Thierry Gallouët
Raphaèle Herbin

2e édition

Mesure, intégration, probabilités

Cours avec plus de 300 exercices corrigés

ellipses

Chapitre 1

Motivation et objectifs

Nous commençons par donner ici un aperçu des motivations de la théorie de l'intégration, en montrant d'abord les limitations de l'intégrale des fonctions continues (sur un intervalle compact de $\mathbb{R}$). L'intégrale de Riemann possède essentiellement les mêmes limitations.

1.1 Intégrale des fonctions continues

Nous présentons ici quelques rappels sur l'intégrale des fonctions continues sur un intervalle compact de $\mathbb{R}$.

Nous nous limitons dans ce paragraphe à l'étude des fonctions définies sur l'intervalle $[0,1]$ à valeurs dans $\mathbb{R}$, par souci de simplicité des notations. Il va de soi que les notions introduites se généralisent à une intervalle $[a,b]$, $a,b \in \mathbb{R}$. Nous allons en fait définir l'intégrale des fonctions dites réglées, au sens de la définition suivante.

Définition 1.1 (Fonction en escalier, fonction réglée) *Soit g une fonction de l'intervalle $[0,1] \subset \mathbb{R}$ dans $\mathbb{R}$.*

— *On dit que g est une* fonction en escalier *s'il existe $p \in \mathbb{N}^*$, une famille $(x_i)_{i\in\{0,\ldots,p\}}$, avec : $x_0 = 0$, $x_i < x_{i+1}$, pour tout $i \in \{0,\ldots,p-1\}$, $x_p = 1$, et une famille $(a_i)_{i\in\{0,\ldots,p-1\}} \subset \mathbb{R}$ tels que*

$$g(x) = a_i, \quad \forall x \in]x_i, x_{i+1}[, \quad \forall i \in \{0,\ldots,p-1\}.$$

— *On dit que g est une fonction* réglée *si elle est limite uniforme d'une suite de fonctions en escalier.*

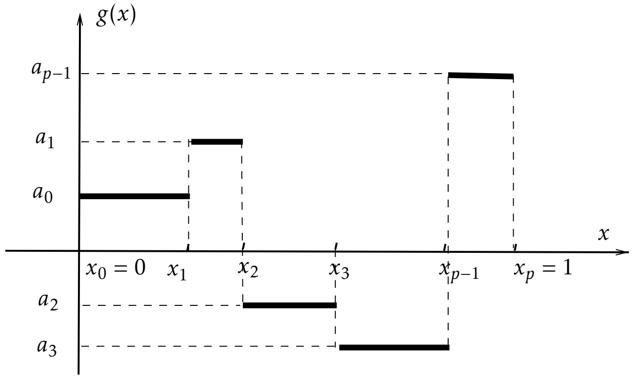

FIGURE 1.1 – Fonction en escalier

La construction de l'intégrale des fonctions réglées permet de définir l'intégrale des fonctions continues car toute fonction continue est réglée. Cette construction peut être décomposée en trois étapes, que nous esquissons ici et qui sont étudiées en détail dans l'exercice 1.2 :

1. *Mesurer les intervalles de* $[0,1]$. Pour $0 \leq \alpha \leq \beta \leq 1$, on pose $m(]\alpha,\beta[) = \beta - \alpha$.
2. *Intégrer les fonctions en escalier.* Avec les notations de cette définition, l'intégrale d'une fonction en escalier est alors

$$\int_0^1 g(x)\,\mathrm{d}x = \sum_{i=0}^{p-1} a_i m(]x_i, x_{i+1}[). \tag{1.1}$$

   On montre que la définition précédente est bien cohérente, au sens où l'intégrale de g ne dépend que du choix de g et non du choix des x_i.
3. *Passer à la limite.* Soit $f : [0,1] \to \mathbb{R}$, une fonction réglée, il existe une suite $(f_n)_{n\in\mathbb{N}}$ de fonctions en escalier convergeant uniformément vers f. On pose $\mathrm{I}_n = \int_0^1 f_n(x)\,\mathrm{d}x$. On peut montrer que la suite $(\mathrm{I}_n)_{n\in\mathbb{N}}$ est de Cauchy. On pose alors

$$\int_0^1 f(x)\,\mathrm{d}x = \lim_{n\to+\infty} \mathrm{I}_n.$$

   On montre que cette définition est cohérente car $\lim_{n\to+\infty} \mathrm{I}_n$ ne dépend que de f et non du choix de la suite $(f_n)_{n\in\mathbb{N}}$.

Notons qu'on a ainsi défini l'intégrale de toute fonction continue de $[0,1]$ à valeurs dans $\mathbb{R}$, puisque toute fonction continue est réglée (voir exercice 1.2, question 2a).

Remarque 1.2 (Intégrale sur un espace de Banach) Un avantage de la construction qu'on vient d'évoquer est qu'elle utilise la convergence des suites de Cauchy, donc le caractère complet de $\mathbb{R}$, et non pas la structure d'ordre de $\mathbb{R}$. Elle permet ainsi de définir sans travail supplémentaire l'intégrale des fonctions continues de $[0,1]$ (ou d'un intervalle compact de $\mathbb{R}$) dans E, où E est un espace de Banach [1] sur $\mathbb{R}$ ou $\mathbb{C}$ (la méthode de construction utilise la structure d'espace de Banach de E, et il peut ne pas y avoir de relation d'ordre sur E). On remplace donc l'espace d'arrivée $\mathbb{R}$ des fonctions qu'on intègre par un espace de Banach E.

Nous verrons dans le paragraphe suivant qu'un inconvénient majeur de cette intégrale est que l'ensemble des fonctions continues, muni de la norme induite par cette intégrale, n'est pas un espace vectoriel normé complet.

Pour définir l'intégrale de fonctions plus générales que les fonctions continues, une méthode classique est la méthode de Riemann, qui permet de définir l'intégrale des fonctions dites *Riemann-intégrables*, (voir l'exercice 5.2) mais ne permet pas non plus d'introduire un espace fonctionnel complet ; de plus, contrairement à l'intégrale des fonctions continues évoquée plus haut et détaillée dans l'exercice 1.2, la construction de l'intégrale de Riemann nécessite la structure d'ordre sur $\mathbb{R}$ et n'a donc qu'un intérêt relatif.

C'est grâce à la méthode de Lebesgue que nous pourrons construire des espaces fonctionnels complets. Nous nous limiterons dans ce cours à l'intégrale de Lebesgue pour des fonctions prenant leurs valeurs dans $\mathbb{R}$, ce qui nous permettra d'utiliser la relation d'ordre dans $\mathbb{R}$.

Pour définir l'intégrale de Lebesgue de fonctions à valeurs dans un espace de Banach de dimension infinie, il faut un travail supplémentaire : c'est l'intégrale de Bochner. Nous référons à [3, 4] pour une introduction. Le cas où l'espace est de dimension finie reste simple car on est alors amené à considérer un nombre fini d'intégrales à valeurs dans $\mathbb{R}$.

Remarque 1.3 (Remarque de terminologie) Dans tout ce document, on utilisera indifférement le terme « fonction » et le terme « application ». Une application (ou une fonction) f de D dans E est la donnée pour tout $x \in D$ de son image par f, notée $f(x)$. (Le domaine de définition de f est donc ici l'ensemble D.) Lorsque nous parlons d'une fonction de $\mathbb{R}$ de $\mathbb{R}$, le domaine de définition de f est donc $\mathbb{R}$ tout entier.

---

1. Un espace de Banach est un espace vectoriel normé complet.

1.2 Insuffisance de l'intégrale des fonctions continues

Dans ce paragraphe, on note E l'ensemble $C([0,1],\mathbb{R})$ des fonctions continues de $[0,1]$ dans $\mathbb{R}$. On a défini dans le paragraphe précédent l'intégrale $\int_0^1 f(x)\,dx$ pour tout $f \in E$ (car l'ensemble des fonctions continues est contenu dans l'ensemble des fonctions réglées).

Théorèmes de convergence.

Un inconvénient important de la théorie de l'intégration exposée ci-dessus est que les théorèmes classiques de convergence pour cette théorie sont peu efficaces. À vrai dire, le seul théorème simple est un résultat de convergence de l'intégrale sous hypothèse de convergence uniforme d'une suite de fonctions. Rappelons tout d'abord les notions de convergence simple et uniforme des suites de fonctions.

Définition 1.4 (Convergence simple et uniforme) *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions de* E,

- $(f_n)_{n\in\mathbb{N}}$ *converge simplement vers f lorsque $n \to +\infty$ si :*

$$\forall \varepsilon > 0, \forall x \in [0,1], \exists N(\varepsilon, x); n \geq N(\varepsilon, x) \Rightarrow |f_n(x) - f(x)| \leq \varepsilon;$$

- $(f_n)_{n\in\mathbb{N}}$ *converge uniformément vers f si :*

$$\forall \varepsilon > 0, \exists N(\varepsilon); n \geq N(\varepsilon), x \in [0,1] \Rightarrow |f_n(x) - f(x)| \leq \varepsilon.$$

Pour la convergence simple, l'entier N peut dépendre de x, alors que pour la convergence uniforme, il ne dépend que de ε, et pas de x. La suite $(f_n)_{n\in\mathbb{N}}$ d'éléments de E définie par $f_n(x) = \frac{x}{n}$ tend simplement et uniformément (sur $[0,1]$) vers 0. On donne à l'exercice 1.1 un exemple de suite qui converge simplement mais pas uniformément.

On rappelle maintenant le théorème classique de convergence de l'intégrale des fonctions continues :

Théorème 1.5 (Convergence de l'intégrale des fonctions continues)

Soient $(f_n)_{n\in\mathbb{N}} \subset E$ et $f \in E$. On a alors :

$$[f_n \longrightarrow f \text{ uniformément lorsque } n \to +\infty] \Longrightarrow \left[\int_0^1 f_n(x)\,dx \to \int_0^1 f(x)\,dx \text{ lorsque } n \to +\infty\right].$$

Ce théorème est assez faible, au sens où l'hypothèse de convergence uniforme est une hypothèse forte. Une conséquence de la théorie de l'intégrale de Lebesgue est le théorème suivant (beaucoup plus fort que le précédent, car il ne demande pas d'hypothèse de convergence uniforme) :

Théorème 1.6 (Convergence dominée de l'intégrale des fonctions continues)
Soient $(f_n)_{n\in\mathbb{N}} \subset \mathrm{E}$*, et* $f \in \mathrm{E}$*. On suppose que*

$$|f_n(x)| \leq \mathrm{C},\ \forall x \in [0,1],\ \forall n \in \mathbb{N},$$

où $\mathrm{C} \in \mathbb{R}_+$ *est fixé, et que* f_n *tend simplement vers* f *quand* n *tend vers* $+\infty$*. On a alors :*

$$\int_0^1 f_n(x)\,\mathrm{d}x \to \int_0^1 f(x)\,\mathrm{d}x \text{ quand } n \to +\infty. \tag{1.2}$$

Par exemple, la suite de fonctions $(f_n)_{n\geq 2}$ définie par

$$f_n(x) = \begin{cases} nx & \text{pour } x \in [0, \frac{1}{n}], \\ n(\frac{2}{n} - x) & \text{pour } x \in]\frac{1}{n}, \frac{2}{n}], \\ 0 & \text{pour } x \in]\frac{2}{n}, 1]. \end{cases} \tag{1.3}$$

(voir figure 1.2) converge simplement mais non uniformément. Elle est dominée par 1, et d'après le théorème 1.6, elle converge. On peut le vérifier directement, car l'intégrale de f_n est facile à calculer et vaut $\frac{1}{n}$. On considère maintenant la suite de fonctions $(g_n)_{n\geq 2}$ définie par $g_n(x) = nf_n(x)$. Cette suite converge toujours simplement vers 0, mais non uniformément, et elle n'est plus dominée. Et de fait, g_n tend simplement vers 0 mais son intégrale vaut 1 et ne tend donc pas vers 0.

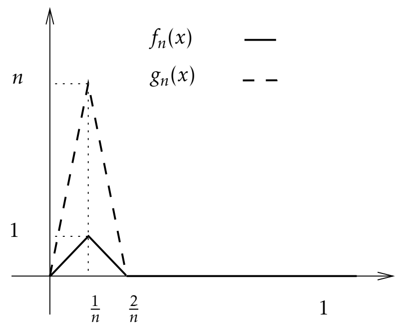

FIGURE 1.2 – Les fonctions f_n et g_n

Le théorème 1.6 est une conséquence immédiate du théorème de convergence dominée de Lebesgue, que nous verrons au chapitre 4, il peut être démontré directement, sans utiliser la théorie de l'intégrale de Lebesgue, mais cela est difficile : nous donnons une technique possible à l'exercice 1.10 ; l'idée essentielle est un passage à la limite sur des suites croissantes de fonctions, qui se retrouve également dans la construction de l'intégrale de Lebesgue. Dans l'exercice 1.10, on introduit des suites croissantes de fonctions continues, et on utilise l'intégrale des fonctions continues.

En revanche, Lebesgue utilise des suites croissantes de fonctions étagées (voir définition 3.5), ce qui permet également d'utiliser la définition de la mesure et donc de s'affranchir de la notion de topologie (voir définition 2.8) sur l'espace de départ pour construire l'intégrale.

Espaces non complets.

Pour $f \in \mathrm{E}$ on pose (en remarquant que $|f| \in \mathrm{E}$ et $f^2 \in \mathrm{E}$) :

$$\mathrm{N}_1(f) = \int_0^1 |f(x)| \mathrm{d}x \text{ et } \mathrm{N}_2(f) = \Big(\int_0^1 (f(x))^2 dx\Big)^{\frac{1}{2}}.$$

Les applications N_1 et N_2 sont des normes sur E (voir l'exercice 1.6). Malheureusement l'espace E muni de la norme N_1 (ou de la norme N_2) n'est pas vraiment intéressant en pratique, en particulier parce que cet espace n'est pas complet (c'est-à-dire qu'une suite de Cauchy n'est pas nécessairement convergente). Ce n'est pas un espace de Banach. La norme N_2 sur E est induite par un produit scalaire mais, muni de cette norme, E n'est pas un espace de Hilbert [2], voir l'exercice 1.6. En fait l'espace vectoriel des fonctions continues de $[0,1]$ dans $\mathbb{R}$ est intéressant lorsqu'il est muni de la norme de la convergence uniforme, c'est-à-dire $\|f\|_u = \sup_{x\in[0,1]} |f(x)|$, avec laquelle il est complet : c'est donc alors un espace de Banach.

Si l'on travaille avec l'ensemble des fonctions réglées plutôt que l'ensemble des fonctions continues, on n'échappe pas vraiment aux inconvénients cités précédemment (N_1 et N_2 sont d'ailleurs alors des semi–normes). On peut aussi généraliser la définition de l'intégrale ci-dessus en améliorant un peu l'étape 3 (passage à la limite), cette généralisation se fait en introduisant les sommes de Darboux , alors que l'intégrale des fonctions continues peut être définie en utilisant seulement les sommes de Riemann). On obtient ainsi la définition de l'intégrale des fonctions dites Riemann-intégrables (voir l'exercice 5.2). En fait cette généralisation est assez peu intéressante, et les inconvénients sont les mêmes que pour l'intégrale des fonctions continues (ou des fonctions réglées).

L'intégrale de Lebesgue va permettre la construction d'espaces de Banach avec les normes N_1 et N_2 (et même de Hilbert avec N_2). Dans le cas des fonctions de $[0,1]$ dans $\mathbb{R}$, ceci pourrait être fait par un procédé de complétion de l'espace E muni de la norme N_1 ou N_2 à partir des suites de Cauchy pour N_1 ou N_2 (procédé semblable à celui qui est utilisé pour construire $\mathbb{R}$ à partir des suites de Cauchy de $\mathbb{Q}$). L'intégrale de Lebesgue va permettre de construire des espaces de Banach en utilisant seulement sur l'espace de départ une structure d'espace mesuré. Cette méthode est en particulier très intéressante pour la théorie des probabilités.

2. Un espace de Hilbert est un espace de Banach dont la norme est induite par un produit scalaire.

1.3 Les probabilités

La théorie des probabilités s'est développée dans le but de modéliser les phénomènes aléatoires, c'est-à-dire de développer un formalisme mathématique pour exprimer les problèmes posés par ces phénomènes. Le terme aléatoire vient du latin *alea* qui signifie en latin jeu de dé ou jeu de hasard ; il est employé pour désigner tous les phénomènes qui semblent être dus au hasard. Il s'oppose au terme déterministe, qui s'applique aux phénomènes dont on connaît l'issue. Le mot hasard vient lui même du mot arabe al-zhar qui veut dire dés, puis par extension chance. On utilisera également le mot stochastique (du grec stokhastikos, qui vise bien) qui est un synonyme d'aléatoire. En anglais, les termes utilisés en théorie des probabilités sont random (hasard, qui vient du français randonnée !) stochastic et aleatory. Par exemple, la chute d'un corps est un phénomène déterministe : pour une position et une vitesse initiale données, on sait parfaitement quelle sera la trajectoire et la vitesse du corps soumis à son poids. Le lancer d'un dé est assimilable à la chute d'un corps, et pourtant, le résultat du lancement du dé est généralement perçu comme aléatoire : on ne sait pas avant l'expérience quel est le nombre entre 1 et 6 que l'on va obtenir, parce qu'on ne connaît pas vraiment les conditions initiales du lancement du dé (position, vitesse) et que, même si on les connaissait, on aurait du mal à calculer rapidement le résultat de ce lancement. Ainsi, de nombreux phénomènes physiques qui ont des causes déterministes sont modélisés à l'aide de modèles au moins en partie aléatoires (en météorologie par exemple). Il existe cependant des phénomènes physiques véritablement aléatoires comme l'interférence d'atomes dans un dispositif à deux fentes d'Young, et de manière plus générale, les phénomènes quantiques (voir à ce sujet le livre grand public [7]).

Une partie importante des phénomènes aléatoires est de nature discrète, c'est-à-dire qu'il existe une injection de l'ensemble des « cas possibles » dans $\mathbb{N}$. Lorsque de plus l'ensemble des cas possibles ou des « éventualités » est fini, le calcul des probabilités se ramène à des problèmes de dénombrement. Lorsque l'ensemble des éventualités est de nature infinie non-dénombrable, on aura besoin, pour définir une probabilité, de la théorie de la mesure. Les liens qui existent entre la théorie des probabilités et la théorie de la mesure et de l'intégration sont nombreux, mais malheureusement, le vocabulaire est souvent différent. Nous essaierons ici de montrer clairement les liens entre les deux théories et de donner systématiquement les termes probabilistes et analystes employés pour les mêmes notions.

1.4 Objectifs

Du point de vue de l'intégration, l'objectif est de construire une théorie de l'intégration donnant des théorèmes de convergence efficaces et de bons espaces fonctionnels, c'est-à-dire des espaces vectoriels normés complets et des espaces hilbertiens. La démarche pour construire cette théorie est décrite au chapitre 4 ; elle est voisine de celle que l'on a utilisée pour l'intégrale des fonctions réglées (ou pour l'intégrale de Riemann, cf. Exercice 5.2).

La théorie de l'intégration que nous allons ainsi obtenir contient, pour les fonctions d'un intervalle compact de $\mathbb{R}$ dans $\mathbb{R}$, la théorie de l'intégrale de Riemann (cf. Exercice 5.2) qui contient elle-même la théorie de l'intégrale des fonctions réglées (et donc la théorie de l'intégrale des fonctions continues).

Du point de vue probabiliste, l'objectif est d'introduire les notions de base et de mettre en évidence les liens entre les outils d'analyse et les outils probabilistes.

1.5 Structure du cours

Ce cours est formé de 11 chapitres (y compris ce chapitre introductif), selon le découpage suivant :

— Le chapitre 2 est une introduction à la théorie de la mesure ; on y définit en particulier l'application λ nécessaire pour mesurer les parties de $\mathbb{R}$. On y introduit aussi les premières notions de probabilités.

— Dans le chapitre 3, on introduit le concept de fonction mesurable, et son synonyme probabiliste, *i.e.* le concept de variable aléatoire, qui est une notion fondamentale pour le calcul des probabilités. On y définit les notions de convergence presque partout et son synonyme probabiliste presque sûre, et de convergence en mesure et son synonyme probabiliste convergence en probabilité.

— On définit au chapitre 4 l'intégrale sur un espace mesuré (suivant les étapes 1 à 3 définies plus haut), et l'espérance des variables aléatoires réelles en théorie des probabilités. On définit également dans ce chapitre la notion de convergence en moyenne.

— On s'intéresse au chapitre 5 aux mesures définies sur les boréliens de $\mathbb{R}$ (c'est-à-dire les parties mesurables au sens de Borel, que l'on aura définie au chapitre 2) et aux propriétés particulières de l'intégrale définies sur $\mathbb{R}$. On y étudie les lois de probabilités de densité.

— On étudie au chapitre 6 les espaces L^p, ensembles des (classes de) fonctions mesurables de puissance p–ième intégrable, et plus particulièrement l'espace L^2, qui est un espace de Hilbert. On donne des résultats de dualité et on introduit les notions de convergence faible et de convergence étroite (pour les probabilités).

— Le chapitre 7 est consacré au produits d'espaces mesurés, à l'intégration de fonctions de plusieurs variables, au produit de convolution.