)| = 1 \text{ donc } \int_X \nu(t_x(D))d\lambda(x) = 1.$$

Ainsi (4.2) n'est pas vérifiée dans ce cas.

REMARQUE 4.1.8. La mesure produit $\mu \otimes \nu$ est elle-même σ-finie puisque

$$X \times Y = \bigcup_{n,m} (X_n \times Y_m)$$

avec $(\mu \otimes \nu)(X_n \times Y_m) = \mu(X_n)\nu(Y_m) < +\infty$ si X_n (resp. Y_m) forment une décomposition de X (resp. Y) en ensembles de mesure finie.

En particulier cela permet de construire par récurrence des mesures portant sur plusieurs facteurs $\mu_1 \otimes \cdots \otimes \mu_n$, si μ_i est une mesure σ-finie sur X_i. La Proposition 4.1.5 montre que cette opération est associative

$$\mu_1 \otimes (\mu_2 \otimes \mu_3) = (\mu_1 \otimes \mu_2) \otimes \mu_3$$

puisque ces deux mesures coïncident sur les rectangles.

Si f est une fonction sur $X \times Y$, on note indifféremment

$$\int_{X \times Y} f d(\mu \otimes \nu) = \int_{X \times Y} f(x, y)d\mu(x)d\nu(y) = \int\int f d\mu d\nu$$

l'intégrale par rapport à la mesure produit.

4.2. Application aux variables aléatoires

Soit (Ω, Σ, P) un espace probabilisé : la mesure P est en particulier finie, donc on peut construire la mesure produit $P^{\otimes n} = P \otimes \cdots \otimes P$ sur Ω^n, qui est encore une mesure de probabilité.

Si X et Y sont des variables aléatoires sur Ω, le vecteur $Z = (X, Y)$ est une application mesurable $\Omega \to \mathbf{C}^2$, avec la tribu $\mathcal{B} \otimes \mathcal{B} = \mathcal{B}_{\mathbf{C}^2}$ sur $\mathbf{C}^2$. Soient $\mu = X(P)$ la loi de X, $\nu = Y(P)$ celle de Y. La mesure $Z(P)$ sur $\mathbf{C}^2$ vérifie

$$Z(P)(C) = P(Z^{-1}(C)) = P(X \in A \text{ et } Y \in B)$$

pour $C = A \times B$ un rectangle mesurable dans $\mathbf{C}^2$. La mesure produit $\mu \otimes \nu$ vérifie, par définition

$$(\mu \otimes \nu)(C) = \mu(A)\nu(B) = P(X \in A)P(Y \in B).$$

D'après la Proposition 4.1.5, on a par conséquent la caractérisation suivante de variables indépendantes :

Lemme 4.2.1. *Soit* (Ω, Σ, P) *un espace probabilisé.*

(1) *Des variables aléatoires* X *et* Y *sont indépendantes si et seulement si*

$$(X, Y)(P) = X(P) \otimes Y(P).$$

(2) *Plus généralement, une famille* $(X_i)_{i \in I}$ *de variables aléatoires est indépendante si et seulement si, pour tout* $n \geqslant 1$ *et tous indices distincts* $i_1, \ldots, i_n$*, on a*

$$(X_{i_1}, \ldots, X_{i_n})(P) = X_{i_1}(P) \otimes \cdots \otimes X_{i_n}(P).$$

Cette caractérisation est souvent plus maniable que la définition originale. Elle permet de démontrer facilement les propriétés telles que celles de la Proposition 1.2.11 et de l'Exercice 1.2.12. Par exemple, pour ce dernier cas, soient (X_i), $1 \leqslant i \leqslant 4$, des variables aléatoires indépendantes, $\varphi_1 : \mathbf{C}^2 \to \mathbf{C}$ et $\varphi_2 : \mathbf{C}^2 \to \mathbf{C}$ des applications mesurables, et posons $Y = \varphi_1(X_1, X_2)$, $Z = \varphi_2(X_3, X_4)$. On veut montrer que Y et Z sont indépendantes. Notons $\psi = (\varphi_1, \varphi_2) : \mathbf{C}^4 \to \mathbf{C}$, et $\mu_i = X_i(P)$ les lois des X_i. On a par (1.10) et indépendance

$$(Y, Z)(P) = \psi(X_1, X_2, X_3, X_4)(P) = \psi_*(\mu_1 \otimes \mu_2 \otimes \mu_3 \otimes \mu_4).$$

Mais on a le lemme utile suivant, énoncé très généralement.

Lemme 4.2.2. (1) *Soient* $(X_i, \mathcal{M}_i, \mu_i)$*, pour* $i = 1$*,* $i = 2$*, des espaces mesurés avec* $\mu_i(X_i) < +\infty$*,* $(Y_i, \mathcal{N}_i)$*, des espaces mesurables et*

$$\psi = (\psi_1, \psi_2) : X_1 \times X_2 \to Y_1 \times Y_2$$

une application mesurable pour les tribus produits. On a alors

$$\psi_*(\mu_1 \otimes \mu_2) = \psi_{1,*}(\mu_1) \otimes \psi_{2,*}(\mu_2)$$

sur $Y_1 \times Y_2$.

(2) *Soit* $(X, \mathcal{M}, \mu)$ *et* $(Y, \mathcal{N}, \nu)$ *des espaces probabilisés et* $p : X \times Y \to X$ *la projection. On a*

$$p_*(\mu \otimes \nu) = \mu.$$

Démonstration. (1) Pour tout $C = A \times B \in \mathcal{N}_1 \otimes \mathcal{N}_2$, on a selon les définitions mêmes

$$\begin{aligned} \psi_*(\mu_1 \otimes \mu_2)(C) &= (\mu_1 \otimes \mu_2)(\psi^{-1}(C)) \\ &= (\mu_1 \otimes \mu_2)(\psi_1^{-1}(A) \times \psi_2^{-1}(B)) \\ &= \mu_1(\psi_1^{-1}(A))\mu_2(\psi_2^{-1}(B)) \\ &= \psi_{1,*}(\mu_1)(A)\psi_{2,*}(\mu_2)(B) \\ &= (\psi_{1,*} \otimes \psi_{2,*})(A \times B) \end{aligned}$$

donc le résultat découle de la Proposition 4.1.5 puisque de plus

$$\psi_*(\mu_1 \otimes \mu_2)(Y_1 \times Y_2) = \mu_1(X_1)\mu_2(X_2) < +\infty.$$

(2) De même on a pour tout $C \in \mathcal{M}$

$$p_*(\mu \otimes \nu)(C) = (\mu \otimes \nu)(p^{-1}(C)) = (\mu \otimes \nu)(C \times Y) = \mu(C)\nu(Y) = \mu(C).$$

□

En utilisant le premier point de ce lemme il vient dans la situation ci-dessus

$$\begin{aligned} (Y, Z)(P) = \psi_*(\mu_1 \otimes \mu_2 \otimes \mu_3 \otimes \mu_4) &= \psi_*((\mu_1 \otimes \mu_2) \otimes (\mu_3 \otimes \mu_4)) \\ &= \varphi_{1,*}(\mu_1 \otimes \mu_2) \otimes \varphi_{2,*}(\mu_3 \otimes \mu_4) \\ &= \varphi_{1,*}(X_1, X_2)(P) \otimes \varphi_{2,*}(X_3, X_4)(P) \\ &= Y(P) \otimes Z(P), \end{aligned}$$

donc Y et Z sont indépendantes par le Lemme 4.2.1.

Les mesures produits permettent aussi de construire des variables aléatoires indépendantes dont les lois sont données.

Proposition 4.2.3. *Soient $n \geqslant 1$ un entier et μ_i, $1 \leqslant i \leqslant n$, des mesures de probabilité sur $(\mathbf{C}, \mathcal{B})$. Il existe un espace probabilisé (Ω, Σ, P) et des variables aléatoires $X_1, \ldots, X_n$ sur Ω telles que la loi de X_i est μ_i et les (X_i) sont indépendantes.*

Démonstration. Soit $(\Omega_i, \Sigma_i, P_i)$ un espace probabilisé quelconque avec une variable aléatoire Y_i telle que $Y_i(P_i) = \mu_i$ (par exemple, $(\Omega_i, \Sigma_i, P_i) = (\mathbf{C}, \mathcal{B}, \mu_i)$ avec $Y_i(z) = z$ convient).

On pose alors $\Omega = \Omega_1 \times \cdots \times \Omega_n$, muni de la tribu Σ produit des Σ_i et de la mesure de probabilité

$$P = P_1 \otimes \cdots \otimes P_n.$$

Soit alors $X_i = Y_i(p_i)$ où $p_i : \Omega \to \Omega_i$ est la i-ème projection. Les variables aléatoires (X_i) répondent à la question.

En effet, tout d'abord on a (Lemme 4.2.2, (2)) $X_i(P) = Y_i(P_i) = \mu_i$, et aussi (Lemme 4.2.2, (1))

$$(X_1, \ldots, X_n)(P) = (X_1, \ldots, X_n)(P_1 \otimes \cdots \otimes P_n) = X_1(P) \otimes \cdots \otimes X_n(P)$$

donc les (X_i) sont indépendantes. □

Cela s'applique en particulier si $\mu_i = \mu$ est fixée, et donne donc des vecteurs indépendants $(X_1, \ldots, X_n)$ arbitrairement longs de variables aléatoires de même loi. Bien entendu, comme la Loi des Grand Nombres du chapitre précédent et des problèmes similaires suggèrent, on voudrait aussi pouvoir fournir un tel énoncé pour une famille *infinie*. Cela est possible, mais demande un peu plus de travail, comme on le verra dans le Théorème 5.6.8.

Voilà cependant une application intéressante de ce premier résultat d'existence.

Théorème 4.2.4. *Soit $f : [0,1] \to \mathbf{R}$ une fonction continue, et pour $n \geqslant 1$ soit B_n le polynôme*

$$B_n = \sum_{k=0}^{n} \binom{n}{k} f\Big(\frac{k}{n}\Big) X^k (1-X)^{n-k} \in \mathbf{R}[X].$$

Alors (B_n) converge uniformément vers f sur $[0,1]$

Il s'agit donc d'une version complètement constructive du théorème d'approximation de Weierstrass dans le cas de l'intervalle $[0,1]$.

Démonstration. La preuve dépend d'une interprétation probabiliste de $B_n(x)$. Soit $x \in [0,1]$ fixé, et supposons données des variables aléatoires $X_1, \ldots, X_n$ à valeurs dans $\{0,1\}$, indépendantes et de même loi $X_i(P) = \mu_x$ telle que

$$\mu_x(\{0\}) = 1 - x \text{ et } \mu_x(\{1\}) = x.$$

Notons que $E(X_n) = E(X_1) = x$ et $V(X_n) = V(X_1) = x(1-x)$.

On a alors la formule

$$B_n(x) = E\Big(f\Big(\frac{X_1 + \cdots + X_n}{n}\Big)\Big). \tag{4.15}$$

En effet, la variable aléatoire $S_n = X_1 + \cdots + X_n$ prend ses valeurs dans $\{0, \ldots, n\}$, donc $f(S_n/n)$ prend ses valeurs dans l'ensemble des $f(k/n)$, $0 \leqslant k \leqslant n$, et

$$E\Big(f\Big(\frac{X_1 + \cdots + X_n}{n}\Big)\Big) = \sum_{k=0}^{n} f\Big(\frac{k}{n}\Big) P(S_n = k). \tag{4.16}$$

L'événement $S_n = k$ correspond à avoir k exactement des valeurs X_i, $1 \leqslant i \leqslant n$, égales à 1, et le reste à 0. Du fait de l'indépendance des (X_i) et de la loi des X_i, on a

$$\begin{aligned} P(S_n = k) &= \sum_{|I|=k} P(X_{i_1} = 1) \cdots P(X_{i_k} = 1) P(X_{j_1} = 0) \cdots P(X_{j_{n-k}} = 0) \\ &= \binom{n}{k} x^k (1-x)^{n-k} \end{aligned}$$

(où $I = \{i_1, \ldots, i_k\}$ parcourt les sous-ensembles de $\{1, \ldots, n\}$ de cardinalité k et $J = \{j_1, \ldots, j_{n-k}\}$ est le complémentaire de I). Ce calcul et (4.16) établissent la formule (4.15).

On a de plus $E(S_n/n) = E(X_1) = x$, et la loi des Grands Nombres rend naturel de penser que $E(f(S_n/n)) \to f(x)$. Mais il n'est pas nécessaire de faire appel aux résultats du chapitre précédent.

Écrivons

$$|B_n(x) - f(x)| = \Big| E\Big(f\Big(\frac{S_n}{n}\Big) - f\Big(E\Big(\frac{S_n}{n}\Big)\Big)\Big)\Big|.$$

L'idée est alors la suivante : si S_n/n est proche de son espérance, la contribution correspondante sera petite par continuité uniforme de f ; et quand au reste, c'est un événement de faible probabilité (à cause de la loi faible des grands nombres, si l'on veut).

Précisément, soit $\varepsilon > 0$ quelconque. Il existe, par continuité uniforme de f, un $\delta > 0$ tel que $|x - y| < \delta$ implique $|f(x) - f(y)| < \varepsilon$. Notons A l'événement

$$A = \{|S_n/n - E(S_n/n)| < \delta\}.$$

On a

$$\Big|\int_A \Big(f\Big(\frac{S_n}{n}\Big) - f(x)\Big) dP\Big| \leqslant \int_A \Big|f\Big(\frac{S_n}{n}\Big) - f(x)\Big| dP \leqslant \varepsilon P(A) \leqslant \varepsilon, \tag{4.17}$$

et sur le complémentaire $B = \Omega - A$ on a

$$\Big|\int_B \Big(f\Big(\frac{S_n}{n}\Big) - f(x)\Big) dP\Big| \leqslant 2\|f\|_\infty P(B), \tag{4.18}$$

et par l'inégalité de Chebychev (3.17) avec $p = 2$ il vient

$$P(B) = P(|S_n/n - E(S_n/n)| \geqslant \delta) \leqslant \delta^{-2} V(S_n/n) = n^{-1}\delta^{-2} V(X_1) \leqslant n^{-1}\delta^{-2}$$

(même calcul que (3.16) et $V(X_1) = x(1-x) \leqslant 1$). En conclusion on a

$$|B_n(x) - f(x)| \leqslant \varepsilon + 2\|f\|_\infty n^{-1}\delta^{-2}.$$

Pour tout n assez grand (ε, et donc δ, restant fixés), on a donc $|B_n(x) - f(x)| < 2\varepsilon$, inégalité valable pour tout x qui montre que B_n converge uniformément vers f comme annoncé. $\square$

Remarque 4.2.5. Cette preuve illustre un certain nombre de points intéressants : ainsi, il n'a jamais été nécessaire de spécifier l'ensemble Ω qui sert de support aux calculs probabilistes, et la construction particulière des variables aléatoires X_n est également sans importance. La loi même peut être variée du moment que $E(X_n) = x$ et $V(X_n)$ est bornée indépendamment de x , et alors on aura l'énoncé plus général

$$E(f(S_n/n)) \to f(x)$$

uniformément sur $[0, 1]$, mais en général, bien entendu, la suite de fonctions $g_n = E(f(S_n/n))$ (qui dépend de x via le choix de la loi des X_n) n'est pas une suite de polynôme.

Le point crucial est donc bien l'existence de la suite de variables indépendantes...

4.3. Le théorème de Fubini–Tonelli

La construction de la mesure produit à la Section 4.1 permet d'écrire la formule d'interversion (4.1) pour une fonction f qui est la fonction caractéristique de $C \in \mathcal{M} \otimes \mathcal{N}$. On en déduit la formule générale par linéarité, positivité et passage à la limite suivant un canevas maintenant habituel.

C'est le contenu du théorème de Fubini-Tonelli, que voici.

Théorème 4.3.1. *Soient $(X, \mathcal{M}, \mu)$ et $(Y, \mathcal{N}, \nu)$ des espaces mesurés σ-finis. On munit $X \times Y$ de la tribu produit.*

(1) *Si $f : X \times Y \to [0, +\infty]$ est mesurable, alors les fonctions positives*

$$x \mapsto \int_Y f(x,y)d\nu(y)$$
$$y \mapsto \int_X f(x,y)d\mu(x)$$

sont mesurables pour $\mathcal{M}$ et $\mathcal{N}$ respectivement, et on a

$$\int_{X\times Y} fd(\mu \otimes \nu) = \int_X \Big(\int_Y f(x,y)d\nu(y)\Big)d\mu(x) = \int_Y \Big(\int_X f(x,y)d\mu(x)\Big)d\nu(y). \tag{4.19}$$

(2) *Si $f : X \times Y \to \mathbf{C}$ vérifie $f \in L^1(\mu \otimes \nu)$, alors pour presque tout $y \in Y$, la fonction $t^y(f) : x \mapsto f(x,y)$ est μ-intégrable, pour presque tout $x \in X$, la fonction $t_x(f) : y \mapsto f(x,y)$ est ν-intégrable, les fonctions*

$$x \mapsto \int_Y f(x,y)d\nu(y)$$
$$y \mapsto \int_X f(x,y)d\mu(x)$$

sont respectivement ν-intégrables et μ-intégrables, et

$$\int_{X\times Y} fd(\mu \otimes \nu) = \int_X \Big(\int_Y f(x,y)d\nu(y)\Big)d\mu(x) = \int_Y \Big(\int_X f(x,y)d\mu(x)\Big)d\nu(y). \tag{4.20}$$

(3) *Si $f : X \times Y \to \mathbf{C}$ est telle que*

$$x \mapsto \int_Y |f(x,y)|d\nu(y)$$

est μ-intégrable, alors $f \in L^1(\mu \otimes \nu)$.

Les opérations $f \mapsto t_x(f)$ (assoçiant une fonction de la variable y à une fonction de deux variables) vérifient les propriétés formelles évidentes suivantes, où $x \in X$ est quelconque :

$$t_x(\alpha f + \beta g) = \alpha t_x(f) + \beta t_x(g)$$
$$t_x(f)^{\pm} = t_x(f^{\pm})$$
$$\text{Si } f \leqslant g \text{ alors } t_x(f) \leqslant t_x(g)$$
$$\text{Si } f_n \to f \text{ en tout point, alors } t_x(f_n) \to t_x(f)$$

(et similairement pour $f \mapsto t^y(f)$).

Démonstration. (1) Comme on l'a déjà remarqué, l'énoncé est valide par définition de la mesure produit si $f = \chi_C$ est la fonction caractéristique de $C \in \mathcal{M} \otimes \mathcal{N}$, puisque alors $t_x(f) = \nu(t_x(C))$ et $t^y(f) = \mu(t^y(C))$. Par additivité pour l'intégrale et les opérations tranche, il reste valide pour toute fonction étagée positive s.

Soit $f \geqslant 0$ quelconque et soit alors (s_n) une suite croissante de fonctions étagées positives telles que $s_n \leqslant f$ et $s_n(x,y) \to f(x,y)$ pour tout $(x,y) \in X \times Y$. D'après les propriétés ci-dessus, pour tout $x \in X$ fixé les suites $t_x(s_n)$ convergent en croissant vers $t_x(f)$. Cela montre que $t_x(f) : Y \to [0,+\infty]$ est mesurable (ce qui est aussi le Corollaire 4.1.3). D'après le théorème de convergence monotone on a d'une part

$$\int_Y t_x(f)d\nu = \lim_{n\to+\infty} \int_Y t_x(s_n)d\nu$$

pour tout x et d'autre part cette limite est également croissante, donc une seconde application du théorème de convergence monotone donne cette fois

$$\begin{aligned}\int_X \int_Y t_x(f)d\nu(y)d\mu(x) &= \lim_{n\to+\infty} \int_X \int_Y t_x(s_n)d\nu(y)d\mu(x)\\ &= \lim_{n\to+\infty} \int_{X\times Y} s_n d(\mu\otimes\nu) \text{ (cas des fonctions étagées)}\\ &= \int_{X\times Y} f d(\mu\otimes\nu),\end{aligned}$$

la dernière étape étant une troisième application du théorème de convergence monotone à la suite $s_n \to f$. Cela montre la première formule de (4.19), et la seconde est vraie en échangeant X et Y.

(2) Soit $f \in L^1(\mu\otimes\nu)$. Par le Corollaire 4.1.3, les fonctions $t_x(f)$ sont mesurables pour tout x. Comme de plus on a d'après (4.19)

$$\int_{X\times Y} |f| d(\mu\otimes\nu) = \int_X \Big(\int_Y |f(x,y)|d\nu(y)\Big)d\mu(x) < +\infty,$$

il s'ensuit que la fonction de x intégrée dans le membre de droite est presque partout finie, c'est à dire que $t_x(f) \in L^1(\nu)$ pour presque tout $x \in X$. En un tel x on a

$$\int_Y f(x,y)d\nu(y) = \int_Y t_x(f^+)d\nu(y) - \int_Y t_x(f^-)d\nu(y), \tag{4.21}$$

et d'après (1) toujours cela montre que $x \mapsto \int f(x,y)d\nu(y)$ (définie presque partout) est mesurable, et comme

$$\int_X \Big|\int_Y f(x,y)d\nu(y)\Big| d\mu(x) \leqslant \int_X \int_Y |f(x,y)|d\nu(y)d\mu(x) < +\infty,$$

cette fonction est même μ-intégrable. Intégrant alors (4.21) sur X, et faisant appel à (1) encore, il vient

$$\begin{aligned}\int_X \int_Y f(x,y)d\nu(y)d\mu(x) &= \int_X \int_Y t_x(f^+)d\nu(y)d\mu(x) - \int_X \int_Y t_x(f^-)d\nu(y)d\mu(x)\\ &= \int_{X\times Y} f^+ d(\mu\otimes\nu) - \int_{X\times Y} f^- d(\mu\otimes\nu)\\ &= \int_{X\times Y} f d(\mu\otimes\nu),\end{aligned}$$

ce qui est (4.20). Comme ci-dessus la seconde formule provient de l'échange de X et Y.

(3) D'après (1), la condition indiquée implique que

$$\int_{X\times Y} |f| d(\mu\otimes\nu) = \int_X \Big(\int_Y |f(x,y)|d\nu(y)\Big)d\mu(x) < +\infty$$

donc que $f \in L^1(\mu\otimes\nu)$. $\square$

Remarque 4.3.2. Le point (3) du théorème doit être vu comme un critère pour pouvoir appliquer le point (2) ; il n'est en effet pas toujours évident de vérifier directement que $f \in L^1(\mu \otimes \nu)$. On montre en pratique une estimation

$$\int_Y |f(x,y)| d\nu(y) \leqslant g(x)$$

avec $g \in L^1(\mu)$.

Exemple 4.3.3. Une des utilisations importantes du théorème ci-dessus est dans l'étude des fonctions définies par des intégrales comme dans la Section 3.3 : soit $h : X \times Y \to \mathbf{C}$ et

$$f(x) = \int_Y h(x,y) d\nu(y).$$

Si l'on peut intervertir les intégrales, il est alors possible dans de nombreux cas d'évaluer

$$\int_X f(x) d\mu(x) = \int_Y \int_X h(x,y) d\mu(x) d\nu(y).$$

Par exemple, considérons la fonction[2]

$$J_0(x) = \frac{1}{\pi} \int_0^\pi \cos(x \sin\theta) d\theta$$

pour $x \in \mathbf{R}$. La fonction est bien définie et continue d'après la Proposition 3.3.1, en particulier mesurable. Comme $|J_0(x)| \leqslant 1$, la fonction $x \mapsto e^{-x} J_0(x)$ sur $[0, +\infty[$ est intégrable.

On a, la justification de l'application du Théorème de Fubini étant immédiate

$$\begin{aligned}\int_0^{+\infty} e^{-x} J_0(x) dx &= \frac{1}{\pi} \int_0^{+\infty} \int_0^\pi e^{-x} \cos(x \sin\theta) d\theta dx \\ &= \frac{1}{\pi} \int_0^\pi \int_0^\infty e^{-x} \cos(x \sin\theta) dx d\theta.\end{aligned}$$

L'intégrale intérieure est élémentaire

$$\begin{aligned}\int_0^\infty e^{-x} \cos(x \sin\theta) dx &= \mathrm{Re}\Big(\int_0^\infty e^{-x(1-i\sin\theta)} dx\Big) \\ &= \mathrm{Re}\Big(\Big[-\frac{1}{1-i\sin\theta} e^{-x(1-i\sin\theta)}\Big]_0^{+\infty}\Big) \\ &= \mathrm{Re}\Big(\frac{1}{1-i\sin\theta}\Big) = \frac{1}{1+\sin^2\theta},\end{aligned}$$

donc il vient

$$\int_0^{+\infty} e^{-x} J_0(x) dx = \frac{1}{\pi} \int_0^\pi \frac{d\theta}{1+\sin^2\theta} = \frac{1}{\sqrt{2}}$$

(faire le changement de variable $u = \tan\theta$ sur $[0, \pi/2[$, l'autre contribution étant égale.)

Exemple 4.3.4. Le théorème de Fubini, associé au Lemme 4.2.1 permet de retrouver facilement les propriétés des variables aléatoires indépendantes déjà utilisées dans le chapitre précédent.

Par exemple, redémontrons le résultat de l'Exercice 2.3.6.

Proposition 4.3.5. *Soit (Ω, Σ, P) un espace probabilisé, X et Y des variables aléatoires intégrables indépendantes. On a alors $XY \in L^1(P)$ et*

$$E(XY) = E(X)E(Y)$$

[2]. Il s'agit de l'exemple le plus simple de fonction de Bessel.

Démonstration. D'après le Lemme 4.2.2, $|X|$ et $|Y|$ sont des variables aléatoires positives indépendantes. Vérifions d'abord la propriété dans le cas où $X \geqslant 0$, $Y \geqslant 0$ ce cas, ce qui permettra de s'assurer que $XY \in L^1(P)$.

Soit μ la loi de X, ν celle de Y, de sorte que $(X,Y)(P) = \mu \otimes \nu$ par le Lemme 4.2.1. Soit $m : \mathbf{C}^2 \to \mathbf{C}$ la multiplication. Par la Proposition 2.3.3, (3), on a

$$\begin{aligned} E(XY) &= \int_\Omega XY dP = \int_\Omega m(X,Y)dP \\ &= \int_{\mathbf{C}^2} md(X,Y)(P) \\ &= \int_{\mathbf{C}^2} xyd(\mu\otimes\nu)(x,y). \qquad (4.22) \end{aligned}$$

Or m est mesurable (car continue) donc par le Théorème de Fubini, (1), on a

$$\int_{\mathbf{C}^2} md(\mu\otimes\nu) = \int_{\mathbf{C}}\int_{\mathbf{C}} xyd\nu(y)d\mu(x) = \Bigl(\int_{\mathbf{C}} xd\mu(x)\Bigr)\Bigl(\int_{\mathbf{C}} yd\nu(y)\Bigr) = E(X)E(Y).$$

Revenant au cas général, ce calcul établit que $XY \in L^1(P)$, et le calcul de $E(XY)$ et celui de l'intégrale (4.22) sont alors valides pour X et Y intégrables par le Théorème de Fubini, (2), ce qui donne le résultat. □

4.4. L'intégrale de Lebesgue sur $\mathbf{R}^d$

Nous considérons maintenant l'exemple particulièrement important de $\mathbf{R}^d$. Si $d \geqslant 1$, on note λ_d (ou simplement λ si la valeur de la dimension d est claire dans le contexte) la mesure

$$\lambda_d = \lambda \otimes \cdots \otimes \lambda$$

produit de d copies de la mesure de Lebesgue sur $\mathbf{R}$. C'est donc une mesure σ-finie sur la tribu produit des tribus boréliennes, qui s'identifie à $\mathcal{B}_{\mathbf{R}^d}$, et qui vérifie par définition

$$\lambda(C) = (b_1 - a_1)\cdots(b_d - a_d) \text{ si } C = [a_1, b_1] \times \cdots \times [a_d, b_d]$$

est un *cube*. Elle est appelée la mesure de Lebesgue en dimension d.

On écrira aussi souvent simplement

$$\int_{\mathbf{R}^d} f(x_1, \ldots, x_d)dx_1 \cdots dx_d = \int_{\mathbf{R}^d} f(x)dx$$

l'intégrale d'une fonction réelle f de d variables $x = (x_1, \ldots, x_d)$ par rapport à cette mesure.

Avec la mesure de Lebesgue viennent les espaces $L^p(d\lambda) = L^p(\mathbf{R}^d)$, $1 \leqslant p \leqslant +\infty$. Il est utile d'avoir des critères simples d'appartenance à L^p par comparaison : le lemme suivant est l'analogue de ceux portant sur la décroissance « polynômiale » nécessaire pour qu'une fonction continue définie sur $\mathbf{R}$ soit intégrable « à l'infini » ou « au voisinage de 0 ».

On note $\|t\|$ la norme euclidienne

$$\|t\| = \Bigl(\sum_{1\leqslant i\leqslant d} |t_i|^2\Bigr)^{1/2} \qquad (4.23)$$

pour $t \in \mathbf{R}^d$.

Proposition 4.4.1. *Soit f une fonction mesurable sur $\mathbf{R}^d$ et $p \in [1, +\infty[$.*

(1) *Si f est bornée et s'il existe une constante $C \geqslant 0$ et $\varepsilon > 0$ tels que*

$$|f(x)| \leqslant C(1 + \|x\|)^{-d-\varepsilon}$$

pour $x \in \mathbf{R}^d$, alors $f \in L^1(\mathbf{R}^d)$.

(2) *Si f est bornée et s'il existe une constante $C \geqslant 0$ et $\varepsilon > 0$ tels que*

$$|f(x)| \leqslant C(1 + \|x\|)^{-d/p-\varepsilon}$$

pour $x \in \mathbf{R}^d$, alors $f \in L^p(\mathbf{R}^d)$.

(3) *Si f est intégrable sur $\{x \mid \|x\| > 1\}$ et s'il existe une constante $C \geqslant 0$ et $\varepsilon > 0$ tels que*

$$|f(x)| \leqslant C\|x\|^{-d+\varepsilon}$$

pour $\|x\| \leqslant 1$, alors $f \in L^1(\mathbf{R}^d)$.

(4) *Si $|f|^p$ est intégrable sur $\{x \mid \|x\| > 1\}$ et s'il existe une constante $C \geqslant 0$ et $\varepsilon > 0$ tels que*

$$|f(x)| \leqslant C\|x\|^{-d/p+\varepsilon}$$

pour $\|x\| \leqslant 1$, alors $f \in L^p(\mathbf{R}^d)$.

Démonstration. Clairement (2) et (4) découlent de (1) ou (3) appliqué à f^p. On va montrer (1), qui concerne l'intégrabilité « à l'infini », par récurrence sur d. Le cas $d = 1$ provient aussitôt du calcul classique des primitives de fonctions $|t|^a$ pour $a \in \mathbf{R}$.

Soit $D = \{x \mid \|x\| \geqslant 1\}$ le complémentaire de la boule unité de $\mathbf{R}^d$. Pour montrer (1) pour f donnée, il suffit par comparaison de montrer que pour tout $a > d$, la fonction positive

$$g(x) = \chi_D(x)\|x\|^{-a}$$

est intégrable. Par le Théorème 4.3.1, on peut écrire

$$\int_{\mathbf{R}^d} g(x)dx = \int_{\mathbf{R}^{d-1}} h(y)dy \tag{4.24}$$

avec

$$h(y) = \int_{\mathbf{R}} \frac{\chi_D(t,y)dt}{(t^2 + \|y\|^2)^{a/2}}$$

Pour $\|y\|^2 \geqslant 1/2$, on écrit

$$\begin{aligned} h(y) &= 2\int_{\sqrt{1-\|y\|^2}}^{+\infty} \frac{dt}{(t^2 + \|y\|^2)^{a/2}} \\ &\leqslant 2\|y\|^{1-a}\int_0^{+\infty} \frac{du}{(1+u^2)^{a/2}} \leqslant C\|y\|^{1-a} \end{aligned}$$

pour une certaine constante $C \geqslant 0$ dépendant de a seulement, puisque $a > d \geqslant 1$ donc l'intégrale en u est « classiquement » convergente.

Pour $\|y\|^2 \leqslant 1/2$, par contre, la condition $(t,y) \in D$ implique $t^2 + \|y\|^2 \geqslant 1$ donc $t \geqslant \sqrt{1-\|y\|^2} \geqslant 1/\sqrt{2}$ et alors

$$h(y) \leqslant 2\int_{1/\sqrt{2}}^{+\infty} \frac{dt}{t^a} \leqslant D$$

où D ne dépend que de a. Ces deux estimations se recoupent pour montrer qu'il existe $C_1 \geqslant 0$ telle que

$$h(y) \leqslant C_1(1 + \|y\|)^{-(a-1)}.$$

Comme $a-1 > d-1$, cela et (4.24) montrent par récurrence que g est intégrable, fournissant ainsi (1). Enfin, le point (3) se démontre exactement de même en utilisant le cas de la dimension 1. □

Remarque 4.4.2. Ce résultat peut aussi bien se démontrer par passage en coordonnées sphériques généralisées, en utilisant la formule de changement de variable discutée ci-dessous ; en dimension 2, il s'agit des classiques coordonnées polaires.

Anticipant un peu sur le thème du chapitre suivant, l'importance de la mesure de Lebesgue provient en partie de son interaction avec les applications différentiables, sous la forme de la formule de changement de variables.

Rappelons pour l'énoncer quelques définitions de calcul différentiel.

DÉFINITION 4.4.3. Soient U, $V \subset \mathbf{R}^d$ des ouverts non-vides, $f : U \to V$ une application.

(1) f est différentiable en $x \in U$ s'il existe une application linéaire $T : \mathbf{R}^d \to \mathbf{R}^d$ telle que

$$f(x+h) = f(x) + T(h) + o(\|h\|)$$

pour $h \in \mathbf{R}^d$ tel que $x + h \in U$. On note $T = D_x(f)$, la *différentielle* de f en x.

(2) f est différentiable sur U si elle est différentiable en tout point, et de classe C^1 si l'application

$$\begin{cases} U \to L(\mathbf{R}^d) \simeq \mathbf{R}^{n^2} \\ x \mapsto D_x(f) \end{cases}$$

est continue sur U.

(3) f est un difféomorphisme (resp. un difféomorphisme de classe C^1) si f est différentiable (resp. de classe C^1) sur U, bijective et si son inverse $g = f^{-1} : V \to U$ est également différentiable (resp. de classe C^1).

(4) Si f est différentiable sur U, le jacobien de f est l'application

$$J_f \begin{cases} U \to \mathbf{R} \\ x \mapsto \det(D_x(f)) \end{cases}$$

(qui est continue sur U si f est C^1).

Autrement dit, la différentielle est la meilleure approximation de f qui soit linéaire, le type d'application le plus simple qui soit.

Le *théorème des fonctions implicites* montre le critère suivant :

PROPOSITION 4.4.4. *Soit U, $V \subset \mathbf{R}^d$ des ouverts non-vides. Une application $f : U \to V$ est un C^1-difféomorphisme si et seulement si f est bijective, de classe C^1, et $D_x(f) \in GL(\mathbf{R}^d)$ pour tout $x \in U$. On a alors*

$$D_y(f^{-1}) = D_{f^{-1}(y)}(f)^{-1} = D_x(f)^{-1}$$

si $y = f(x)$ et

$$J_g(y) = J_f(x)^{-1}. \tag{4.25}$$

REMARQUE 4.4.5. Concrètement, si f est donnée par des « formules », c'est à dire

$$f(x) = (f_1(x_1, \ldots, x_d), \ldots, f_d(x_1, \ldots, x_d))$$

où $f_j : \mathbf{R}^d \to \mathbf{R}$, alors f est C^1 si et seulement si les dérivées partielles des f_j existent, sont des fonctions continues sur U, et $D_x(f)$ est l'application linéaire correspondant à la matrice (dite jacobienne[3]) des dérivées partielles

$$D_x(f) = \Big(\frac{\partial f_i}{\partial x_j}\Big)_{i,j},$$

le jacobien en étant le déterminant.

Si $d = 1$, la différentielle (en x) se réduit à une application linéaire en une variable, c'est à dire $y \mapsto ay$ avec $a \in \mathbf{R}$, et le nombre a (qui « est » la matrice dans ce cas) est évidemment la dérivée $f'(x)$ de f en x.

On dira souvent que f est un « changement de variables » $y_i = f_i(x_1, \ldots x_d)$, l'inverse de f donnant la formule pour revenir aux anciennes variables.

EXEMPLE 4.4.6. (1) Les changements de variable les plus simples sont les applications linéaires elles-mêmes : si $T : \mathbf{R}^d \to \mathbf{R}^d$ est linéaire, alors elle est différentiable sur $\mathbf{R}^d$ avec $D_x(f) = T$ en tout $x \in \mathbf{R}^d$, et c'est un difféomorphsime si et seulement si T est inversible.

(2) Les translations $\tau : x \mapsto x + a$ avec $a \in \mathbf{R}^d$ fixé sont également des difféomorphismes, avec $D_x(\tau) = \mathrm{Id}$ pour tout x.

3. À ne pas confondre avec les jacobins qui guillotinèrent les girondins, ni avec le capitaine Jacobi qui ramena d'Orient le Faucon Maltais.

(3) Les « coordonnées polaires » sont un changement de variables important. Soit $d = 2$, $V = \mathbf{R}^2 - \{0\}$. Chaque $x = (a, b) \in V$ peut être décrit de façon unique par sa distance à l'origine $r = \sqrt{a^2 + b^2} > 0$ et l'angle $\theta \in]-\pi, \pi]$ entre la demi-droite des abscisses et celle passant par x. L'inverse de ce changement de variable est

$$\varphi \begin{cases} U \to V \\ (r, \theta) \mapsto (r\cos\theta, r\sin\theta), \end{cases}$$

où, donc, $U =]0, +\infty[\times]-\pi, \pi] \subset \mathbf{R}^2$. Noter que U n'est pas ouvert, mais par restriction φ fournit une bijection $\varphi : U_1 \to V_1$ avec $U_1 =]0, +\infty[\times]-\pi, \pi[$, $V_1 = \{(x, y) \mid y \neq 0 \text{ ou } x > 0\}$. On voit que φ est différentiable ; la matrice jacobienne est

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

dont le déterminant vaut $J_\varphi(r, \theta) = r > 0$, ce qui montre que φ ainsi restreinte est un C^1-difféomorphisme.

Du point de vue de l'intégration, la « partie manquante » (l'axe des x négatif, les $(x, 0)$ avec $x < 0$) est de toute manière négligeable, puisque de mesure nulle.

Nous énonçons plusieurs formes équivalentes de la formule de changement de variables : chacune apparaît naturellement dans de nombreuses situations.

Théorème 4.4.7. *Soit $d \geqslant 1$ et $\varphi : U \to V$ un C^1-difféomorphisme entre deux ouverts de $\mathbf{R}^d$.*

(1) *Si f est une fonction positive ou intégrable sur V, alors on a*

$$\int_V f(x)d\lambda(x) = \int_U f(\varphi(y))|J_\varphi(y)|d\lambda(y),$$

c'est à dire, dans le cas où f n'est pas $\geqslant 0$, que si l'une des deux fonctions est intégrable, alors l'autre l'est, et la formule est vérifiée.

(2) *Si f est une fonction positive ou intégrable sur V, alors on a*

$$\int_U f(\varphi(x))d\lambda(x) = \int_V f(y)|J_\varphi(\varphi^{-1}(y))|^{-1}d\lambda(y) = \int_V f(y)|J_{\varphi^{-1}}(y)|d\lambda(y).$$

(3) *La mesure image de la mesure de Lebesgue par φ est*

$$\varphi_*(d\lambda(x)) = |J_{\varphi^{-1}}(y)|d\lambda(y). \tag{4.26}$$

La démonstration sera donnée dans le chapitre suivant : elle se comprend mieux en effet dans le cadre général des mesures boréliennes. Nous expliquons juste ici l'équivalence de ces formulations. Pour passer de (1) à la première forme de (2), il suffit d'appliquer (1) à

$$g(x) = f(x)|J_\varphi(\varphi^{-1}(x))|^{-1}$$

au lieu de f. La formule (4.25) montre que

$$g(x) = f(x)|J_{\varphi^{-1}}(x)|,$$

ce qui fournit la dernière formule de (2), à moins que l'on ne préfère appliquer (1) directement à l'inverse φ^{-1} de φ.

Cette symétrie dit aussi que (2) implique (1). Quand à (3), elle est équivalente à (2) puisque

$$\int_V f(y)\varphi_*(d\lambda)(y) = \int_U f(\varphi(x))d\lambda(x)$$

d'après (2.7), (2.10) pour toute fonction intégrable.

Remarque 4.4.8. (1) La valeur absolue est présente dans les formules parce que l'intégrale de Lebesgue ignore les questions d'orientation. Prenons $\varphi(x) = -x$ pour $x \in \mathbf{R}$, avec différentielle -1 et jacobien $|J(x)| = 1$, et comparer la formule

$$\int_{[a,b]} f(x) d\lambda(x) = \int_{[-b,-a]} f(-y) d\lambda(y)$$

pour $a < b$ avec celle traditionnelle de la formule de changement de variable habituelle ($y = -x$, $dy = -dx$) pour l'intégrale de Riemann sur un intervalle compact

$$\int_a^b f(x) dx = -\int_{-a}^{-b} f(-y) dy = \int_{-b}^{-a} f(-y) dy.$$

Noter cependant que le jacobien J_φ ne s'annule pas sur U puisque φ est un difféomorphisme, et donc si l'ouvert U est *connexe*, le signe de $J_\varphi(x)$ sera le même pour tout $x \in U$, de sorte que $|J_\varphi(x)| = \varepsilon J_{\varphi(x)}$ avec $\varepsilon = \pm 1$ indépendant de x.

(2) L'interprétation géométrique du facteur jacobien est très simple et intuitive. En effet, pour T une application linéaire bijective, il est bien connu que la valeur absolue $|\det(T)|$ est le volume (c'est à dire la mesure d-dimensionnelle) de l'image par T du cube unité (c'est particulièrement évident si T est diagonalisable) : cela correspond d'ailleurs tautologiquement au cas $U = [0,1]^d$, $V = T(U)$ et $f = 1$ de la formule de changement de variable

$$\mathrm{Vol}(T(U)) = \int_{T(U)} f(x) d\lambda(x) = \int_U |\det(T)| d\lambda(y) = |\det(T)|.$$

Le facteur jacobien en x est alors, heuristiquement, le « coefficient de dilatation » infinitésimal en x, et il est naturel que la mesure dx se transforme selon (4.26), en y réfléchissant quelques instants et en faisant quelques dessins...

Exemple 4.4.9. (1) Soit T une application linéaire bijective. Il vient

$$\int_{\mathbf{R}^d} f(T(x)) d\lambda(x) = \frac{1}{\det(T)} \int_{\mathbf{R}^d} f(x) d\lambda(x).$$

(2) Soit $f : \mathbf{R}^2 \to \mathbf{C}$ une fonction intégrable. On veut l'intégrer en coordonnées polaires (Exemple 4.4.6, (3)). Comme $Z = \{(a,0) \mid a \leqslant 0\} \subset \mathbf{R}^2$ est de mesure nulle, on a

$$\int_{\mathbf{R}^2} f(x) d\lambda_2(x) = \int_{\mathbf{R}^2 - Z} f(x) d\lambda_2(x)$$

et d'après l'Exemple 4.4.6, (3) il vient, conservant les mêmes notations :

$$\begin{aligned}\int_{\mathbf{R}^2 - Z} f(x) d\lambda_2(x) &= \int_{V_1} f(r\cos\theta, r\sin\theta) r dr d\theta \\ &= \int_{-\pi}^{\pi} \int_0^{+\infty} f(r\cos\theta, r\sin\theta) r dr d\theta.\end{aligned}$$

Le théorème de Fubini s'applique pour justifier d'écrire ainsi l'intégrale, ou avec l'ordre de r et θ échangé.

Supposons de plus que f est *radiale*, c'est à dire est fonction de la distance à l'origine r seulement : il existe $g : [0, +\infty[\to \mathbf{C}$ telle que $f(x,y) = g(\sqrt{x^2 + y^2}) = g(r)$. Dans ce cas, on a

$$\int_{\mathbf{R}^2} f(x) d\lambda_2(x) = \int_{-\pi}^{\pi} \int_0^{+\infty} g(r) r dr d\theta = 2\pi \int_0^{+\infty} g(r) r dr,$$

en effectuant l'intégration sur θ d'abord.

Voici une application importante.

Proposition 4.4.10. *Soit μ la mesure sur $\mathbf{R}$ donnée par*

$$\mu = e^{-\pi x^2} dx.$$

Alors μ est une mesure de probabilité. De plus, si X est une variable aléatoire dont la loi est μ, on a

$$E(X) = 0, \text{ et } V(X) = \frac{1}{2\pi}.$$

Démonstration. Soit $f : \mathbf{R}^2 \to [0, +\infty[$ la fonction définie par

$$f(x, y) = e^{-\pi(x^2+y^2)}.$$

Il s'agit d'une fonction positive, donc son intégrale existe dans $[0, +\infty]$ sans autre justification, et c'est une fonction radiale avec $g(r) = e^{-\pi r^2}$. Il vient

$$\int_{\mathbf{R}^2} e^{-\pi(x^2+y^2)} d\lambda_2(x) = 2\pi \int_0^{+\infty} e^{-\pi r^2} r dr = \left[-e^{-r^2} \right]_0^{+\infty} = 1.$$

Mais on a aussi, par application du théorème de Fubini

$$\int_{\mathbf{R}^2} e^{-\pi(x^2+y^2)} d\lambda_2(x) = \int_{\mathbf{R}} e^{-\pi x^2} \Big(\int_{\mathbf{R}} e^{-\pi y^2} d\lambda(y) \Big) d\lambda(x) = I^2$$

où

$$I = \int_{\mathbf{R}} e^{-\pi x^2} dx = \int_{\mathbf{R}} d\mu(x).$$

En comparant, on trouve bien $\int_{\mathbf{R}} d\mu(x) = 1$, donc μ est une mesure de probabilité.

Il reste à calculer son espérance et sa variance. Or

$$E(X) = \int_{\mathbf{R}} x d\mu(x) = \int_{\mathbf{R}} x e^{-\pi x^2} dx = 0$$

car la fonction $x \mapsto xe^{-x^2}$ est intégrable et impaire, et

$$V(x) = \int_{\mathbf{R}} x^2 d\mu(x) = \int_{\mathbf{R}} x^2 e^{-\pi x^2} dx = \frac{1}{2\pi} \int_{\mathbf{R}} e^{-x^2} dx = \frac{1}{2\pi}$$

par intégration par partie. □

Plus généralement, la mesure

$$\mu_{a,\sigma} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-a)^2}{2\sigma}} d\lambda(x)$$

pour $a \in \mathbf{R}$, $\sigma > 0$, est une mesure de probabilité d'espérance a et de variance σ (cf. Remarque 3.2.11). Cela se déduit de la proposition par des changements de variables simples (translation, dilatation) laissés en exercice ; voir aussi le Chapitre 9.

CHAPITRE 5

Intégration et fonctions continues

5.1. Introduction

Lorsque X est un espace topologique, on dispose de la tribu borélienne sur X, mais aussi d'une classe de fonctions privilégiée, celle des fonctions continues $X \to \mathbf{C}$. Il est donc naturel de se poser la question des interactions entre l'intégration par rapport à une mesure borélienne et les fonctions continues (qui sont mesurables d'après le Corollaire 1.1.9).

Pour pouvoir en dire quelque chose d'intéressant, il faut cependant des hypothèses sur X et sur les mesures considérées. En effet, en toute généralité il se peut d'une part que l'espace vectoriel $C(X)$ des fonctions continues sur X soit réduit aux fonctions constantes, et d'autre part que, même si $C(X)$ est de taille raisonnable, les fonctions continues ne soient presque jamais intégrables.

Par exemple, si $X = \mathbf{R}$, et si ν est la mesure de comptage, qui est en particulie définie sur la tribu borélienne $\mathcal{B}_{\mathbf{R}}$, on a

$$\int_{\mathbf{R}} f d\nu(x) = +\infty$$

pour toute fonction continue $f \geqslant 0$ non identiquement nulle : il existe par continuité une constante $c > 0$ et un intervalle ouvert $I \neq \emptyset$ tel que $f(x) \geqslant c > 0$ pour $x \in I$, donc

$$\int_{\mathbf{R}} f d\nu(x) \geqslant c\nu(I) = +\infty.$$

La proposition élémentaire suivante indique un cas, très général et couramment rencontré, dans lequel les fonctions continues *à support compact* sont intégrables. Rappelons que si $f : X \to \mathbf{C}$ est une fonction continue sur X, le support de f est l'ensemble fermé dans X défini par

$$\operatorname{supp}(f) = \bar{V} \text{ où } V = \{x \mid f(x) \neq 0\}.$$

On a donc la propriété caractéristique

$$f(x) = 0 \text{ si } x \notin \operatorname{supp}(f),$$

(mais il est tout à fait possible que $f(x) = 0$ sans que x n'appartienne au support de f, comme un dessin sur $\mathbf{R}$ convaincra aussitôt le lecteur).

Définition 5.1.1. Soit X un espace topologique. On note $C_c(X) \subset C(X)$ le $\mathbf{C}$-espace vectoriel des fonctions continues à support compact.

Si X est compact, on a donc $C_c(X) = C(X)$, mais par exemple les fonctions constantes non-nulles n'appartiennent pas à $C_c(\mathbf{R})$.

Il est clair que $C_c(X)$ est en effet un espace vectoriel : on a par exemple

$$\operatorname{supp}(\alpha f + \beta g) \subset \operatorname{supp}(f) \cup \operatorname{supp}(g).$$

Une propriété importante d'une fonction $f \in C_c(X)$ est d'être bornée : on note $\|f\|_\infty = \sup\{|f(x)| \mid x \in X\}$ pour $f \in C_c(X)$, et cela donne une norme sur $C_c(X)$. Le fait que $\|f\|_\infty < +\infty$ provient du fait que

$$\sup\{|f(x)| \mid x \in X\} = \sup\{|f(x)| \mid x \in \operatorname{supp}(f)\}$$

par définition, et on sait qu'une fonction continue sur un compact est bornée (et atteint sa borne supérieure, ce qui sera donc aussi vrai ici).

Proposition 5.1.2. *Soit X un espace topologique, μ une mesure borélienne* finie sur les compacts*, c'est à dire que $\mu(K) < +\infty$ pour tout $K \subset X$ compact. Alors l'application*

$$\begin{cases} C_c(X) \to \mathbf{C} \\ f \mapsto \int_X f d\mu(x) \end{cases}$$

est une forme linéaire positive *sur $C_c(X)$, c'est à dire telle que*

$$f \geqslant 0 \text{ implique } \Lambda(f) \geqslant 0.$$

Démonstration. Tout cela est évident, dès lors que Λ est bien définie, c'est à dire que les fonctions continues à support compact sont intégrables. Pour montrer cela, remarquons que si f est continue alors $f^{\pm}$ le sont également, et si f est à support compact, $f^{\pm}$ le sont également. Il suffit donc de montrer que $\int f d\mu < +\infty$ pour $f \geqslant 0$ dans $C_c(X)$.

Soit donc $f \geqslant 0$ continue à support compact $K = \mathrm{supp}(f)$. D'après la remarque ci-dessus on a $\|f\|_\infty < +\infty$ et alors

$$f \leqslant \|f\|_\infty \chi_K \text{ sur } X, \text{ donc } \int_X f d\mu(x) \leqslant \|f\|_\infty \mu(K) < +\infty$$

d'après l'hypothèse. □

Remarque 5.1.3. Sous les hypothèses ci-dessus, on peut donc dire que $C_c(X) \subset L^1(\mu)$, mais il faut noter qu'il s'agit d'un abus de langage (puisque $L^1(\mu)$ est un espace de classes de fonctions, par définition, cf. Définition 3.1.1).

Si μ possède la propriété suivante

$$\text{Tout ouvert } V \subset X \text{ non-vide est de mesure } \mu(V) > 0$$

alors la restriction de la « projection » $C_c(X) \to L^1(\mu)$, qui associe à f sa classe d'équivalence dans L^1, est injective puisque dans ce cas une fonction continue nulle presque partout est identiquement nulle : en effet, tout ensemble dont le complémentaire est de mesure nulle est alors dense dans X, le complémentaire ne pouvant contenir un ouvert non-vide.

Cette hypothèse semble raisonnable mais peut très bien être fausse, par exemple pour la mesure borélienne

$$\mu = (1 - \chi_{[-1,1]}) d\lambda$$

sur $\mathbf{R}$, où λ est la mesure de Lebesgue. Cette mesure est certainement finie sur les compacts, mais par exemple on a

$$\mu(]-1/2, 1/2[) = 0,$$

et deux fonctions f_1 et f_2 continues qui ne diffèrent que sur $]-1/2, 1/2[$ définissent la même fonction dans $L^1(\mu)$.

On va maintenant s'intéresser à deux questions, en apparence indépendantes, mais qui sont liées.

(1) Y-a-t-il une réciproque à cette proposition ? Autrement dit, si on a une forme linéaire Λ positive sur $C_c(X)$, est-elle forcément donnée par « intégration contre une mesure borélienne » ? L'intérêt de cette question est qu'une réponse affirmative fournirait une méthode de construction de mesures non-triviales, si l'on sait construire des formes linéaires Λ intéressantes. Mais cette dernière tâche, qui ne requière que l'étude de fonctions continues, donc relativement régulières en comparaison avec les fonctions mesurables quelconques, peut être plus simple. Le meilleur exemple est celui donné par l'intégrale de Riemann, car toute fonction continue à support compact admet une intégrale de Riemann sur $\mathbf{R}$, et cela définit une forme linéaire positive. La mesure correspondante sera bien évidemment la mesure de Lebesgue, et on verra qu'effectivement cela en donne une construction. Un autre exemple important est le Théorème 5.6.8 ci-dessous.

(2) Si les fonctions continues permettent de caractériser la mesure, est-ce que cela signifie que l'on peut approcher arbitrairement toute fonction μ-intégrable par une fonction continue à support compact ?

5.2. Le théorème de représentation de Riesz

Dans cette section nous donnons la réponse à la première question soulevée ci-dessus : celle-ci est affirmative, si l'espace topologique X vérifie une condition supplémentaire.

Ce résultat s'appelle le théorème de représentation de Riesz.

THÉORÈME 5.2.1. *Soit X un espace topologique* localement compact.[1] *Soit $\Lambda : C_c(X) \to \mathbf{C}$ une forme linéaire positive.*

Il existe une tribu $\mathcal{M} \supset \mathcal{B}_X$ et une mesure complète μ sur $\mathcal{M}$ telle que μ est finie sur les compacts et

$$\Lambda(f) = \int_X f d\mu(x) \text{ pour } f \in C_c(X).$$

De plus, il existe une telle μ vérifiant les propriétés supplémentaires suivantes :

(1) *Pour tout $E \in \mathcal{M}$, on a*

$$\mu(E) = \inf\{\mu(U) \mid U \supset E \text{ est un ouvert contenant } E\} \tag{5.1}$$

(2) *Pour tout $E \in \mathcal{M}$ qui est ouvert ou de mesure finie, on a*

$$\mu(E) = \sup\{\mu(K) \mid K \subset E \text{ est un compact}\}. \tag{5.2}$$

On reviendra plus bas sur le problème de l'unicité de la mesure μ. Il faut comparer les propriétés (5.1), (5.2) à la définition des mesures régulières (Définition 1.4.1).

Avant d'expliquer la preuve de ce théorème, explicitons son application à la construction de la mesure de Lebesgue, qui est très instructive.

EXEMPLE 5.2.2. Soit $X = \mathbf{R}$ et Λ la forme linéaire positive donnée par l'intégrale de Riemann. Appliquant le théorème de Riesz, on obtient une mesure borélienne μ telle que

$$\Lambda(f) = \int_{-\infty}^{+\infty} f(x)dx = \int_{\mathbf{R}} f(x)d\mu(x)$$

(avec à gauche l'intégrale de Riemann) pour $f \in C_c(\mathbf{R})$. Cette mesure est alors la mesure de Lebesgue dont l'existence a été précédemment admise (Théorème 1.3.1).

En effet, μ est une mesure borélienne complète d'après le théorème même et il suffit donc de vérifier que

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

pour tout réels $a \leqslant b$. On peut évidemment supposer $a < b$. Notons $\ell = b - a$. Pour $n > (2\ell)^{-1}$ considérons les fonctions continues à support compact f_n et g_n définies comme suit :

$$f_n(x) = \begin{cases} 1 & \text{si } a \leqslant x \leqslant b \\ 0 & \text{si } x \leqslant a - 1/n \text{ ou } x \geqslant b + 1/n \\ nx - (na - 1) & \text{si } a - 1/n \leqslant x \leqslant a \\ -nx + (nb + 1) & \text{si } b \leqslant x \leqslant b + 1/n \end{cases}$$

et

$$g_n(x) = \begin{cases} 1 & \text{si } a + 1/n \leqslant x \leqslant b - 1/n \\ 0 & \text{si } a \leqslant x \text{ ou } x \geqslant b \\ nx - na & \text{si } a \leqslant x \leqslant a + 1/n \\ -nx + nb & \text{si } b - 1/n \leqslant x \leqslant b \end{cases}$$

(il est utile de faire un dessin ; ces fonctions sont évidemment linéaires par morceaux, et continues).

Par construction on a

$$g_n \leqslant \chi_{[a,b]} \leqslant f_n$$

[1]. C'est à dire que tout point admet un système fondamental de voisinages compacts.

pour $n > (2\ell)^{-1}$ et en intégrant par rapport à μ il vient par monotonie

$$\Lambda(g_n) = \int g_n d\mu \leqslant \mu([a,b]) \leqslant \int f_n d\mu = \Lambda(f_n)$$

et les deux intégrales (de Riemann) de f_n et g_n sont élémentaires (il y a deux demi-rectangles de largeur $1/n$ et de hauteur 1 ajoutés ou retranchés au rectangle $[a,b] \times [0,1]$)

$$\Lambda(f_n) = (b-a) + \frac{1}{n} \text{ et } \Lambda(g_n) = (b-a) - \frac{1}{n}$$

donc, faisant $n \to +\infty$, il vient $\mu([a,b]) = b-a$.

L'hypothèse de compacité locale dans le théorème est justifiée par la nécessité d'avoir un espace $C_c(X)$ « assez gros ». La proposition suivante rappelle les énoncés d'existence de fonctions continues à support compact qui seront utilisées dans la preuve.

Proposition 5.2.3. *Soit X un espace topologique localement compact.*

(1) *Si $V \subset X$ est ouvert, $K \subset V$ est compact, il existe $f \in C_c(X)$ telle que*

$$\chi_K \leqslant f \preccurlyeq \chi_V \tag{5.3}$$

où $f \preccurlyeq \chi_V$ signifie que $f \leqslant \chi_V$ et $\operatorname{supp}(f) \subset V$*. On écrit aussi parfois $f \preccurlyeq V$ simplement.*

(2) *Soient K_1 et K_2 des compacts disjoints dans X. Il existe $f \in C_c(X)$ telle que* $0 \leqslant f \leqslant 1$, *$f(x) = 0$ si $x \in K_1$ et $f(x) = 1$ si $x \in K_2$.*

(3) *Soit $K \subset X$ un compact, V_1, …, V_n un recouvrement ouvert fini de K. Pour toute fonction $g \in C_c(X)$, il existe des fonctions $g_i \in C_c(X)$ telles que* $\operatorname{supp}(g_i) \subset V_i$ *et*

$$\sum_{i=1}^{n} g_i(x) = g(x) \text{ pour } x \in K.$$

De plus, si $g \geqslant 0$, on peut choisir $g_i \geqslant 0$.

Remarque 5.2.4. On peut avoir $f \leqslant \chi_V$ sans avoir $\operatorname{supp}(f) \subset V$, ce qui explique l'introduction de la notation $f \preccurlyeq \chi_V$. En effet, la condition $\operatorname{supp}(f) \subset V$ est importante pour la preuve du théorème de Riesz (cf. par exemple la preuve du Lemme 5.2.6 ci-dessous, en particulier ce qui suit l'introduction des ouverts U_n).

Démonstration. Ce sont des résultats standards. Pour (1), rappelons qu'une construction simple existe si X est un espace métrique : soit $W \subset V$ un voisinage ouvert relativement compact de K, et soit $F = X - W$ son complémentaire fermé. On peut poser

$$f(x) = \frac{d(x,F)}{d(x,F) + d(x,K)}.$$

Le point (2) se déduit de (1) en l'appliquant à $K = K_2$ et V un voisinage de K_2 disjoint de K_1.

Pour (3), on montre classiquement que des fonctions $f_i \in C_c(X)$ existent telles que $0 \leqslant f_i \preccurlyeq \chi_{V_i}$ et $1 = f_1 + \cdots + f_n$. Cela correspond donc au cas $g = 1$, et le cas général s'en déduit en posant $g_i = gf_i$. Si $g \geqslant 0$, on a bien $g_i \geqslant 0$. ☐

Là encore, il est utile de faire un dessin pour visualiser cela. Pour $g = 1$, les fonctions g_i du point (3) sont appelées une *partition de l'unité* sur K, relative au recouvrement V_i.

Preuve du théorème de Riesz. Nous allons écrire la preuve dans le cas où X est compact. Le passage au cas général n'est pas très difficile (voir par exemple [**R**, Ch. 2]). On a donc $C_c(X) = C(X)$.

Les propriétés supplémentaires (5.1) et (5.2) annoncées pour μ permettent de deviner comment définir μ lorsqu'elle est connue sur les ouverts ou sur les compacts. L'exemple de l'intégrale de Riemann, quand à lui, suggère une définition possible pour un ouvert, par exemple, en approchant χ_U par des fonctions continues à support compact.

On pose donc

$$\mu^+(U) = \sup\{\Lambda(f) \mid f \in C(X) \text{ et } 0 \leqslant f \preccurlyeq \chi_U\} \tag{5.4}$$

(le choix de la relation $f \preccurlyeq \chi_U$ au lieu de $f \leqslant \chi_U$ est un point subtil puisque, une fois la mesure construite, on voit a posteriori, par monotonie que $\mu(U) = \sup\{\Lambda(f) \mid f \leqslant \chi_U\}$).

On a alors $\mu^+(U) < +\infty$ pour tout U, et même plus précisément si $U \subset V$ sont des ouverts, il vient

$$0 \leqslant \mu^+(U) \leqslant \mu^+(V) \leqslant \mu^+(X) = \Lambda(1) < +\infty$$

La positivité de $\mu^+(U)$ provient de la positivité de Λ, la seconde inégalité de $\chi_U \leqslant \chi_V$ donc $f \preccurlyeq \chi_U$ implique $f \preccurlyeq \chi_V$. L'inégalité suivante est le cas particulier $V = X$, et enfin l'égalité $\mu^+(X) = \Lambda(1)$ est due au fait que $1 = \chi_X \in C(X)$.

On choisit alors (5.1) pour définir μ^+ en général. Noter que la définition a un sens pour *tout* $E \subset X$. Soit donc

$$\mu^+(E) = \inf\{\mu^+(U) \mid U \supset E \text{ est ouvert}\}$$

(le + est une indication du fait que μ^+ est ce qui s'appelle parfois une « mesure extérieure »). On a donc pour tout E

$$\mu^+(E) \leqslant \mu^+(X) = \Lambda(1).$$

Bien que $\mu^+(E)$ soit partout définie, elle n'est pas dénombrablement additive pour tout E, mais seulement pour une sous-classe d'ensembles, qui évidemment formera la tribu $\mathcal{M}$ cherchée. La définition de $\mathcal{M}$ est motivée par (5.2) : puisqu'il n'y a pas de raison que cette formule soit valide pour tout E, on pose

$$\mu^-(E) = \sup\{\mu^+(K) \mid K \subset E \text{ compact}\}$$

$$\mathcal{M} = \{E \subset X \mid \mu^+(E) = \mu^-(E)\},$$

de sorte que l'on peut définir $\mu : \mathcal{M} \to [0, +\infty[$ (même $[0, \Lambda(1)]$) par

$$\mu(E) = \mu^+(E) = \mu^-(E) \text{ pour } E \in \mathcal{M}.$$

La propriété caractéristique de $E \in \mathcal{M}$ est donc que, pour tout $\varepsilon > 0$, il existe un compact K et un ouvert U tels que $K \subset E \subset U$ et

$$\mu^+(K) \geqslant \mu(E) - \varepsilon, \quad \mu^+(U) \leqslant \mu(E) + \varepsilon. \tag{5.5}$$

La preuve sera complétée par la démonstration des résultats suivants :

(1) $\mathcal{M}$ contient les ouverts.
(2) $\mathcal{M}$ est une tribu.
(3) μ est une mesure sur $\mathcal{M}$.
(4) Si $f \in C(X)$, alors f est μ-intégrable et la forme linéaire Λ est donnée par

$$\Lambda(f) = \int_X f(x)d\mu(x).$$

En pratique, on va mélanger un peu ces différentes étapes. (L'additivité de μ sera connue plus tôt que (2)).

Le premier lemme est un simple énoncé de propriétés complètement élémentaires.

Lemme 5.2.5. (1) *La forme linéaire Λ est monotone : si $f \leqslant g$, alors $\Lambda(f) \leqslant \Lambda(g)$.*
(2) *Si $E \subset F$, on a $\mu^+(E) \leqslant \mu^+(F)$.*
(3) *On a $\emptyset \in \mathcal{M}$, $X \in \mathcal{M}$ et $\mu(\emptyset) = 0$, $\mu(X) = \Lambda(1)$.*

Démonstration. Tout cela est évident. Pour (1), on écrit $\Lambda(g) - \Lambda(f) = \Lambda(g - f) \geqslant 0$ par positivité. ☐

Lemme 5.2.6. *La mesure extérieure μ^+ est dénombrablement sous-additive, c'est à dire que*

$$\mu^+\Bigl(\bigcup_{n \geqslant 1} E_n\Bigr) \leqslant \sum_{n \geqslant 1} \mu^+(E_n)$$

pour tout $E_n \subset X$.

Démonstration. On montre d'abord que si U_1 et U_2 sont des ouverts dans X et $U = U_1 \cup U_2$, on a

$$\mu^+(U) \leqslant \mu^+(U_1) + \mu^+(U_2). \tag{5.6}$$

Soit $g \in C(X)$ telle que $0 \leqslant g \preccurlyeq \chi_U$ et $K = \mathrm{supp}(g)$, qui est compact. D'après la Proposition 5.2.3, (3), il existe g_1, $g_2 \in C(X)$ telles que $g_i \preccurlyeq \chi_{U_i}$ et $g = g_1 + g_2$. Il vient par linéarité de Λ

$$\Lambda(g) = \Lambda(g_1) + \Lambda(g_2) \leqslant \mu^+(U_1) + \mu^+(U_2).$$

En prenant la borne supérieure sur g, on déduit donc (5.6). Par récurrence cela s'étend à la sous-additivité pour une réunion finie d'ouverts.

L'inégalité (5.6) étant acquise, soit $E_n \subset X$ pour $n \geqslant 1$. Fixons $\varepsilon > 0$ quelconque. On peut choisir, par définition, des ouverts $U_n \supset E_n$ tels que

$$\mu^+(U_n) \leqslant \mu^+(E_n) + \frac{\varepsilon}{2^n}$$

pour $n \geqslant 1$. Soit U la réunion des U_n, $f \geqslant 0$ telle que $f \preccurlyeq \chi_U$. Le support de f est inclus dans U,[2] et par compacité il est donc inclus dans une réunion *finie* des U_n, notée $V = U_1 \cup \ldots \cup U_N$. Cela donne donc $f \preccurlyeq \chi_V$ et par conséquent le premier cas permet d'écrire les inégalités

$$\Lambda(f) \leqslant \mu^+(V) \leqslant \sum_{n=1}^{N} \mu^+(U_n) \leqslant \sum_{n \geqslant 1} \mu^+(U_n) \leqslant \sum_{n \geqslant 1} \mu^+(E_n) + \varepsilon.$$

Comme cela vaut pour toute $f \preccurlyeq \chi_U$, il vient

$$\mu^+(U) \leqslant \sum_{n \geqslant 1} \mu^+(E_n) + \varepsilon,$$

et donc par monotonie

$$\mu^+\Bigl(\bigcup_{n \geqslant 1} E_n\Bigr) \leqslant \mu^+(U) \leqslant \sum_{n \geqslant 1} \mu^+(E_n) + \varepsilon.$$

Enfin, en faisant $\varepsilon \to 0$, on trouve le résultat. □

Lemme 5.2.7. *Tout compact est dans $\mathcal{M}$ et tout ouvert est dans $\mathcal{M}$. De plus, si $K \subset X$ est compact, on a*

$$\mu(K) = \inf\{\Lambda(f) \mid \chi_K \leqslant f\}. \tag{5.7}$$

Là encore, on peut comparer avec l'exemple de l'intégrale de Riemann.

Démonstration. Si K est compact, il est clair que $K \in \mathcal{M}$. Soit alors f telle que $\chi_K \leqslant f$, c'est à dire que f est positive et $f \geqslant 1$ sur K. Soit α tel que $0 < \alpha < 1$ et $\varepsilon > 0$. Posons

$$V_\alpha = \{x \mid f(x) > \alpha\} \subset X.$$

L'ensemble V_α est un ouvert contenant K, et donc $\mu(K) \leqslant \mu^+(V_\alpha)$. Il existe $g \preccurlyeq \chi_{V_\alpha}$ telle que $\mu^+(V_\alpha) \leqslant \Lambda(g) + \varepsilon$. Or $g \preccurlyeq \chi_{V_\alpha}$ implique

$$\alpha g \leqslant \alpha \chi_{V_\alpha} \leqslant f,$$

d'où

$$\mu(K) \leqslant \mu^+(V_\alpha) \leqslant \Lambda(g) + \varepsilon \leqslant \Lambda(\alpha^{-1} f) + \varepsilon = \alpha^{-1}\Lambda(f) + \varepsilon,$$

d'où $\mu(K) \leqslant \Lambda(f)$ en faisant $\alpha \to 1$, et $\varepsilon \to 0$, et

$$\mu(K) \leqslant \inf\{\Lambda(f) \mid \chi_K \leqslant f\}$$

puisque f est quelconque telle que $\chi_K \leqslant f$

Réciproquement, pour tout $\varepsilon > 0$ il existe V ouvert tel que $V \supset K$ et

$$\mu^+(V) \leqslant \mu(K) + \varepsilon.$$

2. C'est là qu'il est important d'utiliser $\preccurlyeq$.

En choisissant (Proposition 5.2.3, (1)), une fonction $f \in C(X)$ telle que $\chi_K \leqslant f \preccurlyeq \chi_V$, on trouve l'inégalité

$$\inf\{\Lambda(f) \mid \chi_K \leqslant f\} \leqslant \Lambda(f) \leqslant \mu^+(V) \leqslant \mu(K) + \varepsilon,$$

puis en faisant $\varepsilon \to 0$, l'égalité (5.7).

Il reste à montrer que tout ouvert V est dans $\mathcal{M}$. Soit $\varepsilon > 0$. Il existe $f \leqslant \chi_V$ vérifiant

$$\mu^+(V) \leqslant \Lambda(f) + \varepsilon.$$

Soit $K = \operatorname{supp}(f)$: pour tout $W \supset K$, on a $f \preccurlyeq \chi_W$, donc $\Lambda(f) \leqslant \mu^+(W)$ par définition, et $\Lambda(f) \leqslant \mu(K)$ en prenant la borne inférieure sur W.

Le compact $K \subset V$ vérifie donc

$$\mu^+(V) < \Lambda(f) + \varepsilon \leqslant \mu(K) + \varepsilon,$$

et comme $\varepsilon > 0$ est arbitraire, on trouve que (5.2) est vérifiée pour V. □

Cela donne en particulier le premier point du programme ci-dessus.

Lemme 5.2.8. *Soient $E_n \in \mathcal{M}$, $n \geqslant 1$ des ensembles disjoints. On a alors*

$$E = \bigcup_{n\geqslant 1} E_n \in \mathcal{M} \text{ et } \mu(E) = \sum_{n\geqslant 1} \mu(E_n).$$

Démonstration. Soit d'abord K_1 et K_2 des compacts disjoints de X. La réunion $K = K_1 \cup K_2$ est encore compacte, donc K, K_1 et K_2 sont dans $\mathcal{M}$ par le lemme précédent. On montre d'abord que

$$\mu(K_1 \cup K_2) = \mu(K_1) + \mu(K_2). \tag{5.8}$$

Soit $\varepsilon > 0$. D'après la Proposition 5.2.3, (2), il existe $f \in C(X)$ telle que $0 \leqslant f \leqslant 1$, f est nulle sur K_1 et vaut 1 sur K_2. D'après le Lemme 5.2.7, il existe $g \in C(X)$ telle que $\chi_K \leqslant g$ et $\Lambda(g) \leqslant \mu(K) + \varepsilon$.

Soit $f_1 = (1-f)g$, $f_2 = fg$. On remarque que $\chi_{K_i} \leqslant f_i$. Par (5.7) et linéarité, on obtient

$$\mu(K_1) + \mu(K_2) \leqslant \Lambda(f_1) + \Lambda(f_2) = \Lambda(f_1 + f_2) = \Lambda(g) \leqslant \mu(K) + \varepsilon.$$

Faisant tendre ε vers 0, il vient $\mu(K) \geqslant \mu(K_1) + \mu(K_2)$, et le lemme 5.2.6 montre (5.8). Là encore, par récurrence on en déduit que μ est finiment additive sur les compacts.

Passant au cas général, soit encore $\varepsilon > 0$ quelconque. Il existe par (5.5) des compacts $K_n \subset E_n$ (donc disjoints) tels que

$$\mu(K_n) \geqslant \mu(E_n) - \frac{\varepsilon}{2^n}.$$

Pour tout $N \geqslant 1$, on a par monotonie et en utilisant le premier cas

$$\mu(E) \geqslant \mu\Big(\bigcup_{n\leqslant N} K_n\Big) = \sum_{n=1}^{N} \mu(K_n) \geqslant \sum_{n=1}^{N} \mu(E_n) - \varepsilon.$$

Faisant alors $N \to +\infty$ et $\varepsilon \to 0$, on obtient $\mu(E) \geqslant \sum \mu(E_n)$, et l'inégalité réciproque est donnée par le Lemme 5.2.6. □

Cela montre qu'il ne reste qu'à vérifier que $\mathcal{M}$ est une tribu pour finir les trois premières étapes.

Lemme 5.2.9. *$\mathcal{M}$ est une tribu.*

Démonstration. On va d'abord montrer que $\mathcal{M}$ est stable par réunion finie, intersection finie, et complémentaire.

Soit $\varepsilon > 0$ et $E \in \mathcal{M}$. Il existe un ouvert V et un compact K tels que $K \subset E \subset V$ et

$$\mu(V) - \varepsilon/2 < \mu(E) < \mu(K) + \varepsilon/2,$$

on a alors $V - K = V \cap (X - K)$ ouvert, donc $V - K \in \mathcal{M}$ et

$$\mu(V - K) < \varepsilon$$

puisque on a la réunion disjointe $V = (V - K) \cup K$.

Soient maintenant E_1, $E_2 \in \mathcal{M}$, $F = E_1 - E_2$. D'après ce qui précède, il existe des ouverts V_1, V_2, des compacts K_1, K_2 tels que

$$K_i \subset E_i \subset V_i \text{ et } \mu(V_i - K_i) < \varepsilon.$$

On a alors

$$F \subset (V_1 - K_2) \subset (K_1 - V_2) \cup (V_1 - K_1) \cup (V_2 - K_2),$$

donc par le Lemme 5.2.6,

$$\mu^+(F) \leqslant \mu(K_1 - V_2) + 2\varepsilon,$$

or $K_1 - V_2 \subset F$ est compact, de sorte que $\mu^-(F) \geqslant \mu(K_1 - V_2) \geqslant \mu^+(F) - 2\varepsilon$. Cela vaut pour tout $\varepsilon > 0$, et donc $F \in \mathcal{M}$.

En particulier, prenant $E_1 = X$, on voit que $\mathcal{M}$ est stable par complémentaire ; comme $E_1 \cup E_2 = (E_1 - E_2) \cup E_2$, qui est une union disjointe, le Lemme 5.2.8 et une récurrence montrent que $\mathcal{M}$ est stable par union finie, et enfin comme d'habitude

$$E \cap F = X - ((X - E) \cup (X - F)) \in \mathcal{M}.$$

Cela étant, on sait aussi par le Lemme 5.2.8 que $\mathcal{M}$ est stable par union dénombrable disjointe. Or toute union dénombrable se réécrit comme une union disjointe :

$$E = \bigcup_{n \geqslant 1} E_n = \bigcup_{n \geqslant 1} F_n$$

avec $F_1 = E_1$ et

$$F_{n+1} = E_{n+1} - \bigcup_{i \leqslant n} E_i.$$

Les F_n sont disjoints, de plus $F_n \in \mathcal{M}$ si $E_n \in \mathcal{M}$ pour tout n (utilisant la stabilité par complémentaire, intersection et union finie), donc $E \in \mathcal{M}$. □

Finalement on a[3] :

Lemme 5.2.10. *Toute fonction continue est μ-intégrable et vérifie*

$$\Lambda(f) = \int_X f(x) d\mu(x). \tag{5.9}$$

Démonstration. Puisque μ est une mesure borélienne, toute fonction f continue sur X est mesurable. De plus $|f| \leqslant \|f\|_\infty$ donc par monotonie

$$\int_X |f| d\mu \leqslant \|f\|_\infty \mu(X) < +\infty$$

et par conséquent $C(X) \subset L^1(\mu)$.

Pour montrer (5.9), on peut supposer f à valeurs réelles par linéarité, et il est suffisant de montrer que

$$\Lambda(f) \leqslant \int_X f(x) d\mu(x), \tag{5.10}$$

puisque alors par linéarité on aura également

$$\Lambda(f) = -\Lambda(-f) \geqslant -\int_X -f(x) d\mu(x) = \int_X f(x) d\mu(x).$$

Comme $\Lambda(1) = \mu(X) = \int d\mu$, on vérifie que quitte à ajouter puis soustraire une constante, on peut supposer $f \geqslant 0$.

Soit $K = \mathrm{supp}(f)$. Fixons $\varepsilon > 0$ et $n \geqslant 1$. L'image $f(X) \subset \mathbf{R}$ est compacte, donc incluse dans un intervalle compact $I = [0, M]$. Considérons les ensembles

$$E_i = f^{-1}(]i/n, (i+1)/n]) \cap K \text{ où } -1 \leqslant i \leqslant Mn.$$

[3]. Rappelons pour que le cas de l'intégrale de Riemann, cette étape a déjà été traitée dans la Section 2.4.

On a donc $E_i \in \mathcal{B}$, les E_i sont disjoints et recouvrent K. Il existe pour chaque i un ouvert $V_i \supset E_i$ tel que $\mu(V_i) \leqslant \mu(E_i) + \varepsilon$, et quitte à rétrécir cet ouvert on peut supposer de plus que

$$V_i \subset f^{-1}(]i/n, (i+1)/n + \varepsilon[).$$

D'après la Proposition 5.2.3, (3), il existe des fonctions $g_i \preccurlyeq V_i$ telles que $\sum g_i = 1$. Tout ces préparatifs étant effectués, on a

$$\Lambda(f) = \Lambda\Big(\sum_i f g_i\Big) = \sum_i \Lambda(f g_i) \leqslant \sum_i \Big(\frac{i+1}{n} + \varepsilon\Big)\Lambda(g_i)$$

car $0 \leqslant f g_i \leqslant ((i+1)/n + \varepsilon) g_i$,

$$\leqslant \sum_i \Big(\frac{i+1}{n} + \varepsilon\Big)(\mu(E_i) + \varepsilon)$$

car $\Lambda(g_i) \leqslant \mu(V_i) \leqslant \mu(E_i) + \varepsilon$. Ici, $\varepsilon > 0$ est arbitraire et donc on faisant $\varepsilon \to 0$ on trouve

$$\Lambda(f) \leqslant \sum_i \frac{i+1}{n}\mu(E_i).$$

Mais la fonction étagée

$$s = \sum_i \frac{i}{n}\chi_{E_i}$$

vérifie $s \leqslant f$ donc

$$\Lambda(f) \leqslant \sum_i \frac{i+1}{n}\mu(E_i) = \int_X s(x) d\mu(x) + \frac{\mu(K)}{n} \leqslant \int_X f(x) d\mu(x) + \frac{\mu(K)}{n}.$$

Finalement, on fait $n \to +\infty$ pour obtenir l'inégalité (5.10) voulue. □

On a donc finalement terminé la preuve du Théorème de Riesz. Il s'agit de loin du résultat le plus délicat de tout le cours : le lecteur ne doit pas être découragé si cela semble très technique et difficile. Quitte à en faire une relecture plus tard, il faut continuer car la suite est en descente... □

Remarque 5.2.11. Dans le cas de la construction de la mesure de Lebesgue sur $[0, 1]$, le dernier lemme peut être omis, puisqu'on a déjà vérifié que dans ce cas on a $\mu([a, b]) = b - a$ et on peut alors faire appel aux résultats de la Section 2.4.

Pour éviter de nombreuses répétitions, il est bon de faire une définition du type de mesures fournies par le théorème de Riesz.

Définition 5.2.12. Soit X un espace topologique localement compact. Une mesure de Radon (positive) sur X est une mesure borélienne finie sur les compacts et vérifiant (5.1) pour tout $E \in \mathcal{B}$ et (5.2) pour tout E ouvert ou de mesure finie.

On peut donc exprimer le théorème de Riesz en disant qu'à toute forme linéaire positive sur X est associée une mesure de Radon μ, telle que l'intégrale par rapport à μ est la forme linéaire en question.

5.3. Unicité de mesures boréliennes, et applications

Étant données deux mesures boréliennes μ_1 et μ_2 sur un espace X, une situation fréquente est de pouvoir comparer leurs valeurs pour $U \subset X$ ouvert (ou $K \subset X$ compact). Si celles-ci sont égales, peut-on alors conclure que $\mu_1 = \mu_2$?

Un exemple similaire à celui donné au début de la Section 5.1 montre qu'il faut une hypothèse : si μ_1 est la mesure de comptage sur $\mathbf{R}$ et $\mu_2 = c\mu_1$, avec $c \geqslant 0$, on a $\mu_1(U) = \mu_2(U)$ pour tout ouvert $U \subset \mathbf{R}$ (car U est soit vide, soit infini).

Une hypothèse suffisante sur X est plus forte que l'hypothèse de compacité locale.

Définition 5.3.1. Un espace topologique localement compact X est dit σ-compact si tout ouvert $U \subset X$ est réunion dénombrable de compacts.

Exemple 5.3.2. Soit $X = \mathbf{R}^n$, $n \geqslant 1$. Alors X est σ-compact. En effet, tout ouvert est réunion (forcément dénombrable) d'une suite de compacts pris parmi la famille elle-même dénombrable suivante :

$$K_{x,a,k} = \prod_{i=1}^{n} \Big[x_i, x_i + \frac{a}{2^k}\Big]$$

où $x = (x_1, \ldots, x_n) \in \mathbf{Q}^n$, $k \geqslant 0$ et $a \geqslant 0$ entiers. La preuve en est laissée en exercice. [Indication : Étant donné $U \subset X$ ouvert, considérer la réunion, disons V, de tout les $K_{x,a,k}$ inclus dans U, de sorte que $V \subset U$ et comme toute boule ouverte contient un certain $K_{x,a,k}$, on obtient $U \subset V$ également.]

Un tout petit peu de soin montre que l'on peut choisir les cubes en question de sorte qu'ils soient d'intérieurs disjoints [Indication : Se limiter à des cubes tels que x_i est de la forme $a2^{-b}$, $a \in \mathbf{Z}$ et $b \geqslant 0$; de tels cubes ont la propriété que si deux d'entre eux sont ne sont pas disjoints, l'un contient l'autre, et tout $x \in U$ admet un système fondamental de voisinages de cette forme.]

Lemme 5.3.3. *Soit X un espace topologique σ-compact, et μ une mesure borélienne sur X, finie sur les compacts. Alors μ est régulière, c'est à dire que* (5.1) *et* (5.2) *sont vraies pour tout $E \in \mathcal{B}$.*

En particulier, cela s'applique à la mesure de Lebesgue.

Démonstration. On considère la forme linéaire

$$\Lambda \,:\, f \mapsto \int_X f(x) d\mu(x).$$

D'après la Proposition 5.1.2, Λ est une forme linéaire positive sur $C_c(X)$. D'après le théorème de Riesz, il existe donc une mesure de Radon (Définition 5.3.1) ν telle que

$$\int_X f(x) d\mu(x) = \Lambda(f) = \int_X f(x) d\nu(x)$$

pour toute $f \in C_c(X)$. Nous allons montrer que, d'une part, l'hypothèse de σ-compacité implique que ν vérifie (5.2) pour *tout* $E \in \mathcal{B}$, et d'autre part que $\mu = \nu$.

Pour le premier point, observons qu'il est évident si X est compact puisque ν est finie sur les compacts. Or on peut écrire

$$X = \bigcup_{n \geqslant 1} X_n \text{ avec } X_n \text{ compact,}$$

et on peut évidemment supposer $X_n \subset X_{n+1}$ (quitte à remplacer X_n par $X_n \cup X_{n-1} \cup \cdots \cup X_1$).

Pour tout $E \in \mathcal{B}$ tel que $\nu(E) = +\infty$, on a la limite croissante

$$\lim_{n \to +\infty} \nu(E \cap X_n) = \nu(E) = +\infty.$$

Soit $N > 0$ et n tel que $\nu(E \cap X_n) \geqslant N$. Par (5.2) pour $E \cap X_n$, il existe un compact $K \subset E \cap X_n \subset E$ tel que

$$\nu(K) \geqslant \nu(E \cap X_n) - 1 \geqslant N - 1,$$

ce qui montre (5.2) pour E.

Cela étant, on a

$$\int_X f d\mu(x) = \int_X f d\nu(x) \text{ pour toute } f \in C_c(X). \tag{5.11}$$

Soit $V \subset X$ un ouvert. Par hypothèse sur X, on peut écrire V comme réunion d'une suite de compacts $K_n \subset V \subset X$, que l'on peut supposer croissante. Pour $n \geqslant 1$, il existe une fonction $f_n \in C_c(X)$ telle que

$$\chi_{K_n} \leqslant f_n \leqslant \chi_V \text{ par Proposition 5.2.3, (1).}$$

Soit $g_n(x) = \sup(f_1(x), \ldots, f_n(x)) \in C_c(X)$. On a $g_{n+1} \geqslant g_n$ et $\lim g_n = \chi_V$ car $g_n(x) = 1$ si $x \in K_n$. D'après le théorème de convergence monotone, il vient

$$\mu(V) = \lim_{n\to+\infty} \int_X g_n(x)d\mu(x) = \lim_{n\to+\infty} \int_X g_n(x)d\nu(x) = \nu(V),$$

donc μ et ν coïncident sur les ouverts.

Soit maintenant E un borélien tel que $\nu(E) < +\infty$. Soit $\varepsilon > 0$; il existe $K \subset E \subset V$, avec K compact et V ouvert, tel que $\nu(V - K) < \varepsilon$ par (5.2).

Or $V - K \subset X$ est ouvert, et donc $\mu(V - K) = \nu(V - K) < \varepsilon$. Alors

$$\mu(E) \leqslant \mu(V) = \nu(V) \leqslant \nu(E) + \varepsilon \text{ et } \nu(E) \leqslant \nu(V) = \mu(V) \leqslant \mu(E) + \varepsilon,$$

d'où le résultat $\mu(E) = \nu(E)$ puisque $\varepsilon > 0$ est arbitraire.

Enfin si $E \in \mathcal{M}$ vérifie $\nu(E) = +\infty$, pour tout $N > 1$ il existe un compact $K \subset E$ tel que $\mu(K) = \nu(K) > N$ (car K est de mesure finie), donc $\mu(E) \geqslant \mu(K) > N$, ce qui montre que $\mu(E) = +\infty = \nu(E)$ dans ce cas aussi. □

La preuve a établi en passant le fait utile suivant :

Corollaire 5.3.4. *Soit X un espace topologique σ compact, μ_1 et μ_2 des mesures de Radon sur X. Si on a*

$$\int_X f(x)d\mu_1(x) = \int_X f(x)d\mu_2(x) \text{ pour } f \in C_c(X),$$

alors $\mu_1 = \mu_2$.

Démonstration. C'est ce que montre la deuxième partie de la dernière démonstration (avec μ_1 et μ_2 au lieu μ et ν) à partir de (5.11). □

Proposition 5.3.5. *Soit X un espace topologique σ-compact, μ_1 et μ_2 des mesures boréliennes sur X finies sur les compacts. Soit*

$$\mathcal{K} = \{E \in \mathcal{B} \mid \mu_1(E) = \mu_2(E)\}.$$

(1) *Si $\mathcal{K}$ contient les compacts, alors $\mu_1 = \mu_2$.*
(2) *Si $\mathcal{K}$ contient les ouverts, alors $\mu_1 = \mu_2$.*

Démonstration. D'après le lemme ci-dessus, les mesure μ_1 et μ_2 sont régulières, donc (5.1) et (5.2) sont valides pour tout $E \in \mathcal{B}$. Si (1) est vraie, on utilise (5.2) pour conclure que $\mu_1 = \mu_2$; si (2) est vraie, on utilise à la place (5.1). □

Le corollaire suivant offre une caractérisation intuitive et utile de la mesure de Lebesgue.

Corollaire 5.3.6. *Soit $X = \mathbf{R}^n$, $n \geqslant 1$.*

(1) *La mesure de Lebesgue λ_n sur X est invariante par translation, c'est à dire que pour tout $E \in \mathcal{B}$ et tout $t \in \mathbf{R}^n$, on a*

$$\lambda_n(t + E) = \lambda_n(E),$$

où bien, si $\tau : x \mapsto x + t$ est la translation par t, on a $\tau_*(\lambda_n) = \lambda_n$.

(2) *Si μ est une mesure borélienne sur $\mathbf{R}^n$ telle que μ est finie sur les compacts et μ est invariante par translation, alors il existe $c \in [0, +\infty[$ tel que $\mu = c\lambda_n$.*

Démonstration. Rappelons que $X = \mathbf{R}^n$ est σ-compact.

(1) Soit $t \in \mathbf{R}^n$ et $\mu = \tau_*(\lambda)$. Il s'agit encore d'une mesure borélienne sur $\mathbf{R}$. Comme τ est un homéomorphisme, l'image réciproque d'un compact est compact, et donc μ est finie sur les compacts. D'après la Proposition 5.3.5, il suffit de montrer que $\mu(U) = \lambda(U)$ pour tout ouvert U.

Considérons d'abord le cas $n = 1$. Un ouvert $U \subset \mathbf{R}$ est union disjointe de ses composantes connexes, qui forment une suite *dénombrable* d'intervalles ouverts $]a_n, b_n[$. Comme $\lambda(]a_n, b_n[) = b_n - a_n$ et $\mu(]a_n, b_n[) = \mu(]a_n - t, b_n - t[) = b_n - a_n$, le résultat se déduit de l'additivité dénombrable des mesures.

Si $n \geqslant 2$, on peut utiliser le théorème de Fubini : procédant par récurrence, la translation s'écrit comme composition de n translations dans chacune des directions des coordonnées, et il suffit de traiter le cas où une coordonnée seulement de t est non-nulle, et par symétrie on peut supposer $t = (t_1, 0, \ldots, 0)$. On a alors (cf. (2.7)) pour $U \subset \mathbf{R}^n$

$$\mu(U) = \int_{\mathbf{R}^{n-1}} \Big(\int_{\mathbf{R}} f(u + t_1, x) d\lambda_1(u) \Big) d\lambda_{n-1}(x)$$

où $f(x, u)$ est la fonction caractéristique de U. Pour x fixé, il s'agit de la fonction caractéristique de la tranche $t_x(U)$, qui est un ouvert dans $\mathbf{R}$. D'après le cas $n = 1$, il vient donc

$$\int_{\mathbf{R}} f(u + t_1, x) d\lambda_1(u) = \int_{\mathbf{R}} f(u, x) d\lambda_1(u),$$

et donc

$$\mu(U) = \int_{\mathbf{R}^n} f(u, x) d\lambda_1 \otimes d\lambda_{n-1} = \lambda_n(U).$$

(2) Pour la réciproque, notons $Q = [0, 1]^n$ le cube unité (compact) dans $\mathbf{R}^n$ et posons $c = \mu(Q) < +\infty$. Soit alors

$$\mathcal{K} = \{E \in \mathcal{B} \mid \mu(E) = c\lambda_n(E)\}.$$

de sorte que $Q \in \mathcal{K}$. D'après la Proposition 5.3.5, il suffit de montrer que $\mathcal{K}$ contient les ouverts de $\mathbf{R}^n$. Remarquons que $\mathcal{K}$ est évidemment stable par réunion dénombrable disjointe.

Remarquons d'abord que si $H \subset \mathbf{R}^n$ est un hyperplan, on a $\mu(H) = 0$, et donc $\mu(S) = 0$ pour tout S inclus dans un hyperplan. En effet, par union dénombrable disjointe et translation, il suffit de vérifier que $\mu(Q') = 0$ où $Q' = [0, 1]^{n-1} \times \{0\} \subset Q$. Soit $\alpha = \mu(Q')$. On a pour tout $n \geqslant 1$ la réunion disjointe

$$\bigcup_{i \leqslant n} (1/i + Q') \subset Q$$

donc $n\alpha \leqslant \mu(Q) < +\infty$, ce qui oblige $\alpha = 0$. (En particulier, puisque $\lambda(H) = 0$ également, les sous-ensembles mesurables d'hyperplans appartiennent à $\mathcal{K}$).

Cela nous permet de conclure, par exemple, que

$$\mu(]0, 1[^n) = \mu([0, 1[^n) = \mu(Q) = c.$$

La remarque essentielle est la suivante : si $E \in \mathcal{K}$ est réunion finie disjointe de translatés de $F \in \mathcal{B}$, alors $F \in \mathcal{K}$: en effet

$$E = \bigcup_{1 \leqslant i \leqslant n} (t_i + F)$$

avec $n \geqslant 1$ implique

$$nc\lambda_n(F) = c\lambda_n(E) = \mu(E) = \sum \mu(t_i + F) = n\mu(F) \text{ donc } \mu(F) = c\lambda_n(F).$$

Appliquant cela à Q et à ses translatés, on en déduit que $\mathcal{K}$ contient tout les cubes ouverts du type

$$K^\circ_{x,k} = \prod_{i=1}^{n} \Big] x_i, x_i + \frac{1}{2^k} \Big[$$

avec $k \geqslant 0$, $x \in \mathbf{R}^n$, puisque $K^\circ_{x,k}$ est, à un morceau de frontière près (qui appartient à $\mathcal{K}$), un translaté de

$$Q'' = \Big[0, \frac{1}{2^k} \Big[^n,$$

alors que $[0, 1[^n \in \mathcal{K}$ est réunion disjointe de 2^k translatés de Q''.

Or tout ouvert U s'écrit

$$U = Z \cup V$$

où V est réunion dénombrable disjointe de cubes du type ci-dessus, et $\mu(Z) = 0$ (car réunion de parties des frontières de ces cubes, dont on a vu qu'elles sont chacune de mesure nulle, puisque incluses dans des hyperplans). Il vient $\mu(U) = \mu(V) = c\lambda_n(V) = c\lambda_n(U)$ donc $U \in \mathcal{K}$. □

Cette propriété d'invariance de la mesure de Lebesgue permet d'obtenir une preuve relativement simple de la formule de changement de variables du chapitre précédent. Commençons par le cas d'un changement de variable linéaire.

Lemme 5.3.7. *Soit $T \in GL(\mathbf{R}^n)$ une application linéaire inversible. On a*

$$T_*(\lambda_n) = |\det(T)|^{-1}\lambda_n.$$

Démonstration. Soit $\mu = T_*(\lambda_n)$: c'est une mesure borélienne finie sur les compacts. On va lui appliquer le résultat précédent. Pour tout $t \in \mathbf{R}^n$, $\tau : x \mapsto x + t$ étant la translation correspondante, on a

$$\tau_*(\mu) = (\tau \circ T)_*(\lambda_n).$$

Or par linéarité de T, il vient

$$(\tau \circ T)(x) = T(x) + t = T(x + s) \text{ où } s = T^{-1}(t),$$

(ce qui est bien défini puisque T est inversible). On écrit cela $\tau \circ T = T \circ \tau'$, τ' étant la translation par s. Donc

$$\tau_*(\mu) = (\tau \circ T)_*(\lambda_n) = T_*(\tau'_*(\lambda_n)) = T_*(\lambda_n) = \mu$$

puisque λ_n est invariante par translation. Ceci montre que μ elle-même est invariante par translation, donc d'après le Corollaire 5.3.6, (2), il existe $c \geqslant 0$ tel que

$$T_*(\lambda_n) = c\lambda_n.$$

En évaluant sur le cube unité Q il vient

$$c = T_*(\lambda_n)(Q) = \lambda_n(T^{-1}(Q)) = |\det(T)|^{-1}$$

d'après l'interprétation géométrique du déterminant (Remarque 4.4.8). □

Soient maintenant U et $V \subset \mathbf{R}^n$ des ouverts et $\varphi : U \to V$ un C^1-difféomorphisme. Pour prouver le Théorème 4.4.7, d'après les remarques suivant son énoncé il suffit de montrer que

$$\varphi_*(d\lambda(x)) = |J_{\varphi^{-1}}(y)|d\lambda(y), \tag{5.12}$$

en tant que mesures sur V.

Pour simplifier, posons $J(y) = |J_{\varphi^{-1}}(y)|$. Les deux mesures $\mu_1 = \varphi_*(d\lambda(x))$ et $\mu_2 = J(y)d\lambda(y)$ sont finies sur les compacts : μ_1 car $\varphi^{-1}(K)$ est compact pour tout compact (φ est un homéomorphisme), et μ_2 car J est continue, donc bornée sur les compacts. Ce sont donc des mesures de Radon. L'idée est d'appliquer la Proposition 5.3.5, mais une astuce permet de ne montrer qu'une inégalité, ce qui demande de modifier cette approche un peu. (Comparer avec la preuve du Lemme 5.2.10).

On commence par le cas des cubes.

Lemme 5.3.8. *Soit $Q = \prod [a_i, b_i]$ un cube dans $\mathbf{R}^n$. On a*

$$\mu_1(Q) \leqslant \mu_2(Q).$$

Démonstration. On commence par une inégalité beaucoup plus simple, mais plus faible : pour toute application linéaire inversible $T \in GL(d, \mathbf{R})$, on a

$$\mu_1(Q) \leqslant |\det(T)|^{-1}M^n\lambda(Q), \text{ où } M = \sup\{\|T \circ D_y\varphi^{-1}\| \mid y \in Q\}. \tag{5.13}$$

Supposons d'abord $T = 1$. D'après le théorème de la valeur moyenne, il est classique que $\varphi^{-1}(Q)$ est inclus dans un cube de diamètre[4] au plus M fois le diamètre de Q. L'inégalité (5.13) est alors évidente. Pour T quelconque, on applique cela à $\varphi \circ T^{-1}$ au lieu de φ :

$$\begin{aligned}\mu_1(Q) &= \lambda(\varphi^{-1}(Q)) = \lambda(T^{-1}(T \circ \varphi^{-1}(Q)) \\ &= T_*(T \circ \varphi^{-1}(Q)) = |\det(T)|^{-1}\lambda(T \circ \varphi^{-1}(Q)) \text{ (Lemme 5.3.7)} \\ &\leqslant |\det(T)|^{-1}M^n\lambda(Q).\end{aligned}$$

4. Pour $Q = \prod [a_i, b_i]$, le diamètre est $\sup |b_i - a_i|$.

Cela étant, soit $\delta > 0$ arbitraire et découpons Q en un nombre fini de cubes fermés d'intérieurs disjoints, disons $Q_1, \ldots, Q_m$, tels que le diamètre de chaque cube est $< \delta$. On applique (5.13) sur chaque Q_i séparément avec $T_i = (D_{x_i}\varphi^{-1})^{-1}$, où x_i est tel que $J(x_i)$ réalise le minimum de J restreinte à Q_i. Puisque

$$|\det(T_i)|^{-1} = |\det(D_{x_i}\varphi^{-1})| = J(x_i)$$

par monotonie et additivité,[5] il vient

$$\mu_1(Q) \leqslant \sum_{i=1}^{m} \mu_1(Q_i) \leqslant \sum_{i=1}^{m} M_i^n J(x_i)\lambda(Q_i) \leqslant M^n \mu_2(Q)$$

avec

$$M_i = \sup\{\|T_i \circ D_y\varphi^{-1}\| \mid y \in Q\} \text{ et } M = \sup M_i$$

(on utilise le fait que la fonction étagée $\sum J(x_i)\chi_{Q_i}$ est $\leqslant J\chi_Q$ par construction).

Lorsque δ tend vers 0 (selon une suite δ_n décroissante, disons), on a $M \to 1$ d'après le choix des T_i puisque φ est de classe C^1. D'où le lemme. □

Lemme 5.3.9. *Soit $E \subset V$ un borélien. On a*

$$\mu_1(E) \leqslant \mu_2(E).$$

Démonstration. On suppose d'abord que $E = W$ est un ouvert dans V. Alors on peut (cf. la seconde partie de l'Exemple 5.3.2) écrire W comme réunion dénombrable de cubes d'intérieurs disjoints et l'inégalité $\mu_1(W) \leqslant \mu_2(W)$ découle du lemme précédent par additivité dénombrable.[6]

Si E est un borélien quelconque, puisque μ_2 est une mesure de Radon, il existe par (5.1) une suite (U_n) d'ouverts contenant E tels que $\mu_2(E) = \lim \mu_2(U_n)$. On a pour tout n

$$\mu_1(E) \leqslant \mu_1(U_n) \leqslant \mu_2(U_n)$$

et en faisant $n \to +\infty$, il vient $\mu_1(E) \leqslant \mu_2(E)$. □

On peut alors terminer aisément la preuve de (5.12), puisque si l'on applique l'inégalité précédente à φ^{-1} au lieu de φ, on trouve l'égalité $\mu_1 = \mu_2$.

5.4. Théorèmes d'approximation

Si μ est une mesure borélienne finie sur les compacts sur X, les fonction continues à support compact sont dans $L^p(\mu)$ pour tout $p \geqslant 1$ (y compris $p = +\infty$), par le même argument menant à (5.1.2).

En général les fonctions intégrables sont assez difficiles à « visualiser » très concrètement. Jusqu'à présent on a, pour l'essentiel, étudié ces fonctions par approximation par des fonctions étagées. Mais un ensemble mesurable général E, donc sa fonction caractéristique $f = \chi_E$, reste un object assez complexe et peu explicite. Il est naturel de se demander si l'on peut améliorer cet état de fait en approchant les fonctions dans $L^p(\mu)$ plutôt par des fonctions continues, considérées comme plus régulières, et plus proches de l'intuition. Le théorème suivant répond à la question.

Théorème 5.4.1. *Soit X localement compact, μ une mesure de Radon sur X*

(1) *Soit $p \in [1, +\infty[$. Alors l'image de $C_c(X)$ dans $L^p(\mu)$ est dense dans $L^p(\mu)$.*

(2) *Soit $p = +\infty$; l'adhérence de $C_c(X)$ dans $L^\infty(X)$ est contenu dans l'image de l'espace $C(X)$ des fonctions continues.*

---

5. Noter que l'on utilise le fait que $\mu_2(H) = 0$ pour les hyperplans, car c'est le cas pour la mesure de Lebesgue, mais cela n'est pas requis pour μ_1, et ce n'est d'ailleurs pas évident – c'est vrai à cause de l'égalité $\mu_1 = \mu_2$ que l'on veut prouver...

6. Voir aussi la dernière note de bas de page.

Remarque 5.4.2. En particulier, on voit que (1) est faux pour $p = \infty$. Pour $X = \mathbf{R}^d$, l'adhérence de $C_c(\mathbf{R}^d)$ dans $L^\infty(\mathbf{R}^d)$ est l'espace $C_0(\mathbf{R}^d)$ des fonctions continues tendant vers 0 à l'infini. En effet, si $f_n \to f$ uniformément, $f_n \in C_c(\mathbf{R}^d)$, on a

$$|f(x)| \leqslant |f(x) - f_n(x)| + |f_n(x)|.$$

Pour $\varepsilon > 0$ fixé, il existe n tel que $\|f - f_n\|_\infty < \varepsilon$ donc

$$|f(x)| \leqslant \varepsilon + |f_n(x)|$$

et pour $\|x\|$ assez grand, $x \notin \mathrm{supp}(f_n)$ donc $|f(x)| < \varepsilon$, ce qui montre que $f \in C_0(\mathbf{R}^d)$.

Pour la réciproque, il suffit d'approcher $f \in C_0(\mathbf{R}^d)$ par une suite du type $f_n = fg_n$ où $g_n \in C_c(\mathbf{R}^d)$ vérifie

$$\chi_{[-n,n]^d} \leqslant g_n \leqslant \chi_{[-2n,2n]^d}.$$

On a en effet trivialement

$$\|f - f_n\|_\infty \leqslant \sup\{|f(x)| \mid \|x\| \leqslant n\} \to 0.$$

Démonstration. Le point (2) est simplement dû au fait qu'une suite de fonctions continues qui converge en norme L^∞ converge uniformément sur X, et donc sa limite est continue.

Pour prouver (1), on se ramène au cas des fonctions caractéristiques d'ensembles mesurables, pour lesquels le résultat provient aisément des conditions de régularité (5.1) et (5.2).

Précisément, notons V l'adhérence dans $L^p(\mu)$ de l'image de $C_c(X)$. Il s'agit donc d'un sous-espace vectoriel fermé de $L^p(\mu)$. D'après le Lemme 5.4.3 ci-dessous, V contient les fonctions caractéristiques d'ensembles mesurables E de mesure finie (condition trivialement équivalente à $\chi_E \in L^p(\mu)$ pour $p < +\infty$).

Par linéarité, V contient les fonctions étagées positives dans $L^p(\mu)$. Soit maintenant $f \geqslant 0$ telle que $f \in L^p(\mu)$ et s_n une suite croissante de fonctions étagées positives telles que $s_n(x) \to f(x)$ pour tout $x \in X$ (Proposition 2.2.4). On a donc $s_n \in L^p(\mu)$ et on va montrer que $s_n \to f$ dans $L^p(\mu)$. Comme V est fermé, $f \in V$ et cela finira de montrer que V contient les fonctions L^p qui sont positives, et donc par linéarité $V = L^p(\mu)$.

On a $g_n = |f - s_n|^p \to 0$ pour tout X. De plus, par l'inégalité

$$(a + b)^p \leqslant 2^{p/q}(|a|^p + |b|^p)$$

($p^{-1} + q^{-1} = 1$, inégalité de Hölder triviale !) il vient

$$|g_n| \leqslant 2^{p/q}(|f|^p + |s_n|^p) \leqslant 2^{1+p/q}|f|^p$$

ce qui permet d'appliquer le théorème de convergence dominée pour conclure

$$\|f - s_n\|_p^p = \int_X |g_n| d\mu \to 0$$

comme désiré. □

Lemme 5.4.3. *Soit $E \subset X$ un ensemble de mesure finie, $1 \leqslant p < +\infty$. Pour tout $\varepsilon > 0$ il existe $f \in C_c(X)$ telle que $\|f - \chi_E\|_p < \varepsilon$.*

Démonstration. Par regularité (5.2) et (5.1), il existe $K \subset E \subset U$, K compact et U ouvert, tels que $\mu(U - K) < \varepsilon$. Soit f telle que $\chi_K \leqslant f \preccurlyeq \chi_U$ (donnée par le lemme d'Urysohn). On a alors

$$\int_X |f - \chi_E|^p d\mu \leqslant \int_{U-K} |f - \chi_E|^p d\mu \leqslant \mu(U - K) < \varepsilon$$

car $|f - \chi_E| \leqslant 1$ sur X et $f = \chi_E$ en dehors de U et dans K. Donc $\|f - \chi_E\|_p < \varepsilon^{1/p}$ et le résultat en découle en changeant un tout petit nombre pour un autre... □

Remarque 5.4.4. Si $X = \mathbf{R}$, ayant montré que $C_c(X)$ est dense dans $L^p(\mathbf{R})$ (pour la mesure de Lebesgue particulièrement), on peut aussi regarder des sous-espaces de $C_c(\mathbf{R})$ formés de fonctions plus régulières, de classe C^k par exemple, $k \geqslant 1$ fixé. On montre le résultat suivant :

Proposition 5.4.5. *Soit μ une mesure de Radon sur $\mathbf{R}$, $1 \leqslant p < +\infty$. Alors pour tout $k \geqslant 1$, l'espace $C_c^k(\mathbf{R}^d)$ des fonctions de classe C^k à support compact, est dense dans $L^p(\mu)$.*

Démonstration. Il suffit par le théorème d'approcher dans $L^p(\mu)$ toute fonction continue à support compact par des fonctions de classe C^k.

Soit $f \in C_c(\mathbf{R})$, $K = \mathrm{supp}(f)$· La fonction f restreinte à K est limite uniforme de polynômes f_n d'après le théorème de Weierstrass (cf. le Théorème 4.2.4 aussi), et on a alors $f_n \to f$ dans l'espace $L^p(K, \mu)$.

Cependant, il n'y a pas convergence dans $L^p(\mathbf{R}, \mu)$ car les polynômes f_n ne sont pas à support compacts (s'ils sont non-nuls...) Pour remédier à cela, on considère $g_n = f_n\varphi$ où φ est une fonction dans $C_c^k(\mathbf{R})$ telle que, par exemple $\chi_{[-N,N]} \leqslant \varphi \leqslant \chi_{[-2N,2N]}$ où $N > 0$ vérifie $K \subset [-N, N]$. On a alors trivialement $g_n \to f$ uniformément sur $\mathbf{R}$ et $g_n \to f$ dans $L^p(\mathbf{R}, \mu)$.

Pour construire φ, il s'agit de relier le plateau sur $[-N, N]$ avec la fonction nulle en dehors de $[-2N, 2N]$, de sorte que la fonction résultante soit de classe C^k et à valeurs dans $[0, 1]$. Cela n'est pas bien difficile et le lecteur est encouragé à comparer sa solution avec l'exemple suivant :

$$\varphi(x) = \begin{cases} 1 & \text{si } -N \leqslant x \leqslant N \\ \left(1 - \left(\dfrac{x}{N} - 1\right)^k\right)^k & \text{si } N \leqslant x \leqslant 2N \\ \left(1 - (-1)^k\left(\dfrac{x}{N} + 1\right)^k\right)^k & \text{si } -2N \leqslant x \leqslant -N \\ 0 & \text{si } |x| \geqslant 2N. \end{cases}$$

□

Voir le chapitre suivant, en particulier le Corollaire 6.4.7, pour une forme plus forte de ce théorème lorsque μ est la mesure de Lebesgue.

5.5. Applications simples

Nous allons présenter ici deux exemples relativement simples, mais importants, de l'utilisation du théorème d'approximation de fonctions L^p par des fonctions continues. L'idée est bien entendu, étant donnée une propriété à vérifier pour toute $f \in L^p$, de le faire d'abord pour f continue à support compact, puis d'étendre le résultat « par continuité » (si cela est possible).

Proposition 5.5.1. *Soit $n \geqslant 1$, $1 \leqslant p < +\infty$. Pour toute $f \in L^p(\mathbf{R}^n)$ on a*

$$\lim_{h \to 0} \int_{\mathbf{R}^n} |f(x + h) - f(x)|^p d\lambda_n(x) = 0. \tag{5.14}$$

Rappelons que $L^p(\mathbf{R}^n)$, sans autre précision, désigne $L^p(\mathbf{R}^n, d\lambda)$.

Remarque 5.5.2. (1) L'énoncé n'est pas du tout évident puisque la fonction qui est intégrée, pour $h \neq 0$, n'a pas de raison d'être petite si f n'est pas continue. Il s'agit donc d'un véritable résultat « en moyenne », et non pas point par point. Ceci se confirme en remarquant que l'analogue est faux pour $p = +\infty$: si $f = \chi_{[0,1]} \in L^\infty(\mathbf{R})$, on a pour tout $h \neq 0$

$$\sup\{|f(x + h) - f(x)|\} = 1,$$

(par exemple $f(0) - f(-h) = 1$ si $h > 0$).

(2) Une reformulation de ce résultat est la suivante : pour tout $h \in \mathbf{R}$, on considère l'opérateur linéaire de translation

$$T_h \begin{cases} L^p(\mathbf{R}^n) \to L^p(\mathbf{R}^n) \\ f \mapsto (x \mapsto f(x + h)). \end{cases}$$

Par l'invariance par translation de la mesure de Lebesgue (Corollaire 5.3.6), on a

$$\|T_h(f)\|_p = \|f\|_p$$

pour $f \in L^p(\mu)$, de sorte que T_h est une isométrie (en particulier, T_h est continu). De plus, $T_{h+j} = T_h \circ T_j$, bien évidemment.

La proposition est équivalente à l'assertion selon laquelle l'application $\rho : h \mapsto T_h$ est continue en 0, avec la norme des opérateurs linéaires sur l'espace d'arrivée. En effet, $T_0 = \mathrm{Id}$, et on peut écrire (5.14) sous la forme

$$\lim_{h \to 0} \|f - T_h(f)\|_p = 0,$$

pour toute $f \in L^p(\mathbf{R}^n)$, ce qui implique $\lim T_h = T_0$ d'après le théorème de Banach-Steinhaus.

Comme $T_{h+j} = T_h \circ T_j$, la continuité de ρ en 0 implique la continuité de ρ sur $\mathbf{R}$.

Démonstration. Remarquons tout d'abord que, étant donnée $f \in L^p(\mathbf{R}^n)$, la fonction $g(x) = f(x+h)$ (pour h fixé) est bien définie dans $L^p(\mathbf{R}^n)$ puisque changer f sur un ensemble de mesure nulle Z change g sur l'ensemble $-h + Z$ qui est aussi de mesure nulle puisque la mesure de Lebesgue est invariante par translation (Corollaire 5.3.6, (1)). Pour la même raison, on a $g \in L^p(\mathbf{R}^n)$ (et même $\|g\|_p = \|f\|_p$, cf. aussi la remarque ci-dessus), et donc $g - f \in L^p(\mathbf{R}^n)$, ce qui montre que les intégrales dans (5.14) sont finies pour tout h.

On va, comme déjà mentionné, supposer d'abord que $f \in C_c(\mathbf{R}^n)$. On écrit

$$\psi(h) = \int_{\mathbf{R}^n} |f(x+h) - f(x)|^p d\lambda_n(x) = \int_{\mathbf{R}^n} \varphi(x,h) d\lambda_n(x)$$

avec $\varphi(x,h) = |f(x+h) - f(x)|^p$ pour $(x,h) \in \mathbf{R}^n \times]-1,1[$. La fonction φ est continue en $h = 0$ pour tout x fixé, puisque f est continue, et

$$0 \leqslant \varphi(x,h) \leqslant 2^{p/q}(|f(x+h)|^p + |f(x)|^p) \leqslant 2^{p/q}\|f\|_\infty^p h(x)$$

où h est la fonction caractéristique d'un compact Q assez grand pour que $-h + \mathrm{supp}(f) \subset Q$ pour tout $h \in]-1,1[$ (si $\mathrm{supp}(h) \subset [-A,A]^n$, on peut prendre $Q = [-A-1, A+1]^n$).

Comme Q est compact, h est une fonction intégrable, et la fonction ψ est donc continue en 0 par la Proposition 3.3.1. Comme $\psi(0) = 0$, c'est exactement (5.14). (On peut aussi donner un argument direct en utilisant la continuité uniforme de f pour estimer l'intégrale).

Soit maintenant $f \in L^p(\mu)$ quelconque et $\varepsilon > 0$. Par le théorème d'approximation, il existe $g \in C_c(X)$ telle que

$$\|f - g\|_p < \varepsilon.$$

On a alors

$$|f(x+h) - f(x)|^p \leqslant 3^{p/q}(|f(x+h) - g(x+h)|^p + |g(x+h) - g(x)|^p + |g(x) - f(x)|^p),$$

donc (utilisant encore l'invariance par translation de λ_n)

$$\int_{\mathbf{R}^n} |f(x+h) - f(x)|^p d\lambda_n(x) \leqslant 3^{p/q}\Big(2\|f - g\|_p^p + \int_{\mathbf{R}^n} |g(x+h) - g(x)|^p d\lambda_n(x)\Big).$$

Utilisant le cas d'une fonction continue, il vient

$$0 \leqslant \limsup_{h \to 0} \int_{\mathbf{R}^n} |f(x+h) - f(x)|^p d\lambda_n(x) \leqslant 3^{1+p/q}\varepsilon^p$$

et comme $\varepsilon > 0$ est arbitraire, le résultat suit. □

L'application suivante est à la transformée de Fourier. Rappelons (cf. Section 3.4) que pour $f \in L^1(\mathbf{R})$ on a défini sa transformée de Fourier $\hat{f} : \mathbf{R} \to \mathbf{C}$

$$\hat{f}(t) = \int_{\mathbf{R}} f(y) e(-yt) dy$$

où $e(z) = e^{2i\pi z}$ pour $z \in \mathbf{C}$. On a montré que $\hat{f}$ est une fonction continue bornée (Proposition 3.4.3).

On va améliorer ce résultat.

Théorème 5.5.3. *Soit $f \in L^1(\mathbf{R})$. Alors la transformée de Fourier de f tend vers 0 quand $|t| \to +\infty$, c'est à dire*

$$\lim_{t\to\pm\infty} \hat{f}(t) = 0.$$

Ce résultat important est connu sous le nom de « lemme de Riemann-Lebesgue ».

Nous allons donner trois arguments différents pour le prouver, afin d'illustrer diverses utilisations de l'intégration.

Première preuve. On fait appel à la Proposition 5.5.1 en écrivant

$$\hat{f}(t) = \int_{\mathbf{R}} f(x)e(-xt)dx = -\int_{\mathbf{R}} f(x)e\Big(-t\Big(x+\frac{1}{2t}\Big)\Big)dx = -\int_{\mathbf{R}} f\Big(y-\frac{1}{2t}\Big)e(-yt)dy$$

car $e(1/2) = e^{i\pi} = -1$ et on peut faire le changement de variable (translation) $x = y - 1/(2t)$. Donc il vient

$$\hat{f}(t) = \frac{1}{2}\int_{\mathbf{R}} \Big(f(x) - f\Big(x - \frac{1}{2t}\Big)\Big)e(-xt)dx,$$

d'où

$$|\hat{f}(t)| \leqslant \int_{\mathbf{R}} \Big|f(x) - f\Big(x+\frac{1}{2t}\Big)\Big|dx.$$

Quand $t \to \pm\infty$, on a $1/(2t) \to 0$, et par (5.14) appliqué pour $p = 1$, il vient

$$\lim_{t\to\pm\infty} \hat{f}(t) = 0.$$

□

Seconde preuve. On se ramène plus directement à des fonctions régulières. Plus précisément supposons que $f \in L^1(\mathbf{R})$ est de classe C^1 et à support compact. D'après la Proposition 3.4.3, (3), on a

$$-2i\pi t\hat{f}(t) = \hat{f}'(t)$$

pour $t \in \mathbf{R}$, ce qui montre que la fonction continue $g(t) = |t|\hat{f}(t)$ est bornée sur $\mathbf{R}$: cela implique que

$$\lim_{t\to\pm\infty} \hat{f}(t) = 0$$

dans ce cas.

Comme $C^1_c(\mathbf{R})$ est dense dans $L^1(\mathbf{R})$ d'après la Proposition 5.4.5, le reste de l'argument est habituel : étant donnée $f \in L^1(\mathbf{R})$ et $\varepsilon > 0$, soit $g \in C^1_c(\mathbf{R})$ telle que

$$\|f - g\|_1 < \varepsilon,$$

alors on a

$$|\hat{f}(t)| \leqslant |\hat{g}(t)| + \int_{\mathbf{R}} |f(y) - g(y)|dy \leqslant |\hat{g}(t)| + \varepsilon$$

d'où

$$\limsup_{t\to\pm\infty} |\hat{f}(t)| \leqslant \varepsilon,$$

et le résultat quand $\varepsilon \to 0$. Alternativement, on peut directement invoquer la Remarque 3.4.5 et le fait que l'espace $C_0(\mathbf{R})$ des fonctions tendant vers 0 à l'infini est fermé dans $C_b(\mathbf{R})$ pour la norme $\|\cdot\|_\infty$. □

Troisième preuve. Commençons par le cas de $f = \chi_{[a,b]}$, fonction caractéristique d'un intervalle compact. On a alors explicitement, pour $t \neq 0$

$$\hat{f}(t) = \int_a^b e(-xt)dx = -\frac{1}{2i\pi t}(e(-bt) - e(-at))$$

donc $|\hat{f}(t)| \leqslant (\pi t)^{-1} \to 0$ quand $|t| \to +\infty$. Par linéarité, cela s'étend aux fonctions en escalier à support compact.

Soit maintenant $f \in C_c(\mathbf{R})$, $K = \mathrm{supp}(f) \subset [-A, A]$ pour un certain $A > 0$. Comme f est uniformément continue sur $[-A, A]$, elle est limite uniforme de fonctions en escalier à support dans $[-A, A]$. Par le même argument que dans la deuxième preuve, ou bien en invoquant la Remarque 3.4.5 comme déjà indiqué, le résultat pour f découle de celui pour les fonctions en escalier.

Et enfin le théorème d'approximation permet de passer exactement de la même manière à toute fonction $f \in L^1(\mathbf{R})$. $\square$

5.6. Applications probabilistes du théorème de Riesz

Dans la Section 3.2 nous avons énoncé le théorème de la limite centrale (Théorème 3.2.10). Son énoncé est plus compréhensible si l'on fait intervenir la notion de « convergence en loi » et sa caractérisation en termes de fonctions de répartition. Définissons ces termes...

Définition 5.6.1. (1) Si X_n est une suite de variables aléatoires de lois μ_n et X (resp. μ) est une variable aléatoire (resp. une mesure de probabilité sur $\mathbf{C}$), alors X_n converge en loi vers X (resp. vers μ) si et seulement si

$$\lim_{n\to+\infty} E(f(X_n)) = \lim_{n\to+\infty} \int_{\mathbf{C}} f(x)d\mu_n(x) = \int_X f(x)d\mu(x) = E(f(X))$$

pour toute fonction $f \in C_c(\mathbf{C})$. On note $X_n \xrightarrow{\text{loi}} X$ ou $X_n \xrightarrow{\text{loi}} \mu$.

(2) Soit X une variable aléatoire *réelle* (resp. μ une mesure de probabilité sur $\mathbf{R}$). La fonction de répartition de X (resp. de μ) est la fonction F_X ou F_μ définie sur $\mathbf{R}$ par

$$F(x) = P(X \leqslant x) = \mu(]-\infty, x]).$$

Remarque 5.6.2. (1) Dans la définition de la convergence en loi, il est utile de remarquer que rien n'oblige les variables aléatoires X_n (ni X) à être définies sur le même espace probabilisé : en effet, seules les *lois* μ_n et μ sont importantes. En particulier, il n'y a pas de sens à se demander en toute généralité si la convergence en loi implique la convergence presque sûre par exemple...

(2) Une loi de probabilité sera souvent « présentée » graphiquement par le graphe de sa fonction de répartition qui permet en effet de visualiser assez bien le comportement probabiliste d'une variable aléatoire dont c'est la loi.

(3) Il n'est pas a priori évident dans la définition de la convergence en loi que la mesure limite (la loi de la variable aléatoire limite dans le cas où $X_n \xrightarrow{\text{loi}} X$) est *unique*. Cela provient du Corollaire 5.3.4 puisque $\mathbf{C}$ est σ-compact.

La proposition suivante montre que l'espace des fonctions continues à support compact comme « fonctions test » peut être remplacé par celui des fonctions continues bornées.

Proposition 5.6.3. *Soit (X_n) une suite de variables aléatoires, μ une mesure de probabilité. Alors on a $X_n \xrightarrow{loi} \mu$ si et seulement si pour toute fonction continue bornée $f \in C(\mathbf{C}) \cap L^\infty(\mathbf{C})$ on a*

$$\lim_{n\to+\infty} E(X_n) = \int_{\mathbf{C}} f(x)d\mu(x). \tag{5.15}$$

Démonstration. Bien entendu la condition énoncée est plus forte que celle de la définition puisque $C_c(\mathbf{C}) \subset L^\infty(\mathbf{C})$. Il s'agit donc de montrer la réciproque. En prenant la partie réelle et la partie imaginaire, on se ramène à montrer (5.15) pour f réelle, continue et bornée, et en écrivant $f = f^+ - f^-$, on peut supposer $f \geqslant 0$.

La preuve est basée de manière essentielle sur le point apparemment anodin que $\mu_n(\mathbf{R}) = \mu(\mathbf{R}) = 1$ pour tout $n \geqslant 1$, qui signifie que (5.15) est vraie pour $f = 1$.

Notons μ_n la loi de X_n, qui est donc une mesure de probabilité sur $\mathbf{R}$. On fixe $N > 0$ entier, et on note h_N la fonction continue à support compact dans $[-2N, 2N]$ telle que $h_N(x) = 1$ pour $-N \leqslant x \leqslant N$ et h_N est linéaire sur $[-2N, -N]$ et $[N, 2N]$. Noter que $0 \leqslant h_N \leqslant 1$.

On a alors pour tout $n \geqslant 1$

$$\int_{\mathbf{R}} f(x)h_N(x)d\mu_n(x) \leqslant \int_{\mathbf{R}} f(x)d\mu_n(x) = \int_{\mathbf{R}} f(x)h_N(x)d\mu_n(x) + \int_{\mathbf{R}} f(x)(1-h_N(x))d\mu_n(x) \tag{5.16}$$

Puisque $fh_N \in C_c(\mathbf{R})$ on a

$$\lim_{n\to+\infty} \int_{\mathbf{R}} f(x)h_N(x)d\mu_n(x) = \int_{\mathbf{R}} f(x)h_N(x)d\mu(x)$$

par l'hypothèse $X_n \xrightarrow{\text{loi}} \mu$. De plus comme $1 - h_N \geqslant 0$, on a

$$\int_{\mathbf{R}} f(x)(1 - h_N(x))d\mu_n(x) \leqslant \|f\|_\infty \int_{\mathbf{R}} (1 - h_N(x))d\mu_n(x)$$

et

$$\int_{\mathbf{R}} (1 - h_N)d\mu_n = 1 - \int_{\mathbf{R}} h_N d\mu_n \to 1 - \int_{\mathbf{R}} h_N d\mu = \int_{\mathbf{R}} (1 - h_N)d\mu$$

(utilisant le fait que μ_n et μ sont des lois de probabilité, et que $h_N \in C_c(\mathbf{R})$).

Passant à la limite $n \to +\infty$ dans (5.16), il vient d'après tout ceci

$$\int_{\mathbf{R}} fh_N d\mu \leqslant \liminf_{n\to+\infty} \int_{\mathbf{R}} f d\mu_n \leqslant \limsup_{n\to+\infty} \int_{\mathbf{R}} f d\mu_n \leqslant \int_{\mathbf{R}} fh_N d\mu + \|f\|_\infty \int_{\mathbf{R}} (1 - h_N)d\mu.$$

Maintenant on fait $N \to +\infty$: comme $h_N(x) \to 1$ pour tout x, on a pour tout $x \in \mathbf{R}$

$$f(x)h_N(x) \to f(x) \text{ et } 0 \leqslant f(x)h_N(x) \leqslant f(x)$$
$$1 - h_N(x) \to 0 \text{ et } 0 \leqslant 1 - h_N(x) \leqslant 1$$

donc par le théorème de convergence dominée

$$\int_{\mathbf{R}} fh_N d\mu \to \int_{\mathbf{R}} f d\mu \text{ et } \int_{\mathbf{R}} (1 - h_N)d\mu \to 0,$$

donc il vient

$$\int_{\mathbf{R}} f d\mu \leqslant \liminf_{n\to+\infty} \int_{\mathbf{R}} f d\mu_n \leqslant \limsup_{n\to+\infty} \int_{\mathbf{R}} f d\mu_n \leqslant \int_{\mathbf{R}} f d\mu,$$

ce qui implique que la limite des $\int f\mu_n$ existe et

$$\lim_{n\to+\infty} \int_{\mathbf{R}} f(x)d\mu_n(x) = \int_{\mathbf{R}} f(x)d\mu(x)$$

comme désiré. □

La fonction de répartition permet d'avoir une visualisation plus concrète d'une mesure de probabilité.

Proposition 5.6.4. (1) *Soit μ une mesure de probabilité sur $\mathbf{R}$. La fonction de répartition F de μ est positive, croissante, et vérifie*

$$\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to+\infty} F(x) = 1,$$

et $\mu(]a, b]) = F(b) - F(a)$ pour tout $a < b$.

(2) *De plus, F admet des limites à droite et à gauche en tout $x \in R$, et elle est continue sauf au plus en une suite finie ou dénombrable de points $x_n \in \mathbf{R}$ caractérisés par $\mu(\{x_n\}) > 0$.*

(3) *La fonction de répartition caractérise μ, c'est à dire que $F_\mu = F_\nu$ implique $\mu = \nu$.*

Démonstration. Le point (1) est évident d'après les propriétés de base des mesures (en utilisant le fait que μ est finie).

L'existence des limites à droite et à gauche est une propriété générale d'une fonction monotone, et la continuité sauf au plus en une suite dénombrable de points une conséquence de cela (si $F(x^-) < F(x^+)$, il existe un rationnel $y(x)$ tel que $F(x^-) < y(x) < F(x^+)$ par exemple...)

Finalement pour tout $x \in \mathbf{R}$ on a

$$\mu(\{x\}) = \lim_{n\to+\infty} \mu(]x - n^{-1}, x + n^{-1}]) = \lim_{n\to+\infty} F(x + n^{-1}) - F(x - n^{-1}) = F(x^+) - F(x^-)$$

d'où la caractérisation des points de discontinuité.

Similairement, si $I =]a, b[$ est un intervalle ouvert on a

$$\begin{aligned}\mu(I) &= \lim_{n\to+\infty} \mu(]a, b - n^{-1}]) = \lim_{n\to+\infty} (F(b - n^{-1}) - F(a)) \\ &= F(b^-) - F(a) \text{ si } -\infty < a < b < +\infty \\ \mu(]-\infty, a[) &= F(a^-) \\ \mu(]a, +\infty[) &= 1 - F(a) \\ \mu(\mathbf{R}) &= 1\end{aligned}$$

ce qui montre que $\mu(I)$ ne dépend que de F. Par additivité dénombrable, si U est un ouvert réunion disjointe de ses composantes connexes $I_n =]a_n, b_n[$, on voit que

$$\mu(U) = \sum_n \mu(I_n)$$

est également déterminée par F. Comme une mesure de probabilité sur $\mathbf{R}$ est une mesure de Radon, on déduit de la Proposition 5.3.5, (2) que F détermine la mesure μ. $\square$

Le dernier point de cette proposition amène deux questions proches : peut-on associer, réciproquement, une mesure de probabilité à toute fonction F (vérifiant les conditions de (1)) ? Et, étant donnée F_μ, comment peut-on calculer, par exemple

$$\int_{\mathbf{R}} f(x)d\mu(x) \ ?$$

Une réponse partielle provient de la formule suivante.

Proposition 5.6.5. *Soit μ une mesure de probabilité sur $\mathbf{R}$, F sa fonction de répartition. Pour toute $f \in C_c^1(\mathbf{R})$, fonction C^1 à support compact, on a*

$$\int_{\mathbf{R}} f(x)d\mu(x) = -\int_{\mathbf{R}} f'(x)F(x)dx.$$

En particulier si X est une variable aléatoire dont la loi est μ, on a

$$E(f(X)) = -\int_{\mathbf{R}} f'(x)F(x)dx$$

Démonstration. Formellement, il s'agit d'une simple intégration par partie, F jouant le rôle de la « primitive » (s'annulant en $-\infty$) de μ.

Pour démontrer cela, considérons l'intégrale de droite

$$\int_{\mathbf{R}} f'(x)F(x)dx.$$

Puisque $0 \leqslant F \leqslant 1$ et $\mathrm{supp}(f')$ est compact, l'intégrale existe certainement. On écrit

$$F(x) = \mu(]-\infty, x]) = \int_{\mathbf{R}} \chi(x, y)d\mu(y)$$

où $\chi : \mathbf{R}^2 \to \mathbf{R}$ est donc la fonction caractéristique du sous-ensemble fermé $D = \{(x, y) \mid y \leqslant x\} \subset \mathbf{R}^2$. On écrit l'intégrale comme une intégrale double

$$\int_{\mathbf{R}} f'(x)F(x)dx = \int_{\mathbf{R}}\int_{\mathbf{R}} f'(x)\chi(x, y)d\mu(y)dx,$$

à laquelle le théorème de Fubini est applicable car

$$|f'(x)\chi(x, y)| \leqslant \|f'\|_\infty \chi_K(x)$$

(avec $K = \mathrm{supp}(f)$) et

$$\int_{\mathbf{R}}\int_{\mathbf{R}} \chi_K(x)d\mu(y)dx \leqslant \lambda(K).$$

Intervertissant l'ordre d'intégration, il vient

$$\begin{aligned}\int_{\mathbf{R}} f'(x)F(x)dx &= \int_{\mathbf{R}}\int_{\mathbf{R}} f'(x)\chi(x,y)dxd\mu(y)\\ &= \int_{\mathbf{R}}\int_{[y,+\infty]} f'(x)dxd\mu(y)\\ &= -\int_{\mathbf{R}} f(y)d\mu(y).\end{aligned}$$

□

Cette formule suggère d'essayer d'interpréter la convergence en loi en termes de fonctions de répartition. C'est effectivement possible et fournit un critère très simple, qui est très souvent pris comme définition de la convergence en loi pour des variables aléatoires réelles.

PROPOSITION 5.6.6. *Soit X_n, $n \geqslant 1$, des variables aléatoires* réelles *de lois μ_n, F_n la fonction de répartition de μ_n, μ une mesure de probabilité sur $\mathbf{R}$, F sa fonction de répartition. On a alors $X_n \xrightarrow{loi} X$ si et seulement si $F_n(x) \to F(x)$ quand $x \to +\infty$ pour tout x* qui est un point de continuité de *F.*

DÉMONSTRATION. Supposons d'abord que $F_n(x) \to F(x)$ en tout point de continuité de F. Soit $f \in C_c^1(\mathbf{R})$. D'après la Proposition 5.6.5, on a

$$E(f(X_n)) = \int_{\mathbf{R}} f(x)d\mu_n(x) = -\int_{\mathbf{R}} f'(x)F_n(x)d\lambda(x).$$

Comme l'ensemble des points de discontinuité de F est au plus dénombrable, donc de mesure nulle, on a $F_n(x) \to F(x)$ presque partout pour la mesure de Lebesgue sur $\mathbf{R}$, et $f'(x)F_n(x) \to f(x)F(x)$ presque partout. Comme de plus

$$|f'(x)F_n(x)| \leqslant \|f'\|_\infty \chi_K$$

avec $K = \operatorname{supp}(f')$, compact, on peut appliquer le théorème de convergence dominée et conclure que

$$-\int_{\mathbf{R}} f'(x)F_n(x)d\lambda(x) \to -\int_{\mathbf{R}} f'(x)F(x)d\lambda(x) = \int_{\mathbf{R}} f(x)d\mu(x).$$

Soit maintenant $f \in C_c(\mathbf{R})$. Soit $\varepsilon > 0$. Il existe $g \in C_c^1(\mathbf{R})$ telle que $\|f-g\|_\infty < \varepsilon$ (cf. la preuve de la Proposition 5.4.5). On peut alors écrire

$$\begin{aligned}\Big|\int_{\mathbf{R}} fd\mu_n - \int_{\mathbf{R}} fd\mu\Big| &\leqslant \Big|\int_{\mathbf{R}} (f-g)d\mu_n\Big| + \Big|\int_{\mathbf{R}} gd\mu_n - \int_{\mathbf{R}} gd\mu\Big| + \Big|\int_{\mathbf{R}} (g-f)d\mu\Big|\\ &\leqslant 2\varepsilon + \Big|\int_{\mathbf{R}} gd\mu_n - \int_{\mathbf{R}} gd\mu\Big|.\end{aligned}$$

d'où le résultat à partir du cas précédent.

Considérons la réciproque maintenant. On va utiliser la Proposition 5.6.3. Soit a un point de continuité de F. Pour tout $\varepsilon > 0$, on considère la fonction continue bornée f telle que

$$f(x) = \begin{cases} 1 & \text{si } x \leqslant a\\ 0 & \text{si } x > a+\varepsilon\end{cases}$$

et f est linéaire sur $[a, a+\varepsilon]$. On a $\chi_{]-\infty,a]} \leqslant f \leqslant \chi_{]-\infty,a+\varepsilon]}$ et donc

$$F_n(a) = \mu_n(]-\infty,a]) \leqslant \int_{\mathbf{R}} fd\mu_n \to \int_{\mathbf{R}} fd\mu$$

par la Proposition 5.6.3, et

$$\int_{\mathbf{R}} fd\mu \leqslant \mu(]-\infty,a+\varepsilon]) = F(a+\varepsilon).$$

Il vient donc

$$\limsup_{n\to+\infty} F_n(a) \leqslant F(a+\varepsilon)$$

et similairement on trouve

$$\liminf_{n\to+\infty} F_n(a) \geqslant F(a-\varepsilon).$$

Comme $\varepsilon > 0$ est arbitraire et que F est continue en a, il vient $F_n(x) \to F(x)$ quand $x \to a$. □

Cela étant, on peut donc reformuler le théorème de la limite centrale comme ceci :

Théorème 5.6.7. *Soit (X_n) une suite de variables aléatoires réelles, indépendantes et uniformément distribuées, telles que $X_n \in L^2(\Omega)$ et d'espérance $E(X_n) = 0$. Soit $\sigma = V(X_n)$ la variance des X_n.*

Alors

$$\frac{S_n}{\sqrt{n}} \xrightarrow{loi} \mu_{0,\sigma}$$

la mesure gaussienne centrée de variance σ.

Notre dernière application probabiliste concerne la construction de suites infinies de variables aléatoires indépendantes de lois données : il s'agit donc de la continuation de la Proposition 4.2.3.

Théorème 5.6.8. *Soit I un ensemble quelconque, et pour $i \in I$ soit μ_i une mesure de probabilité borélienne sur $\mathbf{C}$. Il existe un espace probabilisé (Ω, Σ, P) et des variables aléatoires X_i sur Ω telles que $X_i(P) = \mu_i$ pour tout $i \in I$, et les variables (X_i) sont indépendantes.*

Démonstration. Il est pratique d'étendre les mesures μ_i à la « sphère de Riemann » $\hat{\mathbf{C}}$, c'est à dire à l'espace topologique

$$\hat{\mathbf{C}} = \mathbf{C} \cup \{\infty\}$$

(ou l'on peut voir l'élément ∞ comme purement formel, ou bien remarquer qu'effectivement si l'on enlève un point à une sphère, il reste un espace homéomorphe au plan complexe), muni de la topologie ou les voisinages ouverts de ∞ sont les complémentaires $U = \mathbf{C} - K$ d'un compact $K \subset \mathbf{C}$. Cet espace topologique est *compact* (preuve élémentaire).

On étend μ_i en des mesure ν_i sur $\hat{\mathbf{C}}$ en posant

$$\nu_i(X) = \mu_i(X \cap \mathbf{C})$$

pour tout borélien $X \subset \hat{\mathbf{C}}$, autrement dit $\nu_i = j_*(\mu_i)$ où $j : \mathbf{C} \hookrightarrow \hat{\mathbf{C}}$ est l'inclusion.

Soit maintenant Z l'espace produit de I copies de $\hat{\mathbf{C}}$:

$$Z = \hat{\mathbf{C}}^I,$$

muni de la topologie produit. D'après le théorème de Tychonov, c'est un espace compact (c'est pour cela qu'on a compactifié $\mathbf{C}$).

Dans l'espace $C(Z)$ des fonctions continues, on considère le sous-espace

$$C_f(Z) = \{f \in C(Z) \mid f \text{ ne dépend que d'un nombre fini de variables}\},$$

c'est à que que $f \in C_f(Z)$ si et seulement si il existe un sous-ensemble finie $\{i_1, \ldots, i_n\} \subset I$ et une fonction $g : \hat{\mathbf{C}}^n \to \mathbf{C}$ telle que

$$f(x_i) = g(x_{i_1}, \ldots, g_{x_{i_n}}). \tag{5.17}$$

Il est évident que $C_f(Z)$ est une algèbre de fonctions stable par conjugaison complexe, contenant les constantes, et que $C_f(Z)$ sépare les points : si $x = (x_i)$ et $y = (y_i)$ sont différents, disons $x_{i_1} \neq y_{i_1}$, la fonction i_1-ème coordonnée, $f(x_i) = x_{i_1}$, est dans $C_f(Z)$ et sépare x et y. On en déduit par le théorème de Stone-Weierstrass que $C_f(Z)$ est dense dans $C(Z)$ pour la norme $\|\cdot\|_\infty$.

Maintenant on va construire une forme linéaire positive Λ sur $C(Z)$ par continuité à partir de $C_f(Z)$: pour

$$f(x) = g(x_{i_1}, \ldots, g_{x_{i_n}})$$

comme ci-dessus, on pose

$$\Lambda(f) = \int_{\mathbf{C}^n} g(t_1, \ldots, t_n) d\nu_{i_1}(t_1) \otimes \cdots d\nu_{i_n}(t_n). \tag{5.18}$$

Il faut vérifier que ceci ne dépend pas de la représentation de f en (5.17) : en effet, on peut changer l'ordre des variables $i_1, \ldots, i_n$, ou bien introduire d'autres variables pour lesquelles f est « constante ». Mais cela ne change pas le terme de droite de (5.18) car la mesure produit est « invariante par permutations » (c'est le théorème de Fubini !), et les ν_i sont des mesures de probabilité, donc une intégration sur un facteur constant supplémentaire multiplie $\Lambda(f)$ par 1.

Clairement Λ est une forme linéaire sur $C_f(Z)$, et on a $\Lambda(f) \geqslant 0$ si $f \geqslant 0$, mais aussi

$$|\Lambda(f)| \leqslant \|f\|_\infty,$$

donc Λ est continue sur $C_f(Z)$ avec la norme L^∞, de sorte que l'on peut prolonger Λ en une forme linéaire sur $C(Z)$, qui reste clairement positive.

Par le Théorème 5.2.1, il existe donc une mesure de probabilité de Radon P sur Z telle que

$$\Lambda(f) = \int_Z f(x) dP(x)$$

pour toute $f \in C(Z)$. Cette mesure P est intuitivement le produit infini des ν_i. En manipulant des projections sur un nombre fini de variables dans l'esprit du Lemme 4.2.2 et de ses applications dans la Section 4.2, on vérifie alors sans peine que les fonctions coordonnées[7] X_i fournissent les variables aléatoires demandées, leur indépendance provenant du fait que, prises en nombre finie, leurs lois vérifient par construction (5.18) la condition du Lemme 4.2.1. □

Exercice 5.6.9. Montrer la généralisation suivante de ce résultat. Soit I un ensemble quelconque, et pour tout multi-indice $\iota = (i_1, \ldots, i_n)$ d'éléments distincts de I, soit μ_ι une mesure de probabilité borélienne sur $\mathbf{C}$, de telle sorte que les deux conditions suivantes sont vérifiées :

(1) Pour tout n, tout ι et toute permutation σ de $\{1, \ldots, n\}$, on a

$$T_{\sigma,*}(\mu_\iota) = \mu_{\iota'}$$

où T_σ est l'application de permutation

$$(x_1, \ldots, x_n) \mapsto (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$$

et $\iota' = (i_{\sigma(1)}, \ldots, i_{\sigma(n)})$.

(2) Si $m \leqslant n$, alors

$$\pi_*(\mu_\iota) = \mu_{\iota'}$$

où $\pi(x_1, \ldots, x_n) = (x_1, \ldots, x_m)$ et $\iota' = (i_1, \ldots, i_m)$.

Montrer qu'il existe alors un espace probabilisé (Ω, Σ, P) et sur Ω des variables aléatoires X_i, $i \in I$, vérifiant

$$(X_{i_1}, \ldots, X_{i_n})(P) = P_\iota$$

pour tout $\iota = (i_1, \ldots, i_n)$. [Indication : Définir ν_ι, Z et $C_f(Z)$ comme pour le Théorème 5.6.8, et

$$\Lambda(f) = \int_{\mathbf{C}^n} g(t_1, \ldots, t_n) d\nu_\iota(t_1, \ldots, t_n)$$

pour $f \in C_f(Z)$ donnée par (5.18). Montrer que Λ est bien définie puis finir comme précédemment.]

[7]. Noter que, par définition de ν_i, la i-ème coordonnée est presque partout $\neq \infty$, et donc la différence entre $\mathbf{C}$ et $\hat{\mathbf{C}}$ est négligeable de ce point de vue. Rigoureusement on devrait prendre $X_i(x) = x_i$ si $x_i \neq \infty$ et $X_i(x) = 0$ – ou toute autre constante – sinon.

CHAPITRE 6

Convolution de fonctions

6.1. Définition

Nous allons maintenant considérer une construction qui est techniquement très importante : celle de la *convolution* de deux fonctions définies sur $\mathbf{R}^d$, $d \geqslant 1$. Formellement, si f et g sont deux telles fonctions, la convolution $f \star g$ de f et g est définie par

$$f \star g(x) = \int_{\mathbf{R}^d} f(x-t)g(t)d\lambda(t) \tag{6.1}$$

pour $x \in \mathbf{R}^d$, où λ est la mesure de Lebesgue. Bien évidemment, il faut imposer sur f et g des conditions pour que l'intégrale ait un sens. On verra plusieurs cas de figure.

Les propriétés formelles de la convolution resteront valides dans tout les cas. Pour les énoncer, la définition suivante est temporairement utile.

Définition 6.1.1. Soient f et g des fonctions complexes mesurables définies sur $\mathbf{R}^d$. Le domaine de convolution de f et g est l'ensemble $\mathcal{C}(f,g)$ des $x \in \mathbf{R}^d$ tels que l'intégrale ci-dessus ait un sens, c'est à dire tels que $t \mapsto f(x-t)g(t)$ soit dans $L^1(\mathbf{R}^d)$. Pour $x \in \mathcal{C}(f,g)$, on pose

$$f \star g(x) = \int_{\mathbf{R}^d} f(x-t)g(t)d\lambda(t).$$

Remarque 6.1.2. Si f et g sont mesurables, la fonction $(x,t) \mapsto f(x-t)g(t)$ est mesurable sur $\mathbf{R}^{2d}$ puisque produit et composition de fonctions mesurables ou continues. Il n'y a donc pas de problèmes de mesurabilité dans ce qui suit.

Lemme 6.1.3. (1) *Soit $x \in \mathcal{C}(f,g)$. Alors $x \in \mathcal{C}(g,f)$ et $f \star g(x) = g \star f(x)$.*

(2) *Soit $x \in \mathcal{C}(f,g) \cap \mathcal{C}(f,h)$, α, $\beta \in \mathbf{C}$. Alors $x \in \mathcal{C}(f, \alpha g + \beta h)$ et*

$$f \star (\alpha g + \beta h)(x) = \alpha f \star g(x) + \beta f \star h(x).$$

(3) *Si $x \notin \operatorname{supp}(f) + \operatorname{supp}(g)$, alors $x \in \mathcal{C}(f,g)$ et $f \star g(x) = 0$, où*

$$A + B = \{x + y \mid x \in A \text{ et } y \in B\}$$

pour A, $B \subset \mathbf{R}^d$.

(4) *On a $\mathcal{C}(f,g) = \mathcal{C}(|f|,|g|)$, et*

$$|f \star g(x)| \leqslant |f| \star |g|(x) \text{ pour } x \in \mathcal{C}(f,g). \tag{6.2}$$

Démonstration. (1) provient de l'invariance de la mesure de Lebesgue par translation (Corollaire 5.3.6) :

$$f \star g(x) = \int_{\mathbf{R}^d} f(x-t)g(t)d\lambda(t) = \int_{\mathbf{R}^d} f(y)g(x-y)d\lambda(y) = g \star f(x).$$

(2) est évident.

(3) Il suffit de remarque que si $x \notin \operatorname{supp}(f) + \operatorname{supp}(g)$, alors pour tout $t \in \mathbf{R}^d$, si $t \in \operatorname{supp}(g)$ alors $x - t \notin \operatorname{supp}(f)$, et donc $f(x-t)g(t) = 0$ pour tout t, donc $f \star g = 0$.

(4) est également tautologique vu la définition même de l'intégrabilité puisque $|f(x-t)g(t)| = |f(x-t)||g(t)|$ et

$$\Big| \int_{\mathbf{R}^d} f(x-t)g(t)d\lambda(t) \Big| \leqslant \int_{\mathbf{R}^d} |f(x-t)||g(t)|d\lambda(t).$$

$\square$

Remarque 6.1.4. Le point (3) est utilisé, évidemment, sous la forme

$$\operatorname{supp}(f \star g) \subset \operatorname{supp}(f) + \operatorname{supp}(g) \tag{6.3}$$

(lorsque f et g sont telles que $f \star g$ est définie partout, comme dans les cas décrits ci-dessous).

6.2. Cas d'existence de la convolution

Comme souvent, on commence par traiter le cas de fonctions positives. Dans un tel cas, l'intégrale (6.1) existe pour tout x, et prend valeur dans $[0, +\infty]$.

Proposition 6.2.1. *Soient f et g des fonctions mesurables positives et $f \star g$ la convolution de f et g à valeurs dans $[0, +\infty]$. On a alors*

$$\int_{\mathbf{R}^d} f \star g(x) d\lambda(x) = \Big(\int_{\mathbf{R}^d} f(x) d\lambda(x)\Big)\Big(\int_{\mathbf{R}^d} g(x) d\lambda(x)\Big). \tag{6.4}$$

Démonstration. Il suffit d'appliquer le théorème de Tonelli (Théorème 4.3.1, (1)) et l'invariance par translation de la mesure de Lebesgue :

$$\begin{aligned}\int_{\mathbf{R}^d} f \star g(x) d\lambda(x) &= \int_{\mathbf{R}^d}\int_{\mathbf{R}^d} f(x-t) g(t) d\lambda(t) d\lambda(x)\\ &= \int_{\mathbf{R}^d} g(t)\Big(\int_{\mathbf{R}^d} f(x-t) d\lambda(x)\Big) d\lambda(t)\\ &= \Big(\int_{\mathbf{R}^d} f(x) d\lambda(x)\Big)\Big(\int_{\mathbf{R}^d} g(t) d\lambda(t)\Big).\end{aligned}$$

□

Le premier cas important d'existence de la convolution est celle de fonctions dans L^1 avec L^p, $p \geqslant 1$.

Théorème 6.2.2. *Soit $f \in L^p(\mathbf{R}^d)$ et $g \in L^1(\mathbf{R}^d)$, $1 \leqslant p \leqslant +\infty$. Alors $\mathcal{C}(f, g)$ contient presque tout x, et la fonction $f \star g$ ainsi définie presque partout est dans $L^p(\mathbf{R}^d)$ et vérifie*

$$\|f \star g\|_p \leqslant \|f\|_p \|g\|_1 \tag{6.5}$$

et si $p = 1$

$$\int_{\mathbf{R}^d} f \star g(x) d\lambda(x) = \Big(\int_{\mathbf{R}^d} f(x) d\lambda(x)\Big)\Big(\int_{\mathbf{R}^d} g(x) d\lambda(x)\Big). \tag{6.6}$$

De plus si f, g, $h \in L^1(\mathbf{R}^d)$, on a

$$(f \star g) \star h = f \star (g \star h). \tag{6.7}$$

Démonstration. Supposons d'abord $p = 1$. Appliquant la proposition précédente aux fonctions positives $|f|$ et $|g|$, on voit que

$$\int_{\mathbf{R}^d} |f| \star |g| d\lambda = \|f\|_1 \|g\|_1 < +\infty.$$

Cela implique que $|f| \star |g|$ est finie presque partout (par Proposition 2.2.2, (1), rappellons-le), et donc d'après le Lemme 6.1.3, (4), $\mathcal{C}(f, g) = \mathcal{C}(|f|, |g|)$ contient presque tout x.

De plus, intégrant (6.2) sur x, il vient

$$\|f \star g\|_1 \leqslant \int_{\mathbf{R}^d} |f| \star |g| d\lambda = \|f\|_1 \|g\|_1 < +\infty$$

par (6.4).

Étant donc assuré que $f \star g \in L^1(\mathbf{R}^d)$, il est loisible de reprendre la preuve de (6.4) et vérifier que les hypothèses du Théorème de Fubini (Théorème 4.3.1, (2)) sont vérifiées, de sorte que ce même calcul donne (6.6).

Un argument similaire justifie l'application du théorème de Fubini pour montrer que

$$\begin{aligned}(f\star g)\star h(x) &= (g\star f)\star h(x)\\ &= \int_{\mathbf{R}^d}(g\star f)(x-t)h(t)d\lambda(t)\\ &= \int_{\mathbf{R}^d}\int_{\mathbf{R}^d} g(x-t-v)f(v)d\lambda(v)h(t)d\lambda(t)\\ &= \int_{\mathbf{R}^d} f(v)(g\star h)(x-v)d\lambda(v)\\ &= f\star(g\star h)(x).\end{aligned}$$

Nous revenons au cas général où $f\in L^p(\mathbf{R}^d)$. Le cas $p=\infty$ est immédiat. Si $p<+\infty$, le cas $g=0$ est également clair avec $f\star 0=0$, supposons donc $g\neq 0$. Considérons la mesure de probabilité sur $\mathbf{R}^d$ définie par

$$\mu=\frac{|g|}{\|g\|_1}d\lambda.$$

D'après l'inégalité de Jensen (3.6) (ou l'inégalité de Hölder), on a

$$\int_{\mathbf{R}^d}|f(x-t)||g(t)|d\lambda(t)=\|g\|_1\int_{\mathbf{R}^d}|f(x-t)|d\mu(t)\leqslant\|g\|_1\Big(\int_{\mathbf{R}^d}|f(x-t)|^p d\mu(t)\Big)^{1/p},$$

donc

$$\begin{aligned}\int_{\mathbf{R}^d}\Big(\int_{\mathbf{R}^d}|f(x-t)||g(t)|d\lambda(t)\Big)^p d\lambda(x) &\leqslant \|g\|_1^p\int_{\mathbf{R}^d}\int_{\mathbf{R}^d}|f(x-t)|^p d\lambda(x)d\mu(t)\\ &= \|g\|_1^p\Big(\int_{\mathbf{R}^d}d\mu(t)\Big)\Big(\int_{\mathbf{R}^d}|f(x)|^p d\lambda(x)\Big)=\|f\|_p^p\|g\|_1^p<+\infty,\end{aligned}$$

par invariance de la mesure de Lebesgue et définition de μ.

Comme précédemment, cela montre que la fonction $(|f|\star|g|)^p$ est finie presque partout, donc que $\mathcal{C}(f,g)$ contient presque tout $x\in\mathbf{R}^d$, et finalement que

$$\|f\star g\|_p\leqslant\|f\|_p\|g\|_1$$

comme annoncé. □

Remarque 6.2.3. On a donc une application

$$\begin{cases}L^1(\mathbf{R}^d)\times L^1(\mathbf{R}^d)\to L^1(\mathbf{R}^d)\\ (f,g)\mapsto f\star g\end{cases}$$

qui est commutative, associative, distributive par rapport à l'addition : cela justifie l'appellation de *produit de convolution*. Par rapport aux propriétés du produit usuel, il faut noter qu'il n'existe pas d'élément unité $\delta\in L^1(\mathbf{R}^d)$ tel que $f\star\delta=f=\delta\star f$. On verra cependant plus bas qu'il existe des *unités approchées* qui ont des applications importantes (Section 6.4).

L'inégalité (6.5) est très importante : elle établit que le produit de convolution, vu comme opérateur bilinéaire, est continu. En particulier, il en découle que si $f_n\to f$ et $g_n\to g$ sont deux suites convergeant dans $L^1(\mathbf{R}^d)$ et $L^p(\mathbf{R}^d)$ respectivement, alors $f_n\star g_n\to f\star g$ dans L^1 également. L'argument est habituel :

$$\begin{aligned}\|f_n\star g_n-f\star g\|_1 &\leqslant \|f_n\star(g_n-g)\|_1+\|(f_n-f)\star g\|_1\\ &\leqslant \|f_n\|_1\|g_n-g\|_p+\|g\|_p\|f_n-f\|_1\to 0\end{aligned}$$

quand $n\to+\infty$ (puisque $\|f_n\|_1$ est bornée).

De même, si $g\in L^1(\mathbf{R}^d)$ est fixée, et $1\leqslant p\leqslant+\infty$, la convolution par g définit un opérateur linéaire

$$\begin{cases}L^p(\mathbf{R}^d)\to L^p(\mathbf{R}^d)\\ f\mapsto f\star g\end{cases}$$

qui par (6.5) est continue de norme $\leqslant \|g\|_1$.

Avant de présenter d'autres cas d'existence, voici une première interprétation « concrète » du produit de convolution, dans le cadre probabiliste. (Voir le Lemme 7.2.8 pour une autre interprétation simple de la convolution).

Proposition 6.2.4. *Soit (Ω, Σ, P) un espace probabilisé. Supposons que X et Y sont des variables aléatoires réelles* indépendantes *telles que*

$$X(P) = f(x)dx, \text{ et } Y(P) = g(x)dx$$

pour certaines fonctions positives f et g. Alors la loi de $X+Y$ est la mesure

$$(X+Y)(P) = (f \star g)(x)dx.$$

Démonstration. Il s'agit de montrer que pour tout borélien $B \subset \mathbf{R}$ on a

$$P(X+Y \in B) = \int_B f \star g(x)dx.$$

Par définition il vient

$$\begin{aligned}\int_B f \star g(x)dx &= \int_B \int_{\mathbf{R}} f(x-t)g(t)dtdx \\ &= \int_E f(u)g(v)dudv\end{aligned}$$

où $E = \{(u,v) \in \mathbf{R}^2 \mid u+v \in B\}$. Comme $(X,Y)(P) = X(P) \otimes Y(P)$ puisque X et Y sont indépendantes (Lemme 4.2.1), ceci vaut

$$\int_E d(X,Y)P = P((X,Y) \in E) = P(X+Y \in B).$$

$\square$

Remarque 6.2.5. Plus généralement on peut définir le produit de convolution de deux mesures sous certaines conditions. Si μ et ν sont des mesures de probabilité, on pose

$$\mu \star \nu = s_*(\mu \otimes \nu)$$

où $s : \mathbf{C}^2 \to \mathbf{C}$ est l'addition $(x,y) \mapsto x+y$, donc $(\mu \star \nu)(B) = (\mu \otimes \nu)(s^{-1}(B))$.

On a $(f(x)dx) \star (g(x)dx) = (f \star g)(x)dx$, et aussi $(X+Y)(P) = X(P) \star Y(P)$ pour toutes variables aléatoires X et Y indépendantes, ce qui généralise la proposition. Noter que dans ce cadre, il existe une « mesure unité » : si δ est la mesure de Dirac en 0 (cf. Exemple 1.2.4, (2)), on a

$$\mu \star \delta = \delta \star \mu = \mu$$

pour toute mesure de probabilité μ. Voir aussi le Chapitre 9.

Le second cas important d'existence des convolutions est le cas de L^p et L^q où p et q sont des exposants complémentaires.

Proposition 6.2.6. *Soit $p \in [1, +\infty]$, q l'exposant complémentaire tel que $1/p + 1/q = 1$. Pour toute $f \in L^p(\mathbf{R}^d)$ et $g \in L^q(\mathbf{R}^d)$, on a $\mathcal{C}(f,g) = \mathbf{R}^d$ et $f \star g \in L^\infty$. De plus*

$$\|f \star g\|_\infty \leqslant \|f\|_p \|g\|_q. \tag{6.8}$$

Démonstration. Il s'agit d'appliquer l'inégalité de Hölder (3.7) : si f et g sont positives on a pour tout $x \in \mathbf{R}^d$

$$\begin{aligned}f \star g(x) &= \int_{\mathbf{R}^d} f(x-t)g(t)d\lambda(t) \\ &\leqslant \|g\|_q \Big(\int_{\mathbf{R}^d} f(x-t)^p d\lambda(t)\Big)^{1/p} \\ &= \|g\|_q \|f\|_p < +\infty\end{aligned}$$

par invariance de la mesure de Lebesgue toujours. Revenant à $f \in L^p$ et $g \in L^q$ générales, on déduit par le Lemme 6.1.3, (4) que $\mathcal{C}(f, g) = \mathbf{R}^d$ et que

$$|f \star g(x)| \leqslant \|f\|_p \|g\|_q$$

pour tout $x \in \mathbf{R}^d$, ce qui donne (6.8). $\square$

REMARQUE 6.2.7. De (6.8), comme précédemment de (6.5), on déduit la continuité de ce produit : si $f_n \to f$ dans L^p et $g_n \to g$ dans L^q, alors $f_n \star g_n \to f \star g$ dans L^∞ (c'est à dire uniformément).

Le dernier cas d'existence que nous considérons (qui n'est pas le dernier possible !) semble un peu plus complexe, mais il s'avère très pratique.

DÉFINITION 6.2.8. Soit X un espace topologique et μ une mesure borélienne finie sur les compacts. Pour $1 \leqslant p \leqslant +\infty$, on note

$$L^p_c(X) = \{f \in L^p(X) \mid \text{supp}(f) = K \cup Z \text{ où } K \text{ est compact et } \mu(Z) = 0\}$$
$$L^p_{loc}(X) = \{f \mid f \in L^p(K) \text{ pour tout compact } K \subset X\}.$$

L'espace L^p_c est l'espace des *fonctions L^p à support compact*, et l'espace L^p_{loc} celui des *fonctions localement dans L^p*.

REMARQUE 6.2.9. L'introduction de Z dans la définition de L^∞ est faite pour que tout se passe bien avec les fonctions définies à un ensemble de mesure nulle près.

PROPOSITION 6.2.10. *Soit $f \in L^1_{loc}(\mathbf{R}^d)$ et $g \in L^\infty_c(\mathbf{R}^d)$. Alors $\mathcal{C}(f, g) = \mathbf{R}^d$, donc le produit de convolution $f \star g$ est défini pour tout x.*

DÉMONSTRATION. Il suffit de supposer $f \geqslant 0$ et $g \geqslant 0$ comme d'habitude. Or

$$\int_{\mathbf{R}^d} f(x-t)g(t)dt = \int_K f(x-t)g(t)dt$$

si $g(x) = 0$ en dehors du compact K (et d'un ensemble Z de mesure nulle). Donc

$$f \star g(x) \leqslant \|g\|_\infty \int_K f(x-t)dt = \|g\|_\infty \int_{x-K} f(u)du < +\infty$$

puisque $x - K = \{x - k \mid k \in K\}$ est compact et $f \in L^1_{loc}$. $\square$

EXEMPLE 6.2.11. Par exemple, toute fonction continue, ou presque partout continue, est dans L^1_{loc}, mais pas en général dans L^1. Dans la situation de la proposition, la fonction $f \star g$ n'est pas bornée en général.

Noter aussi que, bien évidemment, les trois cas où $f \star g$ a été construite dans cette section ne sont pas exclusifs : il est tout à fait possible que deux, voire les trois, s'appliquent (par exemple, si f et g sont continues à support compact) ; en appliquant les résultats propres à chacune des situations décrites, on obtient alors d'autant plus d'information sur $f \star g$.

6.3. Propriétés de régularisation de la convolution

Si l'on voit $f \star g$ comme une intégrale dépendant du paramètre x, la fonction à intégrer est $h(x, t) = f(x - t)g(t)$, qui du point de vue des résultats de la Section 3.3 est extrêmement simple : à t fixé, c'est une « translation » de f, multiplié par la constante $g(t)$. On peut donc s'attendre à ce que $f \star g$ hérite des propriétés de régularité de f. Et comme $f \star g = g \star f$, la convolution peut même avoir toutes les propriétés de f et de g à la fois. Cela se justifie aisément et fournit une méthode très fructueuse pour construire des fonctions très régulières.

Nous donnons plusieurs exemples de cela. Rappelons la notation des multi-indices pour les dérivées partielles : si $\alpha = (\alpha_1, \ldots, \alpha_d)$ avec $\alpha_i \geqslant 0$ des entiers, on note

$$|\alpha| = \alpha_1 + \cdots + \alpha_d \tag{6.9}$$

et

$$\partial_\alpha = \frac{\partial^{|\alpha|}}{\partial x^\alpha} = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}}, \tag{6.10}$$

vu comme opérateur agissant sur les fonctions de classe $C^{|\alpha|}$ sur un ouvert de $\mathbf{R}^d$.

PROPOSITION 6.3.1. *Soit $k \geqslant 1$, $f \in L^1(\mathbf{R}^d)$ et $\varphi \in C^k(\mathbf{R}^d)$ telle que*

$$\partial_\alpha \varphi \in L^\infty$$

pour tout α tel que $|\alpha| \leqslant k$, c'est à dire que toutes les dérivées partielles d'ordre $\leqslant k$ sont bornées. Alors $f \star \varphi \in C^k(\mathbf{R}^d)$ et, pour tout $|\alpha| \leqslant k$, on a

$$\partial_\alpha(f \star \varphi) = f \star (\partial_\alpha \varphi).$$

DÉMONSTRATION. Tout d'abord, toutes les convolutions rencontrées $f \star \partial_\alpha \varphi$, avec $|\alpha| \leqslant k$, relèvent du cas $p = 1$, $q = \infty$ de la Proposition 6.2.6 et sont donc des fonctions dans $L^\infty(\mathbf{R}^d)$. (D'après la Proposition 6.3.3 ci-dessous, ce sont mêmes des fonctions continues).

En utilisant la définition des dérivées partielles on se ramène aussitôt au cas $d = 1$.[1] Dans ce cas, par récurrence sur k il suffit de traiter le cas $k = 1$. On a pour tout $x \in \mathbf{R}$

$$f \star \varphi(x) = \varphi \star f(x) = \int_{\mathbf{R}} f(t)\varphi(x-t)d\lambda(t),$$

qui est de la forme considérée par le critère de dérivabilité sous le signe d'intégration de la Proposition 3.3.3, avec

$$\frac{\partial}{\partial x} f(t)\varphi(x-t) = f(t)\varphi'(x-t).$$

Comme on a la relation de domination

$$|f(t)\varphi'(x-t)| \leqslant \|\varphi'\|_\infty |f(t)|$$

avec $f \in L^1(\mathbf{R})$, la Proposition 3.3.3 donne le résultat. □

EXEMPLE 6.3.2. Soit par exemple $f \in C_c^k(\mathbf{R})$ et $g \in C_c^m(\mathbf{R})$. On peut appliquer la proposition d'abord à f et g pour voir que $f \star g$ est de classe C^m sur $\mathbf{R}$. Mais on peut aussi échanger le rôle de f et g, et conclure que $f \star g \in C_c^{k+m}(\mathbf{R})$ (cf. (6.3) pour le support), avec par exemple

$$(f \star g)^{(k+m)} = f^{(k)} \star g^{(m)} \ ;$$

noter que pour les premières dérivées on a le choix et

$$(f \star g)' = f' \star g = f \star g'.$$

Une autre propriété de régularisation est plus cachée, puisque les fonctions intervenant n'auront pas de régularité apparente.

PROPOSITION 6.3.3. *Soit $1 \leqslant p \leqslant +\infty$, q l'exposant complémentaire. Pour toute $f \in L^p(\mathbf{R}^d)$, $g \in L^q(\mathbf{R}^d)$, la fonction $f \star g \in L^\infty$ est uniformément continue sur $\mathbf{R}^d$. Si de plus $1 < p, q < +\infty$, alors*

$$\lim_{\|x\| \to +\infty} f \star g(x) = 0.$$

DÉMONSTRATION. Supposons d'abord $p < +\infty$. Soit $x \in \mathbf{R}^d$, $h \in \mathbf{R}^d$. On majore aussitôt par

$$\begin{aligned} |f \star g(x+h) - f \star g(x)| &\leqslant \int_{\mathbf{R}^d} |f(x+h-t) - f(x-t)||g(t)|d\lambda(t) \\ &\leqslant \|g\|_q \Big(\int_{\mathbf{R}^d} |f(x+h-t) - f(x-t)|^p d\lambda(t) \Big)^{1/p} \end{aligned}$$

[1]. Il y a une petite subtilité liée au fait que $f \in L^1$ n'implique pas que les fonctions partielles du type $g(x) = f(x, x_2, \ldots, x_d)$ avec x_i, $i \geqslant 2$, fixés, sont dans $L^1(\mathbf{R})$ pour tout (x_i). L'argument fubinesque simple permettant de contourner cela est laissé à l'imagination du lecteur...

par l'inégalité de Hölder encore. Par la substitution linéaire $x - t = u$, l'intégrale en t est

$$\int_{\mathbf{R}^d} |f(x+h-t) - f(x-t)|^p d\lambda(t) = \int_{\mathbf{R}^d} |f(u+h) - f(u)|^p d\lambda(u).$$

Cette quantité, disons $\varepsilon(h)$, est maintenant indépendante de x, et tend vers 0 quand $h \to 0$ d'après la Proposition 5.5.1. Comme

$$|f \star g(x+h) - f \star g(x)| \leqslant \varepsilon(h)^{1/p} \|g\|_q$$

on voit que $f \star g$ est bien uniformément continue.

Si $p = \infty$, on a $q = 1$ et donc $g \star f = f \star g$ est uniformément continue par le cas précédent...

Finalement, si p et $q \neq \infty$, il existe par le Théorème 5.4.1 des suites f_n et g_n de fonctions continues à support compact convergeant dans L^p et L^q respectivement vers f et g ; alors (Remarque 6.2.7) on a $f_n \star g_n \to f \star g$ uniformément, et comme $f_n \star g_n \in C_c(\mathbf{R}^d)$ (cf. (6.3)), on a (Remarque 5.4.2)

$$\lim_{\|x\| \to +\infty} f \star g(x) = 0.$$

□

Exemple 6.3.4. Une application classique de ce résultat est la suivante. Soient A, $B \subset \mathbf{R}^d$ des ensembles quelconques de mesure de Lebesgue non-nulle : $\lambda(A) > 0$, $\lambda(B) > 0$. Alors l'ensemble $C = A + B$ est d'intérieur non-vide, autrement dit il existe $c = a + b \in C$ et un $r > 0$ tel que la boule de centre c et de rayon r est incluse dans C. Pour $d = 1$, la somme de deux ensembles de mesure non-nulle contient donc forcément un intervalle ouvert non-vide ! Rappelons que tant A que B peuvent tout à fait être d'intérieur vide eux-mêmes (par exemple $A = \mathbf{R} - \mathbf{Q} \subset \mathbf{R}$).

La preuve est extrêmement simple en utilisant la convolution. En effet, on peut supposer $0 < \lambda(A), \lambda(B) < +\infty$, si besoin est en considérant $A \cap D$ et $B \cap D$, où D est une boule de rayon assez grand.

Soient alors $f = \chi_A$, $g = \chi_B$. Puisque A et B sont de mesure finie, $f \in L^2$, $g \in L^2$. La Proposition 6.3.3 montre que $h = f \star g \geqslant 0$ est uniformément continue. Comme f et g sont positives on a

$$\int h dx = \int f dx \int g dx = \lambda(A)\lambda(B) > 0$$

donc $h \geqslant 0$ n'est pas nulle. Soit c tel que $h(c) > 0$. Par continuité, il existe une boule U centrée en c de rayon $r > 0$ telle que $h(x) > 0$ pour tout $x \in U$. Or, comme pour le Lemme 6.1.3, (3), on a $h(x) = 0$ pour $x \notin A + B$ donc $U \subset A + B$.

Un exemple extrême de régularisation qui est assez éclairant (quoique pas très significatif pour les applications) est le suivant :

Exercice 6.3.5. Soit $f \in L^\infty_c(\mathbf{R})$ et $g \in \mathbf{C}[x]$ un polynôme. Montrer que $f \star g$ existe et est un polynôme.

6.4. Approximation par convolution

Les propriétés de régularité des produits de convolution s'avèrent d'autant plus significatives qu'il est possible d'approcher une fonction f donnée par des produits de convolution $f \star \varphi$, où φ est une « unité approchée », essentiellement universelle, dont la régularité (support compact, dérivabilité, etc) peut être extrêmement grande.

Commençons par justifier qu'il n'existe pas d'unité de convolution δ dans $L^1(\mathbf{R}^d)$, bien que cela ne soit logiquement pas nécessaire.

Lemme 6.4.1. *Supposons que $\delta \in L^1(\mathbf{R}^d)$ vérifie $f \star \delta = f$ pour toute $f \in L^1(\mathbf{R}^d)$. Alors on a*

(1) $\delta(x) = 0$ *pour presque tout* $x \neq 0$, (2) $\int \delta(x) d\lambda(x) = 1$.

Bien entendu, ces deux conditions ne sont pas compatibles pour $\delta \in L^1(\mathbf{R}^d)$, puisque la première implique que δ est nulle presque partout, donc son intégrale devrait être nulle ! Mais si l'on relaxe quelque peu la première condition, par exemple en demandant $\delta(x) = 0$ pour $\|x\| > \eta$, avec η petit, on obtient des conditions parfaitement compatibles, et on peut s'attendre à ce que $f \star \delta$ soit, alors, « proche » de f.

Un gain inattendu est que les unités approchées peuvent être pratiquement aussi régulières que désiré, et que $f \star \delta$ peut hériter de cette régularité : cela peut donc fournir des approximations très régulières de f ! (Ce qu'une unité δ « exacte » avec $f \star \delta = f$ ne peut pas donner...)

Remarque 6.4.2. Si on utilise la transformation de Fourier (plus exactement le Lemme 7.2.8), on s'aperçoit que pour que δ soit une unité de convolution, il faudrait que $\hat{\delta}$ soit l'unité pour la multiplication ordinaire, c'est à dire $\hat{\delta} = 1$. Mais c'est impossible car 1 n'est pas une fonction qui tend vers 0 à l'infini (cf. le Théorème 5.5.3 et le Lemme 7.1.2), propriété valide pour la transformée de Fourier d'une fonction intégrable : cela redonne donc l'inexistence de δ.

Démonstration. La condition est que pour toute $f \in L^1(\mathbf{R}^d)$, on a

$$f(x) = \int_{\mathbf{R}^d} f(x-t)\delta(t)d\lambda(t)$$

pour presque tout x. En prenant $f(x) = \chi(-x)$ où χ est la fonction caractéristique de $N = \{x \mid \delta(x) < 0\}$, et $x = 0$, on trouve

$$0 \leqslant f(0) = \int_{\mathbf{R}^d} f(-t)\delta(t)d\lambda(t) = \int_N \delta(t)d\lambda(t)$$

ce qui implique que N est de mesure nulle et donc $\delta \geqslant 0$ presque partout.

En prenant ensuite f la fonction caractéristique d'un cube $[-A, A]^d$, et $x = 0$ encore, on trouve que

$$1 = \int_{[-A,A]^d} \delta(t)d\lambda(t)$$

pour tout A. Cela implique la condition (2), mais également (1) puisque pour tout $B > A$ on a

$$0 = \int_{[-B,B]^d - [-A,A]^d} \delta(t)d\lambda(t)$$

avec $\delta \geqslant 0$, donc $\delta = 0$ sur tout ensemble $[-B, B]^d - [-A, A]^d$, ce qui implique $\delta(x) = 0$ hors de $x = 0$. ☐

La preuve permet de deviner quel type de fonctions pourra fournir une bonne approximation à δ. On utilisera ici la définition suivante :

Définition 6.4.3. Une suite de Dirac dans $\mathbf{R}^d$ est une suite (φ_n) de fonctions positives telles que

$$\int_{\mathbf{R}^d} \varphi_n(t)d\lambda(t) = 1 \text{ pour tout } n \geqslant 1 \tag{6.11}$$

$$\lim_{n\to+\infty} \int_{\|t\|>\eta} \varphi_n(t)d\lambda(t) = 0 \text{ pour tout } \eta > 0. \tag{6.12}$$

où l'on note $\|t\|$ la norme euclidienne (4.23) sur $\mathbf{R}^d$.

Remarque 6.4.4. On peut remplacer dans (6.12) la norme par n'importe quelle norme équivalente ; et on sait que dans l'espace de dimension finie $\mathbf{R}^d$, toutes les normes le sont. Par exemple, il est parfois utile de prendre

$$\|t\|_\infty = \max |t_i| \text{ ou } \|t\|_1 = \sum_{1\leqslant i\leqslant d} |t_i|.$$

Commençons par des exemples.

Exemple 6.4.5. (1) Soit $\varphi_n = (2n)^d \chi_{[-n^{-1},n^{-1}]^d}$. C'est une suite de Dirac, puisque (6.11) est évident et

$$\int_{\|t\|>\eta} \varphi_n(t) d\lambda(t) = 0$$

pour tout $n > \eta^{-1}$, donc la limite dans (6.12) est bien nulle.

(2) Plus généralement, soit φ_n une suite quelconque de fonctions L^1 positives telles que (6.11) est vraie, et telles que $\mathrm{supp}(\varphi_n) \subset \{\|x\| < \varepsilon_n\}$ avec $\varepsilon_n \to 0$ quand $n \to +\infty$. Alors (φ_n) est une suite de Dirac pour la même raison qu'en (1).

(3) La condition (6.12) permet des fonctions plus générales que celles dont le support est et compact « tend vers $\{0\}$ » comme en (2). La condition équivaut à dire que $\varphi_n \to 0$ dans $L^1(\{\|t\| > \eta\})$ pour tout $\eta > 0$. Il s'agit, intuitivement, que l'essentiel de la « masse » de la mesure de probabilité $\mu_n = \varphi_n(t) d\lambda(t)$ se concentre à l'origine quand $n \to +\infty$.

Soit par exemple $\varphi \geqslant 0$ une fonction L^1 quelconque telle que $\|\varphi\|_1 > 0$. On pose $\psi = \varphi / \|\varphi\|_1$, et

$$\psi_n(t) = n^d \psi(nt)$$

pour $n \geqslant 1$. Alors (ψ_n) est une suite de Dirac. (Remarquer que l'exemple (1) est de ce type).

En effet, par changement de variable linéaire $u = nt$, dont le jacobien est $\det(n) = n^d$, on a

$$\int_{\mathbf{R}^d} \psi_n(t) d\lambda(t) = n^d \int_{\mathbf{R}^d} \psi(nt) d\lambda(t) = \int_{\mathbf{R}^d} \psi(u) d\lambda(u) = 1$$

ce qui donne (6.11). Quand à (6.12), on a par le même changement de variable

$$\int_{\|t\|>\eta} \psi_n(t) d\lambda(t) = \int_{\|t\|>n\eta} \psi(t) d\lambda(t).$$

Comme ψ est intégrable et

$$\bigcup_{n\geqslant 1} \{t \mid \|t\| \leqslant n\eta\} = \mathbf{R}^d,$$

on a

$$\lim_{n\to+\infty} \int_{\|t\|>n\eta} \psi(t) d\lambda(t) = 0$$

(Lemme 2.3.10), ce qui donne donc (6.12).

Un exemple important consiste à prendre $\psi(t) = e^{-\pi\|t\|^2}$, puisque d'après la Proposition 4.4.10 et le théorème de Fubini on a

$$\int_{\mathbf{R}^d} e^{-\pi\|t\|^2} d\lambda(t) = \int_{\mathbf{R}^d} \exp\Bigl(-\pi \sum_{1\leqslant i\leqslant d} t_i^2\Bigr) dt_1 \cdots dt_d = \Bigl(\int_{\mathbf{R}} e^{-\pi t^2} d\lambda(t)\Bigr)^d = 1.$$

On a maintenant le théorème d'approximation suivant qui justifie les remarques faites au début de cette section :

Théorème 6.4.6. *Soit (φ_n) une suite de Dirac fixée.*

(1) *Pour tout p, $1 \leqslant p < \infty$, et toute $f \in L^p(\mathbf{R}^d)$, on a*

$$f \star \varphi_n \to f \text{ dans } L^p(\mathbf{R}^d).$$

(2) *Si $f \in L^\infty$ est continue en $x \in \mathbf{R}^d$, alors*

$$f \star \varphi_n(x) \to f(x),$$

et si $f \in L^\infty$ est uniformément continue

$$f \star \varphi_n \to f \text{ uniformément sur } \mathbf{R}^d.$$

Démonstration. On vérifie d'abord que chacun des produits de convolution écrits dans l'énoncé sont bien définis : pour (1), comme $\varphi_n \in L^1$, on a $f \star \varphi_n \in L^p(\mathbf{R}^d)$ d'après le Théorème 6.2.2, tandis que pour (2), d'après la Proposition 6.2.6 avec $p = 1$, $q = \infty$, et la Proposition 6.3.3, la fonction $f \star \varphi_n$ est uniformément continue sur $\mathbf{R}^d$, ce qui justifie de parler de ses valeurs ponctuelles.

Le début de l'argument est commun aux deux cas (et à bien d'autres situations similaires) : on utilise (6.11) pour écrire

$$f \star \varphi_n(x) - f(x) = \int_{\mathbf{R}^d} (f(x-t) - f(x))\varphi_n(t)d\lambda(t) \tag{6.13}$$

pour tout $n \geqslant 1$ (pour presque tout x dans le cas (1)). Puis, considérant que $f(x-t)-f(x)$ sera « petit » pour t proche de 0 (en moyenne ou littéralement), et que la contribution des grandes valeurs de t sera négligeable par (6.12), on va séparer l'intégration en deux morceaux, $\|t\| > \eta$ et $\|t\| \leqslant \eta$.

Considérons d'abord le cas (1) avec $p < +\infty$. Appliquant l'inégalité de Hölder (ou celle de Jensen) pour les fonctions positives 1 et $t \mapsto |f(x-t) - f(x)|$, et la mesure de probabilité $\varphi_n(t)d\lambda(t)$, on a pour presque tout x

$$|f \star \varphi_n(x) - f(x)|^p \leqslant \int_{\mathbf{R}^d} |f(x-t) - f(x)|^p \varphi_n(t)d\lambda(t),$$

d'où

$$\|f \star \varphi_n - f\|_p^p \leqslant S_\eta + T_\eta$$

où, pour $\eta > 0$, on a posé

$$S_\eta = \int_{\mathbf{R}^d} \int_{\|t\| \leqslant \eta} |f(x-t) - f(x)|^p \varphi_n(t)d\lambda(t)d\lambda(x)$$
$$T_\eta = \int_{\mathbf{R}^d} \int_{\|t\| > \eta} |f(x-t) - f(x)|^p \varphi_n(t)d\lambda(t)d\lambda(x).$$

En appliquant le théorème de Tonelli, il vient

$$S_\eta = \int_{\|t\| \leqslant \eta} \Bigl(\int_{\mathbf{R}^d} |f(x-t) - f(x)|^p d\lambda(x)\Bigr)\varphi_n(t)d\lambda(t).$$

D'après la Proposition 5.5.1, pour tout $\varepsilon > 0$ il existe $\eta > 0$ tel que pour $\|t\| < \eta$, on a

$$\int_{\mathbf{R}^d} |f(x-t) - f(x)|^p d\lambda(x) < \varepsilon,$$

donc pour un tel η (fixé) on trouve

$$S_\eta \leqslant \varepsilon \int_{\|t\| \leqslant \eta} \varphi_n(t)d\lambda(t) \leqslant \varepsilon$$

(par (6.11) et monotonie).

Quand à T_η, on a $|f(x-t) - f(x)|^p \leqslant 2^{p/q}\{|f(x-t)|^p + |f(x)|^p\}$, donc par Tonelli encore et invariance de la mesure de Lebesgue par translation il vient

$$T_\eta \leqslant 2^p \|f\|_p^p \int_{\|t\| > \eta} \varphi_n(t)d\lambda_n(t).$$

D'après (6.12), pour tout n assez grand on aura

$$T_\eta \leqslant \varepsilon,$$

et finalement pour n assez grand il vient

$$\|f \star \varphi_n - f\|_p^p \leqslant 2\varepsilon,$$

$\varepsilon > 0$ étant arbitraire, on a convergence de $f \star \varphi_n$ vers f dans L^p.

Considérons maintenant le cas (2) pour un x fixé. À partir de (6.13), on a maintenant

$$|f \star \varphi_n(x) - f(x)| \leqslant S_\eta + T_\eta$$

avec

$$S_\eta = \int_{\|t\| \leqslant \eta} |f(x-t) - f(x)| \varphi_n(t) d\lambda(t)$$

$$T_\eta = \int_{\|t\| > \eta} (|f(x-t)| + |f(x)|) \varphi_n(t) d\lambda(t).$$

Par continuité de f en x, pour tout $\varepsilon > 0$, il existe η tel que si $\|t\| < \eta$, alors $|f(x-t) - f(x)| < \varepsilon$. Pour ce choix de η, il vient alors

$$S_\eta \leqslant \varepsilon$$

et

$$T_\eta \leqslant 2\|f\|_\infty \int_{\|t\| > \eta} \varphi_n(t) d\lambda(t).$$

La preuve de (2) se termine alors exactement comme précédemment. Finalement, si f est uniformément continue et $K \subset \mathbf{R}^d$ est compact, alors on peut choisir η indépendamment de x dans l'argument précédent, et la majoration devient une borne uniforme pour $\|f \star \varphi_n - f\|_\infty$. □

Nous allons déduire de ce théorème un corollaire important.

Corollaire 6.4.7. *L'espace $C_c^\infty(\mathbf{R}^d)$ des fonctions C^∞ à support compact est dense dans $L^p(\mathbf{R}^d)$ pour tout p, $1 \leqslant p < +\infty$.*

Il est évident que C_c^∞ est un espace vectoriel tel que $C_c^\infty(\mathbf{R}^d) \subset C_c(\mathbf{R}^d) \subset L^p(\mathbf{R}^d)$ pour tout p, $1 \leqslant p \leqslant +\infty$.

Pour la preuve il suffit d'avoir à sa disposition une seule fonction $\psi \in C_c^\infty(\mathbf{R}^d)$ convenable...

Lemme 6.4.8. *Il existe $\psi \in C_c^\infty(\mathbf{R}^d)$ telle que $\psi \neq 0$ et $\psi \geqslant 0$.*

Corollaire 6.4.9. *Il existe une suite de Dirac (φ_n) telle que $\varphi_n \in C_c^\infty(\mathbf{R}^d)$ pour tout $n \geqslant 1$.*

Démonstration. Soit ψ comme dans le lemme. Alors $\|\psi\|_1 > 0$, puisque si $\psi(x_0) > 0$, par continuité on a $\psi(x) > 0$ pour x dans un voisinage de mesure > 0 de x_0. On construit alors φ_n comme dans l'Exemple 6.4.5, (3) : soit

$$\varphi = \frac{\psi}{\|\psi\|_1} \in C_c^\infty(\mathbf{R}^d),$$

qui vérifie $\varphi \geqslant 0$ et $\int \varphi(x) d\lambda(x) = 1$, et

$$\varphi_n(x) = n^d \varphi(nx).$$

On a $\varphi_n \in C_c^\infty(\mathbf{R}^d)$, et d'après l'exemple déjà mentionné, la suite (φ_n) est une suite de Dirac. □

Preuve du théorème. Soit p, $1 \leqslant p < +\infty$. Soit $f \in L^p(\mathbf{R}^d)$ et $\varepsilon > 0$. Il existe $g \in L_c^p(\mathbf{R}^d)$ telle que $\|f - g\|_p < \varepsilon$: en effet, on peut soit choisir $g \in C_c(\mathbf{R}^d)$ vérifiant cela (Théorème 5.4.1), soit remarquer que $g_n = \chi_n f$, où χ_n est la fonction caractéristique de la boule de rayon $n > 0$, est une suite de fonctions dans $L_c^p(\mathbf{R}^d)$ ($|g_n| \leqslant |f|$) telle que $g_n \to f$ dans $L^p(\mathbf{R}^d)$, donc pour n assez grand, g_n convient.

Remarquons aussi que $g_n \in L^1(\mathbf{R}^d)$ puisque la mesure de la boule de rayon n est finie (cf. la Proposition 3.1.17).

Soit maintenant (φ_n) une suite de Dirac dans $C_c^\infty(\mathbf{R}^d)$. D'après le Théorème 6.4.6, (1), on a $g \star \varphi_n \to g$ dans $L^p(\mathbf{R}^d)$; d'après la Proposition 6.3.1, $g \star \varphi_n \in C^\infty(\mathbf{R}^d)$; et enfin, d'après (6.3), $g \star \varphi_n$ est à support compact : en définitive, donc, $g \star \varphi_n \in C_c^\infty(\mathbf{R}^d)$, et $g \star \varphi_n \to g$ dans L^p.

Par conséquent pour n assez grand on a

$$\|f - g \star \varphi_n\|_p \leqslant 2\varepsilon,$$

d'où le résultat. $\square$

Il ne reste qu'à montrer le Lemme 6.4.8.

Preuve du Lemme. Soit f la fonction définie sur $\mathbf{R}$ par

$$f(x) = \begin{cases} e^{-1/x} & \text{pour } x \geqslant 0 \\ 0 & \text{pour } x < 0. \end{cases}$$

Clairement, f est C^∞ sur $\mathbf{R} - \{0\}$. On vérifie par récurrence que

$$f^{(k)}(x) = P_k(x^{-1})f(x)$$

pour $k \geqslant 0$ et $x > 0$, avec P_k un polynôme réel de degré k. Il en découle que $f^{(k)}(x) \to 0$ quand $x \to 0$ pour tout k, et par récurrence on vérifie alors que f est C^∞ sur $\mathbf{R}$. Bien entendu, $f \geqslant 0$ et $f \neq 0$.

Maintenant, on peut poser

$$\psi(t) = f(1 - \|t\|^2),$$

pour $t \in \mathbf{R}^d$: c'est une fonction positive, non identiquement nulle, et comme $1-\|t\|^2 = 1-\sum |t_i|^2$ est C^∞, la fonction ψ est C^∞ par composition. Le support de ψ est la boule centrée en 0 de rayon 1, de sorte que ψ convient. $\square$

CHAPITRE 7

La transformée de Fourier

Nous avons déjà vu la définition et certaines propriétés simples de la transformée de Fourier d'une fonction $f \in L^1(\mathbf{R})$ dans la Section 3.4, et démontré l'important théorème de Riemann–Lebesgue (Théorème 5.5.3) selon lequel $\hat{f} \in C_0(\mathbf{R})$, c'est à dire tend vers 0 à l'infini.

Dans ce chapitre on reprendra la théorie dans le cadre des fonctions définies sur $\mathbf{R}^d$, et on ira au-delà avec en particulier la formule d'inversion et l'extension de la transformée de Fourier aux fonctions L^2.

7.1. Rappels et extension à $\mathbf{R}^d$

Soit $d \geqslant 1$ un entier. On notera

$$<x, y> = \sum_i x_i \bar{y}_i$$

le produit scalaire usuel sur $\mathbf{C}^d$, et sa restriction à $\mathbf{R}^d$. On a donc $\|x\| = (<x, x>)^{1/2}$.

Dans ce chapitre, on notera dx (ou dt, dy...) la mesure de Lebesgue sur $\mathbf{R}^d$.

Définition 7.1.1. Soit $f \in L^1(\mathbf{R}^d)$. Sa transformée de Fourier est la fonction définie par

$$\hat{f}(t) = \int_{\mathbf{R}^d} f(x) e(-<x, t>) dx. \tag{7.1}$$

Le lemme suivant reprend les propriétés de régularité dans ce cadre un peu plus général.

Lemme 7.1.2. *L'opérateur linéaire*

$$\mathcal{F} \quad \begin{cases} L^1(\mathbf{R}^d) \to L^\infty(\mathbf{R}^d) \\ f \mapsto \hat{f} = \mathcal{F}(f) \end{cases}$$

est continu, de norme $\leqslant 1$, *et son image est incluse dans* $C_0(\mathbf{R}^d)$, *où*

$$C_0(\mathbf{R}^d) = \{f : \mathbf{R}^d \to \mathbf{C} \mid \lim_{\|t\| \to +\infty} f(t) = 0\}.$$

Démonstration. La continuité et l'inégalité $\|\mathcal{F}(f)\| \leqslant 1$ sont évidentes. L'analogue du lemme de Riemann–Lebesgue se démontre comme (par exemple) avec la première preuve qui en a été donnée, puisque la Proposition 5.5.1 est valide dans $\mathbf{R}^d$:

$$\hat{f}(t) = \frac{1}{2} \int_{\mathbf{R}} \Big(f(x) - f\Big(x - \frac{1}{2t}\Big)\Big) e(-<x, t>) dx$$

où $1/(2t)$ est le vecteur $\frac{1}{2}(1/t_1, \ldots, 1/t_d) \in \mathbf{R}^d$, lequel tend vers 0 quand $\|t\| \to +\infty$, donc

$$|\hat{f}(t)| \leqslant \int_{\mathbf{R}} \Big| f(x) - f\Big(x + \frac{1}{2t}\Big)\Big| dx \to 0$$

quand $\|t\| \to +\infty$. □

7.2. Propriétés formelles

Nous considérons ici une série de propriétés du type « si on change f d'une certaine manière, $\hat{f}$ est changée d'une autre manière. » On va voir apparaître une symétrie certaine dans les énoncés...

LEMME 7.2.1. *Soit* $f \in L^1(\mathbf{R}^d)$.

(1) *Soit* $h \in \mathbf{R}^d$ *et* $g(x) = f(x+h)$. *Alors* $g \in L^1(\mathbf{R}^d)$ *et on a*

$$\hat{g}(t) = e(<h,t>)\hat{f}(t) \text{ pour tout } t \in \mathbf{R}^d. \tag{7.2}$$

(2) *Soit* $h \in \mathbf{R}^d$ *et* $g(x) = e(<h,x>)f(x)$. *Alors* $g \in L^1(\mathbf{R}^d)$ *et on a*

$$\hat{g}(t) = \hat{f}(t-h) \text{ pour tout } t \in \mathbf{R}^d. \tag{7.3}$$

(3) *Soit* $T \in GL(d, \mathbf{R})$, *et* $g(x) = f(T(x))$. *Alors* $g \in L^1(\mathbf{R}^d)$ *et*

$$\hat{g}(t) = \frac{1}{|\det(T)|}\hat{f}(T^{-1}(t)) \text{ pour tout } t \in \mathbf{R}^d, \tag{7.4}$$

en particulier si $g(x) = f(\alpha x)$ *avec* $\alpha \in \mathbf{C}^\times$, *on a* $\hat{g}(t) = |\alpha|^{-d}\hat{f}(\alpha^{-1}t)$.

DÉMONSTRATION. Il s'agit seulement de changements de variable très simples : par exemple pour (1)

$$\int_{\mathbf{R}^d} f(x+h)e(-<x,t>)dx = \int_{\mathbf{R}^d} f(u)e(-<u-h,t>)du = e(<h,t>)\hat{f}(t),$$

pour (2)

$$\int_{\mathbf{R}^d} e(<h,x>)f(x)e(-<x,t>)dx = \hat{f}(t-h),$$

et pour (3) on a (Théorème 4.4.7, (2))

$$\int_{\mathbf{R}^d} f(T(x))e(-<x,t>)dx = \int_{\mathbf{R}^d} f(y)e(-<T^{-1}(y),t>)|\det(T)|^{-1}dy.$$

□

Le second lemme est le lien entre transformée de Fourier et différentiation (cf. le « prototype » dans la Proposition 3.4.3, (2) et (3)). Rappelons les notations (6.9) et (6.10) pour les multi-indices $\alpha = (\alpha_1, \ldots, \alpha_d)$ et les dérivées partielles, auxquelles on ajoute la notation

$$x^\alpha = x_1^{\alpha_1} \cdots x_2^{\alpha_d} \text{ pour } x = (x_1, \ldots, x_d) \in \mathbf{R}^d, \tag{7.5}$$

(on peut donc dire, par exemple, que les monômes X^α forment une base de l'espace $\mathbf{C}[X_1, \ldots, X_d]$ des polynômes en d variables). Tout les multi-indices sont supposés vérifier $\alpha_i \geqslant 0$ pour tout i.

LEMME 7.2.2. *Soit* $k \geqslant 0$ *un entier.*

(1) *Soit* $f \in C^k(\mathbf{R}^d)$ *telle que pour tout* α *satisfaisant* $|\alpha| \leqslant k$, *on a* $\partial_\alpha f \in L^1(\mathbf{R}^d)$, *et pour* $|\alpha| < k$, *on a* $\partial_\alpha f \in C_0(\mathbf{R}^d)$. *Alors pour tout* α *tel que* $|\alpha| \leqslant k$, *on a*

$$\mathcal{F}(\partial_\alpha f)(t) = (2i\pi t)^\alpha \hat{f}(t). \tag{7.6}$$

(2) *Soit* $f \in L^1(\mathbf{R}^d)$ *telle que pour tout* α *vérifiant* $|\alpha| \leqslant k$ *on a* $x^\alpha f(x) \in L^1(\mathbf{R}^d)$. *Alors* $\hat{f}$ *est de classe* C^k *et pour tout* α *tel que* $|\alpha| \leqslant k$, *on a*

$$\partial_\alpha \hat{f} = (-2i\pi)^{|\alpha|}\mathcal{F}(x^\alpha f).$$

DÉMONSTRATION. Dans les deux cas, il est clair qu'il suffit de traiter le cas $k = 1$ pour ensuite en déduire le cas général par récurrence, puisque les hypothèses se propagent. Cela signifie qu'il suffit de considérer les dérivées partielles par rapport à chaque variable, c'est à dire α a une seule composante $= 1$ et les autres nulles.

Le point (2) peut alors se ramener à $d = 1$ par définition des dérivées partielles, et correspond à la Proposition 3.4.3.

Pour le point (1), supposons d'abord $d = 1$. Alors

$$\mathcal{F}(f')(t) = \int_{\mathbf{R}} f'(x)e(-xt)dx = \Big[f(x)e(-xt)\Big]_{-\infty}^{+\infty} + 2i\pi \int_{\mathbf{R}} xf(x)e(-xt)dx$$

par intégration par partie ; le premier terme est nul par hypothèse ($f(x) \to 0$ quand $x \to \pm\infty$), et le second donne la formule dans ce cas particulier. Pour $d > 1$, on intègre par partie par rapport à la variable correspondant à α. $\square$

Remarque 7.2.3. Un résumé de ce lemme est le suivant : la transformée de Fourier « échange » les deux types de régularité d'une fonction, la régularité au sens de dérivabilité, et celle au sens d'une décroissance plus ou moins rapide « à l'infini », mesurée par des puissances inverses de $\|x\|$. En effet, notons le lemme élémentaire suivant :

Lemme 7.2.4. *Soit $k \geqslant 0$ un entier. Alors on a $x^\alpha f(x) \in L^1(\mathbf{R}^d)$ pour tout multi-indice α tel que $|\alpha| \leqslant k$ si et seulement si $g(x) = (1 + \|x\|)^k f(x)$ est intégrable. De même $x^\alpha f(x)$ est bornée, ou tend vers 0 à l'infini, pour tout α tel que $|\alpha| \leqslant k$ si et seulement si $(1 + \|x\|)^k f(x)$ est bornée, ou tend vers 0 à l'infini.*

Démonstration. Le calcul est trivial dans le cas $d = 1$ et devient juste un peu technique pour $d > 1$; on laisse au lecteur le soin de traiter le premier cas pour s'assurer de cette simplicité. $\square$

Alors (1) dit que si f est C^k (et vérifie les hypothèses, par exemple si f est à support compact), alors $(2i\pi t)^\alpha \hat{f}(t)$ est une transformée de Fourier, donc tend vers 0 à l'infini. Cela donne un taux de décroissance explicite de $\hat{f}$, en utilisant le lemme :

$$\hat{f}(t) = o((1 + \|t\|)^{-k})$$

quand $\|t\| \to +\infty$ (rappellons que cette notation signifie exactement que $\lim(1+\|t\|)^k \hat{f}(t) = 0$.)

Le dernier lemme suggère l'introduction de l'espace dit des fonctions de Schwartz.

Définition 7.2.5. L'espace de Schwartz $\mathcal{S}(\mathbf{R}^d)$ est l'espace vectoriel des fonctions C^∞ sur $\mathbf{R}^d$ telles que pour tout entier $k \geqslant 0$ et tout multi-indice α on a

$$\lim_{\|x\| \to +\infty} (1 + \|x\|)^k \partial_\alpha f(x) = 0.$$

La propriété essentielle est qu'il s'agit d'un « grand » espace de fonctions très régulières, stable à la fois par dérivation d'ordre arbitraire et par transformation de Fourier. Intuitivement on parle de « fonctions tendant vers 0 à l'infini plus vite que tout polynôme ainsi que toutes leurs dérivées. »

Proposition 7.2.6. (1) *On a $\mathcal{S}(\mathbf{R}^d) \subset L^p(\mathbf{R}^d)$ pour tout $p \geqslant 1$, et $\mathcal{S}(\mathbf{R}^d)$ est dense dans $L^p(\mathbf{R}^d)$ si $1 \leqslant p < +\infty$.*

(2) *Pour tout multi-indice α, la dérivée partielle d'ordre α correspondante est une application linéaire*

$$\partial_\alpha : \mathcal{S}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d).$$

(3) *La transformée de Fourier induit une application linéaire*

$$\mathcal{F} : \mathcal{S}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d).$$

Démonstration. Tout cela est fort simple. Pour (1), on peut appliquer la Proposition 4.4.1, (2) puisque $x \mapsto (1 + \|x\|)^A f(x)$ est bornée sur $\mathbf{R}^d$ pour tout $A > 0$. La densité de $\mathcal{S}(\mathbf{R}^d)$ dans L^p peut se remarquer à partir du fait que $C_c^\infty(\mathbf{R}^d) \subset \mathcal{S}(\mathbf{R}^d)$ et que (Corollaire 6.4.7) $C_c^\infty(\mathbf{R}^d)$ est dense dans $L^p(\mathbf{R}^d)$.

Le point (2) est évident, et (3) provient du Lemme 7.2.2. $\square$

EXEMPLE 7.2.7. On a déjà fait remarquer que $C_c^\infty(\mathbf{R}^d) \subset \mathcal{S}(\mathbf{R}^d)$, mais il existe de nombreuses fonctions dans $\mathcal{S}(\mathbf{R}^d)$ qui ne sont pas à support compact. Un exemple crucial est la fonction gaussienne

$$f(x) = e^{-\pi\|x\|^2}.$$

En effet, f est certainement C^∞ sur $\mathbf{R}^d$, et on vérifie facilement par récurrence que

$$\partial_\alpha f(x) = P_\alpha(x)f(x)$$

où $P_\alpha \in \mathbf{C}[x_1, \ldots, x_d]$ est un polynôme. La décroissance

$$(1 + \|x\|^k)|\partial_\alpha f(x)| \to 0 \text{ quand } \|x\| \to +\infty,$$

est alors immédiate.

La dernière propriété formelle concerne la convolution : c'est l'une des propriétés clé de la transformation de Fourier.

LEMME 7.2.8. *Soient f, $g \in L^1(\mathbf{R}^d)$. Alors*

$$\mathcal{F}(f \star g)(t) = \hat{f}(t)\hat{g}(t)$$

pour tout $t \in \mathbf{R}^d$.

Autrement dit, la transformée de Fourier échange le produit de convolution défini sur $L^1(\mathbf{R}^d)$ et le produit « point par point » habituel des fonctions.

DÉMONSTRATION. D'après la Remarque 6.2.3, la fonction $f \star g$ est dans $L^1(\mathbf{R}^d)$ et on peut donc calculer sa transformée de Fourier. Pour presque tout $x \in \mathbf{R}^d$ on a

$$f \star g(x) = \int_{\mathbf{R}^d} f(y)g(x-y)dy.$$

Multipliant par $e(-<x,t>)$ et intégrant sur $x \in \mathbf{R}^d$, on obtient

$$\begin{aligned}
\mathcal{F}(f \star g)(t) &= \int_{\mathbf{R}^d} \Big(\int_{\mathbf{R}^d} f(y)g(x-y)dy\Big) e(-<x,t>)dx \\
&= \int_{\mathbf{R}^d}\int_{\mathbf{R}^d} f(y)g(x-y)e(-<x,t>)dydx \\
&= \int_{\mathbf{R}^d} f(y)\int_{\mathbf{R}^d} g(x-y)e(-<x,t>)dxdy \\
&= \int_{\mathbf{R}^d} f(y)e(-<y,t>)dy \int_{\mathbf{R}^d} g(x)e(-<x,t>)dx
\end{aligned}$$

par changement de variable, le théorème de Fubini étant applicable puisque

$$\int_{\mathbf{R}^d} |f(y)g(x-y)e(-<x,t>)|dy = \int_{\mathbf{R}^d} |f(y)g(x-y)|dy$$

est intégrable en x d'après (6.5). □

7.3. La formule d'inversion de Fourier

Le problème considéré dans cette section est *l'inversion* de la transformée de Fourier, c'est à dire la question de retrouver f à partir de $\mathcal{F}(f)$. Ce problème est très important dans les applications : d'une part, il comprend le problème de savoir si $\hat{f}$ détermine de façon unique f (c'est à dire de savoir si $\mathcal{F} : L^1(\mathbf{R}^d) \to C_0(\mathbf{R}^d)$ est injective) ; et si c'est le cas, comment obtenir effectivement f à partir de $\hat{f}$?

Nous commençons par un exemple simple illustrant l'intérêt d'une telle solution. Considérons g une fonction donnée sur $\mathbf{R}^d$. Soit Δ le « laplacien » dans $\mathbf{R}^d$, c'est à dire

$$\Delta f = \sum_{j=1}^{d} \frac{\partial^2 f}{\partial x_j^2}$$

pour f de classe C^2. Supposons que l'on veuille résoudre l'équation aux dérivées partielles d'inconnue f

$$\Delta f - f = g. \tag{7.7}$$

Il s'agit là en apparence d'un problème complexe si $d > 1$ (si $d = 1$, c'est une équation différentielle ordinaire). La transformation de Fourier, jointe aux propriétés formelles de la section précédente, fournit une réponse formelle élégante : si (7.7) est vérifiée, et si cela est autorisé, calculons la transformée de Fourier de chaque côté. Par linéarité et à l'aide de (7.6) on trouve pour tout $t = (t_j)$

$$\sum_{j=1}^{d} (2i\pi t_j)^2 \mathcal{F}(f)(t) - \mathcal{F}(f)(t) = \mathcal{F}(g)(t),$$

c'est à dire

$$-(1 + 4\pi^2 \|t\|^2)\hat{f}(t) = \hat{g}(t).$$

L'équation (7.6) est devenue triviale à résoudre en termes de $\hat{f}$: on a (formellement)

$$\hat{f}(t) = -\frac{\hat{g}(t)}{1 + 4\pi^2\|t\|^2}.$$

On voit donc l'intérêt de pouvoir alors, si possible, re-calculer f à partir de $\hat{f}$...

Une analyse des propriétés formelles de la transformation de Fourier, ou bien l'interprétation « physique » (décomposition en fréquences...), voire le cas des séries de Fourier qui est plus classique (cf. le chapitre suivant) suggèrent la formule

$$f(x) = \int_{\mathbf{R}^d} \hat{f}(t) e(<x, t>) dt$$

qui ne diffère de $\mathcal{F}(\hat{f})$ que par un changement de signe. Il est évident qu'une telle formule ne peut exister que sous certaines hypothèses restrictives : en effet, le membre de droite requière pour être défini que $\hat{f} \in L^1(\mathbf{R}^d)$, ce qui est une condition assez forte (et pas du toute évidente à vérifier pour f donnée) ; et de plus, si c'est le cas, le terme de droite est clairement une fonction continue tendant vers 0, ce qui impliquerait que f en soit une...

On notera dorénavant $\bar{\mathcal{F}}$ l'application linéaire $L^1(\mathbf{R}^d) \to C_0(\mathbf{R}^d)$ telle que

$$\bar{\mathcal{F}}(f)(x) = \int_{\mathbf{R}^d} f(t) e(<x, t>) dt$$

qui vérifie

$$\bar{\mathcal{F}}(f)(x) = \mathcal{F}(f)(-x) = \overline{\mathcal{F}(\bar{f})(x)},$$

$\|\bar{\mathcal{F}}(g)\|_\infty \leqslant \|f\|_1$ et toutes les propriétés de la transformée de Fourier elle-même à des changements de signe près. (Noter que si f est paire alors $\bar{\mathcal{F}}(f) = \mathcal{F}(f)$.)

Théorème 7.3.1. *Soit $f \in L^1(\mathbf{R}^d)$ telle que $\hat{f} \in L^1(\mathbf{R}^d)$. Alors on a la formule d'inversion*

$$f(x) = \bar{\mathcal{F}}(\hat{f})(x) = \int_{\mathbf{R}^d} \hat{f}(t) e(<x, t>) dt \tag{7.8}$$

pour presque tout $x \in \mathbf{R}^d$ et en particulier f coïncide presque partout avec $\bar{\mathcal{F}}(\hat{f})$, une fonction dans $C_0(\mathbf{R}^d)$. Si $f \in C_0(\mathbf{R}^d) \cap L^1(\mathbf{R}^d)$, alors la formule d'inversion est valide pour tout $x \in \mathbf{R}^d$.

Remarque 7.3.2. (1) En termes plus abstraits, on peut dire que $\bar{\mathcal{F}}$ est l'inverse de l'application linéaire $\mathcal{F}$, sur le sous-espace $L^1(\mathbf{R}^d) \cap C_0(\mathbf{R}^d)$. Pour avoir un énoncé plus symétrique, on peut se restreindre à $\mathcal{S}(\mathbf{R}^d)$, et alors $\bar{\mathcal{F}}$ est une application

$$\bar{\mathcal{F}} : \mathcal{S}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d)$$

telle que

$$\bar{\mathcal{F}} = \mathcal{F}^{-1} \text{ c'est à dire } \bar{\mathcal{F}}(\mathcal{F}(f)) = f \text{ et } \mathcal{F}(\bar{\mathcal{F}}(f)) = f \text{ pour } f \in \mathcal{S}(\mathbf{R}^d).$$

(2) Si la normalisation choisie pour définir la transformée de Fourier est différente, la formule d'inversion de Fourier doit être légèrement modifiée. Par exemple, si on utilise

$$\hat{f}(t) = \int_{\mathbf{R}^d} f(x)e^{-i<x,t>}dx$$

alors

$$f(x) = \frac{1}{(2\pi)^d}\int_{\mathbf{R}^d} \hat{f}(t)e^{i<x,t>}dt,$$

et si on utilise

$$\hat{f}(t) = \frac{1}{(2\pi)^{d/2}}\int_{\mathbf{R}^d} f(x)e^{-i<x,t>}dx$$

alors

$$f(x) = \frac{1}{(2\pi)^{d/2}}\int_{\mathbf{R}^d} \hat{f}(t)e^{i<x,t>}dt.$$

La première idée est d'analyser $\bar{\mathcal{F}}(\hat{f})$ en utilisant la définition de la transformée de Fourier et d'échanger l'ordre d'intégration dans l'intégrale double qui apparaît : formellement

$$\begin{aligned}
\bar{\mathcal{F}}(\hat{f})(x) &= \int_{\mathbf{R}^d} \hat{f}(t)e(<x,t>)dt \\
&= \int_{\mathbf{R}^d} \Big(\int_{\mathbf{R}^d} f(y)e(-<y,t>)dy\Big)e(<x,t>)dt \\
&= \int_{\mathbf{R}^d} f(y)K(x-y)dy = f \star K(y)
\end{aligned}$$

avec

$$K(y) = \int_{\mathbf{R}^d} e(<y,t>)dt.$$

Ce calcul n'est cependant *pas justifié* car on voit aussitôt que K « n'existe pas » : la fonction $t \mapsto e(<y,t>)$, étant de module constant 1, n'est pas intégrable sur $\mathbf{R}^d$. D'ailleurs, le résultat du calcul montre que la formule d'inversion serait essentiellement équivalente à dire que K est une unité pour le produit de convolution, qui n'existe pas.

Mais cela suggère l'idée suivante : utiliser « à la place de K » une suite de Dirac K_n (cf. Définition 6.4.3) ; si cela est possible, on devrait avoir $f \star K_n \to f$ (cf. Théorème 6.4.6) et on peut s'attendre, puisque la transformation de Fourier échange produit ordinaire et de convolution, à ce que l'autre terme tende vers $\bar{\mathcal{F}}(\hat{f})$.

Précisément, soit (ψ_n) une suite de fonctions fixée, telle que $\psi_n \in \mathcal{S}(\mathbf{R}^d)$. On va prendre pour ψ_n une suite convergeant vers 1 et on s'attend à ce que la transformée de Fourier (conjuguée) $\bar{\mathcal{F}}(\psi_n)$ se comporte comme une suite de Dirac.

Considérons alors $g_n = \hat{f}\psi_n$ et $f_n = \bar{\mathcal{F}}(g_n)$ qui est bien définie car $\hat{f} \in L^1(\mathbf{R}^d)$ et ψ_n est bornée. On a alors le calcul suivant

$$\begin{aligned}
f_n(x) &= \int_{\mathbf{R}^d} g_n(t)e(<x,t>)dt \\
&= \int_{\mathbf{R}^d} \hat{f}(t)\psi_n(t)e(<x,t>)dt \\
&= \int_{\mathbf{R}^d} \Big(\int_{\mathbf{R}^d} f(y)e(-<y,t>)dy\Big)\psi_n(t)e(<x,t>)dt \\
&= \int_{\mathbf{R}^d} f(y)\bar{\mathcal{F}}(\psi_n)(x-y)dy = f \star \bar{\mathcal{F}}(\psi_n)(y)
\end{aligned}$$

qui est maintenant correct d'après le théorème de Fubini car

$$\int_{\mathbf{R}^d} |f(y)\psi_n(t)e(<x-y,t>)|dt \leqslant |f(y)|\|\psi_n\|_1$$

qui est intégrable.

Il ne reste alors qu'à montrer le lemme suivant :

Lemme 7.3.3. *Il existe une suite de fonctions* (ψ_n) *telles que* ψ_n *est une fonction de Schwartz pour tout* n *et*

(1) *On a* $\psi_n(x) \to 1$ *pour tout* x, $\|\psi_n\|_\infty \leqslant 1$*;*

(2) *La suite* $(\bar{\mathcal{F}}(\psi_n))$ *est une suite de Dirac.*

Admettons cela. On a alors $g_n(t) = \hat{f}(t)\psi_n(t) \to \hat{f}(t)$ pour tout $t \in \mathbf{R}^d$ et cette convergence est dominée par $|g_n(t)| \leqslant |\hat{f}(t)|$ qui est supposée intégrable. D'après le théorème de convergence dominée on a donc $g_n \to \hat{f}$ dans L^1, et puisque $\bar{\mathcal{F}}$ est continue, cela donne $f_n \to \bar{\mathcal{F}}(\hat{f})$ dans L^∞, c'est à dire uniformément.

Similairement, d'après le théorème d'approximation par convolution, on a $f_n = f \star \bar{\mathcal{F}}(\psi_n) \to f$ dans L^1. Il existe donc une sous-suite f_{n_k} telle que $f_{n_k}(x) \to f(x)$ presque partout quand $k \to +\infty$ (cf. Proposition 3.1.4, (2)), et en comparant, on voit que pour presque tout $x \in \mathbf{R}^d$ on a

$$f(x) = \lim_{k\to+\infty} f_{n_k}(x) = \bar{\mathcal{F}}(\hat{f})(x).$$

C'est donc la formule d'inversion de Fourier, valide presque partout. Si f est continue, comme on sait que $\bar{\mathcal{F}}(\hat{f})$ est aussi continue, ces deux fonctions continues coïncidant presque partout doivent être égales.

Il ne reste qu'à montrer le lemme. Noter que si la formule d'inversion est vraie pour une suite de Dirac donnée, on pourrait poser $\psi_n = \mathcal{F}(\varphi_n)$ où (φ_n) est une suite de Dirac. On s'est donc, en quelque sorte, ramené à vérifier (7.8) seulement pour une suite de fonctions.

On se réduit à un énoncé encore plus simple.

Lemme 7.3.4. *Il existe une fonction de Schwartz* ψ *sur* $\mathbf{R}^d$ *telle que* $\psi(0) = 1$, $\|\psi\|_\infty \leqslant 1$, *et* $\bar{\mathcal{F}}(\psi) \in \mathcal{S}(\mathbf{R}^d)$ *est une fonction positive vérifiant*

$$\int_{\mathbf{R}^d} \bar{\mathcal{F}}(\psi)(x)dx = 1. \tag{7.9}$$

Cette fonction étant donnée, posons en effet $\psi_n(t) = \psi(t/n)$. Il vient $\psi_n(t) \to \psi(0) = 1$ (par continuité) pour tout t, et

$$\varphi_n(x) = \bar{\mathcal{F}}(\psi_n)(x) = n^d \bar{\mathcal{F}}(\psi)(nx)$$

(cf. Lemme 7.2.1, (3), adapté à $\bar{\mathcal{F}}$). On a $\varphi_n \geqslant 0$ puisque $\bar{\mathcal{F}}(\psi) \geqslant 0$ par le lemme, et de plus

$$\int_{\mathbf{R}^d} \bar{\mathcal{F}}(\psi)(x)dx = 1$$

par (7.9). On est alors dans la situation de l'Exemple 6.4.5, (3), qui montre que (φ_n) est une suite de Dirac. Cela fournit la preuve du Lemme 7.3.3 et donc du Théorème 7.3.1.

Preuve du Lemme 7.3.4. Soit d'abord $d = 1$ et $\psi_1(x) = e^{-\pi x^2}$ la fonction gaussienne (cf. Proposition 4.4.10). On a $\psi_1 \in \mathcal{S}(\mathbf{R})$ (Exemple 7.2.7), $\psi_1(0) = 1$ et $\|\psi_1\|_\infty = 1$.

De plus, ψ_1 est solution de l'équation différentielle

$$y' + 2\pi xy = 0. \tag{7.10}$$

Prenant la transformée de Fourier de cette relation, on trouve (cf. Lemme 7.2.2, (1) et (2))

$$2i\pi t\mathcal{F}(\psi_1)(t) + 2\pi \frac{\mathcal{F}(\psi_1)'(t)}{-2i\pi} = 0 \text{ c'est à dire } 2i\pi t\mathcal{F}(\psi_1)(t) + i\mathcal{F}(\psi_1)'(t) = 0,$$

autrement dit la transformée de Fourier de ψ_1 est également solution de (7.10). Comme

$$\mathcal{F}(\psi_1)(0) = \int_{\mathbf{R}} \psi_1(t)dt = 1 = \psi_1(0) \text{ (Proposition 4.4.10)}$$

on en déduit que $\mathcal{F}(\psi_1) = \psi_1$. De plus ψ_1 est paire donc $\bar{\mathcal{F}}(\psi_1) = \mathcal{F}(\psi_1) = \psi_1$. Ainsi ψ_1 répond certainement aux conditions demandées quand $d = 1$.

Ensuite pour $d \geqslant 1$ un entier quelconque, il suffit de poser

$$\psi(x) = e^{-\|x\|^2} = \prod_{j=1}^{d} e^{-\pi x_j^2} = \prod_{j=1}^{d} \psi_1(x_j).$$

En effet $\psi \in \mathcal{S}(\mathbf{R}^d)$ et par Fubini et le cas $d = 1$ il vient aussitôt

$$\bar{\mathcal{F}}(\psi)(t) = \prod_{j=1}^{d} \bar{\mathcal{F}}(\psi_1)(t_j) = \psi(t),$$

donc ψ convient. □

Exemple 7.3.5. (1) Puisque $\mathcal{F}$ induit par restriction une application $\mathcal{S}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d)$ et $\mathcal{S}(\mathbf{R}^d) \subset L^1(\mathbf{R}^d)$ (cf. Proposition 7.2.6), la formule d'inversion s'applique pour toute $f \in \mathcal{S}(\mathbf{R}^d)$.

(2) En plus de son intérêt théorique, la formule d'inversion (7.8) permet des calculs d'intégrales « concrètes ». Par exemple, soit $f(x) = e^{-2\pi|x|}$ pour $x \in \mathbf{R}$. C'est une fonction intégrable (mais pas de classe C^1 : elle n'est pas dérivable en 0). Sa transformée de Fourier est aisée à calculer

$$\begin{aligned}
\mathcal{F}(f)(t) &= \int_{\mathbf{R}} e^{-2\pi(ixt+|x|)} dx \\
&= \int_0^{\infty} e^{-2\pi x(1+it)} dx + \int_0^{\infty} e^{-2\pi x(1-it)} dx \\
&= \Big[-\frac{e^{-2\pi x(1+it)}}{2\pi(1+it)} - \frac{e^{-2\pi x(1-it)}}{2\pi(1-it)} \Big]_0^{\infty} \\
&= \frac{1}{2\pi}\Big(\frac{1}{1-it} + \frac{1}{1+it} \Big) \\
&= \frac{1}{\pi}\frac{1}{1+t^2}.
\end{aligned}$$

Comme $\hat{f}$ est intégrable sur $\mathbf{R}$, on a donc

$$e^{-2\pi|x|} = \bar{\mathcal{F}}(\hat{f})(x) = \frac{1}{\pi}\int_{\mathbf{R}} \frac{e(xt)}{1+t^2} dt$$

pour $t \in \mathbf{R}$. Cette dernière formule n'est pas aussi simple à établir directement que la précédente.

On remarque bien sur cet exemple que $\hat{f}$ est C^{∞}, ce qui correspond au fait que f décroît plus vite que tout polynôme à l'infini, mais $\hat{f}$ ne décroît que comme t^{-2}, car f n'est pas C^{∞}.

7.4. La transformée de Fourier dans $L^2(\mathbf{R}^d)$

Comme $\mathbf{R}^d$ est de mesure infinie, l'espace $L^2(\mathbf{R}^d)$ n'est pas inclus dans $L^1(\mathbf{R}^d)$ (cf. la Remarque 3.1.16 par exemple) et donc, pour $f \in L^2(\mathbf{R}^d)$, l'intégrale (7.1) n'est pas définie en général.

Il s'avère cependant possible de définir une transformation de Fourier au sens L^2, et bien que la définition puisse paraître de prime abord assez compliquée, la théorie résultante est en fait à bien des égards plus agréable et plus symétrique que la transformée de Fourier des fonctions intégrables.

L'outil abstrait pour aboutir à cela est le résultat bien connu de prolongement d'applications linéaires définies sur un sous-espace dense d'un espace de Banach.[1]

Lemme 7.4.1. *Soit B un espace de Banach dont on note $\|\cdot\|$ la norme, $E \subset B$ un sous-espace* dense *de B. Soit T une application linéaire définie sur E à valeurs dans B, telle qu'il existe $C \geqslant 0$ vérifiant*

(7.11) $$\|T(v)\| \leqslant C\|v\| \text{ pour tout } v \in E.$$

[1]. Un cas particulier d'un énoncé beaucoup plus général, évidemment.

Alors il existe une unique application linéaire $T_B : B \to B$ telle que
(1) *Pour $v \in E$, on a $T_B(v) = T(v)$.*
(2) *Pour tout $v \in B$, on a $\|T_B(v)\| \leqslant C\|v\|$, autrement dit T_B est continue de norme $\leqslant C$.*

Rappel de la preuve. Puisque E est dense dans B, pour tout $v \in B$ il existe une suite (v_n), $v_n \in E$, telle que $v = \lim v_n$. On pose alors $T_B(v) = \lim T(v_n)$. L'existence de cette limite provient du fait que B est *complet* et de (7.11) : on a

$$\|T(v_n) - T(v_m)\| \leqslant C\|v_n - v_m\|$$

donc $(T(v_n))$ est de Cauchy dans B.

De plus la limite est indépendante du choix de la suite « approximante » (v_n) : si $w_n \to v$, on a

$$\|T(v_n - w_n)\| \leqslant C\|v_n - w_n\| \to 0.$$

Cette unicité implique que T_B prolonge T et que T_B est linéaire ; de plus (7.11) s'étend simplement à B par passage à la limite.

Enfin, l'unicité de T_B est claire, puisqu'elle doit être continue, on doit bien avoir $T_B(v) = \lim T_B(v_n) = \lim T(v_n)$ pour toute suite (v_n) à valeurs dans E qui converge vers v. $\square$

Nous allons appliquer ceci à l'espace de Hilbert $L^2(\mathbf{R}^d)$ et à la transformée de Fourier restreinte à un sous-espace. On pourrait choisir pour celui-ci simplement $E = L^2(\mathbf{R}^d) \cap L^1(\mathbf{R}^d)$, qui est dense dans $L^2(\mathbf{R}^d)$, mais il est instructif (quoique pas très important dans ce cas) de prendre plutôt $E = \mathcal{S}(\mathbf{R}^d)$. Cela illustre le fait que l'étude d'un espace de fonctions de type L^p peut se faire souvent en se restreignant à un sous-espace de fonctions extrêmement régulières, du moment que l'on garde un contrôle de la norme des opérations effectuées, c'est à dire de leur continuité.

Par la Proposition 7.2.6, (1), $\mathcal{S}(\mathbf{R}^d)$ est un sous-espace dense de $L^2(\mathbf{R}^d)$, et par loc. cit. (3), on a l'application linéaire « transformée de Fourier »

$$\mathcal{F} : \mathcal{S}(\mathbf{R}^d) \to \mathcal{S}(\mathbf{R}^d) \subset L^2(\mathbf{R}^d).$$

Nous allons alors montrer une formule du type (7.11). Plus précisément même :

Proposition 7.4.2. (1) *Soient f, $g \in \mathcal{S}(\mathbf{R}^d)$. On a alors*

$$\int_{\mathbf{R}^d} f(x)\hat{g}(x)dx = \int_{\mathbf{R}^d} \hat{f}(x)g(x)dx, \tag{7.12}$$

ou bien de manière équivalente

$$<f, \hat{g}> = <\hat{f}, g>,$$

notant $<\cdot,\cdot>$ le produit scalaire dans $L^2(\mathbf{R}^d)$.

(2) *Pour toute $f \in \mathcal{S}(\mathbf{R}^d)$, on a*

$$\|\mathcal{F}(f)\|_2 = \|f\|_2 \text{ et } \|\bar{\mathcal{F}}(f)\|_2 = \|f\|_2 \tag{7.13}$$

Démonstration. Soient f et g des fonctions de Schwartz. On a alors

$$\begin{aligned}\int_{\mathbf{R}^d} f(x)\hat{g}(x)dx &= \int_{\mathbf{R}^d} f(x)\Big(\int_{\mathbf{R}^d} g(y)e(-<y,x>)dy\Big)dx \\ &= \int_{\mathbf{R}^d\times\mathbf{R}^d} f(x)g(y)e(-<x,y>)d(x,y)\end{aligned}$$

(où $d(x,y)$ est la mesure de Lebesgue sur $\mathbf{R}^d \times \mathbf{R}^d$), grâce au théorème de Fubini qui est trivialement applicable puisque f et g sont intégrables. Mais cette expression est manifestement symétrique en f et g, d'où la formule (7.12). L'autre formule de (1) est équivalente en conjuguant l'un des arguments.

Pour montrer (2) on applique ce qui précède à g telle que $\hat{g} = \bar{f}$ de sorte que

$$\int_{\mathbf{R}^d} f(x)\hat{g}(x)dx = \|f\|_2^2.$$

Cela signifie (par la formule d'inversion, cf. la Remarque 7.3.2) que $g = \bar{\mathcal{F}}(\hat{g}) = \bar{\mathcal{F}}(\bar{f}) = \overline{\mathcal{F}(f)}$ puisque le conjugué de $e(x)$ est $e(-x)$. Donc

$$\int_{\mathbf{R}^d} \hat{f}(x) g(x) dx = \int_{\mathbf{R}^d} |\hat{f}(x)|^2 dx = \|\hat{f}\|_2^2,$$

d'où la première partie de (7.13) en comparant, la seconde s'en déduisant en l'appliquant à $g(x) = f(-x)$. □

Corollaire 7.4.3. *Il existe une unique application linéaire continue*

$$\mathcal{F}_2 : L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)$$

telle que $\mathcal{F}_2(f) = \mathcal{F}(f)$ *si* $f \in \mathcal{S}(\mathbf{R}^d)$ *et de plus* $\mathcal{F}_2$ *est une* isométrie *de* $L^2(\mathbf{R}^d)$*, c'est à dire*

$$<\mathcal{F}_2(f), \mathcal{F}_2(g)> = <f, g> \text{ ou bien } \|\mathcal{F}_2(f)\|_2 = \|f\|_2$$

pour $f,\ g \in L^2(\mathbf{R}^d)$.

De même il existe une unique application linéaire continue

$$\bar{\mathcal{F}}_2 : L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)$$

telle que $\bar{\mathcal{F}}_2(f) = \bar{\mathcal{F}}(f)$ *si* $f \in \mathcal{S}(\mathbf{R}^d)$*, qui est une isométrie et est l'inverse* $\bar{\mathcal{F}}_2 = \mathcal{F}_2^{-1}$ *de la précédente.*

Démonstration. Il suffit d'appliquer le Lemme 7.4.1 (avec $C = 1$) pour obtenir l'existence et unicité de $\mathcal{F}_2$ et $\bar{\mathcal{F}}_2$ à partir de (7.13). Comme $\mathcal{F}(\bar{\mathcal{F}}(f)) = \bar{\mathcal{F}}(\mathcal{F}(f)) = f$ pour $f \in \mathcal{S}(\mathbf{R}^d)$, les mêmes relations restent vraies par continuité pour $f \in L^2(\mathbf{R}^d)$. □

La relation

$$\int_{\mathbf{R}^d} |\mathcal{F}_2(f)(t)|^2 dt = \int_{\mathbf{R}^d} |f(t)|^2 dt \tag{7.14}$$

est appelée souvent la *formule de Plancherel* ou de Parseval.

Remarque 7.4.4. En continuation de la Remarque 7.3.2, (2), si la transformée de Fourier est normalisée différemment, il est possible qu'elle ne soit pas une isométrie sur L^2 : par exemple, avec la normalisation

$$\hat{f}(t) = \int_{\mathbf{R}^d} f(x) e^{-i<x,t>} dx$$

on trouve $\|\hat{f}\|_2 = (2\pi)^{d/2} \|f\|_2$.

Par continuité, la plupart des propriétés de la transformée de Fourier s'étendent à L^2. Par exemple, si $f \in L^2(\mathbf{R}^d)$ et $h \in \mathbf{R}^d$, la fonction $g(x) = f(x + h)$ est bien définie dans L^2 avec $\|g\|_2 = \|f\|_2$ (autrement dit, l'application linéaire $f \mapsto g$ est une isométrie), et à partir de (7.2) on trouve

$$\mathcal{F}_2(g)(t) = e(<h, t>) \mathcal{F}_2(f)(t).$$

De même, si $g(x) = e(<h, x>) f(x)$ on a $g \in L^2(\mathbf{R}^d)$ et $\|g\|_2 = \|f\|_2$ et

$$\mathcal{F}_2(g)(t) = \mathcal{F}_2(f)(t - h) \text{ (comparer avec (7.3)).}$$

La notation $\mathcal{F}_2$ est temporaire : elle est d'abord nécessaire pour faire remarquer que, si $f \in L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$, on peut définir $\mathcal{F}(f) \in C_0(\mathbf{R}^d)$ mais aussi $\mathcal{F}_2(f) \in L^2(\mathbf{R}^d)$. Il n'est pas entièrement évident qu'il existe une relation entre ces deux fonctions (par exemple, parce que les deux espaces d'arrivée sont différents). Le résultat suivant montre cependant qu'elles coïncident en un sens précis.

Proposition 7.4.5. *Soit* $f \in L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$*. On a alors*

$$\mathcal{F}_2(f)(x) = \mathcal{F}(f)(x) \text{ pour presque tout } x \in \mathbf{R}^d,$$

et en particulier $\hat{f}$ *appartient à* $L^2(\mathbf{R}^d)$*. De même*

$$\bar{\mathcal{F}}_2(f)(x) = \bar{\mathcal{F}}(f)(x) \text{ pour presque tout } x \in \mathbf{R}^d.$$

Démonstration. Par continuité (cf. la preuve du Lemme 7.4.1), on peut calculer

$$\mathcal{F}_2(f) = \lim_{n\to+\infty} \mathcal{F}(f_n)$$

où $f_n \in \mathcal{S}(\mathbf{R}^d)$ est une suite quelconque convergeant en norme L^2 vers f. On va considérer la suite

$$f_n = (f\chi_n) \star \varphi_n$$

où χ_n est la fonction caractéristique du cube $[-n,n]^d$ et (φ_n) est une suite de Dirac fixée, avec $\varphi_n \in C_c^\infty(\mathbf{R}^d)$. D'après les propriétés de la convolution, on a alors $f_n \in C_c^\infty(\mathbf{R}^d) \subset \mathcal{S}(\mathbf{R}^d)$ (cf. la preuve du Corollaire 6.4.7).

On a alors $f_n \to f$ en norme L^2 *et* en norme L^1 d'après le Théorème 6.4.6 : on écrit

$$\|f_n - f\|_a \leqslant \|(f\chi_n - f) \star \varphi_n\|_a + \|f \star \varphi_n - f\|_a,$$

(pour $a \in \{1,2\}$) et le second terme tend vers 0 par le théorème, tandis que le premier vérifie

$$\|(f\chi_n - f) \star \varphi_n\|_a \leqslant \|f\chi_n - f\|_a \|\varphi_n\|_1 = \|f\chi_n - f\|_a = \Bigl(\int_{\|x\|_\infty > n} |f(x)|^a dx\Bigr)^{1/a}$$

(cf. (6.5) et la définition d'une suite de Dirac). Cette dernière expression converge vers 0 puisque $|f|^a$ est intégrable et que les cubes $[-n,n]^d$ forment une suite croissante recouvrant $\mathbf{R}^d$ (Lemme 2.3.10).

La convergence de f_n vers f en norme L^1 implique que $\mathcal{F}(f_n) \to \mathcal{F}(f)$ en norme L^∞, c'est à dire uniformément (Lemme 7.1.2) ; et la convergence de f_n vers f en norme L^2 implique que $\mathcal{F}(f_n) = \mathcal{F}_2(f_n) \to \mathcal{F}_2(f)$ en norme L^2 (continuité de $\mathcal{F}_2$).

D'après la Proposition 3.1.4, (2), il existe une sous-suite (f_{n_k}) telle que $\mathcal{F}_2(f_{n_k})$ converge vers $\mathcal{F}_2(f)$ presque partout. Ainsi, pour presque tout $x \in \mathbf{R}^d$ on a donc

$$\mathcal{F}(f_{n_k})(x) \to \mathcal{F}(f)(x) \text{ (par convergence uniforme)}$$
$$\mathcal{F}_2(f_{n_k})(x) \to \mathcal{F}_2(f)(x).$$

Puisque $\mathcal{F}(f_n) = \mathcal{F}_2(f_n)$, on en déduit $\mathcal{F}(f)(x) = \mathcal{F}_2(f)(x)$ pour presque tout x comme désiré.

Le cas de la transformée de Fourier conjuguée s'en déduit comme d'habitude. ☐

À cause de cette proposition, on peut noter $\mathcal{F}_2(f) = \mathcal{F}(f)$ si $f \in L^2(\mathbf{R}^d)$ sans créer d'ambiguïté.

La Proposition 7.4.5 permet, en pratique, d'utiliser des suites de fonctions intégrables (plutôt que de Schwartz) pour calculer $\mathcal{F}(f)$ si $f \in L^2(\mathbf{R}^d)$. En voici un exemple typique.

Proposition 7.4.6. *Soit* $f \in L^2(\mathbf{R}^d)$*. Pour* $n \geqslant 1$*, soit*

$$\varphi_n(t) = \int_{[-n,n]^d} f(x) e(-<x,t>) dx = \mathcal{F}(f\chi_n)$$

$$\psi_n(x) = \int_{[-n,n]^d} \mathcal{F}(f)(t) e(<x,t>) dt = \bar{\mathcal{F}}(\mathcal{F}(f)\chi_n).$$

où χ_n *est la fonction caractéristique de* $[-n,n]^d$*.*

Alors on a

$$\lim_{n\to+\infty} \varphi_n = \mathcal{F}(f) \text{ dans } L^2(\mathbf{R}^d), \quad \lim_{n\to+\infty} \psi_n = f \text{ dans } L^2(\mathbf{R}^d),$$

Démonstration. Il suffit de remarquer que, le cube $[-n,n]^d$ étant de mesure finie, la restriction de f (resp. $\mathcal{F}(f)$) à celui-ci y est intégrable (Proposition 3.1.17) donc $f\chi_n \in L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$ (resp. $\mathcal{F}(f)\chi_n \in L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$), de sorte que φ_n et ψ_n sont des transformées de Fourier au sens L^1 et L^2 simultanément (et ces deux sens coïncident par la Proposition 7.4.5).

Comme $f\chi_n \to f$ (resp. $\mathcal{F}(f)\chi_n \to \mathcal{F}(f)$ en norme L^2, la conclusion provient simplement de la continuité de $\mathcal{F}$ (resp. $\bar{\mathcal{F}}$) sur $L^2(\mathbf{R}^d)$. ☐

CHAPITRE 8

Séries de Fourier

8.1. Fonctions périodiques

La théorie des séries de Fourier est l'analogue de celle des intégrales de Fourier pour des fonctions périodiques. Historiquement, elle a même plutôt précédé celle du chapitre précédent.

Nous allons étudier ici les fonctions 1-périodiques, c'est à dire telles que $f(x+1) = f(x)$. Comme pour la transformée de Fourier, d'autres choix de période sont parfois utilisés, très souvent celle-ci est fixée à 2π. L'adjectif « périodique » dans ce qui suit signifiera toujours périodique de période 1, sauf mention explicite du contraire.

Pour A un espace vectoriel fixé, notons $\mathrm{Per}(A)$ l'espace vectoriel des applications $f : \mathbf{R} \to A$ telles que $f(x+1) = f(x)$ pour tout $x \in \mathbf{R}$. Clairement, cette condition est équivalente à la condition $f(x+n) = f(x)$ pour tout $n \in \mathbf{Z}$ (récurrence sur n).

On peut « interpréter » de diverses manières l'espace $\mathrm{Per}(A)$, et chaque point de vue est utile.

(1) On peut voir $\mathrm{Per}(A)$ comme ci-dessus, un espace de fonctions définies sur $\mathbf{R}$. Si A un espace topologique, cela permet de définir les fonctions périodiques *continues* ; si $A = \mathbf{C}$, cela permet aussi de définir les fonctions périodiques dérivables, de classe C^k, etc...

(2) Une fonction périodique est connue à partir de sa restriction g à $I = [0,1]$, voire à tout intervalle « fondamental » $[\alpha, \alpha+1]$ pour $\alpha \in \mathbf{R}$ fixé, puisque pour tout $x \in \mathbf{R}$, il existe un $m \in \mathbf{Z}$ tel que $x - m \in [\alpha, \alpha+1]$ donc $f(x) = f(x-m) = g(x-m)$. On peut même se restreindre à l'intervalle $[\alpha, \alpha+1[$, et on a des isomorphismes

$$\mathrm{Per}(A) = \{g : [\alpha, \alpha+1[\to A\} = \{g : [\alpha, \alpha+1] \to A \mid g(\alpha) = g(\alpha+1)\}.$$

Dans cette description, les fonctions continues périodiques s'identifient au sous-espace C_p de $C([\alpha, \alpha+1])$ défini par

$$C_p = \{g \in C([\alpha, \alpha+1]) \mid g(\alpha) = g(\alpha+1)\},$$

et les fonctions de classe C^k s'identifient à

$$C_p^k = \{g \in C^k([\alpha, \alpha+1]) \mid g^{(j)}(\alpha) = g^{(j)}(\alpha+1) \text{ pour } 0 \leqslant j \leqslant k\},$$

en particulier cet espace est *distinct* de $C^k(\mathbf{R}) \cap C_p$: il faut que toutes les dérivées d'ordre $\leqslant k$ de g se « recollent » exactement entre α et $\alpha+1$. Par exemple, si

$$f(x) = 1 - |x| \text{ pour } -1/2 \leqslant x \leqslant 1/2$$

et f est prolongéee à $\mathbf{R}$ par périodicité (fonction « triangle »), alors f n'est pas dans C_p^1 bien que sa restriction à $[0,1]$ puisse sembler l'être à première vue...

Une fonction périodique est dite intégrable si sa restriction à un intervalle fondamental $[\alpha, \alpha+1]$ l'est. On définit ainsi les espaces L^p de fonctions périodiques

$$L^p(I) = \{f \in \mathrm{Per}(\mathbf{C}) \mid \int_0^1 |f(x)|^p dx < +\infty\},$$

en pensant à $I = [0,1[$ avec la mesure de Lebesgue. Comme $L^p(I)$ est isomorphe à $L^p([0,1[, dx)$, il vérifie toutes les propriétés habituelles (complétude, etc...)

Noter qu'il faut se restreindre à un intervalle car aucune fonction périodique non-nulle n'est intégrable sur $\mathbf{R}$. De plus le choix de l'intervalle fondamental $[\alpha, \alpha+1]$ n'a pas d'importance :

Lemme 8.1.1. (1) *Si* $f : \mathbf{R} \to \mathbf{C}$ *est* 1*-périodique et n'est pas nulle presque partout, alors* $f \notin L^1(\mathbf{R})$.

(2) *Si* $f \in L^1(I)$*, alors pour tout* $\alpha \in \mathbf{R}$ *on a*

$$\int_0^1 f(x)dx = \int_\alpha^{\alpha+1} f(x)dx. \tag{8.1}$$

Démonstration. Pour (1), on a

$$\int_{[-n,n]} |f(x)|dx = 2n \int_0^1 |f(x)|dx$$

pour f périodique et ceci tend vers $+\infty$ sauf si

$$\int_0^1 |f(x)|dx = 0,$$

condition qui implique que f est nulle presque partout, d'abord sur $[0,1]$ puis sur $\mathbf{R}$ par périodicité.

Pour (2), soit $m \in \mathbf{Z}$ tel que $m \leqslant \alpha < m+1$. On écrit

$$\begin{aligned}\int_\alpha^{\alpha+1} f(x)dx &= \int_\alpha^{m+1} f(x)dx + \int_{m+1}^{\alpha+1} f(x)dx \\ &= \int_{\alpha-m}^1 f(x+m)dx + \int_0^{\alpha-m} f(x+m+1)dx \\ &= \int_0^1 f(x)dx \text{ par périodicité.}\end{aligned}$$

□

La formule (8.1) est l'analogue pour les fonctions périodiques de l'invariance de la mesure de Lebesgue sur $\mathbf{R}^d$ par translation, dont on vu l'importance lors de l'étude de la convolution et de la transformée de Fourier.

(3) Soit K le cercle unité dans le plan complexe. Alors $\mathrm{Per}(A)$ est isomorphe à l'espace des applications $g : K \to A$ par la correspondance

$$g(z) = f(\theta) \tag{8.2}$$

pour $\theta \in \mathbf{R}$, $z \in K$ tels que $z = e(i\theta) = e^{2i\pi\theta}$. Comme θ est bien défini à un entier $n \in \mathbf{Z}$ près, (8.2) définit bien une fonction sur K à partir d'une fonction périodique et vice-versa.

Pour ce point de vue, l'espace C_p des fonction périodiques continues correspond à $C(K)$, l'espace des fonctions continues sur K.

(4) La définition la plus élégante est de considérer l'espace topologique quotient $X = \mathbf{R}/\mathbf{Z}$ de $\mathbf{R}$ par son sous-groupe $\mathbf{Z}$: alors les fonctions périodiques f s'identifient aux fonctions g sur X par

$$g(p(x)) = f(x) \text{ où } p : \mathbf{R} \to X \text{ est la projection}$$

et il existe sur X une mesure borélienne naturelle μ permettant de définir $L^p(X) \simeq L^p(I)$ tel que

$$\int_X g(y)d\mu(y) = \int_0^1 f(x)dx$$

si f correspond à g. De plus X est un groupe et on a

$$\int_X g(y)d\mu(y) = \int_X g(y+\alpha)d\mu(y)$$

pour tout $\alpha \in X$ fixé, ce qui correspond à (8.1).

Nous utiliserons surtout les deux premiers points de vue, qui sont les plus concrets. Remarquer cependant la différence entre la façon de définir une fonction périodique vérifiant une propriété telle que la continuité, et la définition d'une fonction périodique intégrable.

8.2. Séries de Fourier et séries trigonométriques

Les fonctions 1-périodiques « fondamentales » sont les exponentielles $e_n : \mathbf{R} \to \mathbf{C}$, où $n \in \mathbf{Z}$, données par

$$e_n(x) = e(nx).$$

En effet, puisque $e(t) = e^{2i\pi t} = 1$ si et seulement si $t \in \mathbf{Z}$, on a $e_n(x+1) = e(nx+n) = e(nx)$, et ce sont les seules fonctions du type $x \mapsto e(ax)$ qui sont périodiques.

Les fonctions e_n sont bornées, sont dans $L^p(I)$ pour tout p, et sont de classe C^∞.

Définition 8.2.1. Soit $f \in L^1(I)$ une fonction périodique intégrable et $n \in \mathbf{Z}$. On pose

$$c_n(f) = \int_0^1 f(x)e_n(-x)dx = \int_0^1 f(x)e(-nx)dx \tag{8.3}$$

et on dit que $c_n(f)$ est le n-ème *coefficient de Fourier* de f.

Puisque $|f(x)e(-nx)| = |f(x)|$, les coefficients $c_n(f) \in \mathbf{C}$ sont effectivement définis pour tout $n \in \mathbf{Z}$.

Une distinction importante avec le cas de la transformation de Fourier est que, puisque l'intervalle fondamental $I = [0, 1[$ est de mesure finie, on a

$$L^\infty(I) \subset L^p(I) \subset L^1(I) \text{ pour tout } p,\ 1 \leqslant p \leqslant +\infty,$$

et en particulier les coefficients de Fourier sont définis pour toute fonction dans L^2 (ou dans L^∞) sans qu'il soit nécessaire de faire appel à aucune définition spéciale.

Voici les propriétés les plus simples de ces coefficients (comparer avec la Section 7.1).

Proposition 8.2.2. (1) *Pour tout $n \in \mathbf{Z}$, et tout p, $1 \leqslant p \leqslant +\infty$, l'application $f \mapsto c_n(f)$ est une forme linéaire continue*

$$c_n : L^p(I) \to \mathbf{C}$$

de norme $\|c_n\| \leqslant 1$.

(2) (Riemann-Lebesgue) *Pour toute $f \in L^1(I)$, on a*

$$\lim_{|n|\to+\infty} c_n(f) = 0.$$

(3) *Si $f \in L^1(I)$, $\alpha \in \mathbf{R}$ et $g(x) = f(x + \alpha)$, alors $g \in L^1(I)$ et*

$$c_n(g) = e(n\alpha)c_n(f).$$

(4) *Si $f \in L^1(I)$, $m \in \mathbf{Z}$ et $g(x) = e(mx)f(x)$, alors $g \in L^1(I)$ et*

$$c_n(g) = c_{n-m}(f).$$

(5) *Si $f \in C^k(\mathbf{R})$ est périodique, alors pour tout j, $0 \leqslant j \leqslant k$, on a $f^{(j)} \in L^1(I)$ et pour tout $n \in \mathbf{Z}$*

$$c_n(f^{(j)}) = (2i\pi n)^j c_n(f).$$

Les démonstrations sont toutes très simples et similaires à celles des propriétés correspondantes de la transformée de Fourier : il serait un excellente exercice d'essayer de les écrire directement sans regarder ce qui suit (d'autant que la preuve est en général plus simple, cf. par exemple les énoncés portant sur la dérivation).

Démonstration. (1) La linéarité de c_n est évidente, et on a par l'inégalité de Hölder

$$|c_n(f)| \leqslant \Big(\int_0^1 dx\Big)^{1/q} \|fe_n\|_p = \|f\|_p.$$

(2) Il est possible d'adapter sans grande difficulté les preuves déjà vues pour la transformée de Fourier. Écrivons par exemple (comparer avec la première preuve du Théorème 5.5.3)

$$c_n(f) = \int_0^1 f(x)e(-nx)dx = -\int_0^1 f(x)e\Big(n\Big(x + \frac{1}{2n}\Big)\Big)dx$$

d'où

$$|c_n(f)| \leqslant \frac{1}{2}\int_0^1 \Big|f(x) - f\Big(x + \frac{1}{2n}\Big)\Big| dx$$

(en utilisant (8.1)). On applique alors le Lemme 8.2.3 ci-dessous, analogue à la Proposition 5.5.1, pour conclure que $c_n(f) \to 0$ quand $|n| \to +\infty$.

(3) et (4) sont immédiats (pour (3) il faut utiliser encore (8.1)).

(5) Si f est de classe C^k, elle et ses dérivées d'ordre $j \leqslant k$ sont continues donc bornées, et donc intégrables. Il suffit de traiter le cas $k = 1$ par récurrence (en utilisant le fait que f est C^k sur $\mathbf{R}$ entier donc les dérivées sont périodiques). On calcule alors

$$c_n(f') = \int_0^1 f'(x)e(-nx)dx = \Big[f(x)e(-nx)\Big]_0^1 + 2i\pi n \int_0^1 f(x)e(-nx)dx = 2i\pi n c_n(f)$$

car le premier terme est nul. □

Voici l'analogue périodique de la Proposition 5.5.1.

Lemme 8.2.3. *Soit $f \in L^1(I)$. Alors on a*

$$\lim_{h\to 0} \int_0^1 |f(x+h) - f(x)|dx = 0.$$

Démonstration. On peut, au choix, adapter l'argument original, ou bien se raccrocher à celui-ci : étant donnée $f : \mathbf{R} \to \mathbf{C}$ périodique et intégrable sur $[0, 1]$, posons $g = f\chi_{[-1/2,3/2]}$. Alors $g \in L^1(\mathbf{R})$ et $g(x) = f(x)$ si $-1/2 \leqslant x \leqslant 3/2$. Il vient pour $|h| < 1/2$

$$\begin{aligned}\int_0^1 |f(x+h) - f(x)|dx &= \int_0^1 |g(x+h) - g(x)|dx \\ &\leqslant \int_{\mathbf{R}} |g(x+h) - g(x)|dx \to 0 \text{ quand } h \to 0\end{aligned}$$

par la Proposition 5.5.1 appliquée à g. □

Associée à la suite $c_n(f)$ des coefficients de Fourier de f est la *série de Fourier* de f :

$$S(f)(x) = \sum_{n\in\mathbf{Z}} c_n(f)e(nx). \tag{8.4}$$

Plus généralement, si (b_n), $n \in \mathbf{Z}$, est une suite quelconque de nombres complexes, on peut considérer la *série trigonométrique*

$$s_b(x) = \sum_{n\in\mathbf{Z}} b_n e(nx), \tag{8.5}$$

vue d'abord formellement, c'est à dire sans hypothèse de convergence. Il est cependant clair que si $s_b(x)$ converge pour tout $x \in \mathbf{R}$, sa limite est une fonction 1-périodique.

Il n'y a pas, *a priori*, de raison pour que ces deux notions coïncident : une série trigonométrique, même convergeant pour tout $x \in \mathbf{R}$, n'est pas forcément une série de Fourier.

Le point de départ historique est le calcul suivant : si $g = s_b$ est une fonction périodique donnée par une série trigonométrique (8.5), on calcule (d'abord formellement)

$$c_m(g) = \int_0^1 g(x)e(-mx)dx = \sum_{n\in\mathbf{Z}} b_n \int_0^1 e((n-m)x)dx \tag{8.6}$$

or on a le lemme fondamental (parfaitement rigoureux lui) :

Lemme 8.2.4. *Pour tout n, $m \in \mathbf{Z}$, on a*

$$\int_0^1 e((n-m)x)dx = \int_0^1 e(nx)\overline{e(mx)}dx = \langle e_n, e_m \rangle = \delta(n, m)$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire dans $L^2(I)$.

Démonstration. On a pour $h = 0$

$$\int_0^1 e_h(x)dx = \int_0^1 dx = 1,$$

et pour $h \neq 0$

$$\int_0^1 e_h(x)dx = \Big[\frac{1}{2i\pi h} e_h(x)\Big]_0^1 = 0$$

par périodicité. Il suffit d'appliquer cela à $h = n - m$. $\square$

Autrement dit, dans l'espace de Hilbert $L^2(I)$, les fonctions e_n, $n \in \mathbf{Z}$ forment une *famille orthonormale*.

Revenant à (8.6), le lemme donne alors

$$c_m(g) = b_m$$

puisque tout les termes $n \neq m$ de la somme sont nuls. Autrement dit, *si le calcul est permis*, c'est à dire si on peut intégrer la série trigonométrique donnée (8.5) terme par terme, ses coefficients b_n sont les coefficients de Fourier de g.

Cela a justifié l'espoir, qui est correct dans de nombreux cas, que réciproquement si $f \in L^1(I)$ est donnée, alors elle est représentée (en un certain sens) par sa série de Fourier (8.4). Par exemple c'est évidemment le cas si $s_b(x)$ est un polynôme trigonométrique (cf. l'exemple ci-dessous).

Commençons par donner un cas où le calcul formel des coefficients de Fourier d'une série trigonométrique est justifié. Puisque la série (8.5) porte sur $n \in \mathbf{Z}$, il faut préciser comment on entend la « sommer ». On pose alors

$$s_{b,n}(x) = \sum_{|k| \leqslant n} b_k e(kx)$$

(un polynôme trigonométrique) qu'on appelle la n-ème *somme partielle* de la série. On dit que la série $s_b(x)$ converge vers $a \in \mathbf{C}$ en un point $x \in \mathbf{R}$ si et seulement si $s_{b,n}(x) \to a$ quand $n \to +\infty$. Attention, si la convergence n'est pas absolue, cette limite peut être différente de celle obtenue en choisissant un ordre différent pour grouper les coefficients.

Proposition 8.2.5. *Soit (b_n), $n \in \mathbf{Z}$, une suite de nombres complexes tels que pour tout $x \in \mathbf{R}$, la série trigonométrique*

$$s_b(x) = \sum_{n \in \mathbf{Z}} b_n e(nx)$$

converge vers $f(x) \in \mathbf{C}$, et telle qu'il existe $g \geqslant 0$ périodique intégrable telle que $|s_{b,n}(x)| \leqslant g(x)$ pour tout n. Alors $f \in L^1(I)$ et

$$c_m(f) = b_m$$

pour tout $m \in \mathbf{Z}$.

Démonstration. Il suffit d'appliquer le théorème de convergence dominée sur $[0, 1[$ à la suite $s_{b,n}e_{-m}$ qui converge vers fe_{-m} point par point et vérifie $|s_{n,b}(x)e(-mx)| \leqslant g(x)$, donc

$$\int_0^1 f(x)e(-mx)dx = \lim_{n \to +\infty} \int_0^1 s_{n,b}(x)e(-mx)dx = b_m$$

en intervertissant la somme et l'intégrale comme dans (8.6). $\square$

Exemple 8.2.6. (1) Si les coefficients (b_n) forment une série absolument convergente, c'est à dire vérifient

$$\sum_{n \in \mathbf{Z}} |b_n| < +\infty$$

(ce qui est analogue à la condition $\hat{f} \in L^1(\mathbf{R}^d)$ qui donne la formule d'inversion pour la transformée de Fourier), la série $s_b(x)$ converge absolument et uniformément pour $x \in \mathbf{R}$. Sa limite est donc une fonction périodique continue f et on peut appliquer la proposition, donc $c_m(f) = b_m$ pour tout m.

Noter qu'il n'y a pas de réciproque : pour f périodique continue, il n'est pas du tout garanti que la somme des coefficients de Fourier converge absolument.

(2) En particulier, ou bien plus directement, si la suite (b_n) est de support finie, c'est à dire si $s_b(x)$ est en réalité un polynôme trigonométrique $f(x)$, on a $c_m(f) = b_m$ et l'identité

$$f(x) = \sum_{m\in\mathbf{Z}} c_m(f)e(mx)$$

est alors tautologique (la série étant une somme finie) ! Il n'est donc pas étonnant que l'on puisse vouloir (et pouvoir) étendre cette formule à des suites ou des fonctions plus générales. (Noter qu'il n'existe pas d'analogue des polynôme trigonométriques pour la transformée de Fourier.)

(3) Si la série $s_b(x)$ converge en tout point vers une somme $f(x)$ et que ses sommes partielles sont uniformément bornées, puisque $L^\infty(I) \subset L^1(I)$, on peut également appliquer la proposition.

Remarque 8.2.7. Pour une fonction f de période 2π au lieu de 1, les fonctions qu'on utilise au lieu de e_n sont plutôt e^{inx}, $n \in \mathbf{Z}$: les coefficients de Fourier d'une telle fonction sont

$$c_n(f) = \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-inx}dx,$$

et la série de Fourier est

$$\sum_{n\in\mathbf{Z}} c_n(f)e^{inx}.$$

Plus généralement, si f est périodique de période $a > 0$ on pose

$$c_n(f) = \frac{1}{a}\int_0^{a} f(x)e^{-\frac{2i\pi n}{a}x}dx$$

et la série de Fourier est

$$\sum_{n\in\mathbf{Z}} c_n(f)e^{\frac{2i\pi n}{a}x}.$$

Il est également assez fréquent de considérer des fonctions définies sur $\mathbf{R}$ et de n'en garder que la restriction sur un intervalle fondamental, étendue par périodicité à $\mathbf{R}$ (cf. l'Exemple 8.3.4).

8.3. Séries de Fourier de fonctions L^2

Nous allons traiter maintenant la théorie dans $L^2(I)$, qui est bien plus simple d'apparence que pour les intégrales de Fourier. Le résultat principal s'énonce fort simplement :

Théorème 8.3.1. *Les fonctions e_n, $n \in \mathbf{Z}$, forment une* base orthonormée *de l'espace de Hilbert $L^2(I)$. Pour toute $f \in L^2(I)$ on a*

$$c_m(f) = \langle f, e_m \rangle, \tag{8.7}$$

et

$$f = \sum_{m\in\mathbf{Z}} \langle f, e_m \rangle e_m \text{ c'est à dire } f(x) = \sum_{m\in\mathbf{Z}} c_m(f)e(mx) \text{ au sens } L^2. \tag{8.8}$$

Démonstration. La formule (8.7) est valide par définition du produit scalaire dans $L^2(I)$, et (8.8) n'est qu'une reformulation du fait que (e_m) est une base. Il suffit donc de prouver ce premier point. On a déjà vu (cf. la remarque après le Lemme 8.2.4) que (e_m) est une famille orthonormale, et il suffit donc de montrer que les (e_m) engendrent $L^2(I)$ au sens hilbertien, c'est à dire que l'espace vectoriel qu'ils engendrent au sens algébrique est dense dans $L^2(I)$. Or cet espace vectoriel est évidemment l'espace P des polynômes trigonométriques. Notons $\overline{P}$ son adhérence dans $L^2(I)$.

Si l'on adopte le point de vue (3) de la Section 8.1, C s'identifie à $C(K)$ et P s'identifie à un sous espace P_1 de $C(K)$, où K est le cercle unité : la fonction e_m correspond à la fonction $z \mapsto z^m$ pour $m \in \mathbf{Z}$ (cf. (8.2)).

On vérifie donc aussitôt que P_1 est une algèbre de fonctions complexes sur K, qui sépare les points de K car e_1 (correspondant à l'identité z) les sépare, et qui est stable par conjugaison complexe car

$$\overline{z^m} = z^{-m}$$

pour $z \in K$. D'après le théorème de Stone-Weierstrass complexe, P_1 est dense dans $C(K)$ pour la norme de la convergence uniforme, c'est à dire la norme L^∞, ce qui signifie que P est dense dans C pour la norme L^∞. Or

$$\|f\|_2 \leqslant \|f\|_\infty$$

pour $f \in L^\infty(I)$, ce qui montre qu'une suite de fonctions $f_n \in L^\infty(I)$ convergeant uniformément vers g converge également en norme L^2 vers g. La densité de P pour la norme L^∞ implique donc que P est dense en norme L^2 dans C.

Donc $\overline{P} \supset \overline{C}$, mais d'après le Théorème 5.4.1, on a $\overline{C} = L^2(I)$, d'où $\overline{P} = L^2(I)$. □

Remarque 8.3.2. On verra dans la section suivante une suite *explicite* de polynômes trigonométriques convergeant uniformément vers une fonction continue périodique donnée (ce qui permettrait donc une preuve du théorème précédent indépendante de celui de Stone-Weierstrass).

Corollaire 8.3.3. *Soient f, $g \in L^2(I)$. Alors on a*

$$\sum_{n\in\mathbf{Z}} c_n(f)\overline{c_n(g)} = \int_0^1 f(x)\overline{g(x)}dx. \tag{8.9}$$

et en particulier on a l'identité de Parseval

$$\sum_{n\in\mathbf{Z}} |c_n(f)|^2 = \int_0^1 |f(x)|^2 dx. \tag{8.10}$$

Démonstration. C'est immédiat d'après le Théorème 8.3.1 et la théorie des espaces de Hilbert puisque dans (8.9), chacun des deux expressions est égale au produit scalaire $<f, g>$ dans $L^2(I)$: celle de gauche, parce que (e_n) est une base orthornomée et $c_n(f) = <f, e_n>$, et celle de droite par définition. □

Exemple 8.3.4. La formule (8.10) est source de nombreuses identités intéressantes. Soit par exemple f la fonction périodique définie par $f(x) = x$ si $-1/2 \leqslant x < 1/2$ (étendue par périodicité à $\mathbf{R}$). Comme f est bornée, elle est dans $L^2(I)$.

On calcule

$$c_0(f) = \int_{-1/2}^{1/2} x dx = 0$$

et si $n \neq 0$, par intégration par partie, on a

$$\begin{aligned}
c_n(f) &= \int_{-1/2}^{1/2} x e(-nx) dx \\
&= \Big[-\frac{x}{2i\pi n} e(-nx)\Big]_{-1/2}^{1/2} + \frac{1}{2i\pi n}\int_{-1/2}^{1/2} e(-nx)dx \\
&= -\frac{1}{4i\pi n}(e^{-i\pi n} + e^{i\pi n}) \\
&= -\frac{1}{2i\pi n}\cos(n\pi) = \frac{(-1)^{n+1}}{2i\pi n},
\end{aligned}$$

tandis que

$$\int_{-1/2}^{1/2} |f(x)|^2 dx = \int_{-1/2}^{1/2} x^2 dx = \frac{1}{12}.$$

La formule de Parseval (8.10) donne donc

$$\frac{1}{12} = \sum_{n \neq 0} \frac{1}{4\pi^2 n^2} \text{ c'est à dire } \sum_{n \geqslant 1} \frac{1}{n^2} = \frac{\pi^2}{6},$$

un résultat originellement dû à Euler (par un calcul spectaculaire et un peu risqué...)

8.4. Sommes partielles et convolution

On considère maintenant le problème de la convergence ponctuelle des séries de Fourier. On va d'abord exprimer les sommes partielles comme une *convolution.*

DÉFINITION 8.4.1. Soit $f \in L^1(I)$ et $n \geqslant 1$. La n-ème somme partielle de la série de Fourier de f est définie par

$$S_n(f)(x) = \sum_{|k| \leqslant n} c_k(f) e(kx) = c_{-n}(f) e(-nx) + \cdots + c_0(f) + \cdots + c_n(f) e(nx).$$

LEMME 8.4.2. *Soit $f \in L^1(I)$. Alors pour tout $n \geqslant 1$ on a*

$$S_n(f)(x) = \int_0^1 f(t) D_n(x-t) dt$$

où

$$D_n(x) = \frac{\sin((2n+1)\pi x)}{\sin(\pi x)}, \text{ et } D_n(0) = 2n+1. \tag{8.11}$$

REMARQUE 8.4.3. Bien entendu, D_n est continue en 0.

DÉMONSTRATION. On a

$$\begin{aligned} S_n(f)(x) &= \sum_{|k| \leqslant n} c_k(f) e(kx) \\ &= \sum_{|k| \leqslant x} \int_0^1 f(t) e(k(x-t)) dt \\ &= \int_0^1 f(t) D_n(x-t) dt \end{aligned}$$

avec

$$D_n(x) = \sum_{|k| \leqslant n} e(kx)$$

(le calcul ne demandant aucune justification puisqu'il s'agit de sommes finies), et il reste à calculer D_n explicitement : il faut sommer une progression géométrique finie de raison $e(x) = e^{2i\pi x}$, de longueur $2n+1$ et de terme initial $e(-nx)$, donc (pour $x \neq 0$, le cas $x = 0$ étant trivial)

$$\begin{aligned} \sum_{|k| \leqslant n} e(kx) &= e(-nx) \frac{1 - e((2n+1)x)}{1 - e(x)} \\ &= e(-nx) \frac{e((n+1/2)x)}{e(x/2)} \frac{-2i \sin(2\pi(n+1/2)x)}{-2i \sin(\pi x)} \\ &= \frac{\sin((2n+1)\pi x)}{\sin(\pi x)}. \end{aligned}$$

□

Cela justifie, si besoin est, la définition suivante :

Définition 8.4.4. Soient f et $g \in L^1(I)$ des fonctions périodiques intégrables. La *convolution* de f et g est

$$f \star g(x) = \int_0^1 f(t)g(x-t)dt$$

pour $x \in \mathbf{R}$.

On peut donc écrire $S_n(f) = f \star D_n$. Les propriétés de cette opération de convolution des fonctions périodiques sont évidemment proches de celles de la convolution étudiée au Chapitre 6. Nous les résumons dans une seule proposition.

Proposition 8.4.5. (1) *Pour toutes f, $g \in L^1(I)$, on a $f \star g \in L^1(I)$ et*

$$\|f \star g\|_1 \leqslant \|f\|_1 \|g\|_1.$$

(2) *On a $f \star g = g \star f$, $f \star (g \star h) = (f \star g) \star h$ et $f \star (\alpha g + \beta h) = \alpha(f \star g) + \beta(f \star h)$ pour f, g, $h \in L^1(I)$, α, $\beta \in \mathbf{C}$.*

(3) *Si f, $g \in L^1(I)$ et $n \in \mathbf{Z}$*

$$c_n(f \star g) = c_n(f)c_n(g). \tag{8.12}$$

(4) *Si $f \in L^1(I)$ et si g est une fonction périodique de classe C^k, $k \geqslant 1$, alors $f \star g$ est de classe C^k et pour $0 \leqslant j \leqslant k$ on a*

$$(f \star g)^{(j)} = f \star g^{(j)}.$$

Démonstration. (1) On a

$$\begin{aligned}\int_0^1 \int_0^1 |f(t)g(x-t)|dtdx &= \int_0^1 |f(t)| \Big(\int_0^1 |g(x-t)|dx \Big) dt \\ &= \Big(\int_0^1 |f(t)|dt \Big) \Big(\int_x^{x+1} |g(u)|du \Big) \\ &= \|f\|_1 \|g\|_1\end{aligned}$$

(en utilisant encore (8.1)), ce qui montre que $f \star g$ existe presque partout, puis que $f \star g$ est dans $L^1(I)$ (elle est trivialement périodique) et vérifie

$$\|f \star g\|_1 \leqslant \|f\|_1 \|g\|_1,$$

comme désiré.

(2) est exactement similaire aux énoncés correspondants pour $\mathbf{R}^d$.

(3) On calcule

$$\begin{aligned}c_n(f \star g) &= \int_0^1 \Big(\int_0^1 f(t)g(x-t)dt \Big) e(-nx)dx \\ &= \int_0^1 f(t) \Big(\int_0^1 g(x-t)e(-nx)dx \Big) dt \\ &= c_n(f)c_n(g).\end{aligned}$$

(4) Il suffit de dériver sous le signe d'intégration en utilisant la Proposition 3.3.3, et le fait qu'une fonction périodique continue est dans $L^1(I)$. $\square$

Remarque 8.4.6. La formule des coefficients de Fourier d'une convolution (8.12) permet de retrouver la formule pour $S_n(f)$: si b_k est la suite (dépendant de n)

$$b_k = \begin{cases} 1 & \text{si } |k| \leqslant n \\ 0 & \text{sinon} \end{cases}$$

on a

$$S_n(f) = \sum c_k(f) b_k e(kx) = f \star E_n, \text{ où } E_n(x) = \sum_k b_k e(kx) = \sum_{|k| \leqslant n} e(kx) = D_n(x).$$

La première idée pour montrer que $S_n(f) \to f$ est de montrer que la suite (D_n) dite des *noyaux de Dirichlet* est une suite de Dirac (périodique) au sens de la définition suivante qui adapte celle sur $\mathbf{R}^d$:

Définition 8.4.7. Une suite (φ_n), $n \geqslant 1$, de fonctions périodiques intégrables est appelée une suite de Dirac si $\varphi_n \geqslant 0$ et

$$\int_0^1 \varphi_n(t)dt = 1 \text{ pour tout } n \geqslant 1 \tag{8.13}$$

$$\lim_{n\to+\infty} \int_{[\eta,1-\eta]} \varphi_n(t)dt = 0 \text{ pour tout } \eta,\ 0<\eta<1/2. \tag{8.14}$$

On peut réécrire (8.14) sous la forme

$$\int_{U_\eta} \varphi_n(t)dt \to 0$$

où $U_\eta = [-1/2,-\eta] \cup [\eta,1/2]$.

Le même schéma de preuve que pour le Théorème 6.4.6 fournit :

Proposition 8.4.8. *Soit (φ_n) une suite de Dirac périodique.*
(1) *Si $f \in L^1(I)$, alors $f \star \varphi_n \to f$ en norme L^1.*
(2) *Si $x \in \mathbf{R}$, $f \in L^\infty(I)$ et f est continue en x, alors $f \star \varphi_n(x) \to x$.*
(3) *Si f est continue sur $\mathbf{R}$, alors $f \star \varphi_n$ converge uniformément vers f sur $\mathbf{R}$.*

Démonstration. Il est possible d'être bref puisque le raisonnement est exactement parallèle à celui employé pour $\mathbf{R}^d$.

On a par (8.13)

$$f\star\varphi_n(x) - f(x) = \int_0^1 (f(x-t)-f(x))\varphi_n(t)dt.$$

Il est pratique ici de prendre $[-1/2,1/2]$ comme intervalle fondamental. Notons encore $U_\eta = [-1/2,-\eta]\cup[\eta,1/2]$.

Soit $\varepsilon > 0$. Si $f \in L^1$, d'après le Lemme 8.2.3 il existe η, $0<\eta<1/2$, tel que $|t|<\eta$ implique

$$\int_{-1/2}^{1/2} |f(x-t)-f(x)|dt < \varepsilon,$$

et donc il vient

$$\begin{aligned}\|f\star\varphi_n - f\|_1 &\leqslant \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} |f(x-t)-f(x)|\varphi_n(t)dtdx\\ &\leqslant \int_{|t|<\eta}\Big(\int_{-1/2}^{1/2}|f(x-t)-f(x)|dx\Big)\varphi_n(t)dt + 2\|f\|_1\int_{U_\eta}\varphi_n(t)dt\\ &\leqslant \varepsilon + 2\|f\|_1\int_{U_\eta}\varphi_n(t)dt\end{aligned}$$

et donc par (8.14), on trouve bien $f\star\varphi_n \to f$ dans $L^1(I)$.

Pour f continue en x, il existe $\eta>0$ tel que $|t|<\eta$ implique $|f(x-t)-f(x)|<\varepsilon$, et donc

$$\begin{aligned}|f\star\varphi_n(x)-f(x)| &\leqslant \int_{|t|<\eta}|f(x-t)-f(x)|\varphi_n(t)dt + 2\|f\|_\infty\int_{U_\eta}\varphi_n(t)dt\\ &\leqslant \varepsilon + 2\|f\|_\infty\int_{U_\eta}\varphi_n(t)dt,\end{aligned}$$

d'où la convergence de $f\star\varphi(x)$ vers $f(x)$. Si de plus f est continue partout, elle est automatiquement bornée, et étant périodique elle est uniformément continue, ce qui signifie que $\eta > 0$

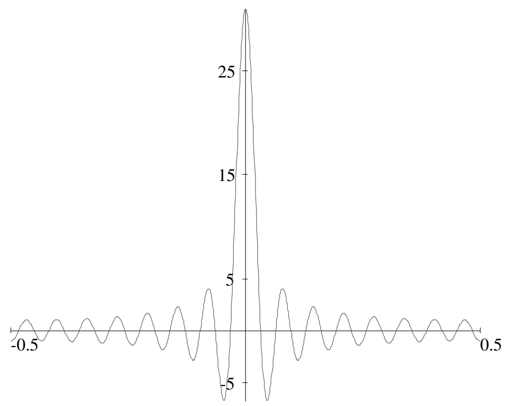
peut être choisi comme ci-dessus ne dépendant que de ε et pas du choix de x. Cela implique aussitôt la convergence *uniforme* de $f \star \varphi_n$ vers f. $\square$

On ne peut appliquer cependant cette proposition directement aux sommes partielles de la série de Fourier de $f \in L^1$ car (D_n) n'est pas une suite de Dirac : précisément, D_n n'est pas positive, et de plus on peut montrer que

$$\|D_n\|_1 \to +\infty$$

(ces nombres sont appelés parfois les *constantes de Lebesgue* ; noter que la preuve de la Proposition 8.4.8 pourrait s'étendre sans changements majeurs si au lieu de $\varphi_n \geqslant 0$ on demandait seulement que la suite $(\|\varphi_n\|_1)$ soit bornée).

Comme illustration, voici le graphe de D_{15} :



8.5. Sommes de Fejér et convergence en moyenne de Césaro

Pour compenser la difficulté apparue à la fin de la Section précédente, une méthode particulièrement élégante est due à Fejér.

L'idée peut se présenter de différentes façons, en voici une : on a vu (c'est l'une des idées de bases de la théorie de l'intégration) que « prendre la moyenne » a (souvent) un effet régularisant (cf. par exemple la loi des grands nombres, les différents théorèmes d'approximation par convolution, etc...) Puisque les sommes partielles $S_n(f)$ sont problématiques, on va les remplacer par une moyenne.

Définition 8.5.1. Soit $f \in L^1(I)$. Les sommes de Fejér $\sigma_n(f)$ sont définies par

$$\sigma_n(f) = \frac{1}{n}(S_0(f) + \cdots + S_{n-1}(f))$$

pour $n \geqslant 1$.

Autrement dit, il s'agit de prendre la moyenne des n premières sommes partielles. L'effet régularisateur apparaît effectivement :

Proposition 8.5.2. *Pour toute $f \in L^1(I)$ et $n \geqslant 1$ on a*

$$\sigma_n(f)(x) = \sum_{|k| \leqslant n} c_k(f)\Big(1 - \frac{|k|}{n}\Big)e(kx) \tag{8.15}$$

$$= \int_0^1 f(t)F_n(x-t)dt \tag{8.16}$$

où
$$F_n(x) = \frac{1}{n}\Big(\frac{\sin^2(n\pi x)}{\sin^2(\pi x)}\Big) \ \textit{et}\ F_n(0) = n.$$

Démonstration. La première formule est très simple : pour x donné, posons $\alpha_n = c_n(f)e(nx)$ pour simplifier ; on a par définition
$$\begin{aligned}
\sigma_n(f)(x) &= \frac{1}{n}\Big(\alpha_0 + (\alpha_{-1} + \alpha_0 + \alpha_1) + \cdots + (\alpha_{-n+1} + \cdots + \alpha_0 + \cdots + \alpha_{n-1})\Big)\\
&= \frac{1}{n}\Big(n\alpha_0(f) + (n-1)(\alpha_{-1}(f) + \alpha_1(f)) + \cdots + (\alpha_{-n+1}(f) + \alpha_{n-1}(f))\Big)\\
&= \frac{1}{n}\sum_{|k|\leqslant n-1} c_k(f)(n-|k|)e(kx)\\
&= \sum_{|k|\leqslant n} c_k(f)\Big(1 - \frac{|k|}{n}\Big)e(kx)
\end{aligned}$$
(puisque les termes avec $|k| = n$ deviennent nuls).

La formule (8.16) s'en déduit avec
$$F_n(x) = \sum_{|k|\leqslant n}\Big(1 - \frac{|k|}{n}\Big)e(kx)$$
(cf. Proposition 8.4.5, (3)) ou bien
$$F_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} D_k(x).$$

Calculons par exemple à partir de cette dernière formule : par (8.11) on a (pour $x \neq 0$)
$$\begin{aligned}
F_n(x) &= \frac{1}{n}\frac{1}{\sin(\pi x)}\sum_{k=0}^{n-1}\sin((2k+1)\pi x)\\
&= \frac{1}{n}\frac{1}{\sin(\pi x)}\operatorname{Im}\Big(\sum_{k=0}^{n-1} e((k+1/2)x)\Big)\\
&= \frac{1}{n}\frac{1}{\sin(\pi x)}\operatorname{Im}\Big(e(x/2)\frac{1-e(nx)}{1-e(x)}\Big)\\
&= \frac{1}{n}\frac{1}{\sin(\pi x)}\operatorname{Im}\Big(e(x/2)\frac{e(nx/2)}{e(x/2)}\frac{\sin(n\pi x)}{\sin(\pi x)}\Big)\\
&= \frac{1}{n}\frac{\sin^2(n\pi x)}{\sin^2(\pi x)}
\end{aligned}$$
(par sommation de progression géométrique encore). □

Le point crucial est alors :

Proposition 8.5.3. *La suite (F_n), $n \geqslant 1$, est une suite de Dirac périodique.*

Démonstration. Il est évident que $F_n \geqslant 0$. De plus, si on choisit $f = 1$, on trouve $S_n(1) = 1$ pour tout $n \geqslant 1$, donc $\sigma_n(1) = 1$ pour $n \geqslant 1$ et
$$1 = \sigma_n(1)(0) = \int_0^1 F_n(t)dt$$
pour tout $n \geqslant 1$.

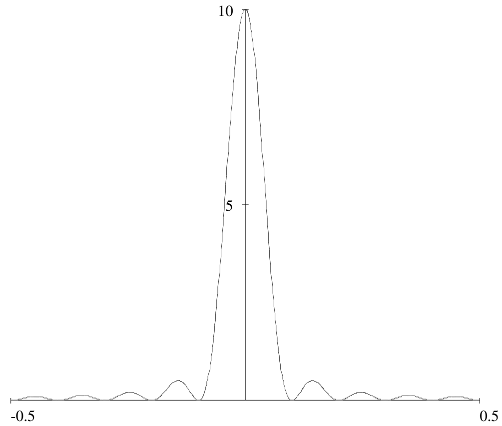
Enfin, si $0 < \eta < 1/2$, et $t \in U_\eta$, on a $|\sin \pi x| \geqslant \sin \pi\eta$, d'où
$$F_n(t) \leqslant \frac{1}{n \sin \pi\eta} \text{ pour } t \in U_\eta$$

et

$$\int_{U_\eta} F_n(t)dt \leqslant \frac{1}{n\sin(\pi\eta)} \to 0.$$

□

Voici le graphe de F_{10} :



On déduit aussitôt quelques corollaires très importants grâce à la Proposition 8.4.5.

Corollaire 8.5.4. (1) *Si $f \in L^1(I)$, alors $\sigma_n(f) \to f$ en norme L^1.*
(2) *Si f est continue périodique, alors $\sigma_n(f) \to f$ uniformément sur $\mathbf{R}$.*

Démonstration. Il suffit maintenant de regrouper (8.16), la Proposition 8.5.3 et la Proposition 8.4.5, (1) ou (3). □

Corollaire 8.5.5. *Soit $f \in L^1(I)$. Si $c_n(f) = 0$ pour tout n, alors $f = 0$ presque partout ; en d'autres termes, l'application « coefficients de Fourier » $f \mapsto (c_n(f))_n$ est injective.*

Démonstration. Si $c_n(f) = 0$ pour tout n, on a $\sigma_n(f) = 0$, et donc $f = \lim \sigma_n(f) = 0$ dans $L^1(I)$. □

Noter que pour $f \in L^2(I)$, cela était déjà une conséquence du Théorème 8.3.1.

On déduit aussi l'autre preuve, constructive, du théorème de Stone-Weierstrass pour les fonctions continues périodiques.

Corollaire 8.5.6. *Les polynômes trigonométriques sont denses dans l'espace des fonctions continues périodiques pour la norme L^∞, et plus précisément $\sigma_n(f) \to f$.*

Démonstration. C'est une simple répétition ! □

8.6. Théorème de Dirichlet

Aussi satisfaisants que soient les résultats de la section précédente, on reste curieux de savoir ce qu'il en est de la convergence des sommes partielles « ordinaires » $S_n(f)$. Il se trouve que la difficulté aperçue par le fait que (D_n) n'est pas une suite de Dirac n'est pas illusoire : il existe des fonctions continues f telles que (par exemple) $S_n(f)$ ne converge pas pour $x = 0$.

Cependant, un peu de régularité supplémentaire permet d'éliminer ce type de phénomène. Voici un des résultats dans cette direction, dû à Dirichlet.

Théorème 8.6.1. *Soit $f \in C^1_p(\mathbf{R})$ une fonction périodique de classe C^1. Alors $S_n(f) \to f$ uniformément sur $\mathbf{R}$.*

Quoique la démonstration originale de Dirichlet ait été différente, nous allons déduire cela des résultats de Fejér (historiquement postérieurs).

Démonstration. Pour tout $x \in \mathbf{R}$ on a (8.15)

$$\sigma_n(f)(x) = \sum_{|k| \leqslant n} c_k(f)\Big(1 - \frac{|k|}{n}\Big)e(kx)$$

et la somme partielle est donnée par

$$S_n(f)(x) = \sum_{|k| \leqslant n} c_k(f)e(kx),$$

donc

$$S_n(f)(x) - \sigma_n(f)(x) = \frac{1}{n} \sum_{|k| \leqslant n} |k| c_k(f)e(kx),$$

et en majorant

$$\|S_n(f) - \sigma_n(f)\|_\infty \leqslant \frac{1}{n} \sum_{|k| \leqslant n} |kc_k(f)|. \tag{8.17}$$

On reconnaît là la moyenne des sommes partielles de la suite de nombres positifs $u_k = |kc_k(f)|$. Comme f est de classe C^1 sur $\mathbf{R}$ on a

$$2i\pi n c_n(f) = c_n(f')$$

par la Proposition 8.2.2, (5) pour tout $n \in \mathbf{Z}$, et donc

$$\lim_{|k| \to +\infty} kc_k(f) = 0$$

puisque $c_k(f') \to 0$ par *loc. cit.*, (2).

Or un lemme bien connu (Lemme 8.6.3 ci-dessous) affirme que si une suite complexe (a_n) converge vers $a \in \mathbf{C}$ alors

$$\frac{1}{n}(a_0 + \cdots + a_{n-1}) \to a \text{ quand } n \to +\infty.$$

Ainsi (8.17) implique que $\|S_n(f) - \sigma_n(f)\|_\infty \to 0$. Comme d'après le théorème de Fejér (Corollaire 8.5.4, (2)) on $\sigma_n(f) \to f$ uniformément (f est continue), il vient

$$S_n(f) \to f \text{ uniformément.}$$

□

Remarque 8.6.2. On peut définir plus généralement, pour une suite complexe (u_n), la notion de *convergence au sens de Cesaro* : on dit que (u_n) converge vers $u \in \mathbf{C}$ en ce sens si

$$v_n = \frac{1}{n}(u_0 + \cdots + u_{n-1}) \to u.$$

Cela s'applique aussi bien sûr aux séries, en considérant la suite des sommes partielles de celles-ci. Le théorème de Fejér dit donc que la série de Fourier d'une fonction continue f converge uniformément vers f *au sens de Cesaro*.

Il est classique que la convergence ordinaire implique la convergence de Cesaro :

Lemme 8.6.3. *Si $u_n \to u$, alors u_n converge vers u au sens de Cesaro et si $\sum u_n = s$, alors la série a pour somme s au sens de Cesaro.*

Démonstration. Il suffit de considérer le cas $u = 0$ en considérant $u'_n = u_n - u$ si nécessaire. Soit

$$\sigma_n = \frac{1}{n}(u_0 + \cdots + u_{n-1}).$$

Soit $\varepsilon > 0$. Par définition de la convergence, il existe N tel que pour tout $n > N$, on a $|u_n| < \varepsilon$. Donc pour $n > N$ il vient

$$|\sigma_n| \leqslant \frac{1}{n}(|u_0| + \cdots + |u_N| + (n - |N|)\varepsilon) \leqslant \frac{b_N}{n} + \varepsilon,$$

où $b_N = |u_0| + \cdots + |u_N|$. Puisque N a été fixé, on a $b_N/n \to 0$ quand $n \to +\infty$, donc il existe $N_1 > N$ tel que si $n > N_1$, alors $|b_N/n| < \varepsilon$.

Par conséquent pour tout $n > N_1$ on a

$$|\sigma_n| \leqslant 2\varepsilon$$

Cela signifie que $\sigma_n \to 0$, ce qui est le résultat désiré. □

Mais il est également bien connu que la réciproque du lemme est fausse : l'exemple le plus classique est la série

$$\sum_{n\geqslant 1} (-1)^{n-1} = 1 - 1 + 1 - 1 + \cdots$$

Puisque le terme général ne tend pas vers 0, cette série diverge. Mais les sommes partielles s_n vérifient

$$s_{2n+1} = 1 \text{ et } s_{2n} = 0 \text{ pour tout } n \geqslant 1$$

donc leurs moyennes de Cesaro sont données par

$$\sigma_{2n} = \frac{1}{2n}(s_1 + \cdots + s_{2n}) = \frac{1}{2}$$

et

$$\sigma_{2n+1} = \frac{1}{2n+1}(s_1 + \cdots + s_{2n+1}) = \frac{n+1}{2n+1} = \frac{1}{2} + \frac{1}{4n+2}$$

ce qui montre que $\sigma_n \to 1/2$, c'est à dire que la série a pour somme $1/2$ au sens de Cesaro.

En pratique, la preuve du théorème de Dirichlet s'est appuyée sur une réciproque partielle du lemme : si une série $\sum u_n$ converge au sens de Cesaro, *et si* elle vérifie la condition supplémentaire que $nu_n \to 0$ quand $n \to +\infty$ (c'est à dire si (u_n) décroît assez vite), alors elle converge (vers la même somme) au sens ordinaire.

Ce type d'argument est très instructif et caractéristique de tout un pan de l'analyse : les sommes partielles « naïves » sont d'un abord malaisé en général. Les sommes de Fejér ou de Cesaro, d'après la formule (8.15), n'en sont pas très éloignées : il s'agit, au lieu de couper brutalement la somme après le n-ème terme, de pondérer les $c_k(f)e(kx)$, $|k| \leqslant n$, par un coefficient qui décroît plus doucement et régulièrement vers 0. La raison cachée de l'efficacité de ce procédé est que la transformée de Fourier, ou les coefficients de Fourier, d'une fonction sont d'autant plus petits que celle-ci est régulière...

Une autre preuve élégante (et plus générale) du théorème de Dirichlet est le sujet du Problème 2 de l'examen final.

Nous terminons par quelques remarques énonçant des résultats autrement plus difficiles que ceux démontrés ici, qui illustrent le fait que le comportement (surtout ponctuel) des séries de Fourier est un sujet très délicat et mathématiquement loin d'être parfaitement compris.

Remarque 8.6.4. (1) Il existe des fonctions continues périodiques f dont la série de Fourier diverge en certains points. On peut même construire des fonctions telles que $S_n(f)(x)$ diverge pour tout x dans un ensemble dense dans $\mathbf{R}$. Et si on considère les fonctions intégrables, Kolmogorov a même montré qu'il existe $f \in L^1(I)$ dont la série de Fourier diverge *partout*.

(2) Par contre, répondant à une question longtemps ouverte, Carleson a montré (1964) que si $f \in L^2(I)$, alors les sommes partielles de f convergent presque partout vers f. (Rappelons qu'on sait que $S_N(f) \to f$ en norme L^2, et d'après le Théorème 3.1.10, il existe donc une *sous-suite* n_k telle que $S_{n_k}(f)(x) \to f(x)$ pour presque tout x ; il n'est pas du tout évident que l'on puisse prendre la même sous-suite pour toute $f \in L^2$). En particulier, si f est continue, elle est dans L^2 et donc ses sommes partielles convergent presque partout : les ensembles de divergence denses mentionnés dans la première remarque sont donc nécessairement de mesure nulle.

CHAPITRE 9

Théorie de Fourier et probabilités

Dans ce chapitre final nous revenons aux probabilités qui ont été délaissées ces derniers temps pour appliquer les résultats et principes concernant la transformée de Fourier à l'étude des variables aléatoires.

9.1. Fonction caractéristique d'une variable aléatoire

Soit (Ω, Σ, P) un espace probabilisé. Nous avons déjà défini au Chapitre 3 la *fonction caractéristique* d'une variable aléatoire réelle X (voir la Remarque 3.4.6). Nous rappellons cette définition ici :

DÉFINITION 9.1.1. (1) Soit X une variable aléatoire *réelle* sur Ω. La fonction caractérisique φ_X de X est la fonction

$$\begin{cases} \mathbf{R} \to \mathbf{C} \\ t \mapsto E(e^{itX}). \end{cases}$$

(2) Soit μ une mesure de probabilité sur $\mathbf{R}$. Sa transformée de Fourier φ_μ est la fonction

$$\begin{cases} \mathbf{R} \to \mathbf{C} \\ t \mapsto \displaystyle\int_{\mathbf{R}} e^{itx} d\mu(x). \end{cases}$$

REMARQUE 9.1.2. (1) Il est important que X soit à valeurs réelles, puisque dans ce cas seulement $|e^{itX}| = 1$ sur Ω, et donc e^{itX} est bornée et par conséquent intégrable. Lorsqu'on veut appliquer les méthodes basées sur la fonction caractéristique à des variables aléatoires complexes, on procède en général en séparant les parties réelles et imaginaires.

(2) Ne pas confondre cette fonction caractéristique avec celle d'un ensemble !

(3) Si $\mu = X(P)$ est la mesure sur $\mathbf{R}$ qui est la loi de X, on a

$$\varphi_X(t) = \int_{\mathbf{R}} e^{itx} d\mu(x) = \varphi_\mu(t),$$

de sorte que les deux notions sont équivalentes.

(4) La normalisation choisie est la plus standard en probabilité : malheureusement elle est en conflit avec celle utilisée pour la transformée de Fourier...

Voici les propriétés les plus simples de la fonction caractéristique.

PROPOSITION 9.1.3. (1) *La fonction $\varphi_X(t)$ est une fonction uniformément continue sur $\mathbf{R}$ telle que $\varphi_X(0) = 1$ et*

$$|\varphi_X(t)| \leqslant 1$$

pour tout $t \in \mathbf{R}$.

(2) *Si $\lambda \in \mathbf{R}$, on a*

$$\varphi_{\lambda X}(t) = \varphi_X(\lambda t) \text{ pour tout } t \in \mathbf{R}.$$

(3) *Si X et Y sont des variables aléatoires réelles indépendantes, alors*

$$\varphi_{X+Y} = \varphi_X \varphi_Y.$$

(4) *Si $X \in L^n$ avec $n \geqslant 1$ entier, alors $\varphi_X \in C^n(\mathbf{R})$ et pour $0 \leqslant j \leqslant n$ on a*

$$\varphi_X^{(j)}(t) = i^j E(X^j e^{itX}), \tag{9.1}$$

et en particulier

$$\varphi_X^{(j)}(0) = i^j E(X^j).$$

Démonstration. (1) La continuité uniforme de φ_X provient de la Proposition 3.3.1 avec une adaptation simple : soit μ la loi de X. On a

$$|\varphi_X(t+h) - \varphi_X(t)| = |E(e^{itX}(e^{ihX} - 1)| \leqslant E(|e^{ihX} - 1|).$$

Or

$$E(|e^{ihX} - 1|) = \int_{\mathbf{R}} |e^{ihx} - 1| d\mu(x).$$

Pour tout $x \in \mathbf{R}$ fixé, $h \mapsto |e^{ihx} - 1|$ est continue (et s'annule) en 0, et majorée par 2 qui est μ-intégrable. Donc on a

$$\lim_{h\to 0} \int_{\mathbf{R}} |e^{ihx} - 1| d\mu(x) = 0$$

ce qui montre la continuité uniforme. L'égalité $\varphi_X(0) = 1$ est valide par définition et la majoration $\|\varphi_X\|_\infty \leqslant 1$ est également évidente.

(2) On a $\varphi_{\lambda X}(t) = E(e^{it\lambda X}) = \varphi_X(\lambda t)$ tautologiquement.

(3) On a

$$\varphi_{X+Y}(t) = E(e^{it(X+Y)}) = E(e^{itX}e^{itY}).$$

Si X et Y sont indépendantes, alors e^{itX} et e^{itY} le sont également (cf. Proposition 1.2.11, (2)), et donc

$$E(e^{itX}e^{itY}) = E(e^{itX})E(e^{itY}) = \varphi_X(t)\varphi_Y(t)$$

par la Proposition 4.3.5.

(4) On procède par récurrence : la formule (9.1) est vraie pour $n = 0$, d'après le point (1) ci-dessus, et si elle l'est pour n, on a

$$\varphi_X^{(n+1)}(t) = \frac{d}{dt} E(X^n e^{itX}) = iE(X^{n+1}e^{itX})$$

d'après la Proposition 3.3.3 si X^{n+1} est intégrable puisque $|X^{n+1}e^{itX}| \leqslant |X|^{n+1}$. De même, $\varphi_X^{(n+1)}$ est continue d'où le résultat final par récurrence. □

Exemple 9.1.4. Voici quelques exemples importants de fonctions caractéristiques. Puisqu'elles ne dépendent que de la loi de la variable aléatoire, on les donne plutôt comme fonction de celles-ci.

(1) Si $\mu = \delta_a$ est la mesure de Dirac en a, alors

$$\varphi_\mu(t) = \int_{\mathbf{R}} e^{itx} d\delta_a(x) = e^{ita}.$$

En particulier, pour $a = 0$, on trouve $\varphi_{\delta_0}(t) = 1$ pour tout t.

(2) Plus généralement, si $x_k \in \mathbf{R}$ et $p_k \geqslant 0$ pour $1 \leqslant k \leqslant n$, avec

$$\sum_{k=1}^n p_k = 1,$$

alors pour la mesure (correspondant à des variables aléatoires discrètes telles que $p_k = P(X = x_k)$)

$$\mu = \sum_{k=1}^n p_k \delta_{x_k},$$

on a

$$\varphi_\mu(t) = \sum_{k=1}^n p_k e^{itx_k}.$$

(3) Soit $\mu = \mu_{a,\sigma}$ la mesure de probabilité gaussienne d'espérance $a \in \mathbf{R}$ et variance $\sigma > 0$,

$$\mu_{a,\sigma} = g_{a,\sigma}(x)dx \text{ avec } g_{a,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-a)^2}{2\sigma}},$$

(cf. Proposition 4.4.10). On a alors

$$\varphi_\mu(t) = \hat{g}_{a,\sigma}\Big(-\frac{t}{2\pi}\Big),$$

et les formules (7.2) et (7.4), avec le fait que

$$\mathcal{F}(g_{0,(2\pi)^{-1}}) = g_{0,(2\pi)^{-1}}$$

(c'est le calcul de la transformée de Fourier de $x \mapsto e^{-\pi x^2}$ effectué dans la preuve du Lemme 7.3.4), montrent (si l'on ne fait pas de faute de calcul...) que

$$\varphi_\mu(t) = e^{iat} e^{-\sigma t^2/2}.$$

Soit X une variable aléatoire dont la loi est $X(P) = \mu$. Alors $X \in L^n$ pour tout n, car $g_{a,\sigma} \in \mathcal{S}(\mathbf{R})$. À l'aide de la Proposition 9.1.3, on peut calculer facilement les *moments* $E(X^n)$ de X si $a = 0$. En effet, on a par (9.1)

$$E(X^n) = i^{-n} \varphi_\mu^{(n)}(0),$$

que l'on trouve en faisant le développement de Taylor de φ_μ en 0 :

$$\varphi_\mu(t) = \sum_{k \geqslant 0} \frac{(-1)^k \sigma^k t^{2k}}{2^k k!},$$

(puisque ici $a = 0$) et donc par comparaison

$$E(X^n) = \begin{cases} 0 & \text{si } n \text{ est impair} \\ \dfrac{(2m)!}{m!}\Big(\dfrac{\sigma}{2}\Big)^m & \text{si } n = 2m \text{ est pair.} \end{cases}$$

On vérifie que cela redonne $V(X) = E(X^2) = \sigma$ (prendre $m = 1$).

(4) Soit $\mu = \mu_a$ la mesure de probabilité

$$\mu_a = \frac{a}{\pi} \frac{1}{a^2 + x^2} dx$$

où $a > 0$ (appelée une *loi de Cauchy*). On remarque que si X est une variable aléatoire dont μ est la loi, alors X *n'est pas intégrable*, puisque $x \mapsto |x|/(1 + x^2)$ n'est pas intégrable sur $\mathbf{R}$.

En utilisant l'Exemple 7.3.5, on vérifie que

$$\varphi_\mu(t) = e^{-a|t|}.$$

Il s'agit bien d'une fonction continue mais pas de classe C^1.

(5) Soit $\mu = \mu_\lambda$ la mesure de probabilité

$$\mu_\lambda = \lambda e^{-\lambda x} \chi_{[0,+\infty[} dx$$

où $\lambda > 0$ (appelée *loi exponentielle*). Une variable aléatoire dont μ est la loi prend donc ses valeurs (presque sûrement) dans $[0, +\infty[$. On calcule aisément

$$\varphi_\mu(t) = \frac{\lambda}{\lambda - it}.$$

En développant $\varphi_\mu(t)$ en série de Taylor en 0

$$\varphi_\mu(t) = \frac{1}{1 - \frac{it}{\lambda}} = \sum_{k \geqslant 0} \frac{i^k t^k}{\lambda^k}$$

pour $|t| < \lambda$, on calcule comme précédemment pour les gaussiennes que

$$E(X^n) = \frac{n!}{\lambda^n}$$

pour $n \geqslant 0$. En particulier $E(X) = \lambda^{-1}$ et

$$V(X) = E(X^2) - E(X)^2 = 2\lambda^{-2} - \lambda^{-2} = \lambda^{-2}.$$

L'importance des fonctions caractéristiques est souvent liée au fait qu'elles fournissent un critère pour la convergence en loi. Rappelons (cf. Définition 5.6.1) que $X_n \xrightarrow{\text{loi}} X$ (X_n converge en loi vers X) signifie que $E(f(X_n)) \to E(f(X))$ pour toute fonction continue à support compact f, ou même pour toute fonction continue bornée, ce qui est équivalent par la Proposition 5.6.3). On a déjà vu une interprétation de cette définition à l'aide de la fonction de répartition des variables aléatoires (cf. Proposition 5.6.6). En voici une autre :

Théorème 9.1.5. *Soit (Ω, Σ, P) un espace probabilisé, (X_n), $n \geqslant 1$, une suite de variables aléatoires réelles, μ une mesure de probabilité sur $\mathbf{R}$. Alors on a $X_n \xrightarrow{loi} \mu$ si et seulement si $\varphi_{X_n}(t) \to \varphi_\mu(t)$ pour tout $t \in \mathbf{R}$.*

L'intérêt de ce résultat vient en partie du fait que la convergence demandée est une convergence point par point sans aucune condition d'uniformité.

Démonstration. Soit $\mu_n = X_n(P)$ et $\varphi_n = \varphi_{\mu_n}$. Supposons $\varphi_n(t) \to \varphi(t)$ pour tout t. On va d'abord montrer que

$$\int_{\mathbf{R}} g(x)d\mu_n(x) \to \int_{\mathbf{R}} g(x)d\mu(x) \tag{9.2}$$

pour toute fonction C^∞ à support compact $g \in C_c^\infty(\mathbf{R})$. Soit alors en effet $f \in C_c(\mathbf{R})$. Il existe $g_m \in C_c^\infty(\mathbf{R})$ telles que $g_m \to g$ uniformément (appliquer le Théorème 6.4.6, (2) à une suite de Dirac dans $C_c^\infty(\mathbf{R})$), et alors (comparer avec la preuve de la Proposition 5.6.6) on a

$$\begin{aligned}\Big|\int_{\mathbf{R}} f(x)d\mu_n(x) - \int_{\mathbf{R}} f(x)d\mu(x)\Big| &\leqslant \int_{\mathbf{R}} |f(x) - g_m(x)|d\mu_n(x) + \Big|\int_{\mathbf{R}} g_m(x)d\mu_n(x) - \int_{\mathbf{R}} g_m(x)d\mu(x)\Big| \\ &\quad + \int_{\mathbf{R}} |g_m(x) - f(x)|d\mu(x) \\ &\leqslant 2\|g_m - f\|_\infty + \Big|\int_{\mathbf{R}} g_m(x)d\mu_n(x) - \int_{\mathbf{R}} g_m(x)d\mu(x)\Big|.\end{aligned}$$

Pour $\varepsilon > 0$ donné, on choisit m tel que $\|g_m - f\|_\infty < \varepsilon$ et alors pour n assez grand la formule (9.2) pour $g_m \in C_c^\infty(\mathbf{R})$ donne

$$\limsup_{n\to+\infty} \Big|\int_{\mathbf{R}} f(x)d\mu_n(x) - \int_{\mathbf{R}} f(x)d\mu(x)\Big| \leqslant \varepsilon$$

d'où finalement la convergence en loi.

Montrons donc (9.2) pour terminer cette première partie de la preuve. Pour cela, on fait appel à la formule d'inversion de Fourier (valide puisque $g \in \mathcal{S}(\mathbf{R})$) pour écrire, pour toute mesure de probabilité ν sur $\mathbf{R}$:

$$\begin{aligned}\int_{\mathbf{R}} g(x)d\nu(x) &= \int_{\mathbf{R}} \Big(\int_{\mathbf{R}} \hat{g}(t)e^{2i\pi xt}dt\Big)d\nu(x) \\ &= \int_{\mathbf{R}} \hat{g}(t)\Big(\int_{\mathbf{R}} e^{2i\pi xt}d\nu(x)\Big)dt \\ &= \int_{\mathbf{R}} \hat{g}(t)\varphi_\nu(2\pi t)dt,\end{aligned}$$

par le théorème de Fubini, dont la justification est immédiate ici puisque $\hat{g} \in \mathcal{S}(\mathbf{R})$. (Cette formule est directement analogue à (7.12)). Maintenant, en tout t on a par hypothèse

$$\hat{g}(t)\varphi_n(2\pi t) \to \hat{g}(t)\varphi_\mu(2\pi t)$$

et de plus $|\hat{g}(t)\varphi_n(2\pi t)| \leqslant |\hat{g}(t)|$ qui est intégrable, donc le théorème de convergence dominé montre que

$$\int_{\mathbf{R}} g(x)d\mu_n(x) = \int_{\mathbf{R}} \hat{g}(t)\varphi_n(2\pi t)dt \to \int_{\mathbf{R}} \hat{g}(t)\varphi_\mu(2\pi t)dt = \int_{\mathbf{R}} g(x)d\mu(x),$$

c'est à dire (9.2).

Réciproquement, si $\mu_n \xrightarrow{\text{loi}} \mu$, alors on sait par la Proposition 5.6.3 que

$$\int_{\mathbf{R}} g(x)d\mu_n(x) \to \int_{\mathbf{R}} g(x)d\mu(x)$$

pour toute fonction continue bornée g. Prenant $g(x) = e^{itx}$ avec $t \in \mathbf{R}$ fixé, qui est continue de module 1, on trouve directement

$$\varphi_n(t) \to \varphi_\mu(t).$$

pour tout $t \in \mathbf{R}$. □

9.2. Application : le théorème central

On a déjà énoncé le théorème central-limite (Théorème 3.2.10) : le revoici sous une forme équivalente (grâce à la Proposition 5.6.6) :

Théorème 9.2.1. *Soit (Ω, Σ, P) un espace probabilisé, (X_n) une suite de variables aléatoires réelles indépendantes et uniformément distribuées, telles que $X_n \in L^2(\Omega)$ et $E(X_n) = 0$. Soit $\sigma = V(X_n)$ la variance des X_n. Alors*

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{loi} \mu_{0,\sigma}$$

la loi normale d'espérance 0 et variance σ.

Démonstration. D'après le critère de convergence en loi donné par le Théorème 9.1.5, et l'Exemple 9.1.4, (3), il suffit de montrer que, notant

$$T_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}},$$

on a

$$\varphi_{T_n}(t) \to e^{-\sigma t^2/2} \tag{9.3}$$

quand $n \to +\infty$ (attention de remarquer que T_n n'est pas la même variable aléatoire que la « moyenne empirique » S_n utilisée dans la preuve de la loi des grands nombres).

Soit μ la loi commune des X_i. D'après la Proposition 9.1.3, (2) et (3), on a

$$\varphi_{T_n}(t) = \varphi_\mu\Big(\frac{t}{\sqrt{n}}\Big)^n$$

puisque les X_i sont indépendantes et de même loi (donc de même fonction caractéristique).

Puisque $X_n \in L^2(\Omega)$, on sait par la Proposition 9.1.3, (4) que φ_μ est C^2. D'après la formule de Taylor on a

$$\varphi_\mu(t) = 1 + \varphi_\mu'(0)t + \frac{\varphi_\mu''(0)}{2}t^2 + t^2\varepsilon(t) \text{ avec } \lim_{t\to 0}\varepsilon(t) = 0$$

donc par (9.1)

$$\varphi_\mu(t) = 1 - \frac{\sigma t^2}{2} + t^2\varepsilon(t)$$

(puisque $E(X_1) = 0$, donc $E(X_1^2) = V(X_1)$, par hypothèse). Ainsi

$$\varphi_{T_n}(t) = \Big(1 - \frac{\sigma t^2}{2n} + \frac{t^2}{n}\varepsilon\Big(\frac{t}{\sqrt{n}}\Big)\Big)^n$$

Si l'on négligeait le « terme d'erreur » contenant $\varepsilon(t^2/n)$, on aurait aussitôt

$$\Big(1-\frac{\sigma t^2}{2n}\Big)^n = \exp\Big(n\log\Big(1-\frac{\sigma t^2}{2n}\Big)\Big) \to \exp(-\sigma t^2/2),$$

(car $\lim x^{-1}\log(1+ax) = a$ quand $x \to 0$) donnant (9.3) comme désiré. Il faut donc simplement justifier que la correction contenant $\varepsilon(t/\sqrt{n})$ est négligeable à la limite.

Mais ceci est aisé : on fixe t, et on utilise l'encadrement

$$-x - x^2 \leqslant \log(1-x) \leqslant -x, \text{ pour } -\frac{1}{2} \leqslant x \leqslant \frac{1}{2}$$

(vérifié immédiatement par calcul de dérivée...) qui montre que si $n > N$, N choisi assez grand pour que

$$\Big|\frac{1}{2} - \frac{\sigma t^2}{2n} + \frac{t^2}{n}\varepsilon\Big(\frac{t}{\sqrt{n}}\Big)\Big| \leqslant \frac{1}{2},$$

(cette quantité tend vers 0 quand $n \to +\infty$), on a

$$-\frac{\sigma t^2}{2n} + \frac{t^2}{n}\varepsilon\Big(\frac{t}{\sqrt{n}}\Big) - \Big\{\frac{\sigma t^2}{2n} - \frac{t^2}{n}\varepsilon\Big(\frac{t}{\sqrt{n}}\Big)\Big\}^2 \leqslant \log\Big(1 - \frac{\sigma t^2}{2n} + \frac{t^2}{n}\varepsilon\Big(\frac{t}{\sqrt{n}}\Big)\Big) \leqslant -\frac{\sigma t^2}{2n} + \frac{t^2}{n}\varepsilon\Big(\frac{t}{\sqrt{n}}\Big)$$

donc

$$-\frac{\sigma t^2}{2} + t^2\varepsilon\Big(\frac{t}{\sqrt{n}}\Big) - \Big\{\frac{\sigma t^2}{2\sqrt{n}} - \frac{t^2}{\sqrt{n}}\varepsilon\Big(\frac{t}{\sqrt{n}}\Big)\Big\}^2 \leqslant \log\varphi_{T_n}(t) \leqslant -\frac{\sigma t^2}{2} + t^2\varepsilon\Big(\frac{t}{\sqrt{n}}\Big).$$

Maintenant, quand $n \to +\infty$, les bornes supérieures et inférieures dans cet encadrement tendent toutes deux vers $-\sigma t^2/2$ (rappelons que t est fixé et $\varepsilon(u) \to 0$ si $u \to 0$...), et par continuité il vient en effet

$$\varphi_{T_n}(t) \to \exp\Big(-\frac{\sigma t^2}{2}\Big).$$

□

Exemple 9.2.2. Il s'agit plutôt d'un contre-exemple. Soit (X_n) une suite de variables aléatoires indépendantes et de même loi, celle-ci étant la loi de Cauchy de paramètre 1 (Exemple 9.1.4, (4)), donc

$$X_n(P) = \frac{1}{\pi}\frac{dx}{1+x^2}.$$

Posons encore $T_n = (X_1 + \cdots + X_n)/\sqrt{n}$. Comme dans la preuve du théorème on a

$$\varphi_{T_n}(t) = \varphi_{X_1}\Big(\frac{t}{\sqrt{n}}\Big)^n = \exp(-|t|\sqrt{n})$$

d'après l'Exemple 9.1.4, (4). Donc $\varphi_{T_n}(0) \to 1$ et $\varphi_{T_n}(t) \to 0$ pour tout $t \neq 0$. Cette fonction limite n'est pas continue, et n'est donc pas une fonction caractéristique : le théorème central limite ne s'étend pas aux variables aléatoires non-intégrables.

Remarque 9.2.3. Et nous terminons le cours par un exemple montrant comment on peut utiliser le théorème central limite pour comprendre la modélisation mathématique du « mouvement brownien » : ce n'est qu'une fenêtre minuscule vers un sujet particulièrement vaste et passionnant.

Historiquement, c'est un botaniste anglais qui décrivit en 1827 les mouvements apparemment complètement désordonnés de petits grains de pollen dans un liquide. La première explication est d'attribuer les incessants changements de direction que l'on observe à des chocs successifs avec les molécules de ce liquide, invisibles à l'œil nu. Bien qu'il puisse paraître difficile d'en dire plus, ces collisions étant par nature aléatoires, c'est un des aspects essentiels de la théorie moderne des probabilités que de parvenir à placer un ordre sur ce type de comportements « stochastiques » ou « chaotiques » : si une trajectoire *individuelle* est effectivement presque impossible à prédire, l'effet global d'un nombre incroyablement grands de telles trajectoires de particules différentes finit par acquérir une régularité susceptible d'être décrite mathématiquement de façon exacte.

Précisément, un modèle mathématique du mouvement brownien fut proposé par Einstein en 1905 (de la même année date la relativité restreinte), et indépendamment par Smoluchowski. En termes modernes, il s'agit de dire qu'il existe un espace probabiliste (Ω, Σ, P) et sur Ω des variables aléatoires B_t à valeurs dans $\mathbf{R}^3$ (de manière équivalente, dans $\mathbf{C}$), où $t \geqslant 0$ est le « temps », vérifiant les quatre propriétés ci-dessous, de sorte qu'intuitivement un grain de pollen correspond à un événement $\omega \in \Omega$, et sa trajectoire est la « courbe » paramétrée

$$t \mapsto B_t(\omega)$$

parcourue par les valeurs « successives » de la famille (B_t) évaluée en ω.

Nous avons mentionné quatre propriétés caractéristiques : en effet, sans plus de conditions, les B_t pourraient être triviales (par exemple, constantes). Ces propriétés sont :

- On a $B_0 = 0$ (c'est à dire $B_0(\omega) = 0$ pour tout ω), que l'on interprète simplement comme une normalisation du point de départ de la particule ; si l'on veut imaginer partir d'un autre point $(x_0, y_0, z_0) \in \mathbf{R}^3$, fixé ou lui-même aléatoire, il suffit de considérer $B_t + (x_0, y_0, z_0)$.
- Pour tout ω, la fonction $t \mapsto B_t(\omega)$ est continue.
- Si $0 \leqslant t < s$, la loi de la variable aléatoire $B_s - B_t$ mesurant le vecteur entre ces deux instants ne dépend que de $s - t$. De plus, pour toute suite finie $0 \leqslant t_0 < t_1 < \cdots < t_n$ de temps, le vecteur des *écarts* successifs correspondant

$$(B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}) \tag{9.4}$$

  est une famille de variables aléatoires indépendantes.
- Chaque B_t est une variable aléatoire *gaussienne*, d'espérance nulle et de variance $V(B_t) = \alpha t$, où $\alpha > 0$ est une constante « physique ». La valeur de α étant mathématiquement sans importance, on peut supposer $\alpha = 1$.

Interprétons les trois dernières conditions, qui sont évidemment cruciales. La seconde dit seulement que les trajectoires sont continues : c'est physiquement naturel.

La troisième correspond au fait que les collisions auxquelles sont soumises les particules sont complètement et uniformément aléatoires (l'atmosphère étant supposée homogène et infinie), et qu'après chaque choc, la direction dans laquelle elles poursuivent leur course est donc aléatoire, et indépendante des directions précédemment choisies. Les composantes de (9.4) décrivant les vecteurs dans $\mathbf{R}^3$ entre différents points de la trajectoire ont donc tout lieu d'être indépendants, et de ne « dépendre » que du temps parcouru...

Admettant cela, la dernière condition est pratiquement imposée par le théorème central-limite : en effet on peut écrire

$$B_t = \sum_{i=1}^{n} (B_{it/n} - B_{(i-1)t/n})$$

pour tout $n \geqslant 1$, ce qui exprime B_t comme somme de n variables aléatoires indépendantes et de même loi d'après ce qui précède. Comme n peut être arbitrairement grand, cela suggère aussitôt que B_t suive une loi normale ; l'espérance doit être nulle pour des raisons de symétrie (aucune direction n'est privilégiée), et quand à la variance, la formule ci-dessus donne aussi

$$V(B_t) = nV(B_{t/n}).$$

Il est facile de vérifier que si $t \mapsto V(B_t)$ est supposée également continue, les seules solutions sont de la forme $V(B_t) = \alpha t$ pour un certain $\alpha \geqslant 0$, et similairement

$$V(B_s - B_t) = \alpha(s - t)$$

pour $0 \leqslant t < s$.

Si l'on prend maintenant le point de vue mathématique, la question devient : existe-t-il effectivement une telle famille de variables aléatoires (B_t) ? Cela n'est nullement évident, et ne pouvait être résolu originellement faute des outils probabilistes basés sur la théorie de la mesure.

Cependant, dans les années 30, N. Wiener et P. Lévy démontrèrent l'existence de tels espaces probabilisés munis de « mouvements browniens ».

En utilisant des résultats tels que le Théorème 5.6.8, il n'est pas trop difficile d'obtenir un exemple d'espace (Ω, Σ, P) avec (B_t), $t \geqslant 0$, vérifiant toutes les conditions demandées, *sauf* la continuité des trajectoires $t \mapsto B_t(\omega)$. Ce dernier point est effectivement plus délicat et dépend de propriétés des lois gaussiennes qui ne seraient pas valides si B_t suivait des lois « arbitraires. »

Partiel – Bordeaux, Avril 2002

Énoncé

Problème 1. (Questions de cours)

(1) Énoncer le théorème de convergence monotone et le théorème de convergence dominée.

(2) Définir l'espérance $E(X)$ et la variance $V(X)$ d'une variable aléatoire réelle X sur (Ω, Σ, P). Quelles formules permettent de les calculer en fonction de la loi $\mu = X(P)$ de X ? Que vaut $E(XY)$ si X et Y sont indépendantes ?

Problème 2. On considère $\mathbf{R}^d$ muni de la mesure de Lebesgue. Soit $K \in L^2(\mathbf{R}^d \times \mathbf{R}^d)$.

(1) Montrer que si $f \in L^2(\mathbf{R}^d)$ l'intégrale

$$T_K(f)(x) = \int_{\mathbf{R}^d} K(x,y)f(y)dy \tag{9.5}$$

existe pour presque tout $x \in \mathbf{R}^d$ (c'est à dire que $y \mapsto K(x,y)f(y)$ est intégrable pour presque tout x), et que la fonction $T_K(f)$ ainsi définie est dans $L^2(\mathbf{R}^d)$, et vérifie

$$\|T_K(f)\|_2 \leqslant \|K\|_2 \|f\|_2.$$

[Indication : On pourra considérer d'abord le cas où $K \geqslant 0$ et $f \geqslant 0$ et montrer alors que $\int |T_K(f)|^2 dx < +\infty$.]

(2) En déduire que T_K est un opérateur linéaire continu $L^2(\mathbf{R}^d) \to L^2(\mathbf{R}^d)$ de norme $\leqslant \|K\|_2$, et que l'application $K \mapsto T_K$ est elle-même continue de norme $\leqslant 1$.

(3) Étant donnée $f \in L^2(\mathbf{R}^d)$ et $\lambda \in \mathbf{C}$, on veut résoudre (si possible) l'équation d'inconnue $g \in L^2(\mathbf{R}^d)$

$$g(x) - \lambda \int_{\mathbf{R}^d} K(x,y)g(y)dy = f(x). \tag{9.6}$$

Montrer que la série

$$g = \sum_{i \geqslant 0} \lambda^i T_K^i(f)$$

(où $T_K^i = T_K \circ \cdots \circ T_K$, composition répétée i fois de l'opérateur T_K) converge dans $L^2(\mathbf{R}^d)$ si $|\lambda| < \|K\|_2^{-1}$ et que sa somme g est une solution de (9.6).

(5) Montrer que g est l'unique solution de (9.6) si $|\lambda| < \|K\|_2^{-1}$. [Indication : Montrer que l'équation $g = \lambda T_K(g)$ n'a pas de solution $\neq 0$ sous cette hypothèse.]

(6) Soient L, $M \in L^2(\mathbf{R}^d \times \mathbf{R}^d)$. Montrer que

$$N(x,y) = \int_{\mathbf{R}^d} L(x,z)M(z,y)dz$$

définit une fonction $N \in L^2(\mathbf{R}^d \times \mathbf{R}^d)$ telle que

$$\|N\|_2 \leqslant \|L\|_2 \|M\|_2$$

et que $T_N = T_L \circ T_M$.

(7) Soit $K_1 = K$ et

$$K_{n+1}(x,y) = \int_{\mathbf{R}^d} K(x,z)K_n(z,y)dz.$$

Montrer que pour toute $f \in L^2(\mathbf{R}^d)$ on a

$$T_K^n(f)(x) = \int_{\mathbf{R}^d} K_n(x,y) f(y) dy,$$

puis que la série

$$R_\lambda = \sum_{n \geqslant 1} \lambda^{n-1} K_n$$

est convergente dans $L^2(\mathbf{R}^d \times \mathbf{R}^d)$ si $|\lambda| < \|K\|_2^{-1}$, et enfin que la solution g de (9.6) construite à la question (4) est donnée aussi par

$$g(x) = f(x) + \lambda \int_{\mathbf{R}^d} R_\lambda(x,y) f(y) dy$$

pour presque tout x.

Problème 3. Soit (Ω, Σ, P) un espace probabilisé, et $X_1, \ldots, X_n$ des variables aléatoires réelles, *indépendantes*, telles que $X_i \in L^2$ et $E(X_i) = 0$. Soit

$$S_k = X_1 + \cdots + X_k \in L^2.$$

(1) Montrer que

$$E(S_n^2) = V(S_n) = \sum_{k=1}^{n} V(X_k).$$

(2) On fixe $\varepsilon > 0$. Pour $1 \leqslant k \leqslant n$, soit A_k l'événement

$$A_k = \{|S_j| < \varepsilon \text{ pour } 1 \leqslant j < k \text{ et } |S_k| \geqslant \varepsilon\}$$

et Y_k sa fonction caractéristique.

Montrer que

$$P(\max_{1 \leqslant k \leqslant n} |S_k| \geqslant \varepsilon) = \sum_{k=1}^{n} P(A_k) \leqslant \varepsilon^{-2} \sum_{k=1}^{n} E(Y_k S_k^2).$$

[Indication : Remarquer que $S_k^2 \geqslant \varepsilon^2$ sur A_k].

(3) Montrer que pour $1 \leqslant k \leqslant n$ on a

$$E(Y_k S_n^2) \geqslant E(Y_k S_k^2) + 2E(Y_k S_k (S_n - S_k)).$$

[Indication : Écrire $S_n = S_n - S_k + S_k$.]

(4) Montrer que $Y_k S_k$ et $S_n - S_k$ sont indépendantes et en déduire que pour $1 \leqslant k \leqslant n$ on a

$$E(Y_k S_n^2) \geqslant E(Y_k S_k^2).$$

(5) Montrer alors que

$$P(\max_{1 \leqslant k \leqslant n} |S_k| \geqslant \varepsilon) \leqslant \varepsilon^{-2} E(S_n^2) = \varepsilon^{-2} \sum_{k=1}^{n} V(X_k).$$

Soit maintenant (X_n), $n \geqslant 1$, une suite de variables aléatoires réelles indépendantes telles que $X_n \in L^2$ et

$$\sum_{n \geqslant 1} V(X_n) < +\infty.$$

(6) Soit B l'événement « la série $\sum X_n$ ne converge pas. » Montrer que

$$B \subset \bigcup_{m \geqslant 1} \bigcap_{N \geqslant 1} B_{N,m} \text{ avec } B_{N,m} = \{\max_{k \geqslant 1} |S_{N+k} - S_N| \geqslant m^{-1}\}$$

où $S_n = X_1 + \cdots + X_n$. [Indication : Utiliser le critère de Cauchy.]

(7) En utilisant le début du problème, montrer que

$$P(B_N) \leqslant m^2 \sum_{k > N} V(X_k).$$

(8) En déduire que pour tout m on a

$$P\Big(\bigcap_N B_{N,m}\Big) = 0$$

puis que la série $\sum X_n$ converge presque sûrement.

(9) Application : si ε_n est une suite de variables aléatoires indépendantes à valeurs dans $\{-1,1\}$ de même loi μ telle que

$$\mu(\{-1\}) = \mu(\{1\}) = \frac{1}{2},$$

alors la série

$$\sum_{n\geqslant 1} \frac{\varepsilon_n}{n}$$

converge presque sûrement.

Corrigé

Problème 1. (1) Convergence monotone : Soit $(X, \mathcal{M}, \mu)$ un espace mesurable. Soit (f_n) une suite de fonctions mesurables positives à valeurs dans $[0,+\infty]$ telles que $f_{n+1} \geqslant f_n$. Soit $f(x) = \lim f_n(x)$. C'est encore une fonction mesurable et

$$\int_X f_n(x)d\mu(x) \to \int_X f(x)d\mu(x).$$

Convergence dominée : soit (f_n) une suite de fonctions intégrables telles que $f_n(x) \to f(x)$ pour tout $x \in X$ et $|f_n| \leqslant g$ pour tout n avec $g \in L^1(\mu)$. Alors $f \in L^1(\mu)$ et

$$\int_X f_n(x)d\mu(x) \to \int_X f(x)d\mu(x).$$

De plus $f_n \to f$ dans L^1, c'est à dire

$$\int_X |f_n - f|d\mu \to 0.$$

(2) Si X est une variable aléatoire réelle alors

$$E(X) = \int_\Omega X(\omega)dP(\omega), \quad V(X) = E((X - E(X))^2).$$

Si μ est la loi de X, on a

$$E(X) = \int_{\mathbf{R}} x d\mu(x), \quad V(X) = \int_{\mathbf{R}} (x - E(X))^2 d\mu(x),$$

et si X et Y sont indépendantes, alors

$$E(XY) = E(X)E(Y).$$

Problème 2. Soit $K \in L^2(\mathbf{R}^d \times \mathbf{R}^d)$.

(1) Si $K \geqslant 0$ et $f \geqslant 0$, toutes les intégrales existent dans $[0,+\infty]$ et on a par Cauchy-Schwarz

$$\Big(\int K(x,y)f(y)dy\Big)^2 \leqslant \|f\|_2^2 \int K(x,y)^2 dy,$$

donc par Tonelli

$$\int T_K(f)(x)^2 dx \leqslant \|f\|_2^2 \int\int K(x,y)^2 dydx = \|f\|_2^2 \|K\|_2^2 < +\infty.$$

Dans le cas général ce calcul montre que la fonction $x \mapsto T_{|K|}(|f|)(x)$ est dans L^2. En particulier elle est finie presque partout, et donc $T_K(f)(x)$ est défini pour presque tout x, et vérifie

$$\|T_K(f)\|_2 \leqslant \|T_{|K|}(|f|)\|_2 \leqslant \|f\|_2 \|K\|_2.$$

(2) Il est évident que $f \mapsto T_K(f)$ est une application (=opérateur) linéaire sur $L^2(\mathbf{R}^d)$. L'inégalité juste démontrée veut exactement dire qu'elle est continue de norm $\leqslant \|K\|_2$.

De même, $T : K \mapsto T_K$ est évidemment linéaire, et comme on a

$$\|T_K\| \leqslant \|K\|_2$$

(la première norme étant celle des applications linéaires $L^2 \to L^2$), cela signifie que $\|T\| \leqslant 1$.

(3) Soit $f \in L^2(\mathbf{R}^d)$ et $\lambda \in \mathbf{C}$. On veut résoudre

$$g(x) - \lambda \int_{\mathbf{R}^d} K(x,y)g(y)dy = f(x). \tag{9.7}$$

On a par récurrence

$$\|T_K^i(f)\|_2 = \|T_K(T_K^{i-1}(f))\|_2 \leqslant \|K\|_2 \|T_K^{i-1}(f)\|_2 \leqslant \cdots \leqslant \|K\|_2^i \|f\|_2.$$

Dans la série $g = \sum \lambda^i T_K^i(f)$, on a donc

$$\|\lambda^i T_K^i(f)\|_2 \leqslant (|\lambda| \|K\|_2)^i \|f\|_2.$$

Par comparaison avec la série géométrique, on voit que c'est une série absolument convergente dans L^2 si $|\lambda| \|K\|_2 < 1$. Sa somme existe donc alors dans L^2 puisque L^2 est complet.

Par continuité de l'application linéaire T_K, on a

$$\lambda T_K(g) = \lambda \sum_{i \geqslant 0} \lambda^i T_K^{i+1}(f) = \sum_{i \geqslant 0} \lambda^{i+1} T_K^{i+1}(f) = g - f,$$

ce qui est équivalent à (9.7).

(5) Si $|\lambda| < \|K\|_2^{-1}$, et si g_1 et g_2 sont des solutions de (9.7), alors $g = g_1 - g_2$ vérifie $T_K(g) = \lambda g$. Or cela implique

$$|\lambda| \|g\|_2 = \|T_K(g)\|_2 \leqslant \|K\|_2 \|g\|_2,$$

ce qui est une contradiction si $\|g\|_2 \neq 0$. Donc $g = 0$, et $g_1 = g_2$.

(6) Soient L, $M \in L^2(\mathbf{R}^d \times \mathbf{R}^d$ et

$$N(x,y) = \int_{\mathbf{R}^d} L(x,z)M(z,y)dz.$$

Si $L \geqslant 0$, $M \geqslant 0$, on a

$$N(x,y)^2 \leqslant \Big(\int L(x,z)^2 dz\Big)\Big(\int M(z,y)^2 dz\Big)$$

et donc en intégrant sur x puis sur y

$$\int N(x,y)^2 dxdy \leqslant \Big(\int L(x,z)^2 dxdz\Big)\Big(\int M(z,y)^2 dzdy\Big) = \|L\|_2 \|M\|_2,$$

ce qui montre comme pour la question (1) qu'en général la fonction N est dans L^2 et vérifie $\|N\|_2 \leqslant \|L\|_2 \|M\|_2$.

Par Tonelli d'abord dans le cas positif puis Fubini, on vérifie que

$$\begin{aligned} T_L(T_M(f))(x) &= \int L(x,y) T_M(f)(y)dy \\ &= \int L(x,y)\Big(\int M(y,z)f(z)dz\Big)dy \\ &= \int \Big(\int L(x,y)M(y,z)dy\Big) f(z)dz \\ &= \int N(x,z)f(z)dz = T_N(f)(x). \end{aligned}$$

(7) Soit $K_1 = K$ et

$$K_{n+1}(x,y) = \int_{\mathbf{R}^d} K(x,z)K_n(z,y)dz.$$

D'après (6), par récurrence sur n on voit que

$$T_K^n(f)(x) = \int_{\mathbf{R}^d} K_n(x,y)f(y)dy.$$

Par (6) encore, on a $\|K_n\|_2 \leqslant \|K\|_2^n$ pour tout n, et donc $\|\lambda^{n-1}K_n\|_2 \leqslant \|K\|_2(|\lambda|\|K\|_2)^{n-1}$, et par comparaison encore la série

$$R_\lambda = \sum_{n\geqslant 1} \lambda^{n-1}K_n$$

converge dans $L^2(\mathbf{R}^d \times \mathbf{R}^d)$ si $|\lambda| < \|K\|_2^{-1}$. Par continuité de $K \mapsto T_K$, on a (égalité d'applications linéaires)

$$\lambda T_{R_\lambda} = \sum_{n\geqslant 1} \lambda^n T_{K_n} = \sum_{n\geqslant 1} \lambda^n T_K^n$$

donc appliqué à f,

$$\lambda T_{R_\lambda}(f) = \sum_{n\geqslant 1} \lambda^n T_K^n(f) = g - f,$$

où g est la solution de (9.7) construite en (3). Cette formule équivaut à

$$g(x) = f(x) + \lambda \int R_\lambda(x,y)f(y)dy$$

pour presque tout x.

Problème 3. Soit (Ω, Σ, P) un espace probabilisé, et $X_1, \ldots, X_n$ des variables aléatoires réelles, *indépendantes*, telles que $X_i \in L^2$ et $E(X_i) = 0$. Soit

$$S_k = X_1 + \cdots + X_k \in L^2.$$

(1) Si X et Y sont indépendantes et L^2, on a $V(X+Y) = V(X) + V(Y)$. Donc $V(S_n) = \sum_{k=1}^n V(X_k)$. On a de plus $V(S_n) = E((S_n - E(S_n))^2) = E(S_n^2)$ parce que $E(S_n) = 0$.

(2) On fixe $\varepsilon > 0$. Pour $1 \leqslant k \leqslant n$, soit A_k l'événement

$$A_k = \{|S_j| < \varepsilon \text{ pour } 1 \leqslant j < k \text{ et } |S_k| \geqslant \varepsilon\}$$

et Y_k sa fonction caractéristique. L'événement $\max\{|S_k| \geqslant \varepsilon\}$ est réunion disjointe des A_k (k indiquant quelle somme $|S_k|$ est la première à dépasser ε) donc sa probabilité est $\sum P(A_k)$. De plus, sur A_k on a $S_k^2 \geqslant \varepsilon^2$, donc

$$P(A_k) = E(Y_k) \leqslant \varepsilon^{-2}E(Y_kS_k^2).$$

(3) On calcule

$$\begin{aligned}
E(Y_kS_n^2) &= E(Y_k(S_n - S_k + S_k)^2) \\
&= E(Y_k(S_n - S_k)^2) + 2E(Y_kS_k(S_n - S_k)) + E(Y_kS_k^2) \\
&\geqslant 2E(Y_kS_k(S_n - S_k)) + E(Y_kS_k^2)
\end{aligned}$$

car le troisième terme est $\geqslant 0$.

(4) On a $Y_kS_k = \varphi_1(X_1, \ldots, X_k)$ et $S_n - S_k = \varphi_2(x_{k+1}, \ldots, x_n)$, où

$$\begin{aligned}
\varphi_1(x_1, \ldots, x_k) &= \chi_k(x_1, \ldots, x_k)(x_1 + \cdots + x_k) \\
\chi_k(x_1, \ldots, x_k) &= \begin{cases} 1 & \text{si } |x_1|, \ldots, |x_1 + \cdots + x_{k-1}| < \varepsilon \text{ et } |x_1 + \cdots + x_k| \geqslant \varepsilon \\ 0 & \text{sinon} \end{cases} \\
\varphi_2(x_{k+1}, \ldots, x_n) &= x_{k+1} + \cdots + x_n.
\end{aligned}$$

Comme $W = (X_1, \ldots, X_k)$ et $Z = (X_{k+1}, \ldots, X_n)$ sont indépendants par hypothèse, les variables $Y_kS_k = \varphi_1(W)$ et $S_n - S_k = \varphi_2(Z)$ sont encore indépendantes. En particulier

$$E(Y_kS_k(S_n - S_k)) = E(Y_kS_k)E(S_n - S_k) = 0$$

car $E(S_n) = E(S_k)$. Avec l'inégalité de (3), on trouve $E(Y_kS_n^2) \geqslant E(Y_kS_k^2)$.

(5) Finalement, il vient

$$\begin{aligned}
P(\max_{1\leqslant k\leqslant n}|S_k|\geqslant\varepsilon) &\leqslant \varepsilon^{-2}\sum_{1\leqslant k\leqslant n}E(Y_kS_k^2)\\
&\leqslant \varepsilon^{-2}\sum_{1\leqslant k\leqslant n}E(Y_kS_n^2)\\
&= \varepsilon^{-2}E\Big(S_n^2\sum_{1\leqslant k\leqslant n}Y_k\Big)\\
&\leqslant \varepsilon^{-2}E(S_n^2)=\varepsilon^{-2}V(S_n)
\end{aligned}$$

la dernière inégalité provenant du fait que $\sum Y_k\leqslant 1$ puisque les A_k sont disjoints.

Soit maintenant (X_n), $n\geqslant 1$, une suite de variables aléatoires réelles indépendantes telles que $X_n\in L^2$ et

$$\sum_{n\geqslant 1}V(X_n)<+\infty.$$

(6) Soit B l'événement « la série $\sum X_n$ ne converge pas. » Appliquant le critère de Cauchy, on trouve que pour tout N,

$$B\subset\bigcup_{m\geqslant 1}\bigcap_{N>0}B_{N,m}$$

où pour $N>0$, $m\geqslant 1$, on pose

$$B_{N,m}=\{\max_{k\geqslant 1}|S_{N+k}-S_N|\geqslant m^{-1}\}$$

(si la série ne converge pas pour ω, il existe $\varepsilon=1/m>0$ tel que pour tout N il existe $M>N$ avec $|S_M-S_N|\geqslant 1/m$).

(7) On déduit du début du problème que pour tout K on a

$$P(\max_{k\leqslant K}|S_{N+k}-S_N|\geqslant m^{-1})\leqslant m^2V(S_{N+K}-S_N)=m^2\sum_{N<k\leqslant N+K}V(X_k).$$

puis en faisant $K\to+\infty$, il vient

$$P(B_{N,m})\leqslant m^2\sum_{k>N}V(X_k).$$

(8) Il s'agit du reste d'une série convergente, donc il tend vers 0 quand $N\to+\infty$, ce qui veut dire que

$$P\Big(\bigcap_N B_{N,m}\Big)=0,$$

pour tout m et donc $P(B)=0$ en faisant la réunion sur m.

(9) L'application est immédiate en posant $X_k=\varepsilon_kk^{-1}$, puisque $V(X_k)=k^{-2}V(\varepsilon^k)=k^{-2}V(\varepsilon_1)$ donc la série des variances converge.

Examen – Bordeaux, Mai 2002

Énoncé

Problème 1. (Questions de cours)

(1) Énoncer la formule d'inversion de Fourier pour une fonction f définie sur $\mathbf{R}^d$ en précisant les conditions de validité.

(2) Énoncer le théorème de convergence de Fejér pour une fonction continue périodique.

Problème 2. Soit $\alpha \in]0,1]$ fixé. Une fonction $f : \mathbf{R} \to \mathbf{C}$ périodique de période 1 est dite höldérienne d'ordre α s'il existe une constante $C(f) \in [0, +\infty[$ telle que

$$|f(x) - f(y)| \leqslant C(f)|x - y|^\alpha$$

pour tous x, $y \in \mathbf{R}$. On note Λ_α l'ensemble des fonctions de ce type.

(1) Montrer que Λ_α est un espace vectoriel inclus dans l'espace des fonctions continues 1-périodiques, tel que $\Lambda_{\alpha'} \subset \Lambda_\alpha$ si $\alpha \leqslant \alpha'$, et contenant l'espace $C^1(I)$ des fonctions périodiques de classe C^1 pour tout $\alpha \in]0,1]$. [Indication : Distinguer les cas $|x - y| \geqslant 1$ et $|x - y| < 1$.]

(2) Soit $f \in \Lambda_\alpha$. Pour tout $h \in \mathbf{R}$, montrer que

$$\sum_{k \in \mathbf{Z}} |c_k(f) \sin 2\pi hk|^2 = \frac{1}{4} \int_0^1 |f(x + h) - f(x - h)|^2 dx.$$

où

$$c_n(f) = \int_0^1 f(x) e(-nx) dx \text{ avec } e(x) = e^{2i\pi x}$$

est le n-ème coefficient de Fourier de f [Indication : Utiliser l'identité de Parseval pour une fonction bien choisie.]

(3) Montrer qu'il existe une constante $D(f)$ telle que

$$\sum_{k \in \mathbf{Z}} |c_k(f) \sin 2\pi hk|^2 \leqslant D(f)|h|^{2\alpha}$$

pour tout $h \in \mathbf{R}$.

(4) Soit $m \geqslant 0$ un entier et $h = 2^{-m-3}$. Pour k tel que $2^m \leqslant |k| < 2^{m+1}$, montrer que

$$|\sin 2\pi hk|^2 \geqslant 1/2.$$

(5) En déduire qu'il existe une constante $E(f) \geqslant 0$ telle que

$$\sum_{2^m \leqslant |k| < 2^{m+1}} |c_k(f)|^2 \leqslant E(f) 2^{-2\alpha m}$$

pour tout $m \geqslant 0$.

(6) Majorer alors à l'aide de ce qui précède

$$\sum_{2^m \leqslant |k| < 2^{m+1}} |c_k(f)|$$

et en déduire que si $\alpha > 1/2$, la série

$$\sum_{k \in \mathbf{Z}} |c_k(f)|$$

est convergente.

(7) En déduire que si $1/2 < \alpha \leqslant 1$ la série de Fourier d'une fonction $f \in \Lambda_\alpha$ converge normalement, donc uniformément, vers f. Quel théorème du cours est un cas particulier de ce résultat ?

Problème 3. Pour $f \in L^1(\mathbf{R})$, $x \in \mathbf{R}$ et $r > 0$ on pose

$$A_r f(x) = \frac{1}{2r} \int_{x-r}^{x+r} f(y)dy.$$

et on note μ la mesure de Lebesgue sur $\mathbf{R}$.

(1) Montrer que pour tout x, $r \mapsto A_r f(x)$ est continue sur $]0, +\infty[$.

(2) Montrer que pour tout $r > 0$, $x \mapsto A_r f(x)$ est mesurable sur $\mathbf{R}$. [Indication : On pourra utiliser le théorème de Fubini et remarquer que $\{(x, y) \in \mathbf{R}^2 \mid x - r \leqslant y \leqslant x + r\} = \{(x, y) \in \mathbf{R}^2 \mid y - r \leqslant x \leqslant y + r\}$.]

On pose maintenant

$$Hf(x) = \sup_{r>0} A_r|f|(x) = \sup_{r>0} \frac{1}{2r} \int_{x-r}^{x+r} |f(y)|dy.$$

(3) Montrer que

$$Hf(x) = \sup_{\substack{r>0 \\ r\in\mathbf{Q}}} \frac{1}{2r} \int_{x-r}^{x+r} |f(y)|dy.$$

et que $x \mapsto Hf(x)$ est une fonction mesurable $\mathbf{R} \to [0, +\infty]$.

Montrer que que pour f, $g \in L^1(\mathbf{R})$ on a

$$H(f + g) \leqslant Hf + Hg.$$

(4) Montrer que si $g \in L^1(\mathbf{R})$ est continue, alors

$$\lim_{r\to 0} \frac{1}{2r} \int_{x-r}^{x+r} |g(y) - g(x)|dy = 0$$

pour tout $x \in \mathbf{R}$ et

$$\lim_{r\to 0} A_r g(x) = g(x).$$

(5) Montrer que si $f \in L^1(\mathbf{R})$ et $y > 0$, on a

$$\mu(\{x \mid |f(x)| > y\}) \leqslant \frac{\|f\|_1}{y}.$$

Pour $f \in L^1(\mathbf{R})$ et $y > 0$, on pose

$$E(f; y) = \{x \in \mathbf{R} \mid Hf(x) > y\}.$$

(6) Soit $f \in L^1(\mathbf{R})$ et $K \subset E(f; y)$ un compact quelconque. Montrer qu'il existe une suite finie de couples $(x_i, r_i) \in \mathbf{R}\times]0, +\infty[$, $i \leqslant n$, avec

$$r_1 \geqslant r_2 \geqslant \ldots \geqslant r_n, \text{ et } K \subset \bigcup_{i\leqslant n} [x_i - r_i, x_i + r_i],$$

tels que

$$\int_{x_i-r_i}^{x_i+r_i} |f(x)|dx > 2yr_i.$$

(7) Montrer que si $i \geqslant j$ et

$$[x_i - r_i, x_i + r_i] \cap [x_j - r_j, x_j + r_j] \neq \emptyset$$

alors

$$[x_i - r_i, x_i + r_i] \subset [x_j - 3r_j, x_j + 3r_j].$$

Montrer alors qu'il existe une sous-suite $(x_{i(j)}, r_{i(j)})$, $j \leqslant k$, telle que les intervalles $[x_{i(j)} - r_{i(j)}, x_{i(j)} + r_{i(j)}]$ sont deux à deux disjoints, et telle que

$$\mu(K) \leqslant 6 \sum_{j \leqslant k} r_{i(j)},$$

(où μ désigne la mesure de Lebesgue). [Indication : Choisir convenablement $i(1)$ puis les autres par récurrence en utilisant les questions précédentes.]

(8) En déduire que

$$\mu(K) \leqslant \frac{3}{y} \|f\|_1 \text{ et } \mu(E(f;y)) \leqslant \frac{3}{y} \|f\|_1.$$

(9) Soit $f \in L^1(\mathbf{R})$ et $g \in C_c(\mathbf{R})$. Pour tout $y > 0$, montrer que

$$\{x \in \mathbf{R} \mid \limsup_{r \to 0} |A_r f(x) - f(x)| > y\} \subset E(f-g; y/2) \cup \{x \mid |f(x) - g(x)| > \frac{y}{2}\}.$$

[Indication : Montrer que $\limsup_{r\to 0} |A_r f(x) - f(x)| \leqslant H(f-g)(x) + |f(x) - g(x)|$.] On rappelle que pour $F :]0, +\infty[\to \mathbf{R}$ on a

$$\limsup_{r \to 0} F(r) = \inf_{r>0} \sup_{0<s<r} F(s).$$

(10) En déduire pour tout n on a

$$\mu(\{x \mid \limsup_{r\to 0} |A_r f(x) - f(x)| > 1/n\}) = 0$$

[Indication : Appliquer ce qui précède à une suite de fonctions g_k approchant convenablement f.]

En déduire que la limite

$$\lim_{r \to 0} A_r f(x)$$

existe et est égale à $f(x)$ pour presque tout x. [Indication : Utiliser la caractérisation : $\lim u_n$ existe et est égale à u si et seulement si $\limsup |u_n - u| = 0$.]

(11) Comment interpréter ce résultat ?

Corrigé

Problème 1. (1) Si $f \in L^1(\mathbf{R}^d)$ est telle que sa transformée de Fourier

$$\hat{f}(t) = \int_{\mathbf{R}^d} f(x) e(-<x,t>) dx$$

est intégrable, alors pour presque tout $t \in \mathbf{R}^d$ on a

$$f(x) = \int_{\mathbf{R}^d} \hat{f}(t) e(<x,t>) dt$$

(2) Si f est une fonction continue périodique de période 1 alors la suites des sommes partielles de Fejér

$$\sigma_n(f) = \frac{1}{n}(S_0(f) + \cdots + S_{n-1}(f))$$

où

$$S_n(f)(x) = \sum_{|k| \leqslant n} c_k(f) e(kx) \text{ avec } c_k(f) = \int_0^1 f(x) e(-kx) dx$$

converge uniformémenent vers f.

Problème 2. (1) Si f, $g \in \Lambda_\alpha$ avec

$$|f(x) - f(y)| \leqslant C(f)|x-y|^\alpha \text{ et } |g(x) - g(y)| \leqslant C(g)|x-y|^\alpha$$

il vient pour $h = af + bg$ (a, $b \in \mathbf{C}$)

$$|h(x) - h(y)| \leqslant (|a|C(f) + |b|C(g))|x-y|^\alpha$$

donc $h \in \Lambda_\alpha$. Puisque $|x-y|^\alpha \to 0$ quand $x \to y$, Λ_α est inclus dans l'espace des fonctions périodiques continues.

Si f est continue périodique, pour $|x-y| \geqslant 1$ on a

$$|f(x) - f(y)| \leqslant 2\|f\|_\infty \leqslant 2\|f\|_\infty |x-y|^\alpha$$

donc toute fonction continue vérifie la condition Höldérienne d'ordre α pour les « points distants ». De sorte qu'il suffit de la vérifier pour $|x-y| < 1$ pour savoir si $f \in \Lambda_\alpha$. Comme $|x-y|^{\alpha'} \leqslant |x-y|^\alpha$ si $0 < \alpha \leqslant \alpha' \leqslant 1$ et $|x-y| < 1$, on a donc alors $\Lambda_{\alpha'} \subset \Lambda_\alpha$.

Enfin si f est de classe C^1 on a par le théorème de la valeur intermédiaire

$$|f(x) - f(y)| \leqslant \|f'\|_\infty |x-y|$$

donc $f \in \Lambda_1$ par définition, et $f \in \Lambda_\alpha$ pour tout $\alpha \in]0,1]$ d'après ce qui précède.

(2) Soit $h \in \mathbf{R}$ fixé et g la fonction

$$g(x) = \frac{1}{2}(f(x+h) - f(x-h))$$

qui est clairement 1-périodique. Comme f est continue, elle est dans $L^2(I)$, et donc $g \in L^2(I)$. D'après l'identité de Parseval pour g on a

$$\frac{1}{4}\int_0^1 |f(x+h) - f(x-h)|^2 dx = \sum_{k\in\mathbf{Z}} |c_k(g)|^2$$

avec

$$\begin{aligned} c_k(g) &= \int_0^1 g(x)e(-kx)dx = \frac{1}{2}\int_0^1 (f(x+h) - f(x-h))e(-kx)dx \\ &= c_k(f)\frac{e(kh) - e(-kh)}{2} = ic_k(f)\sin(2\pi kh), \end{aligned}$$

donc on a bien

$$\sum_{k\in\mathbf{Z}} |c_k(f)\sin 2\pi hk|^2 = \frac{1}{4}\int_0^1 |f(x+h) - f(x-h)|^2 dx.$$

(3) Si $C(f)$ est telle que $|f(x) - f(y)| \leqslant C(f)|x-y|^\alpha$ on a

$$|g(x)| = |f(x+h) - f(x-h)| \leqslant 2^\alpha C(f)|h|^\alpha$$

donc d'après la précédente question

$$\sum_{k\in\mathbf{Z}} |c_k(f)\sin 2\pi hk|^2 = \frac{1}{4}\int_0^1 |f(x+h) - f(x-h)|^2 dx \leqslant 2^{2(\alpha-1)}C(f)^2|h|^{2\alpha}$$

et $D(f) = 2^{2(\alpha-1)}C(f)$ convient.

(4) Soit $m \geqslant 0$ un entier, $h = 2^{-m-3}$ et k tel que $2^m \leqslant |k| < 2^{m+1}$. On a donc $\pi/4 \leqslant 2\pi|h|k \leqslant \pi/2$ d'où aussitôt $|\sin 2\pi hk|^2 \geqslant 1/2$.

(5) On a par positivité (on rajoute des termes positifs à la somme !) pour $h = 2^{-m-3}$ comme ci-dessus

$$\begin{aligned} \sum_{2^m \leqslant |k| < 2^{m+1}} |c_k(f)|^2 \leqslant 2 \sum_{2^m \leqslant |k| < 2^{m+1}} |c_k(f)\sin 2\pi hk|^2 \leqslant 2\sum_{k\in\mathbf{Z}} |c_k(f)\sin 2\pi hk|^2 \\ \leqslant 2D(f)|h|^{2\alpha} = 2^{1-6\alpha}D(f)2^{-2\alpha m} \end{aligned}$$

donc $E(f) = 2^{1-6\alpha}D(f)$ convient.

(6) D'après l'inégalité de Cauchy-Schwartz on a

$$\begin{aligned} \sum_{2^m \leqslant |k| < 2^{m+1}} |c_k(f)| &\leqslant 2^{m/2}\Big(\sum_{2^m \leqslant |k| < 2^{m+1}} |c_k(f)|^2\Big)^{1/2} \\ &\leqslant E(f)^{1/2}2^{m(1/2-\alpha)} \end{aligned}$$

et en réarrangeant la série (à termes positifs) sur $k \in \mathbf{Z}$ en ces segments « dyadiques » on trouve

$$\sum_{k\in\mathbf{Z}} |c_k(f)| \leqslant E(f)^{1/2} \sum_{m\geqslant 0} 2^{m(1/2-\alpha)}$$

qui converge si $1/2-\alpha<0$ c'est à dire $\alpha>1/2$. est convergente.

(7) Puisque la série des coefficients de Fourier converge absolument pour $1/2<\alpha\leqslant 1$, il est évident que la série de Fourier de f converge alors normalement, donc uniformément. Cela permet de montrer que les coefficients de Fourier de sa somme sont égaux à ceux de f par intégration terme à terme, et donc que cette somme est égale à f (partout car il s'agit de fonctions continues) par unicité des coefficients de Fourier.

Puisque $C^1(I)\subset\Lambda_\alpha$, on en déduit en particulier de nouveau le théorème de Dirichlet selon lequel la série de Fourier d'une fonction $f\in C^1(I)$ converge vers f.

Problème 3. Pour $f\in L^1(\mathbf{R})$, $x\in\mathbf{R}$ et $r>0$ on pose

$$A_r f(x) = \frac{1}{2r}\int_{x-r}^{x+r} f(y)dy.$$

et on note μ la mesure de Lebesgue sur $\mathbf{R}$.

(1) On a

$$A_r f(x) = \frac{1}{2r}\int_{\mathbf{R}} g(y,r)dy$$

où $g(y,r)=f(y)\chi_{[x-r,x+r]}(y)$. Comme $|g(y,r)|\leqslant|f(y)|$ qui est intégrable et est continue en r pour presque tout y fixé, la continuité de $A_r f(x)$ comme fonction de r est immédiate.

(2) Comme aussi

$$g(y,r)=f(y)\chi_{[x-r,x+r]}(y)=f(y)\chi_{[y-r,y+r]}(x)$$

la mesurabilité comme fonction de x pour r fixé provient du théorème de Fubini.

Soit

$$Hf(x)=\sup_{r>0} A_r|f|(x)=\sup_{r>0}\frac{1}{2r}\int_{x-r}^{x+r}|f(y)|dy.$$

(3) On a

$$Hf(x)=\sup_{\substack{r>0\\ r\in\mathbf{Q}}}\frac{1}{2r}\int_{x-r}^{x+r}|f(y)|dy.$$

car d'après la première question, pour x fixé, $r\mapsto A_r|f|(x)$ est une fonction continue de r. Comme pour $r\in\mathbf{Q}$ fixé, $x\mapsto A_r|f|(x)$ est mesurable par la seconde question, la borne supérieure sur $r\in\mathbf{Q}$, $r>0$ (ensemble dénombrable) l'est également.

Pour tout r on a trivialement $A_r(|f+g|)(x)\leqslant A_r|f|(x)+A_r|g|(x)$ donc en prenant la borne supérieure on a $H(f+g)(x)\leqslant Hf(x)+Hg(x)$ pour tout x.

(4) Soit $g\in L^1(\mathbf{R})$ continue, et $x\in\mathbf{R}$. Pour tout $\varepsilon>0$, il existe δ tel que si $|y|<\delta$ on a $|g(x)-g(y)|<\varepsilon$. Alors pour $r<\delta$, il vient

$$0\leqslant\frac{1}{2r}\int_{x-r}^{x+r}|g(y)-g(x)|dy\leqslant\varepsilon$$

ce qui montre par définition que

$$\lim_{r\to 0}\frac{1}{2r}\int_{x-r}^{x+r}|g(y)-g(x)|dy=0.$$

De plus on a alors

$$|A_r g(x)-g(x)|=\Big|\frac{1}{2r}\int_{x-r}^{x+r}(g(y)-g(x))dy\Big|\leqslant\frac{1}{2r}\int_{x-r}^{x+r}|g(y)-g(x)|dy\to 0$$

(5) On a $\mu(\{x\mid |f(x)|>y\})\leqslant y^{-1}\|f\|_1$ pour $f\in L^1$ et $y>0$ directement par l'inégalité de Markov-Bienaymé-Chebychev.

Soit $f \in L^1(\mathbf{R})$, $y > 0$ et $E(f;y) = \{x \in \mathbf{R} \mid Hf(x) > y\}$.

(6) Soit $K \subset E(f;y)$ compact. Pour tout $x \in K$, il existe par définition de $E(f;y)$ et de $H(f)$ un $r(x) > 0$ tel que

$$\int_{x-r(x)}^{x+r(x)} |f(x)|dx > 2yr(x). \tag{9.8}$$

Les intervalles ouverts $[x - r(x), x + r(x)]$ recouvrent K, donc par compacité on peut en extraire un sous-recouvrement fini. Si on ordonne ensuite les « rayons » $r(x_i)$ correspondant par ordre décroissant, on trouve la suite finie de couples (x_i, r_i) demandée.

(7) Si $i \geqslant j$ on a $r_i \leqslant r_j$ par construction, donc si $[x_i - r_i, x_i + r_i]$ rencontre $[x_j - r_j, x_j + r_j]$ on trouve

$$[x_i - r_i, x_i + r_i] \subset [x_j - 3r_j, x_j + 3r_j].$$

simplement par l'inégalité triangulaire.

On choisit $i(1) = 1$ puis par récurrence, si $i(1),\ldots,$ $i(j-1)$ ont été choisis de sorte que les conditions demandées soient satisfaites ont prend $i(j) > i(j-1)$ le plus petit indice i (s'il existe) tel que $[x_i - r_i, x_i + r_i]$ ne rencontre aucun des intervalles précédemment sélectionnés. (Si i n'existe pas, on arrête la construction de la suite).

On obtient donc des intervalles disjoints par construction. De plus pour tout $x \in K$, il existe i tel que $x \in [x_i - r_i, x_i + r_i]$ et il existe j tel que $r_i \leqslant r_{i(j)}$ et

$$[x_i - r_i, x_i + r_i] \cap [x_{i(j)} - r_{i(j)}, x_{i(j)} + r_{i(j)}] \neq \emptyset$$

(car sinon on aurait pu continuer la construction de la suite). Alors $x \in [x_{i(j)} - 3r_{i(j)}, x_{i(j)} + 3r_{i(j)}]$ d'après la question précédente. Donc on trouve

$$K \subset \bigcup_j [x_{i(j)} - 3r_{i(j)}, x_{i(j)} + 3r_{i(j)}] \text{ donc } \mu(K) \leqslant 6 \sum_{j \leqslant k} r_{i(j)}.$$

(8) Par (9.8) on a $r_{i(j)} \leqslant y^{-1}\|f\|_1/2$ donc $\mu(K) \leqslant 3y^{-1}\|f\|_1$. Comme cela vaut pour tout compact $K \subset E(f;y)$, la régularité de la mesure de Lebesgue implique $\mu(E(f;y)) \leqslant 3y^{-1}\|f\|_1$ aussi.

(9) Soit $f \in L^1(\mathbf{R})$ et $g \in C_c(\mathbf{R})$. On a

$$|A_r f(x) - f(x)| = |A_r(f-g)(x) + A_r g(x) - g(x) + g(x) - f(x)|$$

donc en prenant la limite pour $r \to 0$ il vient par la question (4)

$$\limsup_{r \to 0} |A_r f(x) - f(x)| \leqslant H(f-g)(x) + |f(x) - g(x)|.$$

Si le membre de gauche est $> y$, il faut que l'un au moins des deux termes de droite soit $> y/2$, et donc

$$\{x \in \mathbf{R} \mid \limsup_{r \to 0} |A_r f(x) - f(x)| > y\} \subset E(f-g; y/2) \cup \{x \mid |f(x) - g(x)| > \frac{y}{2}\}.$$

(10) Soit $n \geqslant 1$ et $y = 1/n$. On a

$$\mu(\{\limsup_{r \to 0} |A_r f - f| > y\}) \leqslant \mu(E(f-g; 1/2n)) + \mu(\{|f-g| > 1/2n\}) \leqslant 8n\|f-g\|_1$$

d'après les questions (5) et (8). Choisissant g dans une suite telle que $g_k \to f$ dans $L^1(\mathbf{R})$, on trouve que l'expression de droite tend vers 0 et donc la mesure en question est nulle.

Par définition de la limite, cela signifie que pour presque tout x on a

$$\lim_{r \to 0} A_r f(x) = f(x).$$

(11) On a donc

$$\lim_{r \to 0} \frac{1}{2r} \int_{x-r}^{x+r} f(y)dy = f(x)$$

pour presque tout $x \in \mathbf{R}$: il s'agit d'une généralisation du théorème de dérivation d'une intégrale indéfinie.

Examen de rattrapage – Bordeaux, Septembre 2002

Énoncé

Problème 1. (Questions de cours)
(1) Énoncer le théorème central-limite.
(2) Énoncer l'inégalité de Hölder.

Problème 2. Soit $X = [0,1]$, λ la mesure de Lebesgue sur X. Soit μ une mesure borélienne finie sur X telle que pour $E \subset X$ on a

$$\lambda(E) = 0 \text{ implique } \mu(E) = 0. \tag{9.9}$$

(1) Pour tout borélien $E \subset X$ on note

$$\nu(E) = \lambda(E) + \mu(E).$$

Montrer que ν est une mesure borélienne finie sur X et que toute fonction mesurable bornée sur X est dans $L^2(X,\nu)$.

(2) Montrer que pour toute fonction mesurable bornée $f : X \to \mathbf{C}$ on a l'inégalité

$$\Big|\int_X f(x)d\mu(x)\Big| \leqslant C\|f\|_2$$

où *la norme* $\|\cdot\|_2$ *est la norme pour l'espace* $L^2(X,\nu)$ et $C \geqslant 0$ est une constante. [Indication : Utiliser l'inégalité de Cauchy-Schwarz.]

(3) Montrer qu'il existe $h \in L^2(X,\nu)$ telle que

$$\int_X f(x)d\mu(x) = \int_X f(x)h(x)d\nu(x)$$

pour toute fonction $f \in L^2(X,\nu)$. [Indication : Utiliser la structure hilbertienne.]

(4) Pour tout borélien $A \subset X$, montrer que

$$0 \leqslant \int_A h(x)d\nu(x) \leqslant \nu(A).$$

En déduire que h est à valeurs réelles puis que $0 \leqslant h(x) \leqslant 1$, ν-presque partout. [Indication : Procéder par contraposition en considérant, par exemple, un ensemble hypothétique A tel que $\nu(A) > 0$ et $\mathrm{Im}(h(x)) > 0$ pour $x \in A$, etc...]

(5) Si g est mesurable et bornée, montrer que

$$\int_X g(x)d\mu(x) = \int_X g(x)h(x)d\lambda(x) + \int_X g(x)h(x)d\mu(x).$$

(6) Soit $Z = \{x \mid h(x) = 1\}$. Montrer à l'aide de la question précédente que $\lambda(Z) = 0$ et en déduire que $\mu(Z) = 0$.

(7) Montrer que pour $n \geqslant 1$ on a

$$\int_X g(x)d\mu(x) = \int_X g(x)(h(x) + h(x)^2 + \cdots + h(x)^n)d\lambda(x) + \int_X g(x)h(x)^n d\mu(x),$$

puis que

$$\int_X g(x)h(x)^n d\mu(x) \to 0$$

et enfin que

$$\int_X g(x)d\mu(x) = \int_X g(x)f(x)d\lambda(x) \tag{9.10}$$

avec $f = h(1-h)^{-1}$ sur $X - Z$ et $f = 0$ sur Z.

(8) Donner un exemple (très simple!) de mesure borélienne finie sur X ne vérifiant pas (9.9) et telle qu'il n'existe pas de fonction f pour laquelle (9.10) soit vérifiée pour toute fonction mesurable bornée.

Problème 3. On rappelle que les notations $L^1(I)$, $C(I)$, etc, se réfèrent aux fonctions périodiques de période 1 sur $\mathbf{R}$ pour lesquelles les coefficients de Fourier sont donnés par

$$c_n(f) = \int_0^1 f(t)e(-nt)dt$$

pour $n \in \mathbf{Z}$, où $e(x) = e^{2i\pi x}$ est 1-périodique.

On note $\mathcal{P}$ l'ensemble des fonctions $f \in L^1(I)$ dont les coefficients de Fourier vérifient

$$c_n(f) \geqslant 0 \text{ pour tout } n \in \mathbf{Z}.$$

(1) Montrer que $e_k(x) = e(kx)$, pour $k \in \mathbf{Z}$, est un exemple de telle fonction.

(2) Montrer que si $f \in \mathcal{P}$ on a $f(x) = f^*(x)$ presque partout, où

$$f^*(x) = \overline{f(-x)}.$$

(3) Soit $f \in \mathcal{P}$. Pour tout polynôme trigonométrique u donné par

$$u(x) = \sum_{n=-N}^{N} c_n(u)e(nx),$$

montrer que

$$\int_0^1 \int_0^1 f(y-x)u(x)\overline{u(y)}dxdy = \sum_{n=-N}^{N} c_n(f)|c_n(u)|^2.$$

En déduire que pour toute fonction continue périodique $u \in C(I)$ on a

$$\int_0^1 \int_0^1 f(y-x)u(x)\overline{u(y)}dxdy \geqslant 0.$$

(4) Soit $g \in L^1(I)$. Montrer que $g \star g^* \in \mathcal{P}$, où $\star$ désigne le produit de convolution de fonctions périodiques.

La suite du problème est indépendante de ce qui précède.

Soit $f \in \mathcal{P}$, telle que de plus il existe $\delta > 0$ et $M \geqslant 0$ de sorte que

$$|f(x)| \leqslant M \text{ pour presque tout } x \in [-\delta, \delta].$$

(5) Pour $n \geqslant 1$, montrer que la n-ème somme partielle de Cesaro de f en 0 est donnée par

$$\begin{aligned}\sigma_n(f)(0) &= \sum_{|k| \leqslant n} c_k(f)\Big(1 - \frac{|k|}{n}\Big) \\ &= \int_{-1/2}^{1/2} f(t)F_n(t)dt\end{aligned}$$

où

$$F_n(t) = \frac{1}{n}\Big(\frac{\sin^2(n\pi t)}{\sin^2 \pi t}\Big).$$

(6) En choisissant une fonction f particulière, calculer l'intégrale

$$\int_{-1/2}^{1/2} F_n(t)dt.$$

(7) Montrer que la suite $(\sigma_n(f)(0))$, $n \geqslant 1$, est une suite positive *bornée*. [Indication : Pour majorer $\sigma_n(f)(0)$, séparer l'intégrale ci-dessus en la partie $[-\delta, \delta]$ et le reste, et majorer séparément en utilisant les propriétés de $F_n(t)$ et l'hypothèse sur f.]

(8) Montrer que la série

$$\sum_n c_n(f)$$

converge absolument et en déduire que f est presque partout somme de sa série de Fourier. [Indication : Utiliser $\sigma_{2n}(f)(0)$ pour majorer la somme partielle d'ordre n de la série.]

Corrigé

Problème 1. (1) Soit (Ω, Σ, P) un espace probabilisé, (X_n) une suite de variables aléatoires réelles indépendantes et uniformément distribuées, telles que $X_n \in L^2(\Omega)$ et $E(X_n) = 0$. Soit $\sigma = V(X_n)$ la variance des X_n. Alors

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \xrightarrow{\text{loi}} \mu_{0,\sigma}$$

la loi normale d'espérance 0 et variance σ.

(2) Soit $(X, \mathcal{M}, \mu)$ un espace mesuré, $p \in [1, +\infty]$, $f \in L^p(\mu)$, q tel que $1/p + 1/q = 1$, $g \in L^q(\mu)$. Alors

$$\|fg\|_1 \leqslant \|f\|_p \|g\|_q,$$

autrement dit (si p, $q < +\infty$)

$$\int_X |f(x)g(x)| d\mu(x) \leqslant \Big(\int_X |f(x)|^p d\mu(x)\Big)^{1/p} \Big(\int_X |g(x)|^q d\mu(x)\Big)^{1/q}$$

et si $p = +\infty$

$$\int_X |f(x)g(x)| d\mu(x) \leqslant \|f\|_\infty \int_X |g(x)| d\mu(x).$$

Problème 2. Soit $X = [0, 1]$, μ une mesure borélienne finie sur X telle que pour $E \subset X$ on a

$$\lambda(E) = 0 \text{ implique } \mu(E) = 0,$$

où λ est la mesure de Lebesgue.

(1) Soit $\nu(E) = \lambda(E) + \mu(E)$ pour $E \subset X$ borélien. Il s'agit clairement d'une mesure borélienne et comme $\nu(X) = 1 + \mu(X) < +\infty$, c'est une mesure finie. Par conséquent si f est mesurable et bornée on a

$$\int_X |f(x)|^2 d\nu(x) \leqslant \|f\|_\infty^2 \nu(X) < +\infty$$

donc $f \in L^2(X, \nu)$.

(2) Soit f mesurable et bornée sur X. On a

$$\begin{aligned}\Big|\int_X f(x) d\mu(x)\Big| &\leqslant \int_X |f(x)| d\mu(x) \leqslant \int_X |f(x)| d\mu(x) + \int_X |f(x)| d\lambda(x) \\ &= \int_X |f(x)| d\nu(x) \leqslant \nu(X)^{1/2} \|f\|_2\end{aligned}$$

par l'inégalité de Cauchy-Schwarz appliquée dans $L^2(X, \nu)$. On peut donc prendre $C = \nu(X)^{1/2}$.

(3) L'inégalité précédente établit que la forme linéaire

$$T(f) = \int_X f(x) d\mu(x)$$

définie sur l'espace vectoriel V des fonctions mesurables bornées est continue et de norme $\leqslant C$ lorsque celui-ci est muni de la norme L^2 pour ν. Comme V est dense dans $L^2(X, \nu)$ d'après les résultats du cours (par exemple, les fonctions continues sont dans V et sont denses dans $L^2(X, \nu)$), il s'ensuit que T se prolonge de façon unique en une forme linéaire continue $\tilde{T}$ sur

$L^2(X,\nu)$. Or ce dernier est un espace de Hilbert, donc toute forme linéaire est de la forme $f \mapsto <f,g>$ pour une certaine $g \in L^2(X,\nu)$. Il existe donc une fonction g telle que

$$\tilde{T}(f) = <f,g> = \int_X f(x)\overline{g(x)}d\nu(x)$$

pour toute $f \in L^2(X,\nu)$. Prenant $h = \bar{g}$, on a le résultat demandé

$$\int_X f(x)d\mu(x) = \int_X f(x)h(x)d\nu(x).$$

(4) Soit $A \subset X$ un borélien. En appliquant la formule ci-dessus à $f = \chi_A$, qui est mesurable bornée, il vient

$$\int_X \chi_A(x)d\mu(x) = \int_X \chi_A(x)h(x)d\nu(x) = \int_A h(x)d\nu(x).$$

Mais de plus

$$\int_X \chi_A(x)d\mu(x) = \mu(A) \leqslant \nu(A) \text{ (par définition)},$$

donc

$$0 \leqslant \int_A h(x)d\nu(x) \leqslant \nu(A). \tag{9.11}$$

Pour en déduire que h est ν-presque partout réelle à valeurs dans $[0,1]$, posons $Z_1 = \{x \mid \mathrm{Im}(h(x)) > 0\}$, $Z_2 = \{x \mid \mathrm{Im}(h(x)) < 0\}$. Si $\nu(Z_1) > 0$, il vient pour $A = Z_1$

$$\mathrm{Im}\Big(\int_A h(x)d\nu(x)\Big) > 0,$$

contredisant (9.11) puisque ce nombre devrait être un réel entre 0 et $\nu(A)$. De même avec $A = Z_2$ on trouve $\nu(Z_2) = 0$, et donc h est ν-presque partout réelle. Soit alors $Z_3 = \{x \mid h(x) < 0\}$ et $Z_4 = \{x \mid h(x) > 1\}$; par un raisonnement similaire on trouve $\nu(Z_3) = \nu(Z_4) = 0$, d'où le résultat.

(5) Soit g mesurable et bornée. On a

$$\int_X g(x)d\mu(x) = \int_X g(x)h(x)d\nu(x) = \int_X g(x)h(x)d\mu(x) + \int_X g(x)h(x)d\lambda(x) \tag{9.12}$$

par définition même de $\nu = \mu + \lambda$.

(6) Soit $Z = \{x \mid h(x) = 1\}$. On applique la formule ci-dessus à $g = \chi_Z$. Il vient

$$\mu(Z) = \int_Z h(x)d\mu(x) + \int_Z h(x)d\lambda(x) = \mu(Z) + \lambda(Z),$$

donc $\lambda(Z) = 0$. *Par hypothèse* sur μ, on a alors $\mu(Z) = 0$. (C'est le seul moment où l'hypothèse est utilisée.)

(7) On a (9.12) pour g mesurable bornée. Mais gh est également mesurable bornée donc on peut appliquer la formule récursivement à l'intégrale de gh par rapport à μ : on trouve

$$\int_X g(x)d\mu(x) = \int_X g(x)h^2(x)d\mu(x) + \int_X g(x)h^2(x)d\lambda(x) + \int_X g(x)h(x)d\lambda(x),$$

puis en répétant on arrive à

$$\int_X g(x)d\mu(x) = \int_X g(x)(h(x) + h(x)^2 + \cdots + h(x)^n)d\lambda(x) + \int_X g(x)h(x)^n d\mu(x),$$

pour tout $n \geqslant 1$.

Comme $h(x) \in [0,1[$, μ-presque partout d'après la question précédente, et que g est bornée, on a

$$g(x)h(x)^n \to 0 \ \mu\text{-presque partout},$$

et comme $|g(x)h(x)^n| \leqslant |g(x)|$ qui est μ-intégrable, le théorème de convergence dominée montre que

$$\lim_{n\to+\infty} \int_X g(x)h(x)^n d\mu(x) = 0.$$

De plus, puisque $h(x) \in [0,1[$, λ-presque partout également, on a

$$h(x) + \cdots + h(x)^n \to \frac{h(x)}{1-h(x)} = f(x)$$

λ-presque partout. Cette suite est de plus croissante. Si g est positive, on en déduit donc d'après le théorème de convergence monotone que

$$\int_X g(x)(h(x) + h(x)^2 + \cdots + h(x)^n) d\lambda(x) \to \int_X g(x)f(x)d\lambda(x).$$

En écrivant $g = g^+ - g^-$ et en appliquant cette dernière formule à g^+ et g^- séparément, on en déduit finalement que

$$\int_X g(x)d\mu(x) = \int_X g(x)f(x)d\lambda(x).$$

(8) Si μ est la mesure de Dirac en 0, on a $\mu(\{0\}) = 1$ bien que $\lambda(\{0\}) = 0$. En prenant $g = \chi_{\{0\}}$, on voit qu'aucune fonction f ne peut donner l'égalité ci-dessus pour μ.

Problème 3. Soit $\mathcal{P}$ l'ensemble des fonctions $f \in L^1(I)$ dont les coefficients de Fourier vérifient

$$c_n(f) \geqslant 0 \text{ pour tout } n \in \mathbf{Z}.$$

Attention, ce n'est pas un espace vectoriel ($-f \notin \mathcal{P}$ si $f \neq 0$ est dans $\mathcal{P}$, par exemple) !

(1) Pour $f = e_k$, on a $c_n(f) = 0$ si $n \neq k$ et $c_k(f) = 1$, donc $e_k \in \mathcal{P}$.

(2) Soit $f \in \mathcal{P}$ et $f^*(x) = \overline{f(-x)}$. On calcule les coefficients de Fourier f^* en utilisant le fait que $\overline{e(x)} = e(-x)$:

$$\begin{aligned} c_n(f^*) &= \int_0^1 f^*(x)e(-nx)dx = \int_0^1 \overline{f(-x)}e(-nx)dx \\ &= \overline{\int_0^1 f(-x)e(nx)dx} = \overline{\int_0^1 f(x)e(-nx)dx} = \overline{c_n(f)} \end{aligned}$$

mais $c_n(f)$ est réel donc $c_n(f^*) = c_n(f)$. Or on sait que l'application « coefficients de Fourier » est injective, et donc $f = f^*$ dans $L^1(I)$, c'est à dire presque partout.

(3) Soit

$$u(x) = \sum_{n=-N}^{N} c_n(u)e(nx),$$

on a alors

$$\begin{aligned} \int_0^1\int_0^1 f(y-x)u(x)\overline{u(y)}dxdy &= \int_0^1\int_0^1 f(y-x)\sum_n c_n(u)e(nx)\sum_m \overline{c_m(u)}e(-my)dxdy \\ &= \sum_{n,m} c_n(u)\overline{c_m(u)} \int_0^1\int_0^1 f(y-x)e(nx-my)dxdy. \end{aligned}$$

En intégrant d'abord par rapport à x on trouve

$$\int_0^1 f(y-x)e(nx-my)dx = e((n-m)y)c_n(f).$$

et l'intégrale sur y est nulle sauf si $n = m$, auquel cas elle vaut 1 (orthonormalité), d'où

$$\int_0^1\int_0^1 f(y-x)u(x)\overline{u(y)}dxdy = \sum_n c_n(f)|c_n(u)|^2.$$

En particulier on a

$$\int_0^1 \int_0^1 f(y-x)u(x)\overline{u(y)}dxdy \geqslant 0$$

pour tout polynôme trigonométrique. Soit maintenant u une fonction continue périodique quelconque. Alors u est limite uniforme sur $[0,1]$ de polynômes trigonométriques u_n. Comme f est intégrable, on vérifie aussitôt par le théorème de convergence dominée que

$$\int_0^1 \int_0^1 f(y-x)u(x)\overline{u(y)}dxdy = \lim_{n\to+\infty} \int_0^1 \int_0^1 f(y-x)u_n(x)\overline{u_n(y)}dxdy$$

et la limite est donc positive.

(4) Soit $g \in L^1(I)$. On sait que $c_n(g \star g^*) = c_n(g)c_n(g^*)$. Mais $c_n(g^*) = \overline{c_n(g)}$ (même calcul que pour la seconde question), donc

$$c_n(g \star g^*) = |c_n(g)|^2 \geqslant 0.$$

Pour la suite on fixe $f \in \mathcal{P}$ telle que $|f(x)| \leqslant M$ pour presque tout $x \in [-\delta, \delta]$ avec $\delta > 0$, $M \geqslant 0$.

(5) On sait par le cours (ou en recalculant !) que pour tout x on a

$$\sigma_n(f)(x) = \sum_{|k|\leqslant n} c_k(f)\Big(1 - \frac{|k|}{n}\Big)e(nx)$$

donc pour $x = 0$ il vient

$$\sigma_n(f)(0) = \sum_{|k|\leqslant n} c_k(f)\Big(1 - \frac{|k|}{n}\Big).$$

Toujours par le cours ou en recalculant, on a

$$\sigma_n(f)(x) = \int_0^1 f(t)F_n(x-t)dt$$

avec

$$F_n(t) = \frac{1}{n}\Big(\frac{\sin^2(n\pi t)}{\sin^2 \pi t}\Big).$$

Pour $x = 0$, observant que F_n est pair et 1-périodique, il vient

$$\sigma_n(f)(0) = \int_{-1/2}^{1/2} f(t)F_n(t)dt.$$

(6) En prenant $f = 1$, on a $\sigma_n(f)(0) = 1$ donc

$$\int_{-1/2}^{1/2} F_n(t)dt = 1.$$

(7) On écrit $J = [-1/2, -\delta] \cup [\delta, 1/2]$ et

$$\int_{-1/2}^{1/2} f(t)F_n(t)dt = \int_J f(t)F_n(t)dt + \int_{-\delta}^{\delta} f(t)F_n(t)dt$$

puis on majore ($F_n \geqslant 0$)

$$\Big|\int_{-1/2}^{1/2} f(t)F_n(t)dt\Big| \leqslant \int_J |f(t)|F_n(t)dt + M\int_{-\delta}^{\delta} F_n(t)dt$$

La seconde intégrale est

$$\int_{-\delta}^{\delta} F_n(t)dt \leqslant \int_{-1/2}^{1/2} F_n(t)dt = 1,$$

et sur J on a

$$F_n(t) \leqslant \frac{1}{n}\frac{1}{\sin^2 \pi\delta}$$

donc
$$\int_J |f(t)|F_n(t)dt \leqslant \frac{1}{n\sin^2\pi\delta}\|f\|_1.$$
Finalement pour tout $n \geqslant 1$ on trouve
$$|\sigma_n(f)(0)| \leqslant M + \frac{1}{n\sin^2\pi\delta}\|f\|_1,$$
qui est borné.

(8) Pour tout k tel que $|k| \leqslant n$, on a par positivité
$$c_k(f) \leqslant 2c_k(f)\Big(1 - \frac{|k|}{2n}\Big)$$
donc
$$\sum_{k=-n}^{n} c_k(f) \leqslant 2\sum_{k=-n}^{n} c_k(f)\Big(1 - \frac{|k|}{2n}\Big) \leqslant 2\sigma_{2n}(f)(0).$$

D'après la question précédente, c'est une suite bornée. Comme $c_n(f) \geqslant 0$, cela signifie que la série $\sum c_n(f)$ à termes positifs a des sommes partielles bornées, c'est à dire qu'elle converge (absolument).

Il en découle que la série de Fourier de f converge et que f est presque partout somme de sa série de Fourier.

Partiel – Bordeaux, Mars 2003

Énoncé

Problème 1. (Questions de cours / applications directes du cours)

(1) Quelles sont les conditions qui définissent une tribu $\mathcal{M}$ sur un ensemble X, et une mesure μ sur $\mathcal{M}$?

(2) Soit $S = \{z \in \mathbf{C} \mid |z| = 1\}$ le cercle unité dans $\mathbf{C}$ et

$$e : [0,1] \to S$$

l'application définie par $e(t) = e^{2i\pi t}$. On considère la mesure borélienne $\mu = e_*(\lambda)$, où λ est la mesure de Lebesgue. Montrer que μ est une mesure de probabilité sur S et calculer

$$\int_S z^n d\mu(z) \text{ pour } n \in \mathbf{Z}.$$

(3) Quelle est la définition de la norme L^p sur $(X, \mathcal{M}, \mu)$, $p \in [1, +\infty[$, et celle de la norme L^∞ ?

Problème 2. Soit J_0 la fonction définie par

$$J_0(x) = \frac{1}{\pi}\int_0^\pi \cos(x \sin\theta) d\theta$$

pour $x \in \mathbf{R}$, $d\theta$ désignant ici la mesure de Lebesgue sur $[0, \pi]$.

(1) Justifier *précisément* que la fonction J_0 est bien définie et que J_0 est continue et bornée sur $\mathbf{R}$.

(2) Calculer $\|J_0\|_\infty$ et montrer que $J_0(x) \neq 0$ si $|x| \leqslant \pi/2$.

(3) Montrer que J_0 est une fonction C^∞ sur $\mathbf{R}$ et que pour $j \geqslant 0$ et $x \in \mathbf{R}$ on a

$$J_0^{(j)}(x) = \begin{cases} \dfrac{(-1)^{j/2}}{\pi}\displaystyle\int_0^\pi (\sin\theta)^j \cos(x\sin\theta) d\theta & \text{si } j \text{ est pair} \\ \dfrac{(-1)^{(j+1)/2}}{\pi}\displaystyle\int_0^\pi (\sin\theta)^j \sin(x\sin\theta) d\theta & \text{si } j \text{ est impair} \end{cases}$$

(on justifiera l'existence de ces intégrales).

(4) Montrer que J_0 est solution de l'équation différentielle

$$xJ_0''(x) + J_0'(x) + xJ_0(x) = 0.$$

[<u>Indication</u> : On pourra faire une intégration par partie.]

(5) Montrer que le développement de Taylor de J_0 en 0 est donné par

$$\sum_{k\geqslant 0} \frac{J_0^{(k)}(0)}{k!} x^k = \sum_{m\geqslant 0} \frac{(-1)^m}{(m!)^2} (x/2)^{2m}.$$

[<u>Indication</u> : Calculer par récurrence

$$\int_0^\pi \sin(\theta)^{2j} d\theta$$

pour $j \geqslant 0$ en écrivant $(\sin\theta)^{2j+2} = (\sin\theta)^{2j} - (\cos\theta)^2(\sin\theta)^{2j}$.]

On admet que pour tout $x \in \mathbf{R}$, on a

$$J_0(x) = \sum_{m \geqslant 0} \frac{(-1)^m}{(m!)^2}(x/2)^{2m}.$$

(6) Montrer que pour tout $a > 0$ la fonction $f(x) = e^{-ax}J_0(x)$ est dans $L^1([0, +\infty[)$ (pour la mesure de Lebesgue notée dx) et que pour $a > 1$ on a

$$\int_0^{+\infty} e^{-ax}J_0(x)dx = \frac{1}{\sqrt{a^2+1}}.$$

[Indication : Montrer que

$$\int_0^{+\infty} e^{-ax}x^{2m}dx = a^{-2m-1}(2m)!,$$

et utiliser (sans démonstration) la formule du binôme

$$\frac{1}{\sqrt{1+h}} = \sum_{k \geqslant 0} \frac{(-1)^k(2k)!}{2^{2k}(k!)^2}h^k$$

pour $|h| < 1$.]

(7) Question hors barême : En utilisant le théorème de Fubini, montrer que la formule du (6) est valide pour tout $a > 0$.

Note : Dans tout cet exercice, la correction prendra particulièrement en compte la précision des détails des justifications effectuées.

Problème 3. On note $L^p(\mathbf{R})$ l'espace L^p pour la mesure de Lebesgue $d\lambda$ sur $\mathbf{R}$, et pour $p \in [1, +\infty]$, $q \in [1, +\infty]$ est le nombre tel que $\frac{1}{p} + \frac{1}{q} = 1$.

(1) Soit $f \in L^p([0, +\infty[)$ avec $p \in [1, +\infty]$. Montrer que

$$F(x) = \int_{[0,x]} f(t)d\lambda(t)$$

est bien définie et que F est une fonction continue sur $[0, +\infty[$ telle que $F(0) = 0$. [Indication : On pourra montrer que F est continue sur $[a, b] \subset [0, +\infty[$ pour tout intervalle fini.]

Dans toute la suite, on considère $f \in L^p([0, +\infty[)$ avec $p \in]1, +\infty[$.

(2) Montrer que pour $x \geqslant 0$, on a

$$\lim_{h \to 0} \int_{[x,x+h]} |f(t)|^p d\lambda(t) = 0.$$

(3) À l'aide de la question précédente, montrer que pour tout $x \geqslant 0$, on a

$$\lim_{h \to 0} \frac{F(x+h) - F(x)}{h^{1/q}} = 0.$$

En déduire

$$(*) \qquad \lim_{x \to 0} \frac{F(x)}{x^{1/q}} = 0$$

[Indication : Majorer $F(x+h) - F(x)$ à l'aide de l'inégalité de Hölder.]

(4) Montrer que pour tout $\varepsilon > 0$, il existe $A_\varepsilon \geqslant 0$ tel que

$$\int_{[A_\varepsilon, +\infty[} |f(t)|^p d\lambda(t) \leqslant \varepsilon^p.$$

(5) En déduire que

$$(**) \qquad \lim_{x \to +\infty} \frac{F(x)}{x^{1/q}} = 0.$$

[Indication : Pour $x \geqslant A_\varepsilon$, écrire $F(x) = F(A_\varepsilon) + \int_{[A_\varepsilon,x]} f(t)d\lambda(t)$.]

(6) Montrer sur un exemple que $(*)$ et $(**)$ ne sont pas toujours vrais si l'on choisit $p = \infty$.

Corrigé

Problème 1. (1) Une tribu $\mathcal{M}$ sur un ensemble X est un ensemble de parties $Y \subset X$ tel que

– On a $\emptyset \in \mathcal{M}$ et $X \in \mathcal{M}$;
– Si $Y \in \mathcal{M}$, alors $X - Y \in \mathcal{M}$;
– Si $Y_n \in \mathcal{M}$, $n \geqslant 1$, est une suite d'ensembles dans $\mathcal{M}$, alors

$$\bigcup_{n\geqslant 1} Y_n \in \mathcal{M} \text{ et } \bigcap_{n\geqslant 1} Y_n \in \mathcal{M}.$$

Une mesure μ sur $\mathcal{M}$ est une application $\mu\,:\,\mathcal{M} \to [0,+\infty]$ telle que $\mu(\emptyset) = 0$ et si $Y_n \in \mathcal{M}$ pour $n \geqslant 1$, avec $Y_n \cap Y_m = \emptyset$ si $n \neq m$, alors

$$\mu\Bigl(\bigcup_{n\geqslant 1} Y_n\Bigr) = \sum_{n\geqslant 1} \mu(Y_n).$$

(2) Soit $S = \{z \in \mathbf{C} \mid |z| = 1\}$ le cercle unité dans $\mathbf{C}$ et

$$e\,:\,[0,1] \to S$$

l'application définie par $e(t) = e^{2i\pi t}$. On considère la mesure borélienne $\mu = e_*(\lambda)$, où λ est la mesure de Lebesgue. Par définition, μ est une mesure sur S pour la tribu borélienne puisque f est continue. De plus

$$\mu(S) = e_*(\lambda)(S) = \lambda(e^{-1}(S)) = \lambda([0,1]) = 1$$

donc μ est une mesure de probabilité.

On sait de plus que pour toute fonction mesurable $f\,:\,S \to \mathbf{C}$ on a

$$\int_S f(z)d\mu(z) = \int_{[0,1]} f(e^{2i\pi t})d\lambda(t)$$

d'où

$$\int_S z^n d\mu(z) = \int_{[0,1]} e^{2i\pi n t} d\lambda(t) = \begin{cases} \Bigl[\frac{1}{2i\pi n} e^{2i\pi n t}\Bigr]_0^1 = 0 & \text{si } n \neq 0; \\ 1 & \text{si } n = 0. \end{cases}$$

(3) La norme L^p, $1 \leqslant p < +\infty$ d'une fonction f sur X est

$$\|f\|_p = \Bigl(\int_X |f|^p d\mu\Bigr)^{1/p}$$

et la norme L^∞ est

$$\|f\|_\infty = \inf\{M \geqslant 0 \mid \mu(\{x \mid |f(x)| > M\}) = 0\}.$$

Problème 2. Soit J_0 la fonction définie par

$$J_0(x) = \frac{1}{\pi}\int_0^\pi \cos(x\sin\theta)d\theta$$

pour $x \in \mathbf{R}$, $d\theta$ désignant ici la mesure de Lebesgue sur $[0,\pi]$.

(1) La fonction J_0 est une intégrale à paramètre $J_0(x) = \int h(x,\theta)d\theta$ avec

$$h(x,\theta) = \pi^{-1}\cos(x\sin\theta).$$

.

D'après le cours, elle est définie et continue en $x_0 \in \mathbf{R}$ si : (i) il existe une fonction $g\,:\,[0,\pi \to [0,+\infty[$, intégrable, telle que $|h(x,\theta)| \leqslant g(\theta)$ pour tout (x,θ), et (ii) pour presque tout $\theta \in [0,\pi]$, la fonction $x \mapsto h(x,\theta)$ est continue en x_0.

Ici, vu la formule, h est continue sur $\mathbf{R} \times [0,\pi]$ entier, donc la condition (ii) est vérifiée. Pour vérifier la première, on majore

$$|h(x,\theta)| = \pi^{-1}|\cos(x\sin\theta)| \leqslant \pi^{-1} \tag{9.13}$$

et on remarque que la fonction constante $1/\pi$ est évidemment indépendante de x et intégrable sur $[0, \pi]$ pour la mesure de Lebesgue.

Les conditions s'appliquant en tout x_0, on voit que J_0 est bien continue sur $\mathbf{R}$ entier. De plus en intégrant l'inégalité (9.13) on trouve

$$|J_0(x)| \leqslant \pi^{-1} \int_0^\pi d\theta = 1.$$

(2) Le dernier calcul montre que $\|J_0\|_\infty \leqslant 1$. De plus on a $J_0(0) = 1$ car $\cos(0) = 1$. Par continuité il existe pour tout $\varepsilon > 0$ un réel $\delta > 0$ tel que $J_0(x) \geqslant 1-\varepsilon$ pour $|x| \leqslant \delta$. Cela montre que

$$\lambda(\{x \mid |J_0(x)| \geqslant 1 - \varepsilon\}) \geqslant \lambda(]-\delta, \delta[) = 2\delta > 0.$$

Par définition de $\|J_0\|_\infty$ on a donc montré que $\|J_0\|_\infty \geqslant 1 - \varepsilon$ pour tout $\varepsilon > 0$, donc $\|J_0\|_\infty \geqslant 1$ en faisant $\varepsilon \to 0$. En définitive, on a donc $\|J_0\|_\infty = 1$.

Si $|x| \leqslant \pi/2$, on a $|x \sin\theta| \leqslant |x| \leqslant \pi/2$ et donc $h(x, \theta) \geqslant 0$ pour tout $\theta \in [0, \pi]$. De plus $|x \sin\theta| \leqslant |x|\sqrt{2}/2$ pour $0 \leqslant \theta \leqslant \pi/4$, ce qui donne

$$J_0(x) \geqslant \frac{1}{\pi} \int_0^{\pi/4} \cos(x\sqrt{2}/2) d\theta = \frac{1}{4} \cos(x\sqrt{2}/2) > 0.$$

(3) Montrons par récurrence sur $k \geqslant 0$ que J_0 est C^k sur $\mathbf{R}$ et vérifie

$$J_0^{(k)}(x) = \begin{cases} \dfrac{(-1)^{k/2}}{\pi} \displaystyle\int_0^\pi (\sin\theta)^k \cos(x \sin\theta) d\theta & \text{si } k \text{ est pair} \\ \dfrac{(-1)^{(k+1)/2}}{\pi} \displaystyle\int_0^\pi (\sin\theta)^k \sin(x \sin\theta) d\theta & \text{si } k \text{ est impair} \end{cases}$$

C'est vrai pour $k = 0$ d'après la question 1. L'admettant pour k, par exemple $k = 2m$ pair, on a

$$J_0^{(k)}(x) = (-1)^m \int_0^\pi h_1(x, \theta) d\theta \text{ avec } h_1(x, \theta) = \frac{1}{\pi} (\sin\theta)^{2m} \cos(x \sin\theta).$$

D'après le cours, cette fonction est dérivable sur $\mathbf{R}$ si h_1 est dérivable par rapport à x pour tout θ fixé et s'il existe $g \geqslant 0$ et $g_1 \geqslant 0$ intégrables sur $[0, \pi]$ telles que $|h_1(x, \theta)| \leqslant g(\theta)$ et $|\frac{d}{dx} h_1(x, \theta)| \leqslant g_1(\theta)$. Or ici on a immédiatement

$$|h_1(x, \theta)| \leqslant \frac{1}{\pi}$$
$$\frac{d}{dx} h_1(x, \theta) = -\frac{1}{\pi} (\sin\theta)^{2m+1} \sin(x \sin\theta) \text{ donc } |\frac{d}{dx} h_1(x, \theta)| \leqslant \frac{1}{\pi}$$

d'où la dérivabilité de $J_0^{(k)}$ avec

$$J_0^{(k+1)} = \frac{(-1)^{m+1}}{\pi} \int_0^\pi (\sin\theta)^{2m+1} \sin(x \sin\theta) d\theta$$

et de plus le même argument dit que $J^{(k+1)}$ est continue. On procède de même si $k = 2m + 1$ est impair pour conclure que J_0 est C^∞ avec la formule annoncée pour les dérivées.

(4) D'après les formules ci-dessus pour $j = 0$ et $j = 2$, on a

$$x(J_0''(x) + J_0(x)) = \frac{x}{\pi} \int_0^\pi (1 - (\sin\theta)^2) \cos(x \sin\theta) d\theta = \frac{1}{\pi} \int_0^\pi x(\cos\theta)^2 \cos(x \sin\theta) d\theta.$$

On observer que $x \cos(\theta) \cos(x \sin\theta)$ est la dérivée *par rapport à* θ de $\sin(x \sin\theta)$ ce qui suggère une intégration par parties. Comme les fonctions intégrées sont bornées et continues, il est effectivement possible d'utiliser la formule connue pour les intégrales de Riemann et de la transcrire pour l'intégrale de Lebesgue puisque ces deux notions coïncident dans ce cadre.

On trouve donc pour tout $x \in \mathbf{R}$

$$\begin{aligned} x(J_0''(x)+J_0(x)) &= \frac{1}{\pi}\Big[\cos(\theta)\sin(x\sin\theta)\Big]_0^\pi - \frac{1}{\pi}\int_0^\pi (-\sin\theta)\sin(x\sin\theta)d\theta \\ &= \frac{1}{\pi}\int_0^\pi (\sin\theta)\sin(x\sin\theta)d\theta = -J_0'(x) \end{aligned}$$

d'après la question précédente.

(5) Pour calculer le développement de Taylor il suffit d'évaluer $J_0^{(k)}(0)$ pour $k \geqslant 0$. On remarque tout d'abord que pour k impair, dans la formule de la question 3, la fonction à intégrer est un multiple de $\sin(x\sin\theta)=0$ pour $x=0$, donc $J^{(2m+1)}(0)=0$ pour tout $m \geqslant 0$.

Plutôt que de calculer $J^{(2m)}(0)$ par intégration (ce qui est très simple également), on peut utiliser l'équation différentielle, ou bien encore écrire

$$\begin{aligned} J_0^{(2m)}(0) &= \frac{(-1)^m}{\pi}\int_0^\pi (\sin\theta)^{2m}d\theta = \frac{(-1)^m}{\pi}\int_0^\pi \Big\{\frac{1}{2i}(e^{i\theta}-e^{-i\theta})\Big\}^{2m}d\theta \\ &= \frac{1}{2^{2m}\pi}\int_0^\pi \sum_{0\leqslant j\leqslant 2m}(-1)^j C_{2m}^j e^{2i(j-m)\theta}d\theta \\ &= \frac{(-1)^m}{2^{2m}}C_{2m}^m = \frac{(-1)^m(2m)!}{2^{2m}(m!)^2}. \end{aligned}$$

Le développement de Taylor est donc

$$\sum_{m\geqslant 0}\frac{J_0^{(2m)}(0)}{(2m)!}x^{2m} = \sum_{m\geqslant 0}\frac{(-1)^m}{(m!)^2}(x/2)^{2m}.$$

<u>On admet</u> que $J_0(x)$ est somme de sa série de Taylor pour tout x.

(6) Pour tout $a>0$, la fonction $g(x)=e^{-ax}$ est intégrable sur $[0,+\infty[$ puisque, par exemple, $x^2e^{-ax}\to 0$. Comme J_0 est bornée la fonction $f(x)=g(x)J_0(x)$ est également intégrable sur $[0,+\infty[$.

On a d'après ce qui précède

$$\int_0^{+\infty} e^{-ax}J_0(x)dx = \int_0^{+\infty} e^{-ax}\Big(\sum_{m\geqslant 0}\frac{(-1)^m}{(m!)^2}(x/2)^{2m}\Big)dx$$

On désire bien sûr échanger la somme et l'intégrale. Comme La série n'est pas à termes positifs, c'est le théorème de convergence dominée qui est l'instrument ad-hoc. Pour l'appliquer il suffit d'après le cours de démontrer que la série est normalement convergence dans L^1, c'est à dire que

$$\sum_{m\geqslant 0}\int_0^{+\infty}\Big|\frac{(-1)^m}{(m!)^2}(x/2)^m e^{-ax}\Big|dx < +\infty$$

Or l'intégrale intérieure vaut

$$\frac{1}{2^{2m}(m!)^2}\int_0^{+\infty} x^{2m}e^{-ax}dx = \frac{1}{a}\frac{(2m)!}{(2a)^{2m}(m!)^2}$$

selon la première indication. (Le calcul de l'intégrale peut se faire en changeant de variable $y=ax$ puis en intégrant par partie et par récurrence; tout cela est justifié par comparaison de l'intégrale de Riemann et de celle de Lebesgue dans ce cas où les intégrales, bien que portant sur $[0,+\infty[$, sont absolument convergentes).

Il s'agit donc d'étudier la série dont cette expression est le terme général u_m. Vu sa forme multiplicative on calcule u_{m+1}/u_m :

$$\frac{u_{m+1}}{u_m} = \frac{(2m+1)(2m+2)}{4m^2}\frac{1}{a^2} \to \frac{1}{a^2}$$

de sorte que si $a > 1$, on a $\lim u_{m+1}/u_m < 1$, ce qui classiquement implique la converge absolue de la série en question. On a donc le droit d'échanger limite et intégrale, et il vient

$$\int_0^{+\infty} e^{-ax} J_0(x)dx = \sum_{m\geqslant 0} \frac{(-1)^m}{2^{2m}(m!)^2} \int_0^{+\infty} e^{-ax} x^{2m} dx = \frac{1}{a} \sum_{m\geqslant 0} \frac{(-1)^m (2m)!}{(2a)^{2m}(m!)^2}$$

d'après le même calcul que précédemment.

On reconnaît dans la dernière série le développement du binôme indiqué dans l'énoncé

$$\sum_{m\geqslant 0} \frac{(-1)^m (2m)!}{(2a)^{2m}(m!)^2} = \frac{1}{\sqrt{1+a^{-2}}}$$

donc

$$\int_0^{+\infty} e^{-ax} J_0(x)dx = \frac{1}{a}\frac{1}{\sqrt{1+a^{-2}}} = \frac{1}{\sqrt{a^2+1}}.$$

Problème 3. (1) Soit $f \in L^p([0,+\infty[)$ avec $p \in [1,+\infty]$ et

$$F(x) = \int_{[0,x]} f(t)d\lambda(t).$$

Bien que f ne soit pas forcément intégrable sur $[0,+\infty[$ (si $p \neq 1$), on a $f \in L^p([0,x])$ par restriction, or $[0,x]$ est de mesure finie et on sait que dans le cas où $\mu(X) < +\infty$, les espaces L^p sont « décroissants » donc $L^p([0,x]) \subset L^1([0,x])$. Cela montre donc que f est intégrable sur $[0,x]$, c'est à dire que F est définie. De plus il est évident que $F(0) = 0$.

Soit $[a,b] \subset [0,+\infty[$ un intervalle compact. On écrit que pour $a \leqslant x \leqslant b$ on a

$$F(x) = \int_{[0,a]} f(t)d\lambda(t) + \int_{[a,b]} h(x,t)d\lambda(t) \text{ avec } h(x,t) = \chi_{[a,x]}(t)f(t).$$

Le premier terme est une constante. Quand au second, montrons que c'est une fonction continue sur $[a,b]$, de sorte que F est continue sur $[a,b]$ également.

Pour tout $x_0 \in [a,b]$, la fonction $x \mapsto h(x,t)$ est continue en x_0 pour tout $t \neq x_0$ fixé, en particulier pour presque tout t. Comme de plus

$$|h(x,t)| \leqslant |f(t)|$$

et que $|f|$ est intégrable sur $[a,b]$ d'après ce qui précède, le résultat du cours implique la continuité en x_0. Comme x_0 est quelconque, on déduit donc que F est continue sur $[a,b]$. Comme a et b sont arbitraires dans $[0,+\infty[$, F est donc continue sur $[0,+\infty[$.

(2) On suppose désormais que $f \in L^p([0,+\infty[)$ avec $p \in]1,+\infty[$. On peut écrire

$$\int_{[x,x+h]} |f(t)|^p d\lambda(t) = F_p(x+h) - F_p(x)$$

avec

$$F_p(x) = \int_{[0,x]} |f(t)|^p d\lambda(t).$$

Il s'agit donc d'une fonction du type de la question précédente, $g = |f|^p$ remplaçant f. Comme $g \in L^1([0,+\infty[)$ par hypothèse, la question (1) montre que F_p est continue sur $[0,+\infty[$. Donc $F_p(x+h) \to F_p(x)$ quand $h \to 0$, c'est à dire que

$$\lim_{h\to 0} \int_{[x,x+h]} |f(t)|^p d\lambda(t) = 0.$$

(3) On a

$$F(x+h) - F(x) = \int_{[x,x+h]} f(t)d\lambda(t),$$

et on peut majorer cela à l'aide de l'inégalité de Hölder

$$\begin{aligned}|F(x+h)-F(x)| &\leqslant \int_{[x,x+h]} |f(t)|d\lambda(t) \leqslant \Big(\int_{[x,x+h]} |f(t)|^p d\lambda(t)\Big)^{1/p}\Big(\int_{[x,x+h]} d\lambda(t)\Big)^{1/q}\\ &= h^{1/q}\Big(\int_{[x,x+h]} |f(t)|^p d\lambda(t)\Big)^{1/p}\end{aligned}$$

donc la question précédente montre que

$$\lim_{h\to 0} \frac{F(x+h)-F(x)}{h^{1/q}} = 0.$$

En particulier en prenant $x=0$, comme $F(x)=0$, il vient

$$\lim_{x\to 0} \frac{F(x)}{x^{1/q}} = 0.$$

(4) Soit $\varepsilon > 0$. Considérons la mesure $\mu = |f|^p d\lambda$ sur $[0,+\infty[$. Il s'agit d'une mesure finie puisque $f \in L^p([0,+\infty[)$. Soit $B_n = [n,+\infty[$ pour $n \geqslant 0$. On a

$$\bigcap_{n\geqslant 0} B_n = \emptyset,$$

donc d'après le cours (en utilisant le fait que μ est finie !)

$$\lim_{n\to+\infty} \mu(B_n) = 0 \text{ c'est à dire } \lim_{n\to+\infty} \int_{[n,+\infty[} |f|^p d\lambda = 0.$$

En particulier puisque $\varepsilon^p > 0$, il existe N tel que si $n \geqslant N$, on a

$$0 \leqslant \int_{[n,+\infty[} |f|^p d\lambda \leqslant \varepsilon^p$$

et on peut prendre $A_\varepsilon = N$.

(5) Pour $x \geqslant A_\varepsilon$ on a

$$F(x) = F(A_\varepsilon) + \int_{[A_\varepsilon,x]} f(t)d\lambda(t)$$

et donc par Hölder encore pour tout $x \geqslant A_\varepsilon$ on a

$$\begin{aligned}|F(x)| &\leqslant |F(A_\varepsilon)| + (x-A_\varepsilon)^{1/q}\Big(\int_{[A_\varepsilon,x]} |f|^p d\lambda\Big)^{1/p}\\ &\leqslant |F(A_\varepsilon)| + x^{1/q}\Big(\int_{[A_\varepsilon,+\infty[} |f|^p d\lambda\Big)^{1/p} \leqslant |F(A_\varepsilon)| + \varepsilon x^{1/q}\end{aligned}$$

d'après le choix de A_ε. Puisque $\varepsilon > 0$ est arbitraire et $F(A_\varepsilon)$ est une constante, il vient

$$\lim_{x\to+\infty} \frac{F(x)}{x^{1/q}} = 0.$$

(6) Si l'on prend $f=1$, on a $f \in L^\infty$, et $F(x) = x$. Comme ici $q=1$, on voit que

$$\lim_{x\to 0} \frac{F(x)}{x^{1/q}} = 1 = \lim_{x\to+\infty} \frac{F(x)}{x^{1/q}}.$$

Examen – Bordeaux, Mai 2003

Énoncé

Problème 4. (Questions de cours)

(1) Donner la définition du produit de convolution h de deux fonctions f et $g \in L^1(\mathbf{R}^d)$. Dans quel espace h existe-t-elle ? Quelle borne peut-on donner pour sa norme ?

(2) Donner la définition de la convergence en probabilité d'une suite de variables aléatoires X_n et énoncer la loi faible des grands nombres.

Problème 5. On désigne par $L^p(\mathbf{R}^d)$ les espaces L^p par rapport à la mesure de Lebesgue.

(1) Soit $p \in [1, +\infty[$ et $f \in L^p(\mathbf{R}^d)$. Montrer que

$$T_f(\varphi) = \int_{\mathbf{R}^d} f(x)\varphi(x)dx = 0$$

existe pour toute $\varphi \in C_c^\infty(\mathbf{R}^d)$.

(2) Si $T_f(\varphi) = 0$ pour toute $\varphi \in C_c^\infty(\mathbf{R}^d)$, montrer que $f = 0$.

Soit $\alpha = (\alpha_1, \ldots, \alpha_d)$ un multi-indice. On note $|\alpha| = \alpha_1 + \cdots + \alpha_d$ et

$$\partial_\alpha f = \frac{\partial^{|\alpha|} f}{\partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}}.$$

On dit que $f \in L^2(\mathbf{R}^d)$ a une dérivée partielle d'ordre α au sens faible dans $L^2(\mathbf{R}^d)$ s'il existe $g \in L^2(\mathbf{R}^d)$ telle que

$$\int_{\mathbf{R}^d} g(x)\varphi(x)dx = (-1)^{|\alpha|} \int_{\mathbf{R}^d} f(x)\partial_\alpha\varphi(x)dx$$

pour toute $\varphi \in C_c^\infty(\mathbf{R}^d)$.

(3) Montrer que la fonction g, si elle existe, est unique.

Lorsque g existe, on dit que c'est la dérivée partielle d'ordre α de f au sens faible et on écrit

$$g = \tilde{\partial}_\alpha f.$$

(4) Montrer que si $f \in C_c^k(\mathbf{R}^d)$ alors f admet des dérivée au sens faible d'ordre α pour tout α tel que $|\alpha| \leqslant k$ et que

$$\tilde{\partial}_\alpha f = \partial_\alpha f.$$

On note H^1 l'ensemble des fonctions $f \in L^2(\mathbf{R}^d)$ telles que pour $1 \leqslant i \leqslant d$, la dérivée partielle par rapport à x_i (i.e. d'ordre $\alpha_i = (0, \ldots, 0, 1, 0, \ldots, 0)$) au sens faible existe dans $L^2(\mathbf{R}^d)$.

(6) Montrer que H^1 est un sous-espace vectoriel de $L^2(\mathbf{R}^d)$ contenant $C_c^\infty(\mathbf{R}^d)$ et que $H^1 \neq L^2(\mathbf{R}^d)$.

(7) Pour $f \in H^1$, soit

$$\|f\|_{H_1}^2 = \int_{\mathbf{R}^d} |f(x)|^2 dx + \sum_{1 \leqslant i \leqslant d} \int_{\mathbf{R}^d} |\tilde{\partial}_{\alpha_i} f(x)|^2 dx.$$

Montrer que $\|f\|_{H_1}$ est une norme sur H^1 provenant d'un produit scalaire que l'on explicitera.

(8) Soit (f_n) une suite de Cauchy dans H^1. Montrer que f_n converge dans $L^2(\mathbf{R}^d)$. Soit f sa limite.

(9) Montrer que pour tout i, f admet une dérivée partielle au sens faible par rapport à x_i et que

$$\tilde{\partial}_{\alpha_i} f = \lim_{n\to+\infty} \tilde{\partial}_{\alpha_i} f_n$$

En déduire que f_n converge vers f dans H^1 et que H^1 est un espace complet. [Indication : Montrer que les suites $(\tilde{\partial}_{\alpha_i} f_n)$ sont également de Cauchy dans L^2.]

(10) Montrer que $C_c^\infty(\mathbf{R}^d)$ est dense dans H^1.

On considère maintenant $d = 2$ et l'espace H^1 correspondant. On identifie $\mathbf{R}$ avec $\{(x, 0)\} \subset \mathbf{R}^2$ et on écrit ∂_x et ∂_y pour les dérivées partielles habituelles.

(11) Soit $f \in C_c^\infty(\mathbf{R}^2)$. Soit $g(x) = f(x, 0)$ pour $x \in \mathbf{R}$. Montrer que

$$\int_{\mathbf{R}} |g(x)|^2 dx \leqslant 2\Big(\int_{\mathbf{R}^2} |\partial_y f(x,y)|^2 dxdy\Big)\Big(\int_{\mathbf{R}^2} |f(x,y)|^2 dxdy\Big)$$

[Indication : Montrer pour tout x que

$$|g(x)|^2 = -\int_0^{+\infty} \partial_y |f(x,y)|^2 dy,$$

calculer la dérivée et intégrer sur x.]

(12) En déduire que $\|g\|_2 \leqslant \|f\|_{H^1}$ puis qu'il existe un opérateur linéaire continu

$$\gamma\, :\, H^1(\mathbf{R}^2) \to L^2(\mathbf{R})$$

tel que $\gamma(f)(x) = f(x, 0)$ si $f \in C_c^\infty(\mathbf{R}^2)$.

Pourquoi cela est-il surprenant ?

Problème 6. Soit (Y_n) une suite de variables aléatoires indépendantes réelles de même loi μ sur $\mathbf{R}$ telle que

$$\mu(\{-1\}) = \mu(\{1\}) = \frac{1}{2} \quad \text{(loi de Bernoulli)}.$$

(1) Montrer que $E(X_n) = 0$, $V(X_n) = 1$ pour tout n, et que $E(e^{uX_n}) = \cosh(u)$ pour tout $u \in \mathbf{R}$.

(2) En déduire que $E(e^{uX_n}) \leqslant e^{u^2/2}$ pour tout $u \in \mathbf{R}$ et $n \geqslant 1$. [Indication : Développer $\cosh(u)$ et $e^{u^2/2}$ en série de Taylor et comparer les coefficients.]

(3) On pose $S_0 = 0$ et $S_n = X_1 + \cdots + X_n$. Montrer que pour tout $a > 0$, $n \geqslant 1$ et $u > 0$ on a

$$P(S_n > a) \leqslant e^{nu^2/2 - ua}$$

[Indication : Remarquer que pour tout $u > 0$ on a $\{S_n > a\} = \{e^{uS_n} > e^{ua}\}$ et estimer la probabilité de ce dernier événement à l'aide de l'inégalité de Tchebychev.]

(4) En déduire que

$$P(S_n > a) \leqslant e^{-a^2/(2n)} \text{ et } P(|S_n| > a) \leqslant 2e^{-a^2/(2n)}.$$

[Indication : Choisir u convenablement.]

(5) Soit A_n une suite d'événements. Montrer que

$$\{\omega \text{ appartient à une infinité de } A_n\} = \bigcap_{n\geqslant 1}\bigcup_{k\geqslant n} A_k.$$

En déduire que si $\sum P(A_n) < +\infty$, alors

$$P(\{\omega \text{ appartient à une infinité de } A_n\}) = 0$$

(6) Soit $c > 1$. Montrer que

$$|S_n| \leqslant c\sqrt{2n\log n}$$

pour tout $n \geqslant 1$ sauf pour un nombre fini d'entre eux. [Indication : Appliquer la question précédente.]

Corrigé

Problème 1. (Questions de cours)

(1) Si f et $g \in L^1(\mathbf{R}^d)$, le produit de convolution h existe dans $L^1(\mathbf{R}^d)$ et est donné par la formule

$$h(x) = f \star g(x) = \int_{\mathbf{R}^d} f(x-t)g(t)dt$$

avec $\|h\|_1 \leqslant \|f\|_1 \|g\|_1$.

(2) On dit qu'une suite de variables aléatoires X_n converge en probabilité vers X si pour tout $\varepsilon > 0$ on a

$$\lim_{n \to +\infty} P(|X_n - X| > \varepsilon) = 0.$$

La loi faible des grands nombres dit que si $X_n \in L^1$ et si (X_n) est une suite de variables aléatoires indépendantes et de même loi μ, alors X_n converge en probabilité vers la variable aléatoire constante égale à l'espérance des X_n.

Problème 2. (1) QUESTION DE COURS. Soit $\varphi_n \in C_c^\infty(\mathbf{R}^d)$ est une suite de Dirac et $f \in L^2(\mathbf{R}^d)$. On a pour presque tout x

$$f \star \varphi_n(x) - f(x) = \int_{\mathbf{R}^d} (f(x-t) - f(x))\varphi_n(t)dt$$

pour tout $n \geqslant 1$. D'après l'inégalité de Cauchy-Schwarz, on a pour presque tout x

$$|f \star \varphi_n(x) - f(x)|^2 \leqslant \int_{\mathbf{R}^d} |f(x-t) - f(x)|^2 \varphi_n(t)dt.$$

Soit $\eta > 0$ et $U_\eta = \{x \mid \|x\| > \eta\}$ et C_η le complémentaire. On a en intégrant

$$\|f \star \varphi_n - f\|_2^2 \leqslant S_\eta + T_\eta$$

où

$$S_\eta = \int_{\mathbf{R}^d} \int_{C_\eta} |f(x-t) - f(x)|^2 \varphi_n(t)dtdx$$

$$T_\eta = \int_{\mathbf{R}^d} \int_{U_\eta} |f(x-t) - f(x)|^2 \varphi_n(t)dtdx.$$

Par Tonelli, il vient

$$S_\eta = \int_{C_\eta} \Big(\int_{\mathbf{R}^d} |f(x-t) - f(x)|^2 dx \Big) \varphi_n(t)dt$$

et l'intégrale intérieure est $< \varepsilon$ pour tout η assez petit d'après un résultat du cours. Comme φ_n est de norme L^1 constante, cela donne $S_\eta \leqslant \varepsilon$ pour tout η assez petit.

Pour T_η, on écrit $|f(x-t) - f(x)|^2 \leqslant 2(|f(x-t)|^2 + |f(x)|^2)$, donc par Tonelli encore et invariance de la mesure de Lebesgue par translation il vient

$$T_\eta \leqslant 4\|f\|_2^2 \int_{U_\eta} \varphi_n(t)dt \to 0 \text{ quand } n \to +\infty.$$

Toute cela montre que $f \star \varphi_n$ converge vers f dans $L^2(\mathbf{R}^d)$.

(2) Soit $f \in L^2(\mathbf{R}^d)$ et $\varphi \in C_c^\infty(\mathbf{R}^d)$. On a $C_c^\infty(\mathbf{R}^d) \subset L^2(\mathbf{R}^d)$ donc la fonction $g = f\varphi$ est produit de deux fonctions L^2, elle est par conséquent dans L^1 avec $\|g\|_1 \leqslant \|f\|_2 \|\varphi\|_2$.

On suppose maintenant que $\int f\varphi = 0$ pour toute $\varphi \in C_c^\infty(\mathbf{R}^d)$. On sait que $C_c^\infty(\mathbf{R}^d)$ est dense dans $L^2(\mathbf{R}^d)$, donc il existe une suite φ_n telle que $\varphi_n \in C_c^\infty(\mathbf{R}^d)$ et $\varphi_n \to \bar{f}$ dans L^2. D'après l'inégalité ci-dessus, on a alors $f\varphi_n \to |f|^2$ dans $L^1(\mathbf{R}^d)$. Par conséquent

$$0 = \int f(x)\varphi_n(x)dx \to \int |f(x)|^2 dx = \|f\|_2,$$

ce qui montre que $f = 0$ dans $L^2(\mathbf{R}^d)$.

(3) Soit $f \in L^2(\mathbf{R}^d)$, α un multi-indice et g_1, $g_2 \in L^2(\mathbf{R}^d)$ des dérivées d'ordre α de f au sens faible. Il vient pour toute $\varphi \in C_c^\infty(\mathbf{R}^d)$

$$\int_{\mathbf{R}^d} g_1(x)\varphi(x)dx = (-1)^{|\alpha|}\int_{\mathbf{R}^d} f(x)\partial_\alpha\varphi(x)dx = \int_{\mathbf{R}^d} g_2(x)\varphi(x)dx$$

par définition, donc si $g = g_1 - g_2 \in L^2(\mathbf{R}^d)$ on a

$$\int_{\mathbf{R}^d} g(x)\varphi(x)dx = 0$$

pour toute φ, ce qui implique que $g = 0$ dans $L^2(\mathbf{R}^d)$ d'après (2).

(4) Soit $f \in C_c^k(\mathbf{R}^d)$, α un multi-indice tel que $|\alpha| \leqslant k$ et $g = \partial_\alpha f$. On a donc $g \in C_c(\mathbf{R}^d)$. On veut montrer que

$$\int_{\mathbf{R}^d} g(x)\varphi(x)dx = (-1)^{|\alpha|}\int_{\mathbf{R}^d} f(x)\partial_\alpha\varphi(x)dx$$

pour $\varphi \in C_c^\infty(\mathbf{R}^d)$. Lorsque $d = 1$ et $\alpha = (1)$, cela s'écrit ($g = f'$ alors)

$$\int_{\mathbf{R}} f'(x)\varphi(x)dx = -\int_{\mathbf{R}} f(x)\varphi'(x)dx$$

ce qui est simplement la formule d'intégration par partie (puisque les limites de φ en $\pm\infty$ sont nulles donc le terme « tout intégré » disparaît). Le cas général s'en déduit par récurrence sur l'ordre de α (il suffit de traiter le cas où α est une dérivée partielle d'ordre 1, par rapport à x_i disons, et en utilisant le théorème de Fubini, cela se ramène au cas de dimension 1, en intégrant par rapport à x_i d'abord et en faisant là l'intégration par partie). Rédaction soignée laissée au lecteur...

Problème 3. Soit (Ω, Σ, P) un espace probabilisé et (X_n) une suite de variables aléatoires indépendantes réelles de même loi μ sur $\mathbf{R}$ telle que

$$\mu(\{-1\}) = \mu(\{1\}) = \frac{1}{2} \quad \text{(loi de Bernoulli)}.$$

(1) Pour toute fonction f on a $E(f(X_n)) = \int f(x)d\mu(x) = \frac{1}{2}(f(-1)+f(1))$. Pour $f(x) = x$, cela donne $E(X_n) = \frac{1}{2}(-1+1) = 0$, pour $f(x) = (x - E(X_n))^2 = x^2$, cela donne $V(X_n) = \frac{1}{2}((-1)^2 + 1^2) = 1$, et enfin pour $f(x) = e^{ux}$, cela donne

$$E(e^{uX_n}) = \frac{1}{2}(e^{-u} + e^u) = \cosh(u).$$

(2) Les développements de Taylor de $e^{u^2/2}$ et $\cosh(u)$ sont, respectivement :

$$e^{u^2/2} = \sum_{k\geqslant 0}\frac{1}{k!}\Big(\frac{u^2}{2}\Big)^k$$

$$\cosh(u) = \sum_{k\geqslant 0}\frac{1}{k!}\frac{1+(-1)^k}{2}u^k = \sum_{k\geqslant 0}\frac{1}{(2k)!}u^{2k}.$$

Ce sont deux séries à termes positifs, et on a

$$(2k)! = 2k(2k-1)\cdots 2\cdot 1 = 2^k k!(2k-1)(2k-3)\cdots 1 \geqslant 2^k k!$$

pour tout $k \geqslant 0$, donc par comparaison il vient $E(e^{uX_n}) = \cosh(u) \leqslant e^{u^2/2}$.

(3) On pose $S_0 = 0$ et $S_n = X_1 + \cdots + X_n$. Soit $a > 0$, $n \geqslant 1$ et $u > 0$. Puisque la fonction $x \mapsto e^{xu}$ est strictement croissante, on a

$$\{S_n > a\} = \{e^{uS_n} > e^{ua}\}$$

et d'après l'inégalité de Markov, on a

$$P(S_n > a) = P(e^{uS_n} > e^{ua}) \leqslant e^{-ua}E(e^{uS_n}) = e^{-ua}E(e^{X_1}\cdots e^{uX_n}).$$

Puisque les X_i sont indépendantes, l'espérance du produit des e^{X_i} est le produit des espérances, et d'après la formule du (1) puis l'inégalité du (2), cela fournit

$$P(S_n > a) \leqslant e^{-ua} E(e^{uX_1})^n = e^{-ua} \cosh(u)^n \leqslant e^{nu^2/2-ua}.$$

(4) Pour $a > 0$ fixé, l'inégalité ci-dessus est valide pour tout choix de $u > 0$. On cherche donc u qui minimise le majorant $e^{nu^2/2-ua}$ pour avoir la meilleure majoration possible. La fonction $h(u) = nu^2/2 - ua$ a comme dérivée $h'(u) = nu - a$, négative pour $u \leqslant a/n$ et positive pour $u \geqslant a/n$. Le minimum de h est donc $h(a/n) = \exp(\frac{a^2}{2n} - \frac{a^2}{n}) = \exp(-\frac{a^2}{2n})$ et ce choix de u donc l'inégalité du (3) donne bien

$$P(S_n > a) \leqslant e^{-a^2/(2n)}.$$

Enfin on a la réunion disjointe pour $a > 0$:

$$\{|S_n| > a\} = \{S_n > a\} \cup \{-S_n > a\}.$$

Si on remarque que $-S_n = (-X_1) + \cdots + (-X_n) = Y_1 + \cdots + Y_n$ avec $Y_i = -X_i$, et que les (Y_n) sont des variables aléatoires indépendantes (car les X_n le sont) et que leur loi est encore μ (car les probabilités $\mu(\{-1\})$ et $\mu(\{1\})$ sont égales), on voit que Y_n peut « remplacer » X_n dans les arguments ci-dessus, donc $P(-S_n > a) = P(S_n > a)$, et l'inégalité du (4) donne donc (au vue de la réunion disjointe ci-dessus)

$$P(|S_n| > a) \leqslant 2e^{-a^2/(2n)}.$$

(5) Soit A_n une suite d'événements et A l'événement « ω appartient à une infinité de A_n ». Un $\omega \in \Omega$ appartient à A si et seulement si pour tout $n \geqslant 1$, on peut trouver un $k \geqslant n$ tel que $\omega \in A_k$. En termes d'ensembles, cela s'écrit exactement

$$A = \bigcap_{n\geqslant 1} \bigcup_{k\geqslant n} A_k.$$

Par monotonie de la mesure, on déduit de cela que pour tout $n \geqslant 1$ on a

$$P(A) \leqslant P\Big(\bigcup_{k\geqslant n} A_k\Big) \leqslant \sum_{k\geqslant n} P(A_k).$$

Si la série $\sum P(A_n)$ converge, la suite de ses restes tend vers 0, et donc en faisant $n \to +\infty$ dans l'inégalité ci-dessus on trouve $P(A) = 0$.

(6) Soit $c > 1$. On considère les événements $A_n = \{|S_n| \leqslant c\sqrt{2n \log n}\}$. On veut montrer que la probabilité que ω appartienne à une infinité de A_n est nulle. D'après (6) il suffit de vérifier que

$$\sum_{n\geqslant 1} P(A_n) < +\infty.$$

Mais par (4) avec $a = c\sqrt{2n \log n}$ (cette valeur de a dépend de n, mais cela est parfaitement permis...), il vient

$$\sum_{n\geqslant 1} P(A_n) \leqslant 2 \sum_{n\geqslant 1} \exp\Big(-\frac{2nc^2(\log n)}{2n}\Big) = 2 \sum_{n\geqslant 1} n^{-c^2}.$$

Comme $c^2 > 1$, il s'agit là effectivement d'une série convergente.

(7) Pour $a > 0$, n fixé et $0 \leqslant i \leqslant n$, on pose

$$A_0 = \{S_1 < a, \ldots, S_n < a\}, \quad A_i = \{S_i \geqslant a \text{ et } S_j < a \text{ si } 0 \leqslant j < i\} \text{ pour } 1 \leqslant i \leqslant n.$$

Il est clair que les A_i sont disjoints et recouvrent Ω : en effet, A_i, $1 \leqslant i \leqslant n$, correspond aux ω pour lesquels i est le premier indice tel que $S_j \geqslant a$, ce qui détermine donc i à partir de ω, et A_0 contient les ω « oubliés », c'est à dire ceux pour lesquels $S_j < a$ pour $1 \leqslant j \leqslant n$. De même cela donne

$$\bigcup_{1\leqslant i\leqslant n} A_i = \{\sup_{1\leqslant i\leqslant n} S_i \geqslant a\}.$$

(8) Soit $R_i = X_{i+1} + \cdots + X_n$. En écrivant que $1 = \sum \chi_{A_i}$, on a

$$E(e^{uS_n}) = \sum_{i=0}^{n} E(e^{S_n}\chi_{A_i}).$$

Par positivité, ceci est supérieur à l'intégrale en omettant A_0, et comme $S_n = S_i + R_n$ par définition, on trouve aussitôt

$$E(e^{uS_n}) \geqslant \sum_{1 \leqslant i \leqslant n} E((e^{uS_i}\chi_{A_i})e^{uR_i}).$$

(9) Pour i fixé, on remarque que S_i et A_i sont définis seulement en fonction de X_j, $1 \leqslant j \leqslant i$, et que R_i est définie seulement en fonction de X_k, $i+1 \leqslant k \leqslant n$. Puisque les (X_n) sont indépendantes les deux variables $e^{S_i}\chi_{A_i}$ et e^{R_i} sont donc indépendantes (fonctions de deux « blocs » de variables indépendantes). Donc

$$E((e^{uS_i}\chi_{A_i})e^{uR_i}) = E(e^{uS_i}\chi_{A_i})E(e^{uR_i}).$$

Le second terme, toujours par indépendance, est $E(e^{uR_i}) = E(e^{uX_1})^{n-i} = \cosh(u)^{n-i} \geqslant 1$. Quand au premier, sur A_i on a $S_i \geqslant a$, et donc par positivité $E(e^{uS_i}\chi_{A_i}) \geqslant P(A_i)e^{ua}$.

Faisant la somme sur $1 \leqslant i \leqslant n$, l'inégalité du (8) implique alors

$$E(e^{uS_n}) \geqslant e^{ua} \sum_{1 \leqslant i \leqslant n} P(A_i).$$

(10) On a par construction

$$\{\sup_{1 \leqslant i \leqslant n} S_i \geqslant a)\} = \bigcup_{1 \leqslant i \leqslant n} A_i = \sum_{1 \leqslant i \leqslant n} P(A_i) \leqslant e^{-ua}E(e^{uS_n}) \leqslant e^{nu^2/2-ua}$$

(cf. (3)). En choisissant u comme précédemment, cela donne

$$P(\sup_{1 \leqslant i \leqslant n} S_i > a) \leqslant P(\sup_{1 \leqslant i \leqslant n} S_i \geqslant a) \leqslant e^{-\frac{a^2}{2n}}.$$

(le facteur 2 dans l'énoncé, bien que correct, n'est pas du tout nécessaire...)

Examen de rattrapage – Bordeaux, Septembre 2003

Énoncé

Problème 1. (Questions de cours)

(1) Énoncer le théorème de Tonnelli et le théorème de Fubini.

(2) Énoncer le théorème de convergence monotone.

Problème 2.

Les questions 6 à 11 du problème ne dépendent que du résultat de la question 4.

On rappelle que la transformée de Fourier d'une fonction f intégrable (par rapport à la mesure de Lebesgue) sur $\mathbf{R}$ est donnée par

$$\mathcal{F}f(\xi) = \hat{f}(\xi) = \int_{\mathbf{R}} f(x)e(-x\xi)dx$$

pour $\xi \in \mathbf{R}$, où $e(z) = \exp(2i\pi z)$ pour $z \in \mathbf{C}$.

(1) Soit $f : \mathbf{R} \to \mathbf{C}$ une fonction intégrable, de classe C^1, telle que $f' \in L^1(\mathbf{R})$ et

$$\lim_{|x| \to +\infty} f(x) = 0.$$

Montrer que $\mathcal{F}(f')(\xi) = 2i\pi\xi(\mathcal{F}f)(\xi)$ pour tout $\xi \in \mathbf{R}$.

(2) Soit $f : \mathbf{R} \to \mathbf{C}$ une fonction intégrable. On note $f_1(x) = xf(x)$. Si f_1 est intégrable, montrer que $\hat{f}$ est de classe C^1 et exprimer $(\mathcal{F}f)'$ en fonction de $\mathcal{F}f_1$.

(3) On note $\mathcal{S}$ l'espace vectoriel des fonctions $f : \mathbf{R} \to \mathbf{C}$ telles que f est C^∞ et de plus pour tout entiers $k \geqslant 0$ et tout $n \geqslant 1$ on a

$$\lim_{|x| \to +\infty} x^n f^{(k)}(x) = 0.$$

Montrer que $\mathcal{S} \subset L^p(\mathbf{R})$ pour tout $p \geqslant 1$, y compris $p = \infty$. Montrer que $\mathcal{S} \neq 0$ et justifier rapidement que $\mathcal{S}$ est dense dans $L^p(\mathbf{R})$ pour $p \neq +\infty$ par rapport à la norme L^p.

(4) Montrer que si $f \in \mathcal{S}$, alors $\hat{f} \in \mathcal{S}$.

(5) Soit $g \in \mathcal{S}$ et $\lambda > 0$. Montrer qu'il existe une unique fonction $f \in \mathcal{S}$ telle que

$$-f'' + 4\pi^2\lambda f = g.$$

[Indication : Si f existe, déterminer quelle devrait être sa transformée de Fourier puis utiliser la formule d'inversion de Fourier.]

Les questions qui suivent ne dépendent que de la question (4).

(6) Soient $f, g \in \mathcal{S}$ et $\hat{f}$, $\hat{g}$ les transformées de Fourier. Montrer que

$$\int_{\mathbf{R}} f(x)\hat{g}(x)dx = \int_{\mathbf{R}} \hat{f}(x)g(x)dx$$

(on justifiera l'existence de ces intégrales).

(7) En déduire que pour $f \in \mathcal{S}$ on a

$$\int_{\mathbf{R}} |f(x)|^2 dx = \int_{\mathbf{R}} |\hat{f}(\xi)|^2 d\xi.$$

(8) Pour une fonction $f \in L^1(\mathbf{R})$ on note

$$R(f) = \inf\Big\{R \geqslant 0 \mid \int_{[-R/2,R/2]} |f(x)|dx \geqslant \frac{1}{2}\int_{\mathbf{R}} |f(x)|dx\Big\}.$$

Montrer que cette borne inférieure existe (et est $< +\infty$) et est atteinte, c'est à dire que

$$\int_{[-R(f)/2,R(f)/2]} |f(x)|dx \geqslant \frac{1}{2}\int_{\mathbf{R}} |f(x)|dx$$

(9) Soit $f \in \mathcal{S}$. Montrer que pour tout $\xi \in \mathbf{R}$ on a

$$|\hat{f}(\xi)| \leqslant 2\sqrt{R(|f|)}\Big(\int_{\mathbf{R}} |f(x)|^2 dx\Big)^{1/2}.$$

(10) Montrer que

$$\int_{\mathbf{R}} |\hat{f}(\xi)|^2 d\xi \leqslant 2R(|\hat{f}|^2)\|\hat{f}\|_\infty^2.$$

(11) En déduire que pour $f \in \mathcal{S}$ non nulle on a

$$R(|f|)R(|\hat{f}|^2) \geqslant \frac{1}{8}.$$

Corrigé

Problème 1. (1) Soient $(X, \mathcal{M}, \mu)$ et $(Y, \mathcal{N}, \nu)$ des espaces mesurés σ-finis. On considère $(X \times Y, \mu \otimes \nu)$.

Théorème de Tonnelli : soit $f : X \times Y \to [0, +\infty]$ mesurable, alors les fonctions positives

$$x \mapsto \int_Y f(x,y)d\nu(y) \text{ et } y \mapsto \int_X f(x,y)d\mu(x)$$

sont mesurables pour $\mathcal{M}$ et $\mathcal{N}$ respectivement, et on a

$$\int_{X\times Y} fd(\mu\otimes\nu) = \int_X \Big(\int_Y f(x,y)d\nu(y)\Big)d\mu(x) = \int_Y \Big(\int_X f(x,y)d\mu(x)\Big)d\nu(y).$$

Théorème de Fubini : soit $f : X \times Y \to \mathbf{C}$ telle que $f \in L^1(\mu \otimes \nu)$, alors pour presque tout $y \in Y$, la fonction $t^y(f) : x \mapsto f(x,y)$ est μ-intégrable, pour presque tout $x \in X$, la fonction $t_x(f) : y \mapsto f(x,y)$ est ν-intégrable, les fonctions

$$x \mapsto \int_Y f(x,y)d\nu(y) \text{ et } y \mapsto \int_X f(x,y)d\mu(x)$$

sont respectivement ν-intégrables et μ-intégrables, et

$$\int_{X\times Y} fd(\mu\otimes\nu) = \int_X \Big(\int_Y f(x,y)d\nu(y)\Big)d\mu(x) = \int_Y \Big(\int_X f(x,y)d\mu(x)\Big)d\nu(y).$$

(2) Soit $(X, \mathcal{M}, \mu)$ un espace mesurable et $f_n \geqslant 0$, $n \geqslant 1$, une suite de fonctions mesurables positives telles que $f_n \leqslant f_{n+1}$. Alors la fonction $f = \lim f_n = \sup f_n$ est mesurable, et

$$\int_X f(x)d\mu(x) = \lim_{n\to+\infty} \int_X f_n(x)d\mu(x).$$

Problème 2. (1) Puisque $f' \in L^1$ on peut calculer sa transformée de Fourier en intégrant par partie. Le terme tout intégré est nul car $\lim f(x) = 0$ à l'infini, et comme la dérivée de $e(-x\xi)$ par rapport à x est $-2i\pi\xi e(-x\xi)$, on trouve

$$\mathcal{F}f'(\xi) = \int_{\mathbf{R}} f'(x)e(-x\xi)dx = 2i\pi\xi \int_{\mathbf{R}} f(x)e(-x\xi)dx = 2i\pi\xi\mathcal{F}f(\xi).$$

(2) On veut dériver $\hat{f}$ sous le signe somme. La fonction intégrée est $h(x,\xi) = f(x)e(-x\xi)$. Pour tout x fixée, c'est une fonction dérivable, et même C^∞, de ξ, avec

$$\frac{d}{d\xi}h(x,\xi) = -2i\pi xf(x)e(-x\xi) = -2i\pi f_1(x)e(-x\xi).$$

On a

$$|h(x,\xi)| \leqslant |f(x)|, \text{ et } \left|\frac{d}{d\xi}h(x,\xi)\right| \leqslant 2\pi|f_1(x)|,$$

qui sont indépendants de ξ. Puisque $f \in L^1$ et $f_1 \in L^1$ par hypothèse, le théorème de dérivabilité des fonctions définies par des intégrales implique que $\mathcal{F}f$ est dérivable et

$$\mathcal{F}f'(\xi) = -2i\pi \int_{\mathbf{R}} f_1(x)e(-x\xi)dx = -2i\pi\mathcal{F}f_1(\xi).$$

On sait de plus que la transformée de Fourier d'une fonction intégrable est continue, et donc $\mathcal{F}f$ est de classe C^1.

(3) Il est clair que $\mathcal{S}$ est un espace vectoriel. Comme f est continue et qu'en prenant $n=1$, $k=0$, on a $xf(x) \to 0$ quand $x \to \pm\infty$, on voit que f est bornée sur $\mathbf{R}$, donc $f \in L^\infty(\mathbf{R})$. Toujours avec $k=0$, mais $n=2$, on a $x^2f(x) \to 0$, en particulier $x^2f(x)$ est bornée, et donc il existe une constante $C \geqslant 0$ telle que

$$|f(x)| \leqslant \frac{C}{x^2}$$

pour $|x| \geqslant 1$. Soit $p > 1$. La fonction $|f|^p$ restreinte à $[-1,1]$ est continue donc intégrable, et pour $|x| \geqslant 1$ on a $|f(x)|^p \leqslant C^px^{-2p}$. Comme $2p > 1$, la fonction $x \mapsto x^{-2p}$ est intégrable, donc $f \in L^p(\mathbf{R})$.

On a $\mathcal{S} \neq 0$ parce que, par exemple, l'espace $C_c^\infty(\mathbf{R})$ des fonctions C^∞ à support compact est évidemment inclus dans $\mathcal{S}$, et l'on sait que cet espace est non seulement $\neq 0$, mais même dense dans $L^p(\mathbf{R})$ pour $p \neq +\infty$. L'espace $\mathcal{S}$ qui le contient l'est donc aussi. (Comme fonction $f \neq 0$ dans $\mathcal{S}$, on peut prendre la gaussienne habituelle $f(x) = e^{-\pi x^2}$.)

(4) Soit $f \in \mathcal{S}$ et $g = \hat{f}$ sa transformée de Fourier. Montrons d'abord que g est C^∞, et remarquons déjà que $f_n(x) = x^nf(x)$ est intégrable pour tout $n \geqslant 0$, puisque continue et que $x^2f_n(x) \to 0$ quand $x \to \pm\infty$.

D'après la question (2), g est C^1, et de plus puisque $f_2 = x^2f(x)$ est intégrable, on peut appliquer de nouveau ce résultat à $g' = -2i\pi\mathcal{F}f_1$, et ainsi de suite par récurrence, on voit que g est C^∞.

Puisque f' est intégrable (prendre $n=2$, $k=1$) et que $f \to 0$ à l'infini, la question (1) implique que $\mathcal{F}f'(\xi) = -2i\pi\xi g(\xi)$. D'après le théorème de Lebesgue-Riemann on a $\lim \mathcal{F}f'(\xi) = 0$ et donc

$$\lim \xi g(\xi) = 0.$$

Si on applique cela à f' au lieu de f ($f' \to 0$ et $f'' \in L^1$, donc c'est possible), on trouve

$$\lim \xi^2 g(\xi) = 0.$$

Encore par récurrence, il vient pour $n \geqslant 1$

$$\lim \xi^n g(\xi) = 0.$$

Remarquons maintenant que $g' = -2i\pi\mathcal{F}f_1$ et que $f_1(x) = xf(x)$ appartient également à $\mathcal{S}$. (Les dérivées de f_1 sont des combinaisons linéaires de produits $x^nf^{(k)}(x)$). On applique ce qui précède à g' et il vient pour $n \geqslant 1$

$$\lim \xi^n g'(\xi) = 0.$$

Finalement une dernière récurrence montre que $g \in \mathcal{S}$.

(5) Soit $g \in \mathcal{S}$ et $\lambda > 0$. Montrons d'abord l'unicité de $f \in \mathcal{S}$ telle que

$$-f'' + 4\pi^2\lambda f = g.$$

En prenant la transformée de Fourier à l'aide de la question (1) appliquée deux fois, on trouve pour $\xi \in \mathbf{R}$

$$4\pi^2(\xi^2+\lambda)\hat{f}(\xi) = \hat{g}(\xi).$$

Comme $\lambda > 0$ on a $\xi^2 + \lambda > 0$ pour tout $\xi \in \mathbf{R}$, et ainsi

$$\hat{f}(\xi) = \frac{\hat{g}(\xi)}{4\pi^2(\xi^2+\lambda)}$$

est déterminée de façon unique. Puisque la transformée de Fourier est injective pour $f \in L^1$ (à cause de la formule d'inversion de Fourier), l'unicité de f en découle.

Pour l'existence, il suffit de remarquer que la fonction

$$h : \xi \mapsto \frac{\hat{g}(\xi)}{4\pi^2(\xi^2+\lambda)}$$

est intégrable puisque $\hat{g} \in \mathcal{S}$ et $|(\xi^2+\lambda)^{-1}| \leqslant \lambda^{-1}$. Donc h admet une transformée de Fourier conjuguée, que l'on appelle f :

$$f = \bar{\mathcal{F}}(h)(x) = \mathcal{F}h(-x) = \int_{\mathbf{R}} h(\xi)e(x\xi)d\xi.$$

Il faut montrer que $f \in \mathcal{S}$ et que f est solution de l'équation demandée.

Puisque $g \in \mathcal{S}$ et $\lambda > 0$, on vérifie facilement en dérivant que $h \in \mathcal{S}$, donc $f \in \mathcal{S}$. De plus d'après la formule d'inversion de Fourier on a

$$\mathcal{F}f(\xi) = h(\xi) = \frac{\hat{g}(\xi)}{4\pi^2(\xi^2+\lambda)},$$

donc d'après la question (1)

$$\mathcal{F}(-f'' + 4\pi^2 f) = \mathcal{F}g.$$

L'injectivité de la transformée de Fourier permet de conclure.

(6) Soient f et g dans $\mathcal{S}$. Donc $\hat{f}$ et $\hat{g}$ sont aussi dans $\mathcal{S}$. Toutes ces fonctions sont donc dans L^2, et par l'inégalité de Cauchy-Schwarz, $f\hat{g}$ et $g\hat{f}$ sont intégrables.

On calcule

$$\begin{aligned}\int_{\mathbf{R}} f(x)\hat{g}(x)dx &= \int_{\mathbf{R}} f(x)\Big(\int_{\mathbf{R}} g(y)e(-yx)dy\Big)dx \\ &= \int_{\mathbf{R}\times\mathbf{R}} f(x)g(y)e(-xy)dxdy\end{aligned}$$

grâce au théorème de Fubini qui est trivialement applicable puisque f et g sont intégrables. Comme cette expression est symétrique en f et g, on a la formule demandée.

(7) Prenant g telle que $\hat{g} = \bar{f}$. Le premier terme devient

$$\int_{\mathbf{R}} |f(x)|^2 dx.$$

D'après la formule d'inversion de Fourier, une telle fonction g existe effectivement et est donnée par $g = \bar{\mathcal{F}}(\hat{g}) = \bar{\mathcal{F}}(\bar{f}) = \overline{\mathcal{F}(f)}$ (en particulier c'est bien une fonction de $\mathcal{S}$), donc le second terme vaut

$$\int_{\mathbf{R}} |\hat{f}(\xi)|^2 d\xi.$$

D'où l'égalité demandée.

(8) Les intervalles $[-n/2, n/2]$ pour $n \geqslant 1$ sont croissants et d'union $\mathbf{R}$. Donc pour toute fonction intégrable f on a

$$\lim_{n\to+\infty} \int_{[-n/2,n/2]} |f(x)|dx = \int_{\mathbf{R}} |f(x)|dx.$$

En particulier il existe $n \geqslant 1$ tel que
$$\int_{[-n/2,n/2]} |f(x)|dx \geqslant \frac{1}{2}\int_{\mathbf{R}} |f(x)|dx,$$
ce qui donne $R(f) \leqslant n < +\infty$.

Par définition de la borne inférieure il existe une suite décroissante $R_n \geqslant R(f)$ telle que $\lim R_n = R(f)$. Pour tout n on a
$$\int_{[-R_n/2,R_n/2]} |f(x)|dx \geqslant \frac{1}{2}\int_{\mathbf{R}} |f(x)|dx.$$
Lorsque $n \to +\infty$, les fonctions $g_n(x) = \chi_{[-R_n/2,R_n/2]}(x)|f(x)|$ convergent point par point vers $g(x) = \chi_{[-R(f)/2,R(f)/2]}|f(x)|$. De plus $|g_n(x)| \leqslant |f(x)|$ qui est intégrable. D'après le théorème de convergence dominée on a donc
$$\lim_{n\to+\infty}\int_{\mathbf{R}} g_n(x)dx = \lim_{n\to+\infty}\int_{[-R_n/2,R_n/2]} |f(x)|dx = \int_{[-R(f)/2,R(f)/2]} |f(x)|dx.$$
Comme chaque terme de la suite est $\geqslant \frac{1}{2}\int |f(x)|dx$, on en déduit que
$$\int_{[-R(f)/2,R(f)/2]} |f(x)|dx \geqslant \frac{1}{2}\int_{\mathbf{R}} |f(x)|dx,$$
c'est à dire que la borne inférieure est atteinte.

(9) Soit $f \in \mathcal{S}$. On a $\hat{f} \in \mathcal{S}$ et
$$\hat{f}(\xi) = \int_{\mathbf{R}} f(x)e(-x\xi)dx.$$
Majorant il vient
$$|\hat{f}(\xi)| \leqslant \int_{\mathbf{R}} |f(x)|dx,$$
donc par la question précédente
$$|\hat{f}(\xi)| \leqslant 2\int_{[-R(f)/2,R(f)/2]} |f(x)|dx,$$
et par Cauchy-Schwarz
$$|\hat{f}(\xi)| \leqslant 2\int_{[-R(f)/2,R(f)/2]} |f(x)|dx \leqslant 2\Big(\int_{[-R(f)/2,R(f)/2]} dx\Big)^{1/2}\Big(\int_{[-R(f)/2,R(f)/2]} |f(x)|^2dx\Big)^{1/2}.$$
On peut majorer la dernière intégrale en l'étendant à $\mathbf{R}$ et on trouve
$$|\hat{f}(\xi)| \leqslant 2\sqrt{R(f)}\Big(\int_{\mathbf{R}} |f(x)|^2dx\Big)^{1/2} = 2\sqrt{R(|f|)}\Big(\int_{\mathbf{R}} |f(x)|^2dx\Big)^{1/2}.$$

(10) On applique de nouveau la question (8) à $|\hat{f}|^2$ cette fois. On a donc
$$\int_{\mathbf{R}} |\hat{f}(\xi)|^2d\xi \leqslant 2\int_{[-R(|\hat{f}|^2)/2,R(|\hat{f}|^2)/2]} |\hat{f}(\xi)|^2d\xi.$$
La fonction intégrée peut être majorée par $\|\hat{f}\|_\infty^2$ et donc
$$\int_{\mathbf{R}} |\hat{f}(\xi)|^2d\xi \leqslant 2R(|\hat{f}|^2)\|\hat{f}\|_\infty^2.$$

(11) Soit $f \neq 0$. En prenant la borne supérieure dans la question (9) et en prenant le carré on trouve
$$\|\hat{f}\|_\infty^2 \leqslant 4R(|f|)\int_{\mathbf{R}} |f(x)|^2dx.$$
D'après (7) cela donne
$$\|\hat{f}\|_\infty^2 \leqslant 4R(|f|)\int_{\mathbf{R}} |\hat{f}(\xi)|^2d\xi$$

et d'après (10), enfin
$$\|\hat{f}\|_\infty^2 \leqslant 8R(|f|)R(|\hat{f}|^2)\|\hat{f}\|_\infty^2.$$
Si $f \neq 0$, on a $\hat{f} \neq 0$ donc on peut diviser par la norme $\|\hat{f}\|_\infty^2$ et obtenir
$$R(|f|)R(|\hat{f}|^2) \geqslant \frac{1}{8}.$$

Bibliographie

[A] W. Appel : *Mathématiques pour la physique*, H&K 2002.

[Bi] P. Billingsley : *Probability and measure*, 3ème édition, Wiley 1995.

[B] N. Boccara : *Intégration*, Ellipses 1995.

[F] G. Folland : *Real Analysis*, Wiley 1984.

[M] P. Malliavin : *Intégration et probabilités, analyse de Fourier et analyse spectrale*, Masson 1982.

[R] W. Rudin : *Analyse réelle et complexe*, 3ème édition, Dunod 1998.